Genomics england

The 100,000 Genomes Project Protocol

January 2017

Issued and approved by the Chief Scientist for Genomics England

Genomics england

About this document

This document sets out the protocol for the development, delivery, and operation of the 100,000 Genomes Project. It also details the patient and clinical benefits, the scientific and transformational objectives, the implementation strategy, as well as the ethical and governance frameworks required for the Project.

For more information please visit **www.genomicsengland.co.uk**

Authors: Mark Caulfield, Jim Davies, Martin Dennys, Leila Elbahy, Tom Fowler, Sue Hill, Tim Hubbard, Luke Jostins, Nick Maltby, Jeanna Mahon-Pearson, Gil McVean, Katrina Nevin-Ridley, Matthew Parker, Vivienne Parry, Augusto Rendon, Laura Riley, Clare Turnbull, Kerrie Woods.

With input from a Northern Ireland perspective from:
Shane McKee, Anne Moffatt, Julie McCarroll

With international and national peer review from:
The Genomics England Rare Disease, Cancer and Sequencing and Annotation Working Groups, the Ethics Advisory Committee, the Science Advisory Committee, and NHS England.

This document (v3.0) received a favourable ethical opinion from the HRA East of England - Cambridge South Research Ethics Committee (REC) on 3rd February 2017. Version 2.0, which this document replaces, was REC approved on 19th February 2015.

NOTE: Health and Social Care (HSC) in Northern Ireland (HSCNI) is the designation of the publicly funded services providing public health and social care services in Northern Ireland. HSC is delivered by a number of organisations including the Public Health Agency (PHA) and a number of health and social care trusts (HSC Trusts). Where the term "NHS" is used to refer to the National Health Service, where applicable this should be regarded as including HSCNI, unless otherwise stated.

Contents

Genomics england

1 Summary and background

1 Summary and background

1.1 The 100,000 Genomes Project

In December 2012, the Prime Minister announced a programme of whole genome sequencing (WGS) as part of the UK Government's Life Sciences Strategy. The principal objective of the 100,000 Genomes Project is to sequence 100,000 genomes from patients with cancer, rare disorders, and infectious disease, and to link the sequence data to a standardised, extensible account of diagnosis, treatment, and outcomes.

The Project is designed to produce new capability and capacity for genomic medicine that will transform the NHS. It will also produce new capability for clinical genomics research. As part of the proposal a secure infrastructure will be established for the protection and analysis of clinical and genomic data. This will be made available for approved academic and industrial research purposes, including those of the contributing clinical organisations from the NHS.[1]

1.2 Genomics England

The Department of Health has established Genomics England as a wholly owned, limited company to deliver the project. Genomics England is working with NHS England (NHSE), Public Health England (PHE), Health Education England (HEE), NHS Trusts, the Northern Ireland Department of Health (DoH NI) and a number of HSCNI organisations. This is to ensure that the project is fully aligned with NHS transformation and sits within a programme of related initiatives in clinical and laboratory genetics, molecular pathology, service innovation, disease registration, clinical audit, training, and technology funding. To identify and enrol participants we have created NHS Genomic Medicine Centres (GMCs).
NHS England has commissed a number of these centres to harness the capability and capacity of the NHS across England to contribute to the Project between 2015 and 2017. The DoH NI, in partnership with the Medical Research Council (MRC), has commissioned the Belfast Health and Social Care Trust (BHSCT) [acting as the delivery entity for a Northern Ireland Genomic Medicine Centre (NIGMC)] to facilitate recruitment of patients in Northern Ireland to the 100,000 Genomes Project.

Genomics England is also working with research groups and funding organisations, to ensure that the new research capability will be fit for purpose and that the data is acquired and managed to appropriate standards. In addition, Genomics England, and its partners will ensure that the tools provided within the secure infrastructure will both accelerate scientific progress and support the focused, interdisciplinary collaboration needed for clinical interpretation and patient benefit. To maximise the value of this programme we have created the Genomics England Clinical Interpretation Partnership (GeCIP) which brings funders, researchers, NHS teams, trainees and potentially industrial partners together to enhance the value of this dataset for healthcare benefit.[1]

Our mission is focused on combining clinical and whole genome sequencing data in rare disease, cancer, and infection. We have developed this infrastructure with the opportunity to expand the Project to other diseases and approaches if separate funding is generated. In particular, our Medical Research Council-funded Data Infrastructure Award is designed to enhance the UK clinical research infrastructure in Genomic Medicine and could be used to store whole genome sequence from other disorders and a range of associated multi-omic datasets. Furthermore, our sequencing contract contains some elasticity to go beyond 100,000 whole genome sequences although we do not have funding for this additional activity at present.

1.3 Whole Genome Sequencing (WGS)

Next generation sequencing using massively parallel sequencing has advanced our understanding of the genetic architecture of disease. To date the focus has been upon exome sequencing, looking at the 1 percent of the genome that codes for proteins. This has resulted in an increased understanding of some of the genetic variants for cancer and rare diseases. As the cost of WGS is falling, it may change the type of genomic investigations undertaken in the NHS and in the research arena. In partnership with the Wellcome Trust, the Wellcome Trust Sanger Institute and Illumina, we are creating the NHS Genomics Medicine Sequencing Centre in Hinxton which will undertake WGS for the main programme. This offers real opportunities for further progress based upon:

- Detection of variants outside exons and a better representation of those inside exons.
- Improved detection of variants on intron/exon boundaries.
- Detection of an enhanced repertoire of variation including insertion and deletions (indels), copy number variants, and structural changes across the genome.
- Identification of variation that may encode other diseases.
- The cost of WGS is falling to the point where the value of a whole genome sequence in terms of information gained may make it more cost-effective than to conduct targeted sequencing.

1.4 About this document

This document sets out the protocol for the development, delivery, and operation of the 100,000 Genomes Project, setting out the patient and clinical benefit, scientific and transformational objectives, the implementation strategy, and the ethical and governance frameworks required.

Genomics england

2 The 100,000 Genomes Project design

2 The 100,000 Genomes Project design

2.1 Project aims

The aims of the project are:

- **Patient benefit:** providing clinical diagnosis and in time, new or more effective treatments for NHS patients.
- **New scientific insights and discovery:** with the consent of patients, creating a database of 100,000 whole genome sequences linked to continually updated long term patient health and personal information for analysis by researchers.
- **Accelerating the uptake of genomic medicine in the NHS:** working with NHSE and other partners to deliver a scaleable WGS and informatics platform to enable these services to be made widely available for NHS patients. In addition, through the Genomics England Clinical Interpretation Partnership (GeCIP), creating a mechanism to both continually improve the accuracy and reliability of information fed back to patients and add to knowledge of the genetic basis of disease.
- **Stimulating and enhancing UK industry and investment:** by providing access to this unique data resource by industry for the purpose of developing new knowledge, methods of analysis, medicines, diagnostics and devices.
- **Increasing public knowledge and support for genomic medicine:** delivering an ethical and transparent programme which has public trust and confidence and working with a range of partners to increase knowledge of genomics.

Rare diseases, cancer and infectious diseases were selected as the focus for the 100,000 Genomes Project as they present high potential for significant health gain from this Project. Focus on these diseases offers the strongest prospect of patient and scientific benefits and the ability to drive the transformation of the NHS in terms of application of genomic medicine. Furthermore, given the current state of knowledge regarding the genetic architecture of these diseases, the application of WGS may enable major new biological insights that will enable new diagnostics and therapeutic innovation.

In preparation for the Project, the Department of Health commissioned a series of Advisory Committee reports (see **www.genomicsengland.co.uk**). These have informed and assisted Genomics England as we established the pilots and have influenced, with expert advice, the design of the main programme.

Genomics england

3 The structure of the 100,000 Genomes Project

3 The structure of the 100,000 Genomes Project

At inception, Genomics England recognised the need for a series of preparatory pilot studies structured to inform decisions for the main project. The project has three phases:

3.1 Phase 1: Sequencing performance, quality and annotation (2014)

Phase 1a: a competitive evaluation of performance, capability, quality, and timeliness to generate whole genome sequence. Several providers entered this competition. Each was provided with five pairs of tumour and germline samples and a separate five germline DNA samples from patients with breast, thyroid, and testicular cancer. The results showed that there were at least two providers with the potential to offer whole genome sequencing (WGS), at sufficient quality and coverage for the main project.

Phase 1b: evaluated the performance and quality of annotation services via an open competition. 29 providers used the read-level data from cancers in Phase 1a and from five rare disease parent-offspring trios to demonstrate their capability to produce high quality variant calling and/or annotated and interpreted variants. The results showed that several providers were able to identify potential pathogenic variants in rare disease patients. In cancer, the ability to call somatic mutations was variable, and the reports were less consistent.

3.2 Phase 2: Pilot programmes (2014-15)

Rare disease pilot: a pilot programme in rare diseases was established in partnership with the National Institute for Health Research (NIHR) BioResource for Rare Diseases, recruiting 2000 people from families with rare diseases. In this pilot, we aimed to recruit the ideal family structure to identify causative gene variants for rare disease. This is a parent-offspring trio based upon two parents and one affected offspring (proband). Other family structures have been included but only in certain disorders, such as, duo family structures (one parent and one offspring) and singletons for disorders of late onset (which will be considered if trios cannot be collected). In this case, there must be a suitable number of family units available with the same phenotype to enhance detection of a pathogenic variant. Samples and data are being collected along with participant phenotypes and the results of DNA sequencing will be released in early 2015. This will be accompanied by NHS clinical feedback and data release via the Genomics England Data Centre.

Cancer pilot: a pilot programme in cancer was established in partnership with Cancer Research UK. This was following ethical approval data and samples from 3000 patients with lung, breast, colon, prostate, and ovarian cancer. These are being collected through centres participating in the Cancer Research UK Stratified Medicines Programme.

Biomedical Research Centres (BRC) cancer pilot: a third pilot programme has been established in collaboration with the National Institute for Health Research Biomedical Research Centres at Oxford, Cambridge, Guys and St Thomas's NHS Foundation Trust, Imperial College London, University College London Hospitals, and the Royal Marsden Hospital. This is focused upon piloting new procedures for sample collection and management, and new standards for clinical data especially molecular and digital pathology.

3.3 Phase 3: Genomics England main programme (2015-2017)

Phase 3: the main Genomics England programme which will run from 2015-2017. This is the focus of this protocol and will build on the lessons from the pilots. As we operationalise the processes and implement the Standard Operating Procedures (SOPs), we will increase our capacity over 2017 to deliver approximately 100 whole genomes per day.

3.3.1 The Genomics England rare disease main programme

The goals of the rare disease programme are:

- To increase discovery of pathogenic variants for rare disease.
- To add value with additional biological insights that build confidence in putative pathogenic variants.
- To enhance the clinical interpretation of WGS in rare disease.
- To develop a programme of functional multi-omics pathways, specifically transcriptomics, epigenetics, micro RNAs and biomarkers.
- To return these findings to the NHS for feedback to patients.
- To create a unique dataset for rare diseases that may enable therapeutic innovation.

Background to rare diseases

There are between 6,000 and 8,000 known rare diseases worldwide. Each disease or syndrome affects less than 0.1 percent of the UK's population, but cumulatively they affect the lives of 3 million people, and are associated with substantial morbidity and mortality. Only 50 percent of the known rare diseases have an existing molecular diagnosis. Of these, 85 percent are associated with a single gene defect, although there may be modifiers elsewhere in the genome that have implications for treatment and outcomes. Recent research has shown that WGS can augment gene discovery by between 25 to 40 percent across a range of rare disease phenotypes (personal communication Whole Genome Sequencing 500). Gene discovery in the 100,000 Genomes Project will create significant opportunities for scientific innovation through our focus on residual unmet need, and our emphasis upon national and international collaborations.[2,3,4,5,6,7]

National and international collaborations in rare diseases

Genomics England will work in collaboration with all important rare disease initiatives. In particular, we will actively support the implementation of the recently published UK Rare Disease Strategy.[2] Some key early partners will include the Deciphering Developmental Disorders (DDD) programme,[4] the NIHR Translational Research Collaboration in Rare Disease (TRC-RD), the Whole Genome Sequencing 500 programme and the NIHR Bioresource in Rare Diseases (BR-RD). DDD is collecting phenotypic and exome data from families with developmental delay across the network of 23 UK Genetic Testing Centres. TRC-RD will undertake deeper phenotyping of rare disease patient cohorts for experimental medicine and interventional studies. BR-RD is undertaking WGS of 8,000 patients with a range of rare diseases. Where possible, we will work with key international programmes including DDD[4] and Orphanet, and complement the work of the International Rare Diseases Consortium (IRDC).[3,4,5,6]

The Project will facilitate the generation of a national data resource of up to 50,000 people. This will consist of families with rare diseases, ideally based upon parent-offspring trios. This will assist in the interpretation of genetic variants of uncertain significance. We hope to create the opportunity to include data from BR-RD, TRC-RD, DDD[4] and other studies within the rare disease component of the Genomics England Data Centre, through the Genomics England Clinical Interpretation Partnership (GeCIP). This is subject to an open application process and ethical approval to combine datasets. This will harness the strength of current UK Rare Disease Programmes and will advance current understanding of rare disease mechanisms. It may also impact upon common diseases that share similar phenotypes. It will also offer opportunities for biomarker, clinical, and interventional studies through industrial partnerships. This initiative is particularly

timely as Public Health England have plans to create a national registry for rare diseases and we aim to partner them to ensure the optimal outcome for patients.

The Project is expected to enable genomically-driven reclassification of rare diseases leading to opportunities to recall patients for deeper phenotyping through TRC-RD. These data will pave the way for functional characterisation of findings thereby adding further value to datasets, improving diagnostic utility and possibly identifying new targets and therapies. The NIHR Office for Clinical Research Infrastructure has identified many rare diseases providing further potential opportunities for therapeutic innovation through academic-industrial partnerships.

The inclusion and exclusion criteria for rare disease areas

Genomics England will evaluate rare disease areas for inclusion in the Project. Each proposed disease will be assessed by scientific experts in our GeCIP (see **Section 5**) and the Genomics England Rare Diseases Working Group who will advise the Science Advisory Committee and the Genomics England Board (see **Appendix 4**).

Inclusion criteria for rare diseases:

- A rare disease with residual unmet diagnostic need identified by a proband within the NHS in England. In addition to typical rare inherited syndromes, this includes rare familial cancers with likely germline predisposition and may include specific individuals with extreme response to specific forms of sepsis (see infectious disease – **Section 3**).
- Prospective collection of samples is very important. However, in very rare diseases, we will consider some legacy collections that appear on the Approved List of Conditions (see **www.genomicsengland.co.uk**), or are approved via the process described above. This is provided that there is a clear unmet diagnostic need and a Genomics England-compliant consent can be/has been obtained.
- Access to appropriate family structures (ideally parent-offspring trio), with suspected Mendelian dominant, recessive, or X-Linked conditions and where feasible, the numbers of families that increase the prospects for successful gene identification.
- Multiple unrelated families with evidence of linkage to the same chromosomal region(s).
- Families with suspected Mendelian conditions with one or more affected individuals in a single generation (likely to be due to *de novo* mutations).
- Other family structures will be considered especially in rare diseases of late onset, including unrelated sporadic patients with no known family history via the Genomics England nomination process.
- Individuals with extreme forms of common disorders (e.g. particularly early onset or unusually severe forms) will also be considered.
- Evidence of core phenotypes compatible with that disease, with evidence of previous genetic testing.
- Particular conditions that have undergone a defined gene panel test or other tests to exclude known genetic mutations.
- Provision of Genomics England informed consent and upload of all phenotypic data to the Genomics England Data Centre.
- Provision of whole blood samples for the extraction of high quality DNA to enable WGS according to the Genomics England Standard Operating Procedures (SOPs) at the Genomic Medicine Centres or the Genomics England Central Biorepository, or providing they pass quality assurance standards.
- DNA must pass quality control (QC) and quality assurance (QA) for local extraction.
- Samples must be provided, managed and transferred according to Genomics England SOPs for multi-omic analyses (RNA, serum, plasma) which will be stored in the Central Bio-repository (see **Appendix 3**).

Exclusion criteria for rare diseases:

- No evidence of an inherited phenotype or rare disease.
- Prior identification of a known causative gene variant in an affected family member.
- Evidence that the NHS in England/HSCNI has no capability to recruit patients for a specific phenotype.
- The absence of valid consent, or unwillingness to give consent or participate in all aspects of the Project (this excludes opt-out options for feedback to participants).
- Inadequate phenotyping or failure to provide data and samples to the Genomics England Central Biorepository within 14 days.
- Probands not under the care of the NHS in England or HSCNI.*
- DNA of insufficient quality for WGS (we may request further samples).
- Where Genomics England has been advised that an adult patient has been deemed by their clinical team **not** to have mental capacity, to consent to join this Project (at the time this consent would be required), then in accordance with the Mental Capacity Act 2005, he or she will **not** be enrolled as a participant in this Project under usual circumstances, unless it has also been determined by the clinical team that his or her lack of capacity is **connected with** their rare disease or its treatment.**

Nomination process for a rare disease to the 100,000 Genomes Project

The Genomics England Rare Diseases Working Group will consider diseases nominated by clinicians, the NHS, Genomic Medicine Centres (including NIGMC), Genomics England Clinical Interpretation Partners, and industrial partners. In Europe, a rare disease is defined as having an incidence of 5/10,000, but the Project will consider rare diseases with high frequency. This is to limit our selection to ultra-rare diseases which would not meet the requirements of the UK Rare Diseases Strategy in terms of equity of access.

A core approved list of diseases that are already approved and included are listed on the Genomics England website (see **www.genomicsengland.co.uk**). This list is by no means exhaustive. Any nomination of a rare disease will need to present evidence of capability and capacity to recruit family units. To nominate a specific rare disease, a Genomics England rare disease nomination form will need to be completed, making the case for inclusion (please refer to **Appendix 5**).

The rare diseases nomination form requires:

- The disease or syndrome name.
- Inheritance pattern and evidence of a heritable phenotype.
- Evidence of unmet diagnostic need and estimated patient numbers across the NHS in England. Evidence of capacity and justification of how these numbers have been calculated.
- A comprehensive list of phenotypic and clinical features, including any diagnostic tests to enable construction of the phenotypic data model for these diseases.
- Numbers of parent offspring trios, family units or affected individuals available. If family units that are not based on two parents and an affected offspring are proposed, some explanation of the rationale for this choice is vital.
- Estimated timescale for recruitment.
- Formal commitment to provide all phenotypes and samples requested for proband and other family members, and NHS numbers on all family members.
- Formal commitment to adhere to all Genomics England Standard Operating Procedures.

* Devolved nations are subject to parallel discussion on funding and access arrangements.
**Adult patients (over 16 years old) in Northern Ireland who have been deemed not to have mental capacity to consent to join this Project (at the time this consent would be required) will not be enrolled as participants under any circumstances.

The process for approval will use the following criteria:

- Unmet diagnostic need.
- Evidence that a critical number of families with the optimal structure can be identified and engaged, and that DNA can be provided from the NHS in England and the HSCNI.
- Evidence of a formal commitment to provide all phenotypes requested and NHS numbers on all family members.
- Timescale for recruitment of completed family units.
- The existence of sufficient resources to perform WGS within the Project, or an external source of funds to complete sequencing in all the families planned for recruitment.

To inform inclusion of families in the rare diseases, we will pay attention to estimates of the numbers needed to detect a pathogenic variant dependent on mode of inheritance, family structure, and penetrance (see **section 3.4** for power simulations).

The approval process for inclusion of rare diseases in the 100,000 Genomes Project

Specific GeCIP domain experts with the Rare Diseases Working Group will review all nominations against the criteria listed above, and make recommendations of appropriate disease areas to the Genomics England Science Advisory Committee (SAC). Any appeal against a decision of the Science Advisory Committee not to accept a rare disease phenotype will be considered once by the Genomics England Board whose decision is final. Anyone from a family with a rare disease who approaches Genomics England directly will be advised to contact their NHS clinical team or general practitioner.

Patient engagement and phenotyping strategy for rare diseases

Genomics England will work in partnership with NHS England and DoH NI (operating through its arm's length body BHSCT) to engage, recruit, and characterise patients with specific rare diseases, with residual unmet diagnostic need. A managed network of NHS Genomic Medicine Centres (GMCs) (including NIGMC) have been established across England and Northern Ireland. It is expected that the English GMCs will work within a performance managed network and within the current commissioning system. Separate arrangements and agreements exist for the establishment, operation and governance of the NIGMC, based on similar contractual arrangements as English GMC. NHS healthcare teams will be responsible for recruitment of patients and family members following agreed Standard Operating Procedures and appropriate fully informed consent. They will also be responsible for the provision and recording of core clinical phenotypic and demographic data on patients and relevant family members. Consent will be obtained via a phenotype acquisition tool in electronic format to provide digital records for access by NHS healthcare teams. Phenotypic data will include all relevant investigations e.g. digital photography, laboratory tests, imaging and extant laboratory or genetic testing.

Clinical phenotyping in rare diseases

For each disease area, a core set of characteristics and data points will be defined in consultation with clinicians and researchers with relevant expertise. In most cases, this will include a number of statements made using a phenotyping ontology. Although the Human Phenotype Ontology (HPO) is being used, other ontologies may also be employed.

Data will be recorded using controlled, clinical terminologies wherever possible for consistency.

The NHS number/Health and Care (H&C) number (for Northern Ireland participants) (and other core demographic data) will be recorded for each participant, including the biological relatives recruited as part of a trio or duo for patients with rare disease. A family tree and history of diseases will be recorded for parents and all offspring. The NHS/H&C number will be used to identify and link to relevant health and social care records, providing additional context for the core phenotypic data.

A secure, web-based, information system will be provided for the collection of this data, removing the need for any additional, bespoke development within participating NHS organisations. An electronic facility will be provided for transmission of the same data from existing information systems. Where organisations have developed their own capacity for capturing and managing the same data, to the same standards, there will be no requirement for additional data entry.

There is an expectation that the phenotypic data will be provided within two weeks of recruitment and samples will not be sent for sequencing until the phenotypic data set, which will vary by disease area, has been supplied.

Each rare disease may have a different set of phenotypic characteristics ranging across clinical features, laboratory tests and imaging. We will use published literature, ontologies and expert advice to create disease specific data models for each disease (**section 3** on power emphasises the importance of precision phenotyping).

Whole genome sequencing strategy and quality assurance for rare diseases

Our current approach for the main programme has been adapted from experience gained in the WGS500 study and the NIHR Bioresource – Rare Diseases, as well as advice from UK and international experts and information from technology providers. Initially, for rare diseases, Genomics England will sequence whole genomes in all members of the family trio (proband, mother, and father) or other family structures approved for inclusion by us. This strategy may be modified as information emerges from the pilot programmes and other efforts worldwide.

DNA sequencing for the main programme has been contracted through Illumina. This was informed by the results of the **Phase 1a** sequence evaluation exercise. Although there is currently only one sequence provider in the programme, we expect that multiple instruments and chemistries will be applied throughout the lifetime of the project as the technology matures. Sequencing standards will therefore be maintained through the requirement to meet platform independent metrics. All germline genomes will be sequenced polymerase chain reaction (PCR) free. This will be to a depth and quality such that 95 percent of the autosomal genome (as defined by the GRCh38 reference genome) will be read by at least 15 or more independent observations each having a quality of >Q30 and mapability of >mapQ20. Genomics England will oversee quality and assurance with a clear standard operating procedure for QA processes, sampling time points, monitoring review, remedial action and change approvals. Other quality metrics will include: duplication rates, sample contamination estimates, insert size, and Guanine and Cytosine distribution (likely to coincide with repetitive regions) which is often associated with difficult to interpret sequence. The data generated will be transferred encrypted via the internet to the Genomics England's Data Centre. Genomics England will capture quality metrics per run and routinely check for the presence of batch effects at the run, flow cell, and lane level (as applicable to the technology) to ensure that sequence quality is maintained and to aid its improvement over time.

Alignment and identification of sequence variation

As part of the sequencing contract, providers will supply sequence data aligned to the reference genome (currently GRCh38) and in a standard file format (currently BAM). They will also identify sequence variants and report them in a standard file format (currently VCF). It is expected that the mapping and calling pipelines will report single nucleotide polymorphisms (SNPs), short indels (<50bp), copy number variants (CNVs) and structural variants (SVs). It is also expected that our pipelines will incorporate processes to jointly call across trio data. This will ensure that missing sequence data is distinguished from that identical to the reference and that putative haplotypes are generated to the extent possible depending on the underlying sequence fragment lengths. Pipelines should also identify regions of the genome with inadequate coverage for reliable variant calling. Genomics England will define clear Standard Operating Procedures and will revise and inform the NHS of any changes as these evolve over the life of the Project.

Genomics England will routinely assess the quality of the mapping and variant calling provided and compare this against alternative algorithms to ensure the data has been processed with the best available pipelines. Periodically, we may run additional tests of annotation suppliers, similar to **Phase 1b**, to blind

test new algorithms and pipelines. These efforts will be supplemented by information from external evaluation efforts and against high quality benchmark genomes (e.g. Platinum Genomes or reference genomes from the Genome in a Bottle Consortium). Where pipelines demonstrate significantly improved results, existing data may be realigned and recalled as appropriate by Genomics England. Sequence providers may also be required to update their contracted alignment and calling pipelines to incorporate such improvements.

Interpretation of sequence variation

In order to provide timely analysis to support interpretation within NHS clinics we will have a core sequence annotation service which may be contracted from external annotation providers through competitive tendering. This will be informed by the results of annotation evaluation exercises starting with the **Phase 1b** exercise. In this respect, GeCIP, where expert research teams and the NHS clinicians work together within the Genomics England Data Centre to shape the eventual report for the NHS, will be a crucial aspect of the Project (see later). The objective is to procure a robust annotation pipeline. As far as current understanding allows, this pipeline will be able to identify a small number of sequence variations that are the most likely causative pathogenic variants of the participant's medical condition. The output from this pipeline will be fed back to clinicians in the NHS in the form of a clear textual report, as well as a list of variants in the form of an annotated VCF format file or other jointly agreed standards. These annotations will also be available to researchers within the Genomics England Data Centre. Given the small number of samples used to evaluate pipelines in **Phase 1b**, it will be necessary to carry out additional evaluations of candidate annotation pipelines in **Phase 2 and 3**.

It is expected that such pipelines will use a variety of algorithms to filter variants and rank the most likely pathogenic candidates. In most cases, both the parents and one or more affected child will be sequenced. Variants confidently identified as heterozygous in the child and as homozygous reference in the parents, and that are absent from all public databases (1000 Genomes, ESP and UK10K) and from other Genomics England samples would be candidate *de novo* causal variants. Pipelines will be expected to evaluate other modes of inheritance, such as compound heterozygous events. They will also identify long regions of homozygosity (LROHs), particularly where consanguinity is suspected, and in such regions, identify potentially pathogenic homozygous variants. In terms of identifying variants most likely to have functional consequences, it is anticipated that pipelines will evaluate variants using various prediction tools (e.g. Ensembl Variant Effect Predictor, Sift, and Polyphen), as well as checking against databases of known pathogenic variants (e.g. ClinVar and ClinGen). Variant annotation will be updated regularly, and snapshots of previous annotation versions will be stored and versioned.

3.3.2 The Genomics England cancer main programme

Goals of the cancer programme

The goals of the cancer whole genome sequencing (WGS) programme are to:

- Use WGS to identify novel driver mutations for cancer and to understand its evolutionary genetic architecture through primary and secondary malignant disease (by multiple biopsy and WGS).
- Partner stratified healthcare programmes and outcome studies with patients from the NHS in England, to enable understanding of WGS benefits in defining predictors of therapeutic response to cancer therapies.
- To use multi-omic approaches including transcriptomics, proteomics and epigenetics to offer additional biological insights into cancer.
- To utilise WGS to identify new pathways for cancer therapies.

Rationale and process for selecting specific cancers for inclusion

Cancer is fundamentally a genetic disorder where mutations (including copy number aberrations, indels

(insertion/deletion variants), complex rearrangements, and non-synonymous substitutions) lead to uncontrolled cellular proliferation. The clinical impact of sequencing technologies has enabled precise definitions of disease, uncovered mechanistic insights into pathogenesis, and identified therapeutic targets based on genetic variation or aberration. Additionally, sequencing approaches have catalogued the complex evolutionary changes that occur in an individual's cancer under the effects of treatment and time. This has demonstrated that there are both expanded clonal populations and the existence of low frequency sub-clones, each with specific genomic architecture.[8,9,10,11,12,13]

Meta-analysis across cancer types has confirmed the importance of 200 key genes in driving cancer.[12,13] However, focusing only on these genes with targeted re-sequencing will be insufficient to significantly impact upon the majority of individuals with cancer. Overall response rates to targeted agents with 'actionable' mutations have been relatively modest. The differences in response between patients with mutations in the same gene suggest complex interactions between identified activated pathways and the genomic environment in which they occur. Moreover, power calculations and saturation analysis suggest that many more cancer genes remain to be found (see **Section 3.4** sample size requirements). In many cancers, the 'actionable' gene may not be the dominant driver. Furthermore, recurrent oncogenic mutations are rare and some of the putative driver mutations are caused by copy number aberrations or complex rearrangements that cannot be globally ascertained with gene panels, thus WGS is required.

The 100,000 Genomes Project will learn from and collaborate with the Cancer Genome Atlas (TCGA) project and the broader International Cancer Genome Consortium (ICGC), who are producing an inventory of genomic, transcriptomic and epigenomic changes in a wide range of different tumour types.[12,13]

A core list of currently approved cancers can be found on the Genomics England website (see **www.genomicsengland.co.uk**), where clinicians, the NHS, Genomic Medicine Centres, Genomics England Clinical Interpretation Partners (GeCIP) and industrial partners can nominate a disease. For any nominated cancer, Genomics England, NHS England and DoH NI (operating through its arm's length body BHSCT) will examine the capability within the NHS and the mechanism for the recruitment of patients. To nominate a specific cancer, a Genomics England cancer nomination form will need to be completed, making the case for inclusion.

The inclusion and exclusion criteria for the cancer programme

Genomics England will individually evaluate each cancer for inclusion in the 100,000 Genomes Project. Each proposed disease will be assessed by scientific experts in specific GeCIP domains, and the Genomics England Cancer Working Group, who will advise the Genomics England Science Advisory Committee, then make a final decision. Any appeal against a decision not to accept a particular cancer will be considered once by the Genomics England Board, whose decision is final. We will also consider collaborations with stratified healthcare programmes in cancer.

Inclusion criteria for cancers:

- A cancer from the Genomics England approved list or an approved nomination by the NHS, academics, or industry that has been identified in a patient within the NHS in England.*
- Evidence of core phenotypes compatible with that disease and capability of the NHS in England/HSCNI to recruit, supply samples, and upload data from patients.
- Access to appropriate high quality DNA from both tumour and germline samples, enabling WGS, including formalin-fixed paraffin-embedded (FFPE) samples that have been in formalin and processed within 24 hours to reduce denaturation of DNA, according to the Genomics England Molecular Pathology SOPs.
- Existence of valid informed consent and upload of the phenotypic data set, including all molecular pathology tests to the Genomics England Data Centre within 14 days.

* It is hoped that devolved nations will join the Project.

- Timelines for recruitment that accord with Genomics England Standard Operating Procedures.
- The existence of sufficient resources to perform WGS within the Project, or an independent source of funds to complete sequencing in all planned recruits e.g. from a GeCIP funder.

Exclusion criteria for cancers:

- The absence of valid consent, unwillingness to give valid consent, or participate in all aspects of the Project (excludes opt-out of feedback).
- Inadequate phenotyping, or failure to upload phenotypes to the Genomics England Data Centre.
- No access to matched tumour and normal DNA samples.
- Patients not under the care of the NHS in England or HSCNI.*
- DNA of insufficient quality for WGS, or failure to adhere to the Genomics England Molecular Pathology SOPs.
- Where Genomics England has been advised that an adult patient has been deemed by their clinical team **not** to have mental capacity to consent to join this Project (at the time this consent would be required), then in accordance with the Mental Capacity Act 2005, he or she will not be enrolled as a participant in this Project under the usual circumstances, unless it has also been determined by the clinical team that his or her lack of capacity is **connected with** their cancer or its treatment.**

Phenotyping of patients participating in the cancer programme

For each disease area, a set of phenotypic characteristics or data points will be defined in consultation with clinicians and researchers with relevant expertise. This core phenotypic data set will be aligned with the key reporting data sets for cancer such as, the Cancer Outcomes and Services Dataset, the Systemic Anti-Cancer Therapy Dataset, and the Radiotherapy Data Set. It will also be aligned with the clinical audit data sets collected for specific cancers: colorectal, lung and prostate, to avoid duplication of effort. Data will be recorded using controlled clinical terminologies, and structured ontologies wherever possible.

The NHS/H&C number (and other core, demographic data) will be recorded for each participant. This information will be used to identify and link relevant health and social care records, providing additional context for the core phenotypic data. It will be used also to identify and link registry and audit data, where this is available.

A secure, web-based information system will be provided for the collection of this data, removing the need for any additional bespoke development within participating NHS organisations. An electronic facility will also be provided for transmission of the same data from existing information systems. Where organisations have developed their own capacity for capturing and managing the same data, to the same standards, there will be no requirement for additional data entry. Genomics England is working in partnership with the NIHR Health Informatics Collaborative (HIC), that is establishing flows or repositories of clinical and laboratory data in key therapeutic areas to known comparable standards, across a number of NHS trusts.

There is an expectation that the phenotypic data will be provided within 14 days. Samples will not be sent for sequencing until genomic and somatic DNA of sufficient quality and a minimum data set, with details of all investigations (including molecular tests on tumour samples – which may vary by disease area) have been supplied. Whole genome sequencing will not be undertaken until the dataset, results, and QA processes are submitted with sample.

DNA quality assurance from cancer samples

For formalin-fixed paraffin-embedded tumour (FFPE) DNA, we will only accept samples that have had associated digital pathology focused upon optimising the cellularity of the tumour, from which DNA is extracted. To avoid sequencing failure because FFPE yields limited DNA, we will apply a delta CQ step within our Biorepository to assess likely success of DNA amplification. In combination with usual optical densitometry, this measures

* It is hoped that other devolved nations will join the Project.
** Adult patients (over 16 years old) in Northern Ireland who have been deemed not to have mental capacity to consent to join this Project (at the time this consent would be required) will not be enrolled as participants under any circumstances.

whether the DNA available is sufficient for WGS. Only DNA that passes this test will enter sequencing.

Additional samples for multi-omic functional studies

We will seek additional funding for appropriate RNA, plasma, and serum allowing the acquisition of expression, methylation, epigenetic data and serum or plasma biomarkers pertaining to the same samples used for WGS. These samples will be collected and stored at the Central Biorepository and will add the value of functional analysis to the WGS. We have a significant opportunity to apply our expertise in detecting temporal changes in cancer genomes (in cell-free tumour DNA in plasma) to develop non-invasive 'liquid biopsy' based on WGS.[14] This could provide a cost-effective means of disease monitoring, to enable innovation in trial design and encourage industrial investment in UK clinical research.

Whole genome sequencing strategy and quality assurance for cancer

Our current approach for the main programme has been adapted with advice from UK and international experts and information from technology providers. Initially, Genomics England will sequence germline samples using the same criteria as for rare diseases samples (see above).

For tumour DNA, we will sequence at 75X coverage. However, in some cancers, greater depth of coverage may be required. We will modify this as information emerges from the pilot programmes and other efforts worldwide. WGS will be performed using the same technology and QC approaches outlined for the rare diseases programme. In addition, we will monitor the frequency of single nucleotide variations (SNVs) potentially caused by DNA alterations created during the FFPE process, (such as G to A, or A to G transitions) due to PCR amplification.

Information governance and identification of sequence variation

For the rare disease programme, sequence providers will carry out initial steps in read mapping and variant calling to deliver BAM (binary alignment map) and variant call files (VCF files), for both germline and tumour samples. In addition to the requirements outlined for rare diseases, we expect pipelines to jointly call across tumour-germline DNA pairs, to identify somatic variants and report the purity of the tumour sample. This is specific to cancer.

Annotation of sequence variation

For rare disease annotation, the intention is to provide the best annotation pipelines which may include external providers, to ensure the generation of timely annotation reports for clinicians. However, given the more difficult nature of annotation in cancer, it is expected that analysis carried out in GeCIP will have a larger input into evolving the annotation process in cancer (see **Section 5**).

It is expected that the annotation process will distinguish between somatic variants, those mutations acquired by the cancer, and cancer susceptibility germline mutations that might increase the overall lifetime risk of cancer in the individual. For germline, annotation will exclude known benign common variants using datasets such as 1000 Genomes, Exome Sequencing Project, and UK10K. The annotation of somatic variants will be expected to highlight those found in known cancer genes and identify them as being likely to result in loss of function. To determine the likely importance of variants, it is expected that reference will be made to general cancer databases such as COSMIC, as well as gene specific databases, such as, the IARC TP53 mutation database and other resources covering RB1 and the BRCA genes where appropriate.

A major component in cancer annotation is analysing the consequence of larger scale genomic changes, such as structural variants, copy number aberrations, loss of heterozygosity and other chromosomal mutational events. Since some will not be unique to the tumour sample, a paired approach to the analysis of the two samples will be important to ensure the origin of the variant is clear. Regarding the annotation of individual sequence variants, the consequence of copy number and structural variants will be assessed for their impact on genes implicated in cancer.

The variability in tumour sample purity, both with respect to contamination with normal tissue and the

likely presence of multiple tumour sub-clones, is a particular challenge to annotation pipelines. It is crucially important that in addition to paired analysis, the tumour is also analysed separately. This will ensure that key somatic lesions are not filtered out by the variant calling algorithm, due to tumour contamination in the germline sample. As the cancer pilot progresses, these issues will be assessed and the pipelines will be modified as required to optimise WGS.

3.3.3 The Genomics England infectious diseases partnership with Public Health England

Infectious diseases are responsible for seven percent of deaths in the UK per annum and eight percent of all hospital bed days. It has been estimated that they cost the UK economy approximately £30 billion per annum (Chief Medical Officer's Report 2011).†

WGS for pathogens (both viruses and bacteria) will soon be adopted for routine management of infectious diseases, providing information on species, virulence, transmission and anti-microbial resistance. Thus there are tremendous opportunities for clinical research in species determination, genotype-to-phenotype prediction, and infection control. In each case, scientific progress will lead directly to significant public health and economic benefits, as well as cost savings to the National Health Service.[15,16]

Scientific aims of the Infectious Disease Programme

Working in partnership with Public Health England, we will address the following scientific objectives:

- Engagement with the taxonomical community to help develop the species knowledge base, with a rational, extensible re-classification of pathogens into species and clinically-important sub-species where genetic variation reveals a wider diversity, on the basis of whole genome and clinical information.
- Development of effective techniques for genotype to phenotype prediction, with a particular emphasis upon anti-microbial resistance.
- Identification of transmission networks based upon phylogenetic trees of sequenced pathogens, using algorithms that take account of complex features in recombination and quasi species in meta-populations.
- Validation of determinations of virulence and anti-microbial resistance, achieved through the linkage of pathogen genomic data with patient health records.
- Identification of specific vulnerabilities to infection that can be used to inform therapeutic developments and help determine preventative action.

Human and pathogen sequencing in the 100,000 Genomes Project

Within the infectious disease workstream of the Project, there are two distinct strands.

(i) Pathogen sequencing:

This is being led by Public Health England, in partnership with Genomics England. There is currently no specific funding from the Project for sequencing of the pathogens. There is significant extant infrastructure to do this within the NHS, which Public Health England will harness to undertake a number of pilots. Currently there is a focus on human immunodeficiency virus, hepatitis C virus, and tuberculosis. The rationale for selection of these is given in the Department of Health Advisory Committees report – Infectious Diseases (see **www.genomicsengland.co.uk**).

(ii) Host sequencing of individuals with severe response to infection:

This has a specific focus on severe response to infection of hosts (e.g. severe response to sepsis). This will include

† Davies S. "Annual Report of the Chief Medical Officer, Volume One, 2011, On the State of the Public's Health" London: Department of Health (2012).

host sequencing funded by Genomics England (and potentially pathogen sequencing, if appropriate, funded by Public Health England, within microbiology services infrastructure). In effect this strand assumes individuals with a severe response can be considered as a rare disease, and are therefore subject to the same requirements and process, as outlined in the protocol (see **section 3.3.1**) for the rare diseases programme (with minor exceptions – see below).

(i) Pathogen sequencing and the role of Genomics England:

As Public Health England is leading the pathogen sequencing pilots, it is not our intention to outline the specific operational aspects of this component of the Project. Rather, the protocol needs to be considered in the light of the potential for the inclusion of data (including WGS raw reads, or BAM files from pathogen sequencing), as a centralised hub for storage and retrieval of genomic data for infectious diseases. This could be for the pilots in the first instance, but expand in the future for all isolates sequenced by Public Health England and/or others on behalf of the NHS. Genomics England would then act as the genomic data warehouse, using experience and expertise from the clinical programme as well as building up a library of sequences from pathogens over time.

As the interpretation of data from infectious agents are dependent on global and national surveillance, dynamic models that link the sequences to global databases are likely to operate for public health functions. However, assembled whole genome sequences with linked metadata could be extracted and mirrored in the structure operated by Genomics England, to establish a mechanism of access for further research in mechanisms of resistance, spread of clonal infections, vaccine efficacy and novel drug design to the benefit of the NHS and taxpayer.

(ii) Host sequencing of individuals with severe response to infection.

The selection of host response to infection phenotypes for the 100,000 Genomes Project

The Public Health England Infectious Disease Genomics Working Group will consider diseases nominated by clinicians, the NHS, Public Health England, industrial partners and patient-facing charities.

This group will advise the Genomics England Science Advisory Committee of eligible phenotypes that should be accepted for WGS. The decision of this committee on inclusion of a particular condition will be ratified by the Genomics England Board, whose decision will be final. When a disease is selected for the Project, we will announce it on our website. In the event that resources limit our ability to include a disease, where possible we will work with the clinical teams, to apply to funders to look at phenotypes of specific response to infection.

Phenotyping and recruitment of patients participating in the sequencing of host response to infection programme

Alternate phenotyping strategies will use the same principles as the rare diseases protocol, i.e. a core phenotypic dataset for each disease will be established by the Genomics England team working with disease experts and the NHS number will be recorded for electronic record linkage.

For some disease groups, if proposed and accepted, it may be more appropriate for identification and recruitment to occur in additional settings to those of the proposed NHS Genomic Medicine Centres (GMCs). For example, as meningococcal infection is a notifiable disease and is required by law to be reported to local Public Health England health protection teams, it may be more efficient to recruit through these systems. The starting point for sample collection would be GMCs, but could be revised in time so that it is in line with local Public Health England systems and requirements. A decision will be made on whether this would be part of a local arrangement between Public Health England and the NHS.

Enhancing the clinical feedback of host whole genome sequencing in infectious diseases

Local validation is unlikely to be required given the knowledge base in this area, but may evolve with

information about the pathogen being fed back to Genomic Medicine Centres/the referring clinician. As would be done routinely, an action plan will be agreed for the care and management of the patient. Because of current state of knowledge in this field, there may be limitation on the likelihood of clinical feedback directly impacting on patient care.

3.4 Sample size considerations in rare disease and cancer

3.4.1 Assumptions for power simulations for the 100,000 Genomes Project

The power simulations we have conducted allow us to estimate sample sizes needed to resolve a rare disease, or identify the driver mutations in a cancer. In the Project, the average sequencing coverage will be 30X for germline and 75X for tumour (the minimum standards of the Project). This results in 99 percent sensitivity to determine single nucleotide changes across the genome for both types of samples (assuming a formalin fixed paraffin extract sample with 50 percent tumour cellularity). We estimated the numbers of samples needed to identify genes underlying rare diseases and cancer under a variety of scenarios. The required numbers vary widely depending on expected mode of inheritance, penetration of a pathogenic variant, genetic heterogeneity of the presentation and, for cancers, mutation rate. In reading these, it is important to remember they are simulations based on conservative scenarios.

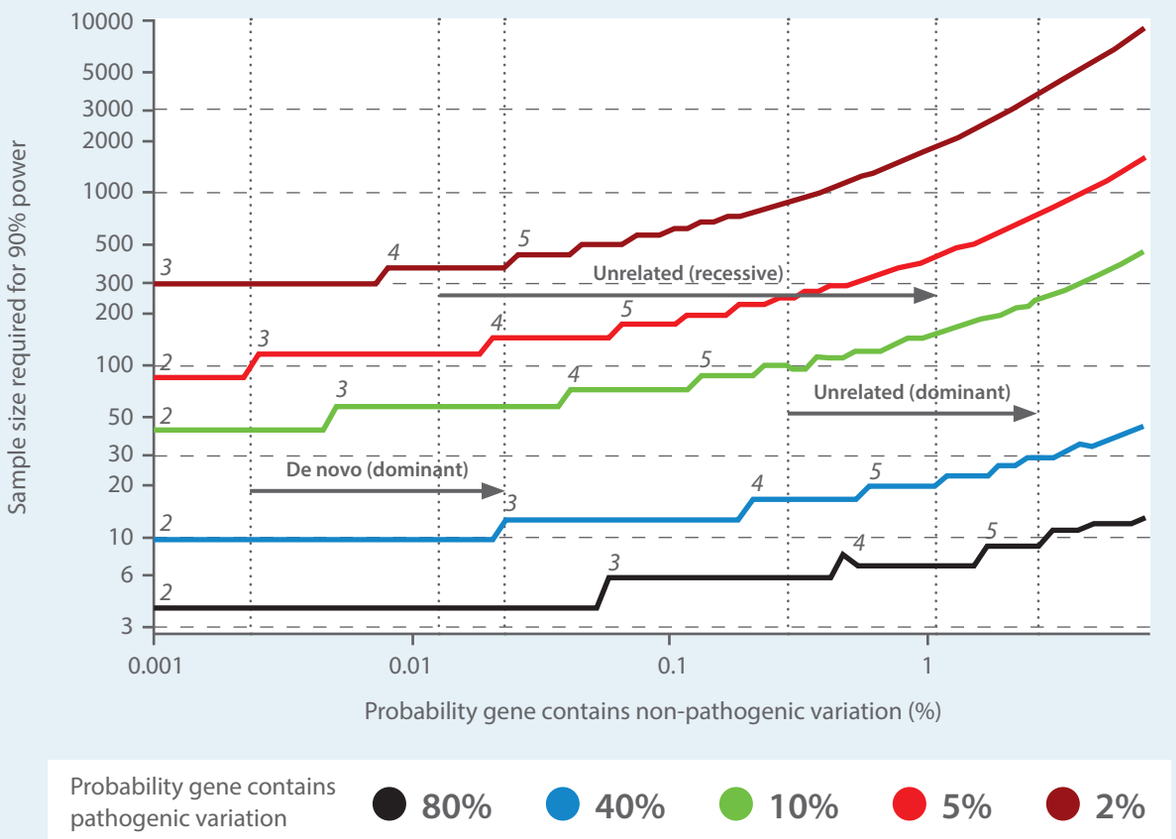
Rare disease sample sizes

Simple Mendelian diseases with little genetic heterogeneity may require 10 cases or less to solve. Diseases caused by many genes, very large genes, or cancers with high mutation rates, can require hundreds, or even thousands of cases to solve. In a rare disease, if we assume that it is caused primarily by *de novo* mutations, and 50 percent of cases share defects in the same gene, then between 30 and 50 trios would be required to have 90 percent power to identify genes that contained 90 percent of causative mutations. We have run simulations under a variety of scenarios:

- ***De novo* dominant diseases:** sequence trios to test for an excess of *de novo* mutations relative to per-gene mutation rate.
- **Inherited dominant diseases:** sequence cases to test for excess of very rare (<0.1% Mutation Allele Frequency (MAF)) mutations compared to healthy individuals. There is a significant increase in power by sequencing affected parent and filtering non-shared variants (dotted arrow in Figure 1).
- **Inherited recessive diseases:** sequence cases to test for an excess of homozygous or compound heterozygous variants from uncommon (<1% MAF) mutations compared to healthy individuals. There are significant increases in power by sequencing one parent and filtering out in-phase variants (see Figure 1).

The power curves in the figures assume that only missense variants are analysed and all variants are treated as equivalent. In practice, analysis may be restricted to certain classes of variation, e.g. loss of function (LoF) variants. This can increase power if the category we choose is highly enriched for pathogenic variation, but can reduce power if a large proportion of cases are explained by variation outside of this category (e.g. by non-LoF missense variants).

Figure 1. Different rare disease experimental designs are marked by arrows below. The number of detected mutations in cases required to pass genome-wide significance ($p < 2 \times 10^{-6}$) is shown by the italic numbers on the power curves (only numbers five or smaller are shown).

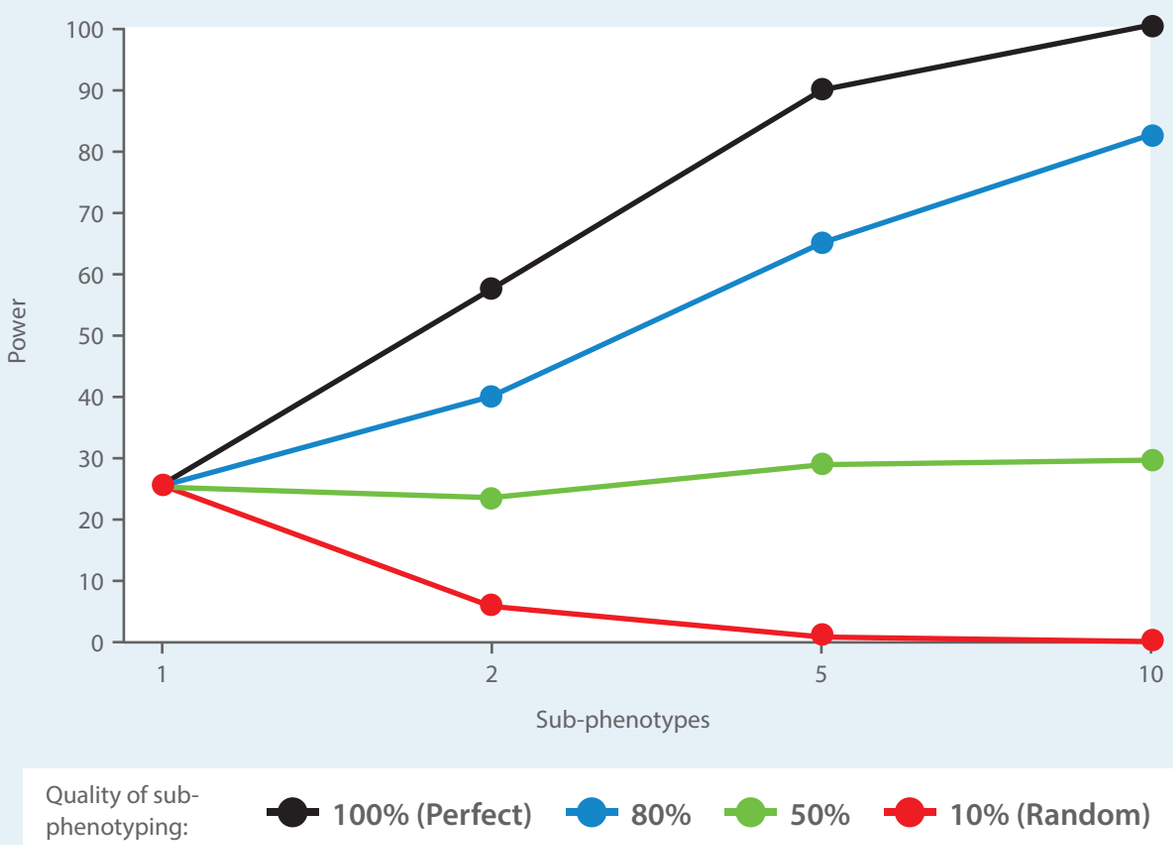


3.4.2 Effect of sub-phenotyping on power

Genetic heterogeneity increases the sample size required to detect disease genes. When this is linked with phenotype heterogeneity, detailed sub-phenotyping can be used to recover some of this lost power, by defining genetically homogenous subgroups. **Figure 2** shows the possible increases in power to detect disease genes for a hypothetical disease caused by mutations in 10 different genes, depending on the number of sub-phenotypes defined and the quality of the sub-phenotyping. The quality of phenotyping ranges from 100 percent (i.e. individuals with mutations in the same gene always have the same sub-phenotype) to 10 percent accuracy (i.e. what would be expected by totally uninformative phenotyping). We assume a sample size of 100 (i.e. approximately 10 mutations per gene).

Perfect sub-phenotyping (genetically fully homogenous subgroups), greatly increases the power, due to the relative simplicity of solving diseases caused by single genes. This increase in power is true, even if the groups are not perfectly homogeneous (e.g. 2 or 5 subgroups in **Figure 3**), or phenotyping is not perfect, but merely highly informative (the blue line in **Figure 3**). However, if the phenotyping is of low quality (where <50 percent of cases with the mutation have the same phenotype, e.g. the orange line in **Figure 3**), splitting samples into sub-phenotypes will actually decrease power, and in this case the phenotype-blind analysis of all individuals would be more powerful. Different scenarios (in terms of mutation rate, number of genes, and sample size) can alter the exact value of the sub-phenotyping accuracy required to boost power. The general message is that good phenotyping greatly boosts power and bad phenotyping is worse than no sub-phenotyping at all.

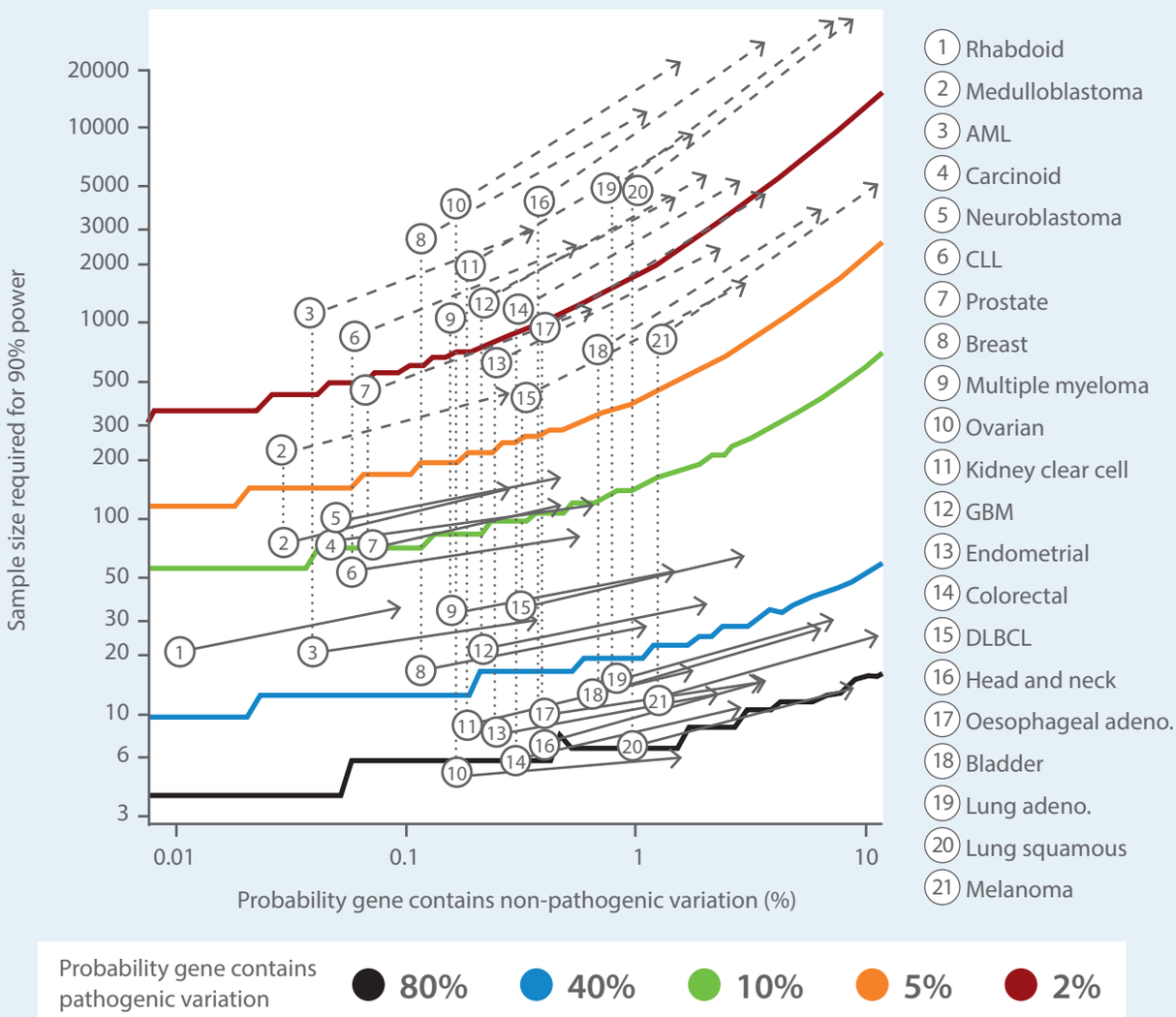
Figure 2. Changes in power as a function of sub-phenotyping detail and accuracy (for one scenario). Assumes 100 samples with a broad phenotype, caused by mutations in 10 genes, each explaining 10 percent of cases. Assumes a 2.7 percent mutation rate, i.e. a hard disease to solve (equivalent to the left hand end of the dominant arrow in **Figure 1**).



3.4.3 Power simulations for the cancer sequencing studies*

To assess the power to detect a gene containing pathogenic coding variation, we characterise a particular study by the rate of neutral variation in healthy individuals (the x-axis below) and the rate of pathogenic variation (the coloured lines below; i.e. the fraction of cases caused by a mutation in the relevant disorder). The following calculations and plots generalise the results of Lawrence *et al* 2014.[9]

Figure 3. Sample sizes required for 90 percent power to detect a disease associated gene under different scenarios at a genome-wide significance level of P = <2 x 10^6. The x-axis shows the probability that the gene contains a relevant missense mutation/variant in a healthy individual (a function of gene size, mutation rate, and genetic model). The coloured lines show the power required for different frequencies of pathogenic mutations in that gene. We looked at two example genes for each tumour, the most commonly mutated known gene, and the most infrequently mutated known gene found in Lawrence *et al.*[9]



In the above figure, the solid arrow depicts the range of sample sizes needed to detect the most commonly mutated gene (e.g. TP53 in ovarian and lung cancer, and APC in colorectal cancer). Studies with sample sizes below these solid arrows will find no significant genes at all. The dashed arrow shows the most infrequently mutated known gene. To start finding many new genes in any of these tumour types, you need to have a sample

* The following section on power calculations was undertaken by Luke Jostins and Gil McVean, University of Oxford 28/07/2014.

size above this dashed arrow. The arrows are sloped to show the increase in sample size you would need, to find more mutagenic genes at this frequency. Although it is likely new genes will be detected as sample sizes increase above the left hand side of the dashed arrow, to be well powered across all genes, you will need sample sizes to exceed those indicated on the right hand side of the dashed arrow.

In cancers where there is only one known gene (rhabdomyosarcoma, carcinoid, and neuroblastoma), and they are expected to be relatively oliogenic, there is only one arrow.

This analysis shows us that lung cancers will be the hardest to solve (dashed arrow 19 and 20 on **Figure 3**). In lung cancer, significant genes containing driver somatic mutations such as TP53 will be detectable in <30 samples. A study that is powered to detect new lung cancer genes would need to have a very large sample size to get over the high mutation rate, e.g. 5,000 cases at minimum, ranging up to >20,000 to pick up larger, or more mutagenic causal genes (the right hand side of the dashed arrows 19 and 20 in **Figure 3**).

The genetic architecture of some cancers such as prostate or medulloblastoma, may require sample sizes of <1,000 to discover many novel lower frequency (1-2 percent) somatic mutations.

Breast cancer has a reasonably low mutation rate, and lots of common driver genes (PIK3CA, PTEN, and P53 are each mutated in about 1/3 of cases), therefore studies ranging between 30-50 samples will find plenty of genes. However, as this is a very heavily studied cancer, many of the rarer genes have already been found. To significantly add to this set and make new discoveries, a study design will need to have 3,000 to 10,000 samples to be well powered.

3.4.4 Declaring significance

Genome-wide significance, correcting for the number of genes in the human genome, is typically defined as $p<2 \times 10^{-6}$. The number of detected mutations in cases required to pass genome-wide significance is shown by the italic numbers on the power curves in **Figure 1** (only numbers five or smaller are shown).

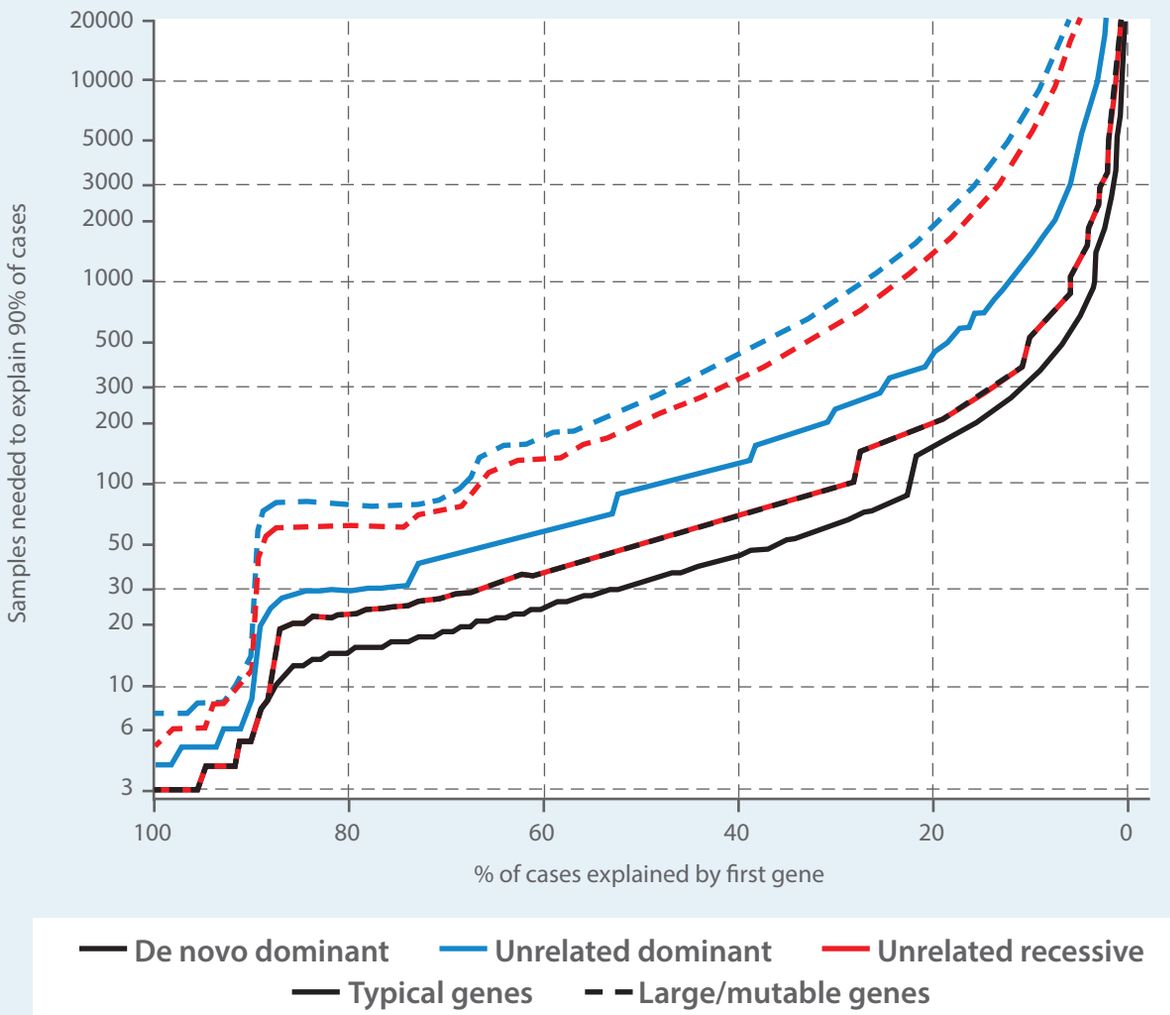
3.4.5 Variation in gene size and mutation rate

The median human gene has 1293 coding bases and a mutation rate of 1.2×10^{-8} bp/generation. The 90th percentile of gene size is 3218 base pairs, and the 90th percentile of mutation rate is 3.9X, the median rate. In **Figure 1**, the arrows for rare disease show the range of background variation, from the median gene on the left, and a large mutable gene on the right. For cancer in **Figure 3**, this is shown by solid or dashed lines.

3.4.6 Genetic architecture, and discovering genes vs solving diseases

We model disease heterogeneity as a geometric distribution, i.e. the fraction of cases caused by mutations in the most commonly mutated gene is x, the fraction explained by the next most commonly mutated gene is x(1-x), and the fraction caused by the nth most common gene is $x(1-x)^{n-1}$. Although this is a crude approximation, it is probably conservative.

Figure 4. Sample sizes required to explain 90 percent of the disease (i.e. have 90 percent power to detect enough genes to explain 90 percent of cases). This increases as the heterogeneity of the disease increases. Disease heterogeneity is modelled by a geometric distribution. The different colours show different experimental designs, and the style of the line shows the mutation rate. The solid lines show power curves for typical genes and the dashed lines represent large mutagenic diseases.



3.4.7 Coverage, sensitivity, and cellularity

Confidently calling a variant requires at least three non-reference reads, including one for each strand. Modelling coverage as an over-dispersed gamma distribution, a 30X genome gives a 99 percent sensitivity to detect germline mutations. A 75X cancer genome gives 99 percent sensitivity at 50 percent cellularity, and 96 percent sensitivity at 30 percent cellularity. Except for the case of low cellularity tumours, the biggest loss of sensitivity will be losing variants to filtering, rather than failing to detect them. All power calculations in **Figures 1** and **3** assume that 90 percent of true mutations are discovered and not filtered out.

3.4.8 Non-coding pathogenic variation

Taking data from the Ensembl regulatory build segmentation, transcription start site associated promoters and strong enhancers, provide 1.75X-2.5X the mutation target per cell type (6X if combined across cell types), compared to coding sequence. These are within the range of size distribution of normal genes. In short, if disorders are caused primarily by regulatory mutations rather than coding mutations, similar power calculations apply. This is assuming that regulatory regions can be well defined *a priori*.

4 Opportunities to extend beyond the primary phenotypes

4 Opportunities to extend beyond the primary phenotypes

4.1 Opportunities to extend beyond rare diseases, cancer and infection

4.1.1 Lifelong electronic health record linkage

The 100,000 Genomes Project's partnership with the NHS is particularly important to deliver high-quality initial phenotyping data; a flow of electronic health data from primary care, hospital, outcomes, registries and social care records; and an opportunity to work with clinicians and patients to acquire further information on the primary conditions, associated comorbidities and outcome.[17]

The evaluation of whole genome sequencing (WGS) data in the context of rich and extended phenotypes derived from electronic health records, such as blood pressure, cholesterol, glucose, and pharmacogenomics, adds significant value. The richness of the Project dataset will allow us to move beyond the primary phenotype of the rare disease, cancer or infectious disease that led to the patient's enrolment to evaluate the WGS in the context of other continuous traits, diseases and response to therapy. To do this, we will learn from Biobank UK, engage with the Oxford Big Data Institute and the Farr Institute.

4.1.2 NHS and public health registry data linkage

Public Health England is responsible for the national and local patient-level registries for several areas relevant to the Project, including cancer and rare diseases. The National Cancer Registration Service collects data on every patient diagnosed with a cancer-registerable condition across England. The data is collected from sources covering the whole pathway from referral, screening, to palliative care. It combines diagnostic, treatment and outcome data, including individual patient reported outcome measures. Data is normalised and quality-assured. There are systems to feedback the data, alongside comparative performance reports, to individual clinical teams and clinicians. This will provide a continual incentive for its improvement.

Plans for a national congenital anomaly and rare disease registration service are well advanced and we anticipate that complete national coverage in England will be in place by the end of 2015. Similar benefits in terms of access to data are likely to accrue (as seen for the linkage to cancer data) as the rare disease registry develops.

As well as being a key source of phenotypic data linkage, this dataset will help inform the genetic makeup of the epidemiology of the diseases (by assessing how representative the sample is of all cancers/rare disease of that type).

4.2 Opportunities beyond whole genome sequencing

Genomics England has create a high quality biological sample resource from participants in support of future health research. An essential element of this collection is to provide a resource that supports scale (power) and diversity (of sample types and analytic options) across different 'omics' platforms. The plan is to collect the following from NHS England Genomic Medicine Centres and the NIGMC:

- Serum and plasma for proteomics and metabolomics.
- Cell free serum for circulating tumour DNA and to assess tumour recurrence.
- Germ-line RNA for transcriptomics.
- Lymphocyte DNA for epigenetics.
- Tumour for RNA expression profiles, tumour epigenetics and proteomics.

Dependent on additional funding, we may collect in the future:

- Cancer cell lines for study or xenotransplantation cancer models (not funded at present).
- Skin biopsies for generation of inducible pluripotent stem cells will be possible under additional consent if funds are raised (not funded at present).

The collection of these samples will ensure that biological samples obtained are suitable for assessing the genome, epigenome, proteome and metabolome in blood and tumour. Analysis of sample assays will not be undertaken immediately after collection, but rather for future research analysis. The storage and processing of samples will account for the assays that are considered most likely to be used in the future. For example in WGS, transcriptomic, epigenetic, metabolomics and proteomic research. These samples will bring potential to look at a complete functional pathway from DNA to the transcription pathway and genetic modifiers to metabolites and the associated development of companion diagnostics and other technologies.

Genomics england

5 The Genomics England Clinical Interpretation Partnership

5 The Genomics England Clinical Interpretation Partnership

The overall aim of the Genomics England Clinical Interpretation Partnership (GeCIP) is to create a thriving, sustainable environment for researchers and clinical (NHS) disease experts, trainees, and a possibly pre-competitive industrial consortium. This will stimulate and encourage new research endeavor and information exchange. This community will analyse and constantly refine the clinical interpretation of the 100,000 Genomes Project dataset. This is with the intention of further improving understanding of findings and their implications for genomic medicine and the clinical setting. GeCIP was launched at the Wellcome Trust on 27 June 2014.

GeCIP may also provide an excellent basis for further research and development in genomic medicine in the UK by academics and industry. All GeCIP users will be required to contribute results and data to the Genomics England Data Centre to enhance the scientific knowledge base. A protected period of six months may be allowed for researchers to work within private domains, after which the data are made available to the community as a whole.

The scientific aims of the GeCIP programme are as follows:

- Enhanced clinical interpretation focused on rare disease, including clinically or genomically-driven deeper phenotyping, novel approaches to interpretation and annotation, validation and functional characterisation of variants, identification of novel therapeutic targets, or repurposing of existing therapies. Rare disease will also be informed by other 'omic' investigations.
- Innovative clinical interpretation in cancer, including multi-omic datasets (e.g. transcriptomics, epigenetics, and proteomics), analysis of circulating tumour DNA, sequential biopsy to address the genetic architecture of cancer, validation and characterisation of variants, identification of novel therapeutic targets, or repurposing of existing therapies.
- Improved clinical interpretation in infectious disease, focused upon individuals with severe outcomes in sepsis, or – in partnership with Public Health England – greater understanding of the spread of antimicrobial resistance, and phylogenetic tracking of transmission across the whole of the health economy.
- Expand the programme to include other disease areas with additional funding, to address specific research questions and opportunities to develop stratified approaches. The Genomics England infrastructure will be designed to facilitate expansion and re-use, and individual GeCIP partnerships can be extended to programmes, with funding, outside of the Project.
- To capitalise upon electronic health records research, exemplified by the rapidly-developing capacity of the Farr Institute and the Oxford Big Data Institute. This can build upon and add value to the clinical, laboratory and health records data, linked to variant call data, which is held securely within the proposed data and computing infrastructure.
- Algorithms, models, and tools for clinical genomics research, data quality assurance, and the annotation, interpretation and presentation of genomic, clinical, and laboratory data in combination, may be developed, evaluated, used, and shared within the proposed infrastructure.

All GeCIP partners will adopt the highest ethical standards in accordance with the terms of the Genomics England consent. This will be with the continued support and advice of the Genomics England Ethics Advisory Committee.

The activities of GeCIP will inform NHS feedback to clinicians and the multidisciplinary teams by providing enhanced data interpretation, additional information on pathogenicity of variants, and functional characterisation. The process of NHS feedback and issue of reports will be managed by Genomics England. The details for this process are outlined in **Section 6**.

Expected pathogenic variants or putative novel disease genes

Some of these findings may be **expected to be pathogenic** or may be of no consequence i.e. variants of unknown or no clinical significance. Genomics England will use the GeCIP domains to evaluate and develop the evidence for presence or absence of pathogenicity. These will also be available to the clinical teams in the NHS who directly care for patients. They will be encouraged to join in the relevant GeCIP domain. It is conceivable that enrolment of additional family members will allow segregation patterns confirming pathogenic status. However, these findings will only be fed back to patients if they transfer from expected pathogenic to known pathogenicity and are validated within the NHS thereby becoming clearly clinically actionable.

5.1 The route by which key funders interact with GeCIP

We are working with key funders to establish the Genomics England Clinical Interpretation Partnership. This offers a mechanism for public, charity, and philanthropic funders and the researchers that they support to engage with the Project and add value in terms of clinical interpretation and collaborative research and training. These will be known as GeCIP funders.

5.2 The route by which funded researchers add value to GeCIP

It is the ambition of Genomics England, and all its funders (GeCIP partners), to see a really strong research and training programme alongside the Project. This will create opportunities for GeCIP researchers and trainees, which will stimulate and support research programmes funded by GeCIP funding partners, through response mode and specific funding calls. The researchers leading these programmes will have access to appropriate patient groups from the NHS in the UK (see later for details related to the Devolved Nations). They will be expected to form and lead appropriate national and international consortia within cognate domains that provide NHS samples, clinical data, and analytical skillsets. This will maximise the value of the dataset value and ensure that the clinical interpretation delivered by the Project remains at the forefront of genomic medicine. This may be achieved in a variety of ways, including:

- Enhanced clinical interpretation focused on rare disease. Omic data will also inform this.
- Innovative clinical interpretation in cancer, including multi-omic datasets.
- Improved clinical and public health interpretation of pathogens and infectious disease.
- Expanding the Project to include other disease areas.
- A researcher or trainee funded specifically to work within the Project by a GeCIP funding partner will become a member of the GeCIP, providing they agree to work according to the Genomics England policies. If a research proposal or training fellowship is planned, it is recommended that applicants contact the Chief Scientist's team, so they can ensure compatibility with the mission and assist where necessary.
- Delivery of health records research using the linked, longitudinal dataset.
- Delivery of algorithms, models, and tools to facilitate interrogation and analysis of linked, longitudinal datasets.
- Work as part of the clinical validation groups to review data and interpret findings (see below).
- Commitment to provide results or findings to the Genomics England Knowledge Base.

These will be known as GeCIP researchers or trainees.

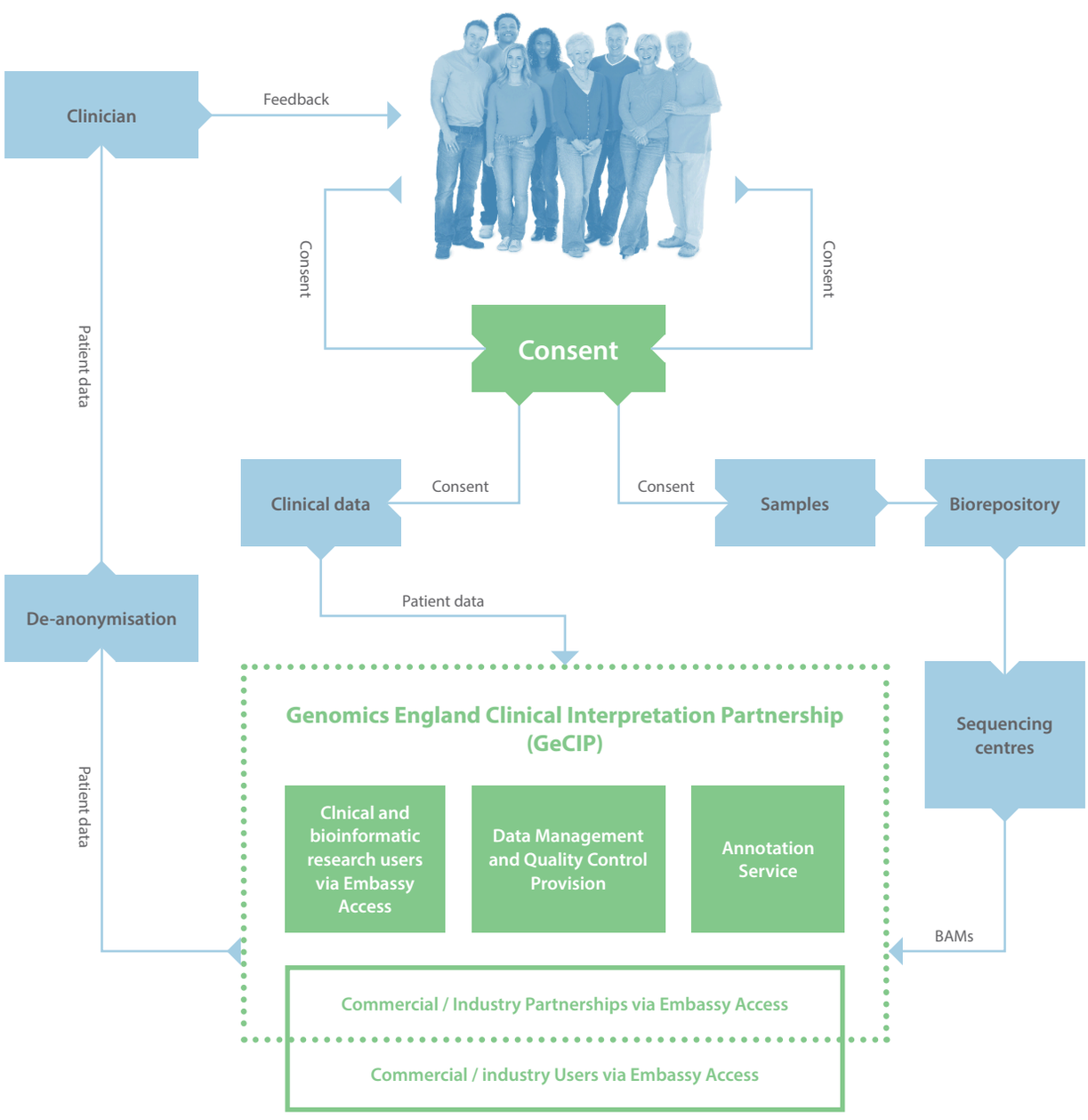
5.3 The route for those who contribute patients and data from the NHS to add value to GeCIP

The GeCIP partnership will be the route for NHS and other organisations who recruit patients, contribute samples, and/or data to the programme, or those funded on specific NHS Training programmes related to Genomics England (NHS Trusts, NHE England, Public Health England, and Health Education England). This offers the opportunity for those in contributing NHS organisations to participate in the analysis and interpretation of sequencing to inform release of clinical reports back to the NHS.

Contribution from the NHS could include:

- NHS contributor who adds value to the GeCIP (e.g. samples, data, annotation, tools etc.).
- Delivery of updated clinical, phenotypic, and laboratory data on patients (through existing Genomics England tools and pathways).
- Working as part of the clinical validation groups to review data and interpret findings.
- Committing results/findings to Genomics England knowledge base.

Figure 5. The data flows that enable the Genomics England Clinical Interpretation Partnership to function.



5.4 Establishing and working within the Clinical Interpretation Partnership

The GeCIP has been organised into a series of disease-specific domains by an open call, in which UK researchers, trainees and the NHS self-organised and created proposed GeCIP domains. Key selection criteria were published in the GeCIP guidance document accompanying the call. Successful GeCIP proposals clearly articulate the value they will add to the clinical interpretation of the Genomics England dataset.

Researchers, NHS clinicians/healthcare professionals and trainees will work as part of **disease-specific or function specific domains** (depending on their area of expertise), contributing patients, phenotypes, knowledge, expertise, and undertaking research to add value. This work will be continuous, as the dataset is updated with patient outcomes and detailed clinical, phenotypic and laboratory data, which is collected during routine NHS care. Some clinicians will work in several domains if they have a broader clinical or laboratory remit.

The list of clinical domains described below is representative of the diseases we are now working on in the Genomics England pilots. It is subject to change at the discretion of Genomics England. In our **call for GeCIP domain proposals**, we encouraged the applicants to self-organise and make their own domain proposals. This may include domains not mentioned below. They will also be encouraged to propose a UK Lead and we will work with applicants to optimise the GeCIP domain as needed.

Table 1 - GeCIP domains.

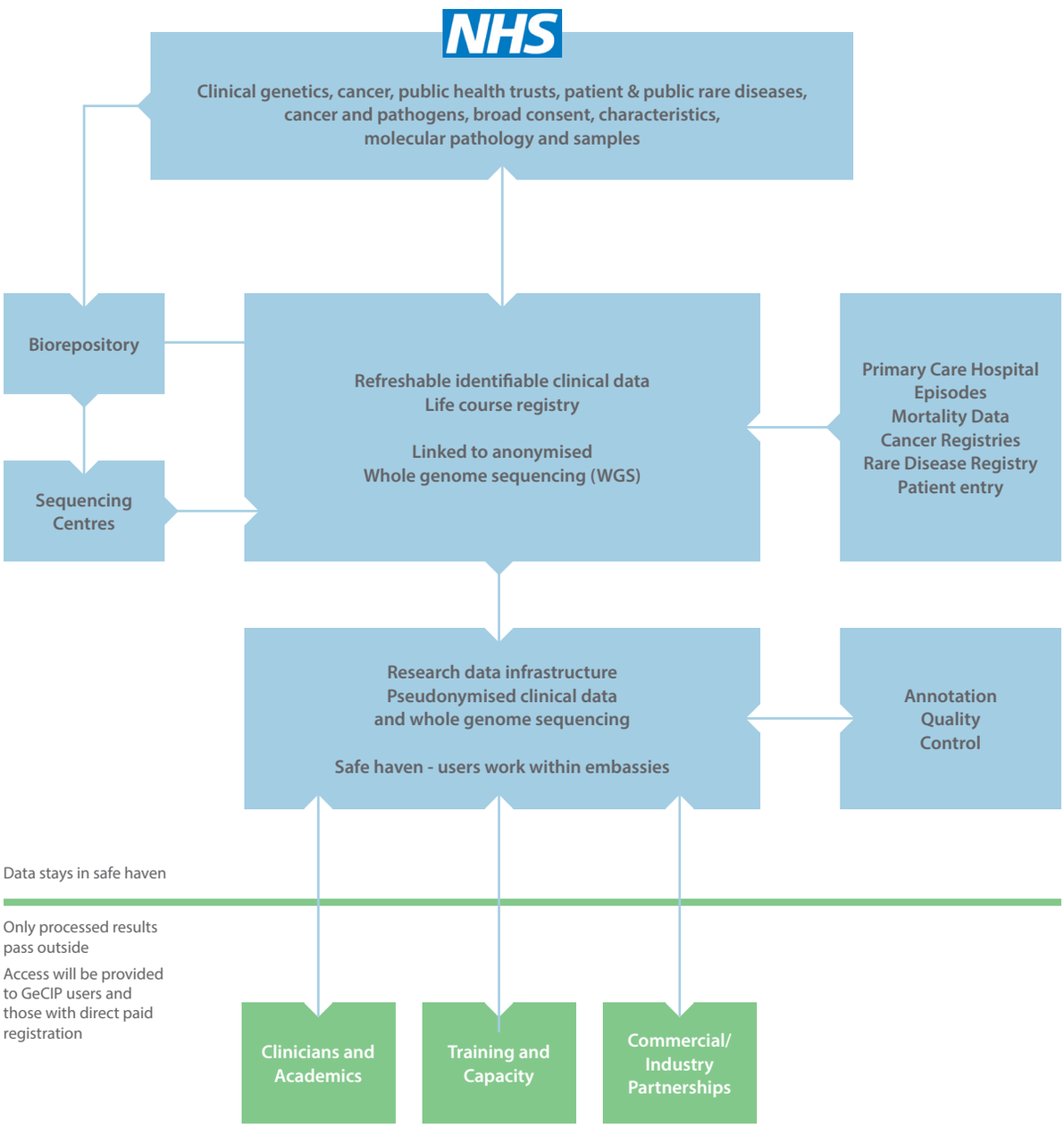
Disease-specific domains		Functional/Cross-cutting domains
Cancers	**Rare inherited diseases**	
Breast	Hearing and Sight	Analytical
Lung	Cardiovascular disease	Ethics and Social Science
Ovarian	Respiratory	
Colorectal	Endocrine and metabolism	
Prostate	Gastroenterology	
CLL/leukaemia	Immunological diseases	
	Neurological and degenerative disease	
	Musculoskeletal	
	Skin	
	Renal	
	Haematology	
	Paediatric	
	Inherited Cancers	
	Severe response to infection	
	Rheumatology	Feedback and Validation

The GeCIP domains can be led and proposed by researchers, NHS clinicians/healthcare professionals and those training anywhere in the UK. They are not confined to England.

Through an investment by DoH NI and the MRC, and in conjunction with Genomics England, the NIGMC has been established within HSC NI. The NIGMC is a virtual centre hosted by BHSCT, and staff contribute through membership of the GeCIP disease and function specific domains.

Figure 6.

Genomics England Clinical Interpretation Operational plan



At this time Genomics England is unable to fund sequencing and analysis for NHS patients from Scotland and Wales. However, these patients will be sequenced if the costs of sequencing and analysis are funded by the Devolved Nation Administration or a GeCIP funder. Patients outside the UK cannot be sequenced in the Project, unless full funding is provided by a funder and they join a GeCIP domain and agree to Genomics England Terms and Conditions.

5.5 Accessing data as a Genomics England Clinical Interpretation user

Individual GeCIP domains relating to a particular research study will be provided with access via a dedicated secure virtual data centre. These domains will have specific sub-sections, or 'embassies', and comprise of the data related to that domain and computer resources housed within the Genomics England infrastructure. These embassies are intended to be seen as areas in which clinical teams, research groups, trainees and pre-competitive industry partners undertake their work. Work undertaken within the embassies is subject to the governance and terms and conditions of Genomics England (see below). Breaches of the governance, or terms and conditions of use, will result in the removal and exclusion of those responsible from the Genomics England Data Centre.

Access will be granted via a data access registration and approval process. Embassies will be generated and hosted within the Genomics England Data Centre in compliance with governance and terms of conditions set out by Genomics England.

All research data will be de-identified in accordance with their REC approved protocols.

GeCIP users will be granted access to data and knowledge held within the 100,000 Genomes Project dataset which relates to the **GeCIP domain** to which they belong. For example, investigators undertaking a research project in ovarian cancer would be granted access to all the anonymised data held within the Genomics England embassy for the GeCIP ovarian cancer domain.

The table below illustrates a possible structure of a domain.

Table 2.

Genomics England Clinical Interpretation Partnership Disease-specific Domain

Activity/Group	Key functions and outputs
Genomics England Chief Scientist's Team	• Oversight, informatics and logistics
GECIP Steering Committee	• Coordination and management of domains
Clinical Interpretation including NHS clinicians	• Highest fidelity dynamic reporting system • Contribution to Interpretation, Validation and Feedback domain
Genomics researchers	• Novel genomic discoveries • Engagement of international collaborators
Multiple phenotypic sub-groups	• Deeper phenotyping and sub-analyses
Analysts and Bio-informaticians	• Novel analytic approaches • Development of analytical and reference resources
Functional characterisation Multi-omics	• Single cell or model functional studies • RNA, epigenetics, proteomics
Trainees and training director	• Research projects and higher degrees through the Genomic Medicine Academy
Precompetitive industry partners	• Academic-industry collaboration to accelerate application of new findings

Note: It is anticipated that precompetitive industry partners will work within the GeCIP domains. All data generated in a GeCIP domain remains available to all.

5.6 The commitment of all data generated to the Genomics England Knowledge Base

Genomics England will encourage and promote an environment of open collaboration. GeCIP members will be expected to comply with this approach to enhance and maximise the value of the Genomics England knowledge base.

In all cases the research results, along with raw data and analytical steps undertaken within the GeCIP domains, must be provided to the Chief Scientist's team, for commitment to the Knowledge Base.

Genomics England is committed to delivering a managed research environment and will negotiate with GeCIP members regarding sharing of intermediate research results within the knowledge base. While it is understood that this is a sensitive area, it is hoped that such sharing will serve to accelerate scientific progress, delivering further patient benefits by enhancing feedback to the NHS.

5.7 GeCIP contribution to interpretation, validation and regulation of findings

The GeCIP domains and committees are expected to collaborate with regulators, including E-Quality Management System, National Institute of Health and Clinical Excellence, Medicines and Healthcare Regulatory Products Agency and NHS England to develop standardised approaches to validate findings from the Project. In addition, this collaboration is expected to support the development of systematic approaches to verification of these scientific findings, evaluation of proposed applications on diagnosis and treatment options and management of impacts upon commissioning strategies and costs. This joined-up-approach will ensure swift translation of clinically validated research findings into patient benefit.

5.8 Overarching GeCIP management and governance

The GeCIP Board, chaired by Professor Dame Kay Davies, will oversee the governance of GeCIP. Membership will also include a GeCIP lead for the NHS contributors relating to cancer, rare disease and infectious diseases as well representatives of major funders. Genomics England will be represented by its Chief Scientist. Further details on the governance structure are available at **www.genomicsengland.co.uk**.

A GeCIP Steering Committee will be established to ensure the smooth running and transfer of knowledge across the GeCIP domains. The membership will consist of leadership from each disease-facing domain. It will be chaired by the Chief Scientist for Genomics England. The GeCIP Steering Committee will directly report to the GeCIP Board. This committee will include representatives of the public, patients and medical research charities.

Genomics england

6 Feedback for participants in the 100,000 Genomes Project

6 Feedback for participants in the 100,000 Genomes Project

It is recognised that the Project, and in particular, GeCIP, will contribute to the building of an evidence base for clinical utility and patient benefit. It is our goal to use this mechanism to improve the relevance of whole genome sequencing (WGS) findings for healthcare during the life of the Project. The Project/GeCIP is committed to regularly publishing peer-reviewed updates on the current state of knowledge in this area. In preparing the Genomics England Feedback Policy, we took full account of all relevant legal and ethical principles including contemporary research and the advice of patients. This has also been informed by Professor Mike Parker's letter to the Chief Medical Officer from the 100,000 Genomes Project Ethics Advisory Working Group (see **www.genomicsengland.co.uk**).

GeCIP will be operated by the Office of the Genomics England Chief Scientist, with two major primary functions: (i) its role with regard to research and the research community (described above) and (ii) responsibility for clinical validation of genetic variants of expected pathogenicity.

6.1 Data feedback to NHS clinicians on Genomics England participants

Diagnostic capability to detect pathogenic variants is expected to evolve over the life of the Project, providing feedback to the NHS via a dynamic reporting system, consisting of annual or bi-annual reports circulated through NHS email systems to named clinicians. These reports will focus entirely on findings of clinical relevance containing the following:

- A Genomics England identifier to be used at the NHS Trust level to identify the individual patient or family member.
- The original clinical diagnosis.
- **Known or pathogenic variants or driver mutations** directly connected to the main disease that led the patient to engage in the Project are clearly actionable. These must be fed back to patients.
- **Secondary findings.** A very limited number of **looked-for known, pathogenic mutations** of high clinical relevance, confined to a very limited list which may be expanded on the advice of our dedicated GeCIP domain, our working groups, the Science Committee and the Ethics Advisory Committee. This will be more limited than that used by the American College of Medical Genetics. It will be possible for patients to opt-out of the feedback of findings of high clinical relevance.
- **Parental carrier status** in cases of rare disease due to bi-allelic germline mutations. It is feasible that this programme will identify double parental carrier status or X-linked maternal carrier with high-risk of a reproductive consequence.

Initial clinical feedback to the NHS teams will be provided as a PDF via secure, encrypted NHS Mail or equivalent to the clinician named by the participating NHS site. NHS England will need to consider how to develop a more systematic means of enabling this feedback, including consideration of care record integration and identifier management in the future.

Staff training, awareness and professional development needs to be planned between Genomics England, NHS England, HSC NI, and Health Education England (as appropriate). This will ensure the benefits of WGS for patients and embed genomic medicine into the NHS for the long term.

6.2 Validation of findings from the 100,000 Genomes Project

Whole genome sequencing is not yet an approach that has attained the specificity and sensitivity of a clinical grade test. One of the goals of this Project will be to drive up the quality of next generation sequencing.

At this time, it is therefore essential that validation of our key diagnostic findings are undertaken within the NHS. It may be necessary to propose a systematic approach to validation of findings from the Project, which can be adopted in the NHS over the life of the Project. This will be done in partnership with GeCIP domains, NHS England, HSC NI and the clinical teams. A part of the role of the NHS Genomic Medicine Centres (GMCs) (including the NI GMC) will be to commit to validation of findings and share these results with Genomics England to enhance the value of the Data Centre.

6.3 Policy for pertinent, secondary, and incidental findings, and feeding back to clinicians and patients

The Project policy on the feeding back of clinically relevant findings to clinicians and patients is summarised below. The Project will take account of the evolving policy and academic debates in the area and has been influenced by these in respect of its current formation.[18]

100,000 Genomes Project policy on feedback to participants' clinical teams

Table 3. Summary of 100,000 Genomes Project policy on feedback of findings:

Type of finding*	Description	Nature of the information to be fed back	How feedback will happen	Approach to consent or refusal
Pertinent finding (also known as a primary finding).	A pertinent (primary) finding is relevant to the explanation, main diagnosis or treatment of the disease for which the patient was referred into the 100,000 Genomes Project by their clinician (i.e. the patient's cancer, rare disease or response to infection).	To be reported as a pertinent (primary) finding, **mutations must be 'Known to be Pathogenic' (KP)** i.e. sequence variation is previously reported and is a *recognised* cause of the disorder. **Mutations that are 'Expected to be Pathogenic' (EP) will be referred to the appropriate GeCIP domain for confirmation. NHS clinicians will be integral to confirmation of pathogenicity through GeCIP. This will not be fed back to patients unless an EP mutation becomes a KP.** i.e. sequence variation is previously *unreported* and is of the type which is *expected* to cause the disorder.	Pertinent (primary) findings will be given to the referring clinician or clinical team by the 100,000 Genomes Project for discussion with the patient after the NHS has validated these findings. It will be emphasised to participants that findings of 'Known Pathogenicity' do not (usually) equal 100% penetrance.	A participant's consent to receive pertinent findings will be integral to their participation in the 100,000 Genomes Project. If such consent is refused, participation will not be permitted.

* Classification of results is adapted by Genomics England taking into account the 'Bioethics Commission's Classification of Individualized Results of Medical Tests', P.27 table 1.2, of the Presidential Commission for the study of Bioethical Issues report, 'Anticipate and Communicate: Ethical Management of Incidental and Secondary Findings in the Clinical, Research, and Direct-to-Consumer Contexts', December 2013. http://bioethics.gov/node/3183. The utilisation of terms by Genomics England has been chosen to provide the greatest possible clarity for Genomics England's activity. Broadly speaking we identify two groups of results; primary/ pertinent and secondary/ non-pertinent. Within the secondary findings group we envisage that some findings will need to be (a) Actively looked for because they reach an evidence/ utility threshold/ on a predefined list, for which consent is sought at the time of testing and that (b) a second group, expected to occur rarely, are of health-related findings not actively sought because they fail to meet the evidence/ utility threshold, but where disclosure may still need to be considered, should they happen to be found. Findings in this second group may have personal utility. Those wanting to access their raw codes might derive this, but Genomics England will not routinely filter for these findings and disclose them.

Type of finding*	Description	Nature of the information to be fed back	How feedback will happen	Approach to consent or refusal
Secondary finding (additional looked-for findings of healthcare importance).	A secondary finding is an additional, looked-for health related finding, that is **not** pertinent to (or a primary cause of) the main condition. It may be found *in addition to* (or *in the absence of)* any pertinent finding.	Genomics England will also look for genomic findings that are known to cause serious conditions for which there is good evidence that knowing about them could influence healthcare. We will start with a limited list of relatively rare conditions, but will adapt this list as more evidence accumulates: • **Hereditary non-polyposis colorectal cancer (HNPCC)/ Lynch syndrome** (genes: mismatch repair genes MLH1, MSH2, MSH6, PMS2) - adult onset** • **Familial adenomatous polyposis (FAP)** (gene: APC) • **MYH-associated polyposis (MAP)** (gene: MutYH) • **Hereditary, breast and ovarian cancer** (genes: BRCA1 and BRCA2) - adult onset • **Von Hippel-Lindau syndrome** (gene: VHL) - child and adult onset • **Multiple endocrine neoplasia type 1** (gene: MEN1) - child and adult onset • **Multiple endocrine neoplasia type 2** (gene: RET) - child and adult onset • **Familial medullary thyroid cancer (FMTC)** (genes: RET and NTRK1) - child and adult onset • **Retinoblastoma** (gene: RB1) - child onset • **Familial hypercholesterolaemia** gene: LDLR- child onset- and also APOB and PCSK9 - child and adult onset*** Please note: the Genomics England approach to findings relating to adult onset conditions in children will be in accordance with the BSGM policy on genetic testing of children** i.e. only the mutations in genes that are known to cause childhood-onset disease will be looked for in the case of minor participants.	Participants will be asked whether they would like these secondary findings to be actively sought. If they consent to this, information relating to secondary findings will be looked for and given to the referring clinician or clinical team for discussion with the participant. (Without this consent these secondary findings will **not** be actively sought.) Following positive or negative confirmation regarding these secondary findings, a health professional will feed back to the participant on these results. Some variants in this gene list may not yet have sufficient evidence for their clinical impact to be known, or clearly predicted. These variants of uncertain significance will not be reported. An important aim of the 100,000 Genomes Project is to generate evidence about the clinical relevance of new findings. Our understanding of the clinical significance of a secondary finding will change over time as more evidence is gathered.	Consent for the feedback of these secondary findings will be by opt-in to feedback. Patients who do not wish to receive information about these findings are free to refuse to consent. Participants can opt-in to consent to the list(s) of identified conditions to be looked-for at the time of consent, plus if they wish, to findings made as that list extends. This will permit results regarding other conditions that meet this criteria in future to be looked-for and fed back.

** This list is based upon that proposed by ACMG plus subsequent expert review at Genomics England.

Green RC *et al.* American College of Medical Genetics and Genomics recommendations for reporting of incidental findings in clinical exams and genome sequencing. Genetics in Medicine 2013; 15(4): 565-574. Available at: https://www.acmg.net/docs/IF_Statement_Final_7.24.13.pdf.

See also ACMG Updates Recommendation on "Opt Out" for Genome Sequencing Return of Results, April 2014. Available at: https://www.acmg.net/docs/Release_ACMGUpdatesRecommendations_final.pdf.

*** British Society for Genetic Medicine 2010. Genetic Testing of Children. Available at: http://www.bsgm.org.uk/media/764635/wgs_discussion-paper.pdf.

Type of finding*	Description	Nature of the information to be fed back	How feedback will happen	Approach to consent or refusal
Carrier status Under certain circumstances carrier status may affect future children.	Autosomal recessive. Where both parents participate, we will be able to identify a limited list of double carrier states where knowledge might affect future family planning. X-linked disorders in women carriers.	Sickle cell anaemia Cystic fibrosis Beta thalassemia Congenital adrenal hyperplasia 21-hydroxylase deficiency Alpha thalassemia Spinal muscular atrophy type I Duchenne muscular dystrophy Adrenoleukodystrophy Haemophillia A (inversion)	Participants will be asked whether they would like these carrier states to be actively sought. If they consent to this, information relating to carrier states will be looked for and given to the referring clinician or clinical team for discussion with the participant. (Without this consent these carrier status will **not** be actively sought.) When either positive or negative results are confirmed regarding these looked-for findings, a health professional will feed back to the participants.	A list of conditions for opt-in looked-for carrier status. For double carrier of recessive feedback will be undertaken for those who opt-in, provided **both parties** in the couple consent to seeking and return of these results **as a couple**. Or a woman carrying an X-linked disorder consents to opt-in to feedback.

Type of finding*	Description	Nature of the information to be fed back	How feedback will happen	Approach to consent or refusal
Incidental findings A sub-category of additional findings that are not actively sought. Also described as 'unsolicited' findings****	This is any secondary health-related finding that has not been included in the gene list above.	In the clinical phase of the 100,000 Genomes Project, such findings will usually not come out of the analysis, but just occasionally it is possible that in diagnosing the Primary finding, an Incidental finding is noted that does not belong to the gene lists above. These examples are expected to be very rare, but will need discussion and evaluation within GeCIP in order to decide whether the evidence is certain enough in regards to clinical relevance that disclosure to the clinician is appropriate. In most cases incidental findings in this group will be found in the research stages of Genomics England and will only be utilised in the clinical setting once they are assigned to the list of secondary findings above.	*Incidental* findings will usually not be sent to clinicians or patients by Genomics England.	The policy on not feeding back incidental findings will be explained to the patient at the time of consent.

**** Anneke L notes: Although these terms have no clear definitions and their meanings may be more or less relevant to the different parties involved. For example, a patient may find a secondary finding unexpected, but a laboratory worker analysing a whole genome should expect to possibly find anything.

Background to the Genomics England feedback policy

At the outset, the sequence data generated by the Project will require a validation step to directly inform clinical management of any individual case. This policy aims to support participating centres in handling the feedback of sequencing results. The Project, as implemented by Genomics England, in conjunction with the Department of Health, NHS England, DoH NI and HSC NI, needs to ensure robust management processes are in place to handle potential clinically significant findings should they arise. This policy will be agreed between Genomics England, the Department of Health, NHSE, and approval will be sought from the NHS Research Ethics Committee.

In common with all medical research involving human participants, research on the information gathered from whole genome sequencing has the potential to generate health-related findings of potential health or reproductive importance to individual participants. The debate about how these findings should be managed is ongoing, particularly when the findings purposely, or inadvertently, go beyond the primary aims of the original investigation, generating results beyond the scope of which the participant may have expected to receive (or consented to receive).

The policy needs to accommodate the hybrid aims of the Project of clinical transformation and future research. There are no national guidelines from the National Research Ethics Service regarding the feedback of clinically significant findings in genetic research. With the exception that, other than that during the consent process, information should be provided about any procedures for feedback of 'individually significant' information, and that any such feedback should be explicitly consented to. Recent policy activity has echoed this.[19] Many of those accessing information, or conducting research on samples or information collected as part of the Project, may have a dual role as both a treating clinician and researcher.

Evidenced views from research participants, health professionals, or relevant patient groups on optimal ways to manage this feedback are acknowledged to be scarce, albeit with evidence that when asked in surveys, the public and research participants welcome the return of life-threatening but treatable health-related research findings to participants.[20,21] Genomics England has put in place a programme of public engagement and social science/ethics research activities in order to contribute to, and learn from, an evidence-base in this developing area of practice. These activities are also supported by the Ethics Advisory Committee (more on this below).

Activities to inform policy development

The relevant work streams have assessed recent evidence and policy contributions in this area.[22] The Project and in particular, the Genomics England Clinical Interpretation Partnership (GeCIP), will contribute to the building of an evidence base for the clinical utility and relevance of the WGS findings, during the life of the Project. GeCIP will provide data and value-added services, to the scientific and clinical research community, in order to improve collective interpretation, knowledge, and understanding of the data. The Project has committed to regularly publishing peer-reviewed updates on the current state of knowledge in this area. The Project will also take account of the evolving policy and academic debates in the area, and has been influenced by these in respect of its current formation.[23]

Genomics England believes that the duration of the 100,000 Genomes Project offers a unique opportunity for:

(i) Consultation on the area with patients, the public, and other key stakeholders on feedback.

(ii) Conducting social science research, including on experiences of, and attitudes to, the feedback of clinically actionable findings.

(iii) Analysis and deliberation, informed by this evidence base, on what would constitute a reasonable and ethically justified policy on the feedback of findings in the context of the NHS.

Genomics England has begun a programme of research, seeking opinions on feedback from patients and their relatives participating in the Project, the healthcare professionals working with them, patient groups and the general public. This research will continue from pilot stage to completion of the Project. The company has taken advice from various stakeholder communities in relation to policy issues.

The key elements of the Genomics England feedback policy

The Wellcome Trust and Medical Research Council call on projects to:

- Have a policy that indicates whether or not health-related findings will be fed back to individuals that can be clearly articulated. And be able to demonstrate the reasoning behind their policy to research participants, funders and the Research Ethics Committee.
- Include clear information on the study policy on the feedback of health-related findings in the consent process.
- In cases where the policy is to provide individual feedback on health-related findings, develop a practical feedback pathway that is adequately resourced[24]

The Presidential Commission for the study of Bioethical Issues recommends that *"researchers have a plan for the ethical management of incidental and secondary findings that arise in research, put in place ahead of time and clearly communicated to potential participants that is vetted and supported by an IRB or expert review structure"*.[25]

The feedback policy for the 100,000 Genomes Project

The Project's policy on the feedback of clinically relevant findings to clinicians and patients might be described as a relatively conservative policy of 'qualified disclosure'.[26] The predefined criteria for feedback are part of the participant consent process. These criteria are expected to remain consistent at least over the life of the Project until 2017. Even within this timeframe, the extent of findings returned are potentially variable, within these criteria, as new evidence emerges.

Genomics England will undertake an initial analysis to identify findings pertinent to the disease or disorder that led to enrolment of the patient as soon as possible after recruitment. This will be based on the policy outlined in **Table 3** above. As relevant knowledge is increased e.g. from published evidence or from the research or GeCIP activity using the Project's dataset, it may become appropriate to report newly-identified *pertinent findings* (pertinent to the condition for which the participant joined the project). Feedback beyond the initial report is covered in the Consent Form.

In the same way, it may become appropriate to look for and report newly-included *secondary findings* where seeking and reporting these has been specifically consented to.

The decisions regarding what are pertinent findings, or should be looked-for secondary findings, will be undertaken within a specific mechanism in GeCIP. The responsibility for making this determination will remain with the Office of the Chief Scientist at Genomics England, in consultation with our working groups, and advisory committees. The decisions around the status of pertinent findings, expected pathogenic mutations, secondary looked-for findings and carrier status to be fed back, will involve the appropriate disease-specific domains within GeCIP. The decisions around expanding the list of looked -for findings (under the agreed criteria) will be made by a specific Feedback and Validation domain within GeCIP. This GeCIP domain will include in its membership: clinical genetics, cancer genomics clinicians and scientists, research, clinical, NHSE and patient representation. The NHS contributing clinicians may interrogate the data on their own patients further if they wish to, and as needed, they may consult GeCIP and their local multi-disciplinary teams on this. Clinicians will undertake to inform the Project of the actions taken in respect of findings and their outcomes, in order to improve knowledge about the disease areas included in the Project.

Defining 'Known' and 'Expected' Pathogenicity

In defining 'Known' and 'Expected' Pathogenicity, the Project has adopted the ACMG laboratory Quality Assurance Committee Working Group on Standards for interpretation of Sequence Variations Revisions 2007.[27] These recommend the following interpretative categories of sequence variations for the purposes of clinical reporting:

Known Pathogenicity

This is a sequence variation that has been previously reported, and is a recognised cause of the disorder. It is recommended that a review of the literature, central mutation databases, (e.g. Human Gene Mutation Database), or appropriate locus-specific databases, to assess the current degree of certainty that the sequence variation is causative of the disorder should be undertaken before reporting.

Expected Pathogenicity

This sequence variation is previously unreported and is of the type which is expected to cause the disorder. Examples include variation that is predicted to result in the introduction of a stop codon; alter the sequence at a splice junction, particularly the invariant AG/GT nucleotides, or delete one or more nucleotides, or exons in such a manner as to lead to a shift in the mRNA reading frame. The Project will define the term 'pathogenic' to reference mutations categorised of either known or expected pathogenicity in regard to feedback, unless otherwise specified.

Feedback of pertinent findings of known pathogenic mutations to the affected participant

Genomics England, and our Clinical Interpretation Partners, will generate pertinent findings which include known pathogenic mutations for the disease, which are clearly clinically actionable. These will be reported to the NHS teams caring for the patients, and they will confirm validity of these pertinent findings and following this will communicate these findings to the participant (and possibly their relatives). Participants are invited to join the Project because they have received NHS care in England or HSC care in Northern Ireland for medical conditions prioritised by the Project. Their relatives may be invited to join as a result of this.

Feedback to other family members

Within their standard clinical care, tests in family members will not be administered without gaining appropriate consent and feeding back the results. As such, if analysis undertaken on the samples and/or information of participants reveals clinically-relevant, pertinent information, they should receive these results. Secondary findings will also be reported back to participating family members (if they have consented to receive these). For example, a patient referred because of a cancer can expect to receive all results relating to possible causes of and treatments for that cancer. Prospective participants must consent to receiving pertinent findings that are clinically actionable as a mandatory part of joining the Project.

Potentially or expected pathogenic mutations

Genomics England will use the GeCIP domains to evaluate and develop the evidence for the presence or absence of pathogenicity. These findings will be available to the clinical teams in the NHS who directly care for patients. It is conceivable that enrolment of additional family members will allow segregation patterns confirming pathogenic status. However, these findings will only be fed back to patients if they transfer to Known Pathogenicity and are validated to become clearly clinically actionable.

Variants of unknown clinical significance (VUS) in pertinent genes

Large numbers of VUS may be identified by WGS, including in known genes associated with the health condition that initiated recruitment into the Project. Where the clinical management of the patient is not likely to alter on the basis of the VUS (if confirmed in a clinically accredited laboratory), it is recommended that VUSs are **not** reported as these are not clinically actionable. Data may emerge in the future that clarifies the pathogenicity of such a variant. However, there is no current obligation on the Project to continue to interrogate the data beyond the lifetime of the Project, after the primary analysis.

Secondary findings: looked-for, clinically relevant major impact

There is a known subset of clinically actionable variants with significant potential to prevent disease morbidity and mortality, if identified before symptoms become apparent. Genomics England has therefore developed two short lists of looked-for secondary, or additional clinically relevant findings for disclosure to patients. This will only be if they opt-in during prior consent process. These lists are:

1. The first is a limited list of germline variants with potentially severe impact, that are clinically actionable causes of rare disease, where a healthcare intervention or screening programme might prevent an untoward outcome[28] i.e. they are known to result in illness or disability that is clinically significant, severely or moderately life threatening **and** clinically actionable.
2. The second of these lists would permit parents within a parent-offspring-trio to have the option to consent to receive feedback on a limited list of rare diseases, where **carrier status** is for example: double carrier status of a known pathogenic variant as a couple e.g. for autosomal recessive conditions, or female carrier of an X-linked known pathogenic variant with high reproductive consequences for future children.

   This list would be limited to carriers of severe autosomal recessive conditions, resulting in illness or disability that is clinically significant, severely or moderately life threatening **and** clinically actionable. Such conditions will be identified in the parents of children with rare diseases recruited to Genomics England, according to the couples' consent.

   Most people are carriers of at least one severe autosomal recessive condition, and this carries no health implications (and so does **not** meet the above criteria for the return of results). Therefore, in these cases, results will only be sent to clinicians to explain to participants e.g. a proband child's parents if *both* parents are found to be carriers of *the same* condition. Results will therefore only be offered as 'couple results' rather than as two individual sets of results.

*Opt-in consent process to receive secondary looked-for findings and carrier status**

Individual adult participants (or their consultees or equivalent in the form of advice), or suitable adult (in regards to child participants) will be asked to give their explicit consent (opt-in) to permit additional findings and to receive the results of these. These will be discussed at the point of consenting or when a consultee's advice is sought.

Consent to searching and return of results in respect of either or both list of conditions is not a mandatory part of joining the Project. In respect of both lists, consent will be given to 'any other conditions that meet these criteria' over the life of the Project as these lists evolve as part of the consent to looking for this information.

These lists will be regularly reviewed and updated by experts within GeCIP (as described above) to guide this policy. The responsibility for this reviewing mechanism will lie with the Office of the Chief Scientist at Genomics England, under their responsibility to provide the service support to GeCIP. Overall responsibility for this process will remain with the Board of Genomics England. Changes to the shortlist of conditions would require the support of NHS England who would need to fund clinical support to patients.

Looked-for additional findings - regarding conditions which onset in adulthood

The adult who gives consent for the inclusion of a child participant in the Project, can also choose to consent (or not) to permit additional findings to be looked-for in regards to the child, and to then receive these results. If this consent is given, these additional findings will only be sought in respect of conditions on the 'additional findings' list which have symptoms which onset in childhood (meaning under the age of 18 years) and which meet the programme's criteria of resulting in illness or disability that is clinically significant and severely or moderately life threatening and clinically actionable. Consent will also be requested to cover the expansion of this list as scientific advice evolves over time, so the initial consent to looked-for findings will also cover 'any other conditions that meet these criteria' over the life of the programme.

*Please note that the Mental Capacity Act 2005 does not apply in Northern Ireland. Mental capacity is currently governed by common law in Northern Ireland and as such this will be applied in terms of the consent process for Northern Ireland participants.

Genomics England will follow recommendations of the British Society for Genetic Medicine on seeking and feeding back on looked-for additional clinically relevant findings with onset in adulthood.[29] This will mean, for example, that additional findings relating to an adult-onset condition will not be looked-for, nor will the results be included in the standard report generated for the child's clinical team, to be shared with a child, nor the parents of any child under the age of 16 (or under the age at which the young person is able to consent on their own behalf to participation in the 100,000 Genomes Project).

However, given that a whole genome sequence will be generated in all cases and these data will be accessible to a patient's treating clinician they could investigate a gene sequence for variants causing a condition of adult onset. Here we would expect clinical teams to follow best practice and guidelines and therefore they would not be prevented from doing this by Genomics England. Additionally, a parent to whom Genomics England had made available a copy of their child's genomic information, (on request, in Variant calling file or BAM file format) would be free to investigate this sequenced information, if they chose to do so via a different service provider. Upon receipt of such a request we will recommend to parents that, they discuss with an appropriate member of their child's clinical team the issues that may be potentially raised in seeking this information. For example in regards to a non-clinically-based service provider offering results pertaining to adult onset conditions in the sample of a child.

Where a child participant in the Project accrues the capacity to consent at age 16 (or before), as part of the standard adult participant consenting process, their Genomic Medicine Centre will be able to ask the young person if they want to receive their looked-for additional findings. This will only be done if they give their consent. This search will include conditions with adult onset as it would with any other adult participant consenting to join the Project and opting-in to receiving this additional information. Any additional findings results meeting the outlined criteria will be returned to the treating clinician.

If a young person does **not** accrue the capacity to consent to participation in the Project on their own behalf by the age of 16, (by which time they are considered to be an adult for the purposes of the Project or where an adult lacks capacity to consent on their own behalf at the time of their potential recruitment), the 2013 'Mental Capacity Act Code of Practice', issued by the Office of the Public Guardian, or the equivalent guidance in effect at the time, wll be followed in England.*

As part of this process we will discuss the standard patient information with their consultee, including the question of whether or not to generate and return results on looked-for additional findings. If the consultee advises that the person would have wanted us to look for and return their additional finding results to their clinician for subsequent action, we will look for these as usual, including in regards to adult onset conditions. If the consultee advises against this, we will not look for or return any additional finding results to the clinician.

If there is an incidental finding in regards to an adult onset condition, this would not be returned to the participant's medical team unless the GeCIP mechanism outlined above determined that in the particular case this should be done.

Clinical access to patient data and feedback

Clinicians from the GMCs (including the NIGMC) who provide direct patient care will have freedom of access to their patients' data in identifiable format and will be able to interrogate both the variant call files and the raw data. As per professional standards, clinicians who do so will be expected to work within their competence and professional ethics in relation to accessing their patients' information.

Adequately resourcing this policy's practical feedback pathway

NHS England, working closely with Genomics England, continues to develop plans with its partner organisations in these and other areas to manage resource requirements. This has allowed the NHS to begin recruitment to the main phase of the Project from February 2015 in England. This includes putting in place measures to confirm results at scale with quality assured tests in a service laboratory, providing extra clinical consultations to explain findings (and in some cases referral for genetic counselling) and training medical staff to deliver these consultations. As part of the Project, Health Education England is actively assessing the training and educational

*Please note that the Mental Capacity Act 2005 and Mental Capacity Act Code of Practice do not apply in Northern Ireland. Mental capacity is currently governed by common law in Northern Ireland and as such this will be applied in terms of the consent process for Northern Ireland participants.

needs around all aspects of transforming the health professional workforce to be able to deliver genomic medicine. Where a patient requires counselling in connection to their participation in this Project (whether this is genetic counselling or another kind of counselling), their clinical care team will refer them appropriately, in line with established practice in clinical care.

The 100,000 Genomes Project feedback timeframe

The return of information to treating clinicians by the Project, in a time frame relevant to treatment decision-making, is a long-term aim in regards to **all** patient groups entering the Project. However, the timeframe for treatment decision-making with regards to the **cancer** patient group is likely to be shorter than with some other groups. In 2014, the expected turnaround time from sample submission to return of results is at least six months. By the end of the Project in 2017, we hope to have progressed to a 15-day turnaround of results.

At the time of consent, all potential participants (and especially those from the cancer patient group) will need to be made aware that, where pertinent findings are found, in the short term, samples and data used in research will not usually be able to be processed and analysed in a way that will reliably inform medical decisions regarding their care, within the time frame that would be required. At all stages of the Project, those discussing and taking consent with participants should have access to up-to-date information regarding the timeframe for return of results.

Specific feedback arrangements for the pathogen sequencing infectious diseases programme

For host sequencing regarding severe response to infection, there are no specific feedback arrangements beyond those outlined for rare diseases.

NHS clinicians' access to their patients data

NHS clinicians with patients in the dataset may require access to data specifically relating to their patients due to an identified clinical need. We recognise that some clinicians will want to look at the latest variant call files and even the raw read data and will grant such access via their NHS Trust embassy.

Genomics england

7 Programme delivery and management

7 Programme delivery and management

7.1 Programme management organisation

The Project will be managed and activity monitored in accordance with the highest standards for an NHS transformation programme management and governance practice in partnership with NHS England. Equivalent arrangements exist in Northern Ireland for management and monitoring of the Project. The constituent projects and workstreams will also follow the same approach, ensuring activity and resources are coordinated and well managed with reporting delivered in a consistent way. This methodology, adopting Programme Management Office (PMO) best practice, supports the early recognition and mitigation of risks and issues. It also ensures associated dependencies can be managed and adjusted in a timely and effective way to ensure a successful outcomes for the programme.

Software and infrastructure development activity will be integrated with the overall programme, although specific methodologies may be adopted where software development and configuration activity is needed to deliver the required informatics infrastructure and functionality.

A Programme Board has been established to oversee the programme activity, monitor progress and act as an initial point of escalation to resolve emerging issues. In addition, the Board will be responsible for overseeing the capture of lessons learned and ensuring, where appropriate, they are fed back into the programme.

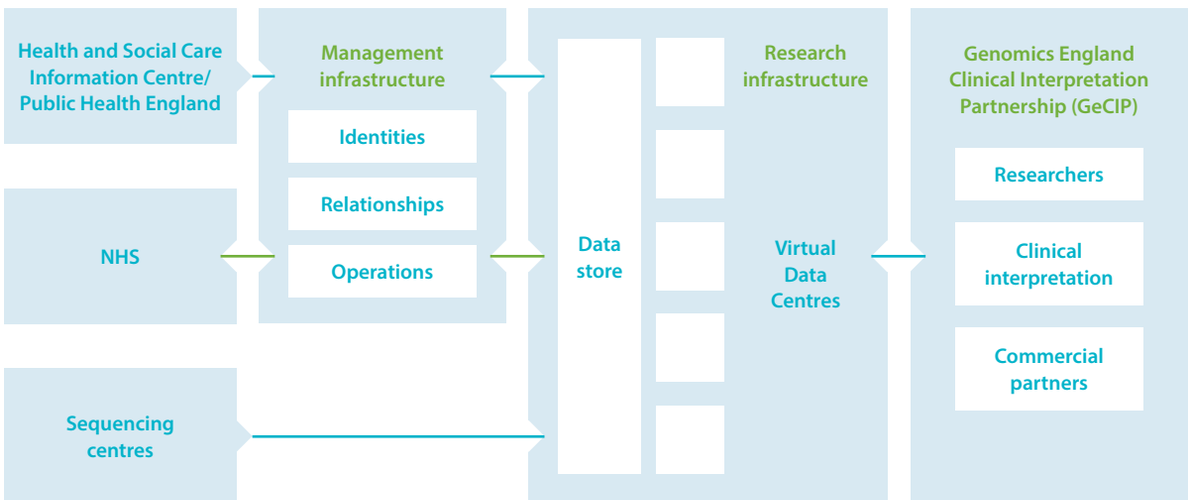
Genomics england

8 Informatics architecture

8 Informatics architecture

The diagram below shows an outline of the informatics architecture for Genomics England. Clinical, laboratory and the health data flow from NHS and other organisations into the Genomics England data management section where it will be de-identified or pseudonymised prior to storage within the research data infrastructure.

Figure 7



Each participant's sample and clinical data will be allocated a unique identifier before processing and inclusion in the Data Centre. The association between this identifier and the patient identifiers will be passed to the management infrastructure, in order that the genomic data resulting from the sample can be correctly re-associated. The sample identifiers will be present within GeCIP, but will not be included in the data made available to researchers in the GeCIP research domain.

Data will be made available within managed, virtual data centres, upon a high-performance computing platform. Each domain will have its own virtual data centre or centres. A different data centre might be created for each research project, programme, or product. The data centres will be able to communicate, not only through access to shared storage, but also using a services architecture based upon open standards. This will facilitate research collaboration and enable organisations to offer services to each other.

8.1 Information governance and security

Patient confidentiality and protection is a key cornerstone of the Project. We wish to promote an active Clinical Interpretation Partnership and encourage many users into the data infrastructure.

A key feature of the Genomics England programme is that individual level data will not be 'released', but will instead be analysed within a secure, monitored environment akin to a reading library (that is where books cannot be taken away but must be read *in situ* in a monitored environment). We will provide summary data on the Genomics England website so that potential users can see the type of data available.

To protect patient confidentiality, access to this environment will be granted only for specific, approved purposes in accordance with informed consent. Any attempted usage beyond the specified purpose may lead to exclusion and possible legal action, supported by automatically-generated electronic evidence.

Genomics England will require:

- An appropriate legal gateway for analysis (for Genomics England, the legal gateway is valid consent based on an ethically approved protocol) from each participant.
- Simple signed data sharing agreements will be expected between all parties accessing and using data as part of the Genomics England pipeline. Genomics England will have a simple electronic agreement system which all users will sign who wish to access the data. This will include electronic verification with their host institution, as used in UK Biobank.

- This agreement will include a description of our rigorous data access, sharing and acceptable uses procedure and approval to ensure compliance with governance requirements across the pipeline and safeguard data confidentiality and patient privacy.

The list below outlines the activities, checks and balances required to ensure Genomics England is processing patient data in an appropriate fashion.

- Delivery of the service model, whereby customers cannot export raw data (only summary results) and data delivered for research use is de-identified.
- Delivery of processes for the achievement of valid consent and accompanying literature for all Genomics England data subjects.
- Including seeking REC approval for the protocol and Patient Information Sheets and consent.
- Delivery and active management of signed data sharing agreements between Genomics England and all its data suppliers, and Genomics England and its users.
- Delivery of NHS Information Governance Toolkit assessment (annual), attainment of Level 2 and continuous information governance improvement plan.
- Delivery of privacy impact assessment and equality impact assessment, and accompanying action plans.
- Delivery of information governance continuous improvement plan, and training and awareness for Genomics England staff.
- Delivery of staff contracts with explicit clauses relating to employee responsibilities to safeguard data integrity and security.
- Where appropriate policies and Standard Operating Procedures (SOPs), based on guidance from the Information Commissioner's Office (ICO) that set out the ways in data may be accessed, handled and disclosed within the secure Data Centre.
- Data access request log and process, and data transfer logs relating to all identifiable data – reviewed regularly to ensure adherence to SOPs.
- Delivery of clear data models for use and plans for effective data management.
- Within the Genomics England Data Centre, information will be held in a de-identified fashion, with the identifier linker file held in a logically separate location.
- Tools for enhanced audit logging and processes to verify customer use of the data and take action if non-compliance is detected.
- Specific techniques utilised for delivery of small datasets, such as obfuscation, to minimise the risk of re-identification.
- Standard OGC ITIL compliant processes for Service Management, including management and control of live service incidents – with 24/7 cover for security incidents or near misses and changes to the infrastructure which holds and processes the data.

A policy and framework will be developed, based on the existing ISO 27001-compliant policies provided by Oxford's NDPH, and advice received from the Genomics England Access Review Committee. For the Terms of Reference and membership of the Access Review Committee see **www.genomicsengland.co.uk**. The scope of the policy includes data access; data integrity and security controls (including encryption of both data and mobile devices, backup, firewalls, anti-virus protection), audit and verification procedures; and formal change management of all security documentation. The scope of the policy spans people, processes, tools and data.

All data transfers and data access requests will be logged and subject to regular scrutiny by the Service Management Team. Escalation procedures will be in place to manage non-compliance if detected, based on the data sharing agreements signed by users (or internal staff/supplier contracts in the case of an internal breach).

Virtual data centres will be provided within the infrastructure, to provide clinical researchers with managed, audited access to data and computing resources. Activity logs will be analysed on a regular basis, with any exceptions flagged to the Service Management Team for further investigation and escalation as required. Additionally, the Genomics England Programme Board will ensure that the Chief Technology Officer and Chief Operating Officer establish adequate procedures are in place for documentation, tracking, monitoring and reporting for overall process of data sharing.

As part of its assurance and approvals plan, Genomics England will seek ethics approval for the research database itself, complete the NHS Information Governance Toolkit on an annual basis and maintain a continuous improvement plan for Information Governance. In addition to the full suite of functional and non-functional tests, penetration (security) testing will be completed. The Genomics England Data Centre will be physically located at more than one site in England.

8.2 Data access, sharing and acceptable uses of information from participants in the 100,000 Genomes Project

The Project's consent discussion and patient literature outlines the plans around data access, sharing and acceptable uses in addition to security, third party and participant access requests to data. This will be accessed in an anonymised format by clinicians, academics and industry in the UK and overseas. The policies relevant to this can be accessed via a Genomics England secure portal.

A full data service will be provided for external users. Users will remotely access the data they need on Genomics England servers via discrete virtual machines – where activity is continually monitored and logged within the secure data centre. Raw data cannot be exported – only summary results of their analysis.

All research users must sign the electronic data access agreement and be members of a healthcare or research organisation. This enshrines all legal and regulatory requirements relevant to the processing of patient information.

Genomics England takes data security and patient privacy extremely seriously. It will take action where a breach, or near breach, is proven. Penalties for non-compliance include Genomics England's revocation of user access, organisation access and reporting the offending activity to the Information Commissioner. We will where possible, notify individual participants affected if we discover data breach notifications, enabling them to exercise their 'right to object' to further processing of their personal data.

All policies, procedures and literature relating to data security and privacy shall be formally placed under change control. Formal policies relating to purposes and uses will be subject to discussion and consultation with stakeholders.

All users must act, at all times, in accordance with the terms of Genomics England's informed consent policy, procedures and literature.

Genomics England welcomes applications for use of its datasets from the commercial, academic and clinical communities, and is keen to create an environment, through the Genomics England Clinical Interpretation Partnership (GeCIP), to enable different kinds of users to work together to share and build new knowledge about the data.

The Project has commissioned participant, patient group and health professional surveys or interviews around issues of data sharing and access. In addition, in relation to these specific issues, stakeholder meetings with a series of groups concerned with issues of privacy and consent have taken place during summer 2014. The views fed back in respect of collection handling and sharing of data in the Project have been incorporated.

Partnership with NHS organisations

The Genomics England team will work closely with colleagues in NHS England (NHSE), Health Education England (HEE) and Public Health England (PHE), DoH NI, PHA and BHSCT in order to deliver the pipeline in an integrated and sustainable manner.

8.3 Partnerships to ensure patient recruitment, sample and data acquisition

NHS England has committed to assume responsibility for funding and delivering certain aspects of the genomics pipeline for the main programme in England. In Northern Ireland, these commitments have been made by DoH NI and MRC with regard to funding and BHSCT in terms of delivery. These aspects include:

- Patient approach and consent process.
- Biological sample acquisition, logistics and tracking.
- Some DNA extraction in accredited centres, where quality assurance parameters have been met.
- Systematic delivery of clinical, phenotypic, laboratory, demographic and administrative data from NHS sources.
- Delivery of identity management function to verify clinician and patient identity and existence of legitimate relationships (necessary for enabling access to identifiable information).
- Validation of WGS identified variants and sharing of data from subsequent investigations.
- Systematic delivery of clinical feedback based on WGS data held and insights gained, and realisation of benefits relating to improvements in patient care.

Depending on progress with this NHS Transformation programme by the end of 2017, the Genomics England Board will make arrangements with the Department of Health for ongoing provision of the full pipeline as appropriate. Based on recommendations from the Genomics England Science Advisory Committee and Board and the NHS, this document will be revised to incorporate any prescribed change to scope.

Separate arrangements, which will be confirmed at a later date, will be put in place for Northern Ireland.

Partnership with the NHS

The collaborations and partnerships with the NHS will deliver high-quality phenotyping data linked to health records data, DNA and samples which contribute to the whole functional genomics pathway. Patients will be initially invited to participate in the Project after identification and selection by eligibility via the NHS clinical team. Patients will then be consented to participate and provide biological samples of either blood, tumour type, or both. All tumour and blood samples will either be processed at the site and then transferred as rapidly as possible to the Genomics England Central Biorepository, or sent immediately to Central Biorepository. It is intended that blood samples will be acquired through routine or a scheduled appointment. Tumour samples will be obtained from tissue taken at time of biopsy or resection. It is anticipated that resected tumours will provide adequate tissue for DNA extraction. All samples will be handled in accordance with the Sample Handling Guidance , using the Institution's Approved Standard Operating Procedures. In some cancers optimal whole genome sequencing may require additional biopsies over and above normal diagnostic practice, to acquire sufficient tissue for DNA extraction.

In addition to the blood and tumour samples taken for whole genome sequencing, an additional sample of blood (see **appendices** for details) will be taken to be stored for future 'omics' research. The specifications and standard operating procedures for the collection, processing, transportation and storage of both the whole genome sequencing sample and the 'omics' samples, has been agreed based on analysis of pilot outcomes but may be updated as further data is available from the programme. However, there are expectations for capability, as well as compliance with national quality standards and regulatory procedures.

For both the NHS Genomic Medicine Centres (GMCs) (including the NI GMC) and the Central Biorepository, there will need to be the capability, capacity or access, for high-speed centrifuge facilities for spin at 16,000 rpm (specifically for cancer plasma samples). This is also required to store samples in -20°C, -80°C, Liquid Nitrogen (LN2) or room temperature storage facilities.

8.4 NHS England requirements and specifications

There are a number of activities required to enable NHS England to deliver aspects of the pipeline described above. These are described or specified in further detail below. Delivery of the project in Northern Ireland is managed through a separate Memorandum of Understanding, Services Agreement and other supplementary agreements.

8.4.1 NHS Genomic Medicine Centres (GMCs)

As of 2016, NHS England have establised 13 GMCs across England following a rigorous tender and accreditation process. This creates the capacity and capability framework that recognises the valuable contributions made by the NHS and patients, and assures maximal performance of sites in order to optimise the whole pipeline. This is entirely separate to current proposals to realigning Genomics Laboratory Services. Following an open process, based on NHS Trusts which may act as hubs on behalf of other NHS Trusts in England, 11 centres were announced on the 22nd December 2014. They were selected based upon the following strict criteria (to be confirmed):

- Evidence of capability and quality to support the 100,000 Genomes Project in terms of eligible patients/families in the specific disease areas of focus, the sample numbers and systems to deliver significant sample flows per annum focused on rare disease, cancer and severe response to infection.
- Clear commitment to deliver specific patient recruitment numbers, to Genomics England timelines and quality requirements.
- Evidence of commitment and capability to obtain the NHS Genomic Medicine Consent for Genomics England locally.
- Evidence of local capability or partnership with centres that can extract high quality germline DNA according to Genomics England Standard Operating Procedures at the right concentration and quality, and provide samples to Genomics England according to specified quality metrics. Tumour DNA will be extracted in a very limited number of centres. All DNA extraction will be quality controlled and assured.
- In cancer, commitment to use Genomics England Standard Operating Procedures for surgical and molecular pathology provision of cancer samples.
- Commitment to record and upload patient characteristics and all data according to disease-specific criteria simultaneously with the sample acquisition.
- Commitment to transfer samples with identifiers to a centralised DNA repository, where a second quality control step will be run.
- Clear commitment to re-supply samples where samples fail quality control.
- Clear commitment to validation of the results within the NHS, and feedback to patients.

The Genomic Medicine Centres will contribute to the programmes in rare diseases, cancer or both. We will release a monthly dashboard of performance of all selected centres. If performance falls short of expectation, the centre designation may be suspended or withdrawn by NHS England. All NHS Genomic Medicine Centres will be members of the Genomics England Clinical Interpretation Partnership (GeCIP) and will have access, to the research data infrastructure at no charge.

The accreditation process involved a submission of a tender including:

- Support from management, clinical and informatics leadership in each disease area and relevant pathology, systems/data and clinical genetics personnel.
- Subject recruitment plans with diseases, sample numbers and timings.
- Informatics plans with details of any necessary technical or organisational change.
- Resource plans for any activities in addition to routine care.
- Local delivery structure including project management and monitoring resources.

As part of mobilisation of sites, Genomics England and NHS England staff as appropriate will visit sites to review processes. This will pay particular attention to the operational commitment at all levels of the organisation, ability to follow required consent and sample handling protocols and capacity to record and export required clinical data. It is essential that these visits are thorough and comprehensive, as over-optimism will result in Genomics England and the clinical sites wasting significant resources.

The Northern Ireland Genomic Medicine Centre (NIGMC)

Through an investment by the DoH NI and the MRC, and in partnership with Genomics England, a NIGMC has been established to facilitate recruitment of participants from Northern Ireland to the Project from early in 2017. This virtual centre is hosted by BHSCT and facilitates recruitment of participants from across all five HSC Trusts in Northern Ireland. Delivery of the Project in Northern Ireland is governed by an overarching Memorandum of Understanding (made amongst DoH NI, BHSCT and Genomics England) and a detailed Services Agreement.

8.4.2 Sample and DNA acquisition and logistics

Sample acquisition and DNA extraction specifications for cancer and rare disease are depicted below.

Figure 8. Cancer sample acquisition processes

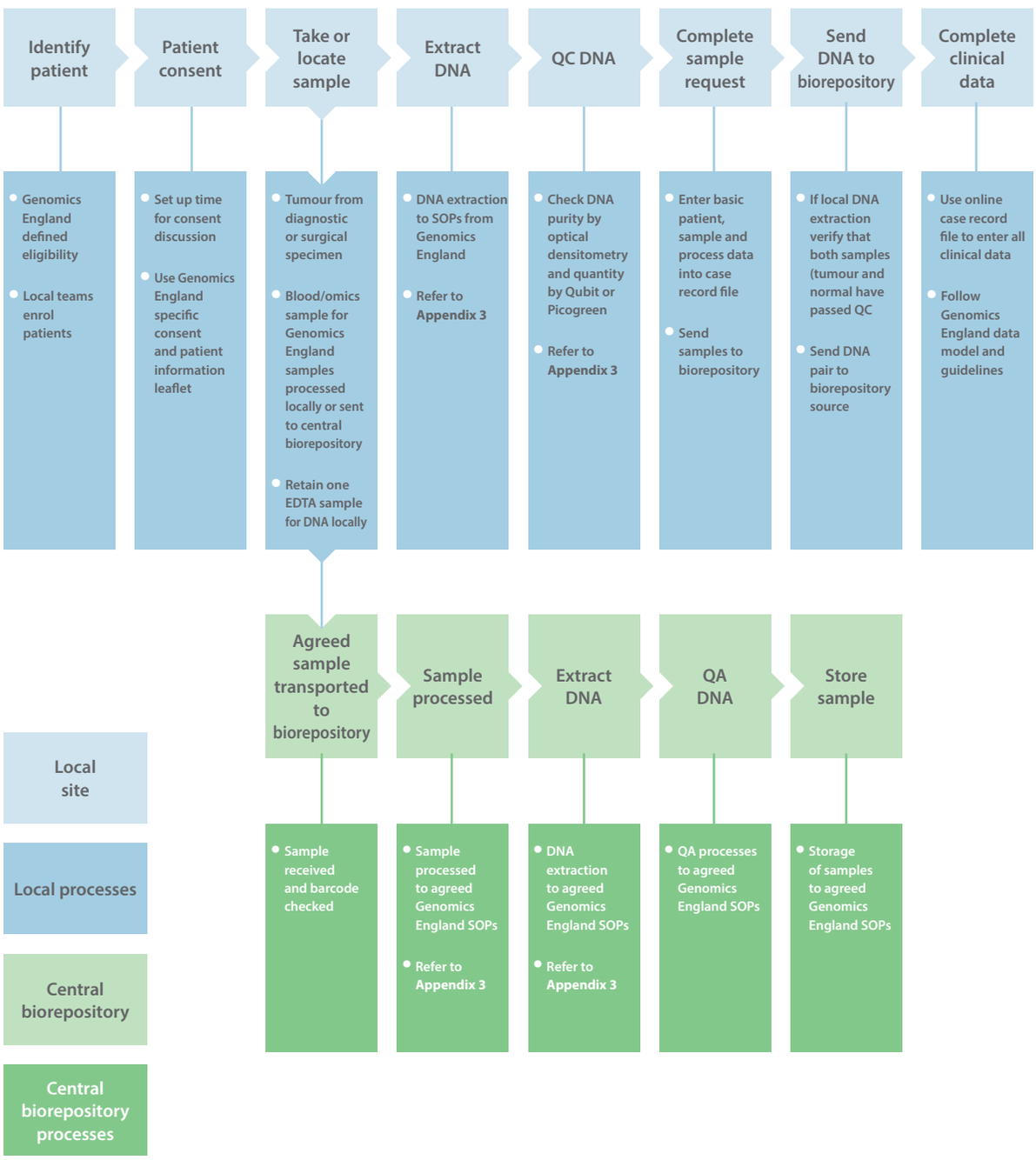


Figure 9. Rare disease sample acquisition process

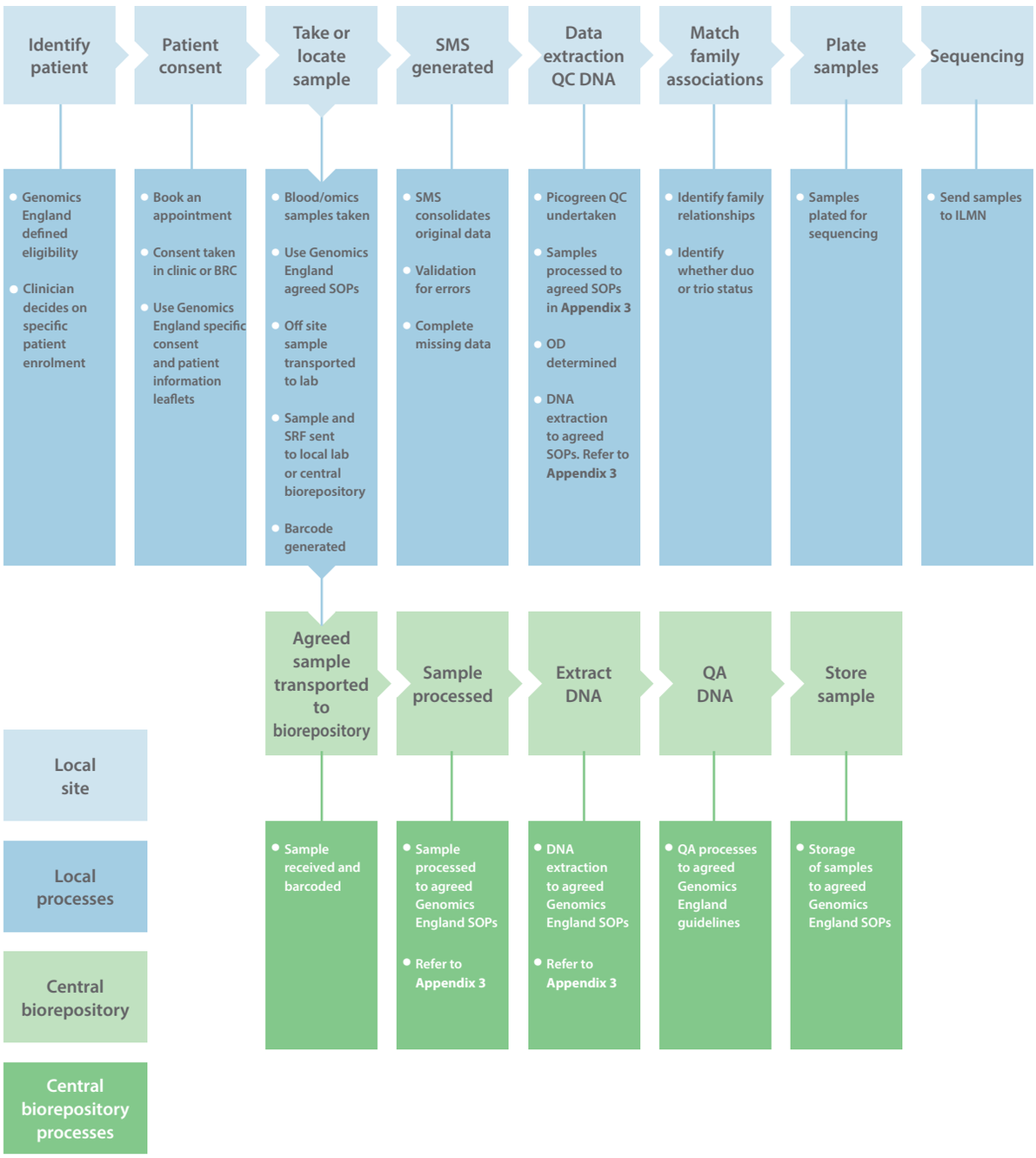


Table 4. DNA extraction requirements for whole genome sequencing

The following defines the output specification that is needed for the DNA that is extracted by the GMC for referral to the Biorepository. DNA should be extracted using a protocol that meets the following specification:

DNA	Specification	Amounts
Extraction	Automated extraction is preferable. Extraction must be carried out according to manufacturer's instructions and the laboratory SOP. Widely used and acceptable systems are: Autopure Gentra, Chemagen System and Qiagen Symphony.	3µg at minimum concentration of 30-100ng/µl.*
Amplification	DNA must **not** be PCR amplified (no Whole Genome Amplified samples).	
Quantification	Quantify determined using Qubit or PicoGreen methodology.	30-100ng/µl.
DNA purity		A260/A280 ratio 1.8-2.0.
DNA buffer	1X TE pH8.0 / as per manufacturers system.**	

DNA from saliva	Specification	Amounts
Extraction	Automated DNA extraction using recommended kit on the same platforms as above.	To be confirmed.
Quantification	Quantify determined using Qubit or PicoGreen methodology.	
Purity		A260/A280 ratio 1.8-2.0.
Amplification	DNA must **not** be PCR amplified (no Whole Genome Amplified samples).	

* We are working in a collaborative manner with Illumina to reduce the quality and quantity of DNA required for successful whole genome sequencing over the course of the project. Stated values are current guidance from Illumina as being required for reliable successful sequencing.

** Where proprietary buffers are included in the automated system these need to have been proven to be compatible with Illumina NGS technology.

Table 5. Specification for DNA extracted from FFPE tumour

The table below defines the output specification that is needed for the DNA that is extracted by the GMC for referral to the Biorepository. DNA should be extracted using a protocol that meets the following specification:

DNA	Specification	Amounts
DNA from FFPE	DNA should be extracted using an automated system that has been demonstrated to yield high quality and quantity of DNA systems - currently in use are Maxwell and Qiagen.	>750ng.*
Amplification	The DNA must **not** be PCR amplified (no Whole Genome Amplified samples).	
Quantification	Quantify using Qubit or PicoGreen.	At least 15ng/µl (max 100ng/µl).
DNA purity	Should be measured.	A260/A280 ratio should be measured.
DNA buffer	1X TE pH8.0 / or as per manufacturers system**	

* We are working in a collaborative manner with Illumina to reduce the quality and quantity of DNA required for successful whole genome sequencing over the course of the project. Stated values are current guidance from Illumina as being required for reliable successful sequencing.

** Where proprietary buffers are included in the automated system these need to have been proven to be compatible with Illumina NGS technology

DNA extracted from blood and tumour must be sent together to the Biorepository, following confirmation that clinical data is available that meets Genomics England requirements as laid out in Genomics England SOPs.

For the main programme, it is expected that Genomics England and NHS England* will work in partnership to identify and implement changes to the molecular pathology pathways within accredited centres to enhance sample quality, in order to optimise the Genomics England pipeline. Such changes will be based on learning derived from phase 2 of the programme and relate specifically to tumour DNA samples from FFPE or fresh frozen preparations. As a result, it should be noted that some of the requirements may change over the life of the programme.

All DNA samples should be accompanied by an electronic and printed sample request form provided by Genomics England, which provides data on the patient and sample itself. It is to be completed at the point the sample is taken and prepared. Genomics England will work with the GMCs to finalise the following dataset and collection tool:

- DNA source – germline/tumour, primary/metastasis.
- Standard Operating Procedure version(s) used for collection, extraction, QC and logistics.
- Tumour type (if cancer).
- Tumour cellularity percentage (if cancer).
- DNA concentration (nanogrammes/microliter).
- DNA volume (microliters).
- DNA QC metrics (to be finalised with sequencer).
- Genomics England identifier for data subject (site ID, local patient ID (NHS Number), sample ID).
- Gender and ethnicity.
- Date of birth.
- Pedigree structure and affectation status (rare disease).
- Version of Consent and Patient Information Sheet used.
- Syndrome or disease name.
- Named site contact and email (to verify receipt of samples by collection hubs, for queries etc.)

8.4.3 Labelling of samples

Each biological sample and data capture form should be labelled with the Genomics England identifier.

8.4.4 Key performance indicators (KPIs)

Key performance indicators will be recorded to monitor the process and also to enable the identification of variables that affect sequence quality. These will need to be agreed.

8.4.5 DNA delivery to sequencer (logistics)

Paired tumour and blood DNA samples extracted in accredited Genomic Medicine Centres or in the Genomics England Central Biorepository. All samples and DNA should be provided to the Genomics England Central Biorepository as soon as possible using a specialist courier. The Biorepository and sequencer will be automatically notified of sample departure via the electronic sample request form.

Genomics England will provide a template and reporting timelines for this work.

Genomics England has committed to a sample schedule with the sequencing companies, and so sample logistics should follow the agreed schedule in order to optimise utilisation of the reserved capacity in the sequencing centre.

*DoH NI and BHSCT in Northern Ireland

8.4.6 Data from sample acquisition site

Locally pseudonymised **sample metadata** will be required to be submitted alongside the DNA samples, to enable sequencing and annotation. This will be within 14 days of a patient or family member providing samples in rare disease and cancer.

Clinical, phenotypic, laboratory, demographic and administrative data will need to be captured locally, alongside or close after sample acquisition takes place. A web-based information system will be provided by Genomics England to support direct data acquisition. Data must be collected based on the agreed data models, which will be refined over time. As a general rule, DNA samples will not be sequenced unless these data have been received into the Genomics England Data Centre.

8.4.7 Data from national NHS and other sources

Genomics England expects to pull in other health data, on its consented patients, into its data centre, to link outcomes data to the WGS and clinical data already held. This may involve ongoing data delivery from other organisations, including but not restricted to the HSCIC, the CPRD, Public Health England and patient communities.

All organisations operating within the pipeline need to provide monitoring and reporting functionality to facilitate the **tracking of samples and data** end to end.

Any proposals for changes to data requirements should be directed to the Genomics England Chief Technology Officer.

9 Genomics England data access and acceptable uses

9 Genomics England data access and acceptable uses

Policies for data access, sharing and acceptable uses have been developed in line with the Medical Research Council (MRC) Policy and Guidance on sharing research data from population and patient studies. These will be based on similar policies within Oxford NDPH (see **http://www.ctsu.ox.uk/research/** for further details), the Wellcome Trust Centre for Human Genetics, and input from the Information Commissioner's Office and the Genomics England Ethics Advisory Committee and patient groups. The Data Access policy is available upon request or via a Genomics England secure portal.

9.1 Users of the 100,000 Genomes Project data

The data access and acceptable uses criteria have been summarised in this section.

Access to data will be controlled according to the provisions of the Data Access policy and all users will sign a Genomics England data access agreement. All Genomics England policies are available via a Genomics England secure portal.

9.1.1 Patient access

Patients will have the right to request that their whole genome sequencing (WGS) data be made available to them. Genomics England will develop a process for delivery of variant call files, on request, following security and validation checks. As part of this work, Genomics England will need to consider the requirements of the Freedom of Information and Data Protection Acts, as well as internal data security policies and ethical rules around disclosure, and the logistics and costs of the data transfer. The subject access request policy is available via a Genomics England secure portal.

We note that a child participant may make a subject access request under s.7 of the Data Protection Act 1998. Genomics England will ensure that the response provided to a child's subject access request is in line with the latest advice in this area from the Information Commissioner's Office, in order to ensure that the response provided is compliant with the Data Protection Act. We have noted the advice that is already available on this issue from the Information Commissioner's Office 'Guide to Data Protection'.

Neither Genomics England nor the NHS accepts any responsibility for any patient-initiated analysis or subsequent interpretation of these data by third parties.

9.1.2 NHS healthcare teams

Findings will be fed back securely to the identified clinician and/or team responsible for caring for the patients. These findings will contain patient identifiable data (see above).

9.1.3 Academic researchers and trainees

They will either be a Genomics England Clinical Interpretation Partnership (GeCIP) user or if not, they will pay to access the data.

9.1.4 Commercial researchers

Models for commercial data access are under development.

9.2 Data access and sharing for the 100,000 Genomes Project

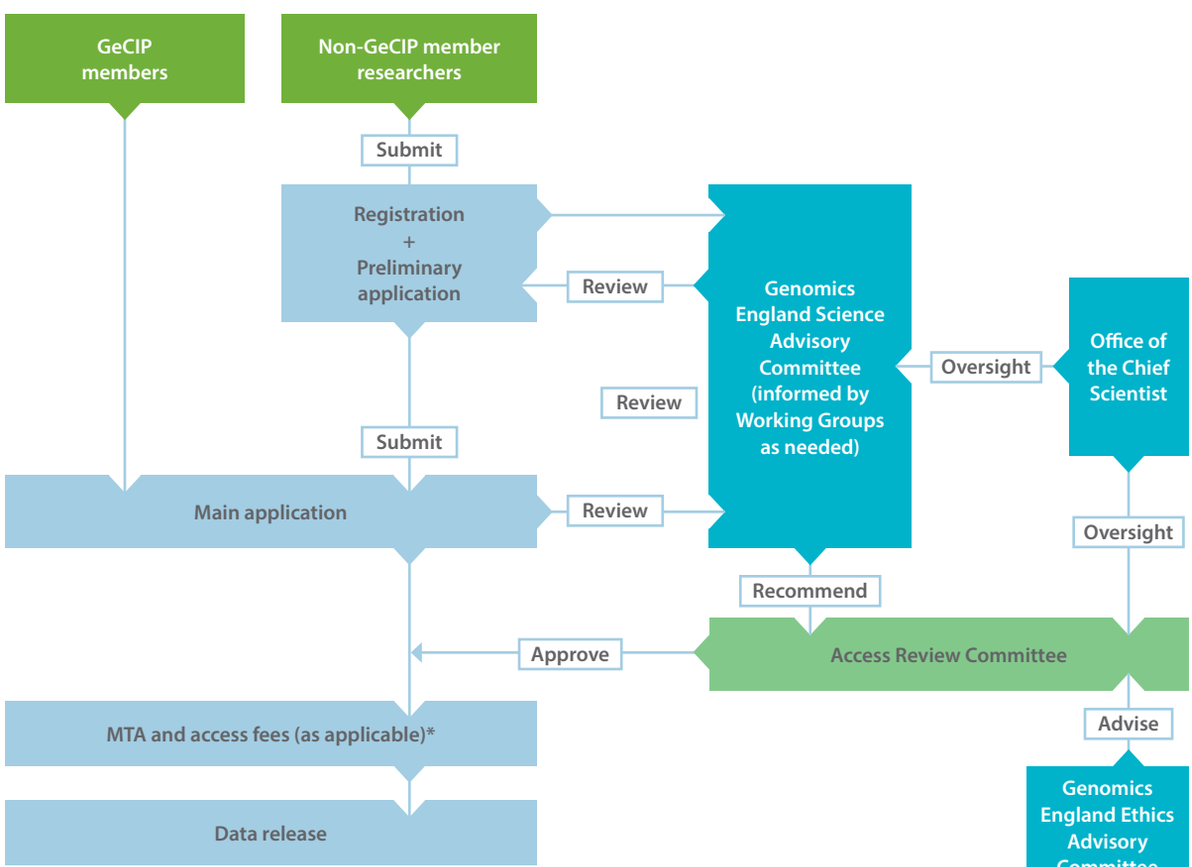
The purpose of the infrastructure is to provide maximal access to researchers, the NHS and those in training, enabling managed access to the data assets created by the Project. The infrastructure will be designed to facilitate the sharing and re-use of data and compute resources, and to promote scientific collaboration in clinical genomics research. Software infrastructure – models, interfaces, and tools – will be provided for the generation of 'analysis ready' datasets.

All requests for data access will be subject to the following considerations:

- Protection of data subjects (honouring commitments made to them, acting within the scope of consent and according to conditions of Research Ethics Committee approval).
- Compliance with legal and regulatory requirements (Data Protection Act 1998, Freedom of Information Act 2000, NHS Act 2006, Health and Social Care Act 2012, the Common Law Duty of Confidentiality, Human Tissue Act 2004 and applicable requirements from organisations affiliated with the Health Research Authority, including Research Ethics Committees and the Confidentiality Advisory Group (CAG)).
- Provision of a signed Genomics England data access agreement to the Access Review Committee (see Information Governance above).
- Prioritisation of access according to resource availability.
- Facilitation of high quality research.

Access to the data and computer infrastructure will be subject to approval by the Genomics England Access Review Committee. There will be a process adapted from that used successfully by UK Biobank, to allow bona fide research access and use within the terms of the Genomics England data access and acceptable uses policy.

Figure 10. Data access application and review process



*Charges are only applicable to non-GeCIP members or industry and reflect the cost of data provision and associated services.

9.2.1 Data sharing requirement

For the purposes of this document, data held shall be defined as either:

i) The data held within the Genomics England Data Centre about participants within the Project (**service data**), or

ii) The data held by Genomics England relating to its business, its customers and users, its staff etc. (**corporate data**).

It is the core business of Genomics England to share **service data**, within an agreed governance framework, in order to deliver access to treating clinicians which may be identifiable for NHS use (see above). It is also part of our core mission to provide access to researchers and scientists. Such data will be anonymised, unless there is a legal basis in place for the disclosure of identifiable information.

Furthermore, Genomics England is bound, through various legislative and regulatory instruments, to deliver data or access to specific requestors for specific purposes e.g. to share data requested under the Data Protection Act (Subject Access Request), or to deliver the results of financial audits to regulators. Further information can be found in the Genomics England Data Access Policy. For these sorts of requirements, both **corporate data** and **service data** might be shared, anonymised or aggregated unless there is a legal basis in place for the disclosure of identifiable information.

As previously noted, sharing of data for prescribed purposes is key to the Genomics England business model. However, decisions on data sharing must always be assessed against the risk to data confidentiality and the privacy of data subjects. All decisions relating to data sharing should be made considering the policy provisions enshrined here.

Where risks to data sharing fall outside of the tolerances set out on the agreed SOPs, Genomics England shall work with its legal counsel and Information Governance and data security leads, to deliver a transparent decision-making process and associated assurance, validation and reporting.

9.2.2 Service data sharing

All data held within the Genomics England Data Centre that relates to the Project data subjects is considered to be service data.

Service data will be shared, within an agreed governance framework, in two ways:

i) With internal staff (under substantive contract of employment) within Genomics England, in order to check data quality and integrity; to undertake agreed cleansing activities; to deliver initial bioinformatics analytics on the data; and to undertake other monitoring, audit and service improvement activity. It is noted that internal access to all data held shall be controlled, monitored and reported upon. Access to identifiable data, which is held separately, must be explicitly logged and approved by the nominated Genomics England Information Governance lead.

ii) With external users and consumers, as follows:

a. In the form of proactive publication of anonymised, aggregated information as summary metadata about the work of Genomics England, according to an ICO approved schema, that meets Freedom of Information requirements on transparency and has a very low risk of re-identification. Further information can be found in the Genomics England Freedom of Information policy.

b. Delivery of access (within the Genomics England Data Centre) to agreed datasets for approved users, including scientists undertaking research into whole genome data. Data will be anonymised, based on the ICO code of practice, unless a legal gateway exists for access to identifiable information.

c. Delivery of limited access to agreed datasets for those requesting information under the Freedom of Information Act (FOI request), the Data Protection Act (SAR request), or under any other regulatory or legal basis for disclosure. Data will be anonymised, based on the ICO code of practice, unless a legal gateway exists for access to identifiable information. Further information can be found in the Genomics England Data Access policy and is available via a Genomics England secure portal.

9.2.3 Corporate data sharing

Corporate data is defined as data that is held by Genomics England that relates to the work and operations of the business, including its organisation, its people, its performance and its decision-making. Certain corporate data may be shared, within an agreed governance framework, in order to meet statutory and regulatory requirements. This includes the proactive publication and maintenance of information about the operations of the business, in line with the provisions of the Freedom of Information Act. All data shall be anonymised, based on the ICO Code of Practice, unless a legal gateway exists for access to identifiable information.

9.3 Acceptable uses for the 100,000 Genomes Project

All users and uses must be consistent with the terms of informed consent policy, procedures and literature pending REC approval.

Genomics England welcomes applications for use of its datasets from the commercial, academic and clinical communities, and is keen to create an environment, through the Genomics England Clinical Interpretation Partnership (GeCIP), to enable different kinds of users to work together to share and build new knowledge about the data that can be translated into new diagnostic and treatment options for patients.

9.3.1 Types of purposes

The following list of purposes will be used by Genomics England to classify the sort of work its users want to undertake. Under the terms of its ethics approval, Genomics England plans to operate an internal process to assess applications and grant access to data that relates to any of the purposes below.

Definition	Explanation
Clinical care	Use of data to directly diagnose / treat a patient.
Hypothesis driven research and development in health and social care - observational	Hypothesis driven review and analysis of data collected within the Genomics England data centre.
Clinical trials feasibility	Queries on the broad location and numbers of patients eligible to participate in interventional research project.
Hypothesis driven research and development in health and social care - interventional	Hypothesis driven review and analysis of agreed cohorts of patients within a research study.
Non hypothesis driven R&D - health	Review and analysis to identify associations between data and variables - to improve our understanding of the causes of disease.
Non hypothesis driven R&D - non health	Use of the data by data scientists to test the effectiveness of tools and models for data analysis.
Public health purposes	Work to gain a population level understanding of the prevalence of disease and corresponding health protection actions where possible.
Other health use - clinical audit	Use of the data to assess whether clinical standards are met.
Other health use - commissioning / commissioning policy	Use of Genomics England data to inform decision--making by any relevant commissioning body / organisation developing commissioning policy. Trust use to undertake local planning and administrative tasks.
Subject access request	Request by the data subject regarding the data Genomics England holds about them.
Tool evaluation and improvement	Use of the data to validate, improve and deliver new annotation tools for WGS data.
Interpretation and validation of the Genomics England Knowledge Base	Use of the data to interpret annotation findings and validate results for use in a clinical context.
Education and training of health and public health professionals	NHS professionals and others learning how to analyse, interpret, and apply genomic medicine.
Deeper phenotyping	Gaining further information on patients with particular genetic results.

9.3.2 Uses that the 100,000 Genomes Project will decline

These types of purposes might be proposed by participants or third parties. They will typically be refused outright following consideration. This will be by the Genomics England Informatics Team and the Genomics England Access Review Committee who will decide upon approval or rejection on a case-by-case basis. Additionally, all requests that are required by court order will be notified to the Genomics England legal counsel and the Executive team.

The Genomics England decision-making process on all occasions shall be documented and transparent, and involve external bodies (e.g. REC) where necessary.

- Paternity testing and level of relatedness testing (though participants can request their data and, as such, if several family members make this request and take the data to a third party to analyse, it is conceivable that the data can be used in this way).
- Parental request for child's data once the child reaches 16 where either the participant assents or dissents.
- Requests in the form of Court Orders will be referred to Genomics England's Legal Counsel as promptly as possible, so that all representations may be made to the court, for example to limit the information requested.
- Uses requiring linkage to employment records, tax records, benefits records or personal life insurance records for non-scientific or non-healthcare related purposes.
- Uses requiring linkage for personal insurance or forensic purposes.
- Purposes that might lead to discrimination against a person or a group of people, based on genetic/genomic characteristics.

Genomics england

10 Genomics England data ownership and intellectual property

10 Genomics England data ownership and intellectual property

Genomics England has received legal advice regarding intellectual property. This recommends that Genomics England owns the combination of the whole genome sequence and the clinical data for the entire dataset from the 100,000 Genomes Project. In addition, Genomics England owns any new intellectual property generated from the data, but we will license this to third parties on favourable terms.

There are very clear reasons why this is essential:

- It ensures that Genomics England Clinical Interpretation Partnership (GeCIP) investigators can collaborate in academic/NHS partnerships and academic industry partnerships without concern for the intellectual property being generated.
- The ready licensing with the capability to include all inventors offers a fair approach to potential intellectual property.

10.1 Genomics England publication policy

Genomics England encourages publication. All publications will be on behalf of Genomics England as a banner heading. The usual rules of authorship will apply and all co-authors will be named. The position of authors on all papers will be based on work done with the application of starred authorship to recognise that there may be multiple authors. The Genomics England team will also be co-authors and will typically share key authorship and corresponding author positions reflecting work done. The decision on authorships will usually be made within GeCIP domains and the GeCIP Steering Committee will advise the GeCIP Board whose decision will be final in the event of disagreement. There may be academic users or industry users who are not GeCIP members and pay to access the data. The same approach will apply to authorship and publication with the GeCIP Steering Committee and Board having oversight.

Acknowledgements in all publications will recognise the contribution of the Department of Health, NIHR and any other GeCIP funder with the following form of words:

"*Genomics England is a wholly owned company of the Department of Health and this programme was made possible by the National Institute for Health Research, NHS England, Public Health England and Health Education England*."

Other GeCIP funders must also be acknowledged.

Genomics england

11 Consent

11 Consent

11.1 The process for consent*

The achievement of informed consent in the context of clinical genomics is recognised to present important practical challenges and require careful ethical consideration. This is also true for the design of the 100,000 Genomes Project, a 'hybrid' initiative combining clinical, research and translational medicine aims.[30,31] The Project will follow established legal and regulatory standards for seeking the informed consent of its participants, or those who could consent on their behalf. It will follow applicable legal and regulatory standards around the potential or actual involvement of adults who lack capacity, including seeking the advice of appropriate consultees regarding their participation, including in regards to participants who join the 100,000 Genomes Project and thereafter are deemed to have lost capacity.[32]

A number of international and national initiatives have recently begun to set out guidelines on good consent practice in genomics projects of this kind. Exemplars of good ethical practice of this kind of approach in the United Kingdom include the Deciphering Developmental Disorders project.[33,34]

In preparing the Genomics England feedback policy, we took full account of all relevant legal and ethical principles including contemporary research and the advice of patients. We have tested our Patient Information Sheet and Consent Materials with our internal Ethics Advisory Committee; a specially convened Health Research Authority Advisory Group of Research Ethics Committee and patients; and where appropriate, families affected by the disorders included in the Project. The approaches described herein and the consent materials have been developed through multiple versions with input from all of these people. This has also been informed by Professor Mike Parker's letter to the Chief Medical Officer from the 100,000 Genomes Project Ethics Advisory Working Group (now renamed the Genomics England Ethics Advisory Committee). This can be found on our website **www.genomicsengland.co.uk**.

*Please note that the Mental Capacity Act 2005 does not apply in Northern Ireland. Mental capacity is currently governed by common law in Northern Ireland and, as such, the process for consent set out in section 11(Consent) will be adapted and applied in line with the Northern Irish common law position for Northern Ireland participants.

The process for approaching and obtaining consent from patients to participate in the Project is shown in the flow diagram below:

Figure 11. Consent process

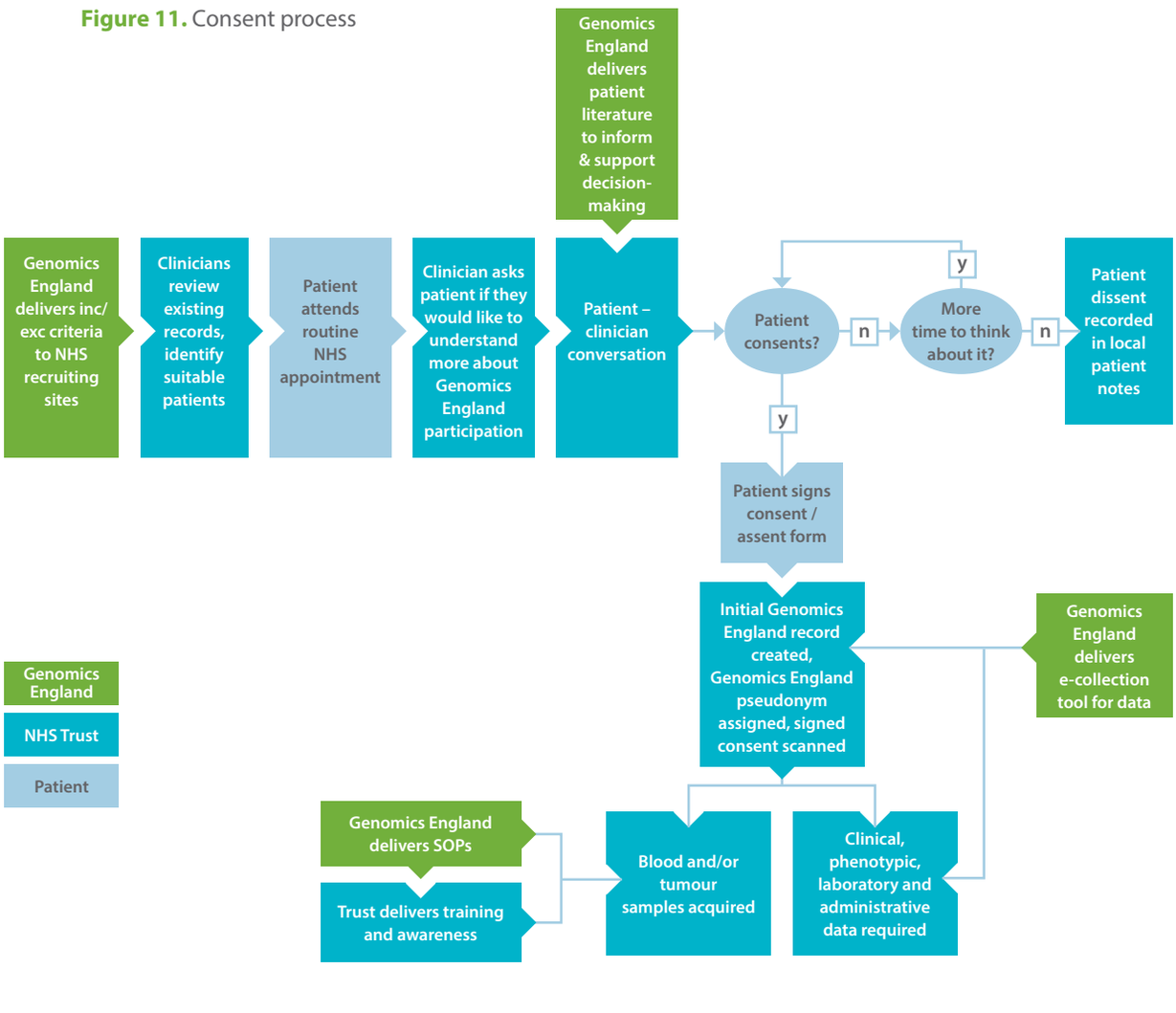


Figure 12. Inclusion process for rare diseases - questions for enrolment

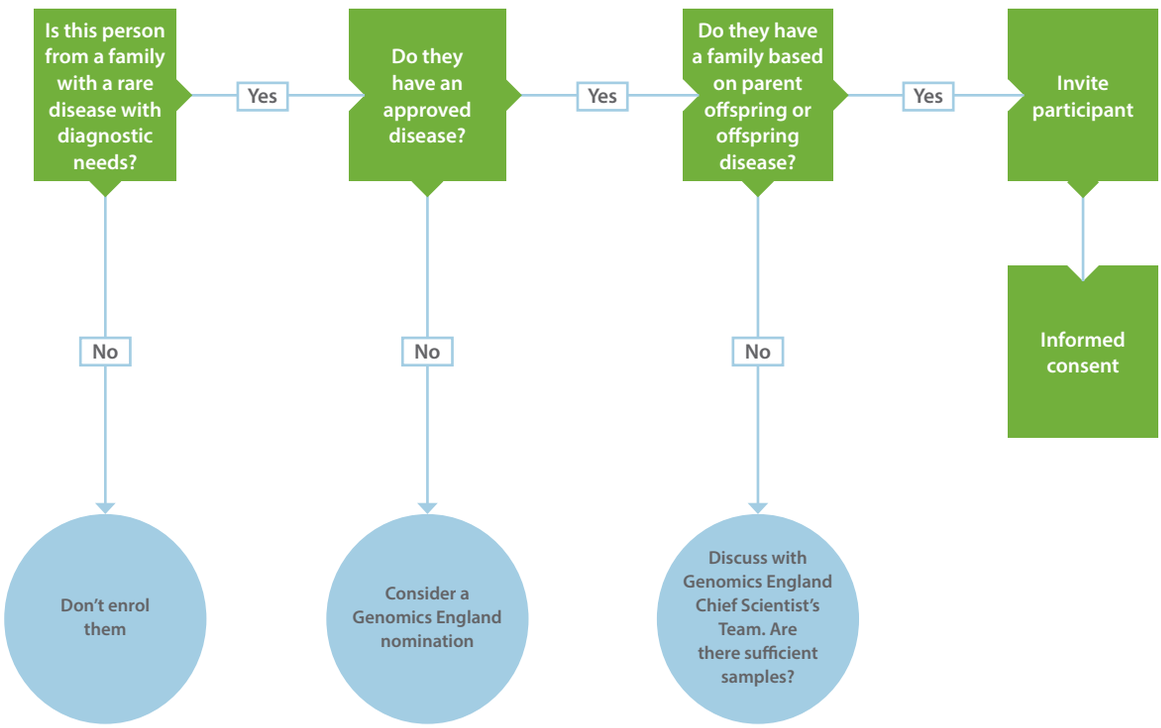
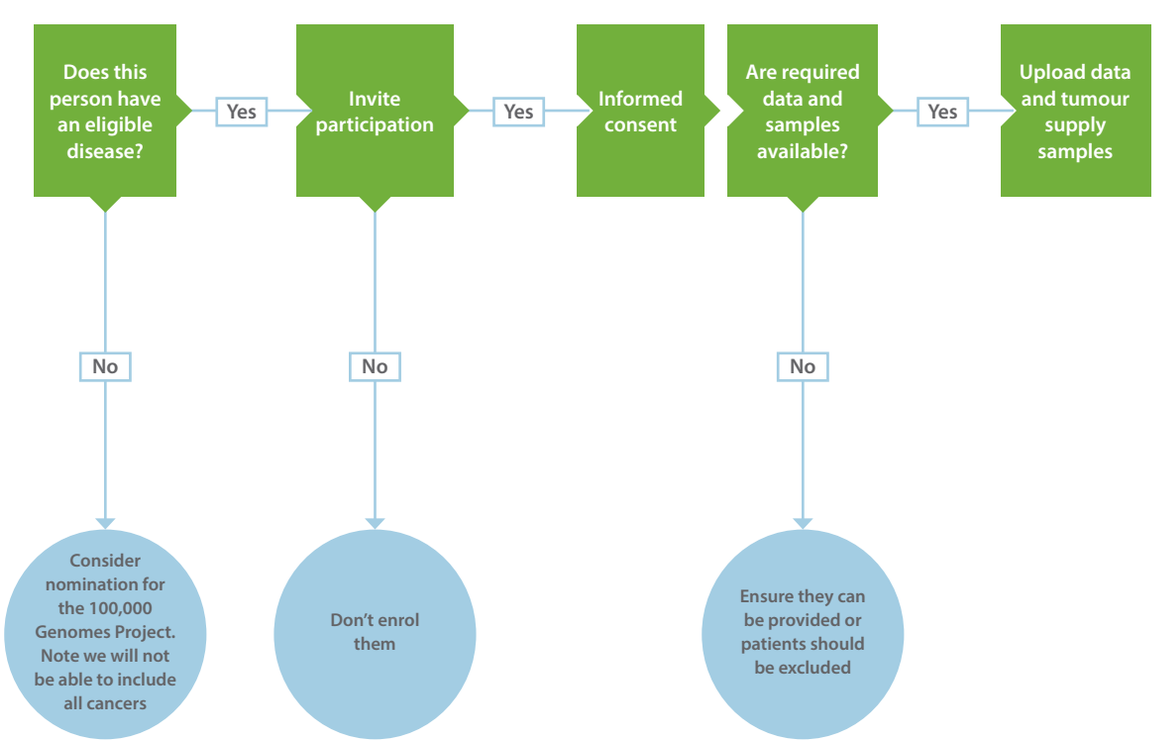


Figure 13. Inclusion process for cancer patients - questions for enrolment



11.1.1 The process for consent in the 100,000 Genomes Project

Patients under the care of the NHS in England or the HSCNI will be referred by their treating clinician for invitation to join the Project, because they have either a (suspected) rare disorder, certain forms of (suspected) cancer or a specific infectious disease, and joining the programme may reveal important information about their health. Eligible relatives of probands in the above categories may also be invited to join the Project to inform research regarding the health of the proband.

In some rare and very specific cases, a deceased participant may be included with the support of their treating clinicians (or their relatives), provided that appropriate phenotypic information and DNA samples are available and appropriate consents are given. His or her relatives may also join the Project in order to gain information about the deceased proband's health condition. The inclusion criteria for the Project are specified elsewhere.

11.1.2 The consent form and patient information literature

Drawing on relevant guidance and exemplars of good practice, the participant materials (i.e. letters of invitation, Patient Information Sheets, and assent and consent forms) have been developed in accordance with the key principles in the Key Ethical Principles document and Ethical Governance Framework, and with the advice of the Ethics Advisory Committee, with comments from the Science Committee, and other stakeholders and the appropriate guidance.[35]

The consent discussion with participants will draw attention to key elements of the Project that may be regarded as having particular ethical, legal or social implications. This discussion will include seeking explicit consent to:

- Mandatory potential whole genome sequencing (WGS), genomic and other "omic' analysis of the donated samples.
- Mandatory linked access to participants' health records in perpetuity.
- Mandatory feedback of pertinent findings.
- Opt-in to feedback of a small number of looked-for additional clinically important findings.
- Mandatory access to pseudonymised participant information and samples by third parties with the appropriate permissions (who may be based in the UK and overseas) and which includes commercial ('for-profit') companies.[36]
- Mandatory agreement to recontact by their clinical team, in order to permit invitations to participate in further studies (including relevant drug trials, or ethics and social research to do with participating in the Project).

The patient literature is also written so that the discussion can draw potential participants' awareness (or that of an adult consenting on behalf of a child or that of a consultee advising in respect of an adult participant who lacks capacity) to key issues or risks from joining the Project, including that:

- For patient groups needing particularly time-sensitive treatments (such as in cancer or extreme response to infection) the results of genomic analysis will usually **not** be able to be returned to their clinician in a timeframe that can influence clinical decision-making around treatment for their pertinent condition. However, the return of clinical information is expected to become swifter during the duration of the Project.
- The return of **any** information of clinical relevance, such as a diagnosis, cannot be guaranteed to participants.
- In general, at this stage in scientific understanding, the significance for health and disease of information revealed by WGS may be relatively uncertain (particularly in the absence of a relevant family history). This knowledge is expected to improve to some extent during the duration of the Project, but findings of unknown significance are quite likely to be generated. These will **not** be returned to the clinician for discussion with the patient except in very exceptional and pre-defined circumstances.

- Should a specific and serious risk to the health of a participant's close family member(s) become apparent via the participant's pertinent findings, or via additional findings (whether these are unexpectedly detected or purposely 'looked-for') and where preventive or ameliorative action is possible in regards to these close relatives, the clinical team is likely to seek to contact the relatives to inform them of this. This may include risks identified after the participant has died.[37] Whilst measures will be in place to keep participants' identity and information confidential, there is an unavoidable, though remote, risk of re-identification through participation.[38]
- The ways in which data security, privacy, access and sharing access to samples and the datasets will be appropriately restricted and monitored.
- Participants can request to see the information held about themselves by the Project and to know the types of third parties who have accessed their data (taking into account all legal obligations under Freedom of Information or Data Protection).
- The process for withdrawing consent to remain in the Project.
- Given the long term nature of the donation, in future, the samples and information donated may be subject to the application of as-yet unanticipated analyses and investigation because of new knowledge.

11.1.3 Giving consent face-to-face

Generally, consent to participate in the Project will be taken as part of a face-to-face discussion by experienced clinicians, study coordinators, nurses or other trained members of staff. This discussion will conform to established good consent practice guidelines.[39] Where possible this should be timed to fit in with scheduled appointments. During the consent discussion, potential participants and people accompanying them will have the opportunity ask questions about the Project or about participating in it. All potential participants (or people consenting on their behalf) will have been given appropriate time to read and consider the Project's patient information literature. Literature may be translated by the Genomic Medicine Centres (undertaken forward and back) where needed in languages appropriate to the patient groups to be recruited. Dependent on demand, Genomic England may commission and hold a few commonly required translations centrally for electronic download.

In order to record their consent, the potential participant (or the person consenting on their behalf, or the person advising about someone else's participation) will need to follow the instructions on the consent/consultee declaration form, to initial the boxes, and to sign and date the final section, adding their own date of birth as an additional confirmation of their identity. We intend to adopt an electronic touch screen for recording consent, adapting processes used by UK Biobank.

11.1.4 Giving consent at a requested home visit

It is ideal if all consent and procedures take place in a suitable healthcare environment. If a participant (or the person who might consent on their behalf) requests that the consent discussion takes place in their home (or in another appropriate healthcare setting) it will be a decision for the Genomic Medicine Centre as to whether this can be provided as an option. If possible, an appropriately trained team member(s) will schedule an appointment convenient to the potential participant(s), and will travel to the agreed location.

11.1.5 Giving consent by post or email

In exceptional cases, some potential participants may be offered the opportunity to give consent to be part of the Project via the return of their completed consent forms to the recruiting clinician or appropriate alternative in the post/via emailed document. A telephone discussion with a suitably trained person will be offered to them as part of this, to take place at a time suited to the potential participant.

This route to seeking consent will only be used where the clinical team has reason to think that seeking postal consent is an appropriate method for the individual concerned and that consent could not be sought face-to-face.

Participants aged 16-18 years

At ages between 16 and 18, it is legally presumed that young people have the ability, or competence to make decisions about their own health care or treatment. When a young person is believed to be competent, consent from those with parental responsibility is not legally required.

On reaching 16 years old, **all** existing participants will be asked to give their own consent to remain in the programme (unless it is deemed by their medical team that they do **not** have the capacity to do so at that time). This also applies to those being invited to join the programme at the age of 16.

In most circumstances, young people consenting on their own behalf will be encouraged to involve the person(s) with parental responsibility for them in their decision-making.[40]

Some participants in the Project who join at younger ages, will then reach the age of 16 as participants and at that point will be deemed not to have capacity to consent (or to refuse), their continued participation in the Project on their own behalf. In such cases, the advice of a personal or legal consultee will be sought regarding their continued involvement, in accordance with the provisions of the Mental Capacity Act 2005 for people in research studies in England and Wales.

Participants under the age of 16 years

Where a potential participant is aged **under 16**, and has **not** been deemed by their clinical team to have attained the capacity to consent to participation on their own behalf, at least one person with parental responsibility for them, or a guardian, will need to provide consent in order for the minor to participate in the Project.[41]

The Project will follow appropriate guidance as to the involvement of children in research.[42]

Patient Information Sheets are available for the Project versioned for 6-10 year olds or 11-15 year olds; and versioned for use either by the proband participant, or by children and young people who join the Project because they are related to a proband.

All children and young people will be given verbal information about the Project by the person conducting the consenting discussion (who should have experience of working with children/young people, in accordance with the minor's age and individual circumstances) and will be given opportunities to ask questions.

The person taking the consent should seek the assent of the child as part of the discussion regarding their participation. If the child or young person is capable of assessing the information provided, they will consider their explicit wishes, including their refusal to take part, or desire to withdraw from the study.[43]

Seeking assent implies that children could have an understanding of the research process and can be informed about what is involved and what they are expected to do. This can be an opportunity for some children to express their opinions and concerns surrounding participation in research, providing them with a formal means to be included or excluded. Where appropriate, the person taking the consent may wish to offer the child the opportunity to evidence their assent via the completion of an Assent Form. These are provided for the use of 6-15 year olds in the Project.

At the time this discussion takes place, the child or young person may not be in a position to, or may not wish to, complete this form. During any discussion around assent, (regardless of whether an Assent Form is offered, or completed), a completed Assent Form will not be considered by the person taking the consent to be the sole or overriding signifier of a child's willingness to participate in the Project.

If, at an age below 16, a young participant **has** been deemed competent by a medical team to consent in their own right to remain in the Project, they will be invited to give this consent. The same will apply to young people being invited to join the programme under the age of 16, who have been deemed competent by a medical team to consent in their own right to joining the Project.[44] Young people consenting on their own behalf under the age of 18 will be encouraged to involve their parent or guardian in their decision, as a matter of good practice.[45]

11.1.6 Seeking consent where participation relates to a health emergency

In specific scenarios and in line with the appropriate guidance, consent may be sought to join a person as a proband into the Project while they are in a health emergency, such as in the case of an extreme response to sepsis.[46]

Adults not able to consent for themselves due to an emergency situation

In a health emergency situation, the law in England and Wales allows adults unable to consent for themselves to be recruited into a project **without** prior advice from a consultee, specifically in regards to that emergency situation. Therefore, and on the advice of the patient's clinical teams, this Project will accept as participants eligible adults who lack capacity due to an emergency situation.

This will include where the timeframe for recruitment does **not** permit the appointment of a consultee **before** including the patient into the Project as a participant. This exceptional route to participation is proposed because the potential or actual clinical benefit to be gained by the patient if he or she is included in the Project is likely to depend on their prompt recruitment. Participation may provide clinically relevant information to inform the patient's care and may also inform the care offered to the patient's close contacts (for example if the health emergency was caused by an infectious disease).

Recruitment would only take place prior to the appointment of a personal or nominated consultee appointed under the principles of the Mental Capacity Act 2005:

- If it is not reasonably practicable to seek advice from a personal or nominated consultee, and
- The procedure has been approved by a NHS Research Ethics Committee; and
- Provided a consultee is consulted **as soon as possible** to seek advice on the participant's likely views and feelings on being included in the Project.[47,48,49]

When the patient's consultee is appointed and consulted, they will be given sufficient information about the Project in an understandable form, to permit their informed advice. (They are free to decide to provide this advice or not.) Advice given by consultees will be recorded by the person seeking the recruitment on a Consultee Declaration form (not a consent form). If the adult then recovers their capacity, as soon as possible they will be asked for their consent to participate in the Project on their own behalf.

The clinical team caring for the participant will also offer the participant themselves with information, appropriate to their current state, about the Project and its risks and benefits.[50]

Children who are unable to consent for themselves and in an emergency situation

In an emergency situation, in England and Wales, the law would allow children and young people to be recruited into the Project specifically in regards to that emergency situation, *without prior consent* from at least one person with parental responsibility or a guardian. These exceptional circumstances for recruitment would include:

- Prior NHS Research Ethics Committee approval for the project.
- The same research question could not be assessed by recruiting from a non-emergency environment, and that the research *is of potential benefit* to the child/young person themselves.[51]
- Someone with parental responsibility for the child/young person is informed about the research *as soon as possible*.
- Their consent (and the child/young person's assent if they are able to give this) is sought *as soon as possible*.
- The person seeking the recruitment (who must have previous experience of working with children/young people) makes it clear to the child/young person or their parent (if the child/young person is not competent) that the child/young person can withdraw (or be withdrawn by their parent) at any time and without needing to give any reason.[52]

Adults consenting on behalf of children/young people who began participation in an emergency will be informed at the time they give consent, that if and when, the child/young person attains/regains capacity, their consent will be sought for participation as would normally be the case.

Participants who regain capacity during their participation in the 100,000 Genomes Project

In all circumstances, any person who attains (or regains) the capacity to decide whether or not to become a participant in the Project will be asked to give us this consent in their own right as soon as is practically possible.

Those acting as consultees for people whilst they were lacking capacity will be informed at the time they gave their advice, that should the person become a participant in the Project and then regain their capacity, that their consent will be sought for participation in their own right.[53]

Recruiting relatives of the proband in emergency situations

Adult or child relatives of the proband would not be recruited if they themselves were in a health emergency situation, prior to the appointment of a personal or nominated consultee, unless in exceptional circumstances. Consent would instead be sought for their participation (or advice from a consultee as appropriate) in the normal way, after the health emergency has passed.

Participants in the 100,000 Genomes Project who lose capacity to consent

Participants who have joined the Project by giving consent for this in their own right, may afterwards lose their capacity to give their informed consent to take part in treatment or research, or this capacity may become fluctuating.

The consent they gave to become part of the 100,000 Genomes Project, at the time when they had capacity fails at the point it is reported to Genomics England that they cease to have this capacity. Once they become aware of a loss of capacity, the Clinical care team should seek to appoint a personal or (if unavailable) a nominated consultee to advise on the patient's wishes, in line with the Mental Capacity Act Code of Practice published by the Office of the Public Guardian.

In order to safeguard the interests of adult patients who join on the basis of their own consent, Genomics England will ask Genomic Medicine Centres to implement a specific check on each participant's capacity (in cases where a participant has consented to join the Project in their own right), around five years after their consent was first given to join the Project. This routine check should be done via appropriately trained staff on the clinical team at the closest appropriate clinical contact time to that five year interval, so as not to cause undue burden to the participant or to their clinical team.

Then, for as long as the person remains in the Project under their own consent, the Genomic Medicine Centres will then be sent a prompt to instigate further routine capacity checks via the clinical team at (approximately) five yearly intervals. This instruction to re-check from the GMCs will be prompted by a reminder issued from the organisation running the 100,000 Genomes Project. (This obligation on Genomics England would also pass to any successor organisation with the responsibility for the ongoing collection of data for the 100,000 Genomes Project. An obligation to participate in this process has also been included in the commissioning of NHS Genomic Medicine Centres.)

Under the Mental Capacity Act as applied in the research setting, we note that there is no legal obligation on researchers to exhaustively ascertain whether or not a participant has lost capacity. For example, if a patient becomes untraceable and falls out of contact with their clinical care team whilst they are also a participant in this programme, it would not be required that Genomics England should require the participant's clinical team to take measures to seek to restore contact with the patient for the purposes of making this capacity check.

In accordance with the Mental Capacity Act, Genomics England will continue to assume that an individual participant retains capacity, even where a capacity check is inconclusive in relation to that particular individual. Examples of this might be that the participant does not respond to the invitation to this check, or he or she does not wish to take part in this discussion, or if a participant becomes uncontactable to the clinical team.

Capacity may also be assumed where the health care professional tasked to conduct the check by the GMC does not return any results from the check to the GMC (to then return to Genomics England).

However, where results have been ascertained by the GMC as to the capacity of any participant, whether via the five yearly capacity check, or as determined in the course of routine clinical care, the GMC must pass this information on to Genomics England. The GMC has discretion as to whether or not to pass this information either directly to Genomics England upon receipt, or at the five yearly point that it is requested.

If a clinician wishes to notify the GMC of a participant's loss of capacity or fluctuating capacity at any intervening point they may of course do so. In normal circumstances this would then prompt the seeking of a personal (or if necessary, nominated) consultee in line with the Mental Capacity Act Code of Practice, published by the Office of the Public Guardian. At the point that Genomics England is informed that the participant has lost capacity, the individual will immediately be removed from the Project as a participant.

If an adult participant is discovered to have lost capacity, (and where no personal or nominated consultee has been appointed) clinical teams must advise their GMC of this fact. The Genomic Medicine Centre will then inform Genomics England of this, in order for Genomics England to remove that individual as a participant in the 100,000 Genomes Project forthwith. For example this means that prospective data collection in relation to that individual will cease once this notification is received. However data already held by Genomics England under prior consent can still be used in accordance with the terms of the consent. In accordance with the principles of the Mental Capacity Act 2005, as applied in the research setting, the affirmative advice of the personal or nominated consultee is then required before the person lacking capacity to be able to be re-joined as a participant into the Project. If the consultee advises against this, clearly the person may not be re-joined into the Project as a participant.

The advice of the Consultee that the person lacking capacity should be re-joined into the Project as a participant will e.g. trigger the resumption of prospective data collection regarding that individual. As with any other participant, this will also permit data collection from recorded information covering any period during which the person was not a participant in the Project.

The advice of the Consultee is also required in each instance in relation to any Project activities which both:

- involve an adult participant who is deemed to lack capacity, and are
- physically intrusive, such as the taking of a blood sample, (unless this activity is taking place within a health emergency situation, in specific circumstances as discussed above).

Where the consultee advises that in this instance the participant would not have wanted to undergo the physically intrusive activity that is proposed, then that activity must not be carried out.

In line with the provisions of the Mental Capacity Act as applied in the research setting, should a consultee advise the participant's clinical team that the participant would have wanted to be withdrawn from the Project (or obviously should the participant recover their capacity and withdraw themselves), this will begin the process of withdrawal without further delay.

A participant's family may sometimes have a different opinion about their participant relative's continued participation in the Project after their relative has lost capacity. This will be handled sensitively by the participant's clinical team, with relatives being encouraged to respect their loved one's wishes, which were given at the time when they were freely able to act upon these. However, we note that the Project has a legal obligation to carefully consider a request made by a representative to withdraw a participant.[54]

Details for how to withdraw are clearly noted on the Patient Information Sheet and will be publicly available on the Project's website.

If a participant in this Project becomes aware that their capacity to give full and informed consent may be likely to lessen, or that their lifespan is likely to be shortened, and they would like to withdraw from the Project whilst they have the capacity to make this decision for themselves, they are of course able to withdraw themselves from the Project at any time and without being required to give a reason. Participants may or may not also

choose to take this opportunity to also inform the clinical team of the names and contact details of individual(s) whom they would like the clinical team to approach as personal consultee(s) in the event that they lose mental capacity. If a participant wishes to 'lodge' this information with their care team in circumstances where they are not considering their imminent or future withdrawal, this is also acceptable.

11.1.7 Seeking consent to use samples and information regarding a person who is deceased

Project participants who die

Some participants will join the Project giving consent for this in their own right and will then die at a later date, whilst they are still participants in the Project. In line with the Human Tissue Act 2004 (in England, Wales and Northern Ireland) and HTA Code of Practice on Consent, (v14.0, July 2014) the participant's consent to join the Project would remain valid even after their death. This provision is explicitly outlined to potential participants at the time of consenting. It is in place to maximise medical researchers' access to deceased participants' samples and health and personal information, thus maximising the potential for increased medical knowledge far beyond the participant's own lifespan.

Although the participant's consent extends beyond their death (e.g. to the collection of recorded health data that emerges after their death or to continued access by approved researchers to the appropriate parts of the participant's information held by the Project), the participant's relatives (including those in qualifying relationships) may sometimes have a different opinion after the participant has died. This view will be handled sensitively by the clinical team of the deceased person with relatives being encouraged to respect their relative's wishes (or in certain cases, to respect the decisions given by the deceased person's nominated representative/nominee).

We note that the decisions made by a nominee of this kind, appointed by the (now deceased) participant to take decisions after the participant's death, cannot be overridden by others, even those in qualifying relationships with respect to the deceased person. We note the Project's legal obligation to consider a request made by a representative to withdraw a deceased participant from inclusion in the project.[55]

It is also explicitly outlined at the time of consenting, that should a specific and serious risk to the health of their close family member(s) become apparent via the participant's pertinent findings, or additional findings and where treatment exists, then the clinical team are likely to seek to contact these family members, to inform them of this. This may include risks that are identified after the participant has died.[56]

People who die before giving consent to join the 100,000 Genomes Project: legacy collections

There may be specific situations where we would like to include samples and available health information from people who have already died, but who are not already participants in this Project. This will generally be in situations where the equivalent information is difficult or impossible to obtain from a living person within the anticipated time frame of the Project (to the end of 2017). This may include where a deceased person affected by one of the disease groups included in the Project has left biobanked samples or health information as a legacy for use in further medical research.

Legacy collection samples may benefit living patients via inclusion in our Project for example by increasing the number of samples available to researchers within a particular disease group. For example, where a child participant in our Project has a particularly rare form of cancer, other participants with the same condition are unlikely to be found in the time frame of the Project. Biobanked samples and information can therefore be particularly precious to researchers in these situations. We note also that the deceased patients who donated these samples were clearly keen that they should be used to potentially help others after they themselves have died.

The inclusion of relevant legacy samples in the Project will only be considered where likely to add substantially to knowledge and potentially offer benefit to patients. The aims and uses to which the Project would put these samples and information would need to be agreed as being scientifically and ethically in accordance with the broad consent given by the patient who donated them to be biobanked. This would be agreed by the relevant parties holding the samples and information and by the Genomics England Science Advisory Committee, the Ethics Advisory Committee and the Genomics England Board.

Adults who die before giving consent to join the Project: seeking consent where a person has (perhaps very recently) died

Where a person with health problems which may relate to the disease groups in the project, has died without receiving a genetic diagnosis for their condition, and their living relatives might benefit from knowing this, we will consider the inclusion of an appropriate sample and their information into the Project, on the advice of their clinical team.

Where the adult concerned has not left specific directions to cover this eventuality, consent from the appropriate person(s) who are in a 'qualifying relationship'[57] to the recently deceased person at the time of their death will be sought, in order to allow us to include the deceased person in the Project in accordance with the Human Tissue Act.[58] This consent will be sought where:

- The deceased person had **not** previously specifically consented to join the project before they died (and also they had not specifically refused, or withdrawn from participation).
- The deceased person's clinical team would like to include their information and tissue in the Project which will require the use or storage of tissue removed from their body after death.
- Genomics England supports the inclusion of the deceased person's samples and information into the Project.

The consent given in regards to the deceased adult's participation in this Project should be recorded using the Consent form in respect of a deceased adult and corresponding Participant Information Sheet in respect of a deceased adult.

An adult who has died in England, Wales and Northern Ireland, may have already nominated someone to represent them after their death and to give consent on their behalf. Those close to the deceased will be asked whether this is the case. If there is such a nominated representative, he or she will be asked by an appropriate person from the deceased person's clinical team to give consent in regard to the deceased person joining the Project. This nominated representative's consent must not be overridden by other individuals, including those who were in a 'qualifying relationship' with the person immediately before their death.[59,60]

In cases where an adult has recently died, their nominee or the person(s) in a qualifying relationship to the deceased person would be approached in a sensitive and timely manner by the healthcare professional seeking their consent. This trained person will use their clinical judgement to decide whether it is appropriate to make this request. They will explain why the deceased person's samples and health information might produce pertinent information which may benefit others in similar situations. Research use of the samples taken from their loved one will be discussed, as well as whether these samples were taken as part of their standard clinical care while they were alive, or if necessary, will be gathered by removing a sample (blood or tissue) from their loved one's body after death, for use in this Project.

Additional looked-for findings will not be offered where the adult in question is deceased.

The healthcare professional seeking this consent is likely to be from the deceased person's clinical care team, or to be a public health professional who is already communicating with the family around necessary public health actions e.g. after a death from invasive meningococcal infections, identifying 'close contacts' and organising appropriate measures for them (such as antibiotics).

Children who die before giving consent to join the Project: seeking consent where a child has (perhaps very recently) died or after a stillbirth, miscarriage or termination of pregnancy at any gestation

If in connection with a (suspected) rare genetic disease, a child[61] or a baby has died or he or she was stillborn (i.e. born showing no signs of life and after more than 24 weeks gestation) or where a pregnancy has ended in miscarriage (i.e. at less than 24 weeks gestation) or a termination of pregnancy; it can be very important to the bereaved parents that the reasons for their loss can be sought by doctors and scientists and that they can be informed of any pertinent results.

Some parents may therefore wish for their child's or fetus' samples and information to be included into this Project, in order to look for a diagnosis of the condition which has affected them, and perhaps also to inform the parent's future reproductive decision-making, or that their close family members'. (We use 'child' as the descriptor in this

Protocol and in the patient literature to refer to a deceased child, baby or fetus at any gestation, although the clinical team should take care to use the same terms used by the woman or couple in conversation with them.)

Provided that the fetus or child's health problems may relate to the disease groups in the Project and the clinical team proposing this inclusion feel that the living relatives might benefit from being able to gain further information in connection with their loss, and where an appropriate sample(s) can be gathered, we will consider the inclusion of the sample(s) and related information into the Project, taking into account the advice of the child's or parent's clinical team.

The mother or parents may also wish for carrier status testing to be carried out as part of their own participation in the Project e.g. as part of a 'trio' in which the parents enter the Project together with the samples and information of the fetus, baby or child who has died. Consent to this carrier testing should be discussed separately as part of the consent process for the parents regarding their own participation and is to be recorded in their own (adult) consent form.

Even where the miscarriage or termination took place when the baby or fetus (depending on the terms the parent wishes to use) was at **under 24 weeks' gestation**, the Consent Form for parents of a (deceased) child with a rare genetic disease should be used to record the consent of the mother with legal parental responsibility for their baby or fetus' samples and information to be included in the project. The mother giving the consent should be asked to read the corresponding Participant Information Sheet for parents of a (deceased) child with a rare genetic disease.

While making a record of the woman's consent for the inclusion of her baby or fetus (who died at **under 24 weeks' gestation**) into the Project would be **legally sufficient** as an entry into her medical notes,[62] unless for some reason the clinical team find entering her consent into her medical notes to be the only appropriate course of action, we would expect the Consent Form for parents of a (deceased) child with a rare genetic disease to be used to evidence her consent. Taking this separate consent is important because it fits in with existing NHS clinical practice in respect of prenatal diagnosis more generally, and also because it makes it explicit that the woman has consented for the analysis of the fetus or baby's sample.

Having the two signed Consent Forms (one for the woman herself and one for the fetus or baby) also allows each sample to be tracked individually with its consent which is a standard requirement of this Project, including more straightforwardly facilitating, e.g. withdrawal of the fetus or baby's samples and information from the Project if requested.

The law requires that where a fetus at **above 24 week's gestation**, or a baby or child aged under 18 years has died, their remains should be handled in accordance with provisions for gaining consent for the use of the tissue of the deceased. The consent required here needs to be given by an adult with legal parental responsibility for the child who has died. Consent to permit the deceased child's inclusion in the Project should therefore to be evidenced by this adult using the Consent Form for parents of a (deceased) child with a rare genetic disease and after reading the corresponding Participant Information Sheet for parents of a (deceased) child with a rare genetic disease.

The consent of one person with parental responsibility is required for a child to be included as a participant in the Project after the child had died (above the age of 24 weeks of gestation), although the consent discussion with the clinical team may involve other family members as appropriate. If no one with parental responsibility for the child is available, then consent will be sought from someone in a qualifying relationship with the child.[63]

In all cases where a child or baby has recently died, or the pregnancy has recently ended in the miscarriage, stillbirth or termination of pregnancy, the appropriate family member(s) will be approached by the person seeking their consent in a sensitive and timely manner. This healthcare professional should use their clinical judgement to decide when and whether it is appropriate to make this request.

The healthcare professional should explain why the child or fetus' (blood or tissue) samples and related health information might produce pertinent information which may benefit their family members or may benefit others in similar situations. Ongoing research use of the samples taken will be discussed, as well as whether the samples that will be used by the Project have already been taken as part of standard clinical care, or if necessary, if they will need to be taken by removing a sample from their child's or baby's body after his or her death (or by

taking a sample from the fetal remains). Additional looked-for findings (about child onset conditions) will not be offered where the child in question is deceased.

If the parents make any specific request in relation to including their child's samples or information in the Project, that the person taking consent is not sure about, the child's clinical team or Genomics England should be contacted about this to resolve any outstanding issues, **before** the consent is taken.

Approaching relatives of the (living) proband to seek consent to participate in the 100,000 Genomes Project

Up-to-date information about the diagnosis and clinical status of other family members can be vital in (for example) determining risks for a proband, particularly when, for example, there appears to be an inherited predisposition to cancer. Accessing this is discussed in 'Consent and confidentiality in genetic practice' (2006, Joint Committee on Genomics in Medicine), the principles of which we have referred to in creating the relevant processes.[64] We note that access to information about other family members is governed by the Data Protection Act 1998 and the Access to Health Records Act 1990.

If the (relative) potential participant is an adult, the clinical care team may give the recruited adult participant (who may be the proband, or may not be e.g. a proband child's parent) a recruitment pack to pass on to their relative, if they are willing to take on this role of passing on information. The recruitment pack will contain a letter of introduction, Patient Information Sheet, Consent Form and a postage-prepaid envelope.

The recruited participant may also be provided with an electronic copy of the Patient Information Sheet and Consent Form for forwarding by email to their relative. To indicate their interest in participating in the Project, the relative can return the slip to say:

- They do NOT wish to be part of the Project.
- They agree to be contacted by a team member to find out more about the Project.
- They would like to be part of the Project and give written consent to participate.

The recruitment pack will contain contact details for the proband's clinical care team and the Project (via Genomics England). Relatives considering joining the Project will have the option to telephone or email for further information. Relatives can then make appointments with the proband's clinical care team to give blood or saliva samples.

If the potential participant is aged under 16 and lacks capacity to consent (dependent on their circumstances) their parents, guardian or at least one person with parental responsibility for the minor will be provided with the relevant Patient Information Sheet and be asked to consent on behalf of the minor.

11.2 Process for withdrawing from the 100,000 Genomes Project

The project will be more successful as a resource if few people withdraw once they have joined. However, participants can withdraw at any time without giving a reason. Participants will be reminded that they are welcome to discuss concerns with their clinical team at any time, or to contact Genomics England directly.

Full patient literature including how to withdraw, will also be available on the Genomics England website. This will list the various options for withdrawal and will provide the contact details for this process. Numbers withdrawing will be monitored by Genomics England, because withdrawal may indicate problems in the consenting or participatory experience.

Participants can withdraw at one of two levels:

'No further contact': this means that the Project would no longer contact the participant via their clinical team, but would have their permission to retain and use information and samples provided previously, and to obtain and use further information from health records. This level of withdrawal leaves the 100,000 Genomes Project dataset intact and will allow researchers to continue to study disease with the goal of improving knowledge and health.

'No further use': in addition to no longer contacting the participant or obtaining further information, any information and samples collected previously would no longer be available to researchers. The Project would

destroy samples (although it may not be possible to trace all distributed sample remnants), and would only hold information for archival audit purposes. Such a withdrawal would prevent information about the participant from contributing to further research, but it would not be possible to remove data from research that had already taken place.

In order to allow the clinical team and the project to action the withdrawal, the participant or those who give consent on their behalf, or another appropriate person (in the case of adults lacking capacity to consent of their own behalf), would need to evidence their wish to withdraw from participation in the Project, by requesting and completing the appropriate Withdrawal Form. This will be either the Full Withdrawal from Participation Form, or Partial Withdrawal from Participation Form.

These forms are available via the participant's clinical team. The relevant person will be asked how they would like (or would like the participant) to withdraw. The differences between the withdrawal options available will be explained by the person actioning the withdrawal with the patient via the use of a withdrawal form.
As expressed in the withdrawal information, they can choose from either:

Option 1) 'No further contact': Partial Withdrawal from participation

Not to be contacted directly by the 100,000 Genomes Project/Genomics England, any further.*

However, previously collected samples from the participant can still be used and collected information will continue to be updated into the Project's databases from the participant's records as usual.**

It is important to note that a participant's clinician might want to discuss with them information found in their information or samples, however, which would still be possible under this option.

OR

Option 2) 'No further use': Full Withdrawal from participation

Not to be a participant in the 100,000 Genomes Project any further.

Genomics England/The 100,000 Genomes Project will:

- Not contact the participant again.
- Put beyond any further use any samples we hold from the participant.

(It may not be possible to trace all divided parts of their samples that have been distributed to approved third party research projects. Any surplus from these samples is routinely destroyed after the third party's research is completed.)

- Put beyond any further use the information we hold about the participant (aside from what is required for audit purposes - we need to retain a record that they were once part of the Project and then withdrew).
- Not retrieve anything further from the participant's health or other records.

* In both options, 'any further' means from the point that Genomics England confirms to the participant's clinician that the withdrawal form has been received by us.

** By December 2017, the full Genomics England dataset shall comprise a number of collected, processed and stored constituent datasets. These include:
- Information relating to patient consent - including signed consent form (from referring NHS trusts and patients)
- Information about sample acquisition, processing, DNA extraction and quality controls (from referring NHS Trusts and Genomics England Biorepository)
- Demographic and administrative data relating to consented data subjects and (referring) NHS trusts. (from referring NHS trusts and patients)
- Phenotypic, clinical, laboratory, imaging, prescribing data collected in the course of routine clinical care (from referring NHS trusts)
- Whole Genome Sequence data – BAMs, VCFs, initial annotation (from sequencing supplier)
- Annotation data (from sequencing supplier and Genomics England Clinical Interpretation Partnership (GeCIP users)
- Interpretation and validation data (from GeCIP users)
- Longitudinal health outcomes data (from the Health and Social Care Information Centre (HSCIC), the Clinical Practice Research Datalink (CPRD), Public Health England etc.)
- Patient reported information e.g. on lifestyle, environment and dispensing of medication (from patients/carers)

A copy of this form would then be given by the clinical team to:

1) The participant who withdraws.
2) The participant's clinical team – to be retained locally in the participant's medical notes.
3) The Project - the clinical team will make Genomics England aware of the withdrawal and send a copy of the withdrawal form electronically for storage.

Participant withdrawal by patient group

Different arrangements are likely to be required by different participant groups to facilitate their withdrawal from the Project.

Withdrawal: Adults who can consent on their own behalf

This group can withdraw on their own behalf by completing the withdrawal form, which their clinical team will then process appropriately. Their withdrawal from the Project is considered to have been actioned from the point that Genomics England confirms to the participant's clinician that the withdrawal form has been received by us.

Withdrawal: Under 16 year olds deemed to have capacity to consent on their own behalf

A decision from a participant in this patient group to withdraw from participation in the Project, will be treated the same as that of an adult (aged over 16 years old), who is able to consent for themselves.

In normal circumstances, those with parental responsibility[65] or the appropriate legally authorised representative will be encouraged to be involved in this discussion, but the final decision about withdrawal rests with the *Gillick* competent young person.[66]

The decision to withdraw may be felt to be against the young person's best interests, but unless the Court of Protection decides that the decision could lead to death or severe permanent injury, which is unlikely to apply here, the young person's withdrawal of consent must be respected.

Withdrawal: Under 16 year olds who lack capacity to consent on their own behalf

Decisions around withdrawal from this Project on the basis of the child or young person's dissent will rely on individual circumstances to some degree. The usual practice around participation in research is not always appropriate in the context of this Project. It is the view of the Project that here, it is more appropriate to follow accepted practice in clinical care in relation to refusal of treatment, because the Project offers potential direct clinical benefits to the proband children and young people who participate. (The General Medical Council and Royal College of Paediatrics and Child Health advise in relation to children's participation in medical research, that if a child or young person 'objects or appears to object in either words or actions',[67] they will not be required to participate in research.)[68]

Decision-making around the clinical care of children or young people who are not Gillick competent should be therefore guided by an assessment of their best interests. It follows from this that it would *not* be legally determinative if such a child or young person withdrew their assent to participate in the Project (or declined to join the Project). Adult/s with parental responsibility retain the legal power to withdraw (or give) their consent to a child's participation in the Project.[69]

Children and young people are made aware via the patient information literature for the Project that their views will be taken seriously and if they don't want to join, or don't want to continue within the Project, that this dissent will 'usually' be accepted.

In practice, this acceptance may depend on the reason that the child or young person gives for seeking to withdraw. Actioning a withdrawal solely on the basis of reluctance expressed to give a blood sample for example, would be likely to entail different considerations than if, say, the child expressed dissent for other reasons.

If a child or young person expresses dissent for reasons which are perhaps not able to be mitigated straightforwardly, and there is no overriding contradictory factor relating to their best interests - especially where direct clinical benefit seems unlikely, then they should be withdrawn from the Project in accordance Iwith their wishes. This should happen even where consent continues from the adult or authority with parental responsibility for that child or young person.

In the case of children or young people who join the Project in order to contribute their information in relation to a proband (such as siblings), the potential for direct clinical benefit to them may be less obvious. In this case, if such a child or young person expresses their wish to withdraw from the Project, this should usually be actioned, on the assumption that benefit may be minimal and withdrawal is not likely to put the health of the sibling child or young person at risk of significant harm.

Withdrawal: Adults who lack capacity to consent on their own behalf

Adults lacking capacity may be recruited to the Project, or they may have joined it under their own consent and then have lost their capacity. In such cases, if the consultee of an adult participant advises that they would wish to be withdrawn, this withdrawal will be actioned, using the 'Withdrawal Form for a consultee to advise on an adult lacking capacity', without delay. This is in accordance of the requirements of the Mental Capacity Act in a research setting.

If an adult participant who lacks capacity implies or expresses their dissent directly, they will be withdrawn from the Project immediately in line with MCA and in accordance with GMC guidance which provides specific considerations to be made about the care and treatment of patients who lack capacity.[70]

11.3 Personal insurance and the participants in 100,000 Genomes Project

Under an open-ended agreement between the Department of Health and the Association of British Insurers the results of whole-genome sequencing carried out in the Project are not disclosable to personal life insurers. This is because they are part of an NHS transformation programme with a significant research element. However, insurers have the normal expectation that patients will disclose relevant family history, other non-genetic diagnostic test results and GPs' reports when applying for new insurance. Under the Genetics and Insurance Concordat and Moratorium, the results of other genetic tests do not need to be disclosed unless the test is for Huntington's Disease for life insurance and the insured sum is over £500,000. The Concordat and Moratorium has recently been extended from November 2017 until the end of 2019 (see attached letter) from the Department of Health and Association of British Insurers website.

Genomics england

12 Communications, stakeholder engagement and patient and public involvement

12 Communications, stakeholder engagement and patient and public involvement

12.1 Communications strategy and governance arrangements

Gaining and retaining public trust and confidence in the 100,000 Genomes Project is critical to its success, both in terms of continued recruitment and wider societal confidence in the use of genomics in medicine. Particular issues are privacy and security in relation to data held by Genomics England, and access to data services by commercial organisations and subsequent commercialisation.

A communications and stakeholder engagement plan has been developed with communications colleagues in DH and with input from communications leads in NHSE, PHE and HEE. A monthly meeting is held to ensure all partners have sight of, and input into, communications about genomics. Above this is a Communications Advisory Board, which has a diverse membership including specialists in the fields of public engagement in science, senior representatives of broadcast, print and online media, and key partners such as the Wellcome Trust, and Cancer Research UK. This body feeds into the Genomics England board through Vivienne Parry and then through to the DH Assurance Board.

12.2 Communications and stakeholder engagement plan

A communications and stakeholder engagement plan has been developed. Whilst some patient and public involvement (PPI) is specific to Genomics England (for instance, in relation to policies around consent and the development of participant literature), there are broader societal concerns about the use of genomics. For instance, in relation to insurance or how much feedback NHS patients should expect, whether it should be used for postnatal screening (and by logical extension, therefore antenatal screening), whether its use is part of a 'responsibility' agenda, which may restrict future access to the NHS etc. These issues require extensive dialogue with the public before genomic medicine is introduced into routine use by the NHS. A number of stakeholders have interests in this area, including NHS England, Health Education England, Public Health England, Department of Health, the Department of Business, Industry and Skills (through its Science and Society division), Department for Education and Science (in relation to the inclusion of genomics in the school curricula), the research councils, medical charities including the Wellcome Trust and other members of the Association of Medical Research Charities, patient support groups and broader interest groups. Genomics England will play a leading role in a coalition of organisations, to develop, initiate and deliver, this essential public dialogue.

12.2.1 Genomics England research projects – seeking the views of patients and the public

Patient and public involvement (PPI) and public engagement

Patient and public involvement is an integral and vital part of the Project. Potential participant views were sought by Genomics England, on key ethical policies relating to consent for rare disease and cancer using a range of experienced, independent specialist market research companies (e.g. Solutions for BAME work, GfK for cancer work, Genetic Alliance UK for rare disease). This feedback was then used in a separate project to inform the design and content of literature. This literature was then extensively tested with potential participants before submission to REC. Literature supporting consent was also tested with professionals involved in the consent process, in order to ensure that there was no divergence in understanding between professionals and patients. Genomics England intends to further evaluate and revise its literature, and patient and public materials at the 10,000 participant recruitment point.

Bearing in mind the need for equity of access, and mindful particularly of those communities disproportionately affected by rare disease, a programme of work with BAME groups was initiated to understand how best to involve these groups in the project. Learning from this work will be fed into both project recruitment and literature.

It is intended that those participants who wish to establish a relationship with the Project will be invited to become part of a 'genome pioneer' group. They will be sent information and news on the use of data and the progress of the Project and invited to become part of what it hoped to be a vibrant online community of 'genome pioneers'.

In addition to Town Hall format public meetings held in London, Cambridge and Newcastle organised by Genomics England, a series of public meetings (London and Birmingham) independently organised by Progress Educational Trust have taken place. One element of work previously commissioned by DH to assess public attitude to data sharing and commercialisation, has been repeated in order to provide directly comparable evidence of the effect, if any, of core data on public attitudes to data sharing.

Genomics england

13 Appendices

13 Appendices

Appendix 1 | References

1. www.genomicsengland.co.uk
2. https://www.gov.uk/government/publications/rare-diseases-strategyTrc and Bioresource
3. Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. Nat Rev Genet. 2013 Oct;14(10):681-91.
4. Bragin E, *et al.* DECIPHER: database for the interpretation of phenotype-linked plausibly pathogenic sequence and copy-number variation. Nucleic Acids Res. 2014 Jan 1;42(1):D993-D1000.
5. http://phenotips.org/
6. http://www.irdirc.org/wp-content/uploads/2013/06/IRDiRC_Policies_Longversion_24May2013.pdf
7. http://www.orpha.net/consor/cgi-bin/index.php
8. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW.Cancer genome landscapes. Science. 2013 Mar 29;339(6127):1546-58. doi:10.1126/science.1235122.
9. Lawrence MS, *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. Nature. 2014 Jan 23;505(7484):495-501.
10. International network of cancer genome projects. International Cancer Genome Consortium, Hudson TJ *et al.* Nature. 2010 Apr 15;464(7291):993-8.
11. Jacobsen A, *et al.* Analysis of microRNA-target interactions across diverse cancer types. Nat Struct Mol Biol. 2013 Nov;20(11):1325-32.
12. Omberg L, *et al.* Enabling transparent and collaborative computational analysis of 12 tumor types within The Cancer Genome Atlas. Nat Genet. 2013 Oct;45(10):1121-6.
13. Cancer Genome Atlas Research Network, Weinstein JN, *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet. 2013 Oct;45(10):1113-20.
14. Murtaza M, *et al.* Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. Nature. 2013 2; 497(7447):108-12.
15. Chapman SJ, Hill AV. Human genetic susceptibility to infectious disease. Nat Rev Genet. 2012 Feb 7;13(3):175-88.
16. Didelot X, Bowden R, Wilson DJ, Peto TE, Crook DW. Transforming clinical microbiology with bacterial genome sequencing. Nature reviews Genetics 2012;13:601-12.
17. Kohane IS. Using electronic health records to drive discovery in disease genomics. Nat Rev Genet. 2011 Jun; 12(6):417-28.
18. MRC/ Wellcome Trust 'Framework on the feedback of health-related findings in research', March 2014 http://www.wellcome.ac.uk/stellent/groups/corporatesite/@policy_communications/documents/web9document/wtp056059.pdf
19. MRC/ Wellcome Trust 'Framework on the feedback of health-related findings in research', March 2014 http://www.wellcome.ac.uk/stellent/groups/corporatesite/@policy_communications/documents/web_document/wtp056059.pdf
20. Wellcome Trust and Medical Research Council (2012) Assessing Public Attitudes to Health Related Findings in Research http://www.wellcome.ac.uk/About-us/Publications/Reports/Public-engagement/WTVM055197.htm
21. MP: We note that survey results may not always map onto actual behaviour or choices when patients are actually offered a test or information. A good example of this is the history of the Huntington's test which more than 80% of people asked said they would take the test if at risk but where in fact fewer than 10% do actually take the test if it is actually available.

22. Sources include:

- Canadian Institutes of Health Research Available at: http://www.pre.ethics.gc.ca/pdf/eng/tcps2/TCPS_2_FINAL_Web.pdf Accessed November 2014.
- Fabsitz *et al.* (2010) Circulation: Cardiovascular Genetics 2010; 3: 574-580.
- Presidential Commission for the Study of Bioethical Issues Available at: http://bioethics.gov/node/3183 Accessed November 2014.
- American College of Medical Genetics and Genomics Available at: https://www.acmg.net/docs/ACMG_Releases_Highly-Anticipated_Recommendations_on_Incidental_Findings_in_Clinical_Exome_and_Genome_Sequencing.pdf Accessed November 2014.
- Wellcome Trust and Medical Research Council Available at: http://www.wellcome.ac.uk/stellent/groups/corporatesite/@policy_communications/documents/web_document/wtp056059.pdf Accessed November 2014.

23. Wellcome Trust and Medical Research Council 'Framework on the feedback of health-related findings in research', March 2014. http://www.wellcome.ac.uk/stellent/groups/corporatesite/@policy_communications/documents/web_document/wtp056059.pdf

24. Wellcome Trust and Medical Research Council 'Framework on the feedback of health-related findings in research', March 2014. http://www.wellcome.ac.uk/stellent/groups/corporatesite/@policy_communications/documents/web_document/wtp056059.pdf

25. The Presidential Commission for the study of Bioethical Issues (2013) 'Anticipate and Communicate: Ethical Management of Incidental and Secondary Findings in the Clinical, Research, and Direct-to-Consumer Contexts', Recommendation 14. http://bioethics.gov/node/3183

26. Bredenoord AL, Onland-Moret NC, Van Delden JJ. Feedback of individual genetic results to research participants: in favour of a qualified disclosure policy. Hum Mutat. 2011 Aug; 32(8): 861-7. http://onlinelibrary.wiley.com/doi/10.1002/humu.21518/pdf

27. Richards, C. Sue, *et al.* "ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007." Genetics in Medicine 10.4 (2008): 294-300.

28. Criteria wording from The Public Health Genomics Foundation (2013) Managing incidental and pertinent findings from WGS in the 100,000 Genomes Project. Recommendations, para 8.5, p. 18. http://www.phgfoundation.org/reports/13799/

29. British Society for Genetic Medicine (2010) Genetic Testing of Children http://www.bsgm.org.uk/media/764635/wgs_discussion-paper.pdf

30. See Privacy and Progress in Whole Genome Sequencing, The Presidential Commission for the Study of Bioethical Issues, 2012 Available at: http://bioethics.gov/sites/default/files/PrivacyProgress508_1.pdf ; Anticipate and Communicate: Ethical Management of Incidental and Secondary Findings in the Clinical, Research, and Direct-to-Consumer Contexts, The Presidential Commission for the Study of Bioethical Issues, 2013. Available at: http://bioethics.gov/node/3183

31. Broad consent versus dynamic consent in biobank research: Is passive participation an ethical problem? European Journal of Human Genetics (2013) 21, 897–902; Available at:http://www.nature.com/ejhg/journal/v21/n9/full/ejhg2012282a.html

32. These include the Human Rights Act (1998), Human Tissue Act (2004), Mental Capacity Act (2005), Declaration of Helsinki (2008, World Medical Association), General Medical Council 'Good Medical Practice' (2013), 'Consent and confidentiality in genetic practice', (2006, Joint Committee on Genomics in Medicine),; 'Informed Consent Elements Tailored to Genomics Research', National Human Genome Research Institute, National Institutes for Health, 2012.

33. The Diagnosing Developmental Disorders project Website is at http://www.ddduk.org/

34. See Realising Genomics in Clinical Practice, PHG Foundation, 2014. Available at: http://www.phgfoundation.org/reports/16447/

35. This includes Health Research Authority (2014) Consent and Participant Information Sheet Preparation Guidance, Available at: http://www.hra.nhs.uk/resources/before-you-apply/consent-and-participation/consent-and-participant-information/ and Good Medical Practice (2013) Consent to research: Giving information in a way that people can understand. Available at: http://www.gmc-uk.org/guidance/ethical_guidance/6465.asp

36. Anonymised defined as per GMC Good Medical Practice (2013) Confidentiality guidance: Glossary: 'Information from which individuals cannot reasonably be identified. Names, addresses, full postcodes or identification numbers, alone or together or in conjunction with any other information held by or available to the recipient, can be used to identify patients'. Available at: http://www.gmc-uk.org/guidance/ethical_guidance/confidentiality_glossary.asp

37. See GMC Good Medical Practice 'Confidentiality guidance: Genetic and other shared information' paras 67-69 http://www.gmc-uk.org/guidance/ethical_guidance/confidentiality_67_69_genetic_and_other_shared_information.asp and GMC Good Medical Practice 'Confidentiality guidance: Disclosure after a patient's death', para 70 http://www.gmc-uk.org/guidance/ethical_guidance/confidentiality_70_72_disclosure_after_patient_death.asp

38. We follow the recommendations of the Expert Advisory Group on Data Access (EAGDA) statement of 2013 here: http://www.wellcome.ac.uk/stellent/groups/corporatesite/@policy_communications/documents/web_document/wtp055972.pdf

39. See GMC Good Medical Practice (2013) Guidance on Consent to Research http://www.gmc-uk.org/guidance/ethical_guidance/5993.asp and Health Research Authority Guidance on consent and participant information sheet preparation guidance http://www.hra-decisiontools.org.uk/consent/index.html , and British Medical Association Consent tool kit http://bma.org.uk/practical-support-at-work/ethics/consent-tool-kit

40. As GMC Good Medical Practice (2013) also notes: See the Declaration of Helsinki and Medicines for Human Use (Clinical Trials) Regulations 2004, which requires parental consent to complement even competent under-16s' agreement to involvement in trials.

41. For further information about parental responsibility, see General Medical Council, Good Medical Practice (2013) 0-18 years guidance: Appendix 2, 1.1, Parents and parental responsibility. Available at: http://www.gmc-uk.org/guidance/ethical_guidance/children_guidance_appendix_2.asp

42. Including Medical research involving children (Medical Research Council, 2004) Guidelines for the ethical conduct of medical research involving children (Royal College of Paediatrics and Child Health: Ethics Advisory Committee, 2000)

43. Health Research Authority Principles of Consent, Children and Young People (England, Wales and Northern Ireland) Childrens/Young people's wishes and assent: Available at: http://www.hra-decisiontools.org.uk/consent/principles-children-EngWalesNI.html

44. Case law suggests that if a young person has sufficient understanding and intelligence to understand fully what is proposed, and can use and weigh this information in reaching a decision (i.e. often called being 'Gillick competent'), he or she can give consent to treatment. See Gillick -v- West Norfolk AHA. 3 All Er 402, at 423–4.

45. Health Research Authority Principles of Consent, Children and Young People (England, Wales and Northern Ireland): Consent to Research http://www.hra-decisiontools.org.uk/consent/principles-children-EngWalesNI.html

46. Guidance on this in England and Wales available at http://www.hra-decisiontools.org.uk/consent/principles-emergency-EnglandandWales.html

47. Health Research Authority, Principles of consent: Emergency Research (England and Wales) Adults not able to consent for themselves in other emergency research, http://www.hra-decisiontools.org.uk/consent/principles-emergency-EnglandandWales.html#two

48. HRA definition of consultee: Under the Mental Capacity Act (enforced in England and Wales) before an adult who lacks capacity to give consent can be included in research, the researcher must take reasonable steps to identify someone to consult (a consultee), to determine if participation in research is appropriate. The consultee must be involved in the person's care, interested in their welfare and must be willing to help. They must not be a professional or paid care worker. They will probably be a family member, but could be another person. A person is not prevented from being a consultee if they are an attorney authorised under a registered Lasting Power of Attorney or are a deputy appointed by the Court of Protection; but that person must not be acting in a professional or paid capacity (for example, person's solicitor). http://www.hradecisiontools.org.uk/consent/glossary.html#C4

49. Mental Capacity Act Code of Practice, 2013 Available at: https://www.gov.uk/government/collections/mental-capacity-act-making-decisions

50. Health Research Authority, Principles of consent: Adults not able to consent for themselves (England and Wales), Participants regaining capacity during your study. Available at: http://www.hra-decisiontools.org.uk/consent/principles-ALC-EnglandandWales.html#three

51. See MRC Ethics Guide: Medical research involving children, Medical Research Council, 2004

52. Health Research Authority, Principles of consent: Emergency Research (England and Wales) children and young people in other emergency research. Available at: http://www.hra-decisiontools.org.uk/consent/principles-emergency-EnglandandWales.html#four

53. Health Research Authority, Principles of consent: Adults not able to consent for themselves (England and Wales) Participants regaining capacity during your study. Available at: http://www.hra-decisiontools.org.uk/consent/principles-ALC-EnglandandWales.html#three

54. HRA Principles of consent: Adults not able to consent for themselves (England and Wales), Available at: http://www.hra-decisiontools.org.uk/consent/principles-ALC-EnglandandWales.html#four Also: HRA Principles of consent: Adults not able to consent for themselves (England and Wales), Losing capacity during participation in research. Available at: http://www.hra-decisiontools.org.uk/consent/principles-ALC-EnglandandWales.html#four

55. HRA, Principles of Consent: Deceased People-consent and participant information sheet preparation guidance. Available at: http://www.hra-decisiontools.org.uk/consent/principles-deceased.html

56. See GMC Good Medical Practice 'Confidentiality guidance: Genetic and other shared information' paras 67-69. Available at: http://www.gmc-uk.org/guidance/ethical_guidance/confidentiality_67_69_genetic_and_other_shared_information.asp and GMC Good Medical Practice 'Confidentiality guidance: Disclosure after a patient's death', para 70 Available at: http://www.gmc-uk.org/guidance/ethical_guidance/confidentiality_70_72_disclosure_after_patient_death.asp , also 'Consent and confidentiality in genetic practice', (2006, Joint Committee on Genomics in Medicine)

57. See section 4.2.6 'Who can give qualifying consent for analysis of DNA in cellular tissue?' in 'Consent and confidentiality in genetic practice', (2006, Joint Committee on Genomics in Medicine), Available at: http://www.rcpath.org/committees/intercollegiate-and-joint-committees/joint-committee-on-genomics-in-medicine/consent-and-confidentiality-in-genetic-practice

58. See Human Tissue Authority Code of Practice (1) Consent, 'Consent requirements - Part 2: Tissue from the deceased'. Available at: http://www.hta.gov.uk/_db/_documents/Code_of_practice_1_-_Consent.pdf

59. See Human Tissue Authority Code of Practice (1) Consent, 'Consent requirements - Part 2: Tissue from the deceased', 'Qualifying relationships' Available at: http://www.hta.gov.uk/_db/_documents/Code_of_practice_1_-_Consent.pdf This states at para 92 that 'If the deceased person has not indicated their consent (or refusal) to post-mortem removal, storage or use of their body or tissue for scheduled purposes, or appointed a nominated representative, then the appropriate consent may be given by someone who was in a 'qualifying relationship' with the deceased person immediately before their death. Those in a qualifying relationship are found in the HT Act in the following order (highest first).

1. *spouse or partner (including civil or same sex partner) The HT Act states that, for these purposes, a person is another person's partner if the two of them (whether of different sexes or the same sex) live as partners in an enduring family relationship.*
2. *parent or child (in this context a child may be of any age and means a biological or adopted child)*
3. *brother or sister*
4. *grandparent or grandchild*
5. *niece or nephew*
6. *stepfather or stepmother*
7. *half-brother or half-sister*
8. *friend of long standing.*

It should be noted that the qualifying relatives for adults in Scotland is different and is set out in the Human Tissue (Scotland) Act.

60. See Health Research Authority: principles of consent: deceased people: http://www.hra-decisiontools.org.uk/consent/principles-deceased.html Also HTA Consent code of practice (v 14.0) 2014, Available at: http://www.hta.gov.uk/legislationpoliciesandcodesofpractice/codesofpractice/code1consent.cfm
61. Under the Human Tissue Act 2004 in England, Wales and Northern Ireland, a 'child' is defined as being under 18 years old. Under the Human Tissue (Scotland) Act 2006, a 'child' is defined as being under 16 years old.
62. This is because in strictly legal terms, the products of a pregnancy of less than 24 weeks gestation are considered to be the mother's tissue. (The law does not distinguish between fetal tissue and other tissue from the living; fetal tissue is regarded as the mother's tissue). For further details please see the Human Tissue Authority, Code of Practice (1) Consent, v 14.0 Updated July 2014, section on Fetal Tissue, paras 171- 175. Available at: http://www.hta.gov.uk/_db/_documents/Code_of_practice_1_-_Consent.pdf
63. See Human Tissue Authority Code of Practice (1) Consent, 'Consent requirements - Part 2: Tissue from the deceased', 'Qualifying relationships' Para 92, Available at: http://www.hta.gov.uk/_db/_documents/Code_of_practice_1_-_Consent.pdf
64. See section 2.4 'Use of Medical records to confirm diagnosis', Consent and confidentiality in genetic practice', Joint Committee on Genomics in Medicine, 2006. Available at: http://www.rcpath.org/committees/intercollegiate-and-joint-committees/joint-committee-on-genomics-in-medicine/consent-and-confidentiality-in-genetic-practice
65. Parental responsibility (2008): Guidance from the British Medical Association.
66. See Gillick -v- West Norfolk AHA. 3 All Er 402, at 423–4.
67. General Medical Council, Good Medical Practice (2013), 0-18 years Guidance: Research: point 38
68. Neena Modi *et al.* Guidance on clinical research involving infants, children and young people: an update for researchers and research ethics committees, Arch Dis Child, Online first, June 9 2014. Available at http://adc.bmj.com/content/early/2014/06/09/archdischild-2014-306444.full?sid=ec16ee7f-5925-4721-9e66-d038b7e5dba7 (accessed 6 October 2014)
69. General Medical Council, Good Medical Practice (2013), 0-18 years Guidance: If a young person refuses ztreatment. Available at http://www.gmc-uk.org/guidance/ethical_guidance/children_guidance_30_33_refuse_treatment.asp (accessed 6 October 2014)
70. General Medical Council, Good Medical Practice (2013), Consent Guidance, making decisions when a patient lacks capacity. Points 75 and 76. Available at http://www.gmc-uk.org/guidance/ethical_guidance/consent_guidance_making_decisions_patient_lacks_capacity.asp (accessed 6 October 2014)

Appendix 2 | List of Abbreviations

Abbreviation	Expansion
CPRD	Clinical Practice Research Datalink
EAC	(Genomics England) Ethics Advisory Committee
HEE	Health Education England
HRA	Health Research Authority
HSCIC	Health and Social Care Information Centre
NDPH	Nuffield Department of Population Health
NGS	Next Generation Sequencing
NHSE	National Health Service England
PHE	Public Health England
QA	Quality Assurance
QMUL	Queen Mary University, London
REC	Research Ethics Committee
WGS	whole genome sequencing
CLL	chronic lymphocytic leukemia
DPA	Data Protection Act
FOI	Freedom of Information Act
ICO	Information Commissioners Office
FF	Fresh Frozen
FFPE	Formalin-fixed, paraffin-embedded
SOP	Standard Operating Procedures

Appendix 3 | Biological samples standard operating procedures

Sample collection

- All samples should be taken specifically for this project, rather than using historic stored samples, except for rare occasions when agreed otherwise with Genomics England.
- Cancer patients should have their blood sample taken prior to biopsy or resection, and within an appropriate period of time to prevent changes due to treatment or blood transfusion.
- Patients with haematological cancers should have a peripheral blood sample taken along with a saliva sample.

All samples must be collected according to the following specification, using the Institution's Approved Standing Operating Procedures and performed by staff with appropriate training, and who receive regular competency checking in these procedures. Institutional safety guidelines for handling of human biological materials and hazardous chemical should also be followed.

Peripheral blood samples are collected for:

- DNA for whole genome sequencing (and storing for potential future epigenetic or other analysis).
- Stabilised blood for future RNA extraction.
- Serum and plasma for future proteomic and metabolomics studies.
- Cell free plasma for future analysis of circulating tumour DNA.

Two samples of blood are taken for DNA extraction, one to be stored at the Genomic Medicine Centre for local testing and validation, and one to be entered into the project and sent to the Biorepository along with the other samples using the centrally procured coordinated transport system.

Sample volumes

Cancer

	DNA (EDTA - purple)	**Plasma** (PST - green)	**Serum** (SST - gold)	**RNA** (Paxgene)	**Saliva** (Oragene)
Adult (14yrs+)	4.5ml x 2 10ml x 1	10ml	10ml	5ml	–
<14yrs	Specific guidance will be given for childhood cancers, please contact Genomics England before enrolling any adolescents or children for other cancer types				
CLL patients	As above	As above	As above	As above	2ml - Follow kit instructions

Rare and Infectious Diseases

	DNA (EDTA - purple)	**Plasma** (PST - green)	**Serum** (SST - gold)	**RNA** (Paxgene)	**Saliva** (Oragene)
Adult (14yrs+)	4.5ml x 2	10ml	10ml	5ml	–
3-14 years	3ml x 2	3ml	3ml	2.5ml	–
0-3 years*	1-3ml	1 ml	1 ml	1ml	–

*In neonates and acutely ill children, clinical discretion should be applied to the volume of blood able to be extracted without risk of adverse events. Where only small volumes of blood are obtained, workflows should be altered to allow 'rescue' of plasma prior to automated DNA extraction.

Prioritisation of samples

All efforts should be made to collect all indicated samples at the same time. If it is not possible to collect all samples, then the priority is to collect blood only for DNA extraction, followed by plasma, RNA (Paxgene) and then serum. If handling small blood samples, then 'rescue' of plasma prior to automated DNA extraction should be performed whenever possible.

Sample labelling

A unique identifier will be produced at the time of consent that should be used, along with NHS number, as the main identifiers within the clinical sites. A minimal sample collection form will be provided for completion of details at the time of collection. The majority of information will be electronically recorded in a tool provided by Genomics England. This will be done by interfacing with existing systems for ease of data transfer. The best methods for identification and labelling are currently being assessed by Genomics England and detailed information on what will be required will be provided at a later date in further guidance.

Sample transport to Genomic Medicine Centres

All samples must be placed in approved standard specimen bags with a collection form to be supplied and transported to the Genomic Medicine Centre processing laboratory(ies) for receipt within 24 hours of sample collection. RNA stabilised blood, and samples for plasma and serum for 'omics' studies, will degrade if they are not processed as soon as possible. Therefore, optimisation of transport logistics to the processing laboratory(ies) within a GMC from all sites of collection should be a priority.

Sample processing

Logging samples

All samples should be entered onto a LIMS and processed as soon as they reach the laboratory. All LIMS system must be able to:

- Create unique barcodes/identifiers based on guidance from Genomics England.
- For each patient codes need to be created for DNA, RNA, serum and plasma samples.
- Track samples at all stages from collection, through processing to dispatch.
- Comply with Information Governance Legislation.
- Export information to provide required data to Genomics England/NHS England.

Information will be provided in further guidance on appropriate tubes and identification to allow ease of handling and tracking when sent on to the Central Biorepository.

Sample processing (see flow diagram below for overview)

- DNA extraction.
- RNA stabilised samples should be aliquoted, and then immediately frozen and stored at -80°C until transported to the Biorepository.
- Serum and plasma samples for proteomics and metabolomics should be separated (centrifugation at 1300-2000 G for 10 minutes) immediately frozen and stored at -80°C until transported to the Biorepository.
- Samples for circulating tumour DNA should be centrifuged at 1300-2000 G for 10 minutes at 4°C. The plasma supernatant should then be immediately transferred in 1ml aliquots to 2ml microtubes and centrifuged on a bench top centrifuge at 14,000rpm for 10 minutes. The plasma supernatant should

then be transferred to new tubes, immediately frozen, and stored at -80ºC until transported to the Biorepository. The buffy coat should also be collected in a separate microtube and frozen at -80ºC until transported to the Biorepository.

- Samples should be aliquoted into a barcoded Universal tube, or other tube as specified in future guidance.

Sample transport to Central Biorepository

- DNA should be transported at room temperature to the Central Biorepository by the agreed transport system.
- All plasma, serum and RNA stabilised blood aliquots should be transported in batches on dry ice at the appropriate temperature, stabilising the packaging by the agreed transport system, at intervals to be advised in further guidance.

Tumour sample collection and processing

Whole genome sequencing success rates from FFPE tumours to date are relatively low, but variable, with relatively limited experience in a number of centres. Evidence and clear best practice are lacking in a number of areas. We will work with centres, along with Illumina, to improve success rates through optimising procesures. It is important centres collaborate with this and follow additonal guidance and protocols produced, as best practice methods are agreed.

When a tumour is received in the laboratory, the following are essential for entry into the project:

- Tumours should have been biopsied or resected prior to treatment, apart from where normal clinical practice is to treat prior to resection or if the sample is from a recurrence (where the original tumour has previously been sequenced).
- A germline blood sample should have been collected prior to resection or biopsy.
- Samples should be placed in formalin for no more than 24 hours before processing to avoid fragmentation and transferred to the laboratory as soon as possible.
- Details of the pre-analytical handling of tissues should be recorded locally in a format that can be uploaded directly into the provided Genomics England tool.

Sample processing by the Genomic Medicine Centre

All samples should be entered onto a LIMS and processed as soon as they reach the laboratory. All LIMS system must be able to:

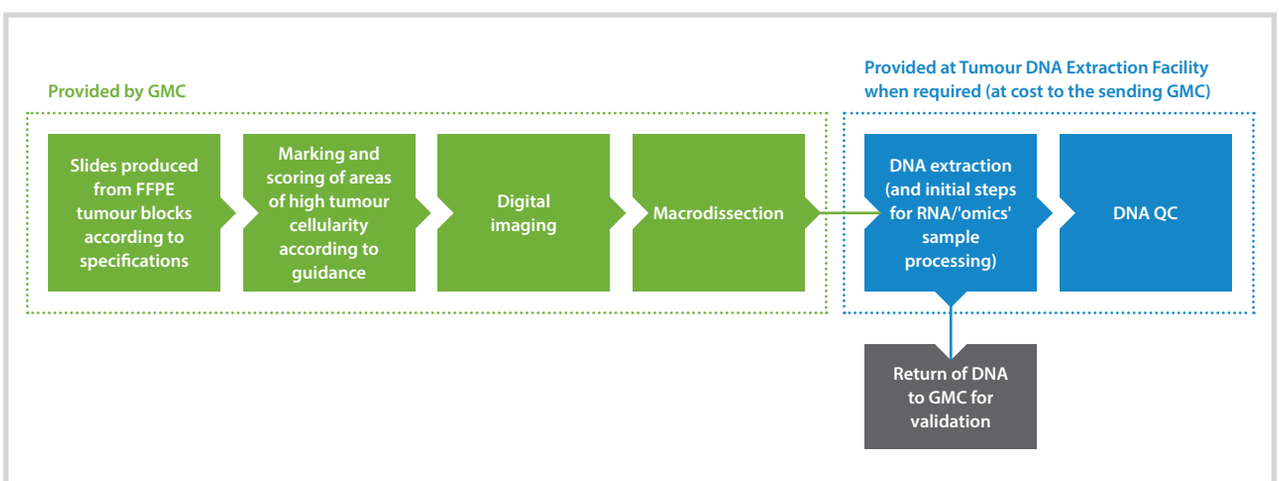
- FFPE samples will only have a single block reference, which should be linked to the patient IDs.
- Create unique barcodes/identifiers according to Genomics England guidance.
- Track samples at all stages from collection, through processing to dispatch.
- Comply with information Governance Legislation.
- Export information to provide required data to Genomics England/NHS England.

Tissue samples should be processed as soon as they reach the laboratory and the following steps undertaken:

- Preparation of slides from FFPE preferably 5-10 um thickness.
- Hematoxylin and eosin (H&E) staining should be used to identifiy and estimate the percentage of neoplastic cells (scored according to illustrated guidance to be provided by Genomics England).
- Areas of high neoplastic content should be marked by a fully trained and competent practitioner in histological assessment for molecular pathology.
- Macrodissection of the marked region of neoplastic tissue should be carried by a trained technologist.
- A digital image from high throughput digital equipment or images of a similar quality from slides showing tumour site selected, should be taken and transferred with the DNA sample to the Biorepository.

Where centres do not achieve defined quality standards, or elect not to perform these steps, DNA extraction and quality control will be performed in a central tumour extraction facility established by Genomics England (this may be one or more GMCs arranged through a separate process):

- A total of at least 60 microns should be processed or transferred to a centrally provided extraction facility for DNA extraction. It is the total volume of tissue that will determine the amount of DNA extracted. This amount should be iterated based on accumulated experience within a site for a particular tumour type to ensure sufficient tumour is provided for DNA extraction.
- DNA requirements and QC processes are described in Annex X.
- Initial stages of processing for RNA extraction, deparaffinisation cell lysis and suspension in RNA stabilising buffer, should be performed as part of the same process as DNA extraction following SOPs to be provided. Due to the need for all processing of the tumour block or biopsy to be done in an RNA free environment, further guidance will be issued on this in the early stages of the programme. Centres should plan towards this being required within 2015 but not at the outset of the programme.
- FFPE tumour scrolls or scrapings of high cellularity not required for DNA/RNA extraction should be forwarded to the Central Biorepository for future 'omics' studies.
- Centres should retain sufficient DNA for any standard testing and validation.



Local storage and Central Biorepository requirements

Local Storage

- DNA should be long-term stored according to current laboratory practices.
- Tumour blocks, slides and other tissue should be stored according to current laboratory practices.
- Temporary storage of other samples should be as detailed in specifications, typically at -80°C.

Central Biorepository

- Genomics England are currently procuring a Central Biorepository to be active through the main programme, commencing by the end of Q2 2015 for all centres. Interim biorepositories within the NHS are being agreed to cover the period until this commences.
- A subset of data will need to be provided to the Biorepository through an interface provided by Genomics England, along with the samples specified through a centrally procured transport service.
- Further guidance on specific labelling and sample containers to ensure smooth processing within centres, through transporting, and then into the Biorepository will be released as soon as possible within this process.

Diagram 1. Cancer (Adult) - Blood Sample Pathway

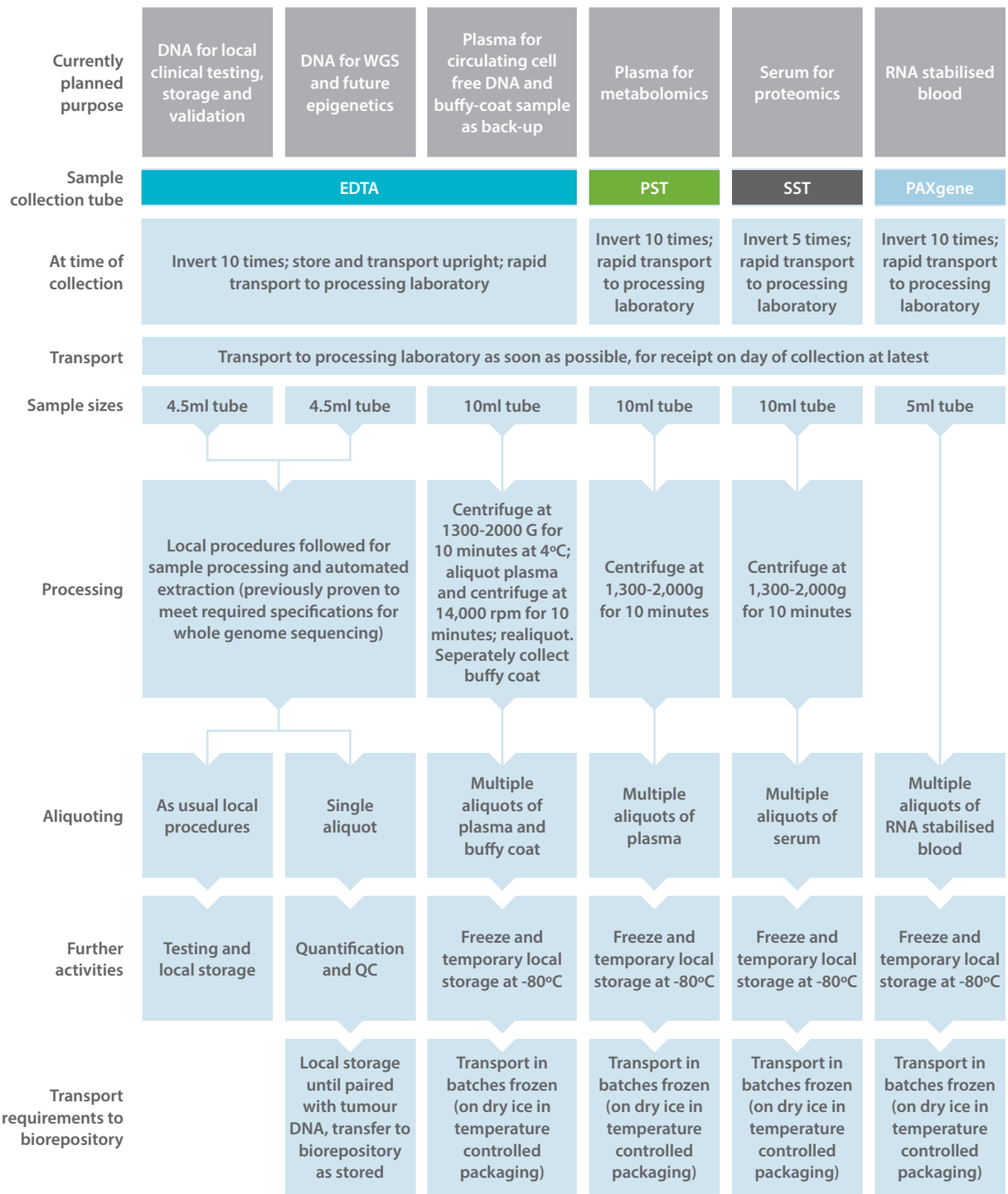
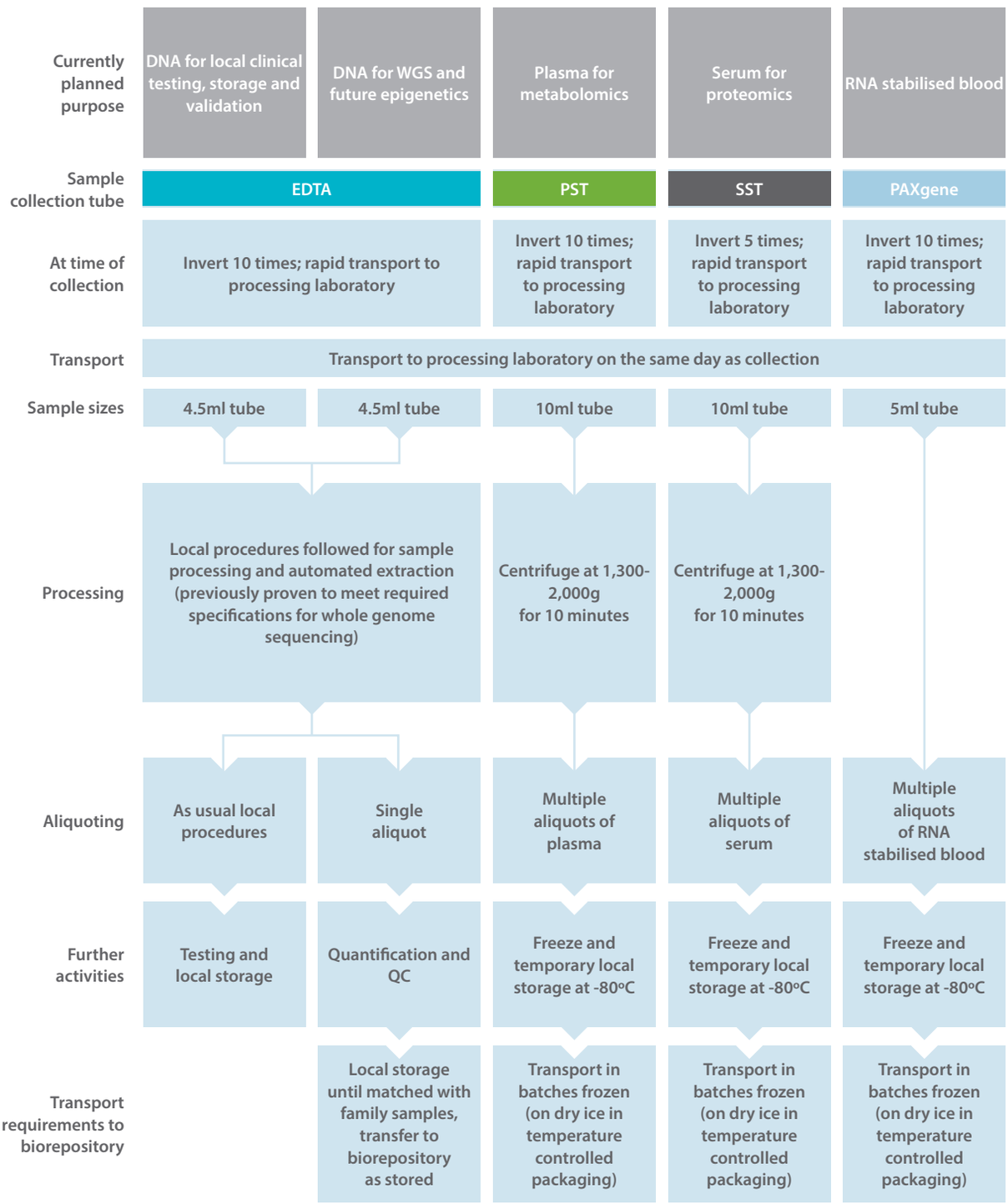
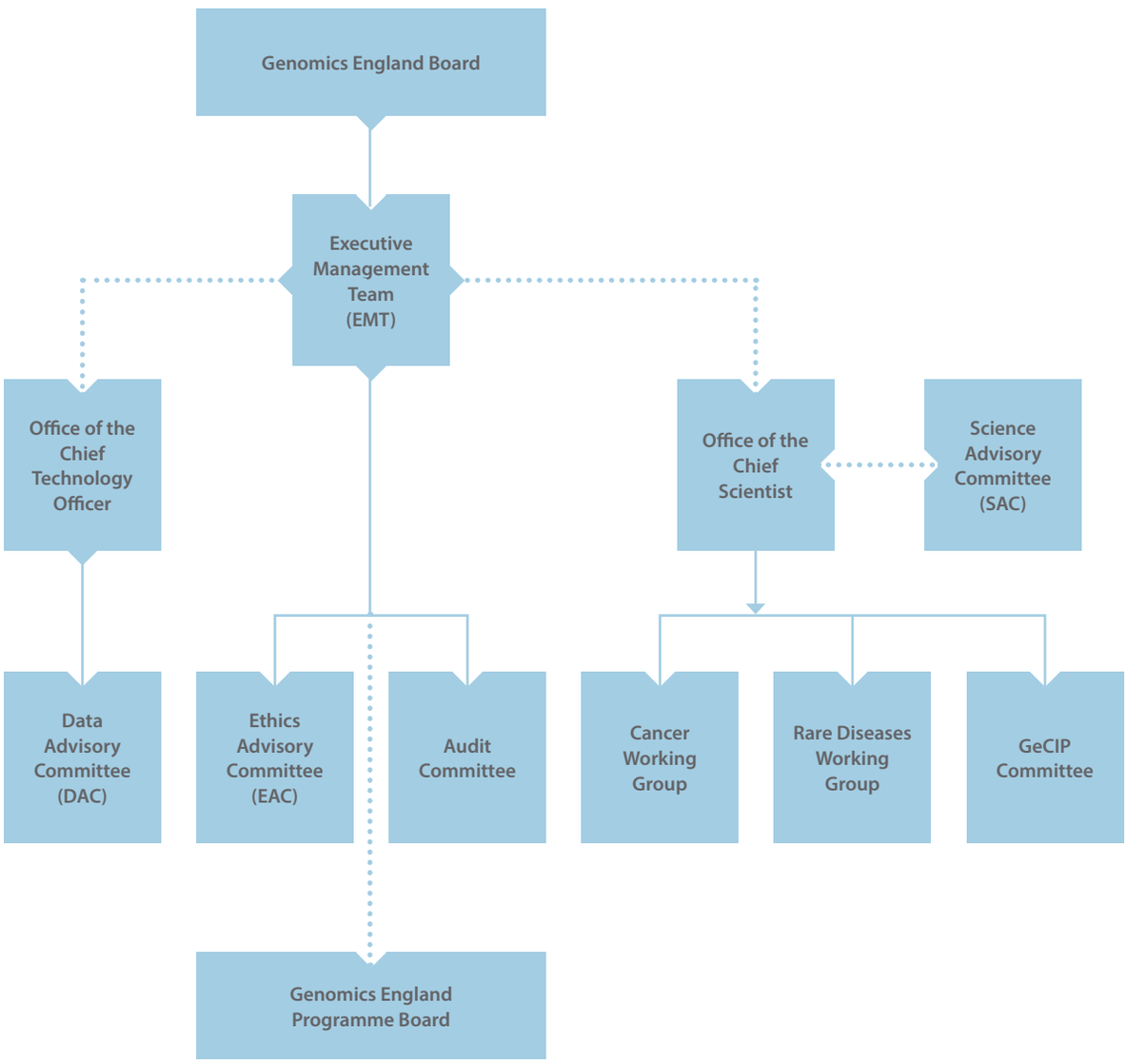


Diagram 2. Rare Diseases (Adult) - Blood Sample Pathway



Appendix 4 | Genomics England Governance and Committees



Version 0.2
Updated 30th September 2014

Appendix 5 | Genomics England Disease List Nomination Process

Please refer to **www.genomicsengland.co.uk** for further details.

Appendix 6 | Letter of Insurance

Department of Health

Richmond House
79 Whitehall
London SW1A 2NS

15th October 2014

Professor Mark Caulfield,
Chief Scientist,
Genomics England Ltd
Charterhouse Square
London

Dear Professor Caulfield

100,000 Genomes Project – disclosure of genome results for insurance purposes

I am writing in my capacity as the genomics policy lead in the Department of Health to clarify the current policy regarding the use of genetic and genomic information by insurance companies. I am aware that this is potentially a concern for participants in the 100,000 Genomes Project, especially given the novel nature of the project and the long-term storage, analysis and feedback that is envisioned for the project.

The use of genetic test information by insurers is the subject of an agreement between the Government and the Association of British Insurers (ABI). It is set out in the Concordat and Moratorium on Genetics and Insurance last updated and published in 2011. For the purposes of this project the key feature is that the insurers will not require the disclosure of:

"a predictive or diagnostic test result acquired as part of clinical research. To avoid doubt, customers may be asked to disclose details of any symptoms, diagnosis or treatment received outside of the clinical research programme, even if those relate to a condition they found out about through the research programme."

We have agreement from the ABI that the wording used in the Patient Information Sheet is in line with their interpretation of the current policy. The ABI are supportive of the aims of clinical research generally and of the 100,000 Genomes Project particularly. I am sure that they are willing to work with the Department and Genomics England to give specific guidance given the high profile nature of this project. The initial patient engagement suggests that further detailed Q&A for patients, clinicians and insurance companies may be necessary around the interpretation of the Concordat and Moratorium in relation to the feedback of secondary findings.

The key issues that the agreement covers are the uses of results from clinical genetic tests for highly penetrant late onset disorders. At present, the results of such tests do not need to be disclosed unless they are approved by an independent panel and are for policies over the financial threshold (£500,000 for life insurance). The only approved test is for Huntington's disease for Life Insurance. Even if the insurance companies were minded to make applications for the other conditions that may be returned as pertinent or incidental findings the timescale is likely to be many months which would give ample notice of a possible need for an amendment of the Patient Information Sheet.

Insurers do expect to be told about symptom, diagnoses or treatment if they request information at the application stage or via a medical report. They may ask for information about family history but many will be willing to consider favourable genetic test results even if they are not approved by the independent panel.

Turning to the longer term, the currently published Concordat and Moratorium states that the agreement will be reviewed in 2014 and will expire in 2017 if not renewed. I am pleased to advise you that we have reached agreement with the ABI that the Concordat and Moratorium will be extended until 2019 with a review in 2016. We are in the process of finalising the announcement and intend for this to be made public in the autumn and before recruitment begins via the NHS Genomic Medicine Centres.

Looking beyond 2019 is difficult as the 2016 review will need to be conducted in light of the priorities of the Government of the day. However, it has been a long-standing position in successive administrations that people should not be deterred from accessing healthcare or research opportunities because of concerns about insurance. In the most recent major review of discrimination legislation, in 2008 the Government decided not to not to extend protection against discrimination on the ground of genetic predisposition because "the existing arrangements for a voluntary moratorium on insurers' use of predictive genetic test results … along with continued monitoring of the use of genetic testing in the UK should provide sufficient reassurance".

It has also been the long-standing policy to ensure that the timing of the periodic reviews of the current position always allow at least 3 years before the end of the Concordat and Moratorium in order that the Government has the opportunity to consider alternatives if either party does not wish to continue with the current voluntary agreements.

I hope that this reassures you that the Department is alert to the concerns about the long-term position on genetics and insurance and that we will work with Genomics England and the ABI to ensure that participants concerns are addressed. In doing this, I would remind you that the original policy announcement establishing the 100,000 Genomes Project stressed the importance of ensuring public trust and confidence and said: "To ensure public confidence in matters of confidentiality and access, this work will be monitored by the Chief Medical Officer for England." As Dame Sally Davies is a Director of Genomics England we will be considering further how she should fulfil this role and how she can address any specific concerns about insurance or other uses of the data in her role advising the Government on medical and scientific matters.

I am copying this letter to Dame Sally Davies and to Felicity Harvey who is the Senior Responsible Officer for the 100,000 Genomes Project.

Please do not hesitate to contact me if I can be of further assistance.

Yours sincerely

Dr Mark Bale
Deputy Director: Genomics, Science and Emerging Technologies
Health Science & Bioethics Division

Genomics england

Contact

Genomics England

Queen Mary University of London

Dawson Hall, Charterhouse Sq.

London EC1M 6BQ

T: 0207 882 6293

E: info@genomicsengland.co.uk

W: www.genomicsengland.co.uk