GeCIP Detailed Research Plan Form

August 2015

Background

The Genomics England Clinical Interpretation Partnership (GeCIP) brings together researchers, clinicians and trainees from both academia and the NHS to analyse, refine and make new discoveries from the data from the 100,000 Genomes Project.

The aims of the partnerships are:

1. To optimise:

- clinical data and sample collection
- clinical reporting
- data validation and interpretation.

2. To improve understanding of the implications of genomic findings and improve the accuracy and reliability of information fed back to patients. To add to knowledge of the genetic basis of disease.

3. To provide a sustainable thriving training environment.

The initial wave of GeCIP domains was announced in June 2015 following a first round of applications in January 2015. On the 18th June 2015 we invited the inaugurated GeCIP domains to develop more detailed research plans working closely with Genomics England. These will be used to ensure that the plans are complimentary and add real value across the GeCIP portfolio and address the aims and objectives of the 100,000 Genomes Project. They will be shared with the MRC, Wellcome Trust, NIHR and Cancer Research UK as existing members of the GeCIP Board to give advance warning and manage funding requests to maximise the funds available to each domain. However, formal applications will then be needed to individual funders. They will allow Genomics England to plan shared core analyses and the required research and computing infrastructure to support the proposed research. They will also form the basis of assessment by the Project's Access Review Committee, to permit access to data. Some of you have requested a template for the research plan which we now provide herewith.

We are only expecting one research plan per domain and have designed this form to contain common features with funder application systems to minimise duplication of effort. Please do not hesitate to contact us if you need help or advice.

Domain leads are asked to complete all relevant sections of the GeCIP Detailed Research Plan Form, ensuring that you provide names of domain members involved in each aspect so we or funders can see who to approach if there are specific questions or feedback and that you provide details if your plan relies on a third party or commercial entity. You may also attach additional supporting documents 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)

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

Application Summary	
GeCIP domain name	**Cardiovascular**
Project title *(max 150 characters)*	**Realising Genomic Cardiovascular Medicine**

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

1. To provide genetic diagnoses to patients and families with cardiovascular diseases enrolled in the 100,000 Genomes Project.

2. To identify new causative genes for phenotypes included within the Project.

3. To functionally characterise causative genetic variants and mechanisms of disease

4. To refine genotype/phenotype relationships and identify factors relevant to presentation, prognosis and response to treatment

5. Where possible, to identify potential for orphan drug or novel therapeutic strategies

6. To equip trainee clinicians with the necessary genomic medicine knowledge and skills for practice in the genomic era.

7. To develop proposals to utilise a proportion of the 30,000 reserved but non-funded genomes secured by Genomics England for cardiovascular genomic medicine studies.

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)*

Rare genetic cardiovascular diseases, also called "Inherited Cardiac Conditions", when considered together affect around 1% of the population. Among rare diseases, as a group they carry a unique susceptibility to sudden death at a young age, usually occurring in otherwise healthy people. They are therefore among the most devastating of rare diseases for affected families. These conditions may affect the functioning of the heart muscle (cardiomyopathies), the regulation of the heart beat (arrhythmias) or cause weakness of major blood vessels (aortopathies). The treatments for these conditions are onerous and in themselves risky: for example the implantation of devices in the body to restart the heart in the event of a sudden cardiac arrest carries a risk of infection and the need for lifelong close followup.

Congenital heart disease, where a baby is born with a structural abnormality of the heart, remains a significant cause of childhood death, may require repeated surgical operations, and sometimes leads to ongoing ill health in adult life together with a shorter lifespan. There is a substantial genetic component to the risk of congenital heart disease. Numbers of adults with congenital heart disease are increasing substantially as more people with the condition survive childhood.

The lymph vessels are a part of the circulatory system of the body with roles in preventing the congestion of body tissues with fluid, and the functioning of the immune system. Genetic abnormalities of these vessels lead to severe problems with swelling of the limbs. Certain genetic

conditions may affect the small blood vessels within the brain and lead to devastating strokes, often at young ages.

A substantial proportion of people with genetic cardiovascular diseases do not have the causative gene for their condition identified. Such genetic diagnoses may inform treatment options, lifestyle choices, and are particularly important in assessing the risk to family members, who otherwise may live many years with uncertainty about their own health and the health of their families.

We will work with the 100,000 Genomes Project data to provide genetic diagnoses to participants. We comprise a diverse group of people who together have expertise in all aspects of genetic cardiovascular disease, from patient care to basic laboratory science. We will harness this expertise to discover new genes in 100,000 Genomes patients, find out how those genes contribute to risk, and look for ways to use that information in best helping families to live with their inherited cardiovascular disease.

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)*

The range of diseases covered by the Cardiovascular GECIP is broad, involving conditions of the myocardium, blood vessels, lymphatics, and lipid metabolism leading to premature atherosclerosis. These may manifest as a result of pathological changes from the earliest period of gestation to late in life. Therefore, methodological approaches will of necessity be heterogeneous; this section outlines the general technical approaches that we will apply in our research, with further details available in the subdomain sections.

1. Testing of 100,000 Genomes candidate variation in population resources held by GECIP members and collaborators. If a variant of unknown significance in a new gene for a particular disorder is identified in the Project, confirmation of the same gene in other families with the disorder not enrolled in the Project will be the key first step towards establishing likely pathogenicity. GECIP participants have extensive patient and family collections in all subdomains of the Cardiovascular GEL ascertainment which chiefly for consent related reasons cannot currently contribute to the GEL Project; nevertheless, these resources could be used to confirm potentially pathogenic candidate genes.

2. Development of novel bioinformatics approaches to the analysis of 100,000 Genomes candidate variation to identify likely causative variants.

3. Functional characterisation of candidate variants using molecular biological, cell biological and where appropriate animal modelling approaches. Where a particular cell type is clearly implicated in disease pathogenesis, iPSC generation followed by differentiation and in-depth cellular phenotyping (or, as technology permits, genome editing using CRISPR/Cas9 to generate mutant cell lines).

4. In-depth re-phenotyping, including multi-omics methods. This will involve high-fidelity clinical characterisation (imaging, electrophysiological, and where appropriate invasive), use of RNA and protein data derived from clinical samples, metabolomics and epigenetic studies, and the curation of a developmental cardiac genetic expression atlas.

5. Detailed longitudinal followup of patients by health record linkage. Investigation of genotype/outcome relationships.

6. Through identifying novel mechanisms from the 100,000 Genomes data, to implement approaches to drug repurposing for rare cardiovascular disease; to use high throughput technology to identify new compounds relevant to rare disease.

7. Analysis of variants in known ICC genes identified as incidental findings in individuals with unrelated conditions, and correlation with clinical phenotypes (e.g. considering penetrance of rare variants in TTN and other genes that have relatively high rates of rare variation)

8. Recall for deep phenotyping of individuals with rare genotypes of interest – for example, those incidentally found to have LoF of a highly constrained gene expressed predominantly in the cardiovascular system, but who have been recruited with an unrelated phenotype

Expected start date	**1/8/2016**
Expected end date	**31/7/2021**

Lead Applicant(s)	
Name	Bernard Keavney
Post	BHF Professor of Cardiovascular Medicine
Department	Institute of Cardiovascular Sciences
Institution	The University of Manchester
Current commercial links	None

Administrative Support	
Name	Caroline Price
Email	Caroline.a.price@manchester.ac.uk
Telephone	0161 2751201

Subdomain leads		
Name	**Subdomain**	**Institution**
Hugh Watkins	Cardiomyopathy	University of Oxford
Perry Elliott	Cardiomyopathy	University College London
Stuart Cook	Cardiomyopathy	Imperial College London
Elijah Behr	Arrhythmias	St George's University of London
Andrew Grace	Arrhythmias	Papworth Hospital NHS FT
Clifford Garratt	Arrhythmias	The University of Manchester
Paul Clift	Familial thoracic aortic aneurysms and dissection	University Hospitals Birmingham NHS FT
Bernard Keavney	Congenital Heart Disease	The University of Manchester
Pia Ostergaard/Sahar Mansour	Primary lymphoedema	St George's University of London
Steve Humphries	Familial hypercholesterolaemia	University College London
Hugh Markus	CADASIL negative small vessel cerebral disease	University of Cambridge
Panos Deloukas	Functional Genomics	Queen Mary University of London
Elijah Behr	Sudden Unexplained Death in the Young (SUDY)/Sudden Infant Death Syndrome (SIDS)	St George's University of London
Mark Caulfield	Blood Pressure	Queen Mary University of London

Detailed research plan - Arrhythmias

Full proposal (total max 1500 words per subdomain)	
Title *(max 150 characters)*	**Arrhythmia syndrome subdomain: Novel diagnostic markers and risk stratifiers in arrhythmia syndromes**

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).*

Three main arrhythmia syndrome phenotypes are currently eligible for whole genome sequencing (WGS): long QT syndrome (LQTS); Brugada syndrome (BrS); and catecholaminergic polymorphic ventricular tachycardia (CPVT). These disorders predispose to sudden death from womb to adulthood and are responsible for much unexplained sudden death (1).

BrS: While a proportion of BrS is sporadic (2), when familial it has appeared to be inherited as an autosomal dominant trait with incomplete penetrance. Mortality due to BrS is usually unheralded and current risk stratification is limited mainly to prior symptoms (3). Whilst it has been associated with mutations in at least 14 different genes the *SCN5A* gene is the only consistent association; with rare variants identified in 21% of probands (1). Thus the majority of BrS remains genetically unexplained, but presents phenotypically as a monogenic disorder. In addition, in many BrS families carrying a *SCN5A* mutation there are clinically affected family members with BrS who are not genetic carriers (4). *SCN5A* variant carriers had greater evidence for conduction disease suggesting a role in disease modification rather than causation with variants in other genes being required for disease development or modification even in patients with *SCN5A* mutations. Furthermore a GWAS in a primary discovery population (n=312) identified common variants in three loci that associated with up to a 28-fold risk of carrying BrS (5). It therefore appears that BrS is likely to have oligogenic inheritance patterns. Importantly, the majority of affected patients are “genotype-negative” at present.

LQTS: Inheritance of LQTS is predominantly Mendelian autosomal dominant but may also be oligogenic in approximately 5% of cases. Incomplete penetrance is common (1) and this may be in part due to genetic modification by common genetic variants which may also modify risk. Three genes, *KCNQ1*, *KCNH2* and *SCN5A*, account for 70-75% of definite LQTS cases (1); the remaining genes account for <5% and are therefore either very rare and/or the evidence for causation is weak; relying upon variant rarity and functional characterisation (1). Around 20% of LQTS is therefore “genotype negative”.

CPVT: The condition is often highly penetrant, predominantly affecting children and usually autosomal dominant in its mode of inheritance. Approximately 60% of probands carry a mutation in the *RYR2* gene, responsible for the cardiac ryanodine receptor. Several other genes have been implicated in a small number of cases (1). Therefore around 30-40% remain “genotype negative” and as yet there have been no genetic modifiers identified.

Thus the genetics of these disorders are variably understood at present and no condition has been ascertained completely. The role for genetic modifiers in risk stratification requires exploration. The subdomain will therefore focus on novel gene discovery, rare variation in regulatory regions, structural variants not assayed by approaches used in routine clinical practice and phenotype associations with follow-on functional work. BrS in particular remains genetically inscrutable despite a clear familial basis in most cases. There are roles for both common and rare genetic variation and a large proband based experiment is needed to identify those haplotypes most linked to the risk of developing BrS.

Hypotheses

1. WGS based linkage studies in selected genotype negative multigenerational families or trios will identify novel genes causative of LQTS, BrS or CPVT.
2. A large 'enriched' singleton vs. control WGS experiment will associate BrS with ultra-rare, rare and common variation in novel loci and existing putative loci.

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.*

Case recruitment and phenotyping

There are extensive existing collections of well characterised cases and families from the Rare Arrhythmia Syndrome Evaluation (RASE) consortium: a UK-wide collection of collaborators. We have received project grant funding from the BHF for a longitudinal deep phenotyping study in BrS (RASE Brugada). We will actively promote GEL recruitment from these cases. The current estimated numbers of *SCN5A* negative BrS probands available through RASE (N=374) are based on a limited number of sites.

Methods

Family selection: For each specific family of interest for WGS based linkage studies we propose selection on the basis of the following inclusion criteria:

- other affected family DNA and phenotype available (at least two over three generations/meioses if autosomal dominant).

  OR

- trio of unaffected parents and severely affected child available (sporadic or recessive)

  PLUS

- strong definite phenotype
- genotype negative based upon GMC eligibility criteria

Family analysis: For trios the child's DNA examined for *de novo* ultra-rare variants or ultra-rare and rare variant homozygosity or compound heterozygosity (one variant from each parent). For large autosomal dominant families all informative family members will undergo WGS. Linkage analysis will be conducted appropriate to the mode of inheritance examining for shared. All efforts will then be made to recruit the whole pedigree and best candidate variants will be further studied for segregation.

BrS cohort study: we aim for 400 *SCN5A* negative singletons meeting inclusion criteria with enrichment for genetic risk by mandating a family history of premature sudden death (<40 years old) or SADS (<65 years old) **and/or** other relatives with a diagnosis of BrS.

Candidate study: Initial assessment of known BrS associated genes (1) and GWAS common loci (5) will be undertaken. This will include examining sequencing data at the GWAS loci to clarify the haplotypes most tightly associated with BrS compared to the relevant tagging SNP and therefore permit more accurate assessment of their likely effects on phenotype. Rare, ultra-rare and novel variation at these loci will also be evaluated for likelihood of pathogenicity based upon similar methods as described above. Novel loci not associated previously with BrS but with known cardiac expression and exhibiting likely pathogenic rare variation will be assessed on a case by case basis.

Segregation studies will then be undertaken for promising candidates in suitable families as above.

GWAS: A case control study will be undertaken comparing probands to controls (control data to be provided from other WG sequencing in UK100K cohort from unaffected and/or non-cardiac carriers) for an increased burden of genetic variation in candidate genes and gene networks. Genome-wide comparisons will be made on different levels and different allele frequency thresholds: individual variants (rare and common); collapsing single gene or sliding 'super locus' model; and biologically active networks/pathways. Functional assessment of variants will be incorporated using combined tools such as SKAT-O.

Evaluation of putative pathogenic variants: These will be evaluated by: confirmation by Sanger sequencing; sporadic or segregation with phenotype; rarity/novelty; prior disease association; nonsense vs. missense; analysis of "genetic constraint"; association with known relevant gene expression profiles; standard in silico methods; and in vitro functional assessment as part of additional research activity (see below).

Replication: Legacy cohorts from RASE collaborators as well as international collaborations (see below) are in existence that may be leveraged according to the phenotype(s) examined for sequencing of novel loci identified in our primary family or BrS association studies. The cohort of ARVC patients recruited in the cardiomyopathy subdomain may serve as an additional replication group for validation of findings in BrS patients linked to conduction slowing and fibrosis.

Proposed additional research activity

- Genomic data will be incorporated into a risk stratification model that includes a novel digital ECG score proposed as part of the RASE Brugada study.
- Blood-based eQTLs for any novel candidate loci will be considered if cardiac expression is relevant (e.g. *SCN5A*). So will markers for inflammation and/or metabolic derangement that may important in phenotype modification or novel phenotypes.
- Novel genes and variants will be considered for *in vitro* functional studies in either simple mammalian models, stem cell derived cardiomyocytes and/or MRC Harwell mouse model collaborations.
- Longitudinal follow-up data are already available in GMC linked cohorts that may be derived from RASE consortium cohorts for assessment of novel genomic markers for risk of SCD.

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.*

All analyses will be conducted in conjunction with the analytical methods domain and GEL bio-informatic support. For meta-analysis and replication we have an agreement of collaboration from Arthur Wilde and Connie Bezzina, AMC, Amsterdam, Netherlands who are conducting a collaborative European WGS study in BrS to which two GeCIP sites are contributing legacy cases.

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

Trainees from across the GECIP labs are being encouraged to use GECIP datasets as the basis of PhD and/or fellowship applications.

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

The GeCIP includes the main clinical cardiology and genetic leads in management of arrhythmia syndromes in the UK, all of whom are also engaged in local GMCs, alongside researchers with substantial expertise in gene discovery, statistical genetics and functional characterisation.

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.*

THE CV domain GECIP has overlapping membership with the variant interpretation domain, as well as with international efforts such as the Clinical Genome Resource (ClinGen) cardiovascular working group. Members are engaged in diverse efforts to improve interpretation through international data aggregation and in silico and in vitro modelling.

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

Identification of novel causative genes will be advantageous to other as yet genotype negative families for diagnostic purposes. Novel genomic associations with sudden risk are also of potential for risk stratification of patients and identification of novel biology for further exploitation, including pharmaco-therapeutic.

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?*

None as yet.

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

(1) Wilde AA, Behr ER. Genetic testing for inherited cardiac disease. Nat Rev Cardiol 2013 Oct;10(10):571-83.

(2) Schulze-Bahr E, Eckardt L, Breithardt G, Seidl K, Wichter T, Wolpert C, et al. Sodium channel gene (SCN5A) mutations in 44 index patients with Brugada syndrome: different incidences in familial and sporadic disease. Hum Mutat 2003 Jun;21(6):651-2.

(3) Raju H, Papadakis M, Govindan M, Bastiaenen R, Chandra N, O'Sullivan A, et al. Low prevalence of risk markers in cases of sudden death due to Brugada syndrome relevance to risk stratification in Brugada syndrome. J Am Coll Cardiol 2011 Jun 7;57(23):2340-5.

(4) Probst V, Wilde AA, Barc J, Sacher F, Babuty D, Mabo P, et al. SCN5A mutations and the role of genetic background in the pathophysiology of Brugada syndrome. Circ Cardiovasc Genet 2009 Dec;2(6):552-7.

(5) Bezzina CR, Barc J, Mizusawa Y, Remme CA, Gourraud JB, Simonet F, et al. Common variants at SCN5A-SCN10A and HEY2 are associated with Brugada syndrome, a rare disease with high risk of sudden cardiac death. Nat Genet 2013 Sep;45(9):1044-9.

Detailed research plan – Congenital heart disease

Full proposal (total max 1500 words per subdomain)	
Title *(max 150 characters)*	**Congenital Heart Disease**

Importance.
Congenital heart disease affects 7/1000 live births. Although most patients born even with serious CHD now reach adult life, CHD is an important cause of lifelong morbidity and premature death. Recurrence risk studies have shown a significant familial component to CHD. Some 20% of all CHD is due to chromosomal and other syndromic conditions (such as Down's syndrome, 22q11 deletion, and Noonan's syndrome) but in the remainder of cases, genetic or environmental causes remain very unclear. As the adult population with CHD continues to grow, it is increasingly important to pursue research into the genetic causes of the condition, in order to optimise genetic counselling that may be given to patients who are considering starting a family. Recent genomic studies of CHD have shown that 5-10% of cases are explained by *de novo* copy number variants, with potentially another 20% of cases with CHD accompanied by other malformations or intellectual disability harbouring *de novo* single nucleotide variants. Given the historically highly lethal nature of CHD, it is likely that causative variants are evolutionarily young (ie rare) or *de novo*, and with relatively high penetrance, thus detectable using sequencing approaches. A recent exome sequencing study suggested that possibly 100 or more genes were involved in the pathogenesis of CHD; of these as yet few are identified. Some recent exome sequencing studies of patients with CHD as part of complex phenotypes have discovered novel syndromic phenotypes. Some whole genome sequencing studies that have included patients with CHD and other features have revealed unusual presentations of recognised syndromes (eg Noonan's, which is in all likelihood significantly underdiagnosed) leading to new diagnoses that impact patient care directly. The above considerations all suggest that whole genome sequencing studies in CHD families will likely discover new risk loci and deliver clinical benefits for affected families.

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.*

We will use the 100,000 Genomes project data in trios and multiplex families to identify potentially pathogenic *de novo* and inherited variation respectively. We will adopt genome-wide approaches but also reduce the genomic search space by referring to lists of candidate genes assembled from our domain members' extensive knowledge of human CHD genetics, mouse models of CHD and gene networks important in heart and vascular development. Potentially causative genes will be sequenced in collections of CHD cases which we have available to replicate the findings. We will further probe causality through studying gene and protein expression patterns in mouse and human developing heart at appropriate stages, and through the study of gene-modified animal and cellular models. We will use the identified genes as means to access and describe new networks of gene-gene interactions important in normal and abnormal heart development. We will investigate the increasingly recognised link between developmental cardiac and neurological phenotypes through followup of health records in the 100,000 Genomes participants. Where appropriate we will also study the effects of identified genes on neural development, which may involve recall of participants for additional phenotyping. We will investigate the role of genetic variants on the development of late complications of CHD, for example heart failure and pulmonary hypertension.

We will raise external funding to perform experiments focused on epigenetics in CHD on the 100K Genomes samples. This is a very under-explored scientific area despite evidence from exome sequencing that genes involved in histone modification may be major susceptibility factors in

CHD, and the proven importance of epigenetics in cardiomyocyte differentiation, maturation and disease.

The major challenges are anticipated to be the incomplete penetrance of likely causative CHD variants that has been observed previously, and the difficulty in modelling the appropriate developmental milieu for functional experiments. The first challenge will be mitigated by the availability of replication samples from groups who are members of the GECIP and their international collaborators. The second will be mitigated by the broad range of animal and cellular modelling expertise available within the GECIP together with the development of novel approaches – this issue is the subject of ongoing intensive research by groups involved in this application, including iPSC based modelling and the development of stem-cell based cardiac organoids.

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.*

International academic collaborations are as per the original GECIP application.

Discussions with other GECIPs have commenced, in particular it is likely there will be common ground between colleagues primarily interested in Neurodevelopmental phenotypes and the congenital heart disease subdomain, as significant numbers of patients will be enrolled in whom both conditions coexist.

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

Trainees will be closely involved in all steps of the research.

MRes and PhD students will carry out projects utilising 100KG Project data

Postdoctoral researchers in multiple disciplines, perhaps most prominently clinical cardiology, clinical genetics, bioinformatics and genetic epidemiology will access the data for their gene discovery projects.

Postdoctoral researchers who are laboratory based will use the variants identified in the project to commence functional studies of the identified genes in cardiovascular development. Junior scientists aiming for independence will be encouraged to make maximum use of the unique project resource in structuring their future career; they will receive mentorship from more experienced members of the subdomain.

We are exploring the possibility of conducting online “Meta-MDT’s” involving trainees in genetics, clinical science and cardiology together with acknowledged experts in multiple centres to bring maximum expertise to bear on the assessment of cases enrolled in the project and the interpretation of potentially significant variants. These will represent significant training opportunities.

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

The group includes leading clinical paediatric and adult congenital cardiologists, clinical geneticists with an interest in CHD, human geneticists focused on congenital heart disease, bioinformaticians and statistical geneticists, laboratory scientists with animal modelling expertise, and stem cell biologists. All necessary expertise to address this project is present.

Many of those involved in this project have worked productively together previously, particularly in human genetic studies of CHD which are of equivalent or larger scale than the likely CHD ascertainment to the 100K project. We will follow similar working practices, focusing on the optimal conduct of tasks by the groups with the most relevant expertise and the generation of added value through pooling our expertise and resources.

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 provide expert heart developmental genetics/developmental biology input to the Validation and Feedback Domain where requested to assist in resolving the specific question of whether a given variant can be considered clinically actionable and fed back to patients. We anticipate that variants suitable for clinical feedback immediately may be concentrated in rather few genes, given the rudimentary state of knowledge about CHD genetics. Rather, it seems likely that most variants will initially be deemed of unknown significance – here the availability of a large replication cohort and extensive international collaboration will be particularly useful in establishing variant pathogenicity.

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

Participants and their families will benefit as clinically actionable variants may be identified by the GECIP research.

Patients with CHD not enrolled in the Project will benefit, as genes will be identified that could be of relevance in their families.

The NHS will benefit from the personalisation of risk counselling that may follow from the success of the research. More distantly, if genetic factors importantly influencing prognosis or risk of particular downstream outcomes are identified, there may be additional benefit to both healthcare providers and patients.

Other researchers and healthcare professionals worldwide will benefit from the availability of a curated database of variants of likely or confirmed pathogenic significance which can be consulted when they encounter patients similar to those enrolled in the Project.

It is possible that the better understanding of heart development that we will gain from studying how it goes wrong may inform approaches to cardiovascular regenerative medicine.

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.*

Homsy et al. 2015 PMID:26785492
Glessner et al. 2014 PMID: 25205790
Zaidi et al. 2013 PMID: 23665959
Al-Turki et al. 2014 PMID: 24702954
Topf et al. 2014 PMID: 25093829
Cordell et al. 2013 PMID: 23708191
Soemedi et al. 2012 PMID: 22939634

Detailed research plan – Primary Lymphoedema

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

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 lymphatic vasculature is important for fluid drainage and immune responses. Dysfunction of lymphatic vessels leads to disturbed tissue fluid balance and development of lymphoedema. The lymphatic system also contributes to immune and inflammatory responses in various other pathologies, including cancer, obesity, hypertension and atherosclerosis.

The majority of lymphoedema occurs as a result of an insult to the lymphatic system by extrinsic factors, such as surgery, infection (e.g. parasitic) or radiotherapy, which is known as 'secondary lymphoedema'. When intrinsic factors, such as genetic alterations, cause lymphoedema, it is called 'primary lymphoedema' (PL). This is a rare but important condition caused by abnormal development of lymph vessels or failure of lymphatic function. The resulting oedema may present at birth, later in childhood, or even in adulthood. Some forms of PL are associated with life-threatening conditions, e.g. pulmonary or intestinal lymphangiectasia as in Hennekam syndrome or acute myeloid leukaemia as in Emberger Syndrome. At present the treatment is palliative not curative, and restricted to physical measures such as compression garments.

Approximately 40% of PL cases are due to an identifiable genetic cause and several sub-types can be attributed to Mendelian gene defects. To date, mutations in 20 genes have been reported to cause forms of hereditary lymphoedema. Through in-depth phenotyping of the patients that attend our specialist primary lymphoedema clinic, we have been instrumental in the identification of ten of the 20 known PL genes (see e.g. Fotiou *et al.* 2015). We have successfully identified causative genes for single gene disorders using a few affected, unrelated patients (i.e. singletons), through extensive deep phenotyping, cohort stratification and whole exome sequencing (WES).

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.*

By working as a consortium we are in an outstanding position to explore the data coming from the 100k Genomes pipeline. Provided we get funding to cover manpower and consumables we will be in a good position to deliver on the following aims and objectives:

(A) Identification of new genes by Next Generation Sequencing (NGS).

The consortium plans to identify novel gene defects underlying PL using DNA from patients and relatives. Variant profiles from patient samples will be analysed with the aim of identifying causal mutations in each sub-group. The analysis plan will utilise a probabilistic framework to prioritise rare variants whose allelic composition is consistent with the observed inheritance patterns, and we will seek to aggregate such variants across groups of individuals with similar phenotypes to test the hypothesis that there is a burden of variants disrupting a specific gene that would warrant further investigation.

Putative pathogenic alleles identified will be verified by Sanger sequencing in both the individual in whom they were observed and family members of the patient to evaluate co-segregation with disease status. Any genes for which we generate evidence that they are the site of pathogenic mutation, either from co-segregation or recurrent observations across unrelated individuals, will be sequenced in extended groups of patients with related phenotypes.

(B) Evaluation of new genes identified and development of pre-clinical models.
We will validate the new candidate genes identified in *(A)* by investigating their function *in vitro* and *in vivo*, and generating new pre-clinical mouse models. The consortium will utilise in vitro studies using various cell lines and primary lymphatic endothelial cells to identify their function and mechanisms of action. Zebrafish will be used for a relatively quick gene validation and analysis of in vivo phenotypes. Mouse models using conventional or CRISPR technologies will be generated and characterised.

The consortium will also work on the characterisation of existing mouse models and we have three possible therapeutic targets we wish to evaluate for the potential as a treatment option for PL. This work will take us closer to the therapeutic goal which is desperately needed for PL, where currently the only available treatments include manual lymph drainage therapies and compression stocking/bandaging.

At present, the major challenge however is the lack of manpower and consumable funds within the consortium to carry out any of the above work, therefore, applications for funding have been submitted to the Wellcome Trust and EU (both January 2016).

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 have set up a consortium with internationally recognised leaders in the field of lymphovascular biology. We feel there is a need for an interdisciplinary collaboration as it will allow the combination of expertise in developmental and pathological lymphangiogenesis with our expertise in clinical practice and genetics research. The members are:
Dr Taija Makinen (Uppsala University, Sweden) - expertise in the analysis of complex developmental vascular phenotypes using genetically modified mice and *in vitro* assays for endothelial cell functions.
Prof Tatiana Petrova (University of Lausanne, Switzerland) - expertise in the transcriptional regulation in human diseases, such as cancer and lymphoedema. She has expertise in mouse molecular genetics, animal disease models, molecular imaging, transcriptomes and ChIPseq analyses.
Prof Kari Alitalo (University of Helsinki, Finland) - expertise in endothelial growth factors and has established preclinical models of therapy. His laboratory is well established for advancing therapies and some of his work has already been translated to clinical development.
Prof Stefan Schulte-Merker (University of Munster, Germany) – expertise in vertebrate organogenesis, in particular the development of the lymphatic and blood vasculature. He has expertise in the use of zebrafish and mouse models.
Prof Michael Simpson (King's College London, UK) - expertise in disease gene discovery and the processing and interpretation of genome sequencing data.

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

There is a regular influx of BSc, MRes, and MSc project students in the Lymphovascular Research Unit at SGUL. Particularly, we will be hosting MSc students from the GEL funded MSc in Genomics Medicine.

We are also hoping to train a research assistant if our WT Collaborative Awards proposal is successful, and 13 PhD students (2 at SGUL) if our MSCA-ITN-ETN proposal is awarded.

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

The Lymphovascular GeCIP consists of Prof Sahar Mansour (consultant Clinical Geneticist and Lead), Dr Pia Ostergaard (non-clinical researcher in lymphovascular disorders and Lead), Professor Peter Mortimer (Head of Lymphoedema Clinical Research SGH/SGUL), Glen Brice (Genetic Counsellor and Research Nurse), Dr Kristiana Gordon (Clinical Lead for lymphovascular medicine SGH), Professor Steve Jeffery (non-clinical geneticist specialising in lymphovascular disorders), all St Georges University of London/St George's Hospital and Dr Vaughan Keeley (Consultant in Palliative Care) at the Royal Infirmary Derby.

All of the above have a national and international reputation in the clinical analysis and management of PL (PM, KG and VK), genetic diagnosis and phenotyping of PL (SM, GB), or the identification of causative genes and investigation of their subsequent functions (SJ and PO). The group at SGUL/SGH operates as a collaborative, and it is the successful integration of their skills which has led to the successes in PL research. We have received 4 grants from the BHF in recent years and have published five new genes in the past three years.

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 are in a prime position to carry out the clinical interpretation for the Lymphovascular GeCIP. We have clinicians and researchers familiar with variant interpretation. We have experience with phenotyping of patients with PL and we have done this for a number of years (see Connell et al 2013).

We also have a large cohort of patients (>550) already recruited to our study. Any new causative genes identified by the 100K project could be looked for in this patients with a similar phenotype within this cohort to confirm the pathogenicity.

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

By advancing our understanding of PL, we can promote better patient care. Simply understanding the cause of the problem offers immediate benefit to our patients. An accurate diagnosis provides

information about the natural history of disease, complications and recurrence risk. It may, in time , lead to targeted treatment of this condition.

One complication of lymphatic dysfunction is recurrent cellulitis (a problem in both primary and secondary lymphoedema), which is amongst the top ten reasons for admission to hospital. Therefore finding a better way of managing disease will greatly benefit the NHS.

Treatment could not only help people with PL but potentially also benefit those suffering from secondary lymphoedema. Of note, WHO estimates that approximately 40 million people worldwide are disfigured and incapacitated by elephantiasis (filariasis-related lymphoedema), therefore, our research approach could be of great significance.

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?*

Our consortium partner, Prof Kari Alitalo, has links with Herantis Pharma for the development of VEGFC gene therapy development. Animal trials are already underway.

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

Connell FC, […], Ostergaard P. The classification and diagnostic algorithm for primary lymphatic dysplasia: An update from 2010 to include molecular findings. *Clinical Genetics*. 2013;84:303-314

Fotiou E, […], Ostergaard P. Novel mutations in piezo1 cause an autosomal recessive generalized lymphatic dysplasia with non-immune hydrops fetalis. *Nat Commun*. 2015;6:8085

Detailed research plan – Famlial Hypercholesterolaemia

Full proposal (total max 1500 words per subdomain)	
Title *(max 150 characters)*	**Familial Hypercholesterolaemia 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).*

Familial Hypercholesterolaemia (FH) is an autosomal dominant disorder, characterised clinically by elevated LDL-cholesterol (LDL-C) levels, and as a consequence premature mortality and morbidity from coronary heart disease (CHD). Once identified, individuals with heterozygous FH can be successfully treated by lipid-lowering agents particularly statins which reduces their CHD mortality to that of the general population. The frequency of FH was historically estimated to be 1/500 but recent population studies from Denmark, and whole exome sequencing from the US have indicated that the true frequency of FH is likely to be closer to 1/250, and this estimate has also been confirmed in the UK 10,000 Genomes Project (SH, MF, VP unpublished). Of the roughly 240,000 FH patients predicted in the UK less than 10% have currently been identified and are being effectively treated in lipid clinics. Cascade testing using DNA information from identified FH patients is recommended by NICE as a cost effective way to increase the number of identified FH patients, but in the absence of a systematic national screening programme most remain unidentified, although the BHF have recently funded 14 FH nurses to carry this out in 11 different centres throughout England. We would therefore expect between 200-400 individuals in the entire 100,000 cohort to be carrying an identifiable FH-causing mutation that can be reported to patients and clinicians for such cascade testing and treatment through genetic services (MW, MW, ZM).

Roughly 93% of patients with identified genetic cause have a mutation in the receptor for low density lipoproteins (*LDLR*) and at UCL we (SH, SL) have curated the worldwide database for such mutations , which currently hold more than 1200 published and confirmed genetic causes of FH. The second most common cause of FH in the UK (~5% of FH patients) is a single mutation in the gene for apolipoprotein B (*APOB*) which prevents efficient binding of the LDL-C particle to the receptor and therefore reduced clearance of LDL-C from the blood by the liver. The other molecular cause of FH is that of mutations in the gene for *PCSK9*, which encodes the PCSK9 protein which is involved in the degradation of the LDL-receptor during its recycling and cellular trafficking. Mutations causing FH are gain-of-function, and we have shown that a common mutation in this gene explains about 2% of patients with FH in the UK (SH, AN, MS, PD, HS). We have recently demonstrated that in the majority of patients with a clinical diagnosis of FH, but where no identified mutation can be found in these three genes, the majority have inherited a greater than average number of LDL-C raising common variants (SH, MF, MS, AN), and therefore cascade testing in these individuals will be significantly less cost effective than in the those with a monogenic cause. There are however at least 10-20% of patients with a clinical diagnosis of FH who have a low polygenic burden, where the genetic explanation for their elevated LDL-C is currently unknown, and in this group searching for a novel gene where mutations are causing FH would be of great value. The identification of any new gene would provide a novel therapeutic target for LDL-C lowering. In the last year there have been several reports of rare causes of FH (SH, MF, VP, and groups in Holland and US) but with all of these genes the contribution appears to be only 1-2% at most of the known mutation group, leaving a considerable number of patients where a new gene discovery could be made. We are confident through this network of being able to provide samples from 100-200 patients for sequencing.

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.*

Although we believe that the elevated CHD risk in patients with a clinical diagnosis of FH is driven mainly through the degree of elevation of LDL-C, the metabolomic and lipidomic profile of patients with different molecular causes of their FH has not been explored in detail. A comparison between monogenic FH and polygenic FH (caused by the co-inheritance of a large number of LDL-C raising SNPs each of modest effect) would also be of great interest. A better understanding of this may lead to more tailored therapy for different patients. With larger numbers it will also be possible to compare profiles in patients with different classes of *LDLR* defects (eg receptor defective (null alleles) vs receptor deficient (missense mutations). It may also be that patients with different molecular causes of their clinical phenotype have different degrees of atherosclerosis which could be detected non-invasively by measuring carotid intima-media thickening or by measuring coronary calcification by EBCT. Such measures are being currently used in the Royal Free hospital by DN and colleagues. Also, the long term clinical outcome of these patients with regard to their lipid lowering response to statin and other therapies (for example PCSK9 inhibitors) and for the development of clinical CHD would be of great interest, and all identified patients will be entered onto the Simon Broome FH register.

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 have strong ties with two relevant UK patient groups namely HEARTUK and the British Heart Foundation. We will develop links with UK Biobank, (PD is co-leader with Prof John Danesh) and with the lipids working group in the UK Cardiometabolic Consortium. This will allow us rapidly to feed in new lipid genes for assessment in FH patients and to explore the utility of polygenic burden tests in the general population. Polygenic burden testing is acknowledged not to lie within the remit of GEL and does not form part of current NHS standards of care although this may change in the future. Linked with this is the international effort of the Global Lipids Genetic Consortium in which PD and others are currently analysing exome chip data from >300K samples.

Several of us have been or are currently on (SH, MS, PD, AN, NC) the advisory boards of major pharmaceutical companies such as MSD, Pfizer, Sanofi and AZ and we will develop a dialogue with them to protect and develop any IP that results from the findings of this work.

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

We will disseminate the findings and experience gained from the analysis to students attending the Genetics of Human Disease MSc at UCL (Director SH, Faculty VP), and through short courses in eHealth and statistics run through the Farr Institute and with colleagues in the London School of Hygiene and Tropical Medicine and QMUL train the genomic medicine staff in the NHS (PD). Several of us (SH, TA, VP) are directing PhD students who will be involved in the molecular analysis of the variants involved

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

The Simon Broome UK FH Register Committee Members have all agreed to be part of this GeCIP (NC, AN, MS, PD, HS) , and between them they have a huge experience in the management of patients with FH. Dr Devi Nair runs a busy Lipid Clinic at the Royal Free Hospital and all these

clinicians will be able to provide patient samples for sequencing, while other sources will be through our contacts with the DNA diagnostic laboratories in Great Ormond Street Hospital (ATB), and Bristol (MW), where next generation sequencing for the FH genes is being carried out routinely. Clinicians who have patients where no mutation can be identified and with a low LDL-C gene score (such testing is currently not part of standard NHS care and not within the remit of GEL) will be approached to submit samples. SH directed the 2010 national audit of FH, served in the 2008 NICE Guidelines for the Management and Identification of FH, and on the NICE Quality Standards published in 2013. Currently he directs the Paediatric Register for children with FH which already has more than 200 enrolled children. He is an international expert in the molecular genetics of FH and with SL at UCL has run the FH Mutation Database for more than 10 years. Currently >10% of identified *LDLR* variants are novel and of unknown significance, and SL has the experience in being able to predict the likely effect on function of the LDL-receptor using open access algorithms, both for missense mutations and splice variants, and colleagues at UCL have an in-house programme which also believe to be of value. We are able to carry out molecular testing of variants likely to affect splicing using an exon-trapping plasmid and will be helped with this by our international collaborator Dr T Leren. Variants identified in the promoter which might be affecting transcription using transfection into a hepatic cell line. Finally, with help from TA and our international collaborator Dr M Bourbon, MF is in the process of developing a Mass Spec based peptide assay to determine the functionality of variants identified in the *APOB* gene, which could affect the binding of LDL to the receptor. The molecular work will be supervised and carried out by MF, a highly qualified molecular geneticist in FH research.

Prof Steve Humphries – UCL (SH)
Dr Sarah Leigh – FH Mutation database UCL (SL)
Dr Marta Futema – Post Doc UCL (MF)
Dr Ruth Lovering – Gene Ontology UCL (RL)
Dr Devi Nair - Royal Free and UCL Hospital (DN)
Prof Andrew Neil - SB Register Oxford (AN)
Dr Mary Seed - SB Register Charing Cross (MS)
Dr Handrean Sorean – SB Register Manchester (HS)
Prof Paul Durrington - SB Register Manchester (PD)
Dr Ian McDowell – SB Register Cardiff (IM)
Dr Nigel Capps - SB Register (NC)
Dr Maggie Williams - DNA Diagnostic Lab Bristol (MW)
Dr Alison Taylor-Beadling - DNA Diagnostic Lab GOSH (ATB)
Prof Tim Aitman – FH Research Edinburgh (TA)
Dr Vincent Plagnol – UCL Bioinformatics (VP)
Prof Zofia Miedzybrodzka – Aberdeen Diagnostics.
Ms Melanie Watson – Clinical Genetics Southampton (MW)
Prof Panos Deloukas – Lipid genetics St Barts (PD)

International collaboration

Norway - Dr Trond Leren – expertise in molecular testing of *LDLR* splicing variants
Portugal - Dr Mafalda Bourbon - expertise in analysis of *APOB* functional variants

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 supplied Genome England with a fully annoted list of reported FH-causing mutations from the UCL LDLR mutation database. We (VP, MF, SH, TA) have considerable

experience in the bioinformatic analysis of NGS data as a result of our involvement in the UK10K project and other on-going research. The identification of subjects in the 100,000 genome project who carry known FH-causing mutations will enable them to be referred to a lipid clinic to lower their elevated LDL-C levels and to manage their CHD risk. Their relatives can also be tested using cascade testing and from each index case we estimate that 3-5 tested first degree relatives will identify 2-3 new FH patients who will similarly benefit from early detection of their risk.

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

This will benefit the NHS because of the reduction in early CHD events.

The discovery of all three molecular causes of FH have led to the development of novel lipid lowering therapies, the most recent of which being monoclonal antibodies to PCSK9. We can predict that the identification of any novel FH-causing gene will also lead to such therapeutic potential.

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 identification of any novel FH-causing gene will lead to significant therapeutic potential. Several of us have been or are currently on (SH, MS, PD, AN, NC) the advisory boards of major pharmaceutical companies such as MSD, Pfizer, Sanofi and AZ and we will develop a dialogue with them to protect and develop any IP that results from the findings of this work.

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

1. M. Benn, G. F. Watts, A. Tybjaerg-Hansen, B. G. Nordestgaard, Familial hypercholesterolemia in the danish general population: prevalence, coronary artery disease, and cholesterol-lowering medication. *The Journal of clinical endocrinology and metabolism* **97**, 3956 (Nov, 2012).
2. D. Marks, M. Thorogood, H. A. Neil, S. E. Humphries, A review on the diagnosis, natural history, and treatment of familial hypercholesterolaemia. *Atherosclerosis* **168**, 1 (May, 2003).
3. B. G. Nordestgaard *et al.*, Familial hypercholesterolaemia is underdiagnosed and undertreated in the general population: guidance for clinicians to prevent coronary heart disease: Consensus Statement of the European Atherosclerosis Society. *Eur Heart J*, (Sep 12, 2013).
4. H. A. Neil, T. Hammond, R. Huxley, D. R. Matthews, S. E. Humphries, Extent of underdiagnosis of familial hypercholesterolaemia in routine practice: prospective registry study. *Bmj* **321**, 148 (Jul 15, 2000).
5. A. Neil *et al.*, Reductions in all-cause, cancer, and coronary mortality in statin-treated patients with heterozygous familial hypercholesterolaemia: a prospective registry study. *Eur Heart J* **29**, 2625 (Nov, 2008).
6. L. Nherera, D. Marks, R. Minhas, M. Thorogood, S. E. Humphries, Probabilistic cost-effectiveness analysis of cascade screening for familial hypercholesterolaemia using alternative diagnostic and identification strategies. *Heart* **97**, 1175 (Jul, 2011).
7. M. A. Umans-Eckenhausen, J. C. Defesche, E. J. Sijbrands, R. L. Scheerder, J. J. Kastelein, Review of first 5 years of screening for familial hypercholesterolaemia in the Netherlands. *Lancet* **357**, 165 (Jan 20, 2001).

8. P. Brice, H. Burton, C. W. Edwards, S. E. Humphries, T. J. Aitman, Familial hypercholesterolaemia: a pressing issue for European health care. *Atherosclerosis* **231**, 223 (Dec, 2013).
9. R. R. Williams *et al.*, Diagnosing heterozygous familial hypercholesterolemia using new practical criteria validated by molecular genetics. *The American journal of cardiology* **72**, 171 (Jul 15, 1993).
10. D. P. Yaram-Smith L, O'Shea S, Dennis G, Bayly G, Taylor A, Day A, Watso M, Giles P, Ayling R, Haralambos K, Whatley S, McDowell I, Williams M, The impact of routine next generation sequencing testing for familial hypercholesterolaemia – 5 months service experience. *Atherosclerosis* **236**, e304 (2014).
11. V. Plagnol *et al.*, A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics*, (2012).
12. P. J. Talmud *et al.*, Use of low-density lipoprotein cholesterol gene score to distinguish patients with polygenic and monogenic familial hypercholesterolaemia: a case-control study. *Lancet* **381**, 1293 (Apr 13, 2013).
13. M. Futema *et al.*, Refinement of Variant Selection for the LDL Cholesterol Genetic Risk Score in the Diagnosis of the Polygenic Form of Clinical Familial Hypercholesterolemia and Replication in Samples from 6 Countries. *Clinical chemistry*, (Nov 20, 2014).
14. E. Usifo *et al.*, Low-density lipoprotein receptor gene familial hypercholesterolemia variant database: update and pathological assessment. *Ann Hum Genet* **76**, 387 (Sep, 2012).
15. M. Futema *et al.*, Analysis of the frequency and spectrum of mutations recognised to cause familial hypercholesterolaemia in routine clinical practice in a UK specialist hospital lipid clinic. *Atherosclerosis*, (Apr 18, 2013).
16. Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. *Nature*, (Dec 10, 2014).
    M. Futema, V. Plagnol, R. A. Whittall, H. A. Neil, S. E. Humphries, Use of targeted exome sequencing as a diagnostic tool for Familial Hypercholesterolaemia. *Journal of medical genetics* **49**, 644 (Oct, 2012).

Detailed research plan Cardiomyopathy

Full proposal (total max 1500 words per subdomain)	
Title *(max 150 characters)*	**Inherited cardiomyopathies:** Understanding the basis of cardiomyopathy not attributable to currently known disease genes

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).*

Prevalence: There are three main inherited cardiomyopathy sub-types, as well as a number of other more or less discrete entities, that together comprise the most prevalent inherited cardiac conditions and indeed are some of the most common, serious inherited diseases of any system: hypertrophic cardiomyopathy (HCM); dilated cardiomyopathy (DCM); and arrhythmogenic right ventricular cardiomyopathy (ARVC).

HCM affects at least 1 in 500 of the population and while the cardiomyopathies are thought to be less prevalent, current estimates are probably underestimates. Some of the rarer and less well characterised cardiomyopathies may be particularly amenable to WGS, for example those with infant or childhood onset and those associated with multi-system phenotypes.

Current state of genetic testing: Recent large scale genetic analyses evaluating the burden of appropriately rare potentially pathogenic variants in the known genes for each cardiomyopathy, in comparison with the background rate of comparable mutation in ExAC reference samples, indicate that: i) the known, robustly supported disease genes account for a minority of each of these cardiomyopathies; ii) a number of other implicated genes most probably do have disease causal mutations but these remain hard to interpret against a background of non-pathogenic rare variants; iii) there is substantial unexplained disease, and hence unmet genetic need, across these disorders.

HCM:
The majority of adult onset HCM appears to be familial, typically with autosomal dominant inheritance with reduced penetrance. In HCM families current gene panel analysis yields pathogenic mutations in ~30-50%; where single probands are studied the yield falls, typically to 20-35%. There are, however, very few large families with multiply affected members with HCM that do not yield diagnostic mutation on current gene tests. This suggests that much unexplained HCM may be caused by low penetrance alleles and that the disease in singletons or nuclear families with a small number of affected individuals, is likely not monogenic.

In neonatal, infant and childhood hypertrophic cardiomyopathy a number of known conditions with diverse inheritance patterns are recognised, yet current genetic testing (of metabolic and structural cardiomyocyte components) again gives a relatively low yield. In these settings there is a higher likelihood of major effect, often *de novo*, gene variants.

DCM:
DCM in adults shows a roughly 50% familial occurrence where first few relatives are systematically screened but the condition is more genetically heterogeneous than HCM and yields of diagnostic pathogenic variants are significantly lower. The one frequently encountered the disease gene is

Titin, *TTN*, where A-band truncation alleles in high PSI exons are found in ~15% of cases vs <0.5% of reference samples. Individuals in the normal population carrying TTN truncations appear to have mild structural cardiac abnormalities and rather low penetrance, but are susceptible to insults including peripartum cardiomyopathy. Many other implicated DCM genes do not show a statistically significant burden of rare variants in cases. Therefore, at present, the utility of diagnostic testing in DCM outside of Titin (and LMNA in early onset DCM) is limited.

ARVC:
The documented disease genes for ARVC involve components of desmosome function. Because pathogenesis is typically via loss of function alleles, interpretation of individually rare variants tends to be easier than in other cardiomyopathies. That said, certain apparent loss of function alleles appear too frequently in the population to be penetrant ARVC alleles and thus further research is required. The majority of singletons with apparently isolated ARVC are gene negative on current panels and a significant proportion of families with multiply affected individuals are also negative, again, indicating scope for further disease gene discovery and for understanding polygenic inheritance. Current data suggest an excess of affected individuals carrying more than one rare, potentially pathogenic variant in known disease genes.

Value for cascade screening and biological insight: Genetic diagnosis is important for all cardiomyopathies because clinical diagnosis can be challenging (e.g. where the cardiomyopathy is a diagnosis of exclusion and there are confounding issues or where mild/borderline features are difficult to interpret and limit the utility of clinical cascade screening). There are already examples of re-purposed drugs undergoing clinical trial in HCM on the basis of molecular insights in the currently known genes and there are new small molecule drugs in the design and evaluation stage targeting precise primary disease mechanisms. Thus there is considerable scope for additional target discovery, target validation, pathway analyses to drive new approaches to therapy. This may be particular pertinent for DCM that often runs a progressive course, is likely as common as HCM but has worse health outcomes and associated healthcare costs.

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.*

Case Recruitment: Current GEL criteria for cardiomyopathy have been designed with the above knowledge in mind to drive recruitment of genetically enriched, informative samples, to include trios for *de novo* or recessive early onset/severe disease, and extended autosomal dominant families (individually rare but probably under-recognised due to the difficultly in tracing remote relatives – which can be addressed by co-ordinated efforts across GMCs) and adequately powered case series of cases enriched for classical disease, early onset, familial occurrence all in the setting of negative gene panels.

Early experience by the three cardiomyopathy lead investigators indicates that WGS will be productive in these settings (e.g. 3 novel adult onset cardiomyopathy genes have been implicated through the WGS 500 programme in Oxford)

Detailed phenotyping: This is a strength of a number of UK cardiomyopathy clinical and research groups and will include conventional ECG and imaging approaches together with novel CMR based imaging (novel tools for detecting aspects of tissue characterisation and machine learning approaches for quantitative data) and systematic blood based metabolomics phenotyping (e.g. as

supported by the NIHR TRC Rare Disease collaboration) in sarcomeric HCM. Existing experience in monogenic cardiomyopathy highlights the utility of qualitative characterisation of cardiac phenotypes beyond heart muscle and systemic phenotypes beyond the heart: identification PRKAG2 (cardiomyopathy in conjunction with pre-excitation and progressive conduction block), CACNA1C (cardiomyopathy in conjunction with bradycardia and QT prolongation).

Data analysis: Techniques will be tailored to the starting material i.e. whether looking for *de novo*, compound heterozygous or identical by descent recessive alleles in infant cardiomyopathy, penetrant heterozygous variants in multiple effected families or cumulative burden of rare variants in individual genes (or pathway related genes) in large case series.

In each instance, findings arising from the project will be evaluated in other clinical and genetic resources, given early opportunities for replication. This includes extensive experience of family based exome and genome sequencing (legacy samples, studies preceding GEL) and availability of WGS on over 200 sarcomere negative HCM probands through the NIHR BioResource Rare Disease programme. For HCM, ARVC and some of the rarer cardiomyopathies initial analyses will interrogate a defined search space based on current mechanistic understanding (as there are a defined, relatively focused, final common pathway in these disorder). This will be harder for DCM and it is anticipated that it will be the presence of additional endophenotypic features (e.g. fibrosis extent and distribution and degree and distribution of ventricular trabeculations) that will allow sufficient homogeneity to define novel disease genes in collected samples.

A number of laboratories in the UK, particularly those led by the three cardiomyopathy leads, have extensive genomic, bioinformatics and functional genomics experience and capability. They are all also closely linked to nationally leading programmes outside of the UK. Reductionist biochemical, cell and animal model based experimental systems (including creation and rescue of mutations by gene editing), are in place, to include studies in isolated cardiomyocytes, iPSC-derived cardiomyocytes, engineered mouse and rat models.

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.*

A significant number of newly implicated cardiomyopathy genes cause a phenotype with extra cardiac features. We will therefore ensure active links with other domains (in particular skeletal myopathy, skin, neurological, metabolic disorders) to ensure joint collection and exploration of families with complex phenotypes that include cardiomyopathy. We have active collaborations with Illumina for gene panel and ICC informatics.

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

Trainees from across the GECIP labs are being encouraged to use the unique opportunities provided by this huge resource to design experimental programmes for higher degrees by research or subsequent career development. Each of the leading labs has substantial track record in training scientists and clinician scientists in this domain.

People and track record. *Explain why the group is well qualified to do this research, how the investigators would work together.*
The lead investigators have established individual and joint track records of discovery and translational research, including identification of key disease genes, introduction of diagnostic (genetic and imaging) tests, and writing of international guidelines

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.*

Cardiomyopathy groups are directly represented on the cross-cutting Validation and Feedback domain and are actively engaged in multinational efforts such as Clin Gen and Clin Var, EXAC consortium.

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

New genetic findings will be of immediate relevance to affected families and individuals and will be of future benefit through identifying new diagnostic and prognostic markers and new targets for potential therapy intervention (especially repositioning of existing agents). It may be that titin DCM is particularly suited to mechanical unloading and recovery is possible for some patients, hence stratified treatment of DCM may be considered. We have a substantial track record of commissioning genetic and genomic tests for cardiac conditions in the NHS.

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?*

Thus far, roll out of diagnostic testing has been within the NHS and therefore not-for-profit. Trials of new agents has required repositioning of existing (usually off-patent) drugs and so limited commercial exploitation as yet.

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

Walsh R and Cook SA. Diagnostic Sequencing for Inherited Cardiovascular Conditions. **Clinical Chemistry**. In Press
Walsh R…..Cook SA Watkins H. Reassessment Of Mendelian Gene Pathogenicity Using 7,855 Cardiomyopathy Cases And 60,706 Reference Samples. **Genetics in Medicine**. In Press.

Diane Fatkin, Lien Lam, Daniel S Herman, Craig C Benson, Leanne E Felkin, Paul JR Barton, Roddy Walsh, Sukru Candan, James S Ware, Angharad M Roberts, Wendy K Chung, Leslie Smoot, Helen

Bornaun, Anne M Keogh, Peter S Macdonald, Christopher S Hayward, JG Seidman, Amy E Roberts, Stuart A Cook, Christine E Seidman. Titin truncating mutations: A rare cause of dilated cardiomyopathy in the young. **Progress in Pediatric Cardiology** 2016 March, 40:41–45

Felkin LE, Walsh R, Ware JS, Yacoub MH, Birks EJ, Barton PJ, Cook SA. Recovery of Cardiac Function in Cardiomyopathy Caused by Titin Truncation. **JAMA Cardiol.** 2016;1(2):234-235. doi:10.1001/jamacardio.2016.0208.

Ware JS, Li J, Mazaika E, Yasso CM, DeSouza T, Cappola TP, Tsai EJ, Hilfiker-Kleiner D, Kamiya CA, Mazzarotto F, Cook SA, Halder I, Prasad SK, Pisarcik J, Hanley-Yanez K, Alharethi R, Damp J, Hsich E, Elkayam U, Sheppard R, Kealey A, Alexis J, Ramani G, Safirstein J, Boehmer J, Pauly DF, Wittstein IS, Thohan V, Zucker MJ, Liu P, Gorcsan J 3rd, McNamara DM, Seidman CE, Seidman JG, Arany Z; IMAC-2 and IPAC Investigators. Shared Genetic Predisposition in Peripartum and Dilated Cardiomyopathies. **N Engl J Med**. 2016 Jan 21;374(3):233-41. doi: 10.1056/NEJMoa1505517. Epub 2016 Jan 6. PMID: 26735901

Hinson JT, Chopra A, Nafissi N, Polacheck WJ, Benson CC, Swist S, Gorham J, Yang L, Schafer S, Sheng CC, Haghighi A, Homsy J, Hubner N, Church G, Cook SA, Linke WA, Chen CS, Seidman JG, Seidman CE. Titin mutations in iPS cells define sarcomere insufficiency as a cause of dilated cardiomyopathy. **Science**. 2015 Aug 28;349(6251):982-6. doi: 10.1126/science.aaa5458. PMID:26315439, PMCID:PMC4618316

Detailed research plan - Aortopathy

Full proposal (total max 1500 words per subdomain)	
Title	Familial Thoracic Aortic Aneurysm Disease

Importance. Thoracic aortic dissection is a cause of sudden death. A significant proportion of individuals have a family history of aortic dissection consistent with an autosomal dominant pattern of inheritance. Existing genetic testing strategies can identify known mutations in around 30-40% of probands tested. Currently known mutations occur in the TGF beta and ACTIN pathways [(1)], and mouse genetic models of these disease have led to the development of targeted therapy using angiotensin receptor antagonists in these patient groups [(2)]. However the majority of patients and families do not have a genetic diagnosis & these are being enrolled into the 100K Genome project.

The objectives of the sub-domain are:

1) The discovery of novel genotypes to provide diagnoses for enrolled patients and families
2) Use in vitro Phenotyping techniques to characterise novel genetic variants
3) Using multi-omic techniques to identify markers of risk and potential new therapeutic targets
4) To determine the prevalence and significance of known genetic variants in the 100K Genome cohort using the 100K Genome dataset
5) To coordinate phenotypic datasets to enrich the known genotype/phenotype descriptors and facilitate comparative phenomics

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.*

Case recruitment: The Aortic GeCIP subdomain involves all clinicians managing and investigating aortopathy patients in England, Rectruitment criteria will drive enrolment of genetically enriched and informative samples. We are successfully recruiting genotype negative probands and families following aortic gene panel testing through NHS GMCs. We will recruit both trios for *de novo* or recessive early onset disease, and extended autosomal dominant families – the potential to coordinate recruitment of geographically dispersed families through the GMC network holds particular promise. WGS will produce candidate variants that can be studied in detail for their significance. Of importance, many participating centres have existing cohorts not suitable for GEL enrolment in whom pathogenicity of newly identified genes in the GEL cohort can be confirmed as and when they are discovered.

Detailed phenotyping: Pre-enrolment phenotyping of GEL recruits will typically have included echocardiography and CMR (in some patients CT) of the thoracic aorta and for some patients other major extrathoracic vessels. Serial imaging will be available as routine on most patients and such data will be mined to determine genotype/phenotype relationships in regard to rate of disease progression. Novel CMR techniques such as 4D flow MRI to characterise abnormal wall shear stress will be used to interrogate the functional characteristics of the aortae of participants with proven causative mutations in different genes. The impact of associated malformations such as bicuspid aortic valve and aortic coarctation on the expressivity of particular genotypes will be investigated. We plan a national dataset to incorporate these and additional phenotypic features to facilitate comparative phenomic research to highlight potential candidate genes of interest.

Systematic blood based metabolomics phenotyping will be undertaken to identify biomarkers particularly associated with the risk of rapid aortic expansion or dissection. Clinical characterisation beyond the aorta will be undertaken to establish relationships between new genes and other connective tissue phenotypes, and other cardiac phenotypes which may coexist with aortopathy (for example the recently described association between fibrillin mutations and coexisting Marfan's syndrome and left ventricular non-compaction cardiomyopathy).

Data analysis: Approaches appropriate to the clinical and family presentation will be used: ie *de novo,* compound heterozygous or IBD recessive alleles, will be sought particularly in early onset presentations; penetrant heterozygous variants in autosomal dominant families; and rare variant burden in genes expressed in aortic tissue in singleton cases. The resources already in hand and international collaborations will facilitate early replication in other clinical and genetic resources. The group includes a number of laboratories with genomics, bioinformatics and functional genomics expertise. Functional characterisation of candidate variants will be carried out using a variety of in vitro techniques, which will prominently include gene editing using CRISPR-Cas9 in cellular and animal models. iPSC lines from patients will be differentiated to functionally relevant cell types (endothelial cells, fibroblasts, smooth muscle cells) to further characterise novel genotypes, CRISPR techniques can further be used in such cells to carry out rescue experiments [(3)]. Prospective followup of patients enrolled in GEL will enable access to aortic tissue when operative treatment is required, and this will provide an opportunity to investigate in detail the *in vivo* cellular phenotypic consequences of identified mutations. Zebrafish morpholino techniques can be used to investigate the effects of candidate variants on aortic development; there are well characterised examples of the use of the fish to model aortic disease.

Biophysical analyses: The group includes experts in the application of nanomechanical engineering techniques. These will be applied to whole explanted aortic tissue to detail mechanisms of dissection initiation and biomechanical phenotyping of whole tissue, and allow genotype/phenotype correlation can be made with known and candidate variants [(4,5)]

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 aortic GeCIP subdomain is working with the Scottish Genomes Partnership (Prof Tim Aitman) & other domain members have established collaborations with the Montalcino Aortic Consortium, GenTac, European Reference Network on Genetic Aortic Aneurysmal Disease (ERN-GAAD) and colleagues in North America and Europe. These are clinical and genetic collaborations.

Analysis of GEL results will be performed in conjunction with the analytical methods domain and using available GEL bio-informatic support

Training. *Describe the planned involvement of trainees in the research and any specific training that will form part of your plan.*
We have offered to deliver modular training in aortic conditions as part of the GMC commitments to training MSc students. We are open to trainees from across the GECIP labs to work with GeCIP members to support PhD and/or fellowship applications

People and track record. *Explain why the group is well qualified to do this research, how the investigators would work together.*
GeCIP members have experience and expertise in all proposed areas of research including novel genotype discovery, new variant analysis, in vitro phenotyping techniques, phenomic dataset

interpretation, biosatistical analysis. There are existing and proposed collaborations within the group and members have a track record of national and international collaboration in the area of genomic research. There is a wealth of clinical experience within the group with surgeons, physicians and clinical geneticists all providing dedicated aortic clinical services from which the cohort of GEL participants have come from.

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.*

The group will work with the validation and feedback domain to examine the likely pathogenicity of novel variants, we can also utilise the existing international collaborations to establish pathogenicity, by examining established pedigrees in the UK and elsewhere.

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

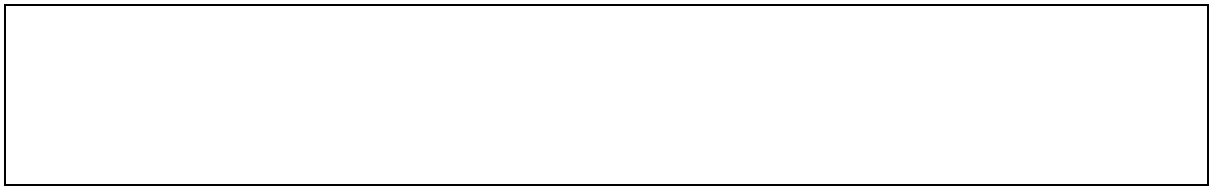
Novel variant discovery and subsequent research will provide a number of benefits: ability to cascade screen relatives of probands, to risk stratify known and novel genotypes to plan clinical management of their aortic condition, with the aim of preventing aortic dissection, reducing the morbidity and mortality of the condition and the associated healthcare costs to the NHS. The goal is prevention of avoidable aortic dissection either by surgical intervention or by the development of targeted novel therapeutic strategies based on known effects on the signalling pathways impacted by the pathogenic mutations, e.g. Angiotensin receptor antagonists in management of Marfan syndrome and Loeys Dietz Syndrome

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 not aware of existing commercial collaboration but it is clear that there could be commercially exploitable results from a number of areas including surgical devices, risk stratification and therapeutic interventions.

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

1. Van Laer et al. Connective tissue disorders with vascular involvement: from gene to therapy *Eur J Pediatr* (2013) 172:997–1005
2. Gallo et al. Angiotensin II–dependent TGF-β signaling contributes to Loeys-Dietz syndrome vascular pathogenesis *J Clin Invest.* 2014;124(1):448–460
3. Sinha et al. Embryonic origins of human vascular smooth muscle cells: implications for in vitro modeling and clinical application. *Cell. Mol. Life Sci*. (2014) 71:2271–2288
4. Akhtar et al. Characterizing the elastic properties of tissues. *Mater Today (Kidlington).* 2011 March ; 14(3): 96–105.
5. Akhtar et al. Localized micro- and nano-scale remodelling in the diabetic aorta. *Acta Biomaterialia* 10 (2014) 4843–4851
6. Gkoutos et al. Computational tools for comparative phenomics; the role and promise of ontologies *Mamm Genome*. 2012 October ; 23(9-10): 669–679



Detailed research plan – Functional Genomics

Full proposal (total max 1500 words per subdomain)	
Title *(max 150 characters)*	**Functional Genomics 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 cardiovascular GECIP comprises subdomains focusing on the genetic analysis of cardiomyopathies, arrhythmias, familial thoracic aortic aneurysms, congenital heart disease, primary lymphoedema, and familial hypercholesterolemia. The six disease subdomains share a common objective to identify likely causative variants and implement approaches to translate these findings in to new therapeutics including drug repurposing, for cardiovascular disease.

The development of novel bioinformatics approaches to analyse variation in the 100,000 Genomes data set is a central objective of the CV GECIP. The Functional Genomics subdomain aims to empower this endeavour through integration of multi-omic data sets that can be generated across many of the analysed patient groups for economy of scale. Based on sample availability we would prioritise the generation of metabolomics profiles followed by transcriptomics analysis - expression profiles for all protein coding genes, microRNAs and long non-coding RNAs - and DNA methylation profiling. Metabolomics can capture the great chemical and functional diversity of the biological processes responsible for the observed phenotype and can be used to explore a wide range of hypotheses. Similarly, expression data can capture perturbations in related genes and lead to the linking of disease phenotypes to specific pathways. Cancer has successfully used the integration of whole genome sequence with other omics and rich clinical data for patient stratification.

In parallel, this subdomain will coordinate efforts to optimise and scale up platforms for functional characterisation of identified mutations where applicable. This may involve establishing induced pluripotent stem cells (hiPSCs) from patients and differentiation to disease relevant cell types; for example cardiomyocytes and endothelial cells will be relevant for several of the analysed disease conditions. Establishing a national collaborative for cellular phenotyping in rare CV disease will accelerate functional work and is very likely to also promote interaction with groups working on common forms of CV disease. The subdomain also has access to the MRC Mouse Network, a member of the International Mouse Phenotyping Consortium to apply CRISPR/cas9 technology for generating mouse knock outs and humanised mouse models.

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.*

In the first instance, the main challenge for this subdomain will be to secure funding to support coordination of activities among the participating groups and the generation of multi-omic data across most of the individuals with whole genome sequence data.

The availability of banked serum and plasma samples by GEL makes metabolomics profiling a preferable first target. On the level of the individual, metabolomics measures can be used to provide a multivariate profile of the individual or a disease. These profiles can be further used to identify the key changes associated with the disease, or the existence of specific subtypes of it. There are high throughput NMR-based platforms for targeted panel analysis e.g. lipid focused which are cost effective but it is also possible to consider a smaller number of samples for a comprehensive scan of over 2000 metabolites (e.g. MS-based platform from Metabolon reports on 2100 including 1100 lipid species). The metabolipidomic unit at Barts is supporting rare disease genomics.

We plan to apply RNA sequencing to affected and unaffected individuals to establish 'healthy' and disease expression profiles for all protein coding genes, microRNAs and long non-coding RNAs in blood. In parallel, we will seek funding to generate DNA methylation profiles from matched DNA samples (collected from the NHS England Genomic Medicine Centres). Blood is an ideal tissue to search for expression and methylation profiles linked to disease and assess them as biomarkers. However, understanding disease biology will require equivalent data sets in additional cell types relevant to the investigated cardiovascular disease phenotypes. To overcome this limitation we will consider patient derived iPSCs differentiated to specific cell types (see below).

For functional characterisation of candidate variants we have access to both human cellular systems and animal models. Many of the participating groups have well established iPSC facilities and robust protocols for differentiation to cardiomyocytes, endothelial cells and smooth muscle cells (e.g. Daniels, Sinha, Tinker, Tao). We will coordinate with parallel efforts such as the NIHR-wide Translational CV Collaborative for pooling functional strengths in high throughput cellular and single cell assays (Deloukas). The availability of whole genome sequence will also enable to assess rare mutations in regulatory regions of the genome. For example, open chromatin maps combined with sequencing have led to the discovery of compound recessive inheritance in the TAR syndrome. Recent methods such as ATAC-Seq allow rapid construction of open chromatin maps and the consortium is well placed to obtain access of primary cells to establish common resources. High-throughput methods such as examining allele-specific effects on chromatin accessibility and massively parallel reporter assays will further aid identification of potentially functional variants. Where potentially pathogenic non-coding variants are prioritised from chromatin and reporter expression profiles, these will be characterised using cell-based CRISPR/Cas9 models followed by examination of gene expression profiles.

Zebrafish models can often provide rapid *in vivo* functional data and morpholino tools can be used for genome editing (Jamshidi). We will coordinate with the MRC Mouse Network (Munroe), a member of the International Mouse Phenotyping Consortium, to prioritise mouse knock-outs for cardiovascular disease genes and use CRISPR/Cas9 technology to

study specific variants. Informatics tools such as exomiser/genomiser can be applied for cross-species comparison with human data.

Several of the investigated rare CV conditions have established disease-specific databases for pathogenic variants (Familial Arrhythmia Network Scotland, Familial Hypercholesterolemia Mutation Database). We will work together with the GEL core informatics team to establish a GECIP database of variants affecting cardiovascular traits linked to the multi-omic data and other functional information.

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.*

In addition to a number of well established collaborations described by the disease groups we anticipate close interaction with the cross-cutting GECIP domains in functional genomics namely 'Quantitative Methods, Machine Learning and Functional Genomics', 'Functional Cross cutting', and 'Functional Effects'. In particular, we will coordinate any efforts aiming to generate omics data across all samples in the 100,000 Genomes project.

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

We envisage a constant interaction with the different cardiovascular disease subdomains of our GeCIP to identify functional genomics projects for BSc, MRes, and MSc students at all participating laboratories. For example, students enrolled in the MSc in Genomic Medicine programme funded by Health Education England and which is run by 10 Universities across the country including QMUL/UCL, KCL/SGUL, Imperial and Manchester will be hosted for their dissertation. In addition, we anticipate hosting PhD students and more senior clinical fellows that will be funded from the new HEE initiative to be launched later this year.

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

The Functional subdomain comprises national and international experts in genomic analyses, iPSC technology, manipulation of zebrafish and mouse models and bioinformatics across the range of rare CV diseases: Professor Panos Deloukas (QMUL), Professor Manuel Mayr (KCL), Professor Paulus Kirchhof (University of Birmingham), Professor Steve Humphries (UCL), Professor Tim Aitman (University of Edinburgh), Professor Patricia B Munroe (QMUL), Professor Andrew Tinker (QMUL), Dr Sanjay Sinha

(University of Cambridge), Dr Yalda Jamshidi (St George's Hospital Medical School), Dr Matthew J Daniels (Oxford University), Dr Tao Wang (The University of Manchester); Dr Andrew Smith (QMUL), Dr Fotios Drenos (University of Bristol), Professor Bernard Keavney (The University of Manchester).

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 leverage the ample expertise of geneticists and clinicians in this subdomain to annotate and interpret candidate variants but overall clinical interpretation will be the task of the specific disease subdomains. But, we anticipate that analysis of integrated omics and phenotypic data across disease phenotypes may also yield common patterns which are clinically relevant to CV disease.

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

Research output of this group will feed directly in to the specific disease subdomains and therefore will benefit patients indirectly. Nonetheless, establishing a national network for functional characterisation of candidate variants for rare CV disease and protocols for scaling up specific analyses will directly benefit researchers in the field. Informatics resources on rare variants will be relevant to the entire international disease genetics community.

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.*

Detailed research plan – CADASIL negative cerebral small vessel disease

Full proposal (total max 1500 words per subdomain)	
Title *(max 150 characters)*	CADASIL negative cerebral small vessel disease

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).*

Stroke describes a syndrome made up of a number of different, and pathophysiologically distinct, subtypes. One of these is cerebral small vessel disease. Cerebral small vessel disease (SVD) describes disease of the small perforating arteries supplying the white matter and deep grey mater nuclei in the brain. It results in a number of pathologies which are well visualised on MRI. These include lacunar infarcts, more diffuse regions of white matter ischaemia, referred to as white matter hyperintensities, and cerebral microbleeds. The major clinical features of SVD are lacunar stroke and vascular cognitive impairment and dementia.

The vast majority of cases of SVD are “sporadic” with polygenic predisposition. The most important risk factor is hypertension. However a small number of cases have Mendelian inheritance. Cerebral small vessel disease is the most common stroke phenotype caused by monogenic diseases. The most common of these is CADASIL, first identified in 1994. More recently a number of other monogenic causes of cerebral SVD have been described including CARASIL, TREX1 cerebral small vessel disease, COL4A1, COL4A2.(1)

Even when these are screened for it has become apparent that there are a number of families with a Mendelian pattern of inheritance in which the typical clinical phenotype and MRI of SVD, but in whom no mutations have yet been identified.

Whole Genome Sequencing (WGS) has opened up this area and offers the potential to answer a number of important research questions.

1) How common are mutations in known familial causes of SVD in patients presenting with suspected familial SVD. When the gene for CADASIL (NOTCH3) was first described we set up a similar programme using Sanger sequencing to answer this question for CADASIL.(2) This led to optimal genetic testing of the disease and the establishment of a national genetic service for CADASIL at St George’s where Hugh Markus worked at the time. There is now the opportunity to carry out similar research on the more recently described and rarer known causes of familial SVD such as CARASIL, TREX1, COL4A1, COL4A2, and others.

2) It is clear that there are other genes underlying familial SVD and we have already identified a number of families in whom these are likely to occur. WGS offers the opportunity to identify these novel genes.

Although rare, familial forms of SVD are telling us a considerable amount about the underlying biology and much of this appears to be relevant not only to sporadic disease but also to familial disease. For example it has been shown that diverse causes of SVD such as CADASIL, CARASIL, COL4A1, and COL4A2, seem to act via mechanisms which converge on disruption of the Extracellular Matrix (ECM) and matrisone.(3) These findings are leading to potential new treatment approaches.(4) Therefore identifying novel genes could have important outcomes for both familial and sporadic disease.

We have already established screening in patients with non-CADASIL suspected familial SVD as part of the BRIDGE project. GEL is extending the ascertainment of this phenotype throughout the NHS in England.

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.*

We will recruit patients with suspected familial SVD in whom in screening for CADASIL has shown that they do not have NOTCH3 mutations. We have already successfully recruited such individuals to the BRIDGE cerebral SVD project. To date we have recruited over 200 of the 250 spaces we have within the BRIDGE Project. Collection has been performed as part of a consortium involving neurology/stroke specialists and geneticists in a number of units including; Cambridge, St George's London, UCL London, Newcastle, Imperial London, Sheffield, Leeds, Glasgow, and Exeter. Our eligibility criteria are shown below.

It is understood that recruitment of CADASIL families will be conducted in the 100,000 Genomes Project until the cap of 50 families is reached; by this point data from the BRIDGE study will be available and will be used to inform the likely benefit of additional recruitment in GEL.

Eligibility criteria

Inclusion criteria

ALL of the follow inclusion criteria must be met.

(A) Clinical features consistent with cerebral small vessel disease: either or both of;
- a) lacunar stroke,
- b) vascular cognitive impairment/dementia

(B) MRI confirmed evidence of cerebral small vessel disease as evidenced by; multiple lacunar infarcts and/or confluent white matter hyperintensities.
(C) Mendelian pattern of inheritance
(D) Age of onset of stroke or dementia < 60 years in index case or first degree family member.
(E) No CADASIL mutations on sequencing of the exons 2-24 of NOTCH3 gene

Exclusion criteria

1. Causes of white matter disease other than cerebral small vessel disease (e.g. multiple sclerosis, vasculitis, leukodystrophy).
2. Cases who have CADASIL, or who not been fully genetically screened for CADASIL

Prior genetic testing:

NOTCH3 screening to exclude CADASIL causing mutations in all exons which can be affected by the disease (2-24) must be performed prior to recruitment.

Screening for the other known Mendelian causes of SVD is not currently widely available and this is not required as an inclusion criteria.

We will collect phenotypic information on individuals (a standard proforma has been developed for this as part of BRIDGE) and will collect MRI scans from individuals recruited. This is important so that we can stratify analysis of the results according to the MRI appearances of SVD. There are a number of MRI appearances including lacunar Infarcts, white matter hyper-intensities, and microbleeds, as well as intercerebral haemorrhage and it is likely that different genes may predispose to differing patterns of disease.

We will use the data collected, and the results from the WGS, to answer the following research questions.

1) How common are mutations in known genes causing familial SVD (excluding CADASIL).?

2) What is the clinical and radiological phenotype of these known causes of SVD?

By collecting detailed clinical information and MRI information we will be able to gain important information on the phenotype of these diseases. Currently there is limited data available worldwide. This may lead to important advances in diagnosis. For example using a similar approach we identified diagnostic MI features of CADASIL which have played a major role in more wide spread recognition of the disease.

3) Can we identify novel genes associated with familial SVD?

If such genes are identified we will follow this up with studies identifying the clinical and radiological phenotype, and functional studies.

A major challenge is determining whether variants discovered are pathogenic. We will address these by using bioinformatic approaches by recruiting wherever possible multiple family members, and for known genes by working with groups who have extensive experience working with specific monogenic forms of SVD.

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.*

All analyses will be conducted in conjunction with the analytics domain and BRIDGE and GEL Bioinformatics support.

We will collaborate with centres with expertise in interpretation of variants within genes causing known forms of familial SVD as in the section above.

We will collaborate with centres carrying out similar studies in other countries through the International Stroke Genetics Consortium (ISGC).

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

We have a number of trainees who are working with the groups. This includes a PhD student (Rhea Tan) who is coordinating the BRIDGE SVD project. We also have a number of Postdoc statisticians.

Trainees from across the GeCIP lab will be encouraged to use the dataset for PhD and Fellowship applications.

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

We have already established a group interested in recruiting patients with familial SVD, and interpreting the WGS findings, as part of the BRIDGE cerebral SVD project. This includes clinicians from Glasgow, Newcastle, Sheffield, Leeds, Cambridge, UCL, St George's London, as well as clinical genetics laboratories in Exeter and St George's.

We are using an informatics pipeline developed as part of the BRIDGE 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.*

We will provide expert clinical and genetic input to the validation and feedback domain where requested to assist in resolving the specific question of whether a given variant can be considered clinically actionable and be fed back to patients. We anticipate that a number of variants will be already known variants in known genes to cause SVD. There will be additional variants in the same genes which will be of unknown significance, we will help in interpretation of these using expertise of the other collaborating groups (above) collaborating in these areas. It seems likely that many variants come up particularly in not previously implicated genes, may be deemed of unknown significance. Here the availability of replication within the International Stroke Genetics Consortium will be particularly useful in establishing variant pathogenicity.

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

Participants and their families will benefit as clinically actionable variants may be identified by the GeCIP research.

Patients with suspected familial SVD who are not enrolled in the project may benefit as genes will be identified that could be of relevance to these families. In addition the information on genotype-phenotype correlations derived from this project will aid in identifying relevant genes in families in future and provide information on an actual cause of these diseases.

The NHS will benefit because we will be able to determine the yield of screening in this patient group and therefore provide clinical algorithms which can be used in the management of such patients and will identify those patients in whom performing NGS is a benefit.

Other researchers and healthcare professionals worldwide will benefit from the availability of the curated database of variants of likely or confirmed pathogenic significance, which can be consulted when they encounter patients similar to those enrolled in the project.

It is possible that novel variants we identify will provide information on the underlying mechanisms for both Monogenic SVD and also for Sporadic SVD.

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?*

None currently in place.

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

1.Tan RY, Markus HS. Monogenic causes of stroke: now and the future. J Neurol.2015 Dec;262(12):2601-16.

2.Markus HS, Martin RJ, Simpson MA, Dong YB, Ali N, Crosby AH, Powell JF. Diagnostic strategies in CADASIL. Neurology. 2002 Oct 22;59(8):1134-8.

3.Joutel A, Haddad I, Ratelade J, Nelson MT. Perturbations of the cerebrovascular matrisome: A convergent mechanism in small vessel disease of the brain? J Cereb Blood Flow Metab. 2016 Jan;36(1):143-57.

4.Capone C, Cognat E, Ghezali L, Baron-Menguy C, Aubin D, Mesnard L, Stöhr H, Domenga-Denier V, Nelson MT, Joutel A. Reducing Timp3 or vitronectin ameliorates disease manifestations in CADASIL mice. Ann Neurol. 2016 Mar;79(3):387-403.

Full proposal (total max 1500 words per subdomain)	
Title *(max 150 characters)*	**Sudden Unexplained Death in the Young (SUDY)/ Sudden Infant Death Syndrome (SIDS)**

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).*

Sudden Unexplained Death in the Young (SUDY) is a term describing unexpected and unexplained death in a previously healthy person aged 1-35 years following a negative post mortem (including toxicology).[1] The definition typically requires the death to be witnessed or to have occurred within 12 hours of them last being seen alive and to have no prior recorded cardiac disease. SUDY may be the first manifestation in a person (or family) of a previously undiagnosed inherited arrhythmia syndrome ("channelopathy"); encompassing a number of cardiac conditions including long QT syndrome (LQTS), Brugada syndrome (BrS) and catecholaminergic polymorphic ventricular tachycardia (CPVT).[1] These conditions are due to mutations in genes encoding cardiac ion channel proteins. They are heterogeneous conditions, typically with an underlying structurally normal heart, and a predisposition to life-threatening ventricular arrhythmias.

Post mortem genetic testing in SUDY cases is also known as a molecular autopsy with variable yields being reported.[2,3] A recent study using next-generation sequencing in 302 SUDY cases reported a yield of pathogenic variants of 13% (predominantly in CPVT and LQTS genes).[2] Current guidelines recommend the consideration of molecular autopsy following SUDY.[1] Given the familial basis to these diseases, first degree family members are recommended to undergo clinical cardiac screening. Recent data shows a combined evaluation of molecular autopsy and family screening increases diagnostic yield to 39%. [2]

Sudden Infant Death Syndrome (SIDS) is defined as "the sudden death of an infant under 1 year of age which remains unexplained after thorough investigation including post-mortem and review of the clinical history". Improved education and public health campaigns have reduced the prevalence of SIDS in the UK to approximately 1 in 2000 live births however this has plateaued in recent years.[4] SIDS remains the most common cause of death for infants in the developed world. There is a large body of evidence to suggest that some SIDS deaths are due to underlying genetic causes. The "triple-risk hypothesis" was proposed in 1994 and encompasses environmental trigger (e.g. sleeping in a cot), critical development period and vulnerable host (i.e. increased genetic susceptibility above that of the general population, such as a young infant with a previously undetected inherited cardiac condition).[5] Previous twin studies have shown a clear genetic basis underpinning some elements of SIDS deaths with one study demonstrating the second twin has a fourfold increased chance of SIDS after the death of the first twin.[6]

The potential pathogenic link between SIDS cases and underlying inherited cardiac conditions (specifically long QT syndrome) was first highlighted in 1998.[7] This study assessed day 3-4 ECGs on neonates and identified long QT intervals in those who subsequently experienced a SIDS death. SIDS case studies and cohort studies have shown evidence of mutations in inherited arrhythmia syndrome genes in infants who had died following SIDS or from infants with successful resuscitation with documented ventricular arrhythmias following "near-SIDS".[8-9] Other reports have suggested underlying putative pathogenic variants in up to 9.4% with the greatest yield seen in *SCN5A*.[10,11] A recent study using whole exome sequencing in 161 European SIDS cases identified rare cardiac genetic variants in 20% of cases.[12] Overall 80-90% of SIDS cases remain "genotype-negative" with a proportion of these deaths suspected to be due to undiagnosed inherited cardiac

and non-cardiac conditions.[13] Current guidelines advise that first degree relatives of SIDS cases can be evaluated for evidence of cardiac disease. The data for this is however extremely weak. [1]

Small SIDS studies have suggested that rather than being due to a single pathogenic variant, in fact the underlying inherited cardiac condition is due to increased burden of common variation. For example in a cohort of Caucasian infants, polymorphisms in *KCNH2* (LQT2 gene) were over-represented in SIDS cases compared to controls by almost 5%.[14] A similar study in black American infants showed a homozygous *SCN5A* (LQT3/BrS gene) polymorphism was associated with a 24-times increased risk of SIDS compared with controls.[15] It is possible that more complex genetics, with an underlying oligogenic basis for disease, accounts for at least a proportion of SIDS cases.

Therefore, the genetics underpinning SIDS and SUDY remains incompletely understood at present and there is an unmet clinical need to determine risk in future pregnancies and family members. The role for common variants requires further study and research. This subdomain project will focus on novel gene and risk pathway discovery in SUDY/SIDS including cardiac, respiratory and neuromuscular genes. This prioritisation is partly based upon unpublished data from exome-wide case control studies from a trans-Atlantic collaboration between SGUL and the Mayo clinic and our own pre-screening of over 500 cases of SIDS and SUDY for established cardiac risk genes. Parents of these cases will be recruited as part of the effort, allied to their clinical evaluation to determine if there is a suitable role for this approach in SIDS. A secondary focus will be on increased burden of common variants in cases compared with controls.

Hypotheses:

1. WGS based trio studies in SUDY/SIDS victims (ages 0-35) and their parents will identify mutations, particularly de novo mutations in genes known to cause underlying inherited cardiac conditions whilst others will be in non-cardiac genes such as respiratory or neuromuscular genes.
2. Clinical cardiac evaluation in families offers little in the way of additional utility in diagnosing the cause of a SIDS death.
3. Common genetic variation predisposes to the risk of SUDY/SIDS.

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) WGS Trio studies

Potential cases aged 0-35 years will be identified from coroner's cases or referred from pathologists to the inherited cardiac conditions clinic at GMC centres, with the study team actively promoting GEL recruitment in the South London GMC and collaborating with the North-West London GMC and the Lullaby trust. Whole genome sequencing (WGS) will be undertaken in trios (concurrent deceased, mother and father) allowing for assessment for familial and *de novo* variants of all potential SUDY/SIDS genes including cardiac, neurological, respiratory and musculoskeletal genes.

There are a significant proportion of SUDY/SIDS cases which may be due to underlying genetic causes that have been missed by previous genetic testing methodologies. There are a number of possible reasons for this knowledge gap; it may be that there are causative genes not yet investigated, or, more likely, there are important genetic variants in gene-regulatory regions such as promoter regions, 3'-untranslated regions or non-coding regions that are not targeted for sequencing with current technologies. Whole genome sequencing provides the opportunity to

sequence the genome beyond the coding regions and discover new genetic variants in previously unexplored regions.

Variants identified as strict ultra-rare variants (allele frequency<0.00005) will be assessed for pathogenicity. Relevant factors when determining pathogenicity include: the predicted effect on protein function (e.g. nonsense, frameshift, splice-site), rarity in large population datasets (e.g. Exome Aggregation Consortium, gnomAD), prediction from in silico software tools and previous functional studies in cell or animal models suggesting significant physiological effects of the variants. We will also compare parent's data with the deceased infant's DNA. Any variants present in the deceased DNA and absent in the parent's DNA will be considered as potential *de novo* SIDS/SUDY causative variants.

It is expected that if a clear causative mutation is identified in a parent or case and permission has been given, a clinical referral will be made if necessary. Otherwise, novel disease-causing variants may be identified using this methodology, potentially leading to the discovery of new genes and mechanisms of disease.

Cardiac evaluation in family members
Following detailed personal and family history and physical examination, detailed clinical information will be retrospectively and prospectively recorded to assess key phenotypic parameters for inherited cardiac conditions (e.g. ECG, echo and ETT as basic investigations and others where indicated). Probable and definite clinical diagnoses will be based on established diagnostic criteria. If there a pathological phenotype is confirmed, first degree family members will be recommended clinical screening with their local inherited cardiac conditions clinics.

Replication:
There is an established transatlantic collaboration for SIDS exome sequencing studies (SGUL and the Mayo Clinic) providing a platform of cases for ongoing replication studies. Our internal SUDY cohort (n=500) will also allow for replications of novel loci identified in our studies. Collaboration with neurology colleagues studying epilepsy and musculoskeletal phenotypes (Sisodiya and Matthews, UCL) is already in place. Novel genes and variants will be considered for in vitro functional studies in animal, cell and in silico models. A number of models are readily available to screen for potential effects including zebra fish, heterologous expression systems with cellular patch clamping and rat myocyte culture.

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.*

All analyses will be conducted in conjuction with the analytical methods domain and GEL bioinformatics support. Collaboration with the neurology GeCIP will be considered if novel suitable neurological targets are identified.

We have ongoing GWAS studies in collaboration with centres in USA, Australia, Denmark and across the UK, using anonymised DNA samples from SIDS victims to look for increased burden of common genetic variation. There is also the established transatlantic SIDS exome sequencing project as described above.

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

Trainees from across the GeCIP labs are being encouraged to use GECIP datasets as the basis of PhD and or fellowship applications. Dr Gray, a post-CCT fellow on an Australian NHMRC Early Career award, will be working on the project and recruitment.

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

The GeCIP includes the main clinical cardiology and genetic leads in management of sudden death cases and SIDS in the UK, all of whom are also engaged in local GMCs, alongside researchers with substantial expertise in gene discovery, statistical genetics and functional characterisation.

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.*

The CV domain GeCIP has overlapping membership with the variant interpretation domain, as well as other collaborative international projects such as ClinGen and Clinvar. Members have established diverse methods to improve variant interpretation through international data aggregation as well as functional and computational studies.

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

New genetic discoveries will be of benefit to the affected families by giving them closure to the cause of the sudden death of their loved one. In addition, the identification of a genetic basis underpinning the SUDY/SIDS death has implications for future pregnancies and other first degree family members who may also be at risk of inherited cardiac conditions.
Novel genomic associations with sudden death risk may also have implications for other patients in terms of risk stratification and prevention of future sudden death events

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?*

None as yet

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

1. Priori SG, Wilde AA, Hoei M, Cho Y, Behr ER, Berul C, Blom N, Brugada J, Chiang CE, Huikuri H, Kannakeril P, Krahn A, Leenhardt A, Moss A, Schwartz PJ, Shimizu W, Tomaselli G, Tracy C. HRS/EHRA/APHRS expert consensus statement on the

diagnosis and management of patients with inherited primary arrhythmia syndromes. Heart Rhythm 2013;10(12):1932-63
2. Lahrouchi N, Raju H, Lodder EM, et al. Utility of Post-Mortem Genetic Testing in Cases of Sudden Arrhythmic Death Syndrome. J Am Coll Cardiol May 02 2017;69:2134-2145.
3. Bagnall RD, Weintraub RG, Ingles J, et al. A Prospective Study of Sudden Cardiac Death among Children and Young Adults. N Engl J Med Jun 23 2016;374:2441-2452.
4. Moon RY, Horne RS, Hauck FR. Sudden infant death syndrome. Lancet Nov 03 2007;370:1578-1587.
5. Filiano JJ, Kinney HC. A perspective on neuropathologic findings in victims of the sudden infant death syndrome: the triple-risk model. Biol Neonate 1994;65:194-197.
6. Getahun D, Demissie K, Lu SE, Rhoads GG. Sudden infant death syndrome among twin births: United States, 1995-1998. J Perinatol Sep 2004;24:544-551.
7. Schwartz PJ, Stramba-Badiale M, Segantini A, Austoni P, Bosi G, Giorgetti R, Grancini F, Marni ED, Perticone F, Rosti D, Salice P. Prolongation of the QT interval and the sudden infant death syndrome. N Engl J Med Jun 11 1998;338:1709-1714.
8. Schwartz PJ, Priori SG, Dumaine R, Napolitano C, Antzelevitch C, Stramba-Badiale M, Richard TA, Berti MR, Bloise R. A molecular link between the sudden infant death syndrome and the long-QT syndrome. N Engl J Med Jul 27 2000;343:262-267.
9. Schwartz PJ, Priori SG, Bloise R, Napolitano C, Ronchetti E, Piccinini A, Goj C, Breithardt G, Schulze-Bahr E, Wedekind H, Nastoli J. Molecular diagnosis in a child with sudden infant death syndrome. Lancet Oct 20 2001;358:1342-1343.
10. Wedekind H, Smits JP, Schulze-Bahr E, et al. De novo mutation in the SCN5A gene associated with early onset of sudden infant death. Circulation Sep 04 2001;104:1158-1164.
11. Ackerman MJ, Siu BL, Sturner WQ, Tester DJ, Valdivia CR, Makielski JC, Towbin JA. Postmortem molecular analysis of SCN5A defects in sudden infant death syndrome. JAMA Nov 14 2001;286:2264-2269.
12. Neubauer J, Lecca MR, Russo G, Bartsch C, Medeiros-Domingo A, Berger W, Haas C. Post-mortem whole-exome analysis in a large sudden infant death syndrome cohort with a focus on cardiovascular and metabolic genetic diseases. Eur J Hum Genet Apr 2017;25:404-409.
13. Wong LC, Behr ER. Sudden unexplained death in infants and children: the role of undiagnosed inherited cardiac conditions. Europace Dec 2014;16:1706-1713.
14. Tester DJ, Ackerman MJ. Sudden infant death syndrome: how significant are the cardiac channelopathies? Cardiovasc Res Aug 15 2005;67:388-396.
15. Plant LD, Bowers PN, Liu Q, Morgan T, Zhang T, State MW, Chen W, Kittles RA, Goldstein SA. A common cardiac sodium channel variant associated with sudden infant death in African Americans, SCN5A S1103Y. J Clin Invest Feb 2006;116:430-435.

Application Summary	
GeCIP domain name	**Cardiovascular**
Project title *(max 150 characters)*	**Rare and Common gene variants for Blood Pressure Phenotypes (rare and common variants)**

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

1. **To elucidate the genetic variants in syndromic forms of blood pressure**
2. **To evaluate these in the population blood pressure across the 100,000 Genomes Cohort when blood pressure phenotypes are present**
3. **To take these into other cohorts to improve power to detect associations e.g. UK Biobank.**

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)*

High Blood Pressure is the commonest risk factor for heart disease and stroke. Rare inherited forms of blood pressure have been identified and in these families these gene discoveries explain why specific therapies can be used to treat their blood pressure. We seek to identify new variants that affect blood pressure (high or low) as these may identify ways to improve blood pressure control and help us to choose therapies that people are most likely to benefit from first time.

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)*

High Blood Pressure is the commonest cardiovascular risk factor and causing an estimated 17 million deaths per year. There are now over 1000 gene loci for blood pressure that have been identified mostly based upon common genetic variants. This project aims to investigate the role of genomewide rare and common variation using primary clinical data in syndromic high or low blood pressure and across the entire Genomics England Dataset using all primary care and secondary care datasets and where possible actual blood pressure records. As part of extending the 100,000 Genomes Project value we will investigate whether rare syndromic or oligogenic effects on blood pressure exist. We also seek approval to do the same across the 500,000 whole genomes to be sequenced via NHS care. The goal is to identify new biological mechanisms and prime new therapeutic innovation for improved blood pressure control for patients.

Expected start date	**February 2019**
Expected end date	**January 2024**

Full proposal (total max 1500 words per subdomain)	
Title *(max 150 characters)*	**Rare and Common gene variants for Blood Pressure Phenotypes (rare and common variants)**

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).*

High blood pressure (BP) is a leading heritable risk factor for stroke and coronary artery disease, responsible for an estimated 7.8 million deaths and 148 million disability life years lost worldwide in 2015 alone[1]. Blood pressure is determined by complex interactions

between life-course exposures and genetic background[2-4]. Previous genetic association studies have identified and validated variants at 274 loci with modest effects on population BP, explaining in aggregate ~3% of the trait variance[5-12]. We recently reported a genome-wide discovery analyses of BP traits - systolic (SBP), diastolic (DBP) and pulse pressure (PP) - in people of European ancestry drawn from UK Biobank (UKB)[13] and the International Consortium of Blood Pressure-Genome Wide Association Studies (ICBP)[11,12]. In a combined one- and two-stage study design to test common and low-frequency single nucleotide polymorphisms (SNPs) with minor allele frequency (MAF) > 1% associated with BP traits. To do this we studied over 1 million people of European descent, including replication data from the US Million Veterans Program N=220,520)[14] and the Estonian Genome Centre, University of Tartu (EGCUT, N=28,742) Biobank[15].

In this studied we identified and confirmed over 1,000 independent signals at 901 loci for BP traits, and the 535 novel loci more than triples the number of BP loci and doubles the percentage variance explained, illustrating the benefits of large-scale biobanks. By explaining 27% of the estimated heritability for BP, we make major inroads into the missing heritability influencing BP level in the population[31]. The novel loci open the vista of entirely new biology and highlight gene regions in systems not previously implicated in BP regulation. This is particularly timely as global prevalence of people with SBP over 110-115 mm Hg, above which cardiovascular risk increases in a continuous graded manner, now exceeds 3.5 billion, of whom over 1 billion are within the treatment range.

The analyses to date have been almost entirely focused on common variants and it is likely amongst the 100,000 Genomes Project we will identify rare variants in both syndromic forms of blood pressure and the entire enrolled population, which will explain new biology and offer new therapies.

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.*

Analyses, techniques and design

We will investigate syndromic blood pressure disorders enrolled in the 100,000 Genomes Project which have Human Phenotype Ontology Terms that indicate high or low blood pressure or hypertension. We will undertake family by family analysis within syndrome and then cohort-wide analyses for burden testing to provide additional evidence for a role in blood pressure. As the primary care data becomes available we will be able to use blood pressure data alongside other secondary data (Hospital Episodes and outcomes) to correlate our findings with morbidity and mortality. Blood pressure data would need to be adjusted for age, age2, body mass index and treatment and if necessary batch effects as some whole genomes are on different chemistries.

Timelines

We would aim to start and conclude the initial syndromic analysis quite quickly over the next 6 months. However, until we have the primary care data and the constituent blood pressure data it will be difficult to specify timelines but hopefully over the next year. We aim to make this analysis continuous across the 500,000 whole genomes to be done in the new service

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.*

Major planned academic collaborators from the Cardiovascular GeCIP:
Sub domain: Functional Effects: Panos Deloukas
The Project lead: Mark Caulfield
CV GeCIP Co-researcher (some of these may not be GeCIP Members: Claudia Cabrera, Michael Barnes, Helen Warren, Fu Ng, Bori Mifsud, Morris Brown, Patricia Munroe, Emma MacGavern, Matthew Traylor (all from Queen Mary)

Renal GeCIP: Danny Gale
Electronic Health GeCIP: Harry Hemingway,

Bori Mifsud: b.mifsud@qmul.ac.uk
Claudia Cabrera: c.cabrera@qmul.ac.uk
Fu Ng: f.ng@qmul.ac.uk
Helen Warren: h.r.warren@qmul.ac.uk
Morris Brown: morris.brown@qmul.ac.uk
Patricia Munroe p.b.munroe@qmul.ac.uk
Michael Barnes m.r.barnes@qmul.ac.uk
Stephen Walsh <ucgbsbw@ucl.ac.uk>
Bernard Keaveney bernard.keavney@manchester.ac.uk
Emma Magavern e.magavern@smd18.qmul.ac.uk

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

Emma McGavern is an Academic Clinical Fellow in Clinical Pharmacology and she is undertaking the Masters in Genomic Health.
Others are welcome

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

Deloukas is a world-leader in cardiovascular complex trait genetics and top 200 most highly cited researcher.
Caulfield, Munroe and Brown are world leaders in genetics of blood pressure either as a complex trait, syndromic blood pressure and adrenal hypertension
Claudia Cabrera, Michael Barnes, Helen Warren, Fu Ng, Bori Mifsud and Matthew Traylor have all published in Nature and other journals on blood pressure genetics.

The investigators have worked and published together productively and possess the expertise required to carry on this project. They will work together via regular group meetings and teleconferences with collaborators.

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

We will investigate the 100,000 Genomes to identify new variants that affect blood pressure. The most recent analyses focus almost entirely on common variants; however, with the availability of the whole genome we have the potential of identifying new rare variants. We will investigate each family and perform burden testing using R, followed by appropriate statistical correction for

multiple testing. We will characterize and annotate the functional variants using standard bioinformatics open-source tools (e.g. bedtools, ANNOVAR, exomiser, GREAT). We seek to integrate the new discoveries with the previously reported blood pressure loci (n>1000) into an systems biology approach, which will explain new biology and potentially offer new therapies. We will then extend the analysis to a cohort-wide spectrum. We wish to use R data libraries of clustering and machine learning algorithms to investigate the cohort's properties and possible hidden structure. These properties could inform us further on the parameters for the downstream analysis. In addition, with the availability of the blood pressure information we would investigate the blood pressure genomic associations using the software PLINK and BOLT and adjusting for age, age^2, body mass index, treatment and batch. With this approach, we expect to identify new biological mechanisms and improve blood pressure control therapeutics for patients

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

It will give more biological insights and may identify or tailor therapy for patients, the NHS and open new avenues of research for scientists across the world.

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 a long way from a drug or repurposing opportunity but if something did arise we would discuss it with Genomics England and consider how best to work with industry. We do not have commercial partners at present

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

1. Forouzanfar, M.H. *et al.* Global Burden of Hypertension and Systolic Blood Pressure 942 of at Least 110 to 115 mm Hg, 1990-2015. *JAMA* **317**, 165-182 (2017).
2. Munoz, M. *et al.* Evaluating the contribution of genetics and familial shared environment to common disease using the UK Biobank. *Nat Genet* **48**, 980-3 (2016).
3. Poulter, N.R., Prabhakaran, D. & Caulfield, M. Hypertension. *Lancet* **386**, 801-12 946 (2015).

5. Cabrera, C.P. *et al.* Exploring hypertension genome-wide association studies findings and impact on pathophysiology, pathways, and pharmacogenetics. *Wiley Interdiscip Rev Syst Biol Med* **7**, 73-90 (2015).
6. Ehret, G.B. *et al.* The genetics of blood pressure regulation and its target organs from association studies in 342,415 individuals. *Nat Genet* **48**, 1171-1184 (2016).
7. Surendran, P. *et al.* Trans-ancestry meta-analyses identify rare and common variants 955 associated with blood pressure and hypertension. *Nat Genet* **48**, 1151-1161 (2016).
8. Liu, C. *et al.* Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci. *Nat Genet* **48**, 1162-70 (2016).
9. Hoffmann, T.J. *et al.* Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nat Genet* **49**, 54-64 (2017).
10. Warren, H.R. *et al.* Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nat Genet* **49**, 403-415. (2017).
11. Wain, L.V. *et al.* Novel Blood Pressure Locus and Gene Discovery Using Genome-Wide Association Study and Expression Data Sets From Blood and the Kidney. *Hypertension* (2017).

12. International Consortium for Blood Pressure Genome-Wide Association Studies *et al.* Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103-9 (2011).
13. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* **12**, e1001779 (2015).
14. Gaziano, J.M. *et al.* Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol* **70**, 214-23 (2016).
15. Leitsalu, L. *et al.* Cohort Profile: Estonian Biobank of the Estonian Genome Center,University of Tartu. *Int J Epidemiol* **44**, 1137-47 (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)*

Participants
All participants within the Cardiovascular domain.

Within other GECIP domains:

- Ciliopathies involving congenital heart disease
- Dysmorphic and congenital abnormality syndromes involving the heart and vessels (for example RASopathies which may present hypertrophic cardiomyopathy or congenital heart disease).
- Neurodevelopmental disorders where participants also have congenital heart disease.
- Ultra-rare undescribed monogenic disorders involving the heart and vessels.
- Participants recruited with other disorders who are unexpectedly found to have rare and potentially deleterious alleles in ICC genes.

Agreement not yet sought from other GECIP domains in respect of these participants.

Data required
- Phenotype data
- Sequence data on proband and all family members.
- Data form will vary between participating groups and by subdomain; we anticipate filtered variant lists, VCF files and BAMfiles will all be required at different times.

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

1. Testing of 100,000 Genomes candidate variation in population resources held by GECIP members and collaborators. If a variant of unknown significance in a new gene for a particular disorder is identified in the Project, confirmation of the same gene in other families with the disorder not enrolled in the Project will be the key first step towards establishing likely pathogenicity. GECIP participants have extensive patient and family collections in all subdomains

of the Cardiovascular GEL ascertainment which chiefly for consent related reasons cannot currently contribute to the GEL Project; nevertheless, these resources could be used to confirm potentially pathogenic candidate genes.

2. Development of novel bioinformatics approaches to the analysis of 100,000 Genomes candidate variation to identify likely causative variants.

3. Functional characterisation of candidate variants using molecular biological, cell biological and where appropriate animal modelling approaches. Where a particular cell type is clearly implicated in disease pathogenesis, iPSC generation followed by differentiation and in-depth cellular phenotyping (or, as technology permits, genome editing using CRISPR/Cas9 to generate mutant cell lines).

4. In-depth re-phenotyping, including multi-omics methods. This will involve high-fidelity clinical characterisation (imaging, electrophysiological, and where appropriate invasive), use of RNA and protein data derived from clinical samples, metabolomics and epigenetic studies, and the curation of a developmental cardiac genetic expression atlas.

5. Detailed longitudinal followup of patients by health record linkage. Investigation of genotype/outcome relationships.

6. Through identifying novel mechanisms from the 100,000 Genomes data, to implement approaches to drug repurposing for rare cardiovascular disease; to use high throughput technology to identify new compounds relevant to rare disease.

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)*

Case and family history

Cardiovascular phenotypes: clinical examination, ECG, imaging data (typically echocardiographic, cardiac MRI and CT appropriate for some conditions) other specialised tests for specific conditions (eg lymphoscintigraphy and venous duplex scan for primary lymphoedema patients).

Non-cardiovascular phenotypes: particularly intellectual impairment, musculoskeletal phenotypes or metabolic derangements, ocular or skin conditions

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)*

No differences currently identified

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)*

A mechanism whereby our own exome sequencing data could be compared with GEL WGS data on the same platform would be helpful. If it will only be possible to access GEL data within the Data Embassy, could it also be possible to upload our data to a shared space? Will tools to allow the analysis of particular fractions of the genome (eg exome captures) be available?

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)*

In addition to phenotypic data collected at baseline by the enrolling GMC, it will be necessary to import additional clinical data which may include imaging, catheterisation results, electrophysiological investigations, exercise tests etc. For patients who have phenotypes spanning more than one domain (eg neurodevelopmental plus congenital heart disease), it will be necessary to access clinical phenotyping data from all the systems involved in the patient's condition.

During the functional analysis it will be necessary to import data which might include transcriptomic profiles, epigenetic profiles, proteomic and metabolomics data, cellular phenotypic data derived from iPSC modelling, and data from novel in vitro analyses eg nanomechanical phenotyping of aortopathy genotypes.

For prospective analyses of outcomes, data from electronic health records will require to be uploaded and updated periodically. If clinical trials of novel genotype-guided therapies are conducted, a facility to upload data on biomarker endpoints and disease endpoints will be required.

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

As the genomic/epigenomic/transcriptomic/proteomic dataset grows in parallel with dense phenotyping and lifecourse information, integrative analyses of these variables seems likely to place the highest demand on computing and storage resources, although these are difficult to quantify at present.

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)*

Metabolomic analyses: eg of coronary sinus blood in hypertrophic cardiomyopathy
Transcriptomic/proteomic analyses: eg of myocardial biopsy samples in dilated cardiomyopathy.
Blood based eQTL analyses for novel candidate loci in arrhythmias .
Epigenomic characterisation of congenital heart disease
Proteomic and transcriptomic approaches to lymphovascular malformation.
Development of biomarkers of elevated risk in aortopathy. In vitro nanomechanical analyses of aneurysm risk genotypes.
Metabolomic and lipidomics profiling of patients with different molecular causes of familial hypercholesterolemia.

Data access and security	
GeCIP domain name	Cardiovascular
Project title *(max 150 characters)*	Realising Genomic Cardiovascular Medicine

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).*

X *Clinical care*

X *Clinical trials feasibility*

X *Deeper phenotyping*

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

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

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

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

X *Non hypothesis driven R&D - health*

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

X *Other health use - clinical audit*

X *Public health purposes*

X *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)

No additional attachments. CVs from new members have been supplied as they have signed up.