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

Application Summary	
GeCIP domain name	**Metabolic and Endocrine Disease**
Project title *(max 150 characters)*	Whole genome sequencing to improve diagnosis and management of inherited metabolic and endocrine disorders

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

Our major objectives are to use data from high-throughput whole genome sequencing to advance the understanding of the aetiology and heterogeneity of a range of important inherited metabolic and endocrine syndromes where full understanding is currently lacking. We will gain insights into disease mechanism and to novel therapeutic opportunities.

Specific aims:

1. We will develop and implement bioinformatics algorithms and pipelines to categorise rare, severe inherited metabolic and endocrine disorders into homogeneous phenotypic groups. This will improve gene identification, assist genotype-phenotype correlations and future stratification prior to intervention studies.
2. We will identify novel causative genes and establish clinically useful risk scores to facilitate identification of novel disease genes, genetic risks and modifying factors to enable NHS diagnostic testing, prediction of disease onset, penetrance and clinical severity.
3. We will use our extensive experience in phenotyping in cellular and animal model systems to study disease mechanism.
4. We will recall and invite affected individuals for further deep metabolic and endocrine phenotyping to better understand how their diseases impact on their in vivo physiology.
5. We will use our extensive nexus of collaborative relations with the biotech and pharm industry to work together to develop novel approaches to therapy of these disorders.
6. We will work together with other GeCIPs and GMCs to train the next generation of scientists, technologists and clinicians in genomic medicine.

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)*

Disorders of metabolism impair the way our cells handle the natural chemical substances including nutrients that are delivered to them. In endocrine diseases the way that cells in the body communicate with each other through the bloodstream malfunctions. Here are many hundreds of rare inherited metabolic and endocrine diseases that have serious effects on health and quality of life and together they bring a large burden of ill health to people in the UK. Many are often diagnosed late and a large proportion currently have no highly effective therapy.

Over the next two years, our linked Genome Medicine Centres (GMCs) will collect DNA and other samples on over 8,000 families with rare inherited metabolic and endocrine diseases for whole genome sequencing. Our GeCIP will optimise genome validation and interpretation, maximise the research potential of whole genome sequencing and improve our understanding of the genetic architecture of rare endocrine and metabolic disease, exploiting this information to identify key pathways forming the foundations for future mechanistic studies and novel therapeutic interventions. We will also train the next generation of NHS scientists, analysts and clinicians in genomic medicine to sustain this thriving healthcare initiative.

Our GeCIP research plan is to build on our well-established NHS, research and collaborative infrastructure to enhance the clinical interpretation and validation of whole genome sequencing. We will collaborate widely with other GeCIPs and form early partnerships with industry to convert

genetic findings into therapeutic benefit. This will ensure that future NHS patients with metabolic and endocrine disorders receive unrivalled diagnostic, genome driven precision healthcare.

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)*

To maximise the collection of rare inherited metabolic and endocrine disorders, over the next two years and beyond, the GeCIP will work closely with the GMCs to enable sequencing from over 8,000 individuals as trios and families. The GeCIP will optimise validation and clinical interpretation; in appropriate cases that information will be taken into mechanistic studies in cells and disease models to better understand the mechanism of disease. We will constantly seek opportunities to translate new discovery into new approaches to treatment.

Experimental Aims

1. Improving Primary Clinical Phenotyping

Using the disease related information provided by recruiting GMC we will develop and collaborate with the cross-cutting electronic records and advanced analytic domains to develop data capture and similarity scoring algorithms. We will group rare metabolic and endocrine diseases into phenotypic entities with meaningful homology entities into homogeneous phenotypic groups, based on clinical features, inheritance and metabolic and endocrine investigations. This will improve our efficiency in the identification of disease genes, genetic risk and modifying factors.

2. Identification of disease causing variants, filtering strategies and the creation of resources for variant prioritisation and the development of new NHS diagnostic tests

This work will share commonalities across all rare disease GeCIPs. After annotation and filtering there will be a significant number of possible variants that remain. Some will be likely pathogenic such as exonic de-novo changes and segregating exonic deletions or potential splice site variants. There will also be many other possibly pathogenic non-synonymous changes. These variants will require the use and creation of resources for prioritisation such as: developing high quality control variant databases, defining the transcribed portion and regulatory/non-coding regions of the human genome within the human CNS and muscle tissue, EQTL and transcriptomic analyses of patient-derived samples and the creation of integrated web-based issue specific resource for variant interpretation. A further important output from this work will be the development of new diagnostic tests.

3. Investigation of disease mechanisms in vitro, in cells and in model organisms

Gene discovery will lead to the identification of novel pathways that will be investigated and modelled to advance our understanding of disease pathophysiology and mechanisms. Depending on the putative gene, a wide range of different biochemical and cellular approaches are available to validate pathogenicity of novel disease genes. We have considerable existing expertise in biochemistry cell biology and animal models within the Domain (Chatterjee, O'Rahilly, Savage, Beales, Dunkel, Gissen, Gloyn, Thakker) and will also collaborate with cross-cutting GeCIPs when appropriate.

4. Training (see research plan)

We plan to train clinical fellows through PhD fellowships, basic scientists through PhD studentships, genomic counsellors within associated GMC's, specialist nurses, NHS genetics technologists where a higher degree such as an MSc can be taken as part of this training and NHS clinical scientists within the STP and HSST programmes.

Expected outputs

1. The systematic capture of phenotypic data in a wide range of rare metabolic and endocrine diseases which will underpin future efforts in stratified medicine.
2. New genes, genetic risk and disease modifying factors for metabolic and endocrine leading to a

new NHS diagnostic tests for these disorders. 3. The identification of novel pathways and advanced understanding of disease pathophysiology. 4. The training of a number of clinicians, nurses, scientists and technologists in genomic medicine.	
Expected start date	**01.06.2016**
Expected end date	**31.05.2021**

Lead Applicant(s)	
Name	Professor Sir Stephen O'Rahilly
Post	Professor of Clinical Biochemistry and Medicine, Director of MRC Metabolic Diseases Unit
Department	WT-MRC Institute of Metabolic Science, Metabolic Research Laboratories
Institution	University of Cambridge
Current commercial links	Pfizer, AstraZeneca, MedImmune, ERX - Membership of scientific advisory boards in the area of pre-clinical to early phase clinical drug development. Aegerion – Chair of international expert panel for compassionate use of leptin therapy.

Administrative Support	
Name	Carole Smith
Email	cjs54@medschl.cam.ac.uk
Telephone	01223 336855

Subdomain leads		
Name	**Subdomain**	**Institution**
Professor Stephen O'Rahilly	**Adipose and Lipid Disorders**	WT-MRC Institute of Metabolic Science, Metabolic Research Laboratories, University of Cambridge
Professor Andrew Hattersley (Professor of Molecular Medicine)	**Diabetes**	Institute of Biomedical and Clinical Science, University of Exeter
Professor Sian Ellard (Professor of Molecular Genetics & Genomic Medicine)	**Hypoglycaemia**	Institute of Biomedical and Clinical Science, University of Exeter
Professor Paul Gissen (Wellcome Trust Senior Clinical Fellow and Consultant in Paediatric Metabolic Diseases)	**Inborn errors**	UCL Institute of Child Health
Professor Krishna Chatterjee (Professor of Endocrinology)	**Endocrine disorders**	WT-MRC Institute of Metabolic Science, Metabolic Research Laboratories, University of Cambridge
Dr Ewan Birney (Senior adviser) / Dr Brian Lam (Cambridge lead)	**Bioinformatics**	European Bioinformatics Institute, Cambridge / WT-MRC Institute of Metabolic Science, Metabolic Research Laboratories

Detailed research plan

Full proposal (total max 1500 words per subdomain)	
Title *(max 150 characters)*	Diabetes

Importance.

Defining the genetic aetiology of monogenic diabetes is important to improve diagnosis and clinical care of patients and also gives critical scientific insights. For the commonest causes of MODY or neonatal diabetes there are specific treatment requirements meaning that diagnostic testing is now part of routine clinical care. The majority of monogenic diabetes result in beta-cell dysfunction: the definition of genes has defined pathways which are critical in beta-cell development, beta-cell destruction or beta-cell dysfunction giving biological insights of potential relevance for Type 2 diabetes. There are still many patients (c30-40%) with phenotypes suggestive of monogenic diabetes (neonatal diabetes, familial diabetes and developmental syndromes including diabetes) that do not have mutations in the known genes showing there are still monogenic forms of diabetes to be discovered.

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.*

Aims: (1) To exploit the opportunities of genome sequencing to identify new genes and regulatory regions for known genes; (2) link genetic information to phenotype and other data from unique UK cohorts.

Specific objectives and outline experimental approaches:

A) To define potential mutations in novel genes and to define non-coding mutations in known genes for monogenic diabetes we will use an approach that relies upon the investigation of variant segregation with disease as well as the nature of the variant. In likely heterozygous disorders the most productive approach is to look for de novo mutations or when there are large multi-generation families to look for shared heterozygous variants in family members with the same phenotype who are distantly related. In recessively inherited conditions we will look for homozygous or compound heterozygous variants in affected subjects that are heterozygous or not present in unaffected family members. Assessment of the likely pathogenicity of the variant will use an integrated bioinformatics approach which will include assessing conservation, likely biological impact of the mutation, presence in unaffected subjects in large datasets etc. In all cases co-segregation within families and replication in additional families will be critical to defining novel genes or novel non coding mutational mechanisms.

B) Investigation of mutation mechanism in instances such as potential splice site, structural variation and copy number variation.

C) Phenotyping of genetically defined patients including biochemistry, metabolomics etc to identify potential biomarkers.

D) Functional effects. The level of investigation will depend on the putative gene function and likely pathogenicity of the variant. Were ever possible we will work with leading international investigators in the gene of interest.

E) Models of disease will be facilitated by discussing novel genetic findings with researchers working on developing experimental animal models and sharing of samples including stem cells

F) With new genes we will investigate the genome data in other GeCIP domains (such as paediatric and neurodevelopmental disorders) and collaborate internationally to identify further patients especially when they are associated with multi-system syndromes

Major Challenge
The major challenge is to ensure that sufficient numbers of appropriate patients with diabetes are referred to the 100,000 genomes project to allow identification of novel genetic aetiologies. We need referrals to have a phenotypes strongly suggestive of monogenic diabetes (such as neonatal diabetes, familial early-onset diabetes or multi-system developmental disorders including diabetes) and for the known genes to have been excluded by genetic sequencing.

This challenge will be mitigated by ensuring appropriate referrals by using the National Network of regional Genetic Diabetes Nurses (run by Shepherd, Ellard, & Hattersley from the GeCIP.and making doctors aware that they have suitable patients when mutations in the known genes are not found in samples sent for diagnostic referrals.

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.*

The major collaborations with other GeCIPs will be with the cross cutting Paediatrics GeCIP.

The major academic collaborations will be with research investigators within the GeCIP (both in the UK (**Ellard, Hattersley, Gloyn, McCarthy, Owen**) and abroad (**Njolstad, Bell**)) who have collections of potential monogenic patients in whom the known genes have been excluded. This will facilitate the crucial early replication when a potential gene is identified and to help assess the phenotypic variation seen with patients with mutation in the gene.

Training. *Describe the planned involvement of trainees in the research and any specific training that will form part of your plan.*
We will work closely with the cross-cutting GeCIP training domain. The sites represented in this GeCIP comprise leading biomedical research centres, each with strong doctoral and post-doctoral training programs. In addition to the continuation of site-based training, the GeCIP would promote integration and cross-training, through workshops organised by GeCIP direct and through partnership with the NHS postgraduate and undergraduate training programmes. In addition, the GeCIP approach to research would be used to develop a common doctoral training program across sites and across subdemains

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

The investigators have a track record of joint publications (see below) and collaboration, with previous/current joint grants from the Wellcome Trust, MRC and NIHR to **Michael Weedon, Andrew Hattersley**, **Sian Ellard, Anna Gloyn, Jorge Ferrer. Mark McCarthy, Catherine Owen, Tim Frayling, Karen Temple, Deborah McKay Khalid Hussain** and **Julian Shield**. **Sian Ellard** leads the UK NHS genetic testing service for MODY and Neonatal Diabetes and with Andrew Hattersley

Michael Weedon, and **Sarah Flanagan** and other collaborators has an international reputation for their MODY and Neonatal diabetes genetics having described 12 novel genes. They have worked with **Jorge Ferrer** and **Innes Barroso** to gain functional insights into genetic discoveries . We have long term (12-20 year) international collaborations with **Graeme Bell** Chicago and **Pal Njolstad** Bergen.

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.*

This GeCIP builds a closer working relationship between diabetes clinicians, clinical geneticists and research teams. This provides an immediate benefit to patients and families in enhancing the quality of information, the interpretation of results and advice during and after genetics testing,

It will enable more sophisticated modelling of disease risk, explain phenotypic variation within and between families and establish the clinical significance of variants found in monogenic diabetes genes.

In the longer term, the GeCIP will establish the most effective diabetes therapies for specific monogenic forms of diabetes. These may come from established glucose lowering therapy or may require specific therapies once the aetiology and hence pathophysiology is known eg anti-Stat3 therapy (developed for Cancer) in patients with diabetes from activating mutations of Stat3.

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

Patients and Clinicians –Improved diagnosis, understanding if apparently unrelated clinical features, guidance on optimal therapy,

Family members predictive and prenatal testing.

NHS organisations cost savings. through appropriate testing of patients, appropriate therapy and avoidance of inappropriate diagnostic testing

Pharmaceutical companies –

Will identify new disease pathways in diabetes that could lead to novel drug targets

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

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

1: De Franco E, **Flanagan SE**, Houghton JA, Lango Allen H, **Mackay DJ, Temple IK,Ellard S, Hattersley AT.** The effect of early, comprehensive genomic testing onclinical care in neonatal diabetes: an international cohort study. Lancet. 2015 Sep 5;386(9997):957-63. PMID: 26231457;

2: Raimondo A, et al ; **Flanagan SE**, Van De Bunt M, **Hattersley AT, Gloyn AL, Ellard S**; International NDM Consortium. Phenotypic severity of homozygous GCK mutations causing neonatal or childhood-onset diabetes is primarily mediated through effects on protein stability. Hum Mol Genet. 2014 Dec 15;23(24):6432-40 PMID: 25015100;

3: **Flanagan SE,** De Franco Eet al atch AM, **Ellard S, Hattersley AT**. Analysis of transcription factors key for mouse pancreatic development establishes NKX2-2 and MNX1 mutations as causes of neonatal diabetes in man. Cell Metab. 2014 Jan 7;19(1):146-54 PMID: 24411943;

4: **Weedon MN**, Cebola I, Patch AM, **Flanagan SE**, De Franco E et al , **Hussain K,** Marsh P, Vallier L, , **Ellard S, Ferrer J, Hattersley AT**. Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. Nat Genet. 2014 Jan;46(1):61-4. PMID: 24212882;

5. **Flanagan SE**, Haapaniemi E, et al Morgan NG, **Ellard S, Hattersley AT**. Activating germline mutations in STAT3 cause early-onset multi-organ autoimmune disease. Nat Genet.2014;46(8):812-4 PMID: 25038750;

6: **Weedon MN, Ellard S**, et al Savage DB, **O'Rahilly S**, Kos K, Loeb LA, **Semple RK, HattersleyAT.** An in-frame deletion at the polymerase active site of POLD1 causes a multisystem disorder with lipodystrophy. Nat Genet. 2013;45(8):947-50. PMID: 23770608;

7: Garin I, Edghill ELet al K**, Flanagan SE**, et al Castaño L**, Ellard S, Ferrer J,** Perez de Nanclares G,**Hattersley AT.** Recessive mutations in the INS gene result in neonatal diabetes through reduced insulin biosynthesis. Proc Natl Acad Sci U S A. 2010;107(7):3105-10. PMID:20133622;

8**: Mackay DJ**, Callaway JL, , et al , **Ellard S, Hattersley AT**, Robinson DO, **Temple IK.** Hypomethylation of multiple imprinted loci in individuals with transient neonatal diabetes is associated with mutations in ZFP57. Nat Genet. 2008 40(8):949-51. PMID: 18622393

9: Pearson ER, Flechtner I, **Njølstad PR**, Malecki MT, **Flanagan SE,** Larkin B, Et al , **Shield J,** Robert JJ, , **Ellard S**, Søvik O, Polak M, **Hattersley AT** Switching from insulin to oral sulfonylureas in patients with diabetes due to Kir6.2 mutations. N Engl J Med. 2006 3;355(5):467-77 PMID: 16885550.

10: **Gloyn AL**, Pearson ER, Antcliff et al , **Frayling TM, Temple IK,Mackay D, Shield JP**, et al **Ellard S,** Njølstad PR, Ashcroft FM, **Hattersley AT**. Activating mutations in the gene encoding the ATP-sensitive potassium-channel subunit Kir6.2and permanent neonatal diabetes. N Engl J Med. 2004 350(18):1838-49. PMID: 15115830.

Full proposal (total max 1500 words per subdomain)	
Title *(max 150 characters)*	Adipose and Lipid 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).*

This subdomain comprises

1) Familial early-onset obesity
2) Familial extreme low body weight
3) Lipodystrophic syndromes
4) Disorders of triglyceride metabolism

These are all illnesses sin which severe, early-onset forms cause considerable morbidity and mortality and where much of the genetic aetiology remains unknown. They also represent illnesses which provide an opportunity to gain biological insights into pathways of great potential relevance for common disorders of major public health significance, such as obesity, insulin resistance, diabetes, hyperlipidaemia and cardio-metabolic risk.

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.*

Objectives: (1) To exploit the opportunities of genome sequencing to identify new genes, disease pathways and risk factors in known genes; (2) link to multiple 'omic samples and phenotype data from unique UK cohorts; and (3) directly inform new therapeutic strategies.

Specific aims:

A) To interpret mutations we will use an approach that relies upon the investigation of variant properties and segregation with disease, in discovery and replication case-sets, spanning classical phenotypes and phenotypic diversity.
B) Investigation of mutation mechanism in instances such as potential splice site, structural variation and copy number variation.
C) With new genes we will investigate the genome data in other GeCIP domains (such as paediatric and neurodevelopmental disorders) and collaborate internationally to identify further patients.
D) Functional effects. The level of investigation will depend on the likely pathogenicity of the variant and the putative gene function.
E) Phenotyping of genetically defined series. Core phenotyping will build a database record of the major functional bioenergetics, metabolic and lipidomics domains.
F) Models of disease will be facilitated by long term acquisition and sharing of samples including stem cells

The GeCIP would coordinate the subdomain's portfolio of genetic research, drawing together for the first time a comprehensive knowledge base of UK studies of genetic-phenotyping, genetic screening, and genetic-stratification in trials. This coordination would seek to develop common core datasets and facilitate material transfer, within and across subdomains.

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 plan to work closely with a number of cross-cutting GeCIPs to help with the aims of this research plan. We have also established international collaborations to replicate or repudiate identified variants. We have extensive links with industry including AstraZeneca/Medimmune whose world headquarters is now being established in Cambridge.

Training. *Describe the planned involvement of trainees in the research and any specific training that will form part of your plan.*
We will work closely with the cross-cutting GeCIP training domain. The sites represented in this GeCIP comprise leading biomedical research centres, each with strong doctoral and post-doctoral training programs. In addition to the continuation of site-based training, the GeCIP would promote integration and cross-training, through workshops organised by GeCIP direct and through partnership with DPUK. In addition, the GeCIP approach to research would be used to develop a common doctoral training program across sites.

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

Sadaf Farooqi and Stephen O'Rahilly lead the team that has discovered the majority of genes currently known to cause human early-onset obesity. They have followed on from gene discovery to enhance mechanistic understanding at the biochemical, cell biological, animal model and whole human level. They were the first to use leptin therapeutically in humans and are currently collaborating with industry in targeted pharmacotherapy for other monogenic forms of obesity. **David Savage and Steve O'Rahilly** have led or participated in the discovery of the majority of genes causing human forms of lipodystrophy, and, as with obesity, have followed on with deep biological studies in cells, animals and humans. On the basis of these discoveries they developed and now run a nationally commissioned NHS service for patients with lipodystrophy. **Steve Humphries** is an internationally leading expert on the genetics of hyperlipidaemia. **Gerome Breen** leads major genetic studies in familial anorexia nervosa.

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.*

This GeCIP builds a closer working relationship between clinicians, clinical geneticists and research teams. This provides an immediate benefit to patients and families in
enhancing the quality of information and advice during and after genetics testing, and in the interpretation of results. It will enable more sophisticated modelling of disease risk, explain phenotypic variation within and between families.
In the longer term, the GeCIP will lead to novel and effective therapies via the clinically validated model systems in preclinical pathways, from cell biology and high-throughput screening programs in early preclinical models through to optimised selection of candidate therapies and target populations in clinical trials for efficacy.

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

Clinicians and patients - Families; diagnosis, predictive, prenatal testing. NHS organisations and cost savings.
Pharmaceutical companies - New disease pathways, targets, stratified patient groups, pre-clinical research models, clinical Trials.

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 have existing extensive collaborative interactions with industry and biotech including Astrazeenca, Medimmune, Pfizer, Rhythm pharmaceuticals which will only grow more substantial as GeCIP data emerges

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

UK10K Consortium et al . The UK10K project identifies rare variants in health and disease. **Nature**. 2015 Oct 1;526 (7571):82-90. . PubMed PMID: 26367797.

Simonds SE et al Leptinmediates the increase in blood pressure associated with obesity. **Cell.** 2014 Dec 4;159(6):1404-16. PubMed Central PMCID: PMC4259491.

Pearce LR et al KSR2 mutations are associated with obesity, insulin resistance, and impaired cellular fuel oxidation. **Cell.** 2013 Nov 7;155(4):765-77. PubMed PMID: 24209692

Shungin D et al New genetic loci link adipose and insulin biology to body fat distribution. **Nature.** 2015 Feb 12;518(7538):187-96. PubMed PMID: 25673412;

Payne F et al Mutations disrupting the Kennedy phosphatidylcholine pathway in humans with congenital lipodystrophy and fatty liver disease. **Proc Natl Acad Sci U S A.** 2014 Jun 17;111(24):8901-6. PubMed PMID: 24889630

Full proposal (total max 1500 words per subdomain)	
Title *(max 150 characters)*	Hyperinsulinism

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).*

Severe, early-onset forms of hyperinsulinism cause considerable morbidity and mortality and the genetic aetiology in up to 50% of cases is unknown. There is an opportunity to gain biological insights into pathways of great potential relevance for common disorders of major public health significance, such as diabetes and insulin resistance.

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.*

Objectives: (1) To exploit the opportunities of genome sequencing to identify new genes, disease pathways and risk factors in known genes; (2) link to multiple 'omic samples and phenotype data from unique UK cohorts; and (3) directly inform new therapeutic strategies.

Specific aims:

A) To interpret mutations we will use an approach that relies upon the investigation of variant properties and segregation with disease, in discovery and replication case-sets, spanning classical phenotypes and phenotypic diversity.
B) Investigation of mutation mechanism in instances such as potential splice site, structural variation and copy number variation.
C) With new genes we will investigate the genome data in other GeCIP domains (such as paediatric and neurodevelopmental disorders) and collaborate internationally to identify further patients.
D) Functional effects. The level of investigation will depend on the likely pathogenicity of the variant and the putative gene function.
E) Phenotyping of genetically defined series including metabolomics.
F) Models of disease will be facilitated by long term acquisition and sharing of samples including stem cells

The GeCIP would coordinate the subdomain's portfolio of genetic research, drawing together for the first time a comprehensive knowledge base of UK studies of genetic-phenotyping, genetic screening, and genetic-stratification in trials. This coordination would seek to develop common core datasets and facilitate material transfer, within and across subdomains.

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 plan to work closely with a number of cross-cutting GeCIPs to help with the aims of this research plan. We have also established international collaborations to replicate or repudiate identified variants.

Training. *Describe the planned involvement of trainees in the research and any specific training that will form part of your plan.*
We will work closely with the cross-cutting GeCIP training domain. The sites represented in this GeCIP comprise leading biomedical research centres, each with strong doctoral and post-doctoral training programs. In addition to the continuation of site-based training, the GeCIP would promote integration and cross-training, through workshops organised by GeCIP direct and through partnership with DPUK. In addition, the GeCIP approach to research would be used to develop a common doctoral training program across sites.

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

The investigators have a track record of joint publications (see below) and collaboration, with previous/current joint grants from the Wellcome Trust and MRC to **Sian Ellard, Khalid Hussain** and **Julian Shield**. **Sian Ellard** leads the UK NHS genetic testing service for hyperinsulinism and with **Sarah Flanagan** has an international reputation for their hyperinsulinism genetics. They have worked with **Khalid Hussain**, **Julian Shield, Rob Semple** and **Indi Banerjee** to gain new insights into genotype/phenotype correlations. We are pleased to be working with **Diva De Leon** from the internationally renowned CHI centre in Philadelphia, US.

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.*

This GeCIP builds a closer working relationship between clinicians, clinical geneticists and research teams. This provides an immediate benefit to patients and families in
enhancing the quality of information and advice during and after genetics testing, and in the interpretation of results. It will enable more sophisticated modelling of disease risk, explain phenotypic variation within and between families.
In the longer term, the GeCIP will lead to novel and effective therapies via the clinically validated model systems in preclinical pathways, from cell biology and high-throughput screening programs in early preclinical models through to optimised selection of candidate therapies and target populations in clinical trials for efficacy.

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

Clinicians and patients - Families; diagnosis, predictive, prenatal testing. NHS organisations and cost savings.
Pharmaceutical companies - New disease pathways, targets, stratified patient groups, pre-clinical research models, clinical Trials.

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

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

1: **Senniappan S**, Alexandrescu S, Tatevian N, Shah P, Arya V, **Flanagan S, Ellard S**, Rampling D, Ashworth M, Brown RE, **Hussain K**. Sirolimus therapy in infants with severe hyperinsulinemic hypoglycemia. N Engl J Med. 2014 Mar 20;370(12):1131-7. doi: 10.1056/NEJMoa1310967. PubMed PMID: 24645945.

2: Kapoor RR, **Flanagan SE**, Arya VB, **Shield JP, Ellard S, Hussain K**. Clinical and molecular characterisation of 300 patients with congenital hyperinsulinism. Eur J Endocrinol. 2013 Mar 15;168(4):557-64. doi: 10.1530/EJE-12-0673. Print 2013 Apr. PubMed PMID: 23345197; PubMed Central PMCID: PMC3599069.

3: **Flanagan SE**, Xie W, Caswell R, Damhuis A, Vianey-Saban C, Akcay T, Darendeliler F, Bas F, Guven A, Siklar Z, Ocal G, Berberoglu M, Murphy N, O'Sullivan M, Green A, Clayton PE, **Banerjee I,** Clayton PT, **Hussain K, Weedon MN, Ellard S**. Next-generation sequencing reveals deep intronic cryptic ABCC8 and HADH splicing founder mutations causing hyperinsulinism by pseudoexon activation. Am J Hum Genet. 2013 Jan 10;92(1):131-6. doi: 10.1016/j.ajhg.2012.11.017. Epub 2012 Dec 27. PubMed PMID: 23273570; PubMed Central PMCID: PMC3542457.

4: Kapoor RR, **Flanagan SE**, James CT, McKiernan J, Thomas AM, Harmer SC, Shield JP, Tinker A, **Ellard S, Hussain K**. Hyperinsulinaemic hypoglycaemia and diabetes mellitus due to dominant ABCC8/KCNJ11 mutations. Diabetologia. 2011 Oct;54(10):2575-83. doi: 10.1007/s00125-011-2207-4. Epub 2011 Jun 15. PubMed PMID: 21674179; PubMed Central PMCID: PMC3168751.

5: **Hussain K**, Challis B, Rocha N, Payne F, Minic M, Thompson A, Daly A, Scott C, Harris J, Smillie BJ, Savage DB, Ramaswami U, De Lonlay P, O'Rahilly S, Barroso I, **Semple RK**. An activating mutation of AKT2 and human hypoglycemia. Science. 2011 Oct 28;334(6055):474. doi: 10.1126/science.1210878. Epub 2011 Oct 6. PubMed PMID: 21979934; PubMed Central PMCID: PMC3204221.

6: Männikkö R, **Flanagan SE**, Sim X, Segal D, **Hussain K, Ellard S**, Hattersley AT, Ashcroft FM. Mutations of the same conserved glutamate residue in NBD2 of the sulfonylurea receptor 1 subunit of the KATP channel can result in either hyperinsulinism or neonatal diabetes. Diabetes. 2011 Jun;60(6):1813-22. doi: 10.2337/db10-1583. PubMed PMID: 21617188; PubMed Central PMCID: PMC3114383.

7: **Flanagan SE**, Patch AM, Locke JM, Akcay T, Simsek E, Alaei M, Yekta Z, Desai M, Kapoor RR, **Hussain K, Ellard S**. Genome-wide homozygosity analysis reveals HADH mutations as a common cause of diazoxide-responsive hyperinsulinemic-hypoglycemia in consanguineous pedigrees. J Clin Endocrinol Metab. 2011 Mar;96(3):E498-502. doi: 10.1210/jc.2010-1906. Epub 2011 Jan 20. PubMed PMID: 21252247; PubMed Central PMCID: PMC3100671.

8: **Flanagan SE**, Kapoor RR, **Banerjee I**, Hall C, Smith VV, **Hussain K, Ellard S**. Dominantly acting ABCC8 mutations in patients with medically unresponsive hyperinsulinaemic hypoglycaemia. Clin Genet. 2011 Jun;79(6):582-7. doi: 10.1111/j.1399-0004.2010.01476.x. PubMed PMID: 20573158; PubMed Central PMCID: PMC3375476.

9: **Hussain K, Flanagan SE**, Smith VV, Ashworth M, Day M, Pierro A, Ellard S. An ABCC8 gene mutation and mosaic uniparental isodisomy resulting in atypical diffuse congenital hyperinsulinism. Diabetes. 2008 Jan;57(1):259-63. Epub 2007 Oct 17. PubMed PMID: 17942822.

10: Pearson ER, Boj SF, Steele AM, **Barrett T**, Stals K, **Shield JP, Ellard S**, Ferrer J, Hattersley AT. Macrosomia and hyperinsulinaemic hypoglycaemia in patients with heterozygous mutations in the HNF4A gene. PLoS Med. 2007 Apr;4(4):e118. PubMed PMID: 17407387; PubMed Central PMCID: PMC1845156.

Full proposal (total max 1500 words per subdomain)

Title *(max 150 characters)*	Inborn Errors of Metabolism

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).*

Inborn Errors of Metabolism (IEM) is a very diverse group of disorders (currently more than 600 individual genes are known to cause IEMs) that can be caused by defects in biochemical pathways located in different intracellular compartments such as peroxisomes, lysosomes, mitochondria or be due to the individual organellar biogenesis defects. Many of these disorders result in neurological deficit, which can be prevented if the condition is diagnosed early. Moreover, in approximately 50% of cases with characteristic clinical picture of IEM the genetic aetiology remains unknown.
There is an opportunity to gain biological insights into pathways of great potential relevance for common disorders of major public health significance, such as neurodegenerative disorders and metabolic syndrome.

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.*

Objectives: (1) To exploit the opportunities of genome sequencing to identify new genes, disease pathways and risk factors in known genes; (2) link to multiple 'omic samples and phenotype data from unique UK cohorts; and (3) directly inform new therapeutic strategies.

Specific aims:
A) To interpret mutations we will use an approach that relies upon the investigation of variant properties and segregation with disease, in discovery and replication case-sets, spanning classical phenotypes and phenotypic diversity.
B) Investigation of mutation mechanism in instances such as potential splice site, structural variation and copy number variation.
C) With new genes we will investigate the genome data in other GeCIP domains (such as paediatric and neurodevelopmental disorders) and collaborate internationally to identify further patients.

D) Functional effects. The level of investigation will depend on the likely pathogenicity of the variant and the putative gene function.
E) Phenotyping of genetically defined series including metabolomics.
F) Models of disease will be facilitated by long term acquisition and sharing of samples including stem cells

The GeCIP would coordinate the subdomain's portfolio of genetic research, drawing together for the first time a comprehensive knowledge base of UK studies of genetic-phenotyping, genetic screening, and genetic-stratification in trials. This coordination would seek to develop common core datasets and facilitate material transfer, within and across subdomains.

Major Challenge
The major challenge is to ensure that sufficient numbers of appropriate patients with IEMs are referred to the 100,000 genomes project to allow identification of novel genetic aetiologies. We need referrals to have a phenotypes strongly suggestive of monogenic IEMs (patients with biochemical abnormalities in conjunction with the symptoms suggestive of IEM e.g. neurodegeneration, movement disorder or epilepsy) and for the known genes to have been excluded by genetic sequencing.

This challenge will be mitigated by ensuring appropriate referrals from the current clinical network of hospitals involved in looking after IEM patients: Birmingham and Manchester Children's Hospitals, Evelina Children's hospital and Addebrooke's hospital. Making doctors aware that they have suitable patients when mutations in the known genes are not found in samples sent for diagnostic referrals.

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 plan to work closely with a number of cross-cutting GeCIPs to help with the aims of this research plan. We have also established international collaborations to replicate or repudiate identified variants.

Training. *Describe the planned involvement of trainees in the research and any specific training that will form part of your plan.*
We will work closely with the cross-cutting GeCIP training domain. The sites represented in this GeCIP comprise leading biomedical research centres, each with strong doctoral and post-doctoral training programs. In addition to the continuation of site-based training, the GeCIP would promote integration and cross-training, through workshops organised by GeCIP direct and through partnership with DPUK. In addition, the GeCIP approach to research would be used to develop a common doctoral training program across sites.

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

The investigators have a track record of joint publications (see below) and collaboration, with previous/current joint grants from the Wellcome Trust, European Union and MRC to **Paul Gissen, Manju Kurian, Wyatt Yue, Peter Clayton, Derralyn Hughes, Tim Cox, Philippa Mills, Sara Mole** and **Kevin Mills**. **Peter Clayton** is the President of the Society for Study of Inborn Errors of Metabolism (the main international IEM professional body) and with **Tim Cox** and **Paul Gissen**

have an international reputation for their genetics of IEM. Most of the members of this subdomain have worked together to gain new insights into genotype/phenotype correlations.

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.*

This GeCIP builds a closer working relationship between IEM clinicians, clinical geneticists and research teams. This provides an immediate benefit to patients and families in enhancing the quality of information and advice during and after genetics testing, and in the interpretation of results. It will enable more sophisticated modelling of disease risk, explain phenotypic variation within and between families.
In the longer term, the GeCIP will lead to novel and effective therapies via the clinically validated model systems in preclinical pathways, from cell biology and high-throughput screening programs in early preclinical models through to optimised selection of candidate therapies and target populations in clinical trials for efficacy.

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

Clinicians and patients - Families; diagnosis, predictive, prenatal testing. NHS organisations and cost savings.
Pharmaceutical companies - New disease pathways, targets, stratified patient groups, pre-clinical research models, clinical Trials.

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

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

1: Tuschl K, Meyer E, Valdivia L, Zhao N, Dadswell C, Abdul-Sada AA, Hung C, Simpson M, Chong W, Jacques J, Woltjer R, Eaton S, Gregory A, Sanford L, Kara E, Houlden H, Cuno S, Prokisch H, Valletta L, Tiranti V, Younis R, Maher ER, Spencer J, Straatman-Iwanowska A, **Gissen P**, Selim L, Pintos-Morell G, Coroleu-Lletget W, Mohammad S, Yoganathan S, Dale R, Thomas M, Rihel J, Bodamer O, Enns C, Hayflick S, **Clayton PT**, **Mills P**, **Kurian M**, Wilson S. Mutations in SLC39A14 disrupt manganese homeostasis and cause childhood-onset parkinsonism-dystonia. Nature Communications (In Press).

2. Heywood WE, Camuzeaux S, Doykov I, Patel N, Preece RL, Footitt E, Cleary M, **Clayton P**, **Grunewald S**, Abulhoul L, Chakrapani A, Sebire NJ, Hindmarsh P, de Koning TJ, **Heales S**, Burke D, **Gissen P**, **Mills K**. Proteomic Discovery and Development of a Multiplexed Targeted MRM-LC-MS/MS Assay for Urine Biomarkers of Extracellular Matrix Disruption in Mucopolysaccharidoses I, II, and VI. Anal Chem. 2015 Dec 15;87(24):12238-44.

3. Stödberg T, McTague A, Ruiz AJ, Hirata H, Zhen J, Long P, Farabella I, Meyer E, Kawahara A,

Vassallo G, Stivaros SM, Bjursell MK, Stranneheim H, Tigerschiöld S, Persson B, Bangash I, Das K, Hughes D, Lesko N, Lundeberg J, Scott RC, Poduri A, Scheffer IE, Smith H, **Gissen P**, Schorge S, Reith ME, Topf M, Kullmann DM, Harvey RJ, Wedell A, **Kurian MA.** Mutations in SLC12A5 in epilepsy of infancy with migrating focal seizures. Nat Commun. 2015 Sep 3;6:8038. doi: 10.1038/ncomms9038.

4. Thomas AC, Williams H, Setó-Salvia N, Bacchelli C, Jenkins D, O'Sullivan M, Mengrelis K, Ishida M, Ocaka L, Chanudet E, James C, Lescai F, Anderson G, Morrogh D, Ryten M, Duncan AJ, Pai YJ, Saraiva JM, Ramos F, Farren B, Saunders D, Vernay B, **Gissen P**, Straatmaan-Iwanowska A, Baas F, Wood NW, Hersheson J, Houlden H, **Hurst J**, **Scott R**, Bitner-Glindzicz M, Moore GE, Sousa SB, Stanier P. Mutations in SNX14 cause a distinctive autosomal-recessive cerebellar ataxia and intellectual disability syndrome. Am J Hum Genet. 2014 Nov 6;95(5):611-21.

5. Hersheson J, Burke D, Clayton R, Anderson G, Jacques TS, **Mills P**, Wood NW, **Gissen P**, **Clayton P**, Fearnley J, **Mole SE**, Houlden H. Cathepsin D deficiency causes juvenile-onset ataxia and distinctive muscle pathology. Neurology. 2014 Nov 11;83(20):1873-5. doi: 10.1212/WNL.0000000000000981. Epub 2014 Oct 8.

6. Dai D, **Mills PB**, Footitt E, **Gissen P**, McClean P, Stahlschmidt J, Coupry I, Lavie J, Mochel F, Goizet C, Mizuochi T, Kimura A, Nittono H, Schwarz K, Crick PJ, Wang Y, Griffiths WJ, **Clayton PT**. Liver disease in infancy caused by oxysterol 7 α-hydroxylase deficiency: successful treatment with chenodeoxycholic acid. J Inherit Metab Dis. 2014 Sep;37(5):851-61.

7. **Mills PB**, Camuzeaux SS, Footitt EJ, **Mills KA**, **Gissen P**, Fisher L, Das KB, Varadkar SM, Zuberi S, McWilliam R, Stödberg T, Plecko B, Baumgartner MR, Maier O, Calvert S, Riney K, Wolf NI, Livingston JH, Bala P, Morel CF, Feillet F, Raimondi F, Del Giudice E, Chong WK, Pitt M, **Clayton PT**. Epilepsy due to PNPO mutations: genotype, environment and treatment affect presentation and outcome. Brain. 2014 May;137(Pt 5):1350-60.

8. Tuschl K, **Clayton PT**, Gospe SM Jr, Gulab S, Ibrahim S, Singhi P, Aulakh R, Ribeiro RT, Barsottini OG, Zaki MS, Del Rosario ML, Dyack S, Price V, Rideout A, Gordon K, Wevers RA, Chong WK, **Mills PB**. Syndrome of hepatic cirrhosis, dystonia, polycythemia, and hypermanganesemia caused by mutations in SLC30A10, a manganese transporter in man. Am J Hum Genet. 2012 Mar 9;90(3):457-66.

9. **Kurian MA**, **Gissen P**, Smith M, **Heales S Jr**, **Clayton PT**. The monoamine neurotransmitter disorders: an expanding range of neurological syndromes. Lancet Neurol. 2011 Aug;10(8):721-33.

10. **Mills PB**, Struys E, Jakobs C, Plecko B, Baxter P, Baumgartner M, Willemsen MA, Omran H, Tacke U, Uhlenberg B, Weschke B, **Clayton PT**. Mutations in antiquitin in individuals with pyridoxine-dependent seizures. Nat Med. 2006 Mar;12(3):307-9.

Full proposal (total max 1500 words per subdomain)	
Title *(max 150 characters)*	Rare Endocrine 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).*

This subdomain encompasses a range of endocrine disorders, in each of which members of the GeCIP have internationally recognised expertise and strong track records of genetic research:

5) Congenital adrenal hypoplasia
6) Familial or syndromic hypoparathyroidism
7) IUGR and IGF abnormalities
8) Disorders of sex development (*release pending*)
9) Early onset familial premature ovarian failure (*release pending*)
10) Congenital hypothyroidism (*submitted*)
11) Isolated hypogonadotrophic hypogonadism (*submitted*)
12) Primary hyperandrogenism (*submitted*)
13) Thyroid Hormone Resistance (*in preparation*)

The groups of disorders encompassed by this subdomain are of early onset, exact a large toll of morbidity and mortality, and in each case there is strong evidence for a significant burden of Mendelian disease whose basis has not yet been elucidated. Critically, many rare endocrine disorders if undiagnosed have lifelong ramifications, profoundly affecting development as well as health in later life, and many are amenable to hormone-based interventions, which may often be transformative. As this domain deals with diseases of hormone production or receptor-mediated hormone action, it is highly likely that identification of new disease genes in this area will not only provide an opportunity to gain novel insights into the mechanisms underlying physiological regulation, but also that new treatments will be suggested for some of these rare diseases in the short to medium term. As many of these endocrine disorders are sentinel features of more complex syndromes, and as many of the pathways implicated are of fundamental importance to many developmental processes, it is also expected that some of the genes identified will also be involved in other clinical disorders including cancer, cardiovascular disease and epilepsy. Rare endocrine disease thus represents a high yield, relatively low risk area for mutation discovery through the 100,000 Genomes project.

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.*

Objectives:

(1) To exploit whole genome sequencing to identify new disease genes and pathways
(2) To link to phenotypic data from unique established UK rare disease cohorts
(3) To link human genomic data to established "omic" datasets from both disease-relevant human populations and model systems.
(3) To assess the contribution of novel disease genes identified to commoner forms of endocrine disease
(3) To directly inform new therapeutic strategies.

Specific aims:

A) To interpret mutations we will use an approach that relies upon the investigation of variant

properties and segregation with disease, in discovery and replication case-sets, spanning classical phenotypes and phenotypic diversity. Many patients submitted via this subdomain will already have been prescreened using extended panels of known and candidate disease-specific genes by GeCIP investigators.
B) Investigation of mutation mechanism in instances such as potential splice site, structural variation and copy number variation. Key transcriptomic resources such as The Human Developmental Biology Resource (www.hdbr.org) will be exploited to assess expression of genes in which mutations are identified in human development and in human cells, with some diseases also benefiting from organ-specific transcriptomic atlases of gene expression during critical early human foetal development.
C) With new genes we will investigate the genome data in other relevant GeCIP domains (such as Paediatric, Endocrine tumours, Neurodevelopmental disorders and Cancer) and collaborate internationally to identify further patients.
D) Functional effects. The level of investigation will depend on the likely pathogenicity of the variant and the putative gene function. Functional analysis of non-exonic variants will include *in vitro* splicing assays for suspected splice variants, reporter gene studies for promoter changes and mRNA/cDNA analysis where possible.
E) Phenotyping of genetically defined series. Deep disease-specific phenotyping will build a database record of the major clinical, metabolic and cellular characteristsics of patients with different pathogenic mutations, in many diseases drawing on the emerging support and infrastructure afforded by the NIHR Rare Disease Translational Research Collaboration.
F) Models of disease will be facilitated by long term acquisition and sharing of samples including stem cells, and will benefit both from existing experimental models of organ-specific disease established by several of the GeCIP investigators and from exploitation of genome editing techniques to create new cellular and animal models

The GeCIP would coordinate the subdomain's portfolio of genetic research, drawing together for the first time a comprehensive knowledge base of UK studies of genetic-phenotyping, genetic screening, and genetic-stratification in trials. This coordination would seek to develop common core datasets and facilitate material transfer, within and across subdomains.

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.*

This GeCIP encompasses distinct disorders of many different endocrine glands. Endocrine perturbations usually have multisystem pleiotropic effects, while, conversely, they may originate from genetic defects affecting fundamental aspects of cell signalling or behaviour (e.g. DNA replication or repair, wnt signalling, cell growth and differentiation) that have complex extra-endocrine phenotypic consequences. In toto this means that strong collaboration within and beyond 100K Genomes will be essential to realise the full potential of the Rare Endocrine Disease Subdomain. The GeCIP investigators broadly align with the diseases so far approved, and in each area there are rich connections to healthcare provision, a wide network of academic collaborators nationally and internationally, to patients and patient advocacy groups, and in some diseases to pharmaceutical or other commercial partners. The GeCIP also has strong links to the NIHR Rare Disease Translational Research Collaboration (Endocrine lead O'Rahilly; deputy Semple) which, it is anticipated, will provide an important conduit for deep phenotyping studies downstream from genomic investigations. Within GeCIP investigators have an established network of national and international academic collaborators, healthcare providers in each of the disease areas indicated, with many also having close ties to patient advocacy groups, which in many cases were set up

with the support of GeCIP investigators. Within the 100K Genomes Project several key areas of collaboration are anticipated:

1. We plan to work closely with the cross-cutting GeCIPs on generic aspects of genomic analysis, functional study, electronic databases, feedback and reporting
2. There is highly likely to be strong collaboration with the paediatrics GeCIP, with whom there is considerable membership overlap, and given current knowledge we anticipate a strong interest in collaboration with cancer and cancer predisposition domains (both endocrine and non-endocrine), based on implication of many Mendelian endocrine disease genes in cell growth/differentiation/DNA damage repair). Collaborations with other GeCIPS will be driven by genetic findings and associated phenotypes, especially where these are complex and multisystem, for example affecting cardiovascular, neurological and gastroenterological function.

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

We will work closely with the cross-cutting GeCIP training domain. The sites represented in this GeCIP comprise leading biomedical research centres, each with strong doctoral and post-doctoral training programs. In addition to the continuation of site-based training, the GeCIP would promote integration and cross-training, through workshops organised by GeCIP direct and through partnership with DPUK. In addition, the GeCIP approach to research would be used to develop a common doctoral training program across sites.

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

Rare Endocrine Disease GeCIP members have an outstanding record of scholarship in each of the disease areas to be studied, showing extensive collaborative working, longstanding histories of serially securing competitive grant funding from all major agencies, and collectively play many national and international leadership roles in research consortia and clinical networks. This is all backed by a very strong track record of publication in leading international journals. Translational credentials are particularly strong, with most GeCIP members being practising clinicians as well as leading research programmes.

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.*

This GeCIP builds a closer working relationship between clinicians, clinical geneticists and research teams. This provides an immediate benefit to patients and families in enhancing the quality of information and advice during and after genetics testing, and in the interpretation of results. It will enable more sophisticated modelling of disease risk, explain phenotypic variation within and between families. In the longer term, the GeCIP will lead to novel and effective therapies via the clinically validated model systems in preclinical pathways, from cell biology and high-throughput screening programs in early preclinical models through to optimised selection of candidate therapies and target populations in clinical trials for efficacy. The assembled rare endocrine disease GeCIP members have long experience of variant interpretation, including both exonic and non exonic variants, and oligogenic disease. They have world leading credentials in their disease areas to interpret genetic variants in the light of detailed understanding of associated phenotypes

and model organism studies, and expect to add significant value to variant interpretation prior to clinical feedback.

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

Clinicians, patients, patient advocacy groups - Early diagnosis, predictive and prenatal testing, avoidance of unnecessary investigation, enhanced primary prevention of complications, possibility of new treatments.

NHS organisations: Cost savings through improved primary prevention and avoidance of unnecessary investigation.

Pharmaceutical companies - New disease pathways and targets, improved patient stratification, new pre-clinical research models, clinical trials.

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 potential for commercial exploitation of new genetic discoveries in endocrine disease is moderate, with mutation discovery in Mendelian endocrine disease often having wider relevance for other, non endocrine diseases (e.g. osteoporosis and epilepsy in hypoparathyroidism; cancer in Mendelian growth disorders; subfertility in pituitary/gonadal/adrenal disorders) Several investigators in this GeCIP have established commercial partnerships in diseases of interest (e.g. Thakker - GSK and NPS/Shire for hypocalcaemia/hypoparathyroidism; Krone – Diurnal for congenital adrenal hyperplasia). None of these collaborations are specifically based on the 100K Genomes project or its findings, however as a group the GeCIP is well placed to develop these existing partnerships as directed by new genomic findings.

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

General Mutation Discovery

- UK10K Consortium et al . The UK10K project identifies rare variants in health and disease. **Nature**. 2015 Oct 1;526 (7571):82-90.
- Taylor JC, et al. (2015). Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. **Nature Genetics**. 2015 47: 717-726.

Disease-specific Mutation Discovery and Characterisation (Selected)

Adrenal

- Novoselova TV et al NNT pseudoexon activation as a novel mechanism for disease in two siblings with familial glucocorticoid deficiency. **J Clin Endocrinol Metab**. (2015) 100:E350-4.
- Guran T et al Rare causes of primary adrenal insufficiency: Genetic and clinical characterization of a large nationwide cohort. **J Clin Endocrinol Metab**. 2015;101(1):284-92
- Hughes et al. MCM4 mutation causes adrenal failure, short stature, and natural killer cell deficiency in humans. **J Clin Invest**. (2012); 122(3):814-20.
- Arboleda VA et al. Mutations in the PCNA-binding domain of CDKN1C cause IMAGE Syndrome. **Nat Genet** 2012;44(7):788-792
- Meimaridou et al. Mutations in NNT, encoding nicotinamide nucleotide transhydrogenase, cause familial glucocorticoid deficiency. **Nature Genetics** (2012) 44:740-742.
- Parajes S et al A novel entity of clinically isolated adrenal insufficiency caused by a partially

inactivating mutation of the gene encoding for P450 side chain cleavage enzyme (CYP11A1). **J Clin Endocrinol Metab.** 2011 Nov;96(11):E1798-806.
- Lawson AJ et al Cortisone-reductase deficiency associated with heterozygous mutations in 11beta-hydroxysteroid dehydrogenase type 1. **Proc Natl Acad Sci U S A**. 2011 Mar 8;108(10):4111-6.
- Noordam C et al Inactivating PAPSS2 mutations in a patient with premature pubarche. **N Engl J Med.** 2009 May 28;360(22):2310-8.
- Metherell et al. Mutations in MRAP, encoding a new interacting partner of the ACTH receptor, cause familial glucocorticoid deficiency type 2. **Nat Genet**. (2005) 37:166-70.

Parathyroid/Calcium
- Hannan FM et al (2015). Adaptor protein-2 sigma subunit mutations causing familial hypocalciuric hypercalcaemia type 3 (FHH3) demonstrate genotype-phenotype correlations, codon bias and dominant-negative effects. **Human Molecular Genetics**, 24: 5079-5092.
- Nesbit MA et al (2013). Mutations affecting G-protein subunit α11 in hypercalcemia and hypocalcemia. **New England Journal of Medicine** 368: 2476-86.
- Grigorieva IV et al (2010). Gata3 deficient mice develop parathyroid abnormalities due to dysregulation of parathyroid-specific transcription factor Gcm2. **Journal of Clinical Investigation** 120: 2144-2155.
- Van Esch H et al (2000). GATA3 haplo-insufficiency causes human HDR syndrome. **Nature** 406:419-422.

Growth/Pituitary
- Gaston-Massuet C et al (2016) Transcription factor 7-like 1 is involved in hypothalamo-pituitary axis development in mice and humans. **Proc Natl Acad Sci USA** (in press).
- David A et al Repair of Aberrant Splicing in Growth Hormone Receptor by Antisense Oligonucleotides Targeting the Splice Sites of a Pseudoexon. **J Clin Endocrinol Metab**. (2010) 95(7):3542-6.
- Rajab A et al (2008) Novel mutations in LHX3 are associated with hypopituitarism and sensorineural hearing loss. **Human Molecular Genetics** 17(14): 2150-9.
- Kelberman D et al (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.
- Woods KS et al (2005) Over- and under-dosage of SOX3 is associated with infundibular hypoplasia and hypopituitarism. **American Journal of Human Genetics** 76(5):833-49.
- Carvalho LR et al (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.
- Metherell LA et al Pseudoexon activation as a novel mechanism for disease resulting in atypical growth-hormone insensitivity. **Am J Hum Genet**. (2001) 69:641-6.
- Dattani MT et al HESX1/Hesx1 associated with septo-optic dysplasia in human and mouse. (1998) **Nature Genetics** 19(2): 125-133.

Disorders of sex development
- Idkowiak J et al A missense mutation in the human cytochrome b5 gene causes 46,XY disorder of sex development due to true isolated 17,20 lyase deficiency. **J Clin Endocrinol Metab**. 2012 Mar;97(3):E465-75.

Gonads/Reproduction/Puberty
- Idkowiak J et al. **J Clin Endocrinol Metab.** 2016 Steroid sulfatase deficiency and androgen activation before and after puberty. Mar 22:jc20154101. [Epub ahead of print]

- Bashamboo A et al Human male infertility associated with mutations in NR5A1 encoding steroidogenic factor-1. **Am J Hum Genet** 2010;87:505-512
- Topaloglu AK et al TAC3 and TACR3 mutations in familial hypogonadotropic hypogonadism reveal a key role for Neurokinin B in the central control of reproduction. **Nat Genet.** 2009 Mar;41(3):354-8.
- Arlt W et al Congenital adrenal hyperplasia caused by mutant P450 oxidoreductase and human androgen synthesis: analytical study. **Lancet**. 2004 Jun 26;363(9427):2128-35.

Thyroid

- Schoenmakers E et al Mutation in human selenocysteine transfer RNA selectively disrupts selenoprotein synthesis. **J Clin Invest.** 2016 Mar 1;126(3):992-6. doi: 10.1172/JCI84747.
- Taylor PN et al Whole-genome sequence-based analysis of thyroid function. **Nat Commun.** 2015 Mar 6;6:5681.
- Sun Y et al (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.
- Bochukova E et al A mutation in the thyroid hormone receptor alpha gene. **N Engl J Med.** 2012 Jan 19;366(3):243-9.

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 the data for all variants in VCF format, and in some cases, raw Fastq or BAM files for re-analysis and somatic mutation detection. In addition, we will require clinical phenotypic data (HPO terminology including modifiers), as well as documents detailing previous genetic testing, imaging, biochemistry, histopathology and any other relevant phenotypic data.

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

- Generate high-quality variant call sets for all samples within the GeCIP domain.
- Collaborate with the cross-cutting GeCIPs to identify structural variation, call copy number variants and detect somatic mosaicism.
- Collaborate with the cross-cutting GeCIP for electronic patient records to obtain more detailed phenotyping information.
- Collaborate with the cross-cutting Advanced Analytics GeCIP to group patients on the basis of phenotypic similarity.
- Annotate variants using empirically-derived data on gene structure and splicing as well as on the basis of existing standard annotation resources.
- Filter variants on the basis of existing control data sets as well as biochemically and endocrinology evaluated control data sets.
- Prioritise variants on the basis of more complex forms of annotation including regulatory loci.

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)*

We will require the data captured by the disease-specific data models (both HPO and non-HPO terms), but given the complexity of phenotyping in endocrinology and metabolism we would also wish to access uploaded original reports when possible. These would include reports on prior genetic testing, imaging, biochemistry and histopathology. We intend to use these documents to

improve the quality of phenotyping data. This will be achieved through the use of text-mining and semantic technologies for annotation and capture of information. This additional phenotypic data will form

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)*

Complementary to Genomics England's NGS bioinformatics pipelines we will also deploy a GATK-based variant calling workflow (https://www.broadinstitute.org/gatk/) to increase confidence for novel variants.

For somatic mutation analysis we will employ Varscan2 (http://varscan.sourceforge.net) and Mutect analysis workflows (https://www.broadinstitute.org/cancer/cga/mutect) to detect somatic mosaicism.

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)*

In addition to tools that are listed in the spreadsheet, we will also require:
- Variant Effect Predictor, or VEP (http://www.ensembl.org/info/docs/tools/vep/index.html) complementary to annovar and provide additional functional annotations.
- Beagle (https://faculty.washington.edu/browning/beagle/b3.html) – for phasing genotype data and for imputation of missing genotype data.
- Varscan2 (http://varscan.sourceforge.net) – complementary variant calling method that in addition can pick up mosaic variants.
- Mutect (https://www.broadinstitute.org/cancer/cga/mutect) - complementary to Varscan2 for somatic mosaic variants.

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 would require the additional reference data listed below:

For GATK and Mutect workflows:
1000G_omni2.5.b37.vcf
1000G_phase1.indels.b37.vcf
1000G_phase3_v4_20130502.sites.vcf
b37_cosmic_v54_120711.vcf
dbsnp_138.b37.vcf
hapmap_3.3.b37.vcf
Mills_and_1000G_gold_standard.indels.b37.vcf

Most files are available from the Broad Institute website.

Reference Human Genome GRCh37 where alignment is originally performed.

For Variant Effect Predictor, we will require:

Data cache for Ensembl V83, plugin for EXAC, 1000G phase 3 MAFs. Data is available to download via VEP installation script.

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

Alignment, GATK and Mutect workflows would require significant computing resources, and intermediate files may take up a significant amount of storage, for instance:
Each genome would require approximately 200Gb of temporary storage space.
Each genome would require approximately 2,400 CPU hours to generate a high-quality variant call set.

At the moment similar workflows are being carried out at the Cambridge High-performance Computing cluster, with NGS data stored in dedicated Lustre storage directly linked to the cluster. Similar computational setup will be required in order to process WGS data generated by Genomics England in an effective manner.

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 believe that the omics samples being collected by Genomics England will be a key resource for testing in silico predictions of pathogenicity. We know that many genes associated with tissue specific disorders are expressed in peripheral blood, making it possible in some cases at least to explicitly test the impact of genetic variants on gene expression and splicing in blood-derived RNA samples (accessible through the collection of PAX gene tubes). Furthermore, many metabolic and endocrine disorders would be expected to result in changes in metabolites detectable in the blood and the storage of plasma samples will be extremely useful. Therefore, we envisage omics samples being an essential part of the process of calibrating and testing in silico predictions of variant pathogenicity.

Data access and security

GeCIP domain name	[from previous entry]
Project title *(max 150 characters)*	[from previous entry]

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*

☐ *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*

☐ *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*

☐ *Tool evaluation and improvement*
Information Governance ☐ 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.