GeCIP Detailed Research Plan Form

Genomics England Clinical Interpretation Partnership (GeCIP) Detailed Research Plan Form

Application Summary	
GeCIP domain name	**Paediatric GECIP**
Project title *(max 150 characters)*	**Clinical Interpretation of Genomics England 100,000 Genomes in Children with Rare Diseases**

Objectives. *Set out the key objectives of your research. (max 200 words)*

1. To identify genetic variations that are responsible for recognisable disease phenotypes within our disease areas of fetal life, growth and endocrine disorders, imprinting disorders, multisystem ciliopathy disorders, infancy and childhood liver disorders, developmental disorders and orphan diseases.
2. Where the volume of genomic data permits, to undertake meta-analyses, e.g. of neurodevelopmental disorders in collaboration with the Neurology GeCIP domain, including analysis of shared genetic aetiology across a range of frequently co-morbid disorders, such as intellectual disability (ID), autism and epilepsy.
3. To explore the range of functional consequences of known or novel genetic variants through a functional discovery pipeline, to evaluate their likelihood as disease causing. This includes gene expression and proteomics studies, cell biological localisation, cell models, and animal models.
4. To build on our previous work to evaluate the utility of incorporating metabolomics, epigenomics and transcriptomics data with WGS data for identifying diagnostic variants.

Lay summary. *Information from this summary may be displayed on a public facing website. Provide a brief lay summary of your planned research. (max 200 words)*

This domain focus is on identifying genetic causes of fetal, early life, and childhood abnormalities. These include birth defects, developmental disorders, growth problems, and disorders affecting several body systems that present in childhood. The aim of this research plan is to bring together existing genetic information, with new information from the 100,000 Genomes project, to find new genetic causes of disease, genetic changes that modify risks for disease, and to identify the mechanisms by which these genetic changes have their effects. We will develop more accurate terms to describe the features of rare diseases. We anticipate that the outcomes from this research plan will be to identify new treatment targets for development in collaboration with academic and commercial partners; to gain more in-depth understanding of disease mechanisms; and to support the training of the next generation of NHS workers in genomic medicine and rare diseases. Most of all we believe this plan will benefit children with conditions that do not yet have a known genetic cause.

Technical summary. *Information from this summary may be displayed on a public facing website. Please include plans for methodology, including experimental design and expected outputs of the research. (max 500 words)*

Our focus will be to exploit the unique opportunities of the datasets to undertake detailed analysis of the genomic architecture of rare diseases. This will include detailed exploration of rare gene variants, combined with extended phenotyping of affected patients to understand how variants in the same gene result in a spectrum of phenotypes.

Data integration:

In addition to the ground breaking analyses that the 100,000 Genomes Project samples will allow, we will also be able to integrate with other valuable sample sets to power analyses that would not

be feasible on any dataset alone. We will harmonise and integrate relevant Genomics England data with existing and already funded genome-wide genetic datasets (e.g. WES/WGS) generated by the sub-domain members on over 35,000 samples from families with Developmental and Orphan Disorders (DODs): DDD, NIHR-SPEED, UK10K, PAGE, Institute of Child Health Neurogenetics Research Group and project-specific datasets.

Data analysis:
We will structure our analyses of these integrated datasets into sub-groups in order to support: gene discovery in defined phenotypic subsets; genotype-phenotype analyses of patients with pathogenic variants in the same gene, or pathway; deep phenotyping of patients (e.g. in similar manner to MRC IMAGINE project); assessment of polygenic contributions to the disorders in our domain; and development of novel analytical methods. These analyses will be carried out in collaboration with disease-facing GeCIPs, including neurology, endocrinology and metabolism, and ear and eye, and cross-cutting GeCIP domains such as the functional cross-cutting and functional effects domains.

Functional studies:
Selected variants will enter a functional discovery pipeline, to evaluate their likelihood as disease causing. We have cell- and organism-based cell models established for evaluation of human gene variants and antibody, siRNA and mutant construct resources to facilitate effective and rapid throughput variant analysis, as well as strong collaborative links with other labs. The typical pipeline will be:
a. *Gene expression studies* in existing collections of patient biosamples, together with in situ studies in mouse and human embryos through the Human Developmental Biology Resource (HDBR); Western blot, qPCR and immunofluoresence confocal microscopy imaging.
b. *Cell biological* localisation and mutational analysis to explore new disease candidates for deficits in cell biological functions.
c. *Cell models*. Cell lines have already been established for some disorders e.g. ciliopathy, imprinting and growth disorders. CRISPR based mutagenesis or siRNA antisense gene depletion will model a simple gene disruption or different prioritised VUS (variants of uncertain significance).
d. *Model organisms*. Human diseases can be modelled in several organisms that display relevant phenotypes, e.g. zebrafish, in which siRNA and CRISPR methods are well established at several of our partner institutions.

Modelling diseases and their modifiers
We will generate biologically accurate, human disease models by human mutation engineering in human embryonic stem cells (hESC), derivation of patient induced pluripotent stem cells (iPSC) and of patient-relevant cell models, e.g. human neuronal and retinal cells to model ciliopathy. We would like to proceed to use these resources to evaluate the effects of novel mutations and genetic modifiers by CRISPR, to trial new therapies directed at specific classes of mutations, and to employ novel high content small molecule screening for identifying therapeutic entry points.

Multiomics studies:
There are existing collections of fibroblasts, serum, urine and other materials, including selected collections of patient iPSCs, for instance with ciliopathy patients, for multiomic studies including RNASeq and proteomics. Through GEL, we will integrate the genomes of individuals with their transcriptomes (blood and circulating microvesicle, iPSC), proteomes (blood, iPSC), lipidomes (plasma) and clinical phenotypes. Combining data, to integrate genomic variation with DNA methylation, gene expression, lipid and protein profiles and resultant phenotype will provide unique insights into variations in pathophysiology for imprinting and other disorders within our

domain. With additional funds available to some group members, and external funding for others, we will extend the generation of transcriptomic, metabolomic and epigenomic profiles on patients and their samples that have been sequenced by Genomics England.

Expected start date	**July 2018**
Expected end date	**December 2020**

Lead Applicant(s)	
Name	Timothy Barrett; Philip Beales
Post	Prof of Paediatrics; Prof of Genetics
Department	Institute of Cancer and Genomic Sciences; Experimental and Personalised Medicine
Institution	University of Birmingham/ University College London
Current commercial links	

Administrative Support	
Name	Rachel Bruton, Ruth Jamieson
Email	r.k.bruton@bham.ac.uk; r.jamieson@ucl.ac.uk
Telephone	0121 414 8504/8557; 02076796165

Subdomain leads		
Name	**Subdomain**	**Institution**
Mark Kilby	Fetal Medicine	Kings College London
Peter Clayton	Growth and Endocrine	Manchester University
Karen Temple	Imprinting Disorders	Southampton University
Hannah Mitchison	Ciliopathy Disorders	University College London
Matthew Hurles	Developmental and Orphan Disorders	University of Cambridge
Richard Thompson	Hepatology	Kings College London
Network support		
Bioinformatics	Jean-Baptiste Cazier	Hywel Williams
Ethics	Jonathan Ives	Angus Clarke
Functional Biology	Dagan Jenkins	Connie Bonnifer
Health Economics	Sue Jowett	Steve Morris
Public engagement	Tess Harris	Alastair Kent
Training	Tim Barrett	Sarah Ennis
Variant validation	Mike Griffiths	Lucy Jenkins
Bioinformatics	Jean-Baptiste Cazier	Hywel Williams

Detailed research plan

Full proposal (total max 1500 words per subdomain)	
Title *(max 150 characters)*	**Genetic causes of fetal abnormalities**
Importance. *Explain the need for research in this area, and the rationale for the research planned. Give sufficient details of other past and current research to show that the aims are scientifically justified. Please refer to the 100,000 Genomes Project acceptable use(s) that apply to the proposal (page 6).*	

The major focus of the Fetal Medicine sub-domain is the study of the genetic causes of fetal structural abnormalities (e.g. central nervous system anomalies, fetal hydrops, skeletal dysplasia, renal anomalies, cardiac anomalies) and severe early onset fetal growth restriction, in the absence of major fetal aneuploidies.

Prenatal diagnosis plays an essential role in contemporary obstetric care. Standard chromosome testing has been available since the 1960s and has been commonly used in the prenatal genetic testing when a fetus has abnormal findings on ultrasound. Quantitative fluorescence PCR and traditional karyotype have been the main means to detect fetal aneuploidies and other relatively larger translocations, deletions or duplications. However, these methods cannot identify microscopic and submicroscopic genomic imbalances. In recent years, there has been an introduction of genomic microarrays that provide a genome wide screen for genomic imbalances at a high resolution, allowing for the detection of microdeletion and microduplication syndromes. This is now the method of prenatal diagnosis when there are fetal structural abnormalities on an ultrasound scan as there is an additional diagnostic yield of up to 9% as compared to the traditional karyotype [1].

Major advances in DNA sequencing technologies and bioinformatics make it possible to do large-scale sequencing for diagnostic purposes as opposed to the traditional approach of sequencing individual gene(s) responsible for a given phenotype. There is however limited research in evaluating DNA sequencing technologies in prenatal diagnosis. Currently, there is a nationwide study evaluating whole exome sequencing (WES) in detecting genetic abnormalities in cases with fetal structural abnormalities (PAGE, PI: Hurles; co-applicant Kilby). Initial results suggest that the WES can provide an additional diagnostic yield of 10-20% [3,4]. However, as a large proportion of cases with fetal structural abnormalities remain unaccounted for by known genes, there is therefore a need to gain experience with and utilise whole genome sequencing (WGS) for the diagnosis of underlying pathological genetic variants in cases of fetal abnormalities. To date, there is only one case report on the utilisation of WGS in prenatal diagnosis [5]. WGS has the potential to provide essential information for expectant parents to make an informed decision with regards to how best to manage the pregnancy, i.e. termination or continuation of pregnancy, as well as to plan for future pregnancies. We therefore propose to focus on fetal samples from pregnancies with varied fetal abnormalities.

Research plans. *Give details of the analyses and experimental approaches, study designs and techniques that will be used and timelines for your analysis. Describe the major challenges of the research and the steps required to mitigate these.*

The samples available to deliver this subdomain research plan are:

- Post-mortem fetal samples collected through the 100,000 genomes project
- Post-mortem fetal samples in existing collections collected with data sharing consent and to appropriate QC standards
- Prenatal samples (chorionic villi, amniotic fluid and fetal blood) from ongoing pregnancies in collections such as the PAGE project (see below), with data sharing consent and with the appropriate QC standards
- Cord blood samples where available as part of existing collections.

Note the PAGE project consent process includes data sharing, consent for WGS, but not for the same additional findings as in the 100,000 Genomes project. However this should not hinder the main focus of the research which is to increase the diagnostic yield in fetal anomalies

Harmonisation of datasets:
Harmonisation of datasets will proceed on an analysis-by-analysis approach. To address specific questions, different combinations of datasets will be required, and integration of data may occur at different levels. For example, for a meta-analysis of coding de novo mutations in fetal disorders we will want to combine datasets which have exome or genome data available for parent-offspring trios (e.g. GEL + PAGE + other datasets), and these data could be integrated at the level of high confidence de novo mutations, rather than entire VCF files. In this scenario, harmonisation will be needed at the level of variant detection (i.e. ensure similar sensitivity and specificity of detection of de novo mutations across datasets) and variant annotation (i.e. software and gene annotation resource). How the data are subsequently integrated for statistical testing will depend to some degree on the precise constraints on data ingress and egress from the GEL data centre. For case/control analyses of probands, a broader set of cohorts can be analysed, and similar harmonisation of variant annotation will be required but the harmonisation of variant detection will be different, and the level at which this harmonisation occurs will depend on the results of initial analyses assessing the comparability of the existing VCF files across studies.

Objective 1
To identify new pathological genetic variants in fetus collections with:

- Severe early onset growth restriction defined as estimated fetal weight <5th percentile diagnosed between 22^{+0} and 28^{+0} weeks' gestation.
- Congenital heart diseases – tetralogy of Fallot, hypoplastic left heart syndrome, pulmonary atresia, transposition of the great vessels, left ventricular outflow tract obstruction disorders and isomerism.
- Fetal hydrops.
- Central nervous system anomalies – migration neuronal defects, severe ventriculomegaly, posterior fossa abnormalities, midline brain abnormalities, microcephaly.
- Ophthalmological disorders – ano/microphthalmia, cataracts.
- Renal disorders - bladder exstrophy, cystic kidney disease, reflux nephropathy.
- Skeletal disorders - craniosynostosis syndromes, skeletal dysplasias.

We will rapidly develop eligibility criteria based on criteria developed for existing projects (PAGE, see below). This will allow a short period for collection of samples from across the Genome Centres. To provide assurance that there will be enough samples to generate meaningful conclusions, we will access the deceased fetal samples from our existing collections. Appropriate ethical consent for WGS and data sharing prospectively, is already in place for these. In addition, the HICF/Wellcome Trust funded 'Prenatal Assessment of Genomes and Exome (PAGE) study has 1,000 trios collected on ongoing pregnancies with varied /heterogenous fetal anomalies. We will perform WGS to compare with our existing data from WES. Finally, in a highly selected sub-cohort of n=50 pregnancies with associated mortality we have both USS findings and full post mortem findings, so that WES and WGS will be performed and compared.

Sequenced data of fetal materials from deceased fetuses with structural abnormalities would be expected to yield specific diagnoses (many of which will be novel) and variants that can be evaluated for their impact on phenotype.

With regard to fetal growth restriction, the sequenced data will be the first step in exploring the feasibility of developing targeted preventative measures or therapies in optimising fetal growth. We also aim to perform integrative *omic* analysis including mutation, copy number variant, DNA methylation, gene expression and microRNA, in the fetal growth restriction cohort. This approach

has the ability to significantly expand our understanding of fetal growth restriction and helps define molecular subtype, develops useful biomarkers to improve detection of high-risk pregnancies, disease management and clinical follow-up of patients.

Objective 2:

A. To review and optimise the process of DNA extraction from prenatal and post-mortem fetal materials to enable culture of fibroblasts that will yield DNA of adequate quality and quantity.
   - Prenatal samples - We anticipate stored fetal DNA will vary in quality and quantity between centres. The variability is most likely due to variability in the fetal samples taken and the types of testing requested. It is likely that more DNA will be needed to deliver WGS than is stored for some prenatal tests. Currently, the laboratories do not routinely evaluate DNA quality and quantify the amount of DNA stored so such data are not available. The first step is to carry out an audit of fetal DNA samples stored across our collaborating centres.
   - Post-mortem samples - DNA extraction from fetal organs. This source of DNA has the potential to yield sufficient DNA of high quality and is not affected by maternal cell contamination like fetal samples taken during pregnancy. However, the quality and quantity may vary according to the gestational age of the fetal samples. Fetal organ extraction is split into two parts - the maceration of tissue, which prepares it for DNA extraction and then the extraction of the DNA from the broken down tissue. The processes for breaking down tissue samples and DNA extraction will be evaluated.
B. This sub-domain will establish a framework in which the processes of (i) recruitment, (ii) consenting for storage of fetal samples, cord blood collection, post-mortem examination and WGS, and (iii) acquisition of fetal material with paired parental blood samples will be streamlined.
   This objective can be achieved by utilising the experience with the nationwide study – PAGE.

Collaborations including with other GeCIPs. *Outline your major planned academic, healthcare, patient and industrial collaborations. This should include collaborations and data sharing with other GeCIPs. Please attach letters of support.*

Renal GeCIP (Daniel Gale)
Neurological (Henry Houlden)
Musculoskeletal (Muhammad Kassim Javaid)
Cardiovascular (Bernard Keavney)
Endocrinology and metabolism GeCIP (Stephen O'Rahilly)

Training. *Describe the planned involvement of trainees in the research and any specific training that will form part of your plan.*

The fetal medicine unit at each GMC has extensive experience of supervising both clinicians and scientists doing postgraduate degrees related to fetal medicine and perinatal genetics.

Harris Birthright Research Centre for Fetal Medicine, King's College Hospital, and the Fetal Medicine Foundation (UK Charity), of which Prof Nicolaides is the Director, provide an excellent training programme for trainees in fetal medicine and perinatal genetics.

People and track record. *Explain why the group is well qualified to do this research, how the investigators would work together.*

Lead and co-lead have extensive experience in carrying out large-scale multicentre research studies involving pregnant women. Each GMC is represented by a fetal medicine specialist and a perinatal geneticist. Most of them are currently involved in PAGE and have experience in collaborating in a nationwide study that is similar to the 100K Genome Project.

Clinical interpretation. *(Where relevant to your GeCIP) Describe your plans to ensure patient benefit through clinical interpretation relevant to your domain. This should specifically address variant interpretation and feedback and your interaction with the cross-cutting Validation and Feedback domain.*

There will be potential benefit in identifying heritable conditions for families who may be considering future pregnancies.

Beneficiaries. *How will the research benefit patients and healthcare institutions including the NHS, other researchers in the field? Are there other likely beneficiaries?*
WGS has the potential to provide essential information for expected parents to make an informed decision with regards to how best to manage the pregnancy, i.e. termination or continuation of pregnancy, as well as to plan for future pregnancies. The work of this domain will be instrumental in identifying pathogenic variants, and in establishing diagnostic protocols.

Commercial exploitation. *(Where relevant to your GeCIP) Genomics England has a very explicit intellectual property policy. We and other funders need to know if the proposed research likely to generate commercially exploitable results. Do you have commercial partners in place?*

References. *Provide key references related to the research you set out.*

1. de Wit MC, Srebniak MI, Govaerts LC, Van Opstal D, Galjaard RJ, Go AT. Additional value of prenatal genomic array testing in fetuses with isolated structural ultrasound abnormalities and a normal karyotype: a systematic review of the literature. Ultrasound Obstet Gynecol. 2014 Feb;43(2):139-46.
2. Carss KJ, Hillman SC, Parthiban V, McMullan DJ, Maher ER, Kilby MD, Hurles ME. Exome sequencing improves genetic diagnosis of structural fetal abnormalities revealed by ult rasound. Hum Mol Genet. 2014 Jun 15;23(12):3269-77.
3. Drury S, Williams H, Trump N, Boustred C, GOSGene, Lench N, Scott RH, Chitty LS. Exome sequencing for prenatal diagnosis of fetuses with sonographic abnormalities. Prenat Diagn. 2015 Oct;35(10):1010-7.
4. Talkowski ME, Ordulu Z, Pillalamarri V, Benson CB, Blumenthal I, Connolly S, Hanscom C, Hussain N, Pereira S, Picker J, Rosenfeld JA, Shaffer LG, Wilkins-Haug LE, Gusella JF, Morton CC. Clinical diagnosis by whole-genome sequencing of a prenatal sample. N Engl J Med. 2012 Dec 6;367(23):2226-32.

Full proposal (total max 1500 words per subdomain)

Title *(max 150 characters)*	**Understanding genetic mechanisms that lead to growth and endocrine disorders**

Importance. *Explain the need for research in this area, and the rationale for the research planned. Give sufficient details of other past and current research to show that the aims are scientifically justified. Please refer to the 100,000 Genomes Project acceptable use(s) that apply to the proposal (page 6).*

The diagnosis of many of the growth and endocrine disorders included in the 100k GP is based on clinical features and biochemical profile, but not on a precise molecular aetiology. Families want and seek precise genetic diagnoses so that they can fully understand what has happened and what may happen to their child. Over the last ~15 years, an increasing number of monogenic growth and endocrine disorders have been elucidated (including those identified by investigators in our sub-domain. What the 100k GP now offers is the opportunity to assign a specific genetic diagnosis to many more children with growth and endocrine disorders.
In addition functional work on genes causing growth and endocrine disorders has significantly extended our understanding of complex phenotypes (e.g. midline/eye abnormalities and the

pituitary and relationships between systems (e.g. growth, cancer and metabolic predispositions. Within the 100k GP, functional work on novel growth & endocrine genes and novel variants in previously identified genes is very likely to help us to identify much broader spectra of phenotypes and enhance our ability to prognosticate in cases where families are desperate to know what lies ahead decade by decade.
Children with growth and endocrine disorders are usually followed over many years: some will receive treatments, e.g. recombinant human (rh-) Growth Hormone or Insulin-like Growth Factor-I. Identification of children with mutations or variants through the 100k GP offers the opportunity for longitudinal study of evolution of phenotype, identification of optimal therapies, and possibly prediction of responses to treatments. Identification of novel molecules implicated in these rare disorders may also lead to new therapies for these disorders.

Research plans. *Give details of the analyses and experimental approaches, study designs and techniques that will be used and timelines for your analysis. Describe the major challenges of the research and the steps required to mitigate these.*

Objective 1:
(A) To identify new mutations in known growth/endocrine genes in children not previously recognised to fit the clinical disease description – i.e. Extending the phenotype. These will be for children with diseases included in the Paediatrics cross-cutting domain such as Kabuki, Noonan, VACTERL-like phenotype.
(B) To identify mutations in new growth/endocrine genes in children without a precise diagnosis – i.e. Generate new knowledge.

Objective 2:
To explore the functional consequences of mutations in known and novel growth/endocrine genes by a range of techniques (dependent on the type of mutation found) - see list below.

Objective 3:
To use existing biological samples (serum / urine / skin fibroblasts etc / mRNA) to profile changes over time and/or in response to therapies, using integrated 'omic and clinical analyses to generate profiles that could inform a personalised medicine approach.

Any novel genomic variant identified will need to be carefully evaluated in the context of the clinical phenotype. Detailed phenotyping through existing medical records will include developmental & system history, detailed anthropometry, biochemistry (GH provocation, IGF-I generation, other hormonal profiling), XRs (including skeletal survey) and MR brain/pituitary. Careful examination of any other systems involved e.g. cardiovascular, respiratory, renal, special senses.
The variant will be assessed by one or more of the following approaches:
For Objective 1

- Public database consultation to assess the frequency of the variant in large population cohorts.
- Bioinformatic analysis to assess the likely impact of the variant on protein function. If the variant is likely to impact on protein function, then further studies will be performed.
- Structural modelling of the mutation using a bioinformatics approach.
- Detailed clinical phenotyping of all children and adults (in collaboration with Endocrine and Metabolism GeCIP) carrying the same novel gene variant.

For Objective 2

- Gene expression analysis in human/murine embryonic/post-natal tissue to study the expression of the gene in the relevant tissues.

- Design of relevant in vitro studies e.g. transient transfection assays, electrophoretic mobility shift assays, cell-based assays (receptor function, signalling, enzyme/metabolite changes). These will depend on the function of the gene that is affected.
- Proteomic approaches using relevant tissues.
- Generation of relevant animal models e.g. zebrafish, Xenopus, Mouse, using either whole animal or conditional transgenesis
- Induced pluripotent stem cells to investigate gene function with a view to cell-based correction of the defect

For Objective 3

Many of the short children with defects in the hypothalamo-pituitary-somatotroph axis will be treated with rhGH and/or rhIGF-I, and therefore data pertaining to response to therapy will be available. In those with these complex growth disorders, response to therapy is often disappointing. It is hoped that the molecular defects and variants identified will lead to the identification of new avenues for therapeutic approaches based on targeting the specific functional impairment rather than bluntly trying to overcome tissue resistance with pharmacological doses of growth factors. One recent example of this is the development of C-type Natriuretic peptide as an inhibitor of MAPK signalling downstream of a constitutively activated FGFR-3 in achondroplasia [Yasoda et al Nat Med 2004].

The children enrolled in the 100k GP are usually followed for many years in Growth / Endocrine / Metabolic clinics, and those enrolled as adults will already have gone through such follow-up. Many of these conditions have evolving phenotypes – e.g. children starting with isolated GH deficiency develop other pituitary hormone deficiencies over time, children with complex growth disorders may develop a metabolic phenotype as adults. There is therefore significant opportunity to follow evolving phenotypes and co-morbidities. This is where collaboration with many of the other GeCIPs, in particular Metabolic and Endocrine, will be very important. In addition recruitment to this part of the work could be enhanced through the members of the British Society of Paediatric Endocrinology & Diabetes, the European Society of Paediatric Endocrinology, the European Reference Networks (which will be forming over 2016 and will include a are endocrine disorders network) and our international collaborators, including Prof Dauber [Cincinnati Center for Growth Disorders], Prof Camper [University of Michigan] and Prof Wit [University of Leiden], who are supporting the Growth GeCIP.

Therefore on these cohorts, we would plan to collect from each child blood and urine samples taken under standardised conditions for Transcriptomic, Metabolomic and Proteomic studies. Facilities for undertaking such work are available in the Centres represented in the Domain, including Manchester, Birmingham & London. Integrated analyses of these datasets will identify key pathways impacting on phenotype and response to treatments, and should identify pathways and 'master' regulators of these pathways that could be novel biomarkers for disease progression or therapeutic targets. For example, around 40-60% of patients with isolated GHD tend to have normal GH secretion on retesting in adolescence. Identification of molecular or proteomic biomarkers that could distinguish between those patients who reverse and those who do not could help target patients more accurately, and possibly prevent unnecessary treatment and the expense incurred therein.

Collaborations including with other GeCIPs. *Outline your major planned academic, healthcare, patient and industrial collaborations. This should include collaborations and data sharing with other GeCIPs. Please attach letters of support.*

Disordered growth and/or hormonal dysfunction may accompany many rare conditions as major or minor phenotypic features. Understanding the mechanism for these abnormalities in particular when they are a minor feature of conditions with other more prominent characteristics will be a great opportunity to broaden our knowledge of pathways that disrupt endocrine function. We therefore anticipate that our sub-domain will collaborate extensively both with the other sub-domains within the Paediatric domain and with the other GeCIPs. We already have a Growth & Development sub-domain within the Endocrinology & Metabolism GeCIP, and all the other GeCIPs will include conditions in which growth ± endocrine abnormalities can occur.

We have secured collaborations internationally with four groups, two in Europe and two in the US, that work in the field of gene discovery in growth and hypothalamic-pituitary disorders:
Dauber & Hwa, Cincinnati Children's Hospital
Wit, Leiden University Medical Center
Camper, University of Michigan
Pitteloud, Lausanne University Hospital

In addition Dattani is a Founder member and Steering group member of the newly established EMA supported EnPrEMA initiative in Paediatric Endocrinology and Diabetes (CADET), which will lead to the formation of clinical networks of paediatric endocrinologists based in Europe and foster collaborations with pharma.
Dattani is also a member of an international consortium of clinical and basic scientists working on disorders of the pituitary gland. Annual workshops are held that foster collaboration in this area of research. The 7th International workshop on pituitary disorders was recently held in Edinburgh.

Our GeCIP has strong connections with patients groups in the UK, in particular the Child Growth Foundation, the Turner Syndrome Support Society, & the Pituitary Foundation. We also have links to family support groups in Europe and the US.

Our GeCIP also has long-standing collaborations with Pharma: (1) Clayton was CI for the Merck Serono PREDICT study looking at the pharmacogenomics of r-hGH treatment in GH deficiency and Turner syndrome, (2) Clayton was CI for the Ipsen EPIGROW study, a pharmacoepidemiological study in Short Stature, which looked at the frequency of growth gene variants in short children with undefined aetiology versus controls, (3) Storr is supported by Ipsen, UK to provide the Growth Hormone Insensitivity Genetic Sequencing Service at the William Harvey Research Institute, London.
We therefore have excellent access to Pharma companies that are interested in Growth & Genomics. In addition through ESPE, its Corporate Liaison Board and annual meeting exhibition sponsors we are connected to most of the pharma companies delivering clinical trials in the growth field.

Training. *Describe the planned involvement of trainees in the research and any specific training that will form part of your plan.*

Currently the Domain supports Dr A Foster, who is a Clinical genetics trainee out of programme enrolled on a PhD supported by the NIHR RD Translational Research Collaboration. Each centre represented in our sub-domain has extensive experience of supervising both clinicians and scientists doing postgraduate degrees related to Rare Disease. Masters courses in Rare Disease and Personalised Medicine (starting 2019) are being planned in London. Masters courses in Genomics and in Bioinformatics are being launched in Manchester. There are highly successful

HEE MSc Genomic Medicine course run in University of Birmingham and University College London/Queen Mary London, for which the contracts have just been renewed. Modules include the compulsory rare diseases module.

Sandoz are providing support for Clinical Fellowships at St Bartholomew's Hospital, which can be linked to Rare Growth Disorder research.

People and track record. *Explain why the group is well qualified to do this research, how the investigators would work together.*

Currently Manchester, London, Edinburgh, Glasgow, Birmingham and Southampton are represented in the sub-domain, but we anticipate membership will expand. London, Manchester & Birmingham run the 3 largest paediatric endocrine units in the UK, including the largest growth clinics. Currently the members include those with the most active growth & pituitary research groups.

We will establish an operational plan to achieve the objectives outlined above, with distribution of methodologies to the labs with the relevant expertise, including our international collaborators; a pipeline to analyse, interpret and validate the results as they filter back to the sub-domain; monthly teleconferences; twice yearly face-to-face meetings; and a process for presenting and publishing data in line with the NHSE policy.

Clinical interpretation. *(Where relevant to your GeCIP) Describe your plans to ensure patient benefit through clinical interpretation relevant to your domain. This should specifically address variant interpretation and feedback and your interaction with the cross-cutting Validation and Feedback domain.*

We have already been actively involved in compiling the gene list in the PanelApp. This is a fast moving field and a number of the genes we have added to the panel have only been reported in 2015. We will therefore need to keep the panel up-to-date based on discoveries reported at the network of International meetings we attend. This will mean that we will be in the best position to provide potential new diagnoses to families / children with new / unrecognised phenotypes. We will be guided by the V&F domain on the best approach to feedback when novel variants are found.

Members of the group have experience in this process from involvement in feedback for the Genetics of Obesity Study (Farooqi, Cambridge) and more recently for the genetics of Congenital Hyperinsulinism (Ellard, Exeter).

Beneficiaries. *How will the research benefit patients and healthcare institutions including the NHS, other researchers in the field? Are there other likely beneficiaries?*
Above all and most importantly, patients & families: understanding the diagnosis and pathogenesis, taking away feelings of guilt, genetic counselling, pre-implantation genetic diagnosis, antenatal diagnosis, prediction of disease evolution.

But also:
Our clinical teams – better understanding of disease pathogenesis
Our trainees –understanding the key role genetics can play in management of RDs
UK research – Patient databases, deep phenotyping, long-term natural history studies; ready access for research studies

International research – stimulating other groups to study new mechanisms discovered in 100k GP, or draw together their own research observations with these newly discovered mechanisms
European reference networks – new EU initiative for 2016
Pharma collaborations – stimulate the development of new drugs, repositioning of established drugs

Commercial exploitation. *(Where relevant to your GeCIP) Genomics England has a very explicit intellectual property policy. We and other funders need to know if the proposed research likely to generate commercially exploitable results. Do you have commercial partners in place?*

No commercial exploitation at present, however it is possible that commercially exploitable discoveries will be found – e.g. a growth / hormonal pathway that is open to the generation of novel therapies.

References. *Provide key references related to the research you set out.*
The following are examples of work done by members of the sub-domain, relevant and related to the Research Objectives listed above.

1. **Dattani MT,** Martinez-Barbera JP, Thomas PQ, Brickman JB, Gupta R, Martensson I, Toresson H, Fox M, Wales JKH, Hindmarsh PC, Krauss S, Beddington RSP, Robinson ICAF (1998) Mutations in the homeobox gene *HESX1/Hesx1* associated with septo-optic dysplasia in human and mouse. *Nature Genetics* **19(2):** 125-133.
2. Carvalho LR, Woods KS, Mendonca BB, Marcal N, Zamparini AL, Stifani S, Brickman JM, Arnhold IJP, **Dattani MT** (2003) A homozygous mutation (I26T) in the engrailed homology domain of HESX1 associated with evolving hypopituitarism due to impaired repressor-corepressor interaction. *Journal of Clinical Investigation* **112 (8):**1192-1201.
3. Kelberman D, Rizzoti K, Avilion A, Bitner-Glindzicz M, Cianfarani S, Collins J, Chong WK, Kirk JM, Achermann JC, Ross R, Carmignac D, Lovell-Badge R, Robinson IC, **Dattani MT** (2006) Mutations within Sox2/SOX2 are associated with abnormalities in the hypothalamo-pituitary-gonadal axis in mice and humans. *J Clin Invest* **116 (9):**2442-2455.
4. Sun Y, Bak B, Schoenmakers N, van Trotsenburg ASP, Oostdijk W, Voshol P, Cambridge E, White JK, le Tissier P, Mousavy Gharavy SN, Martinez-Barbera JP, Stokvis-Brantsma WH, Vulsma T, Kempers MJ, Persani L, Campi I, Bonomi M, Beck-Peccoz P, Zhu H, Davis TME, Hokken-Koelega ACS, Del Blanco DG, Rangasami JJ , Ruivenkamp CAL, Laros JFL, Kriek M, Kant SG, Bosch CAJ, Biermasz NR, Appelman-Dijkstra NM, Corssmit EP, Hovens GCJ, Pereira AM, den Dunnen JT, Breuning MH, Hennekam RC, Chatterjee KK*, **Dattani MT***, Wit JM*, Bernard DJ* (*Co-Senior Authors) (2012) Loss-of-function mutations in IGSF1 cause a novel X-linked syndrome of central hypothyroidism and testicular enlargement *Nature Genetics* **44(12):** 1375-1381.
5. Stevens A, Hanson D, Whatmore A, Destenaves B, Chatelain P, **Clayton P**. Human growth is associated with distinct patterns of gene expression in evolutionarily conserved networks. BMC Genomics. 2013 Aug 13;14:547.
6. Lancaster MA, et al., Jackson AP, Knoblich JA (2013) Cerebral organoids model human brain development and microcephaly. Nature. 501:373-379
7. **Clayton P**, Chatelain P, Tatò L, Yoo HW, Ambler GR, Belgorosky A, Quinteiro S, Deal C, Stevens A, Raelson J, Croteau P, Destenaves B, Olivier C. A pharmacogenomic approach to the treatment of children with GH deficiency or Turner syndrome. Eur J Endocrinol. 2013 Jul 29;169(3):277-89.

8. **Clayton P**, Bonnemaire M, Dutailly P, Maisonobe P, Naudin L, Pham E, Zhang Z, Grupe A, Thiagalingam A, Denèfle P; EPIGROW Study Group. Characterizing short stature by insulin-like growth factor axis status and genetic associations: results from the prospective, cross-sectional, epidemiogenetic EPIGROW study. J Clin Endocrinol Metab. 2013 Jun;98(6):E1122-30.
9. Stevens A, Bonshek C, Whatmore A, Butcher I, Hanson D, De Leonibus C, Shaikh G, Brown M, O'Shea E, Victor S, Powell P, Settle P, Padmakumar B, Tan A, Odeka E, Cooper C, Birch J, Shenoy A, Westwood M, Patel L, Dunn BW, **Clayton P**. Insights into the pathophysiology of catch-up compared with non-catch-up growth in children born small for gestational age: an integrated analysis of metabolic and transcriptomic data. Pharmacogenomics J. 2014 Aug;14(4):376-84.
10. Martin CA, Ahmad I, Klingseisen A, Hussain MS, et al inc Jackson AP*. Mutations in PLK4, encoding a master regulator of centriole biogenesis, cause microcephaly, growth failure and retinopathy (2014). Nature Genetics 46:1283-92.
11. de Bruin C, Mericq V, Andrew SF, van Duyvenvoorde HA, Verkaik NS, Losekoot M, Porollo A, Garcia H, Kuang Y, Hanson D, **Clayton P**, van Gent DC, Wit JM, Hwa V, Dauber A (2015) An XRCC4 splice mutation associated with severe short stature, gonadal failure, and early-onset metabolic syndrome. J Clin Endocrinol Metab. May;100(5):E789-98.
12. Gaston-Massuet C, McCabe MJ, Scagliotti V, Young RM, Carreno G, Gregory LC, Jayakody SA, Pozzi S, Gualtieri A, Basu B, Koniordou M, Wu CI, Bancalari RE, Rahikkala E, Veijola R, Lopponen T, Graziola F, Turton J, Signore M, Mousavy Gharavy SN, Charolidi N, Sokol SY, Andoniadou CL, Wilson SW, Merrill BJ, ***Dattani MT**, *Martinez-Barbera JP (*Co-Senior Authors) (2016) Transcription factor 7-like 1 is involved in hypothalamo-pituitary axis development in mice and humans. *Proc Natl Acad Sci USA (in press)*.
13. Harley, ME, Murina, O, et al. inc Jackson, AP* (2016) TRAIP promotes DNA damage response during genome replication and is mutated in primordial dwarfism. Nature Genetics. 48:36-43

Full proposal (total max 1500 words per subdomain)

Title

(max 150 characters) **Imprinting Disorders Subdomain**

Importance. *Explain the need for research in this area, and the rationale for the research planned. Give sufficient details of other past and current research to show that the aims are scientifically justified. Please refer to the 100,000 Genomes Project acceptable use(s) that apply to the proposal (page 6).*

The imprinting subdomain proposes genome analysis in patients with clinical imprinting disorders (ImpD), with emphasis on both coding and noncoding gene products, cis-acting distal or proximal regulatory sequences, cis-acting genomic rearrangements, and trans-acting mutations. Currently, the clinical heterogeneity and epigenetic aetiology of ImpDs fragments existing clinical cohorts, hinders molecular diagnosis of ImpDs and impedes stratified care. Moreover, the phenotypes of ImpDs overlap with many existing recruitment categories (intellectual delay and abnormal behavior and growth disturbance).

The imprinting subdomain will, in conjunction with the 100,000 genomes project, seek externally funded collaborative funding for epigenetic analyses in order to identify undiagnosed ImpD patients within the 100K project. This will delineate an expanded phenotype of ImpD and enable effective translation of novel molecular diagnostic ImpD services into NHS practice.

Research plans. *Give details of the analyses and experimental approaches, study designs and techniques that will be used and timelines for your analysis. Describe the major challenges of the research and the steps required to mitigate these.*

1. Novel Disease gene Discovery:

A: Trans-acting mutations in ultra-rare ImpD cases cause general disturbance of imprinting, and these mutations have yielded vital insights into the mechanisms of imprinting and epigenetics. The co-ordination of research effort to find more of these rare mutations will maintain critical UK advantage in this fast-advancing field. These patients can already access 100,000 genomes project.

B: Genome analysis to increase diagnostic yield in patients with clinical imprinting disorder phenotype but no molecular (genetic or epigenetic) diagnosis. Clinical genetics partners have significant numbers of ImpD patients who remain undiagnosed under current targeted testing strategies. We propose genome analysis in these patients, with particular interest in imprinted gene products and developmental pathways shared with them, to identify novel molecular causes of disease.

C: There is a growing literature of genetic variants in mothers associated with ImpDs in offspring: maternal-effect mutations may account for 50% of so-called multi-locus imprinting disorders, particularly in children with atypical phenotypes, or mothers with a clinical history of reproductive difficulty. Analysis focused on the maternal genome is essential to identify maternal variation predisposing to offspring developmental disorders.

2. Enhancing molecular diagnosis of inherited disease by WGS within imprinted loci:

In five of the eight well-known ImpDs, targeted sequencing has identified a proportion of cases with cis-acting genetic mutations causing epigenetic errors. Such analysis is limited by the cost of targeting and the fragmentation of patient cohorts, and is not available in NHS service. WGS analysis of these patients will identify the genomic factors critical for transmission of epigenetic signatures, leveraging research and clarifying diagnosis through comparison with non ImpD patients.

3. Functional annotation of non-coding regulatory elements:

A: By integrating deep-phenotyping and WGS in ImpD patients and non-ImpD patients (already within 100K project) we will identify genetic variants modifying ImpD presentation, clinical severity, prognosis and risk of co-morbidities, eg: (i) GH unresponsiveness and pathologically high IGF1 in some SRS cases; (ii) risk of childhood cancer in BWS; (iii) predisposition to adult diabetes in TNDM; (iv) predisposition to metabolic syndrome, diabetes, and stroke in adults with SRS. Identification of genetic risk for these morbidities will likely shed light on population-based risk for related lifecourse diseases including diabetes and cancer. Moreover, comparison of the genomic mutation load between imprinted and non-imprinted loci will contribute to the ongoing debate about the evolutionary origins of genomic imprinting.

B: Methylation signatures of imprinting disturbance will be defined for prognostication. Imprinting disturbances are not always concordant with clinical affectedness in ImpD patients, seriously challenging prognostication and management. Using the principles developed for patients with mutations in chromatin modifiers (eg Aref-Eshghi et al, 2017) we will define blood methylation signatures in patients with ImpDs. This will support prognostication, potentially highlight optimal therapies, and define common pathways affected by ImpDs and related disorders.

C: Identification of novel imprinting effects: Currently 48 imprinted DNA regions are identified on 18 of 22 autosomes, but only 8 imprinted regions are associated with 'classical' ImpDs. Agnostic methylomic analysis will be performed to identify methylation disturbance at non-classical loci, and map phenotypic associations. This work will also capture the effects of transient imprints and

metastable epialleles, which new research suggests may potential environmental modulation of early developmental trajectory.

4. Multiomics studies:
Given the focal congenital (germline and zygotic) epigenetic disturbances in imprinting disorders, an ImpD cohort represents a unique human cohort to dissect the relationship of genome and epigenome.
A: We will match genomes with epigenomes in patient samples, to identify epigenetic signatures associated with specific mechanisms of disordered imprinting, and to identify novel genetic and epigenetic mutations in ImpDs characterised by growth disturbances. ImpD methylomes matched with genomes will also be analysed in relation to population genomic and epigenomic variation with age, gender and lifecourse disease (external funding required).
B: We will integrate the genomes of individuals with their transcriptomes (blood and circulating microvesicle), proteomes (blood), lipidomes (plasma) and clinical phenotypes. This will distinguish causal (epi)genomic variation from that secondary to the presenting and developing phenotypes seen in ImpDs, many of which are of fundamental interest to human health (obesity, growth, cancer). New imprinted RNA species, including noncoding RNA and microRNA, will be identified. Combining data, to integrate genomic variation with DNA methylation, gene expression, lipid and protein profiles and resultant phenotype will provide unique insights into variations in ImpD pathophysiology (external funding required).

Collaborations including with other GeCIPs. *Outline your major planned academic, healthcare, patient and industrial collaborations. This should include collaborations and data sharing with other GeCIPs. Please attach letters of support.*

We will collaborate closely with cross-cutting GECIPs, particularly those focused on evolving the current pipelines for variant detection, annotation and filtering; this will be crucial for detecting non-coding variation causing dysregulation of imprinted genes. Several of us are co-located within the Epigenomics GeCIP and the nascent UK-GFP and will foster the symbiotic links needed to optimise the value of multi-omic studies. We also anticipate collaborating closely with colleagues within the DOD subdomain, to capture individuals with currently-unrecognised genetic mutations causing epigenetic disease.

Training. *Describe the planned involvement of trainees in the research and any specific training that will form part of your plan.*

Both clinical and scientific members of the Imprinting subdomain have mature expertise in designing and delivering training in clinical and molecular epigenetics, including:

- Supervision of students at all levels in epigenetic projects, including EU Marie Curie studentship programmes
- Module and programme leadership across several of the sister MSc programmes in Genomic Medicine developed by Genomics England and Health Education England (including Epigenetics and Epigenomics module)
- Organisation and leadership in the “ID-Schools” delivered through the EU COST network on imprinting
- Regular contribution to ESHG and UK dysmorphology schools on clinical genetics

This established expertise will act as a framework for building an integrated network of training resources in cutting-edge research and medical epigenetics. This will equip trainees with the knowledge and skills to interact with genomic and epigenomic DNA, in a way that is fit for purpose in their careers in this rapidly-developing field.

People and track record. *Explain why the group is well qualified to do this research, how the investigators would work together.*

Imprinting disorders are 'younger' as a field than larger and more established subdomains of genetic or clinical research:– this gives us a striking advantage in terms of tight-knit collaborations and extremely close and cordial contact across Europe and beyond. We are closely integrated in a communication network (www.imprinting-disorders.eu) that supports regular meetings between clinicians, researchers and patient support groups across the EU. Network members are close collaborators in funding proposals (including FP7, H2020 and Marie Skłodowska-Curie actions) and specifically in translational collaborations between clinicians and basic scientists. Temple, Maher, Mackay and Moore are all internationally recognised for ImpD research and have played leading roles in the establishment of the European Congenital Imprinting Disease Network. Furthermore, key Imprinting subdomain researchers are closely aligned with NHS clinical genetic and molecular diagnostic laboratories in Cambridge, London and Southampton/Salisbury and well placed to ensure translation of research into NHS practice.

Clinical interpretation. *(Where relevant to your GeCIP) Describe your plans to ensure patient benefit through clinical interpretation relevant to your domain. This should specifically address variant interpretation and feedback and your interaction with the cross-cutting Validation and Feedback domain.*

We will collaborate with the cross-cutting Validation and Feedback domain, to inform the development of interpretation resources needed to identify noncoding regulatory variation associated with ImpDs. This will provide new genetic diagnoses, probably including both new causes of known disorders and new causes of new clinical entities identified through integration of genomic, epigenomic and clinical data.

Beneficiaries. *How will the research benefit patients and healthcare institutions including the NHS, other researchers in the field? Are there other likely beneficiaries?*

The primary beneficiaries of this work will be patients and their families through provision of new diagnoses, and the consequent stratification of prognosis, counselling and access to treatment. The lean and focused nature of our networking will ensure rapid dissemination and translation, particularly to NHS professionals.
Beyond this, we are using congenital ImpDs as a unique paradigm for genetic predisposition to epigenetic disease. Our findings – including biological mechanisms and genetic aetiologies – will inform research into population-level lifecourse diseases such as cancer and diabetes which are a burgeoning burden on health and wealth. In consequences the long-term beneficiaries of this research will include policy-makers and populations.

Commercial exploitation. *(Where relevant to your GeCIP) Genomics England has a very explicit intellectual property policy. We and other funders need to know if the proposed research likely to generate commercially exploitable results. Do you have commercial partners in place?*

We are already working as a network with commercial providers focused on the development of new diagnostic tools. Our aim is to improve the accuracy of genetic/epigenetic diagnosis and improve technical robustness; there will also be substantial cost reductions leading to greater diagnostic capacity and equity of access to testing.

In terms of management and therapeutics, identification of new aetiologic/pathological mechanisms will define opportunities to develop novel therapeutic targets. We will engage with industry to ensure these opportunities are explored fully.

References. *Provide key references related to the research you set out.*

1. Eggermann T, Perez de Nanclares G, Maher ER et al. Imprinting disorders: a group of congenital disorders with overlapping patterns of molecular changes affecting imprinted loci. Clin Epigenetics. 2015 Nov 14;7:123.

2. Meyer E, Lim D, Pasha S et al: Germline mutation in NLRP2 (NALP2) in a familial imprinting disorder (Beckwith-Wiedemann Syndrome). PLoS Genet 2009; 5: e1000423.

3. Docherty LE, Rezwan FI, Poole RL et al. Mutations in NLRP5 are associated with reproductive wastage and multilocus imprinting disorders in humans. Nat Commun. 2015 Sep 1;6:8086.

4. Begemann M, Rezwan FI, Beygo J et al. Maternal variants in NLRP and other maternal-effect proteins are associated with multi-locus imprinting disturbance in offspring. J Med Genet. 2018, in press.

5. Eggermann T, Binder G, Brioude F, Maher ER et al. CDKN1C mutations: two sides of the same coin. Trends Mol Med. 2014;20:614-22.

6. Abi Habib W, Azzi S, Brioude F et al. Extensive investigation of the IGF2/H19 imprinting control region reveals novel OCT4/SOX2 binding site defects associated with specific methylation patterns in Beckwith-Wiedemann syndrome. Hum Mol Genet. 2014 Nov 1;23(21):5763-73

7. Beygo J, Elbracht M, deGroot K, et al. (2015). Novel deletions affecting the MEG3-DMR provide further evidence for a hierarchical regulation of imprinting in 14q32. Eur J Hum Genet 23:180–188

8. Demars J, Shmela ME, Khan AW et al. Genetic variants within the second intron of the KCNQ1 gene affect CTCF binding and confer a risk of Beckwith-Wiedemann syndrome upon maternal transmission. J Med Genet. 2014 Aug;51(8):502-11.

Full proposal (total max 1500 words per subdomain)

Title

(max 150 characters) **Improving diagnosis and recognition of ciliopathies and identifying therapeutic targets**

Importance. *Explain the need for research in this area, and the rationale for the research planned. Give sufficient details of other past and current research to show that the aims are scientifically justified. Please refer to the 100,000 Genomes Project acceptable use(s) that apply to the proposal (page 6).*

Ciliopathies are incurable inherited diseases collectively affecting 1 per 200 births. A lack of formal clinical diagnostic measures, variable disease features and multiple organ involvement hamper clearcut clinical diagnosis, with patients subject to multiple clinic visits, delayed or missed diagnosis, and suboptimal disease management that impacts upon morbidity and lifespan. Mutations in >100 genes cause ciliopathies [1], with extensive genetic and phenotypic variability within and between different subtypes [2]. This large and complex disease spectrum of > 30 major subtypes has numerous disease variants that range from chronic diseases where lifelong disease management for better quality of life and survival is the priority, to lethal developmental malformations demanding better prenatal diagnosis options. Variable penetrance, epistatic alleles and triallelic inheritance are all proposed to influence ciliopathies, suggesting an extensive influence of modifier alleles [3-5]. Ciliopathies are a paradigm of rare disease complexity that promises to benefit from the GEL Project, addressing the unmet need of unrecognised ciliopathy patients, offering new cell biological clues and ideas for new gene-based therapeutics.

Research plans. *Give details of the analyses and experimental approaches, study designs and techniques that will be used and timelines for your analysis. Describe the major challenges of the research and the steps required to mitigate these.*

Background. Our overall aim is to optimise the interpretation of VUS and maximise the benefits of genomic sequencing for ciliopathy patients. In-house cilia resources offer significant advantages for this: (i) the Ciliome. We are significantly assisted in this by (i) definition of around 1,000 genes considered sufficient for ciliary functions and resources such as Syscilia Gold Standard, CILDB, Cilia Proteome [11-13]; (ii) curated databases of ciliopathy exome-level variants, collaborative pre-publication gene lists of cilia genes/proteins from model ciliate organisms and candidate ciliopathy disease genes from our own and collaborative human genetics studies; (iii) cell-based and organism-based ciliated model systems (including iPSC-differentiated organoids for relevant tissues) for gene and protein expression studies and gene depletion/knockout studies; (iv) human ciliated tissue expression resource (Wellcome-MRC HDBR). (v) The recent award of a Wellcome Collaborative Grant (to Beales & Jenkins) to study the function of missense mutations in skeletal ciliopathies (start date October 2018). We also benefit from longstanding collaborative links to cilia researchers including human genetics labs worldwide, for cross-linking of novel variants and genes; and wider resources for rare disease discovery e.g. FORGE, the NIH Undiagnosed Diseases Network, and the UK's Rare Diseases Translational Research Collaboration, as well as making use of public resources such as Decipher and Matchmaker Exchange.

Specific aims:

1. Bioinformatic analysis to reveal the genome beyond the prescribed gene panel analysis performed by GEL, and putative modifier loci.
GEL genomes will be reanalysed at several levels, applying expertise of the ciliopathy literature and mining of in-house databases including cilia mutations and genes of interest from human genetic studies and studies of model organisms and various Omics studies by collaborators e.g. the EU FP7-funded SYSCILIA ciliary protein network. We will use ciliary localisation of gene products, knock-down studies, known functions or those inferred from large scale gene-protein network studies, to determine the nature of any required subsequent functional workup. Genome level insight into variants that can affect the larger structural organisation of genes, or the gene regulatory regions (promoters, enhancers, splicing mechanism) would be totally novel for ciliopathies and can move us towards wider understanding of the influence of modifier loci. We are already working on GEL pilot data with the Validation and Feedback GeCIP team (Bill Newman) on developing new genome interpretation alogorithms to detect such variants (Hywel Williams) The prior protein-protein complex network of the cilium developed through systems biological analyses (SYSCILIA) will provide an invaluable starting resource here, as well as for example a database of genes arising from siRNA library work that modify cilia length and intraflagellar transport [6], and databases of gene affected by the FOXJ1 multiciliogenesis transcription factor [7, 8]. The interpretation of modifier loci will likely require access to larger patient numbers through datamining our own or collaborative exomes. Probably 10% of ciliopathy diseases are explained by larger CNVs, and the option to detect interesting larger scale genomic rearrangements arising from genomic analysis is an exciting challenge. We may expect milder, regulatory mutations in surviving patients affected by more severe/lethal nonmotile ciliopathies.

2. Review of clinical phenotypic information on ciliopathy patients to enhance diagnostic likelihood and initiate genotype-phenotype databasing.
We will combine gene functional information with the clinical information available to determine whether additional features and disease outcomes are consistent with the new candidate genes,

or unusual variants in known genes. For example, previous work established that distinct gene defects underlying Jeune syndrome cause end-stage renal failure and retinal degeneration, versus symptoms restricted to milder skeletal phenotypes [5, 9]. Clinical-phenotype-genotype correlation will be essential in examination of putative modifier loci, and we will expand and curate our current ciliopathy database to document clinical features associated with different gene defects, a challenge that includes development of clinical scoring measures and merging of clinical datasets across different clinical centres. Expanding this database will develop an important resource of identified mutations and their clinical manifestations, working towards improved diagnosis of difficult variants and optimal family counselling with more meaningful prognostic indicators. This will form an important basis for work towards understanding disease modifier alleles, and the overall allelic burden underlying ciliopathy diseases.

3. Functional studies to evaluate prioritised variants of uncertain significance.
Selected variants will enter a functional discovery pipeline, to evaluate their likelihood as disease causing. We have cell- and organism-based cell models established for evaluation of human gene variants and many cilia antibody, siRNA and mutant construct resources to facilitate effective and rapid throughput variant analysis, as well as strong collaborative links with other ciliopathy labs e.g. ciliary proteomic network analysis (Ronald Roepman, Marius Ueffing) [10]. We conduct a range of tests for ciliary functions tailored to the ciliopathy under study including high resolution immunofluoresence, electron microscopic structural work, measurement of intraflagellar transport and cilia length and number measures. The typical pipeline recently supported the recent Wellcome Trust award will be:
a. *Proteomics (affinity/SILAC) studies* introducing patient mutations in to ciliary proteins to determine their effects on complex formation and stability.
b. *Gene expression studies* in patient ciliated tissues and in mouse and human embryo tissue (HDBR Resource); Western blot, qPCR and immunofluoresence imaging.
c. *Cell biological* localisation and mutational analysis to explore new disease candidates for deficits in cell biological functions e.g. predicted ciliary protein-protein interactions.
d. *Cell models*. We can model ciliopathy diseases using ciliated fibroblasts and urine-derived ciliated renal organoids (UDRO), or cell lines including RPE1, IMCD3. We also have 50+ iPS cell lines (via the Sanger Institute HiPSCi study) from BBS patients with known mutations some of which have been successfully differentiated into neurons, kidney and retinal organoids. Cellular phenotypes and deficits are already established for some known ciliary components, allowing analysis of the effects of interesting variants using similar functional outputs. CRISPR based mutagenesis or siRNA antisense gene depletion will model a simple gene disruption or VUS.
d. *Model organisms*. We can model human ciliopathy diseases in a number of model organisms that display ciliopathy phenotypes such as cilia dysmotility, cystic kidneys, craniofacial defects, neuronal tube and neural crest defects e.g. zebrafish, in which siRNA and CRISPR methods are well established at UCL. We have used *Paramecium*, mice and fly models to test novel genes for mRNA rescue complementation of the human ortholog, as well as ciliary mutants such as ENU zebrafish libraries and IMPC gene-trap mouse models (several already in-house).

4. Human multiomic analysis of ciliopathies and modifiers of ciliopathies
We have capacity to collect to expand on our current collections of patient fibroblasts, serum, urine and other materials (e.g. iPSC) for multiomic studies including RNASeq, proteomics. These materials will be essential to understanding the genetic complexity of ciliopathy diseases, and how different diseases genes within a single ciliopathy class are associated with distinct phenotypic variation and progression of disease, between families but even within families of multiple affected children. e.g. *C21orf2* mutations were recently associated with lethal Jeune syndrome or isolated retinal dystrophy [6]. To assist this work we have already started to generate biologically accurate, human ciliopathy models by human mutation engineering in hESC,

derivation of patient iPS and patient ciliated cells. These allow for derivation of lineage specific differentiated cells and organoids e.g. retinal, kidney, bone – to assess the effects of human mutations and expand knowledge about disease variability and impact of cilia defects such as impaired Hedgehog signalling per mutation/per cell type. Moreover, these human studies are complemented by derivation of new *in vivo* models of ciliopathies**.** Ultimately we aim to use these resources to evaluate the effects of genetic modifiers by CRISPR, to trial new therapies directed at specific classes of mutations, and to employ novel high content small molecule screening for identifying therapeutic entry points.

Collaborations including with other GeCIPs. *Outline your major planned academic, healthcare, patient and industrial collaborations. This should include collaborations and data sharing with other GeCIPs. Please attach letters of support.*

We can expect that ciliopathy variant diseases may emerge from identification of novel ciliary variants that arise in different GeCIPs. The following GeCIPs are predicted the most likely to reveal shared cilia related genomic data arising from the GEL Project:

Hearing and Sight GeCIP (Andrew Webster)
Renal GeCIP (Daniel Gale)
Respiratory (Eric Alton)
Neurological (Henry Houlden)
Musculoskeletal (Muhammad Kassim Javaid)
Cardiovascular (Bernard Keavney)
Endocrine and Metabolism (Stephen O'Rahilly)

Training. *Describe the planned involvement of trainees in the research and any specific training that will form part of your plan.*
We will include access to the following training opportunities:

- Great Ormond Street Hospital BRC – PhD programmes in Experimental Medicine and Springboard Fellowships
- Involvement in the BRC Early Careers Network
- Masters courses in Human Genetics at UCL, as well the HEE funded Genomic Medicine Masters run between UCL-Queen Mary University (contract recently renewed)
- New UCL MSc course in Personalised Medicine (to start in 2019)
- Short courses in Applied Genomics run six times a year at UCL Institute of Child Health

Early Careers Network – UCLP AHSC Personalised Medicine Domain and new UCL Institute of Precision Medicine

People and track record. *Explain why the group is well qualified to do this research, how the investigators would work together.*

Clinical and research expertise in ciliopathies in the UK is world-leading led by key personnel included in this subdomain.
Non-motile ciliopathy clinical and fundamental research is led is Phil Beales, Dagan Jenkins and Hannah Mitchison, Primary Ciliary Dyskinesia genetics research is led by Hannah Mitchison (who is also a member of the Respiratory GeCIP domain).
Core expertise in high resolution imaging and fundamental cilia biology in London and Leeds (Beales and Knight (QMUL) and also in Leeds, Colin Johnson); in model systems such as zebrafish in Sheffield (Jarema Malicki) and London (Beales, Knight).
Translational research programmes are underway at the Institute of Child Health/GOSH (e.g. gene therapy, read-through therapies).

Clinical Services for Ciliopathies
There are three nationally commissioned clinic services for ciliopathies notably:
Primary Ciliary Dyskinesia (Leicester, London, Southampton, Leeds-Bradford)
Bardet-Biedl Syndrome (London and Birmingham)
Alstrom Syndrome (Birmingham)

In 2019, it is likely a further service will be commissioned by NHSE to provide multidisciplinary care and management of patients with very rare ciliopathies (e.g. Jeune, Joubert, Mainzer-Saldino syndromes etc.)

The sub-domain will operate with mailing lists and regular virtual meetings of both individual sub-groups and the entire domain.

Clinical interpretation. *(Where relevant to your GeCIP) Describe your plans to ensure patient benefit through clinical interpretation relevant to your domain. This should specifically address variant interpretation and feedback and your interaction with the cross-cutting Validation and Feedback domain.*

We have already been actively involved in GEL and PanelApp data model development for the ciliopathies, and we maintain a gene list for ciliopathies that requires constant updating as the cilia-disease gene discovery literature rapidly expands. As already noted, these links are crucial to ensure patient relevance and benefit for counselling and prognosis. We are already interacting with the GEL V&F domain to maximise the benefits for ciliopathy variant interpretation.

Beneficiaries. *How will the research benefit patients and healthcare institutions including the NHS, other researchers in the field? Are there other likely beneficiaries?*

Patients will benefit through better diagnosis of ciliopathies that are highly heterogeneous and remain poorly recognised and under diagnosed. This project will assist in understanding the true incidence of ciliopathies, especially the impact of milder alleles in larger disease phenotypes such as retinal dystrophy and cystic kidney disease, and will open up improved NGS-based genetic diagnosis for more accurate counselling for families. Moreover, new cell biological clues about ciliopathies will prompt novel genetic therapeutics and pharmocogenomic approaches.

Commercial exploitation. *(Where relevant to your GeCIP) Genomics England has a very explicit intellectual property policy. We and other funders need to know if the proposed research likely to generate commercially exploitable results. Do you have commercial partners in place?*

GOSH and UCL have a commercial partnership with Congenica (Cambridge) to use software for genetics diagnosis and develop new tools for Omics data integration and analysis (via InnovateUK grant). Gene therapy programme for Bardet-Biedl syndrome is funded by Apollo Therapeutics.

References. *Provide key references related to the research you set out.*

1. Baker, K. and P.L. Beales, *Making sense of cilia in disease: the human ciliopathies.* Am J Med Genet C Semin Med Genet, 2009. **151C**(4): p. 281-95.
2. Travaglini, L., et al., *Expanding CEP290 mutational spectrum in ciliopathies.* Am J Med Genet A, 2009. **149A**(10): p. 2173-80.
3. Badano, J.L., et al., *Dissection of epistasis in oligogenic Bardet-Biedl syndrome.* Nature, 2006. **439**(7074): p. 326-30.
4. Davis, E.E., et al., *TTC21B contributes both causal and modifying alleles across the ciliopathy spectrum.* Nat Genet, 2011. **43**(3): p. 189-96.

5. Schmidts, M., et al., *TCTEX1D2 mutations underlie Jeune asphyxiating thoracic dystrophy with impaired retrograde intraflagellar transport.* Nat Commun, 2015. **6**: p. 7074.
6. Wheway, G., et al., *An siRNA-based functional genomics screen for the identification of regulators of ciliogenesis and ciliopathy genes.* Nat Cell Biol, 2015. **17**(8): p. 1074-87.
7. Choksi, S.P., et al., *Systematic discovery of novel ciliary genes through functional genomics in the zebrafish.* Development, 2014. **141**(17): p. 3410-9.
8. Newton, F.G., et al., *Forkhead transcription factor Fd3F cooperates with Rfx to regulate a gene expression program for mechanosensory cilia specialization.* Dev Cell, 2012. **22**(6): p. 1221-33.
9. Halbritter, J., et al., *Defects in the IFT-B component IFT172 cause Jeune and Mainzer-Saldino syndromes in humans.* Am J Hum Genet, 2013. **93**(5): p. 915-25.
10. Benmerah, A., et al., *The more we know, the more we have to discover: an exciting future for understanding cilia and ciliopathies.* Cilia, 2015. **4**: p. 5.
11. Gherman, A., E.E. Davis, and N. Katsanis, *The ciliary proteome database: an integrated community resource for the genetic and functional dissection of cilia.* Nat Genet, 2006. **38**(9): p. 961-2.
12. Arnaiz, O., et al., *Remodeling Cildb, a popular database for cilia and links for ciliopathies.* Cilia, 2014. **3**: p. 9.
13. van Dam, T.J., et al., *The SYSCILIA gold standard (SCGSv1) of known ciliary components and its applications within a systems biology consortium.* Cilia, 2013. **2**(1): p. 7.

Full proposal (total max 1500 words per subdomain)

Title *(max 150 characters)*	**Developmental and Orphan Disorders (DOD) sub-domain**

Importance. *Explain the need for research in this area, and the rationale for the research planned. Give sufficient details of other past and current research to show that the aims are scientifically justified. Please refer to the 100,000 Genomes Project acceptable use(s) that apply to the proposal (page 6).* ***[152 words]***

The major focus of the DOD sub-domain is the study of the genetic causes of both neurodevelopmental disorders (e.g. those associated with intellectual disability, autism, structural brain abnormalities and early-onset neurological disorders) and multi-system developmental disorders.

Numerous large-scale genome-wide genetic studies using both array-based and sequence-based technologies indicate that developmental disorders, as defined broadly above, represent a fundamental and uniquely informative domain of disorders for genetic research. These studies have demonstrated shared genetic aetiology across a broad range of frequently co-morbid neurodevelopmental disorders, such as intellectual disability (ID), autism and epilepsy. Moreover, the variable penetrance and expressivity of many multi-system disorders means that patients with identical mutations can have quite different clinical presentations. The proposed domain is ideal for conducting 'in-the-round' meta-analyses, e.g. of neurodevelopmental disorders, as well as sub-analyses of more specific disorders. Multi-system developmental disorders represent a major area of clinical need, with ID being the most frequently observed phenotype.

Research plans. *Give details of the analyses and experimental approaches, study designs and techniques that will be used and timelines for your analysis. Describe the major challenges of the research and the steps required to mitigate these.* ***[268 words]***

Data integration:

We will harmonise and integrate relevant Genomics England data with existing and already funded genome-wide genetic datasets (e.g. WES/WGS) generated by the sub-domain members on over 35,000 samples from families with DODs (DDD, NIHR-SPEED, UK10K, PAGE, Institute of Child Health Neurogenetics Research Group). In addition, we are willing to commit additional funds available to the Group to generate transcriptomic, metabolomic and epigenomic profiles on patients sequenced by Genomics England, using samples provided by the GMCs.

Harmonisation of datasets will proceed on an analysis-by-analysis approach. To address specific questions, different combinations of datasets will be required, and integration of data may occur at different levels. For example, for a meta-analysis of coding de novo mutations in neurodevelopmental disorders we will want to combine datasets which have exome or genome data available for parent-offspring trios (e.g. GEL + DDD + other datasets), and these data could be integrated at the level of high confidence de novo mutations, rather than entire VCF files. In this scenario, harmonisation will be needed at the level of variant detection (i.e. ensure similar sensitivity and specificity of detection of de novo mutations across datasets) and variant annotation (i.e. software and gene annotation resource). How the data are subsequently integrated for statistical testing will depend to some degree on the precise constraints on data ingress and egress from the GEL data centre. For case/control analyses of probands, a broader set of cohorts can be analysed, and similar harmonisation of variant annotation will be required but the harmonisation of variant detection will be different, and the level at which this harmonisation occurs will depend on the results of initial analyses assessing the comparability of the existing VCF files across studies.

Data analysis:
We will structure our analyses of these integrated datasets into sub-groups so as to make most effective use of the diverse expertise represented within the over 300 clinicians, researchers, trainees and clinical scientists within the DOD sub-domain, drawing upon our experience of coordinating highly-collaborative, multi-disciplinary research projects. The activities undertaken by specific sub-groups will vary, but will include:

• gene discovery in defined phenotypic subsets,
• genotype-phenotype analyses of patients with pathogenic variants in the same gene, or pathway,
• deep phenotyping of patients (e.g. MRC IMAGINE project)
• psychiatric record linkage across all Genomics England participants to fully characterize patient phenotypes, led by the NIHR BRC for Mental Health (Matthew Hotopf)
• assessing polygenic contributions to these disorders
• development of novel analytical methods
• characterization of the molecular processes generating germline pathogenic mutations
• functional analyses: investigating molecular, cellular or animal models
• identifying opportunities to develop novel therapeutic strategies (e.g. high-throughput drug screening, molecular chaperones, gene therapy
• improving NHS service delivery and developing and validating new tests

A major challenge will be a lack of power due to high genetic heterogeneity and sample size limitations, hence the importance of integrating Genomics England data with other relevant datasets.

Collaborations including with other GeCIPs. *Outline your major planned academic, healthcare, patient and industrial collaborations. This should include collaborations and data sharing with other GeCIPs. Please attach letters of support.* ***[114 words]***

We envisage collaborating with other GeCIPs focused on patients with related aetiologies to those studied in the DOD sub-domain. We anticipate collaborations with several GeCIPs combining expertise on non-syndromic and syndromic forms of developmental disorders, e.g. working with the Cardiology GeCIP on congenital heart defects, and the Neurology GeCIP in relation to the diseases with substantial shared genetic aetiology with the intellectual disabilities and other related disorders being studied in the Paediatric GECIP (e.g. epilepsies, movement disorders and neurodevelopmental disorders).

We will also collaborate with the Germline Mutation sub-domain of the cross-cutting Population Genomics GECIP (lead by Matthew Hurles), which will be evaluating and improving de novo mutation detection and filtering. We will also collaborate with other cross-cutting GECIPs that are improving variant detection, annotation and filtering beyond the core data resource provided by the standard Illumina calling pipeline.

Training. *Describe the planned involvement of trainees in the research and any specific training that will form part of your plan**. [209 words]***

Training within the DOD sub-domain will be integrated within the wider training plan for the Paediatric GECIP.

Members of the DOD sub-domain have a strong track record in designing and delivering training in clinical genomics, including:
• Supervision of masters and PhD students in bioinformatics and genetic data analysis
• Leading modules within the MSc programmes being developed as part of the Genomics England Initiative
• 'Fundamentals of Clinical Genomics' – biennial 3 day course sponsored by the Wellcome Trust, attended by consultants and trainees in Clinical Geneticists and other Specialties (~75 participants)
• Genomic Medicine for Clinicians – biennial course
• Annual DDD Collaborators' meeting for clinicians, scientists and research nurses
• UK dysmorphology meeting for trainees – annual meeting
• RSM course – Genomics of Paediatric Disease
• Genomics of Rare Disease – annual international conference (150-250 delegates)

Building on this expertise, the DOD sub-domain will ensure that clinical and research trainees working in disciplines contributing to the DOD sub-domain will participate in the research and will have access to courses and workshops that will equip them with the knowledge and skills to work with genomic data and an understanding of the scale and complexity of genomic variation and the resources and bioinformatics skills required to analyse it.

People and track record. *Explain why the group is well qualified to do this research, how the investigators would work together. **[213 words]***

The DOD sub-domain integrates several highly productive and collaborative clinical research initiatives, with over 300 clinicians, researchers and trainees with broad expertise in large-scale genetic analyses and clinical phenotyping:

• The Deciphering Developmental Disorders (DDD) collaboration of all 23 regional genetics services in the UK, which includes roughly 200 (>95% of) Consultant Clinical Geneticists who see patients with DDs, led by Matthew Hurles and Helen

Firth.
• The NIHR – SPEED collaboration with multiple cases of retinal dystrophy or Mendelian diseases presenting to paediatric neurology led by Lucy Raymond.
• The Paediatric Neurogenetics Network of cases with movement disorders, neurometabolic and neurodegenerative conditions, structural brain malformations and epilepsy led by Manju Kurian
• Adult Epilepsies with co-morbid ID or Autism, led by Michael Johnson
• The neurodevelopmental psychiatric group led by Gerome Breen, including expertise on autism and adult ID (Nick Bass and Andre Strydom) and record linkage (Matthew Hotopf).
• The Community Genomics Group, a UK-led study aiming to translate genomic findings into population benefit from investigations of neurodevelopmental disorders in diverse international populations, led by Andrew Crosby and Emma Baple.
• The Prenatal Assessment of Genomes and Exomes (PAGE) that is applying exome and genome sequencing to improve clinical management of developmental disorders detected by fetal ultrasound, led by Matthew Hurles.

A more comprehensive list of individuals requiring data access will be provided at a later stage.

The sub-domain will operate with mailing lists and regular virtual meetings of both individual sub-groups and the entire domain.

Clinical interpretation. *(Where relevant to your GeCIP) Describe your plans to ensure patient benefit through clinical interpretation relevant to your domain. This should specifically address variant interpretation and feedback and your interaction with the cross-cutting Validation and Feedback domain.*

We will work closely with the cross-cutting Validation and Feedback domain both to improve resources (e.g. Panel App) necessary to achieve a high diagnostic yield within the 100,000 Genomes Project, and to identify additional diagnostic variants that may have been missed by the standard variant calling, annotation and interpretation pipelines in the patients allocated to the DOD sub-domain. These additional diagnoses will include discovering new disorders as well as additional variants in already known DOD-associated genes that have been missed by the standard pipeline (e.g. complex structural variants, mosaic mutations).

Beneficiaries. *How will the research benefit patients and healthcare institutions including the NHS, other researchers in the field? Are there other likely beneficiaries?*

The primary beneficiaries will be patients and their families, through the provision of accurate genetic diagnoses. The know-how generated by this sub-domain will be disseminated through training to ensure this know-how benefits the NHS and its patients more broadly.

Commercial exploitation. *(Where relevant to your GeCIP) Genomics England has a very explicit intellectual property policy. We and other funders need to know if the proposed research likely to generate commercially exploitable results. Do you have commercial partners in place?*

The DOD sub-domain will work collaboratively to generate pre-competitive resources of utility to commercial organisations focused on diagnostic and/or therapeutic development. We will liaise with representatives of these organisations to ensure that these resources meet their requirements. On the diagnostic side, examples of pre-competitive resources generated by the members of the group include the DDG2P gene list that has been used by genome interpretation companies to improve their services (e.g. Congenica), and by genome technology companies to

improve their products, for example, optimized cost-effective assays (e.g. Oxford Gene Technology).

On the therapeutic side, we will identify opportunities to develop novel therapeutic targets and strategies and engage with industry to ensure these opportunities are explored more fully (e.g. existing collaboration between UCB and Michael Johnson on new drug target discovery in epilepsy).

Full proposal (total max 1500 words per subdomain)

Title **Unravelling the genetics of liver disease in children**
(max 150 characters)

Importance. *Explain the need for research in this area, and the rationale for the research planned. Give sufficient details of other past and current research to show that the aims are scientifically justified. Please refer to the 100,000 Genomes Project acceptable use(s) that apply to the proposal (page 6).* ***[148 words]***

The first year of life is the one with the most frequent presentation with liver disease. Infants present with different manifestations of liver dysfunction; jaundice, hepatocellular failure or hypoglycaemia being the most common. Some diseases have clear aetiologies, but these are the minority. The patients with the greatest unexplained diseases are those presenting with isolated cholestasis and those with liver failure.

The identification of new causes of early onset liver disease has been highly productive in recent years. However there remains a cohort of patients "neonatal hepatitis" and "idiopathic cholestasis". Slightly more than half of children presenting in the first 5 years of life, with liver failure, do not have a clear cause.

The genes that have been identified as causing neonatal onset liver disease in recent years have already lead to a better understanding of liver physiology and pathophysiology. Many such patients are already in clinical trials.

Research plans. *Give details of the analyses and experimental approaches, study designs and techniques that will be used and timelines for your analysis. Describe the major challenges of the research and the steps required to mitigate these.* ***[349 words]***

Objective 1:
Identify new genetic causes of cholestatic liver disease. This will largely be through the identification of genes not previously recognised as causing early onset liver disease.
Objective 2:
Identify new genetic causes of liver failure occurring in the first 5 years of life.
Objective 3:
Establish the functional consequences of variants identified through objectives 1 and 2.

Objectives 1 and 2 will follow similar plans. The extensive bioinformatics experiences in the collaborating centres, combined with the tools available through GEL will be combined to extract the highest likelihood of success. This is essentially activity that is already going on, on a daily basis, with existing cohorts of families/children. However we will work with GEL to urgently develop eligibility criteria for these 2 disorders (early onset liver failure; and early onset cholestatic liver disease), so that sufficient samples can be collected during the remainder of the GEL project.

Objectives 1 and 2 will be supplemented by existing data sets within the collaborating laboratories.
Objective 3 will follow several different lines. The existing close working relationships at all sites with liver histopathology will allow localisation and distribution of proteins expressed by disease associated genes. As liver tissue from 100k Genome patients will not generally be available tissue from local patients found to have variants in the same genes will be used. All centres have extensive biobanking and histological archiving, that makes this very feasible.
Several in vitro testing systems are currently in use, and others are currently being developed. Previously expression of genes implicated in neonatal liver disease have been expressed by use in vitro, using a number of different cell systems. This has allowed the study of splicing defects, protein trafficking abnormalities and membrane transport processes. These tools will continue to be used as appropriate.
Hepatocyte and cholangiocyte like cells have both been derived from iPSC in the last few years. These are currently being used in our laboratories to explore the consequences of genetic variants derived from WES data. Cells have been created from patient fibroblasts, and more recently through genetic modification of normal cells. The latter will be ideal for the GeCIP related work, as suitable patient material will not be available in most cases.
These methods will allow robust characterisation of genetic variants and their functional consequences in a systematic and timely manner.

Collaborations including with other GeCIPs. *Outline your major planned academic, healthcare, patient and industrial collaborations. This should include collaborations and data sharing with other GeCIPs. Please attach letters of support.* ***[44 words]***

We have previously shown that many of the genetic liver diseases presenting in early life have milder variants presenting in later childhood and in adults. The contribution to adult disease will be explored through the Gastro/Hepatology GeCIP domain of which we are also members.

Training. *Describe the planned involvement of trainees in the research and any specific training that will form part of your plan****. [31 words]***

Our laboratories already train basic scientists and clinical trainees. In addition several HSST posts have been applied for the HEE and others will be sought over the course of the project.

People and track record. *Explain why the group is well qualified to do this research, how the investigators would work together.* ***[26 words]***

The paediatric hepatology community worldwide is small, and has a strong track record of collaboration. We have coordinated numerous national and international laboratory and clinical studies.

Clinical interpretation. *(Where relevant to your GeCIP) Describe your plans to ensure patient benefit through clinical interpretation relevant to your domain. This should specifically address variant interpretation and feedback and your interaction with the cross-cutting Validation and Feedback domain.* ***[34 words]***

In all collaborating centres the research laboratories are either integrated with diagnostic laboratories, or work closely with them. This has previously meant that research findings can be translated into diagnostic tests with great efficiency.

Beneficiaries. *How will the research benefit patients and healthcare institutions including the NHS, other researchers in the field? Are there other likely beneficiaries?* ***[57 words]***

Children with liver disease, and their families deserve to understand the cause of their illness. This will inevitably lead to clinical trails and eventually better treatment. If nothing else the identification of genetic variants underlying different liver diseases has already saved huge expense in excluding other disorders through conventional testing. This will continue to reduce diagnostic costs.

Commercial exploitation. *(Where relevant to your GeCIP) Genomics England has a very explicit intellectual property policy. We and other funders need to know if the proposed research likely to generate commercially exploitable results. Do you have commercial partners in place?*

References. *Provide key references related to the research you set out.*

A gene encoding a liver specific ABC transporter gene is mutated in progressive familial intrahepatic cholestasis. Sandra S. Strautnieks, Laura N. Bull, Alexander S. Knisely, et al. Nature Genetics 20, 233-238 (1998)

Mutations in VPS33B, encoding a regulator of SNARE-dependent membrane fusion, cause arthrogryposis - renal dysfunction - cholestasis (ARC) syndrome. Paul Gissen, Colin A. Johnson, Neil V. Morgan, et al. Nature Genetics 36, 400-404 (2004)

Severe bile salt export pump deficiency: 82 different ABCB11 mutations in 109 families. Sandra S. Strautnieks, Jane A. Byrne, Ludmila Pawlikowska, et al. Gastroenterology 134, 1203-1214 (2008)

Missense mutations and single nucleotide polymorphisms in ABCB11 impair BSEP processing and function or disrupt pre-mRNA splicing. JA Byrne, SS Strautnieks, G Ihrke, et al. Hepatology 49, 553-567 (2009)

Differences in presentation and progression between severe FIC1 and BSEP deficiencies. Ludmila Pawlikowska, Sandra Strautnieks, Irena Jankowska, et al. Journal of Hepatology 53, 170-178 (2010)

Mutations in TJP2 cause progressive cholestatic liver disease. Melissa Sambrotta, Sandra Strautnieks, Efterpi Papouli, et al. Nature Genetics 46, 326-328 (2014)

Massive gene amplification drives pediatric hepatocellular carcinoma caused by bile salt export pump deficiency. Fabio Iannelli, Agnese Collino, Shruti Sinha, et al. Nature Communications doi:10.1038/ncomms4850 (2014)

Hepatocellular Carcinoma Associated with Tight-Junction Protein 2 Deficiency. Shengmei Zhou, Paula M. Hertel, Milton J. Finegold, et al. Hepatology DOI: 10.1002/hep.27872 (2015)

Data requirements

Data scope. *Describe the groups of participants on whom you require data and the form in which you plan to analyse the data (e.g. phenotype data, filtered variant lists, VCF, BAM). Where participants fall outside the disorders within your GeCIP domain, please confirm whether you have agreement from the relevant GeCIP domain. (max 200 words)*

We will require full genotype (BAM and VCF files) and full phenotype data for all patients matching the HPO terms that we specify for neurodevelopmental disorders and congenital anomalies.

Once we have access to the data and have a better understanding of where the potential productive overlaps lie with respect to other GECIP domains, we will seek agreement from the relevant domains

Data analysis plans. *Describe the approaches you will use for analysis. (max 300 words)*

Gene discovery analyses will be informed by 'burden' analyses that identify specific classes of variation (e.g. de novo mutations predicted to result in protein truncation, or biallelic mutations that ablate transcription factor binding sites within annotated regulatory elements) that are enriched within patients compared to controls, and/or unaffected parents. These analyses are particularly informative for exploring the non-coding space.

Gene discovery analyses will be conducted using robust statistical approaches that assess the significance of enrichment within individual functional elements of specific classes of variants (e.g. de novo truncating mutations in gene X). Rigorous criteria will be used to determine whether individual elements exceed genome-wide significance thresholds, as exemplified by recent dominant and recessive gene discovery papers from the DDD study.

Key phenotype data. *Describe the key classes of phenotype data required for your proposed analyses to allow prioritisation and optimisation of collection of these. (max 200 words)*

Comprehensive and consistent HPO phenotyping, especially of rare and unusual phenotypes.

Clinical photos are also useful for identifying characteristic dysmorphism associated with individual syndromes.

In previous analyses, we have shown that the recruiting clinician's suspicion of a particular phenotypically defined syndrome can also be highly informative. In other words, "does this patient's presentation remind you of a defined genetic disorder?" With the answer limited to disorders with MIM numbers.

Alignment and calling requirements. *Please refer to the attached file (Bioinformatics for 100,000 genomes.pptx) for the existing Genomics England analysis pipeline and indicate whether your requirements differ providing explanation. (max 300 words)*

Based on our current experience of analysing WES and WGS datasets, no single variant calling workflow ever identifies all pathogenic variants (e.g. structural variants, mosaic variants). Therefore, only once we have access to the data will we be able to assess how much additional value (e.g. improved sensitivity for detecting causal variants) might be yielded by implementing additional calling algorithms. Our initial impressions are that the current pipeline might best be supplemented by including calling of UPD, mosaic sequence and structural variants and complex structural variants. If we can demonstrate that inclusion of some or all of these additional algorithms is valuable within the DOD sub-domain then we will work with the Genomics England Bioinformatics team to roll out these supplementary variant calling algorithms across all individuals and not just those in this sub-domain.

Providing DP4 information within the VCF files themselves will reduce the amount of computation that needs to be performed on the BAM files themselves, and therefore would lower the compute requirements for this sub-domain.

Additional centralised datasets that would also be of use and would lower the compute requirements for the project are:

- Allele frequency data for all variants detected within the Genomics England data.
- Estimated variant-specific error rates for every base in the genome.
- Comprehensive results of experimental validation of diagnostic mutations performed by GMCs - both true positive and false positive mutations are valuable.

Tool requirements and import. *Describe any specific tools you require within the data centre with particular emphasis on those which are additional to those we will provide (see attached excel file List_of_Embassy_apps.xlsx of the planned standard tools). If these are new tools you must discuss these with us. (max 200 words)*

The current list of tools to be provided covers our requirements, as currently envisaged. If further tools are required we will work with Genomics England to get these installed.

Data import. *Describe the data sets you would require within the analysis environment and may therefore need to be imported or accessible within the secure data environment. (max 200 words)*

We will need to import summary genotype and phenotype data from external genome-wide (e.g. WES and WGS) datasets, to enable meta-analyses. This need not require the import of BAM files or entire VCF/GVCF files, but, more likely lists of candidate genes/variants and counts of classes of variants in specific functional elements.

Once greater clarity is provided on how the export of summary data from the Genomics England Data Centre will managed, it will be possible to plan these meta-analyses in greater detail.

Computing resource requirements. *Describe any analyses that would place high demand on computing resources and specific storage or processing implications. (max 200 words)*

The most major computing requirements are likely to be those that entail applying additional variant calling algorithms. As described above, only once we have access to data can we evaluate what additional variant calling algorithms it would make sense to implement. By contrast, other analyses (e.g. gene discovery) will have much lower storage or processing implications. Again, clarity over what level of summary data can be exported is required to inform whether planned meta-analyses are conducted within or external to the Genomics England Data Centre.

Omics samples

Analysis of omics samples. *Summarise any analyses that you are planning using omics samples taken as part of the Project. (max 300 words)*

We have successfully conducted a pilot project (see below) to evaluate the additional utility of incorporating proteomics, metabolomics, epigenomics and transcriptomics data in addition to WGS data for identifying diagnostic variants. This pilot project involved 170 patients, and was funded from a combination of resources including NIHR TRC-RD and the GOSH Biomedical Research Centre.

The use of Omics samples banked through GEL provides a vital resource to build out our NIHR-funded Omics programme for Rare Diseases (HIGH5). This study has helped to build the SoPs and data integration tools for the pilot mentioned above. It's purpose is to:

- Combine genomic, clinical and molecular phenotype information in patients with rare diseases
- Identify biomarkers
    - Disease progression – specific complications
    - Response to treatment
    - Stratify patients for appropriate / new treatments
- Identify key molecular pathways and modifiers
    - Novel targets for therapy

HIGH5 rare disease cohorts recruited are:

- Wilm's Tumour
    - Identify prognostic markers for tumour severity / relapse
- Juvenile Dermatomyositis
    - Molecular characterisation of disease subgroups
    - Stratify groups for trials for new therapies
- Mitochondrial disease
    - Markers for development of disease complications
    - Early prediction of cardiac or neurological phenotypes
- Very Early Onset Inflammatory Bowel Disease
    - Molecular characterisation of disease pathophysiology
    - Identify markers for treatment / surgical options
    - Silver Russell Syndrome
    - Molecular mechanisms in growth restriction
- Usher Syndrome
    - Understanding of molecular pathology and mechanisms behind phenotypic variability
- Bardet-Biedl Syndrome
    - Understand the basis of variable disease penetrance and expressivity found even within multiplex families
    - Search for modifiers
    - Stratify patients to potential treatment categories

MultiOmics study of these conditions has generated over 70TB of data including genomics, proteomics and transcriptomics datasets. The pilot has allowed us to generate unique SoPs for optimal sample acquisition, data generation and storage. We have assembled a cadre of bioinformaticians expert in Omics dataset integration and clinical interpretation. The 5 year renewal of the GOSH BRC in 2017 has allowed us to expand this programme beyond the pilot and we are currently reviewing expressions of interest that utilise GEL participants for whom deep phenotyping could benefit refined diagnoses, stratification or determination of therapeutic targets.

Data access and security	
GeCIP domain name	Paediatrics
Project title *(max 150 characters)*	**Clinical Interpretation of Genomics England 100,000 Genomes in Children with Rare Diseases**

Applicable Acceptable Uses. *Tick all those relevant to the request and ensure that the justification for selecting each acceptable use is supported in the 'Importance' section (page 3).*

☐ *Clinical care*

☐ *Clinical trials feasibility*

X *Deeper phenotyping*

☐ *Education and training of health and public health professionals*

☐ *Hypothesis driven research and development in health and social care - observational*

☐ *Hypothesis driven research and development in health and social care - interventional*

X *Interpretation and validation of the Genomics England Knowledge Base*

☐ *Non hypothesis driven R&D - health*

☐ *Non hypothesis driven R&D - non health*

☐ *Other health use - clinical audit*

☐ *Public health purposes*

☐ *Subject access request*

X *Tool evaluation and improvement*

Information Governance

X The lead and sub-leads of this domain will read and signed the Information Governance Declaration form provided by Genomics England and will submit by e-mail signed copies to Genomics England alongside this research plan.

Any individual who wishes to access data under your embassy will be required to read and sign this for also. Access will only be granted to said individuals when a signed form has been processed and any other vetting processes detailed by Genomics England are completed.

Other attachments

Attach other documents in support of your application here including:

- a cover letter (optional)
- CV(s) from any new domain members which you have not already supplied (required)
- other supporting documents as relevant (optional)