Study Protocol

Project title: The IBD BioResource

Full Title: The UK Inflammatory Bowel Disease BioResource: Progressing from Genetics to Function and Clinical Translation in Crohn's Disease & Ulcerative Colitis in adults, young people and children

Contents

1. Introduction

The IBD BioResource has been set up by the UK IBD Genetics Consortium and the NIHR BioResource to collect biological samples and data from a cohort of up to 50,000 individuals with Crohn's disease (CD) or ulcerative colitis (UC). Inflammatory Bowel disease, or IBD, is the collective term for these conditions. There are three arms of recruitment:

1/ The Main cohort comprises individuals with established CD, UC or IBD type unclassified (IBDU). Both clinical and self-reported phenotype data will be collected, alongside plasma, serum and DNA samples for genetic analysis, including whole Genome Sequencing.

2/ The Inception cohort aims to recruit a sub-set of 1,000 individuals recently diagnosed with IBD or suspected of having IBD. More detailed samples including stool, biopsy tissue and whole blood for RNA and longitudinal follow-up of their medical course will be obtained from clinical records and from the subjects themselves.

3/ The paediatric IBD (PIBD) BioResource aims to recruit 800 paediatric IBD patients, including 150 newly diagnosed, to establish an overall resource of 5,000 paediatric onset IBD patients (incorporating adults who have already joined the IBD BioResource but had paediatric onset IBD). Biological samples collected will include blood (or saliva) and biopsies collected during routine clinic appointments, and stool samples completed at home. Longitudinal follow-up of the participants' medical course and surgery information will be obtained from clinical records and from the subjects / parents themselves.

Both panels of adult and paediatric IBD patients will be integrated within the existing cohort of the NIHR BioResource, with its National Coordinating Centre based in Cambridge. The PIBD arm of the IBD BioResource is based in Oxford but is directly aligned to the IBD BioResource and will be operating under a collaborative agreement established between Cambridge University Hospital / CUH (the sponsor for IBD BioResource) and Oxford University.

The NIHR BioResource provides researchers with access to groups of recallable volunteers and / or sample sets, tailor-made to the research question at hand, though not only for IBD research. The goal is to establish a sampling frame and build a panel from which cohorts of patients can be selected on the basis of their genotype, phenotype and/or healthcare records for downstream data access studies, or to be invited for observational or interventional research activities

referred to as 'Stage 2' studies. [see Parkes M and IBD BioResource Investigators. IBD BioResource: an open-access platform of 25 000 patients to accelerate research in Crohn's and Colitis. *Gut*. 2019 Jul 3. pii: gutjnl-2019-318835.] Patients can be invited to up to 8 studies per year with a maximum of 4 of them being face-to-face studies.

Invitation of patients for Stage 2 studies and data access applications are managed by the NIHR BioResource. Both types of studies have to be approved by the NIHR BioResource Scientific Advisory Committee (SAB) prior to inviting patients and by the Data Access Committee prior to granting access to data. Separate ethics approval and consent is required for Stage 2 studies where appropriate.

In addition, DNA, blood, stool and biopsy samples will be used in on-going genetic and genomic analyses by the UK IBD Genetics Consortium (UKIBDGC).

2. Aims of the study

The IBD BioResource aims to support studies looking at how genes and environmental factors influence IBD and response to therapy. By gaining more information on the genes involved in CD and UC, by understanding the differences in their function and integrating information regarding environmental influences including the microbiota, we hope to gain insights regarding:

- causal mechanisms
- potential new therapies
- treatment approaches for IBD to reduce the burden of disease
- potential preventative strategies

By working together with the NIHR BioResource, the IBD BioResource benefits from access to a robust and tested system of re-call for volunteers to take part in medical research studies based on their genotype, phenotype, health, education and social care records.

3. Background

3.1 Genetic prelude

The last 15 years has seen major progress in identifying susceptibility genes for Crohn's disease and ulcerative colitis. The UK IBD Genetics consortium, which is chaired by PI Miles Parkes, has played a key role in many of the pivotal studies

which have been published in a sequence of reports in *Nature* and *Nature Genetics* [1-11].

The discovery of genes that influence risk of IBD and how the disease behaves, and interrogating the functions of these genes and their variants, has a very important role in helping to develop new interventions and treatments for IBD. As a step toward these goals, it is essential to be able to correlate variation in genes with variations in characteristics such as disease behaviour, gene expression and immune response and treatment response.

Classical genetic approaches narrow down areas of association to genes and single nucleotide polymorphisms or haplotypes, but causality can only be inferred if separate functional studies demonstrate an effect of the implicated variant on a key biological pathway known to lead to disease. This is relatively easy to do when rare mutations give rise to extreme phenotypes, but is much harder for variants that are more common and which are neither necessary nor sufficient to cause disease. It is, however, these more common variants that account for much of the inherited risk of common diseases including IBD. Thus, as an adjunct to genetic association studies, it is important to establish studies in which the functional or therapeutic significance of genetic variants can be investigated. When a variant is very common, it may be possible to do this in an unselected population. However, for most variants this would be inefficient: it would be preferable to specifically select the population for functional and clinical studies based on carriage of the variant of interest.

3.2 Integrating successful current BioResources

The NIHR Cambridge BioResource was established in 2005. This now has >17,000 healthy volunteers who have donated a sample of DNA and who are available for recall by genotype and/or phenotype. Members of our IBD research team have used this resource in our studies interrogating the function of IBD risk variants [12-13].

The IBD BioResource described in this application works in collaboration with the NIHR BioResource (with its National Coordinating Centre in Cambridge and a total of eighteen local BioResource centres across England) by integrating a patients with Crohn's disease or ulcerative colitis – such that we have available a large cohort of patients recallable by genotype (or disease phenotype) for future studies, markedly enhancing the efficiency by which we can undertake genotype-stratified IBD studies. The NIHR IBD BioResource aims to further increase the size of its patient cohort - hence increasing the total number of

volunteers that can be invited to take part in research studies and increasing power to study interactions and the impact of low frequency and rare variants.

Recruitment of patients into the NIHR BioResource is covered in the NIHR BioResource (Research Tissue Bank) Protocol, IRAS ref: 313104 (REC Ref 22/EE/0230). The IBD BioResource-associated patient information sheets (PIS) and consent forms (CF) are based on those used by the NIHR BioResource – with some modifications made to incorporate the additional information required for a large cohort of subjects with IBD and also the genetics analysis which is integral to our proposal. The IBD BioResource material makes it clear that the IBD BioResource sits within the NIHR BioResource.

In addition, the current proposal aims to encompass the samples, clinical information, genotyping and sequencing data that have been generated in the course of our earlier work, as part of the core activity of the constituent groups within the UK IBD Genetics Consortium (UKIBDGC). UKIBDGC and IBD BioResource are closely integrated, with substantially overlapping investigators for both branches. The previous IBD genetics studies received ethics committee approval in earlier applications and substantial amendment proposals have been approved to respective REC's (for ref 05/Q0108/355 and 03/5/012) to allow roll-over of samples and data from these studies and collections into the proposed new IBD BioResource study.

3.3 Specifics of the IBD BioResource

Unlike the main NIHR BioResource cohort, the IBD BioResource includes a gene discovery programme as well as more detailed sampling from the adult and paediatric inception cohorts (see below) and many IBD-specific questions on both patient questionnaires and clinical phenotype form. The genetic analysis, which is integral to the proposal and to the work of the UKIBDGC, is more than simply the means by which subjects are identified for stage 2 studies, which is the case for most subjects in the NIHR BioResource programme. Although the ethics issues of consent for genotyping / sequencing for the IBD BioResource are essentially the same as the ethics issues of consent for genotyping / sequencing for IBD gene discovery we feel that the emphasis is different. Furthermore, for our future IBD gene discovery programme we will aim to use DNA samples from the IBD BioResource rather than continuing recruitment of a separate cohort for UKIBDGC studies.

In this protocol, we describe the basis for the IBD BioResource and summarise the principles under which subsequent 'Stage 2' research studies will be

undertaken. However, as with the existing NIHR BioResource of healthy and patient volunteers, the ethical issues of the future functional and clinical studies will be particular to the study in question, and each of these Stage 2 research studies will require its own Ethics Committee application and permission prior to inviting BioResource participants. We are not seeking blanket approval for these studies with this application.

Of note the lead clinician of the Paediatric IBD BioResource, Professor Holm Uhlig is based at Oxford. The set up of paediatric recruiting sites, enrolment process of paediatric patients, study coordination and data monitoring will be undertaken by their coordinator team who will report progress to the IBD BioResource Management Committee. PIBD will use the same databases that adult IBD use for clinical and demographic data storage. All PIBD study related documents (i.e. Site Master files) will be located in Oxford while copies of PIBD consent forms will be stored in Cambridge with the potential of secure file sharing option with Oxford.

A key issue relates to the notification of

- Pharmacogenetic findings – relating to variants which we, and others, are increasingly finding and which correlate with IBD treatment outcomes in terms of efficacy and adverse events
- Additional or secondary findings with potential clinical implications - Our policy in the IBD BioResource is directly aligned to the policy and protocol implemented at Genomics England and which has received REC approval (Ref 14/EE/1112). In formulating this we have been in discussion with Michael Parker, Professor of Bioethics at the University of Oxford and Director and Chair of the Ethics Advisory Group for Genomics England. In accordance with the Genomics England protocol, we will offer potential volunteers the option of receiving feedback (or not) if they carry variants of potential medical significance in a specific list of 'pathogenic' genes which is monitored and updated by Genomics England/NHS England NHS Genomic Medicine Service. The research pathway for monogenic IBD has been specified by a recent position paper on Genomic diagnosis and care co-ordination for monogenic inflammatory bowel disease in children and adults: consensus guideline on behalf of the British Society of Gastroenterology and British Society of Paediatric Gastroenterology, Hepatology and Nutrition (Kammermeier et al. Lancet gastroenterol Hepatol 2023) and endorsed by patient stakeholders.

We recognize that the genetic analysis that we are doing is in a research setting and not an accredited NHS genetics laboratory – and we have formulated our approach accordingly. In particular, information relating to pharmacogenetics will only be used in a research context and in the form of a clinician research survey. Information pertaining to these notifications are clearly defined and explained in both the adult and paediatric PIS.

In addition to functional studies, the IBD BioResource will also prove of benefit in the study of the differential impacts of lifestyle and other environmental factors, including the microbiota, by genotype. Observational studies have been established for investigating gene-lifestyle interaction through case-control studies nested within large cohorts (EPIC-Europe) and cross-sectional quantitative trait studies (e.g. Ely, Fenland). However, the causal inference from these observational studies would be greatly enhanced by evidence of differential response to intervention. An example is the demonstration of an apparent gene-diet interaction for the common Pro12Ala polymorphism in PPARγ [14]. The case is strengthened by this combination of biologically relevant environment-gene interaction plus a demonstrable link between the polymorphism and a discrete disease outcome [13]. Testing the hypothesis of dietary intervention and how genes and environmental factors interact would be greatly enhanced by being able to access a cohort of individuals of known genotype to study.

4. Sample size for added value

A very large sample size is important for a number of reasons, as detailed in the appendix. In summary it will allow detailed evaluation of the contribution of low frequency and rare variants to disease susceptibility (for which sample sizes in the 10's of thousands are required); evaluation of the role of genetics in differences in disease course in IBD overall and in its constituent sub-phenotypes; an assessment of the role of gene-gene interaction (which for variants of modest frequency and low effect size, such as is typical in common disease, requires very large sample sizes) – and an assessment of gene-environment interactions (where similar arguments apply). Currently the IBD BioResource has recruited > 42,000 subjects but there is definite scientific advantage in continuing recruitment. It is also true that many patients with IBD want to participate in this effort and we do not want to deny them that opportunity. One of the key funders for the IBD BioResource is Crohn's and Colitis UK, the national patient support group, who see the IBD BioResource as a key vehicle extending to patients the opportunity to participate in research regardless of where they live.

5. Stage 1 activity – Recruitment to the IBD BioResource panel

5.1 Eligibility criteria

For adult cohorts, volunteers must be 16 years of age or above. For paediatric cohorts, volunteers must be less than 16 years of age There is no exclusion on the basis of race, sex or disease status. Patients with mental and learning disabilities are included, provided they are deemed to have capacity to provide informed consent. However, patients with significant cognitive impairment will be excluded from the study.

Specific eligibility criteria for the adult and paediatric main cohort:

- Patients diagnosed with Crohn's disease,
- Patients diagnosed with UC and all its constituent sub-phenotypes
- IBDU: Patients with indeterminate colitis or IBD type unspecified (IBDU)
- Patients with monogenic IBD

Specific eligibility criteria for the adult and paediatric Inception cohort:

- Patients diagnosed with CD or UC or IBDU within ~12 months;

Or

- Patients suspected of having IBD based on the combination of symptoms, examination findings and baseline investigations

5.2 Route to recruitment

a- Adult cohort

Potential volunteers may become aware of the IBD BioResource through social media platforms (Website, Twitter and Facebook) and given the opportunity to complete a paper or electronic 'Expression of Interest' (EOI; through a database link) to get more information about taking part in the study.

Interested volunteers and/or screened eligible patients will be recruited through their IBD clinical care team in participating hospitals across the UK, either directly on site or remotely (e.g., over the phone, through patient information systems/hospital-based electronic healthcare platforms, by email, by post or through virtual consultations). Some will be among current participants in our UK IBD Genetics Consortium (UKIBDGC) studies while others may have participated in other large IBD studies (e.g., CONSTRUCT, TOPICC, CLARITY etc.).

The majority, however, will be new to IBD research. Many will be recruited in district hospitals. IBD BioResource allows these patients an opportunity to contribute to advancing IBD research. Patient recruitment will be overseen by the NIHR BioResource and supported by research nurses including CRN nurses (the study is an NIHR portfolio study).

Patients willing to join the IBD BioResource will be able to do so through the following routes:

- They will be able to join while attending their routine clinic appointment (most are on at least twice yearly follow-up in out-patients). Potential volunteers will be supplied with a letter of invitation and patient information sheet explaining the study requirements - either through the post ahead of their appointment or at the time of IBD clinic attendance. Where subjects have had sufficient time to consider involvement, and had the chance to discuss this if they wish (IBD clinics in participating hospitals will have a research nurse / CRN nurse available to answer questions re the IBD BioResource), they can be recruited at the time of clinic attendance. For many participants this will be the most convenient route, and the blood sample can be taken at the time of routine clinic bloods.

- Patients will, of course, be free to take the information away with them if they wish to consider things further. If they subsequently decide they are willing to participate they can then either join up at the time of their next clinic appointment; or they can contact their local IBD research team and arrange an appointment.

- For participating hospitals where face-to-face IBD Clinics remain minimal in a post-pandemic era, recruiting and consenting of patients can be made at other hospital locations. These may include but are not limited to: Infusion clinics, Endoscopy departments, Inpatients, Research Centres or other hospital departments where the patient is due to visit.

- We anticipate that many of our existing UKIBDGC recruits, having been notified of the IBD BioResource by post or by email and provided with the patient information sheet, will be willing to provide written or electronic consent. Many of these individuals have been attending our clinics for regular review and management for years, and the vast majority are positively enthusiastic about research that might help others with CD and UC in the future. We already hold extensive phenotype and genotype data

on these individuals which, once they have consented, can be used to populate the IBD/NIHR BioResource database. This will allow the IBD/NIHR BioResource to continue establishing a sizable cohort of patients recallable for Stage 2 studies. Where additional blood samples are required these can be obtained at subsequent clinic visits.

- For certain sites who have large cohorts of UKIBDGC patients and for whom existing holdings of frozen blood or DNA samples and data are present, these may be transferred to the IBD BioResource (following informed consent), without the need for retaking samples unless the quality and viability of these samples has been compromised. A portion of thawed / extracted DNA may be returned to these sites, if required.

- Others, hearing about the IBD BioResource through other means (e.g. word of mouth, social media platforms or posters/flyers displaying information in IBD clinics, GP surgeries or Research Centres) will be able to submit a paper or electronic EOI and send this either to the central team or to their local IBD team. They will then be contacted by email, post or phone, and provided with further information about the study. An appointment will be arranged either via email, phone or virtual consultation with a research nurse to answer questions and then, if willing, sign consent, fill in data collection forms and provide a blood sample.

- Patients for the Inception cohort will be ascertained through a variety of routes which may vary between hospitals according to how services are configured. For example, through triage of clinic referrals; in the out-patient clinic itself; patients admitted with symptomatic flares or known or suspected IBD; or at the time of endoscopy. Patients who are included on the basis of suspected IBD but who turn out not to have IBD on further investigation can remain in the BioResource as healthy controls.

- Initial recruitment of patients for the Inception cohort will be led by clinical staff at participating hospitals who will collect baseline samples and data. The responsibility for following up the patient will then pass to the central coordinator team who will contact the recruited patient by phone or email to collect data at 6-, 12-, 24- and 36-months post consent. In addition, a stool sample collection kit will be sent to patients at 6 months post consent by the central team. Both adult and paediatric IBD

coordinators will contact the nurses 2 months post consent to confirm the patient's diagnosis.

b - Paediatric cohort

Potential paediatric volunteers will be recruited through their paediatric IBD clinical care team in participating hospitals across the UK. The paediatric cohort in the UK IBD BioResource will facilitate a centralised national recallable cohort of at least 800 paediatric IBD patients, including 150 newly diagnosed. Patient recruitment will be overseen by the Oxford co-ordinating centre in close collaboration with the adult IBD BioResource central team and supported by CRN nurses at local sites. The study is supported by Crohn's in Childhood Research Association (CICRA). Patients willing to join the paediatric IBD BioResource will be able to do so through the following routes:

- The local IBD clinical care team will identify potential participants and send by post/email a letter of invitation and the patient information sheet (parent/guardian and appropriate age group PIS), or give it to them at their follow up hospital appointment, explaining what joining the PIBD BioResource involves. The member of the research team will allow them time to consider involvement and give them the opportunity to ask any questions they may have about joining. If the child and parent/guardian are both willing for the child to join (age specific assent criteria and forms), the participant can be recruited at the time of clinic attendance. For many participants this will be the most convenient route as the blood sample (or saliva) can be taken at the time of routine clinic bloods, and participants provided with a stool sample kit to take home. (See section 5.4 Samples, for more information about sample collection)

- Potential participants can, of course, take whatever time they need to consider joining. If they subsequently decide they are willing to participate, they can either join the PIBD BioResource at the time of their next clinic appointment; or they can contact their local IBD team and arrange an appointment.

- Others, hearing about the PIBD BioResource through other means (e.g. word of mouth, social media platforms, poster displaying information in IBD clinics) will be able to submit an 'Expression of Interest' form and send this to their local paediatric IBD team. They will then be sent a formal letter of invitation and the patient information sheet (parent/guardian and age-appropriate form and assent form) and an

appointment made with a member of the research team to discuss the BioResource in greater detail and answer any questions.

- Paediatric patients who are included on the basis of suspected IBD but who turn out not to have IBD on further investigation can remain in the BioResource as healthy controls if they choose.

- We anticipate that many of our existing paediatric onset IBD (now adult) recruits, who will be notified of the IBD BioResource by post or email providing them with the PIS, will be willing to provide written / electronic consent. Many of these individuals have been attending our adult IBD clinics for regular review and management for years, and the vast majority are positively enthusiastic about research that might help others with CD and UC in the future. Sites who have paediatric onset IBD patients and for whom existing holdings of frozen blood or DNA samples and data are present, these may be transferred to the paediatric IBD/NIHR BioResource database (following informed consent), without the need for retaking samples unless the quality and viability of these samples has been compromised. A portion of thawed / extracted DNA may be returned to these sites, if required. This will allow the IBD/NIHR BioResource to rapidly establish a sizable cohort of paediatric IBD patients recallable for Stage 2 studies

c – Informed Consent process

Adult cohort
Participants will sign up to the study after being provided with the PIS and having had the opportunity to ask questions. Patients will be given the option to complete a paper consent form as provided during their clinic attendance consultation or electronically via a database link (i.e. remotely). Patients will be provided a paper or electronic copy of the consent form.

Paediatric cohort
Participants will sign up to the study after being provided with the parent/guardian PIS and age appropriate PIS, having had the opportunity to ask questions. Both the parent/guardian and the child/young person will be asked to sign a paper or electronic version of the consent form (age-appropriate assent form if under 16 years of age). Electronic version of the consent form can be completed through the use of electronic devices on site.

For sites that are provided with a tablet to take e-consent with the participant, the parent and child/young person will read the relevant PIS on paper or on the tablet. The parent/guardian will be offered the option of completing the consent form on paper or on the tablet, and the child/young person will complete a paper assent form. If they complete a paper consent they will be provided with a paper copy, or if e-consent is completed on tablet they will be sent an electronic version of the document.

When the pediatric participant turns 16, they will be invited by email/post to participate in the adult IBD BioResource. They will be sent an adult IBD Participant Information Sheet and invited to complete a Consent Form to confirm their decision to continue. They will also be provided with the option of withdrawing from the PIBD BioResource if they so choose.

5.3 Data collection and storage

a – Questionnaires (adult and paediatric participants)
Participants and parent/guardian of participant (if paediatric) will be asked to fill in a data collection sheet for personal contact details and basic demographic information.

In addition participants and parent/guardians will be asked to complete a health and lifestyle questionnaire either on paper (which they may complete in clinic or take home with them and return in a pre-paid envelope) or via a secure, personalised, online data collection form, a link to which will be sent via email. For PIBD, there may be an option of completing the health and lifestyle questionnaire on electronic devices provided in the clinic. This will include questions relating to their IBD as well as more general fields relating to the NIHR BioResource. In addition to the health and lifestyle questionnaire, participants/parent/guardians will also be invited to complete follow up surveys concerning surgery and medication, on a yearly basis via email to ensure that the IBD BioResource database is up to date.

For the Inception cohort, participants (or parent/guardians if paediatric) will be contacted by email at 6-, 12-, 24- and 36-months post consent or to complete an online Patient Reported Outcome Measure (PROM). This will include questions relating to changes in medication, any surgeries the patient has had since the previous PROM was completed and their experience of their symptoms over the time period. The email will contain a link which will be personal to the individual,

allowing them to access an online data collection tool securely. A follow up email reminder (maximum 2) will be sent if the questionnaires (health and lifestyle, PROM and survey) have not been completed within 2 weeks. A text message reminder will be sent for stool sample collection at 6 months post consent. Some subjects may also be asked to complete an additional PROM at the time of future disease re-evaluation or treatment change; or when undergoing colonoscopy as part of their clinical care.

b - Clinical data

Protocolised clinical phenotype data will be extracted from the hospital record by the clinician or research nurse and entered onto an encrypted system developed through the NIHR BioResource (the NIHR BioResource IT development team has developed the 'front end' and the back-end comprises a MySQL server hosted in the NHS environment by Cambridge University Hospital NHS Foundation Trust and will regularly 'dump' selected data to the NIHR BioResource database). Core fields will include IBD type, location and behaviour, complications, co-morbidities, family history, smoking history, surgical data and drug therapy outcomes. In addition to these IBD related core fields, other more general fields will be included so that this can be incorporated into the NIHR BioResource database allowing for these patients to be re-called for any type of research study approved by the NIHR BioResource Steering Committee. Where possible, teams will be asked to update the phenotype and therapy data, particularly at those selected sites doing biopsies on main cohort participants for example at the time of treatment change.

For the inception cohort there will be protocolised collection of clinical data at the point of consent relating to disease activity, treatment response and development of complications. For patients that were recruited to the IBD BioResource up to September 2021, this will also involve the collection of follow up clinical data at 12-, 24- and 36-months post recruitment. This will be entered onto the encrypted system by the recruiting hospital site.

c - Healthcare records and other databases

NIHR / IBD BioResource is a partner of the Health Data Research UK IBD digital research hub called 'Gut Reaction'. This is led by Cambridge University Hospitals NHS Foundation Trust and is developing methodologies to link routinely collected clinical (e.g. medication), laboratory (e.g. blood results), imaging (e.g. endoscopy and radiology), and histopathology data with NIHR IBD BioResource data (e.g. genomics, clinical and health and lifestyle information) and PROM data from the IBD Registry. In addition to getting health and lifestyle information from

patients and clinical data from our recruiting sites, we will be able to obtain data from hospital records, GP records and centrally held health records where applicable (including but not limited to NHS Digital databases regarding hospital admissions, diagnoses, surgeries and cause of death; Public Health England databases; prescribing databases; national cancer registries and IBD Registry; and data held at the Office for National Statistics). In addition, for the PIBD BioResource, information from social care and educational records will be accessible.

After seeking advice from patient and public involvement events (Patient involvement in Research day – Crohn's & Colitis UK) and from the Gut Reaction Patient Advisory Committee, we have included this information in the PIS to be clear and transparent about it: '...allow us to collect, store and analyse health information about you. This involves accessing existing and future hospital records, GP records and centrally held health records (including but not limited to NHS Digital databases regarding hospital admissions, diagnoses, surgeries and cause of death; Public Health England databases; prescribing databases; national cancer registries and IBD Registry; and data held at the Office for National Statistics such as COVID-19 data for example)

All data will ultimately be uploaded to the secure NIHR BioResource database. From here it will be used for two distinct purposes:

- to enable patient selection and invitation to ethically approved Stage 2 research studies (for example where an investigator needs to access a homogenous group of Crohn's disease patients with a specific sub-phenotype). Invitation to such stage 2 studies will be carried out by the NIHR BioResource.

and

- to align (in a de-personalised format) the phenotype and treatment response data to the analysis of, for example, DNA and microbial sequencing data - to identify relevant associations and correlations. The IBD/NIHR BioResource is adopting the "5 Safes" model for data access – that allows access to data for approved researchers, with approved projects, datasets, analysis settings and outputs. While limited data downloads will still be available in heavily de-personalised data, approved researchers will also be able to access and analyse the

requested de-personalised data through a secure, on-line trustworthy research environment (TRE).

All genotype and sequence data is stored in de-personalised form at the Wellcome Trust Sanger Institute in their secure compute farm or secure cloud storage. The NIHR BioResource and Wellcome Trust Sanger Institute IT teams are working together to enable efficient and secure interrogation by the NIHR BioResource team of IBD BioResource genotypes. For our IBD cohort we currently hold GWAS data with 1,000 Genomes imputation at >8,000,000 SNPs on ~15,000 subjects, whole genome sequence data on ~4,000 patients and whole exome sequence on a further 32,000 subjects. IBD BioResource database and all previous UKIBDGC GWAS and iChip data will be deposited to the European Genome-Phenome Archive (EGA).

d- National Genomic Research Library

The National Genomic Research Library is a secure national database of de-identified genomic and health data that is managed by Genomics England. Approved researchers can use the samples and data in a de-personalised form. The adult NIHR IBD BioResource Participant Information Sheet includes a paragraph describing the National Genome Research Library and includes an optional consent box for the sharing of genomic data through this platform.

5.4 Samples:

All subjects will be asked to provide blood samples (adult participants 15-20mls; paediatric participants 7-14mls, depending on size and age) or saliva (if blood not obtainable) at IBD BioResource enrolment - for DNA and serum/plasma extraction.

Subjects in the adult inception cohort and all paediatric participants will, in addition, be asked to provide stool samples (for microbiota analysis), additional blood samples (for RNA extraction), and, if they are undergoing colonoscopy as part of their clinical care, additional biopsies (up to 6 additional biopsies from up to 2 sites) for 'omics' analyses.

In selected sites main cohort subjects may also be asked to provide additional blood samples (10ml) and stool samples at the time of future disease re-evaluation or treatment change; and, if they are undergoing colonoscopy as part of their clinical care, up to 6 additional biopsies from up to 2 sites (total 12 extra biopsies) for 'omics' analyses. Typically at IBD colonoscopy up to 24 biopsies are routinely taken for histopathological assessment. Biopsy samples (in RNA later

or frozen) will be submitted to the Wellcome Sanger Institute in de-personalised form via Royal Mail or courier. For some paediatric participants e.g. with rare immune or epithelial problems who have an endoscopy (gastroscopy or colonoscopy), the biopsy and associated blood samples will be collected at local sites and either processed (fresh/frozen/storage medium) for approved studies and/or stored in the Oxford GI Biobank.

There is a possibility that additional samples may need to be taken if the original sample is depleted to the point that it is not possible to extract DNA. These sampling activities will be coordinated by the research nurses at each recruitment site and only when required.

The Cambridge-based study coordinators and the Oxford-based study coordinator will be responsible for the collection of additional stool samples at 6 months post consent for the adult (newly diagnosed patients only) and paediatric cohorts (all patients), respectively.

All samples will be taken and labelled with a unique barcode for de-personalisation. Blood samples for DNA and serum extraction will be submitted either to the local NIHR BioResource (where the hospital is affiliated with such a centre) – or to the National Biosample Centre (NBC) at Milton Keynes or an accredited 3rd party lab, from where they will be forwarded (after processing) to NBC. Samples will be processed using established SOP's for DNA and serum extraction. Genotyping or sequencing of the DNA samples will be required for downstream genetic association studies run by UKIBDGC or prior to subsequent Stage 2 research studies led by other investigators. Stage 2 studies involving participant recall have to be previously ethically approved and then reviewed and approved by the NIHR BioResource Steering Committee.

For the adult inception cohort and all paediatric participants, stool samples (in a postage-compatible preserving buffer kit such as Omni-gene GUT), and biopsy samples (in RNA later or frozen) will be submitted to the University of Edinburgh Wellcome Trust Clinical Research Facility in de-personalised form via Royal Mail or courier. For specific projects that involve functional studies, organoid work or single cell studies, biopsies and/or blood will be stored in specific freezing medium and stored at dry ice, -80 freezer or liquid nitrogen prior to shipment to the Wellcome Sanger Institute for processing and analysis.
For subprojects approved by the PIBD BioResource Steering committee, some biological samples may be processed or stored and distributed by local sites for subsequent analysis, according to their HTA licence. For instance, in paediatric participants with rare immune or epithelial problems, biopsy samples and/or

bloods will be transferred to the Oxford GI Biobank, where they will be stored and processed in accordance with the Oxford GI BioBank protocol (REC reference: 21/YH/0206).
Stool samples will be processed for extraction of DNA and RNA, with RNA converted to cDNA and all samples barcoded and stored at -80 in a dedicated freezer prior to analysis using meta-genomic and 16s ribosomal RNA sequencing.

5.5 Reporting and feedback of results

We are planning to undertake a programme of whole genome and exome sequencing using the samples in the IBD BioResource, to complete our interrogation of the contribution of common, rare and structural variants to IBD susceptibility. Where necessary (for example, pending the availability of the sequence data) additional genotyping may be done prior to a Stage 2 research study to identify potential participants to take part in the subsequent study. It is not planned to routinely feedback the results from genetic or other biochemical tests obtained from donated samples, but note comments above (section 3.3) regarding detection and handling of further genetic findings that may be medically actionable. Where 'incidental' findings are made of medically significant, treatable conditions then – if the subject has opted for feedback - the supervising medical team will be contacted by the IBD BioResource team with the suggestion to arrange formal testing through an accredited NHS diagnostics laboratory with subsequent relay of results by an experienced clinician or genetics counsellor.

Specify monogenic IBD Variant feedback
In rare instances, we will offer patients feedback of genetic variants with a likely strong impact on their disease severity and/or treatment (monogenic IBD; likely pathogenic or pathogenic variants in actionable monogenic IBD genes, after review of a steering group, to be confirmed and followed up in the NHS setting).

We will also study specific pharmacogenetic variants known to correlate with IBD treatment outcomes in terms of efficacy and adverse events. These are most relevant to clinicians who are about to initiate new therapies for IBD. Alongside assessing the impact of these variants on outcomes we will survey the clinicians to assess their views about the clinical utility of knowing about such pharmacogenetic variants and whether this may affect their prescribing practice. Formal testing of pharmacogenetics variants will be sought through an NHS accredited diagnostics lab.

5.6 Participant safety

During the establishment of the IBD BioResource, there is minimal potential for harm to volunteers as blood draw or (for the adult inception cohort and all paediatric participants) the taking of additional biopsies at the time of clinically indicated colonoscopy/endoscopy are the only interventions proposed and both are extremely safe. Each individual recall 'Stage 2' study will require a separate protocol and separate ethics approval.

6. The IBD BioResource funding and Organisational Structure

The IBD BioResource is funded by the Medical Research Council with additional support from the National Institute for Health Research (NIHR), the Wellcome Trust, Crohn's and Colitis UK, the Helmsley Charitable Trust and Industry partners. Crohn's in Childhood Research Association (CICRA) awarded the grant for Translational Science in Paediatric Inflammatory Bowel Disease - The UK IBD BioResource (PIBD BioResource).

Both the IBD and PIBD BioResources sit within the NIHR BioResource. The IBD BioResource has its own management committee, meeting every 2 weeks and co-chaired by Prof. Miles Parkes (Consultant Gastroenterologist in Cambridge). The PIBD BioResource will have a committee team who will review research proposals put forward by PIs at local sites, in conjunction with the NIHR Steering Committee.

We maintain a close working relationship with the NIHR BioResource central team based in Cambridge (i.e. the NIHR BioResource National Coordinating Centre). The NIHR BioResource is a federation of currently eighteen local NIHR BioResources, including the NIHR BioResource Centre Cambridge, and has its headquarters (i.e. National Coordinating Centre) in Cambridge. Each local NIHR BioResource Centre has its internal organisation, including a Scientific Advisory Board which reviews Researchers' applications for local Stage 2 studies.

The local BioResource Centres recruit participants using the national NIHR BioResource (Research Tissue Bank - RTB) ethics and study protocol and its governance structure. Stage 2 study applications can also be submitted by researchers directly to the national NIHR BioResource (RTB) for non-local studies, where they are reviewed by the NIHR BioResource Steering Committee. A local BioResource centre may recommend to researchers applying to it to re-apply to the central/national NIHR BioResource for their

Stage 2 proposed study, where it may not be feasible for that study to be run locally (e.g. if the local cohort with the inclusion criteria in question is too small). IBD clinics based in the same institutions as one of the eighteen current NIHR BioResource centres may, where appropriate, feed their samples into their local NIHR BioResource facility, and these samples may be processed and stored locally; the remainder of the country will post their samples to an accredited 3rd party lab. Ultimately all samples will be moved to an approved NIHR Biosample Centre.

Although it is not directly within the remit of this particular application, we summarise below how the BioResource is organised in relation to future studies:

7. Stage 2 activity-Research studies involving BioResource volunteers and/or their data

The IBD BioResource panel is open to access from researchers with ethically approved research studies. Any researcher wishing to utilise the BioResource will complete an application form indicating the phenotype or gene variant(s) of interest, it's/their likely frequency and information about what is known about its/their functional significance and association with disease. They will also summarise the nature of the experiment they propose.

These stage 2 studies may involve a range of possible options – for example

- access to specific sequence data on a specific subset of samples for detailed analysis without the need to recall participants or collection of additional clinical material
- the participation of patients recalled by genotype and/or phenotype for extra sampling e.g. the taking of one or more blood or stool samples,
- more intensive physical examinations or analyses through to potentially clinical interventions. The latter will require the study leads to set up separate clinical trials agreements with the home hospital of the BioResource volunteer (many of IBD BioResource recruiting centres have extensive experience and active clinical trials programmes).

The NIHR BioResource team will handle the process of recruitment of individuals to each new Stage 2 study.

For studies requiring recall of patients by genotype and/or phenotype, in accordance with the study requirements sequencing and or genotyping or serum sample assay will be carried out by the NIHR BioResource on all or the relevant subset of IBD samples to identify those subjects meeting genotype- or

biomarker- specific inclusion criteria. Having been identified samples carrying the relevant genotypes (usually those who are homozygous for risk genotypes and those homozygous for non-risk at specific locus of interest) or serotypes a collective list of barcodes will be passed to the BioResource volunteer administration team (who will not link e.g. risk / non-risk genotypes to identifiers). These potential volunteers will be invited to participate in the specific Stage 2 study for which they have been identified (due to carrying relevant genotypes or serological markers) by way of an invitation letter or email outlining details of the study. They will also be sent the approved PIS for the Stage 2 study together with any other relevant study information (e.g. questionnaires). It is entirely up to them whether they wish to participate in any given Stage 2 study. The BioResource team, by controlling the process of recruitment, will be in a position to regulate the number of invitations any individual receives, and the number of recall studies that volunteers participate in each year.

Patients who are willing to participate in a particular Stage 2 study for which they are eligible, and to which they have been invited, will contact the NIHR BioResource team and an appointment will be arranged for them to attend for sampling. To protect volunteer anonymity, where the study design does not require the researcher to have direct contact with the participants and the recall can be facilitated by the BioResource, all blood sampling will be carried out by trained nurses or phlebotomists working under the direction of the NIHR BioResource. This will happen in venues appropriate for venepuncture (usually an NIHR BioResource or a local Clinical Research Facility). Informed consent will be always be taken prior to further blood sampling for Stage 2 studies – using the study-specific consent form and information sheet. Copies will be kept with volunteers' records by the BioResource team with an anonymised proof of consent form being passed on to the researcher for their records to show evidence that consent has been obtained for each sample they receive.

The researcher will conduct the study blind to genotype and only receive genotype information at the end of the study when they have completed all measurements. The genotype to phenotype link will be made without individual identifiers. In this way, individual identity and genotype will be kept separate. Also the blinding ensures that there is no bias, either in the genotyping or in the functional study.

Anonymity of IBD BioResource subjects will be maintained where possible and the approved Stage 2 study design allows for it; but e.g. for intervention studies it will be necessary for the team directly undertaking the research to meet with

IBD BioResource subjects. As per all participant-facing Stage 2 studies, all such work will require its own ethics application and approval; and all such work will be undertaken by physicians working to GCP principles.

8. Appendix: detailed rationale for sample size

In order to ensure sufficient numbers of participants homozygous for low frequency alleles or possessing combinations of alleles can be recruited into any future stage 2 study, the size of the BioResource is crucial. We estimate that it is realistic to recruit a minimum 50,000 individuals and that this will provide adequate power for a broad range of Stage 2 studies.

This size of BioResource will benefit a range of future studies and applications that we anticipate for the cohort. Examples of these are given below.

1. through increased sample size and sampling from non-European origin groups aid fine mapping and identification of causal variants in >300 confirmed IBD GWAS loci - work by UKIBDGC + partners in the International IBD Genetics Consortium
2. significantly increase power for follow-up of low-frequency and rare variants identified by our current IBD whole genome sequencing programme at the Sanger Institute (Barrett and Anderson with UKIBDGC)
3. provide expanded cohorts of IBD patients for testing and validation of markers identified in current UKIBDGC pharmacogenetic discovery programme and allow for predicting therapy outcome
4. allow selection of IBD patients by genotype or specific IBD phenotype for study of IBD pathophysiology using primary human cells. The functional effects of IBD-associated variants and their environmental modifiers will be further characterized, including on pathways relating to autophagy, ER stress, microbial sensing and IL23 response. The importance of studying primary human cells of the appropriate subset is emphasized by the cell-type specificity of eQTL effects (and indeed the prevalence of inverse eQTL effects in different cell types) [16]. Several investigators allied to UKIBDGC are undertaking such work including: Simmons / Powrie, Oxford; Lamb/Mansfield, Newcastle; Lee/Kaser, Cambridge, and PIBD bioresource.
5. the functional effects of IBD-associated variants on gene expression will also be studied using single cell sequencing on biopsies and blood samples from IBD BioResource participants (collaborative study with

Wellcome Sanger Institute); and the correlation with treatment response will be ascertained.

6. provide cohorts of patients for further development and validation of biomarkers of disease course / stratification based on GWAS, transcriptomic (see inception cohort below) and proteomic analysis. Glean biological insights from their functional interrogation in primary human cells; and lead into biomarker stratified intervention studies.
   This would build on existing work by Lee/Parkes/Smith, Cambridge [17] and provide a validation cohort for the many stratification/biomarker studies planned or occurring in the UK and elsewhere (including PANTS Ahmad, Exeter and Hart, Imperial; IBD-BIOM and IBD-CHARACTER Satsangi, Edinburgh)
7. enable *in vitro* testing of novel/re-purposed drugs to interrogate their effect on relevant IBD pathogenic pathways stratified by genotype. This work, undertaken in collaboration with industrial partners, will precede proof of concept clinical testing. IBD BioResource will be an important channel of communication with patients, and will aid recruitment to subsequent experimental and stratified medicine studies in Crohn's disease and ulcerative colitis.
8. provide a platform to study the role of gut microbiota in IBD pathogenesis. Distinguishing cause vs correlation is complex, given the many confounders in patients who already have intestinal inflammation. At present UKIBDGC studies here are at an early stage, comprising:
   - an existing (Crohn's & Colitis Foundation of America funded) collaboration between Parkes and Virgin, Washington University, St Louis, of patients in early IBD flare for stool sampling, viral enrichment, Illumina HiSeq and bioinformatic analysis to detect viral triggers. We have published early findings [18] but need to extend the size of the sample set studied.
   - a pilot study comparing microbial composition of stool with samples obtained from the ileal and colonic mucosa at colonoscopy (Hart, Imperial; Hold, Aberdeen; Jostins/ Barrett, Walker/ Parkhill, Sanger) as a prelude to future specific IBD / microbiome grant application. A future 'stage 2' study might for example target stool collection <48 hours of flare, for 16S rRNA sequencing or metagenomic methods. This study is critical for identifying microbial triggers of IBD flares, but is impossible without cohorts of patients recruited prospectively, and primed with immediate access to research teams in the event of a flare.

   - A new study – IBD-RESPONSE – which will look at the impact of the gut microbiota on response to biologic therapies in ~1300 patients with IBD. All participants will also be invited to join the IBD BioResource to enable long term data collection and incorporation of host genetic data into the prediction modelling of treatment response.
9. use the inception cohort of 1,000 individuals under investigation for IBD (some of whom will be newly diagnosed with IBD) – an important component of the IBD BioResource for providing more detailed samples (including stool + whole blood for RNA, alongside DNA and serum) unconfounded by drug treatment or effects of surgery with which to identify biomarkers of disease severity and course, and microbial triggers of disease flares in individuals newly diagnosed with IBD. We aim for protocolised longitudinal collection of a minimum of 2 years clinical data (including outcomes and response to therapies) and repeat sampling of stool.

Given the range of potential applications and experiments, the goal of recruiting 1,000 individuals under investigation for IBD reflects a balance of feasibility and adequate statistical power for an initial set of studies (the successful NIHR Cambridge BioResource of >17,000 volunteers initially set the same target). Where possible we will expand the sample size further, boosting the size of important phenotypic or genotypic strata to allow well powered analyses of disease subsets. Statistical power otherwise tends to be limiting in many sub-phenotype / stratified analyses. For example in 1,000 IBD patients we would anticipate ~500 UC and ~500 CD, the latter comprising ~30% ileal (n=150), 20% colonic (n=100) and 50% ileo-colonic (n=500) – important given evidence emerging from our work in the International IBD Genetics Consortium that ileal and colonic CD have distinct pathogenesis [17].
The following are examples of existing or potential future studies, which would benefit from the availability of the inception cohort. Such an inception cohort would:

1. provide accurate prospective clinical data to enhance phenotype-genotype analyses, including response to drug therapies for pharmacogenetic analysis (UKIBDGC core). In this respect it will provide both a valuable validation panel for existing programmes and a platform for a new series of studies. In 1,000 cases we would expect
   - ~450 UC patients to be treated with mesalazine; 50% remission
   - ~400 IBD patients with thiopurine; 30% remission
   - ~200 CD patients with anti-TNF; 40% remission

Hence a study seeking a proteomic or transcriptomic biomarker of thiopurine efficacy could access inception samples from ~120 subjects who gained remission after treatment. This would, for example, provide 95% power at P<0.01 to identify a biomarker with frequency 30% correlating with an OR of 3 for remission.

2. deliver clinical data and bloods unconfounded by the effects of chronic disease or drug treatments, for epigenetic analysis (Satsangi, Edinburgh; Parkes, Cambridge); and for proteomic and immunological analyses – with the option to resample the same individuals to obtain fresh material for cell-based studies (Simmons, Oxford; Kaser, Cambridge).
3. deliver biometric data unconfounded by the effects of chronic disease or drug treatments for epidemiologic analyses (Hawkey, Nottingham);
4. deliver whole blood RNA + clinical outcomes data for development / validation of transcriptomic biomarkers (Lee/Smith, Cambridge)
5. include environmental data for genetic interaction analysis (Barrett/Anderson, Sanger)

It would provide a unique resource of stool samples to detect microbial triggers of IBD. Previous work suggests that, unlike *Helicobacter pylori* in peptic ulcer, IBD is not caused by a single microbial species. However, early studies have been underpowered to detect low abundance taxa such as are known to trigger IBD in animal models, and none have sampled large cohorts at diagnosis. Our inception cohort would be well powered and provide samples un-confounded by the effects of drugs, surgery or changes in diet. Assuming access to the NIHR Cambridge BioResource of healthy controls, study of our population would have 80% power to detect associations between common (>25%) microbial elements with an OR=1.4, and rare elements (<1% frequency) with OR >3, at p<0.01. Through longitudinal sampling the CD inception cohort of 500 cases would provide 96% power to detect a microbe present in 2% of cases in remission rising to 10% (an OR of 5) on significant relapse at p<0.01.

9. References

1. Parkes, M., et al., *Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility.* Nat Genet, 2007. **39**(7): p. 830-2.
2. Wellcome Trust Case Control Consortium., *Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.* Nature, 2007. **447**(7145): p. 661-78.

3. Barrett, J.C., et al., *Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease.* Nat Genet, 2008. **40**(8): p. 955-62.
4. UK IBD Genetics Consortium and WTCCC., *Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the HNF4A region.* Nat Genet, 2009. **41**(12): p. 1330-4.
5. Fisher, S.A., et al., *Genetic determinants of ulcerative colitis include the ECM1 locus and five loci implicated in Crohn's disease.* Nat Genet, 2008. **40**(6): p. 710-2.
6. Franke, A., et al., *Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci.* Nat Genet, 2010. **42**(12): p. 1118-25.
7. Jostins, L., et al., *Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease.* Nature, 2012. **491**(7422): p. 119-24.
8 Huang H et al. Fine mapping IBD loci to single variant resolution. Nature Jul 2017; 547 (7662); 173-178
9. Lee JC, et al. Genome-wide association study reveals distinct genetic contributions to prognosis and susceptibility in Crohn's disease. *Nature Genetics*. 2017 Feb;49(2):262-268
10. Luo Y, et al. Exploring the genetic architecture of inflammatory bowel disease by whole-genome sequencing identifies association at ADCY7. *Nature Genetics*. 2017 Feb;49(2):186-192
11. de Lange KM, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nature Genetics*. 2017 Feb;49(2):256-261.
12. Ellinghaus, D., et al., *Association between variants of PRDM1 and NDP52 and Crohn's disease, based on exome sequencing and functional studies.* Gastroenterology, 2013. **145**(2): p. 339-47.
13. Lee, J.C., et al., *Human SNP links differential outcomes in inflammatory and infectious disease to a FOXO3-regulated pathway.* Cell, 2013. **155**(1): p. 57-69.
14. Barroso, I., et al., *Dominant negative mutations in human PPARgamma associated with severe insulin resistance, diabetes mellitus and hypertension.* Nature, 1999. **402**(6764): p. 880-3.
15. Altshuler, D., et al., *The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes.* Nat Genet, 2000. **26**(1): p. 76-80.

16. Fairfax, B.P., et al., *Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles.* Nat Genet, 2012. **44**(5): p. 502-10.
17. Lee, J.C., et al., *Gene expression profiling of CD8+ T cells predicts prognosis in patients with Crohn disease and ulcerative colitis.* J Clin Invest, 2011. **121**(10): p. 4170-9.
18. Norman, J.M., et al., *Disease-specific alterations in the enteric virome in inflammatory bowel disease.* Cell, 2015. **160**(3): p. 447-60.
19. Cleynen, I., et al., *Inherited determinants of Crohn's disease and ulcerative colitis phenotypes: a genetic association study.* Lancet, 2016. **387**(10014): p. 156-67.