Received: 25 October 2024 | Revised: 10 September 2025 | Accepted: 11 October 2025

DOI: 10.1111/bph.70254

INVITED REVIEW

UK Biobank: Transforming drug discovery and precision medicine

Jelena Bešević[1,2] | Saredo Said[3] | Reka Nagy[3] | Yalda Jamshidi[3] | Christopher D. Whelan[4] | Lauren Carson[2] | Martin K. Rutter[2,5,6] | Adam J. Lewandowski[1,2] | Mark Effingham[2] | Rory Collins[1,2] | Ben Lacey[1,2] | Naomi E. Allen[1,2]

[1]Oxford Population Health (Nuffield Department of Population Health), University of Oxford, Oxford, UK

[2]UK Biobank, Greater Manchester, UK

[3]Human Genetics Centre of Excellence, Novo Nordisk Research Centre Oxford (NNRCO), Oxford, UK

[4]Neuroscience Data Science, Johnson & Johnson Innovative Medicine, Cambridge, Massachusetts, USA

[5]Division of Diabetes, Endocrinology and Gastroenterology, School of Medical Sciences, University of Manchester, Manchester, UK

[6]Diabetes, Endocrinology and Metabolism Centre, Manchester Royal Infirmary, Manchester University NHS Foundation Trust, NIHR Manchester Biomedical Research Centre, Manchester Academic Health Science Centre, Manchester, UK

Correspondence
Ben Lacey, Oxford Population Health (Nuffield Department of Population Health), Big Data Institute Building, University of Oxford, Oxford, UK
Email: ben.lacey@ndph.ox.ac.uk

Funding information
The core funding for UK Biobank is provided by the UK Medical Research Council, Wellcome, British Heart Foundation, Cancer Research UK, and the National Institute for Health Research (grant ref 223600/Z/21/Z). The funders had no role in the design; in the collection, analysis or interpretation; in the writing of the report; or in the decision to submit the paper for publication.

UK Biobank is a large-scale, prospective study with extensive genetic and phenotypic data on half a million individuals. Volunteers, aged between 40 and 69 years, were recruited between 2006 and 2010 from the general population of the United Kingdom. At recruitment, participants completed a series of questionnaires on a range of factors (including lifestyle and medical history), physical measurements were taken and biological samples were collected for long-term storage. Large-scale assays have been undertaken (including biochemical assays, genotyping, whole exome and whole genome sequencing, as well as proteomics and metabolomics) with potential for further assays to be performed on stored samples in the future. The participants provided consent for linkage to their health-related records to identify health outcomes over time. The UK Biobank study, with its vast collection of genetic data, has enabled researchers worldwide to identify new drug targets for common diseases of middle and older age, and progress towards precision medicine. As the UK Biobank resource matures, its value to health-related research will continue to grow. Thousands of researchers worldwide are actively using UK Biobank data to improve our understanding of the prevention, diagnosis and treatment of a wide range of diseases.

KEYWORDS
drug discovery; multi-omics (genomics, proteomics, metabolomics); precision medicine; population-based cohort studies; translational pharmacology

Abbreviations: BMI, body mass index; COVID-19, coronavirus disease 2019; DEXA, dual-energy X-ray absorptiometry; DNA, deoxyribonucleic acid; GWAS, genome-wide association study; IMD, indices of multiple deprivation; LDL, low-density lipoprotein; MRI, magnetic resonance imaging; MR, Mendelian randomisation; NHS, National Health Service; NICE, National Institute for Health and Care Excellence; NMR, nuclear magnetic resonance; NULISA, nucleic acid-linked immunosorbent assay; PCR, polymerase chain reaction; pQTL, protein quantitative trait locus; PheWAS, phenome-wide association study; PRS, polygenic risk score; QRISK2, cardiovascular disease risk prediction algorithm; RAP, Research Analysis Platform; UKB-RAP, UK Biobank Research Analysis Platform; UKBDRS, UK Biobank Dementia Risk Score; WES, whole-exome sequencing; WGS, whole-genome sequencing.

1 | INTRODUCTION

The development of new pharmacological therapies is a lengthy multi-year, multi-phase process with high rates of attrition at early clinical stages (Phases 1 and 2), most commonly because a drug demonstrates unfavourable pharmacokinetics or unacceptable toxicity, or has little or no clinically meaningful efficacy (Sun et al., 2022). Drugs with human genetic support for their targets are more than twice as likely to gain approval than those without genetic support (King et al., 2019; Minikel et al., 2024). Consequently, there has been a shift in recent years towards the increasing use of human genetic data in drug discovery and development.

The identification and prioritisation of drug targets during the pre-clinical stage is a key point in the creation of pharmacological therapies that can be optimised through the interrogation of large biomedical resources. Large-scale population studies with genetic, molecular and health outcome data are uniquely suited to investigate disease-causing genetic variants and identify perturbations in biological pathways that may lead to the development of novel therapeutics.

UK Biobank is one of the world's leading resources, available to academic and commercial researchers worldwide, and contains a wealth of genetic, proteomic, metabolomic, lifestyle, environmental and health outcome data on half a million UK adult volunteers. The study's large size, breadth of data and long-term follow-up of health outcomes, coupled with its easy accessibility to researchers worldwide, have made UK Biobank a unique resource to better understand disease biology with potential implications for the discovery of novel drug targets.

This review illustrates how UK Biobank has developed over the last two decades, and how it facilitates drug discovery research and progress towards tailored therapy.

2 | THE UK BIOBANK RESOURCE

The UK Biobank is a very large, population-based prospective study established to allow detailed investigations of the genetic and non-genetic determinants of the diseases of middle and old age. At the time of its inception, it was and remains a pioneering study within the health research landscape (Sudlow et al., 2015).

Between 2006 and 2010, half a million people from across the United Kingdom were recruited to the study. All participants provided information about their lifestyle and medical history, underwent a series of physical measures and provided biological samples (blood, urine and, for a subset, saliva) for long-term storage (Downey & Peakman, 2008; Elliott et al., 2008). The participants gave consent for these samples to be used for health-related research purposes (even after incapacity or death), to be re-contacted for future assessments, and for their health to be followed over time by linking to their health-related records. Ethical approval for the study was provided by the North West Multi-Centre Research Ethics Committee as a Research Tissue Bank. The UK Biobank's independent Ethics Advisory Committee provided guidance and advice on ethical issues arising from the usage of the resource.

In the United Kingdom, most of the health care provided to the population is via the National Health Service (NHS). UK Biobank has established electronic linkage to national death and cancer registries, hospital inpatient admissions and, for ~230,000 participants, primary care records (up until 2016/2017, depending on the data provider), which includes coded data on diagnoses, symptoms, prescriptions (including prescription date, drug code, drug name and quantity, where available) and laboratory results. Efforts are underway to extend the primary care linkage to all half a million UK Biobank participants, because these data are particularly useful for the study of conditions that are unlikely to result in hospitalisation but nonetheless have significant morbidity (such as diabetes, mental health conditions and dementia). To support COVID-19 research, UK Biobank obtained linkage to the national polymerase chain reaction (PCR) test data for all participants and access to primary care records under an emergency legislation notice issued by the government of the United Kingdom to enable vital research into COVID-19. As a result, over 300 research papers have been published on the determinants and consequences of COVID-19 using UK Biobank data (Bešević et al., 2023; Pavey et al., 2022; Raisi-Estabragh et al., 2020).

3 | ENHANCING DATA COLLECTION IN UK BIOBANK

Since recruitment, UK Biobank has undertaken a series of enhancements to the resource (Figure 1). Of particular interest to researchers has been the generation of assay data for all half a million study participants from the stored biological samples. This includes over 30 key biochemistry markers, which were selected for inclusion because they are either known diagnostic markers or important risk factors for disease (e.g. **cholesterol**, sex hormones and markers of kidney and liver function). Haematology data also are available for the whole cohort, as well as leukocyte telomere length, a marker of biological ageing (Codd et al., 2022). Data on antibodies for a range of infections are currently available for 10,000 participants with plans to expand to larger numbers of participants, owing to clear evidence of their value for assessing the role of viral and other infections in the development of chronic diseases (Mentzer et al., 2022).

Importantly for drug discovery, UK Biobank possesses a wealth of genetic data. Genotyping data are available for all participants via a custom-built genotyping array with 850,000 variants directly measured and more than 90 million imputed (Bycroft et al., 2018). This dataset has had an impact on genome-wide association study (GWAS) research and the generation of polygenic risk scores (PRS) for disease risk stratification. Given the significance of genetic data to drug discovery, whole exome sequencing, which measures genetic variation within the coding regions of the genome, was funded by an industry consortium (Szustakowski et al., 2021). Moreover, whole genome sequencing data, which encompasses the entire DNA sequence of individuals, was made available on the whole cohort at the end of 2023 (Li et al., 2023), funded by government, charity and industry, highlighting the high confidence in the value of these UK Biobank

FIGURE 1 A selection of significant milestones in the conduct of UK Biobank. Note: The above timeline does not include all enhancements, data releases or significant milestones. Abbreviations: IMD (indices of multiple deprivation), NMR (nuclear magnetic resonance), PCR (polymerase chain reaction) and RAP (research analysis platform).

data to future biomedical research. Other industry-led enhancements to the resource include UK Biobank's growing 'omic' data. For example, a consortium of 13 biopharmaceutical companies funded the generation of proteomic data using the Olink Explore 3072 platform, which measures the relative concentrations of about 3000 circulating proteins across 54,000 participants (Dhindsa et al., 2023; Sun et al., 2023). Plans were announced in early 2025 to extend these proteomic measures to the full cohort, including repeated measures on the 100,000 participants who provided a second blood sample at an imaging assessment some years after their initial visit. Furthermore, metabolomics data from Nightingale Health, comprising about 250 metabolites, predominantly lipids, are available for ~300,000 participants (Julkunen et al., 2023), with data on the full cohort expected to become available to researchers in 2025.

Additional enhancements to UK Biobank have included a series of online questionnaires sent to ~330,000 participants with a known e-mail address. These questionnaires aim to collect more detailed data on exposures that were not captured at the recruitment visit (e.g. diet [Galante et al., 2016] and occupational history [Bradbury et al., 2018]) and on health outcomes that are not easily captured via healthcare records (e.g. cognitive function, mental health, pain, sleep, autism, prosopagnosia and aphantasia; Brailean et al., 2020; Davis et al., 2020; Wang et al., 2022).

Between 2012 and 2013, UK Biobank undertook a repeat assessment of 20,000 participants to allow researchers to adjust for biases because of a single baseline measurement of an exposure being a potentially inaccurate characterisation of long-term average (or 'usual') levels (UK Biobank, 2013). Between 2013 and 2016, 100,000 UK Biobank participants wore a wrist-worn accelerometer for 7 days to obtain objective measures of physical activity and sleep (Doherty et al., 2017; UK Biobank, 2016), to allow for more accurate assessment of the impact of physical activity on future disease risk. A subset of about 3000 participants also wore the accelerometers at multiple time points over the course of a year to assess the impact of seasonality on physical activity and sleep.

In 2014, UK Biobank started the world's largest multi-modal imaging assessment, aiming to image 100,000 of its participants (Littlejohns et al., 2020). The imaging visit includes magnetic resonance imaging (MRI) scans of the brain, heart and body (neck-to-knee); a carotid ultrasound scan to obtain measures of the large arteries; and a dual energy X-ray absorptiometry (DEXA) scan to measure bone density and fat distribution. The participants also completed the questionnaires, had physical measurements taken and provided biological samples at the recruitment visit. Up to 36,000 participants wore a cardiac monitor for 14 days to obtain continual measures of the electrical activity of the heart (as well as physical activity and sleep), which will provide valuable objective data on subclinical atrial fibrillation and other common cardiac arrhythmias. The sheer size of the imaging study has necessitated the development of automated pipelines to generate image-derived phenotypes (such as regional brain volumes, cardiac structure and function, and body fat measures) (Alfaro-Almagro et al., 2018; Linge et al., 2018; Liu et al., 2021; Mojtahed et al., 2019; Sorokin et al., 2022; UK Biobank, 2024; Wilman et al., 2017) that can be used by both specialists and non-specialists in imaging data to investigate the associations between the structure and function of organs and various exposures and health-related outcomes.

To facilitate understanding of the relationship between changes in organ structure and function as the cause or consequence of disease, UK Biobank started a repeat imaging study in 2019 to collect a second series of MRI scans in up to 60,000 participants. This will enable large-scale assessment of genetic and lifestyle determinants of changes in imaging-derived measures in middle-to-old age and identification of early markers related to subsequent disease risk. Details of all the data available in UK Biobank can be found on the UK Biobank Data Showcase (online).

4 | A GLOBAL PLATFORM FOR PHARMACEUTICAL RESEARCH

UK Biobank data are available for access, upon registration, to all *bona fide* researchers whether they are based in academic, commercial, governmental or charity sectors all over the world. Thousands of researchers worldwide are currently using UK Biobank data (Figure 2). To access UK Biobank data, all researchers must submit an application, detailing their research question and the potential public health impact of their proposed project. Data only applications are reviewed by the scientific team to ensure the proposed research is health

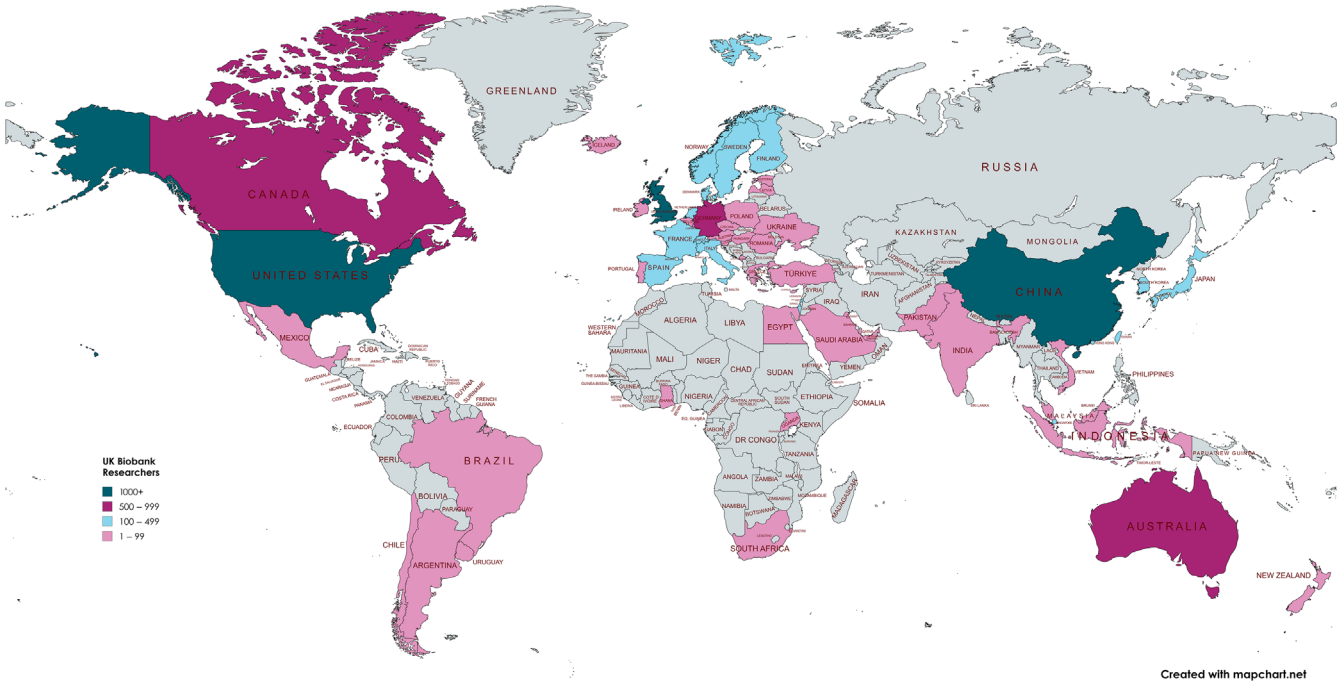

FIGURE 2 Worldwide map illustrating where UK Biobank data is being used.

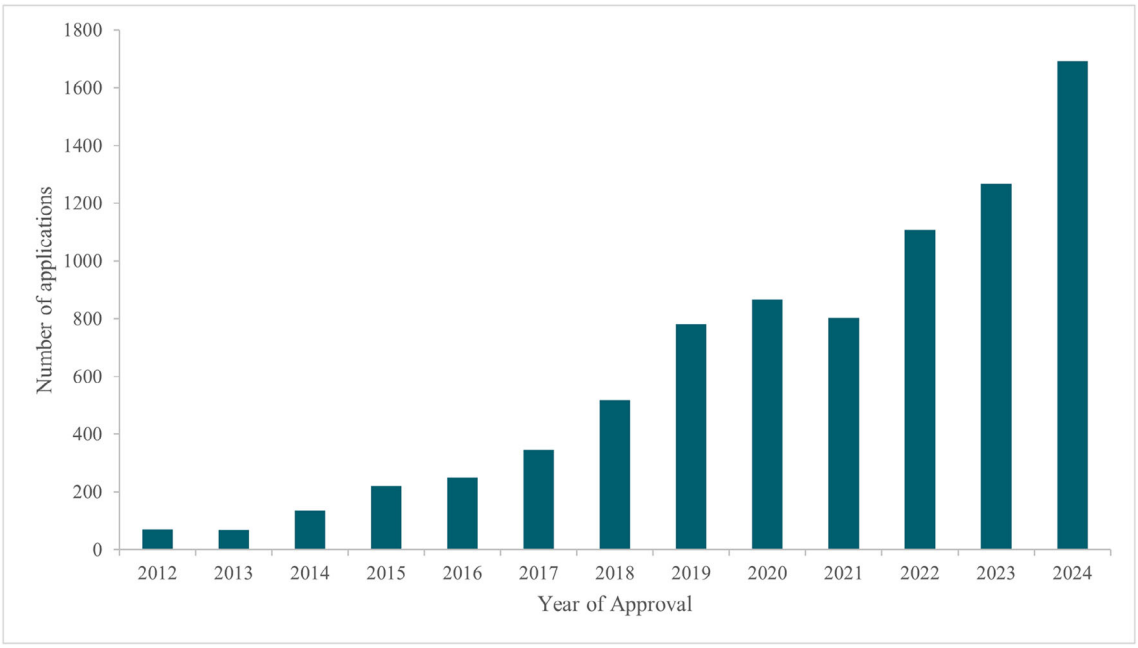

FIGURE 3 Number of applications to use UK Biobank data by year.

related and in the public interest, and applications that require further discussion are considered by the Access Committee. Because of the limited availability of biological samples and their depletable nature, access to samples is carefully considered by the Access Committee to ensure the sample utility is maximised (Conroy et al., 2019). Further details on how to apply and the registration process can be found on the UK Biobank website.

UK Biobank data became available for health-related research in 2012; since then, the number of research projects has increased year on year (Figure 3). There is no preferential access afforded to any researcher, and the use of the resource is on the same basis for all researchers (whether from academia or industry), who undergo rigorous background and international sanctions checks. Charges associated with accessing the data are in place to cover the costs of

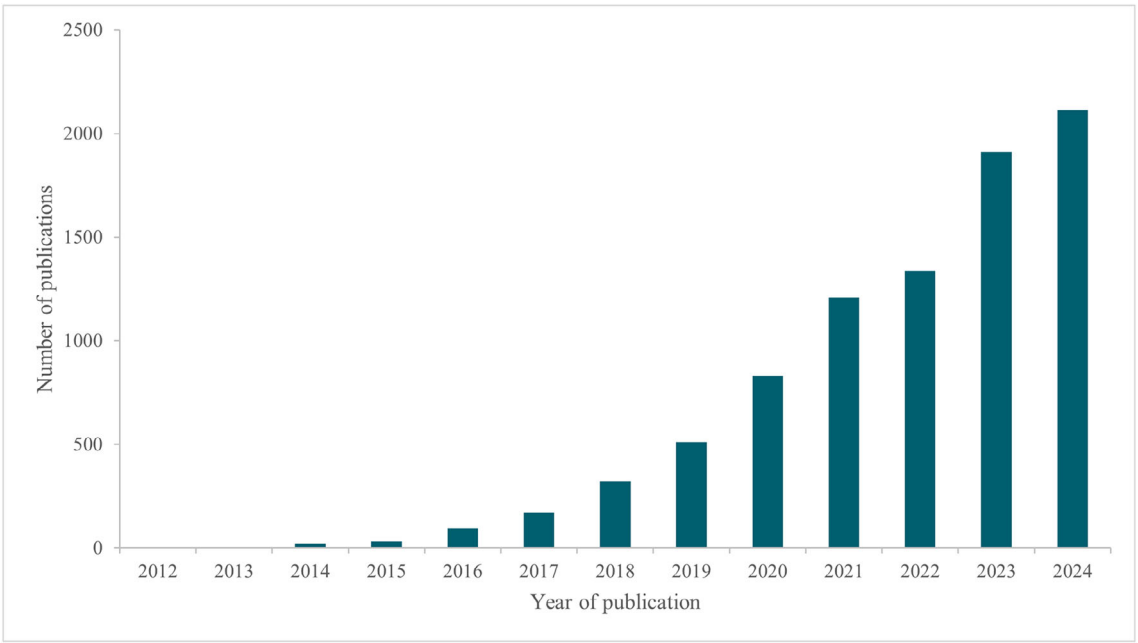

FIGURE 4 Number of publications using individual-level UK Biobank data, by year. Note: Publications using summary statistics, or those about UK Biobank, are not included in this graphic. Details of all publications can be found at https://biobank.ndph.ox.ac.uk/showcase/docs.cgi?id=2.

servicing access to the resource only, rather than the costs incurred during the setting up of the study.

UK Biobank requires researchers who use the resource to make their findings publicly available, and the number of publications resulting from use of UK Biobank data is growing every year (Figure 4). Researchers are required to return any newly generated data, including data derived from analyses of images or assay data, so that other researchers may use them. Recognising that not all researchers have the necessary facilities to process the large volumes of data (which, in any event, are too large to disseminate), in 2021 UK Biobank launched the UK Biobank Research Analysis Platform (UKB-RAP), a cloud-computing platform that brings the researcher to the data. The platform is currently delivered by DNAnexus, and the data are housed by Amazon Web Services (AWS; London). Researchers can apply for the UK Biobank Platform Credits Programme, funded by AWS, to cover compute costs associated with using the platform, if needed. In 2024, the UKB-RAP became the principal method through which researchers can access and analyse data, ensuring rapid data access, enhanced data control, and compliance with potential future governance requirements on the use of NHS health-care record data. Significant efforts are ongoing to improve the UKB-RAP's functionality and add supplementary capabilities tailored to different researcher requirements.

UK Biobank launched the Global Researcher Access Fund in 2024, supported by industry partners, to democratise worldwide access to UK Biobank data for researchers from institutes based in less wealthy countries. This fund can be used to cover the costs associated with accessing the UK Biobank resource, and the same researchers also are eligible for the UK Biobank Platform Credits Programme, meaning that, through industry support, these researchers can have all their costs to access and analyse UK Biobank data covered.

5 | UK BIOBANK DATA FOR DRUG DISCOVERY AND DEVELOPMENT

The wealth of health outcome, genetic and molecular data available in UK Biobank means that it is arguably the world's leading resource to investigate the biological pathways that underlie disease development. In particular, the scale of the genetic and proteomic data, combined with the accessibility of UK Biobank's data, has positioned the study as a leading resource for research into drug development, including drug target identification and validation, investigation of drug efficacy, and refinement of drug prescribing towards precision medicine. Although it is still too premature to see UK Biobank's real impact on drug development (given the recent availability of genome sequencing and proteomics data), the amount of research in this area suggests that drug discovery and development will accelerate significantly over the next few years based on these data.

5.1 | Genetics-driven target identification

Identifying potential drug targets using genetic data often requires large population sizes or smaller populations known to be enriched for the trait or functional genetic variant of interest. Having exome sequencing data available for half a million participants in UK Biobank has been crucial to identify potential drug targets for osteoporosis (Zhou et al., 2023), neuropsychiatric diseases (Deng et al., 2024) and liver disease (Verweij et al., 2022). UK Biobank genotyping data have contributed to several large-scale common variant GWAS meta-analysis of disease and disease relevant traits (Aragam et al., 2022; Ghouse et al., 2024). Results from such studies can inform on human

relevance of putative drug targets as well as discovery of novel drug targets. It is important to note that not all targets identified in GWAS will result in the development of suitable therapeutics as some low impact genetic variants may be associated with a disease, but their biological effect on the protein may be negligible so that it does not warrant replication. To date, rare variants account for the most widely publicised of genetics-guided targets in drug discovery (Minikel et al., 2020). For example, data from UK Biobank, along with other population biobanks, has led to the identification of genetic variants associated with body mass index (BMI), including rare protein truncating variants in ***GPR75*** (Akbari et al., 2021). Individuals with one functional copy of the *GPR75* gene because of protein truncating variants had a 1.8 kg/m^2 lower BMI and were 54% less likely to be obese than individuals with two functional *GPR75* copies (Akbari et al., 2021). As a result of this discovery, pharmaceutical companies are now developing small molecules to mimic the effects of the naturally occurring protein truncating variants, with a view to developing new drugs to promote weight loss (Regeneron Pharmaceuticals Inc, 2021).

Whereas whole-exome sequencing data can provide valuable genetic insights, they have some limitations in predicting the phenotype of some drug-metabolising enzymes. This is particularly the case for those enzymes that are highly regulated by intergenic (i.e., non-coding) variants, contain complex structural genetic variants (such as large copy number variations) or whose expression is strongly influenced by epigenetic factors (such as methylation, histone modifications and post-translational changes)—none of which can be captured with whole-exome sequencing. Data from whole-genome sequencing on all half a million participants were made available in 2023 and will provide more insight into the role of genetic variation on drug metabolism. Similarly, data from long-read sequencing on 50,000 participants will be made available in due course.

5.2 | Proteogenomic driven target identification

In addition to genetic data, proteomic data are invaluable for drug target identification and development. This is because most drugs act by targeting proteins in the human body and modulating their activity (with gene therapies being notable exceptions). Thus, while the identification of genes associated with disease is valuable to drug discovery, the successful development of the *right drug* for the *right patient* at the *right time* (Ashley, 2015; Collins & Varmus, 2015; Kosorok & Laber, 2019) will depend upon a clearer understanding of the relationship between those genes and their functional products (i.e. proteins). Furthermore, it is important to know not only the relationship between genetic variants and traits of interest but also the functional consequences of these genetic variants on target genes, including whether and how they modify gene expression, and consequently protein levels. In UK Biobank, proteomic data have been linked to genetic data to obtain protein quantitative trait loci (pQTLs)—genetic loci that are associated with the levels of circulating proteins (Sun et al., 2023). This, in turn, enables statistical genetic approaches such as colocalisation and Mendelian Randomisation (MR), which leverages genetic variants (e.g. pQTLs), to investigate the putative causal associations of an exposure (e.g. protein concentrations) on an outcome (e.g. cardiovascular disease). Where there is evidence of pQTLs associating with a disease, this could indicate potential drug targets and inform on a putative drug's mechanisms of action, thereby accelerating the drug development pipeline.

Several studies have used UK Biobank pQTL data, together with other datasets, to identify potential new drug targets. For example, MR approaches have been used in several studies to identify proteins potentially causally associated with various health outcomes. This includes five proteins (CD160, DNPH1, LAYN, LRRC37A2 and **TLR1**) potentially causally related to breast cancer (Mälarstig et al., 2023), REG1A and REG1B associated with pancreatic cancer (Lyu et al., 2024), three proteins (CFH, B3GNT8 and CFHR4) associated with diabetic retinopathy (Yuan et al., 2024), and thirteen proteins associated with coronary heart disease, seven of which were causally associated with myocardial infarction, including **PCSK9** and tyrosine-protein kinase **FES** (Sun et al., 2024). MR and protein–protein interaction analyses have been employed to identify therapeutic targets for myasthenia gravis within the **BLyS** / **APRIL** pathway, which is targeted by approved medicines like **belimumab** and **telitacicept** (Ouyang et al., 2024). Dozens of additional putative causal links between circulating proteins and complex illnesses have been reported in the literature (Hu et al., 2024; Si et al., 2024; Tao et al., 2024; Zhang et al., 2024; Zhu et al., 2024), which is likely to grow further over the coming years. Metabolomics data also are useful in predicting and/or confirming the effect of drug targets. For example, genetically predicted inhibition of LDL-cholesterol lowering targets, **HMGCR** and PCSK9, is associated with (expected) lower lipoprotein levels but had a negligible effect on glycoprotein acetyls levels, an NMR-derived biomarker of systemic inflammation (Richardson et al., 2022).

5.3 | Drug repurposing

The observation that drugs with genetic evidence supporting their efficacy are more likely to be approved has inspired the use of genetic data to identify targets for drug repurposing. This involves identifying novel indications for existing approved drugs and, as a result, accelerates the time to drug approval as the pre-clinical safety checks have already been completed.

For example, GWAS summary statistics, including those utilising UK Biobank data, were used to conduct MR and colocalisation analyses to investigate the plasma proteome as a source of therapeutic targets identifying >100 potentially causal associations between various proteins and health outcomes (Zheng et al., 2020). Drugs, both approved and in development, were associated with other phenotypes beyond the ones for which they were developed, suggesting there are wider opportunities for drug repurposing, such as the potential repurposing of Kinlytic (**urokinase**) for the potential treatment of inflammatory bowel disease (Zheng et al., 2020). Further, the genetic and proteomic data have identified specific immune-related proteins, many of which have drugs that are approved or are in advanced

clinical trials for cardiovascular and autoimmune conditions, are also implicated in a range of neuropsychiatric conditions, strengthening the rationale for the repurposing of specific drugs for these traditionally difficult-to-treat conditions (Dardani et al., 2024).

6 | UK BIOBANK AND TAILORED THERAPY

Medical treatments are generally designed for the average patient, but it is well known that not all patients will have the same response to, or success with, a given treatment. Therapy that is tailored to the individual (often termed personalised, or precision, medicine) involves the identification of individuals, or groups of individuals, who have distinct manifestations of disease or differential responses to treatment, so that their treatment can be tailored accordingly. Such a tailored approach is largely limited in the NHS to the treatment of some rare diseases and cancers. For example, **tamoxifen** is provided to women who specifically have **oestrogen receptor**-positive breast cancer (NICE, 2018), and **ivacaftor** is provided specifically for the treatment of cystic fibrosis patients who have ***CFTR*** gene variants (Davis et al., 2012; NHS England, 2023). Whilst the application of UK Biobank to drug discovery is currently more established than its use in tailoring therapy, the availability of large-scale genomic and longitudinal data provides an important opportunity for future research in this area, even though this work remains in its infancy.

6.1 | Pharmacogenomics

Pharmacogenomics refers to the identification and study of genetic variants that influence drug metabolism, transport, efficacy and toxicity. It can help identify different patient subpopulations to enable a personalised treatment approach. A personalised approach to drug prescription based on the patient's genetic and/or molecular makeup can improve drug safety and efficacy, and potentially reduce healthcare costs. Several UK Biobank studies have investigated known pharmacogenes and their effect on drug responses (McInnes et al., 2021; Sangkuhl et al., 2023; Wendt et al., 2021), and others have gone beyond known pharmacogenes to identify novel variants associated with drug response (Lavertu et al., 2020; Sadler et al., 2024). Indeed, research from UK Biobank suggests that almost everyone has actionable pharmacogenetic variants (across the 14 genes studied) that could affect treatment response (McInnes et al., 2021), highlighting the potential importance of integrating genetic information into tailored treatment strategies.

Although not focussing on the drug target gene itself, the identification of individuals who metabolise drugs differently also is important in tailoring treatment. For example, associations between certain pharmacogenes and drug responses have shown that the dosage required by patients varied according to their genetic profile, notably those with a particular ***CYP2C19*** variant requiring a lower **warfarin** maintenance dose (McInnes & Altman, 2021). As such, genetic profiling could aid the optimisation of warfarin prescription, although it is worth noting that other factors, such as diet, liver function and other drug–drug interactions may have a large impact on dosage requirements over the long term. Having repeated longitudinal measurements of relevant traits before, during and after treatment is important to monitor the efficacy of such tailored approaches.

6.2 | Research into drug adherence

UK Biobank data also have been used to study the genetic determinants of medication use that might explain inter-individual variability in treatment response. For example, a large number of genetic loci (most of which are shared with the underlying risk factors) are associated with long-term medication adherence and response in hyperlipidaemia, hypertension and type 2 diabetes (Kiiskinen et al., 2023), offering potential strategies for personalised medicine and improved prevention of cardiovascular diseases. Other research has shown that individuals with a high genetic risk of schizophrenia and who have a well-known variant in the ***SLCO1B1*** gene tend to have a lower adherence to statin therapy (as well as an increased risk of statin-induced myopathy) and were more likely to discontinue treatment (Türkmen et al., 2022). Conversely, certain variants in the ***ABCB1*** and ***ABCG2*** genes are associated with improved statin adherence, most likely through increased **simvastatin** plasma concentration (Malki et al., 2021). Whereas these findings highlight the potential benefit of using genetics to tailor interventions to optimise treatment efficacy, socio-demographic factors, such as education and income, also are likely to be effective as indicators of medication adherence (including statin therapy) as genetic predisposition (Hope et al., 2019).

6.3 | Research into side effects

The availability of large-scale studies with a wealth of genetic and clinical data is now facilitating research into the effects of drug targets on a wide range of health outcomes using phenome-wide association studies (PheWAS), that investigate the effect of a genetic variant that reflects a drug target across the whole phenotype in an agnostic (or unbiased) way, to identify pleiotropic effects that may lead to adverse side effects. This type of analysis has shown, for example, that a variant in the ***interleukin 6 receptor*** (*IL6R*) gene, which has a similar biochemical effect as **IL-6** blocking therapy, is associated with both a lower risk of aortic aneurism, which is related to its primary indication of giant cell arthritis, and also with a higher risk of atopic dermatitis and other dermatological conditions (Cai et al., 2018). Other research using this approach has found, for example, that a genetic variant in the ***PNPLA3*** gene is not only associated with an increased risk of liver disease and raised liver enzymes but also is associated with lower cholesterol levels and statin use, acne, gout and gallstones, suggesting that therapeutic inhibition of PNPLA3 could treat liver diseases, as well as being beneficial for other aspects of health (Diogo et al., 2018). A similar approach, which used genetic

instruments as a proxy for the use of statins (or other lipid-lowering medications), has shown no evidence of adverse disease associations and, if anything, a suggestive beneficial association with brain health (denoted with a positive association with hippocampal volume) (Pham et al., 2023). However, other research has found that a higher genetically predicted statin response was associated with a small increased risk of intracerebral haemorrhage risk (Mayerhofer et al., 2022), highlighting how different analytical methods can produce different results. In particular, a PheWAS approach, when combined with data on the genetic determinants of protein levels and tissue-specific genetic expression, can be particularly useful to inform drug side effects, for example, in clinical trials (Duffy et al., 2020). Other research focussed on well-known pharmacogenetic alleles have identified associations between certain cytochrome P450 enzyme polymorphisms and incidence of side effects with specific drugs. For example, CYP2C19 intermediate-metabolising patients taking **citalopram** had significantly lower likelihood of having herpes zoster than normal metabolisers, providing insights into how treatment strategies could be tailored to individuals based on genetics (McInnes & Altman, 2021).

6.4 | Disease risk prediction

PRS (polygenic risk scores) estimate the combined effects of genetic variants on a phenotype of interest, for example, the risk of developing a disease or the levels of a biological marker (such as, for example, LDL cholesterol). Identifying genetically predicted 'at risk' individuals can inform which population subgroups may require monitoring or early intervention (such as screening) to prevent disease onset or reduce disease severity through optimised treatment or lifestyle changes.

Whereas traditionally the role of genetics in precision medicine has been largely in relation to rare, single gene variants with a significant burden, UK Biobank data have been instrumental to research investigating the cumulative genetic disease risk burden originating from common genetic variants. UK Biobank data have been used to develop risk scores for many health conditions, almost all of which have a sizeable genetic component (Forgetta et al., 2020; Inouye et al., 2018; Riveros-Mckay et al., 2021; Sukcharoen et al., 2020). An early example of this was in 2018, where researchers found that a particular genetic profile present in 8% of the population was associated with a three times higher risk of coronary artery disease, compared to the genetic profiles of the remainder of the study population (Khera et al., 2018). Further, the prevalence of these common risk variants in the population was found to be 20 times higher than rare monogenic variants that confer a comparable risk. Such scores could therefore be used to identify individuals who are at high genetic risk of developing heart disease but do not carry monogenic disease risk variants (Khera et al., 2018). When considering the benefit, it is important to note that the same relative reduction in risk may translate to a larger absolute risk reduction in some population subgroups. For example, a study using UK Biobank exome sequencing data identified carriers of pathogenic familial hypocholesteraemia variants and investigated the effects of lifestyle on subsequent risk of disease. The absolute risk reduction among carriers of pathogenic variants who maintained a favourable lifestyle was more than twice that observed among non-carriers (Fahed et al., 2022).

The NHS in England undertook a pilot study to assess the clinical utility of incorporating PRS into clinical risk scores in primary care. Integrating PRS into the QRISK2 score (which includes demographic, lifestyle and clinical information) to predict the 10-year risk of cardiovascular disease resulted in changes to the care and management of ~30% of patients that had a higher integrated risk score compared to their QRISK2 score (Fuat et al., 2024). PRS may be of particular value for conditions for which there are few identified modifiable risk factors (e.g. prostate cancer), or where the PRS is more discriminatory at younger ages. Not all disease risk scores require genetic information. For example, a dementia risk score based on clinical and sociodemographic information in the UK Biobank population ('UKBDRS') outperformed existing risk scores developed in other cohorts, with and without the addition of *APOE4* gene status (Anatürk et al., 2023). Protein-based scores also hold promise for more powerful prediction of incident diseases. For example, combining circulating proteins like GFAP, NEFL, **GDF15** and LTBP2 with demographic information can predict 10-year all-cause dementia with similar accuracy to that of models containing data from cognitive function tests (Guo et al., 2024). Protein scores also have been shown to outperform PRS-based risk prediction for type 2 diabetes, likely as a result of the protein score capturing genetic and environmental risk, whereas PRS only capture genetic risk (Gadd et al., 2024). Others have used proteomics to better predict the 10-year incidence of over 200 diseases compared with existing clinical diagnostic models (Carrasco-Zanini et al., 2024). More recently a novel research tool has been developed to predict risk of >3000 diseases leveraging UK Biobank biochemistry, protein and health outcome data. The tool can identify cases incorrectly classified as controls, increasing statistical power for genetic discovery with implications for drug target discovery (Garg et al., 2024).

7 | LIMITATIONS

It is important to acknowledge the inherent limitations of using UK Biobank, and similar studies, for drug development. It is well documented that the UK Biobank population is predominantly White, with 88% of the cohort genetically clustered into the 'White-British' group (Bycroft et al., 2018). To date, most genetic studies have been conducted in high-income countries on White-European populations, with less representation of other ancestries. Lack of diverse population ancestries in genetic research can limit the discovery of genetic variants associated with complex diseases, complicating early disease detection and diagnosis, as well as slowing genetics-guided drug discovery (Sirugo et al., 2019). Additionally, this limitation may hinder the generalisability of findings and the development of precision medicine interventions that effectively cater to individuals from varied ancestries, potentially leading to disparities in health-care outcomes. African ancestry populations are particularly genetically diverse and have been important in the discovery of drugs for the treatment of

dyslipidaemia. Loss of function variants in the *PCSK9* gene, associated with lower levels of plasma LDL, were found to occur in 2% of African Americans but in less than 0.1% of European Americans (Cohen et al., 2005). Genetically diverse population studies are needed to enhance the possibility of identifying novel disease variants and drug targets. It is important to note that while the ethnic distribution of UK Biobank is reflective of the population structure of the United Kingdom at the time of recruitment, this has changed over time and the overall numbers of ethnic minority groups is relatively small. New studies currently being established, such as Our Future Health in the United Kingdom and All of Us in the United States, are seeking to recruit more diverse populations and empower the discovery of novel variants associated with disease.

UK Biobank participants were consented on the basis that they would not receive any feedback about their individual results (UK Biobank tests are conducted for research purposes only and are not diagnostic); as such, it is not possible to re-contact participants to either invite them to engage in study enhancements or recruitment into other studies, based on the information that is not known to them. In the context of genetic pharmacoepidemiology, this prevents participant recall by genotype. Recall by genotype would be valuable to investigate phenotypes not assessed within UK Biobank. For example, it would be informative to study **insulin** response among individuals with putative loss of function variants in genes relevant to diabetes. This is of particular interest in drug discovery research where recruitment of participants with variants of interest (notably rare loss of function variants) would allow detailed research into disease biology that could lead to improved treatments and personalised health care.

Research into some rare diseases may be impeded because of being a small number of cases in the UK Biobank population. For example, there will be relatively few participants that have rare, inherited forms of developmental disorders, or disorders that frequently result in death in childhood or early adulthood. Nonetheless, researchers are using UK Biobank data for some rare disease research, including to enhance phenotyping of rare diseases (Patrick et al., 2022) and to investigate the penetrance of rare variants by comparing the phenotypic presentation of those with a particular rare disease and the asymptomatic variant carriers (Blair & Risch, 2024; Fitzsimmons et al., 2024).

Measurements from one point in time are subject to random errors that may lead to the underestimation of the strength of association with health outcomes (regression dilution bias), similarly within-person variation over time also will lead to regression dilution (Clarke et al., 1999). Using repeat measures from a subset of the cohort allows for correction of regression dilution. Repeat measures are available for 20,000 participants from the repeat assessment as well as for participants who have attended an imaging visit, but for most of the cohort, their exposures, physical measurements and biological samples were collected only at one time point. UK Biobank plans to invite as many surviving participants as possible to a repeat of the baseline assessment which would allow research into how changes in exposures over mid-to-late life affect disease risk. Lack of multiple longitudinal assessments prevents research that tracks disease progression over time. For some variables of interest, data are available through medical records, but this only includes individuals with chronic or acute medical conditions that have consulted a physician. Biological factors that trigger disease onset may not be the same as those that drive disease progression, and therapies are generally 'reactive', tailored to manage disease progression rather than prevent disease onset. As such, the lack of longitudinal data for the whole cohort limits research into factors influencing disease progression and, consequently, limits the discovery of drug targets that could modulate disease progression.

The UK Biobank participants are known to be, on average, slightly healthier and wealthier than the general population of the same age range, details of which can be found elsewhere (Fry et al., 2017). For example, the proportion of current smokers at the time of recruitment among 45–54-year-old men was 15% in UK Biobank and 22% in the general population. In contrast, 95% of the cohort are of White ethnicity, which was similar to the national population of the same age range at the time of recruitment (although which declined in recent years). Furthermore, children and young people are not included, and pregnant women are underrepresented, owing to the older age range at recruitment. Whereas the study population is not wholly representative of the general population from which it was sampled, relationships between exposures and diseases may nonetheless be more widely generalisable. Generalisability may be assessed by examining consistency across subgroups and by comparing findings with other studies (Fry et al., 2017), and analytical techniques are now widely established to identify and correct for such sampling bias where necessary (Allen et al., 2024; Lash et al., 2014).

Tissue-specific gene expression or protein levels would be of interest to drug discovery; however, as tissue samples collected as part of routine health care are not currently collected in UK Biobank, it is only possible to investigate the circulatory levels of proteins. Finally, reliance on electronic health records for outcome ascertainment can further exacerbate inequalities, as sociodemographic inequalities are associated with attendance to health clinics (Kerr et al., 2023) and the associated reporting. However, linkage to electronic medical records provides comprehensive coverage and is likely to be less subjected to bias than reliance on self-report.

8 | FUTURE DIRECTIONS

In addition to the planned repeat assessment of all participants, UK Biobank is planning to undertake a 'Brain Health Study' to enable research into neurodegenerative disorders. The participants diagnosed with, or with symptoms of, a neurodegenerative disorder will be invited to an assessment centre to undertake a series of tests (e.g. brain MRI, cognitive assessment, blood sample and remote monitoring devices) to enable researchers to identify disease subtypes. Such a study will generate more precise phenotypic and functional data to enrich routinely available medical records (e.g. hospital inpatient admissions and primary care data), allowing research into the

biology of neurodegenerative disease subtypes with potential implications for disease-specific treatment and prevention. To enable research into molecular subtypes of tumours, the UK Biobank intends to collect tumour digital histopathology images to enable researchers to infer tissue-specific gene expression levels. The UK Biobank is also undertaking additional data linkages to NHS datasets to enhance health outcome characterisation, including clinical audits and registry data, which will allow research into specific disease subtypes.

Proteomic data are particularly valuable to drug discovery given that proteins are the most common targets of pharmaceutical agents. Generating protein measurements at several time points will facilitate investigations of how protein levels vary over time and in relation to certain exposures or disease progression. The extension of the proteomics assay to the whole cohort, including repeat samples from the imaging visits, will increase sample sizes and statistical power for biomarker discovery and for research into the temporal dynamics of proteins throughout the progression of common diseases. The Olink assay is a dual antibody-based platform whose latest iteration measures approximately 5400 proteins, capturing only a fraction of the human proteome. The unmeasured proteome may be of great importance to certain conditions, so additional proteomics technologies, such as aptamer (Schneider et al., 2022) or NULISA (Feng et al., 2023) assays, may be warranted in the future. Certain proteins with potential clinical and diagnostic value are not included in the Olink panel, such as **amyloid beta 40 and 42**, as well as pTau-181, pTau-217 and other phosphorylated **tau** species, which are of interest for research into neurodegeneration (Ashton et al., 2024). Next generation mass spectrometry based technologies (Blume et al., 2020; Guzman et al., 2024) could be used to detect proteins with post-translation modifications, protein isoforms and other proteins not included in targeted assays (Duff et al., 2024).

9 | CONCLUSIONS

The UK Biobank is one of the more widely used biomedical research resources globally. The combination of genetic, proteomic and other assay data with health outcomes, on a scale unmatched by any other study, is unprecedented. Because of the easy accessibility of UK Biobank data, it has been used extensively to advance health research. As the cohort ages, more incident diseases occur, and further assays are conducted on the biological samples, the more valuable the UK Biobank resource will become for all types of health-related research including for drug target identification, characterisation and prioritisation, as well as precision medicine.

9.1 | Nomenclature of Targets and Ligands

Key protein targets and ligands in this article are hyperlinked to corresponding entries in https://www.guidetopharmacology.org and are permanently archived in the Concise Guide to PHARMACOLOGY 2023/24 (Alexander et al., 2023).

AUTHOR CONTRIBUTIONS

Jelena Bešević: Conceptualization (equal); project administration (lead); writing—original draft (lead); writing—review and editing (equal). **Saredo Said:** Writing—original draft (equal); writing—review and editing (equal). **Reka Nagy:** Writing—original draft (equal); writing—review and editing (equal). **Yalda Jamshidi:** Writing—original draft (equal); writing—review and editing (equal). **Christopher D. Whelan:** Writing—original draft (equal); writing—review and editing (equal). **Lauren Carson:** Writing—review and editing (equal). **Martin K. Rutter:** Writing—review and editing (equal). **Adam J. Lewandowski:** Writing—review and editing (equal). **Mark Effingham:** Writing—review and editing (equal). **Rory Collins:** Writing—review and editing (equal). **Ben Lacey:** Conceptualization (equal); writing—review and editing (equal). **Naomi E. Allen:** Conceptualization (equal); writing—original draft (equal); writing—review and editing (equal).

ACKNOWLEDGEMENTS

We would like to thank all UK Biobank participants for their contributions to this resource. Thank you also to Lish Motley for producing Figures 1 and 2. Thank you to Ben Holbrook for providing the data for Figure 2.

CONFLICT OF INTEREST STATEMENT

JB, LC, MKR, AJL, RC, BL and NEA are members of the scientific, access and/or executive team of UK Biobank. SS, RN and YJ are employed by Novo Nordisk. CDW is an employee and stockholder of Johnson & Johnson Innovative Medicine.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article because no new data were created or analysed in this study.

ORCID

Ben Lacey https://orcid.org/0000-0003-0139-2934

REFERENCES

Akbari, P., Gilani, A., Sosina, O., Kosmicki, J. A., Khrimian, L., Fang, Y. Y., Persaud, T., Garcia, V., Sun, D., Li, A., Mbatchou, J., Locke, A. E., Benner, C., Verweij, N., Lin, N., Hossain, S., Agostinucci, K., Pascale, J. V., Dirice, E., … Lotta, L. A. (2021). Sequencing of 640,000 exomes identifies GPR75 variants associated with protection from obesity. *Science*, *373*(6550), eabf8683. https://doi.org/10.1126/science.abf8683

Alexander, S. P. H., Kelly, E., Mathie, A. A., Peters, J. A., Veale, E. L., Armstrong, J. F., Buneman, O. P., Faccenda, E., Harding, S. D., Spedding, M., Cidlowski, J. A., Fabbro, D., Davenport, A. P., Striessnig, J., Davies, J. A., Ahlers-Dannen, K. E., Alqinyah, M., Arumugam, T. V., Bodle, C., … Zolghadri, Y. (2023). The Concise Guide to PHARMACOLOGY 2023/24: Introduction and Other Protein Targets. *British Journal of Pharmacology*, *180*, S1–S22. https://doi.org/10.1111/bph.16176

Alfaro-Almagro, F., Jenkinson, M., Bangerter, N. K., Andersson, J. L. R., Griffanti, L., Douaud, G., Sotiropoulos, S. N., Jbabdi, S., Hernandez-Fernandez, M., Vallee, E., Vidaurre, D., Webster, M., McCarthy, P., Rorden, C., Daducci, A., Alexander, D. C., Zhang, H., Dragonu, I., Matthews, P. M., … Smith, S. M. (2018). Image processing and quality

control for the first 10,000 brain imaging datasets from UK Biobank. *NeuroImage*, *166*, 400–424. https://doi.org/10.1016/j.neuroimage.2017.10.034

Allen, N. E., Lacey, B., Lawlor, D. A., Pell, J. P., Gallacher, J., Smeeth, L., Elliott, P., Matthews, P. M., Lyons, R. A., Whetton, A. D., Lucassen, A., Hurles, M. E., Chapman, M., Roddam, A. W., Fitzpatrick, N. K., Hansell, A. L., Hardy, R., Marioni, R. E., O'Donnell, V. B., … Collins, R. (2024). Prospective study design and data analysis in UK Biobank. *Science Translational Medicine*, *16*(729), eadf4428. https://doi.org/10.1126/scitranslmed.adf4428

Anatürk, M., Patel, R., Ebmeier, K. P., Georgiopoulos, G., Newby, D., Topiwala, A., de Lange, A.-M. G., Cole, J. H., Jansen, M. G., Singh-Manoux, A., Kivimäki, M., & Suri, S. (2023). Development and validation of a dementia risk score in the UK Biobank and Whitehall II cohorts. *BMJ Mental Health*, *26*(1), e300719. https://doi.org/10.1136/bmjment-2023-300719

Aragam, K. G., Jiang, T., Goel, A., Kanoni, S., Wolford, B. N., Atri, D. S., Weeks, E. M., Wang, M., Hindy, G., Zhou, W., Grace, C., Roselli, C., Marston, N. A., Kamanu, F. K., Surakka, I., Venegas, L. M., Sherliker, P., Koyama, S., Ishigaki, K., … Butterworth, A. S. (2022). Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. *Nature Genetics*, *54*(12), 1803–1815. https://doi.org/10.1038/s41588-022-01233-6

Ashley, E. A. (2015). The precision medicine initiative: A new national effort. *JAMA*, *313*(21), 2119–2120. https://doi.org/10.1001/jama.2015.3595

Ashton, N. J., Brum, W. S., Di Molfetta, G., Benedet, A. L., Arslan, B., Jonaitis, E., Langhough, R. E., Cody, K., Wilson, R., Carlsson, C. M., Vanmechelen, E., Montoliu-Gaya, L., Lantero-Rodriguez, J., Rahmouni, N., Tissot, C., Stevenson, J., Servaes, S., Therriault, J., Pascoal, T., … Zetterberg, H. (2024). Diagnostic accuracy of a plasma phosphorylated tau 217 immunoassay for Alzheimer disease pathology. *JAMA Neurology*, *81*(3), 255–263. https://doi.org/10.1001/jamaneurol.2023.5319

Bešević, J., Lacey, B., Callen, H., Omiyale, W., Conroy, M., Feng, Q., Crook, D. W., Doherty, N., Ebner, D., Eyre, D. W., Fry, D., Horn, E., Jones, E. Y., Marsden, B. D., Peto, T. E. A., Starkey, F., Stuart, D., Welsh, S., Wood, N., … Allen, N. (2023). Persistence of SARS-CoV-2 antibodies over 18 months following infection: UK Biobank COVID-19 serology study. *Journal of Epidemiology and Community Health*, *78*(2), 105–108. https://doi.org/10.1136/jech-2023-220569

Blair, D. R., & Risch, N. (2024). Dissecting the reduced penetrance of putative loss-of-function variants in population-scale biobanks. Medrxiv, 2024.2009.2023.24314008. https://doi.org/10.1101/2024.09.23.24314008

Blume, J. E., Manning, W. C., Troiano, G., Hornburg, D., Figa, M., Hesterberg, L., Platt, T. L., Zhao, X., Cuaresma, R. A., Everley, P. A., Ko, M., Liou, H., Mahoney, M., Ferdosi, S., Elgierari, E. M., Stolarczyk, C., Tangeysh, B., Xia, H., Benz, R., … Farokhzad, O. C. (2020). Rapid, deep and precise profiling of the plasma proteome with multi-nanoparticle protein corona. *Nature Communications*, *11*(1), 3662. https://doi.org/10.1038/s41467-020-17033-7

Bradbury, K. E., Young, H. J., Guo, W., & Key, T. J. (2018). Dietary assessment in UK Biobank: An evaluation of the performance of the touchscreen dietary questionnaire. *Journal of Nutritional Science*, *7*, e6. https://doi.org/10.1017/jns.2017.66

Brailean, A., Curtis, J., Davis, K., Dregan, A., & Hotopf, M. (2020). Characteristics, comorbidities, and correlates of atypical depression: Evidence from the UK Biobank Mental Health Survey. *Psychological Medicine*, *50*(7), 1129–1138. https://doi.org/10.1017/s0033291719001004

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., Cortes, A., Welsh, S., Young, A., Effingham, M., McVean, G., Leslie, S., Allen, N., Donnelly, P., & Marchini, J. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature*, *562*(7726), 203–209. https://doi.org/10.1038/s41586-018-0579-z

Cai, T., Zhang, Y., Ho, Y.-L., Link, N., Sun, J., Huang, J., Cai, T. A., Damrauer, S., Ahuja, Y., Honerlaw, J., Huang, J., Costa, L., Schubert, P., Hong, C., Gagnon, D., Sun, Y. V., Gaziano, J. M., Wilson, P., Cho, K., … Liao, K. P. (2018). Association of interleukin 6 receptor variant with cardiovascular disease effects of interleukin 6 receptor blocking therapy: A phenome-wide association study. *JAMA Cardiology*, *3*(9), 849–857. https://doi.org/10.1001/jamacardio.2018.2287

Carrasco-Zanini, J., Pietzner, M., Davitte, J., Surendran, P., Croteau-Chonka, D. C., Robins, C., Torralbo, A., Tomlinson, C., Grünschläger, F., Fitzpatrick, N., Ytsma, C., Kanno, T., Gade, S., Freitag, D., Ziebell, F., Haas, S., Denaxas, S., Betts, J. C., Wareham, N. J., … Langenberg, C. (2024). Proteomic signatures improve risk prediction for common and rare diseases. *Nature Medicine*, *30*, 2489–2498. https://doi.org/10.1038/s41591-024-03142-z

Clarke, R., Shipley, M., Lewington, S., Youngman, L., Collins, R., Marmot, M., & Peto, R. (1999). Underestimation of risk associations due to regression dilution in long-term follow-up of prospective studies. *American Journal of Epidemiology*, *150*(4), 341–353. https://doi.org/10.1093/oxfordjournals.aje.a010013

Codd, V., Denniff, M., Swinfield, C., Warner, S. C., Papakonstantinou, M., Sheth, S., Nanus, D. E., Budgeon, C. A., Musicha, C., Bountziouka, V., Wang, Q., Bramley, R., Allara, E., Kaptoge, S., Stoma, S., Jiang, T., Butterworth, A. S., Wood, A. M., Di Angelantonio, E., … Samani, N. J. (2022). Measurement and initial characterization of leukocyte telomere length in 474,074 participants in UK Biobank. *Nature Aging*, *2*(2), 170–179. https://doi.org/10.1038/s43587-021-00166-9

Cohen, J., Pertsemlidis, A., Kotowski, I. K., Graham, R., Garcia, C. K., & Hobbs, H. H. (2005). Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. *Nature Genetics*, *37*(2), 161–165. https://doi.org/10.1038/ng1509

Collins, F. S., & Varmus, H. (2015). A new initiative on precision medicine. *New England Journal of Medicine*, *372*(9), 793–795. https://doi.org/10.1056/NEJMp1500523

Conroy, M., Sellors, J., Effingham, M., Littlejohns, T. J., Boultwood, C., Gillions, L., Sudlow, C. L. M., Collins, R., & Allen, N. E. (2019). The advantages of UK Biobank's open-access strategy for health research. *Journal of Internal Medicine*, *286*(4), 389–397. https://doi.org/10.1111/joim.12955

Dardani, C., Robinson, J. W., Jones, H. J., Rai, D., Stergiakouli, E., Grove, J., Gardner, R., McIntosh, A. M., Havdahl, A., Hemani, G., Smith, G. D., Richardson, T. G., Gaunt, T. R., & Khandaker, G. M. (2024). Immunological drivers and potential novel drug targets for major psychiatric, neurodevelopmental, and neurodegenerative conditions. medRxiv, 2024.2002.2016.24302885. https://doi.org/10.1101/2024.02.16.24302885

Davis, K. A. S., Coleman, J. R. I., Adams, M., Allen, N., Breen, G., Cullen, B., Dickens, C., Fox, E., Graham, N., Holliday, J., Howard, L. M., John, A., Lee, W., McCabe, R., McIntosh, A., Pearsall, R., Smith, D. J., Sudlow, C., Ward, J., … Hotopf, M. (2020). Mental health in UK Biobank – Development, implementation and results from an online questionnaire completed by 157 366 participants: A reanalysis. *BJPsych Open*, *6*(2), e18. https://doi.org/10.1192/bjo.2019.100

Davis, P. B., Yasothan, U., & Kirkpatrick, P. (2012). Ivacaftor. *Nature Reviews Drug Discovery*, *11*(5), 349–350. https://doi.org/10.1038/nrd3723

Deng, Y.-T., Wu, B.-S., Yang, L., He, X.-Y., Kang, J.-J., Liu, W.-S., Li, Z.-Y., Wu, X.-R., Zhang, Y.-R., Chen, S.-D., Ge, Y.-J., Huang, Y.-Y., Feng, J.-F., Zhu, Y., Dong, Q., Mao, Y., Cheng, W., & Yu, J.-T. (2024). Large-scale whole-exome sequencing of neuropsychiatric diseases and traits in 350,770 adults. *Nature Human Behaviour*, *8*(6), 1194–1208. https://doi.org/10.1038/s41562-024-01861-4

Dhindsa, R. S., Burren, O. S., Sun, B. B., Prins, B. P., Matelska, D., Wheeler, E., Mitchell, J., Oerton, E., Hristova, V. A., Smith, K. R., Carss, K., Wasilewski, S., Harper, A. R., Paul, D. S., Fabre, M. A., Runz, H., Viollet, C., Challis, B., Platt, A., … Petrovski, S. (2023). Rare variant

associations with plasma protein levels in the UK Biobank. *Nature*, *622*(7982), 339–347. https://doi.org/10.1038/s41586-023-06547-x

Diogo, D., Tian, C., Franklin, C. S., Alanne-Kinnunen, M., March, M., Spencer, C. C. A., Vangjeli, C., Weale, M. E., Mattsson, H., Kilpeläinen, E., Sleiman, P. M. A., Reilly, D. F., McElwee, J., Maranville, J. C., Chatterjee, A. K., Bhandari, A., Nguyen, K.-D. H., Estrada, K., Reeve, M.-P., … Runz, H. (2018). Phenome-wide association studies across large population cohorts support drug target validation. *Nature Communications*, *9*(1), 4285. https://doi.org/10.1038/s41467-018-06540-3

Doherty, A., Jackson, D., Hammerla, N., Plotz, T., Olivier, P., Granat, M. H., White, T., van Hees, V. T., Trenell, M. I., Owen, C. G., Preece, S. J., Gillions, R., Sheard, S., Peakman, T., Brage, S., & Wareham, N. J. (2017). Large scale population assessment of physical activity using wrist worn accelerometers: The UK Biobank study. *PLoS ONE*, *12*(2), e0169649. https://doi.org/10.1371/journal.pone.0169649

Downey, P., & Peakman, T. C. (2008). Design and implementation of a high-throughput biological sample processing facility using modern manufacturing principles. *International Journal of Epidemiology*, *37*-(Suppl 1), i46–i50. https://doi.org/10.1093/ije/dyn031

Duff, E. P., Zetterberg, H., Heslegrave, A., Dehghan, A., Elliott, P., Allen, N., Runz, H., Laban, R., Veleva, E., Whelan, C. D., Sun, B. B., & Matthews, P. M. (2024). Plasma proteomic evidence for increased Alzheimer's disease-related brain pathology after SARS-CoV-2 infection. medRxiv. 2024.2002.2002.24302132. https://doi.org/10.1101/2024.02.02.24302132

Duffy, Á., Verbanck, M., Dobbyn, A., Won, H.-H., Rein, J. L., Forrest, I. S., Nadkarni, G., Rocheleau, G., & Do, R. (2020). Tissue-specific genetic features inform prediction of drug side effects in clinical trials. *Science Advances*, *6*(37), eabb6242. https://doi.org/10.1126/sciadv.abb6242

Elliott, P., Peakman, T. C., & Biobank, U. K. (2008). The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. *International Journal of Epidemiology*, *37*(2), 234–244. https://doi.org/10.1093/ije/dym276

Fahed, A. C., Wang, M., Patel, A. P., Ajufo, E., Maamari, D. J., Aragam, K. G., Brockman, D. G., Vosburg, T., Ellinor, P. T., Ng, K., & Khera, A. V. (2022). Association of the interaction between familial hypercholesterolemia variants and adherence to a healthy lifestyle with risk of coronary artery disease. *JAMA Network Open*, *5*(3), e222687. https://doi.org/10.1001/jamanetworkopen.2022.2687

Feng, W., Beer, J. C., Hao, Q., Ariyapala, I. S., Sahajan, A., Komarov, A., Cha, K., Moua, M., Qiu, X., Xu, X., Iyengar, S., Yoshimura, T., Nagaraj, R., Wang, L., Yu, M., Engel, K., Zhen, L., Xue, W., Lee, C.-j., … Ma, X.-J. (2023). NULISA: A proteomic liquid biopsy platform with attomolar sensitivity and high multiplexing. *Nature Communications*, *14*(1), 7238. https://doi.org/10.1038/s41467-023-42834-x

Fitzsimmons, L., Beaulieu-Jones, B., and Kobren, S. N. (2024). Phenotypic overlap between rare disease patients and variant carriers in a large population cohort informs biological mechanisms. Medrxiv. 2024.04.18.24305861. https://doi.org/10.1101/2024.04.18.24305861

Forgetta, V., Keller-Baruch, J., Forest, M., Durand, A., Bhatnagar, S., Kemp, J. P., Nethander, M., Evans, D., Morris, J. A., Kiel, D. P., Rivadeneira, F., Johansson, H., Harvey, N. C., Mellström, D., Karlsson, M., Cooper, C., Evans, D. M., Clarke, R., Kanis, J. A., … Richards, J. B. (2020). Development of a polygenic risk score to improve screening for fracture risk: A genetic risk prediction study. *PLoS Medicine*, *17*(7), e1003152. https://doi.org/10.1371/journal.pmed.1003152

Fry, A., Littlejohns, T. J., Sudlow, C., Doherty, N., Adamska, L., Sprosen, T., Collins, R., & Allen, N. E. (2017). Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *American Journal of Epidemiology*, *186*(9), 1026–1034. https://doi.org/10.1093/aje/kwx246

Fuat, A., Adlen, E., Monane, M., Coll, R., Groves, S., Little, E., Wild, J., Kamali, F. J., Soni, Y., Haining, S., Riding, H., Riveros-Mckay, F., Peneva, I., Lachapelle, A., Giner-Delgado, C., Weale, M. E., Plagnol, V., Harrison, S., & Donnelly, P. (2024). A polygenic risk score added to a QRISK®2 cardiovascular disease risk calculator demonstrated robust clinical acceptance and clinical utility in the primary care setting. *European Journal of Preventive Cardiology*, *31*(6), 716–722. https://doi.org/10.1093/eurjpc/zwae004

Gadd, D. A., Hillary, R. F., Kuncheva, Z., Mangelis, T., Cheng, Y., Dissanayake, M., Admanit, R., Gagnon, J., Lin, T., Ferber, K. L., Runz, H., Ferber, K. L., Foley, C. N., Marioni, R. E., Sun, B. B., & Biogen Biobank, T. (2024). Blood protein assessment of leading incident diseases and mortality in the UK Biobank. *Nature Aging*, *4*, 939–948. https://doi.org/10.1038/s43587-024-00655-7

Galante, J., Adamska, L., Young, A., Young, H., Littlejohns, T. J., Gallacher, J., & Allen, N. (2016). The acceptability of repeat Internet-based hybrid diet assessment of previous 24-h dietary intake: Administration of the Oxford WebQ in UK Biobank. *The British Journal of Nutrition*, *115*(4), 681–686. https://doi.org/10.1017/s0007114515004821

Garg, M., Karpinski, M., Matelska, D., Middleton, L., Burren, O. S., Hu, F., Wheeler, E., Smith, K. R., Fabre, M. A., Mitchell, J., O'Neill, A., Ashley, E. A., Harper, A. R., Wang, Q., Dhindsa, R. S., Petrovski, S., & Vitsios, D. (2024). Disease prediction with multi-omics and biomarkers empowers case-control genetic discoveries in the UK Biobank. *Nature Genetics*, *56*(9), 1821–1831. https://doi.org/10.1038/s41588-024-01898-1

Ghouse, J., Sveinbjörnsson, G., Vujkovic, M., Seidelin, A. S., Gellert-Kristensen, H., Ahlberg, G., Tragante, V., Rand, S. A., Brancale, J., Vilarinho, S., Lundegaard, P. R., Sørensen, E., Erikstrup, C., Bruun, M. T., Jensen, B. A., Brunak, S., Banasik, K., Ullum, H., Verweij, N., … Stender, S. (2024). Integrative common and rare variant analyses provide insights into the genetic architecture of liver cirrhosis. *Nature Genetics*, *56*(5), 827–837. https://doi.org/10.1038/s41588-024-01720-y

Guo, Y., You, J., Zhang, Y., Liu, W. S., Huang, Y. Y., Zhang, Y. R., Zhang, W., Dong, Q., Feng, J. F., Cheng, W., & Yu, J. T. (2024). Plasma proteomic profiles predict future dementia in healthy adults. *Nature Aging*, *4*(2), 247–260. https://doi.org/10.1038/s43587-023-00565-0

Guzman, U. H., Martinez-Val, A., Ye, Z., Damoc, E., Arrey, T. N., Pashkova, A., Renuse, S., Denisov, E., Petzoldt, J., Peterson, A. C., Harking, F., Østergaard, O., Rydbirk, R., Aznar, S., Stewart, H., Xuan, Y., Hermanson, D., Horning, S., Hock, C., … Olsen, J. V. (2024). Ultra-fast label-free quantification and comprehensive proteome coverage with narrow-window data-independent acquisition. *Nature Biotechnology*, *42*, 1855–1866. https://doi.org/10.1038/s41587-023-02099-7

Hope, H. F., Binkley, G. M., Fenton, S., Kitas, G. D., Verstappen, S. M. M., & Symmons, D. P. M. (2019). Systematic review of the predictors of statin adherence for the primary prevention of cardiovascular disease. *PLoS ONE*, *14*(1), e0201196. https://doi.org/10.1371/journal.pone.0201196

Hu, D., Jiang, J., Zhang, Q., & Lin, Z. (2024). Exploring the potential causal effects of myopia: A phenome-wide Mendelian randomization association study. Medrxiv, 2024.2006.2024.24309385. https://doi.org/10.1101/2024.06.24.24309385

Inouye, M., Abraham, G., Nelson, C. P., Wood, A. M., Sweeting, M. J., Dudbridge, F., Lai, F. Y., Kaptoge, S., Brozynska, M., Wang, T., Ye, S., Webb, T. R., Rutter, M. K., Tzoulaki, I., Patel, R. S., Loos, R. J. F., Keavney, B., Hemingway, H., Thompson, J., … Samani, N. J. (2018). Genomic risk prediction of coronary artery disease in 480,000 adults: Implications for primary prevention. *Journal of the American College of Cardiology*, *72*(16), 1883–1893. https://doi.org/10.1016/j.jacc.2018.07.079

Julkunen, H., Cichonska, A., Tiainen, M., Koskela, H., Nybo, K., Makela, V., Nokso-Koivisto, J., Kristiansson, K., Perola, M., Salomaa, V., Jousilahti, P., Lundqvist, A., Kangas, A. J., Soininen, P., Barrett, J. C., & Wurtz, P. (2023). Atlas of plasma NMR biomarkers for health and disease in 118,461 individuals from the UK Biobank. *Nature Communications*, *14*(1), 604. https://doi.org/10.1038/s41467-023-36231-7

Kerr, G., Greenfield, G., Hayhoe, B., Gaughran, F., Halvorsrud, K., Pinto da Costa, M., Rehill, N., Raine, R., Majeed, A., Costelloe, C., Neves, A. L., & Beaney, T. (2023). Attendance at remote versus in-person outpatient appointments in an NHS Trust. *Journal of Telemedicine and Telecare*, *31*(5), 721–731. https://doi.org/10.1177/1357633X231216501

Khera, A. V., Chaffin, M., Aragam, K. G., Haas, M. E., Roselli, C., Choi, S. H., Natarajan, P., Lander, E. S., Lubitz, S. A., Ellinor, P. T., & Kathiresan, S. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature Genetics*, *50*(9), 1219–1224. https://doi.org/10.1038/s41588-018-0183-z

Kiiskinen, T., Helkkula, P., Krebs, K., Karjalainen, J., Saarentaus, E., Mars, N., Lehisto, A., Zhou, W., Cordioli, M., Jukarainen, S., Rämö, J. T., Mehtonen, J., Veerapen, K., Räsänen, M., Ruotsalainen, S., Maasha, M., Niiranen, T., Tuomi, T., Salomaa, V., … FinnGen. (2023). Genetic predictors of lifelong medication-use patterns in cardiometabolic diseases. *Nature Medicine*, *29*(1), 209–218. https://doi.org/10.1038/s41591-022-02122-5

King, E. A., Davis, J. W., & Degner, J. F. (2019). Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. *PLoS Genetics*, *15*(12), e1008489. https://doi.org/10.1371/journal.pgen.1008489

Kosorok, M. R., & Laber, E. B. (2019). Precision medicine. *Annual Review of Statistics and Its Application*, *6*(6), 263–286. https://doi.org/10.1146/annurev-statistics-030718-105251

Lash, T. L., Fox, M. P., MacLehose, R. F., Maldonado, G., McCandless, L. C., & Greenland, S. (2014). Good practices for quantitative bias analysis. *International Journal of Epidemiology*, *43*(6), 1969–1985. https://doi.org/10.1093/ije/dyu149

Lavertu, A., McInnes, G., Tanigawa, Y., Altman, R. B., & Rivas, M. A. (2020). LPA and APOE are associated with statin selection in the UK Biobank. bioRxiv, 2020.2008.2028.272765. https://doi.org/10.1101/2020.08.28.272765

Li, S., Carss, K. J., Halldorsson, B. V., & Cortes, A. (2023). Whole-genome sequencing of half-a-million UK biobank participants. Medrxiv, 2023.2012.2006.23299426. https://doi.org/10.1101/2023.12.06.23299426

Linge, J., Borga, M., West, J., Tuthill, T., Miller, M. R., Dumitriu, A., Thomas, E. L., Romu, T., Tunon, P., Bell, J. D., & Dahlqvist Leinhard, O. (2018). Body composition profiling in the UK Biobank imaging study. *Obesity (Silver Spring)*, *26*(11), 1785–1795. https://doi.org/10.1002/oby.22210

Littlejohns, T. J., Holliday, J., Gibson, L. M., Garratt, S., Oesingmann, N., Alfaro-Almagro, F., Bell, J. D., Boultwood, C., Collins, R., Conroy, M. C., Crabtree, N., Doherty, N., Frangi, A. F., Harvey, N. C., Leeson, P., Miller, K. L., Neubauer, S., Petersen, S. E., Sellors, J., … Allen, N. E. (2020). The UK biobank imaging enhancement of 100,000 participants: Rationale, data collection, management and future directions. *Nature Communications*, *11*(1), 2624. https://doi.org/10.1038/s41467-020-15948-9

Liu, Y., Basty, N., Whitcher, B., Bell, J. D., Sorokin, E. P., van Bruggen, N., Thomas, E. L., & Cule, M. (2021). Genetic architecture of 11 organ traits derived from abdominal MRI using deep learning. *eLife*, *10*, e65554. https://doi.org/10.7554/eLife.65554

Lyu, J., Jiang, M., Zhu, Z., Wu, H., Kang, H., Hao, X., Cheng, S., Guo, H., Shen, X., Wu, T., Chang, J., & Wang, C. (2024). Identification of biomarkers and potential therapeutic targets for pancreatic cancer by proteomic analysis in two prospective cohorts. *Cell Genomics*, *4*(6), 100561. https://doi.org/10.1016/j.xgen.2024.100561

Mälarstig, A., Grassmann, F., Dahl, L., Dimitriou, M., McLeod, D., Gabrielson, M., Smith-Byrne, K., Thomas, C. E., Huang, T.-H., Forsberg, S. K. G., Eriksson, P., Ulfstedt, M., Johansson, M., Sokolov, A. V., Schiöth, H. B., Hall, P., Schwenk, J. M., Czene, K., & Hedman, Å. K. (2023). Evaluation of circulating plasma proteins in breast cancer using Mendelian randomisation. *Nature Communications*, *14*(1), 7680. https://doi.org/10.1038/s41467-023-43485-8

Malki, M. A., Dawed, A. Y., Haywood, C., Doney, A., & Pearson, E. R. (2021). Utilizing large electronic medical record data sets to identify novel drug-gene interactions for commonly used drugs. *Clinical Pharmacology and Therapeutics*, *110*(3), 816–825. https://doi.org/10.1002/cpt.2352

Mayerhofer, E., Malik, R., Parodi, L., Burgess, S., Harloff, A., Dichgans, M., Rosand, J., Anderson, C. D., & Georgakis, M. K. (2022). Genetically predicted on-statin LDL response is associated with higher intracerebral haemorrhage risk. *Brain*, *145*(8), 2677–2686. https://doi.org/10.1093/brain/awac186

McInnes, G., & Altman, R. B. (2021). Drug Response Pharmacogenetics for 200,000 UK Biobank Participants. *Pacific Symposium on Biocomputing*, *26*, 184–195.

McInnes, G., Lavertu, A., Sangkuhl, K., Klein, T. E., Whirl-Carrillo, M., & Altman, R. B. (2021). Pharmacogenetics at scale: An analysis of the UK Biobank. *Clinical Pharmacology and Therapeutics*, *109*(6), 1528–1537. https://doi.org/10.1002/cpt.2122

Mentzer, A. J., Brenner, N., Allen, N., Littlejohns, T. J., Chong, A. Y., Cortes, A., Almond, R., Hill, M., Sheard, S., McVean, G., Board, U. K. B. I. A, Collins, R., Hill, A. V. S., & Waterboer, T. (2022). Identification of host-pathogen-disease relationships using a scalable multiplex serology platform in UK Biobank. *Nature Communications*, *13*(1), 1818. https://doi.org/10.1038/s41467-022-29307-3

Minikel, E. V., Karczewski, K. J., Martin, H. C., Cummings, B. B., Whiffin, N., Rhodes, D., Alföldi, J., Trembath, R. C., van Heel, D. A., Daly, M. J., Alföldi, J., Armean, I. M., Banks, E., Bergelson, L., Cibulskis, K., Collins, R. L., Connolly, K. M., Covarrubias, M., Cummings, B. B., … Genome Aggregation Database, C. (2020). Evaluating drug targets through human loss-of-function genetic variation. *Nature*, *581*(7809), 459–464. https://doi.org/10.1038/s41586-020-2267-z

Minikel, E. V., Painter, J. L., Dong, C. C., & Nelson, M. R. (2024). Refining the impact of genetic evidence on clinical success. *Nature*, *629*(8012), 624–629. https://doi.org/10.1038/s41586-024-07316-0

Mojtahed, A., Kelly, C. J., Herlihy, A. H., Kin, S., Wilman, H. R., McKay, A., Kelly, M., Milanesi, M., Neubauer, S., Thomas, E. L., Bell, J. D., Banerjee, R., & Harisinghani, M. (2019). Reference range of liver corrected T1 values in a population at low risk for fatty liver disease—A UK Biobank sub-study, with an appendix of interesting cases. *Abdominal Radiology*, *44*(1), 72–84. https://doi.org/10.1007/s00261-018-1701-2

NHS England. (2023). Commissioning statement: Ivacaftor, tezacaftor/ivacaftor, lumacaftor/ivacaftor and elexacaftor/tezacaftor/ivacaftor for cystic fibrosis. Retrieved July 12, from https://www.england.nhs.uk/publication/commissioning-statement-ivacaftor-tezacaftor-ivacaftor-lumacaftor-ivacaftor-and-elexacaftor-tezacaftor-ivacaftor-for-cystic-fibrosis/

NICE. (2018). Adjuvant endocrine therapy for invasive breast cancer. Retrieved July 12 from https://www.nice.org.uk/guidance/ng101/chapter/Recommendations#endocrine-therapy

Ouyang, Y., Chen, Y., Chen, K., Tang, Z., Shi, G., Qu, C., Zhang, K., & Yang, H. (2024). Mendelian randomization and colocalization analysis reveal novel drug targets for myasthenia gravis. *Human Genomics*, *18*(1), 43. https://doi.org/10.1186/s40246-024-00607-7

Patrick, M. T., Bardhi, R., Zhou, W., Elder, J. T., Gudjonsson, J. E., & Tsoi, L. C. (2022). Enhanced rare disease mapping for phenome-wide genetic association in the UK Biobank. *Genome Medicine*, *14*(1), 85. https://doi.org/10.1186/s13073-022-01094-y

Pavey, H., Kulkarni, S., Wood, A., Ben-Shlomo, Y., Sever, P., McEniery, C., & Wilkinson, I. (2022). Primary hypertension, antihypertensive medications and the risk of severe COVID-19 in UK Biobank. *PLoS ONE*, *17*(11), e0276781. https://doi.org/10.1371/journal.pone.0276781

Pham, K., Mulugeta, A., Lumsden, A., & Hyppönen, E. (2023). Genetically instrumented LDL-cholesterol lowering and multiple disease outcomes: A Mendelian randomization phenome-wide association study in the UK Biobank. *British Journal of Clinical Pharmacology*, *89*(10), 2992–3004. https://doi.org/10.1111/bcp.15793

Raisi-Estabragh, Z., McCracken, C., Bethell, M. S., Cooper, J., Cooper, C., Caulfield, M. J., Munroe, P. B., Harvey, N. C., & Petersen, S. E. (2020). Greater risk of severe COVID-19 in Black, Asian and Minority Ethnic populations is not explained by cardiometabolic, socioeconomic or behavioural factors, or by 25(OH)-vitamin D status: Study of 1326 cases from the UK Biobank. *Journal of Public Health*, *42*(3), 451–460. https://doi.org/10.1093/pubmed/fdaa095

Regeneron Pharmaceuticals Inc. (2021). Regeneron and AstaZeneca to research, develop and commericialize new small molecule medicines for obesity. Retrieved 24/06 from https://investor.regeneron.com/news-releases/news-release-details/regeneron-and-astrazeneca-research-develop-and-commercialize-new

Richardson, T. G., Leyden, G. M., Wang, Q., Bell, J. A., Elsworth, B., Davey Smith, G., & Holmes, M. V. (2022). Characterising metabolomic signatures of lipid-modifying therapies through drug target Mendelian randomisation. *PLoS Biology*, *20*(2), e3001547. https://doi.org/10.1371/journal.pbio.3001547

Riveros-Mckay, F., Weale, M. E., Moore, R., Selzam, S., Krapohl, E., Sivley, R. M., Tarran, W. A., Sørensen, P., Lachapelle, A. S., Griffiths, J. A., Saffari, A., Deanfield, J., Spencer, C. C. A., Hippisley-Cox, J., Hunter, D. J., O'Sullivan, J. W., Ashley, E. A., Plagnol, V., & Donnelly, P. (2021). Integrated polygenic tool substantially enhances coronary artery disease prediction. *Circulation: Genomic and Precision Medicine*, *14*(2), e003304. https://doi.org/10.1161/CIRCGEN.120.003304

Sadler, M. C., Apostolov, A., Cevallos, C., Ribeiro, D. M., Altman, R. B., and Kutalik, Z. (2024). Leveraging large-scale biobank EHRs to enhance pharmacogenetics of cardiometabolic disease medications. medRxiv 2024.04.06.24305415. https://doi.org/10.1101/2024.04.06.24305415

Sangkuhl, K., Whaley, R., Woon, M., Keat, K., Gong, L., Whirl-Carrillo, M., Ritchie, M., & Klein, T. (2023). P662: Estimating UK biobank population-specific PGx allele and phenotype frequencies using PharmCAT. *Genetics in Medicine Open*, *1*(1), 100727. https://doi.org/10.1016/j.gimo.2023.100727

Schneider, D. J., Lynch, S. A., Gelinas, A. D., Ostroff, R. M., Rohloff, J. C., Williams, P., Janjic, N., & Drolet, D. W. (2022). Chapter 8—SOMAmer reagents and the SomaScan platform: Chemically modified aptamers and their applications in therapeutics, diagnostics, and proteomics. In P. H. Giangrande, V. de Franciscis, & J. J. Rossi (Eds.), *RNA Therapeutics* (pp. 171–260). Academic Press. https://doi.org/10.1016/B978-0-12-821595-1.00007-5

Si, S., Liu, H., Xu, L., & Zhan, S. (2024). Identification of novel therapeutic targets for chronic kidney disease and kidney function by integrating multi-omics proteome with transcriptome. *Genome Medicine*, *16*(1), 84. https://doi.org/10.1186/s13073-024-01356-x

Sirugo, G., Williams, S. M., & Tishkoff, S. A. (2019). The missing diversity in human genetic studies. *Cell*, *177*(1), 26–31. https://doi.org/10.1016/j.cell.2019.02.048

Sorokin, E. P., Basty, N., Whitcher, B., Liu, Y., Bell, J. D., Cohen, R. L., Cule, M., & Thomas, E. L. (2022). Analysis of MRI-derived spleen iron in the UK Biobank identifies genetic variation linked to iron homeostasis and hemolysis. *American Journal of Human Genetics*, *109*(6), 1092–1104. https://doi.org/10.1016/j.ajhg.2022.04.013

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., & Collins, R. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Medicine*, *12*(3), e1001779. https://doi.org/10.1371/journal.pmed.1001779

Sukcharoen, K., Sharp, S. A., Thomas, N. J., Kimmitt, R. A., Harrison, J., Bingham, C., Mozere, M., Weedon, M. N., Tyrrell, J., Barratt, J., Gale, D. P., & Oram, R. A. (2020). IgA nephropathy genetic risk score to estimate the prevalence of IgA nephropathy in UK Biobank. *Kidney International Reports*, *5*(10), 1643–1650. https://doi.org/10.1016/j.ekir.2020.07.012

Sun, B. B., Chiou, J., Traylor, M., Benner, C., Hsu, Y. H., Richardson, T. G., Surendran, P., Mahajan, A., Robins, C., Vasquez-Grinnell, S. G., Hou, L., Kvikstad, E. M., Burren, O. S., Davitte, J., Ferber, K. L., Gillies, C. E., Hedman, A. K., Hu, S., Lin, T., … Whelan, C. D. (2023). Plasma proteomic associations with genetics and health in the UK Biobank. *Nature*, *622*(7982), 329–338. https://doi.org/10.1038/s41586-023-06592-6

Sun, D., Gao, W., Hu, H., & Zhou, S. (2022). Why 90% of clinical drug development fails and how to improve it? *Acta Pharmaceutica Sinica B*, *12*(7), 3049–3062. https://doi.org/10.1016/j.apsb.2022.02.002

Sun, Z., Yun, Z., Lin, J., Sun, X., Wang, Q., Duan, J., Li, C., Zhang, X., Xu, S., Wang, Z., Xiong, X., & Yao, K. (2024). Comprehensive mendelian randomization analysis of plasma proteomics to identify new therapeutic targets for the treatment of coronary heart disease and myocardial infarction. *Journal of Translational Medicine*, *22*(1), 404. https://doi.org/10.1186/s12967-024-05178-8

Szustakowski, J. D., Balasubramanian, S., Kvikstad, E., Khalid, S., Bronson, P. G., Sasson, A., Wong, E., Liu, D., Wade Davis, J., Haefliger, C., Katrina Loomis, A., Mikkilineni, R., Noh, H. J., Wadhawan, S., Bai, X., Hawes, A., Krasheninina, O., Ulloa, R., Lopez, A. E., … Ye, Z. (2021). Advancing human genetics research and drug discovery through exome sequencing of the UK Biobank. *Nature Genetics*, *53*(7), 942–948. https://doi.org/10.1038/s41588-021-00885-0

Tao, T., Mo, X., & Zhao, L. (2024). Identifying novel potential drug targets for endometriosis via plasma proteome screening. *Frontiers in Endocrinology*, *15*, 1416978. https://doi.org/10.3389/fendo.2024.1416978

Türkmen, D., Masoli, J. A. H., Kuo, C. L., Bowden, J., Melzer, D., & Pilling, L. C. (2022). Statin treatment effectiveness and the SLCO1B1*5 reduced function genotype: Long-term outcomes in women and men. *British Journal of Clinical Pharmacology*, *88*(7), 3230–3240. https://doi.org/10.1111/bcp.15245

UK Biobank. (2013). Repeart Assessment (2012–2013). Retrieved July 9, from https://biobank.ndph.ox.ac.uk/showcase/exinfo.cgi?src=repeat_assessment_2012

UK Biobank. 2016. Physical Activity Monitor (Accelerometer). Retrieved July 9, from https://biobank.ndph.ox.ac.uk/showcase/refer.cgi?id=131600

UK Biobank. (2024). UK Biobank Brain Imaging. Retrieved July 9, from https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/brain_mri.pdf

Verweij, N., Haas, M. E., Nielsen, J. B., Sosina, O. A., Kim, M., Akbari, P., De, T., Hindy, G., Bovijn, J., Persaud, T., Miloscio, L., Germino, M., Panagis, L., Watanabe, K., Mbatchou, J., Jones, M., LeBlanc, M., Balasubramanian, S., Lammert, C., … Lotta, L. A. (2022). Germline mutations in CIDEB and protection against liver disease. *The New England Journal of Medicine*, *387*(4), 332–344. https://doi.org/10.1056/NEJMoa2117872

Wang, K., Liu, H., Liu, J., Han, L., Kang, Z., Liang, L., Jiang, S., Meng, N., Chen, P., Xu, Q., Wu, Q., & Hao, Y. (2022). Factors related to irritable bowel syndrome and differences among subtypes: A cross-sectional study in the UK Biobank. *Frontiers in Pharmacology*, *13*, 905564. https://doi.org/10.3389/fphar.2022.905564

Wendt, F. R., Koller, D., Pathak, G. A., Jacoby, D., Miller, E. J., & Polimanti, R. (2021). Biobank scale pharmacogenomics informs the genetic underpinnings of simvastatin use. *Clinical Pharmacology & Therapeutics*, *110*(3), 777–785. https://doi.org/10.1002/cpt.2260

Wilman, H. R., Kelly, M., Garratt, S., Matthews, P. M., Milanesi, M., Herlihy, A., Gyngell, M., Neubauer, S., Bell, J. D., Banerjee, R., & Thomas, E. L. (2017). Characterisation of liver fat in the UK Biobank cohort. *PLoS ONE*, *12*(2), e0172921. https://doi.org/10.1371/journal.pone.0172921

Yuan, W., Xu, W., Xu, X., Qu, B., & Zhao, F. (2024). Exploration of potential novel drug targets for diabetic retinopathy by plasma proteome screening. *Scientific Reports*, *14*(1), 11726. https://doi.org/10.1038/s41598-024-62069-0

Zhang, K., Chidiac, B., Zhao, V., Noordeen, S., Warrier, V., Butterworth, A., Perry, B., Murray, G., & Osimo, E. (2024). Associations of 2,922 genetically predicted plasma protein levels with mental illness and response to treatment. PREPRINT (Version 1) available at Research Square. https://doi.org/10.21203/rs.3.rs-4313341/v1

Zheng, J., Haberland, V., Baird, D., Walker, V., Haycock, P. C., Hurle, M. R., Gutteridge, A., Erola, P., Liu, Y., Luo, S., Robinson, J., Richardson, T. G., Staley, J. R., Elsworth, B., Burgess, S., Sun, B. B., Danesh, J., Runz, H., Maranville, J. C., … Gaunt, T. R. (2020). Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nature Genetics*, *52*(10), 1122–1131. https://doi.org/10.1038/s41588-020-0682-6

Zhou, S., Sosina, O. A., Bovijn, J., Laurent, L., Sharma, V., Akbari, P., Forgetta, V., Jiang, L., Kosmicki, J. A., Banerjee, N., Morris, J. A., Oerton, E., Jones, M., LeBlanc, M. G., Idone, V., Overton, J. D., Reid, J. G., Cantor, M., Abecasis, G. R., … Regeneron Genetics, C. (2023). Converging evidence from exome sequencing and common variants implicates target genes for osteoporosis. *Nature Genetics*, *55*(8), 1277–1287. https://doi.org/10.1038/s41588-023-01444-5

Zhu, J., Liu, H., Gao, R., Gong, R., Wang, J., Zhou, D., Yu, M., & Li, Y. (2024). Genetic-informed proteome-wide scan reveals potential causal plasma proteins for idiopathic pulmonary fibrosis. *Thorax*, *2024*, 221398. https://doi.org/10.1136/thorax-2024-221398

How to cite this article: Bešević, J., Said, S., Nagy, R., Jamshidi, Y., Whelan, C. D., Carson, L., Rutter, M. K., Lewandowski, A. J., Effingham, M., Collins, R., Lacey, B., & Allen, N. E. (2026). UK Biobank: Transforming drug discovery and precision medicine. *British Journal of Pharmacology*, *183*(2), 234–248. https://doi.org/10.1111/bph.70254