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

Application Summary	
GeCIP domain name	**Neurology and Neurodegeneration**
Project title *(max 150 characters)*	**Translational Neurogenetics: Using Whole Genome Sequencing to Diagnose, Understand and Develop Treatments for Neurological Disorders**

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

The major objectives of the Neurology and Neurodegeneration research plan are to advance and optimise whole genome sequencing validation, interpretation and interrogation. This will significantly accelerate our understanding of disease pathophysiology and key mechanistic pathways enabling the future development of new therapies.

Specific aims:

1. Development of data capture and similarity scoring algorithms to delineate rare and severe neurological and neurodegenerative disorders into homogeneous phenotypic groups. This will improve gene identification, enable genotype-phenotype correlations and future stratification of patients into 'trial ready' groups.
2. Identification of novel disease genes, genetic risks and modifying factors to enable comprehensive NHS diagnostic testing, prediction of disease onset and severity.
3. Gene discovery will lead to the identification of novel pathways that will be investigated and modeled to advance our understanding of disease pathophysiology and mechanisms.
4. Training to develop the next generation of NHS technologists, scientists and clinicians in genomic medicine to sustain a thriving effective team for the future.
5. Future collaboration with industry to use pathway discovery and identify new medicines, vaccines, and pharmacogenomics to deliver precision patient treatments. The overlap between "sporadic" and Mendelian disease in some areas means that pathway directed therapies will expand the use of genome discovery to benefit multifactorial and acquired disease.
6. Genomes have been recruited to the 100,000 genomes project for two specific categories of mental illness: "Schizophrenia Plus Other Features" and "Severe Familial Anorexia". In addition, there are hundreds of other participants with records of mental illness. The neurology domain will also study mental health conditions, enabling them to identify genetic associations with cohorts of patients with mental illness, as well as to perform research using only the clinical records and associated demographic data.

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)*

Neurological conditions like dementia, Parkinson's and motor neuron disease are a major, and increasingly prevalent cause of death, but unlike cancer and cardiovascular disorders, there are no disease modifying treatments. Historically, there has been far less funding for research in neurology as compared to cancer. The overarching objective of the Neurology and Neurodegeneration GeCIP research plan is to build on our well-established NHS, research and collaborative infrastructure to enhance the clinical interpretation and validation of whole genome sequencing. We will collaborate widely with other GeCIPs and form early partnerships with industry to convert genetic findings into therapeutic benefit. This will ensure that future NHS patients with neurological and neurodegenerative diseases receive unrivalled diagnostic, genome-driven precision healthcare.

Over the next two years, our linked Genome Medicine Centres (GMCs) will collect DNA and "omic" samples on over 8,000 trios and families with neurological and neurodegenerative diseases for whole genome sequencing. Our GeCIP will optimise genome validation and

interpretation, maximise the research potential of whole genome sequencing and improve our understanding of the genetic architecture of rare neurological disease, exploiting this information to identify key pathways forming the foundations for future mechanistic studies and novel therapeutic interventions. Imperative for the future, we will train the next generation of NHS scientists, analysts and clinicians in genomic medicine to sustain this thriving healthcare initiative.

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

To maximise the collection of rare and aggressive neurological disorders, over the next two years and beyond, the Neurology GeCIP will work closely with the GMCs to enable sequencing from over 8,000 individuals as trios and families. The GeCIP will optimise validation, interpretation, interrogation of variants and training in genomic medicine.

Experimental Aims

1 – Capturing disease phenotypes (Disease specific and phenotype research plan).

We will develop and collaborate with the cross-cutting electronic records and advanced analytic domains to develop data capture and similarity scoring algorithms. Demarcate rare and severe neurological and neurodegenerative disorders into homogeneous phenotypic groups, based on the clinical features, inheritance and investigations such as brain MRI. This will improve our efficiency in the identification of disease genes, genetic risk and modifying factors.

2 - Identification of disease causing variants, filtering strategies and the creation of resources for variant prioritisation and the development of new NHS diagnostic tests (see research plans).

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

3 – Investigation of disease pathways and creation of disease models.

Gene discovery will lead to the identification of novel pathways that will be investigated and modeled to advance our understanding of disease pathophysiology and mechanisms. Depending on the putative gene, a wide range of different biochemical and cellular approaches are available to validate pathogenicity of novel disease genes through collaboration with cross-cutting GeCIPs and in-house. For example, if a channel gene defect is highly likely then we will carry out mutagenesis and express the channel in an *in-vivo* system such as the oocyte and patch clamp the mutant and wild type channel (e.g. UCL Kullmann laboratory).

4 – Training (see research plan)

Genomic medicine is a relatively new area, particularly in neurology. We plan to train clinical fellows through PhD fellowships, non-clinical PhD studentships, genetic counsellors within associated GM's, NHS technologists where a higher degree such as an MSc can be taken as part of this training and NHS clinical scientists. This aspect of our research plan will be imperative for the delivery of future benefit to the NHS and public.

Expected outputs

1. New genes, genetic risk and disease modifying factors for neurologic disease leading to a wave of new NHS diagnostic tests for neurological disorders. **2.** The identification of novel pathways and advanced understanding of disease pathophysiology. **3.** Advanced and systematic capture of deep phenotypes. **4.** Training of a number of NHS staff, clinicians and scientists in genomic medicine.

Expected start date	**01/06/2016**
Expected end date	**31/05/2021**

Lead Applicant(s)	
Name	Henry Houlden (GeCIP co-lead)
Post	Professor of Neurology and Neurogenetics
Department	Molecular Neuroscience and Neurogenetics Laboratory
Institution	The National Hospital for Neurology and Neurosurgery and UCL Institute of Neurology, Queen Square, London WC1N 3BG.
Current commercial links	No formal links but collaborations

Name	Patrick Chinnery (GeCIP co-lead)
Post	Professor of Neurology
Department	Neurology and MRC Mitochondrial Biology Unit
Institution	University of Cambridge
Current commercial links	No formal links but collaborations

Name	Huw Morris (GMC and training lead)
Post	Professor of Neurology
Department	Clinical Neurology
Institution	Royal Free London, The National Hospital for Neurology and Neurosurgery and UCL Institute of Neurology.
Current commercial links	No formal links but collaborations

Administrative Support	
Name	Ifrah Lidow
Email	ifrah.lidow@uclh.nhs.uk
Telephone	020 3448 4068

Subdomain leads		
Name	**Subdomain**	**Institution**
James Rowe	Dementia and MND	University of Cambridge
Simon Mead	Dementia and MND	UCL
Chris Shaw	Dementia and MND	King's College London
Patrick Chinnery	Mitochondrial Disorders	University of Cambridge
Arjune Sen	Paroxysmal Neurological Disorders	University of Oxford
Sanjay Sisodiya	Paroxysmal Neurological Disorders	UCL
Mike Johnson	Paroxysmal Neurological Disorders	Imperial College
David Bennett	Paroxysmal Neurological Disorders	University of Oxford
Geoff Woods	Paroxysmal Neurological Disorders	University of Cambridge
Henry Houlden	Paroxysmal Neurological Disorders	UCL

Huw Morris	Movement Disorders	Royal Free and UCL
Oliver Bandmann	Movement Disorders	University of Sheffield
Nick Wood	Movement Disorders	UCL
Mike Hanna	Neuromuscular Disorders	UCL
Kate Bushby	Neuromuscular Disorders	University of Newcastle
Andrea Nemeth	Ataxia, chorea and other hyperkinetic movement disorders	University of Oxford
Patrick Chinnery	Ataxia, chorea and other hyperkinetic movement disorders	University of Cambridge
Henry Houlden	Ataxia, chorea and other hyperkinetic movement disorders	UCL
Tom Warner	Genome sequencing in very rare inherited neurological disorders: Identifying overlapping common pathological pathways	UCL
Evan Reid	Genome sequencing in very rare inherited neurological disorders: Identifying overlapping common pathological pathways	University of Cambridge
Henry Houlden	Genome sequencing in very rare inherited neurological disorders: Identifying overlapping common pathological pathways	UCL
Volker Straub	Understanding disease and deep phenotyping	University of Newcastle
Mina Ryten	Transforming the diagnosis and understanding of neurological disorders through the creation of resources for variant prioritisation	King's College London
Alan Pittman	Transforming the diagnosis and understanding of neurological disorders through the creation of resources for variant prioritisation	UCL
Chris Ponting	Transforming the diagnosis and understanding of neurological disorders through the creation of resources for variant prioritisation	University of Oxford
Huw Morris	Training subdomain – Neurological disorders	UCL
James Polke	Developing a streamlined system to interpret and report genome sequencing results in the diagnostic laboratory	UCLH

The sample collection and genome sequencing is an essential part of the project but equally important is the analysis and translation of these findings, as described in the research plans here. This will require substantial financial support both in the early stages to co-ordinate clinicians and research groups and also in the long-term to fully investigate and sustain this initiative.

Data requirements

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

We will require:

- BAM files
- VCF files
- Phenotypic data (HPO terminology including modifiers)
- Uploaded documents detailing previous genetic testing
- Uploaded documents detailing imaging, electrophysiology, histopathology and other data.

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

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

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

We will require the data captured by the disease-specific data models (both HPO and non-HPO terms), but given the complexity of phenotyping in neurology, we would also wish to access uploaded original reports when possible. These would include reports on prior genetic testing, imaging, electrophysiology and histopathology. We intend to use these documents to improve the quality of phenotyping data. This will be achieved through the use of text-mining and semantic technologies for annotation and capture of information. This additional phenotypic data will form

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

We have reviewed the existing NGS analysis pipeline for Genomics England and in addition to this; we would also employ an additional complementary variant calling pipeline, as per GATK good practice (https://www.broadinstitute.org/gatk/guide/best-practices.ph). These steps include the following:

- Base Quality Score Recalibration
- Local realignment around INDELS
- HaplotypeCaller (cohort analysis workflow)

- Variant Quality Score Recalibration
- Genotype-refinement (incorporating pedigree information and population structure*)*

This callset refinement ends with a VCF file. Processing involves using meta-data to assess and improve genotyping accuracy, attach additional information, evaluate the overall quality of the callset and to help weed out false positive genotype calls. Extra data resources required for this are detailed in the data import section.

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

The vast majority of the tools that we require are available as per the List_of_Embassy_apps.xlsx*,* However, we would also require the following applications:

- Variant Effect Predictor (http://www.ensembl.org/info/docs/tools/vep/index.html) – an annotation tool complementary to annovar to facilitate the prioritization of variant analysis (eg. loss of rare functional variants)
- Beagle (https://faculty.washington.edu/browning/beagle/b3.html) – for phasing genotype data and for imputation of missing genotype data.
- Varscan2 (http://varscan.sourceforge.net) – complementary variant calling method that in addition can pick up mosaic variants.

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

We would require the following resources/reference data sets to be available the in data environment:

- resource:hapmap - hapmap_3.3.hg19.sites.vcf
- resource:omni - 1000G_omni2.5.b37.sites.vcf
- resource:1000 - n1000G_phase1.snps.high_confidence.vcf
- resource:dbsnp - dbsnp_135.b37.vcf
- resource:mills -snMills_and_1000G_gold_standard.indels.hg19.sites.vcf
- resource:dbsnp - dbsnp_138.hg19.vcf
- FASTA reference genome used in the original alignment.

These are available from here: https://www.broadinstitute.org/gatk/download/

We would also require the hg19 Variant Effect Predictor data cache (Ensemble v77) available from http://www.ensembl.org/

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

GATK variant calling pipeline would place a high demand on computation resources, and the intermediate files would take up a significant amount of storage space:

- Each genome would require approximately 200Gb of temporary storage space.
- Each genome would require approximately 2,400 CPU hours to generate the high-quality variant call set that we require; optimised to take advantage of the parallel computing environment with multiple-data threading.

We do not deem that the downstream analysis of the variant data in R studio will place excessive high demand on computing resources.

Omics samples

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

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

Detailed research plan

Full proposal (total max 1500 words per subdomain)

Title *(max 150 characters)*	**Dementia and Motor Neurone 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).*

One in six people aged 80 and over have dementia, while early onset dementias (<65 years old) affect around 40,000 people in the UK. The health and social care costs exceed £18billion per annum (Fineberg et al, 2013) and yet there are currently no proven effective disease-modifying treatments for dementia and few symptomatic treatment options.

Alzheimer's disease (AD) and frontotemporal dementia (FTD) are the commonest causes of dementia and thought to explain 50–70% of all cases. After age, a family history of AD/FTD is the most important risk factor and studies of dizygotic and monozygotic twin pairs have revealed heritability estimates of between 60% and 80%. In AD and FTD the proportion of genetically undefined small families and patients with early onset disease outweighs those with a known disease gene.

Our greatest understanding of AD, FTD and other causes of dementia has come from the identification of inherited disease genes. In AD, three autosomal dominant genes; *APP*, *PSEN1*, and *PSEN2* account for around 65% of early onset disease. Important risk factors for late onset AD also exist, such as ApoE4. In FTD 40% of patients have a family history and the important genes identified include MAPT, GRN and C9ORF72. Overall, mutations in GRN, MAPT and C9ORF72 together account for 17% of FTD cases (mutations in VCP and CHMP2B are rare, each explaining less than 1% of the familial FTD).

Motor neurone disease (MND, also called amyotrophic lateral sclerosis, ALS) is a relentlessly progressive neurodegenerative disorder with just three years average survival. The lifetime risk is 1 in 300 and it is the commonest reason people seek euthanasia (Johnson, 2006). Around 10% of MND is familial and a further 10% are early onset and likely genetically predisposed. Two major genes in C9ORF72 and SOD1 account for around 45% of the familial cases. C9ORF72 expansions are also found in 6% of sporadic MND cases, reflecting the later onset age dependant low penetrance of this mutation (Smith et al 2013). Overlap in the pathology and genetics of MND and FTD is so great that they are now considered phenotypic variants of a common disease spectrum.

The previous contribution to research of sequencing and GWAS studies

Substantial progress has been made towards understanding the genetic architecture of AD, FTD and MND but a significant proportion of the single gene genetic variance remains unidentified, as above. Genome-wide association studies (GWASs) have been successful in AD but rely on the use of many tens of thousands of samples. In addition to ApoE4, the replicable associations for several new genetic risk factors account for only a small proportion of genetic risk.

One powerful approach to identify AD risk variants with large effect sizes has been to study families with a strong history of AD using exome and genome sequencing. Although expensive it has revealed important results, even in the small numbers examined. In addition protective APP variants have been identified as well as new AD risk genes such as TREM2 both with a '2 to 4' fold increase in risk of developing AD. Large-scale genome sequencing will yield a wave of important

disease genes and greater insight into mechanistic pathways.

Important issues in dementia and MND

The young onset, poor prognosis, very high health and social care costs, and absence of effective therapies define dementia and MND as priority areas for UK and international research. Through GEL we are collecting multiple familial cases AD, FTD and other rare forms of dementia. The depth of phenotyping is essential and will exploit parallel studies at single sites (BioMOX, PiPPIN), disease specific collaborative programs (GENFI, PROSPECT) and national resources (Dementia Platform UK - DPUK, BioBANK).

Why further genome sequencing and investigation is required

The identification of defective genes and mapping of disease pathways will allow us to stratify patient cohorts and discover new drug targets, leveraging the benefit of genome sequencing in neurodegenerative disorders. Based on smaller scale studies we are confident of the identification of (i) new genes as causes of dementia and importantly new disease causing pathways, (ii) new genetic defects in known genes and their promotors, as well as new types of genetic defects such as structural defects and *de-novo* mutations, and (iii) new guidelines on genotype-phenotype relationships that facilitate clinical genetic diagnostics and clinical trials.

These discoveries will support functional investigations of the new genes and genetic pathways, their biochemistry and cell biology. It will create new and clinically-validated disease models, and underpin the work of pre-symptomatic trials to prevent dementia (eg. PRENI and related consortia). The data benefit clinical academic research, but also engage pharma partnerships and stimulate patient facing NHS services leading to better care and to genetically defined patients available to invite to clinical trials.

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

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

Specific aims:

A) To interpret mutations we will use an approach that relies upon the investigation of variant properties and segregation with disease, in discovery and replication case-sets, spanning classical phenotypes and phenotypic diversity.

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

C) With new genes we will investigate the genome data in other GeCIP domains (such as paediatric and neurodevelopmental disorders) and collaborate internationally to identify further patients.

D) Functional effects. The level of investigation will depend on the likely pathogenicity of the variant and the putative gene function.

E) Phenotyping of genetically defined series. Core phenotyping will build a database record of the major functional neurological, motor, cognitive and psychiatric symptom domains. Deep phenotyping of fluid biomarkers, brain imaging, cognition and physiology will exploit current data sharing mechanisms (e.g. DPUK) and attract new funding where required, drawing on proven models in the TRC-D and DPUK network and internationally.

F) Models of disease will be facilitated by long term acquisition and sharing of derived stem cells, including but not restricted to StemBancc and DPUK schemes.

Identifying therapies from pathways. An example using a known gene is the Progranulin mutations causing familial frontotemporal dementia. In FTD, up to 40% of probands report a strong relevant family history (of FTD or MND), only half of whom have a known single gene mutation/expansion. On the commonest three mutations is of Progranulin, leading to a loss of function. The International GenFI study, based in the UK is undertaking deep phenotyping and longitudinal study of ~200 cases and asymptomatic first-degree relatives. A phase II clinical trial is now underway, led by Forum Pharmaceuticals with support from GenFI sites, to test a histone deacetylase (HDAC) inhibitor FRM-0334, to reverse the loss of function.

Another example is the Dominantly Inherited Alzheimer Network Trials Unit (DIAN-TU). More generally, in neurodegeneration there is an increasing appreciation of a common disease mechanism associated with the misfolding, seeded propagation and toxicity of specific brain proteins (eg. alpha synuclein in Parkinson's disease, prion protein in prion disease etc.). Some experimental therapeutics in clinical or preclinical development simply target the expression from the gene encoding the disease-associated protein.

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

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

We plan to work closely with a number of cross-cutting GeCIPs to help with the aims of this research plan. We have also established international collaboration to replicate or repudiate identified variants with the major sequencing laboratories of Professors Andy Singleton and we collaborate with the largest global consortia in gene discovery in dementia; Professor Julie Williams, Cardiff University, and other international colleagues, in GERAD, IGAP and PERADES.

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

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

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

Sub-domain leads
Prof James Rowe, Prof Simon Mead and Prof Chris Shaw are research group leader and experts in

dementia and ALS. All three have extended experience in genetics, genome sequencing and the functional investigation of disease genes. As the genome sequencing proceeds they will work together to distribute the various tasks needed to examine and interpret the genetic data.

Domain leads

Prof Henry Houlden and Prof Patrick Chinnery, both have an interest in neurogenetics and to maximise the opportunity of the 100,000 genomes 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.*

This GeCIP builds a closer working relationship between neurology clinicians, clinical geneticists and research genetics teams. This provides an immediate benefit to patients and families in enhancing the quality of information and advice during and after genetics testing, and in the interpretation of results. It will enable more sophisticated modelling of disease risk, explain phenotypic variation within and between families.

In the longer term, the GeCIP will lead to novel and effective therapies via the clinically validated model systems in preclinical pathways, from cell biology and high-throughput screening programs in early preclinical models through to optimised selection of candidate therapies and target populations in clinical trials for efficacy.

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

Clinicians and patients - Families; diagnosis, predictive, prenatal testing. NHS organisations and cost savings.

Pharmaceutical companies - New disease pathways, targets, stratified patient groups, pre-clinical research models, clinical Trials.

Neurogenetic laboratories - New tests.

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?*

Potential new gene mutations and pathways may be patentable. Involving industry will be a helpful to fund and accelerate further work

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

Smith BN et al, The C9ORF72 expansion mutation is a common cause of ALS+/-FTD in Europe and has a single founder. Eur J Hum Genet. 2013 Jan;21(1):102-8.

Johnston C. A. et al, ALS in an urban setting, J Neurology 2006.

Fineberg NA et al. The size, burden and cost of disorders of the brain in the UK. J Psychopharmacol 2013 Sep;27(9):761-70.

Title *(max 150 characters)*	**Paroxysmal Neurological Disorders**

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

Summary

Paroxysmal neurological conditions include epilepsy, episodic movement disorders, headache and episodic pain. These are some of the commonest neurological conditions seen but in general they are poorly understood and underdiagnosed. They affect people across their lifespan meaning that the socioeconomic impact of these diseases is vast. There is also sudden unexpected death in epilepsy (SUDEP) where there is growing evidence of a genetic basis for SUDEP, with contributions that may be both monogenic (Bagnall et al, 2015) and distributed across the genome (Leu et al, 2015) – WGS would be an ideal way to determine individual risk of SUDEP and genetic cause in one test. The treatment of these conditions is limited and many patients are resistant to multiple forms of medication and some conditions such as cluster headache are extremely difficult to treat.

The previous contribution to research of next generation sequencing and GWAS studies

Familial and extreme phenotypes of paroxysmal disorders are rare, but in the past the identification of disease genes have revealed important information on the pathophysiology of these conditions and therapeutic pathways. There has been considerable effort in the past directed at understanding the basis of these conditions, for example, international consortia for epilepsy already exist (Lancet Neurol, 2014), but these have so far not used sequencing technologies, which are the methods that have led to the major advances. Moreover, paroxysmal disorders, such as epilepsy, are multifaceted conditions and seizures are a manifestation of many different insults to the brain. Therefore only projects with the reach of the 100,000 Genomes Project will prove sufficiently powered in identifying genes that may, for example, underlie the unsolved familial focal, familial generalised epilepsy or the epileptic encephalopathies. In addition the scope of the 100,000 genomes project will help to overcome the difficulties in capturing large numbers of trios and families and the careful genetic analysis needed with variable disease penetrance. Moreover, while there are frequently novel genes identified as potential causal to epileptic encephalopathy, the 100,000 Genomes Project (with large numbers of trios captured and deep phenotyping of those families) will allow us to more accurately characterise which of these mutations are common enough to warrant targeted functional analysis with a view to novel molecular therapeutics.

Important issues in paroxysmal neurological disorders

A distinct advantage in the analysis of genome sequencing data from trios and families with paroxysmal disorders is that the disease gene is highly likely to have an important function in the synaptome, such as coding for a channel subunit or having a role in the synaptic vesicle delivery/recycling machinery. We expect to identify a number of new disease genes, genetic defects in known genes and in a series of deeply phenotyped genetically proven patients. This will support functional work to investigate new genes and the development of disease models, identify cohorts of genetically defined patients to support clinical trials and the development of new genetic tests.

In addition we will train the next generation of clinicians and scientists in the interpretation and application of genomic and, in turn, WGS will provide causal and pharmacogenomic variants in one test. WGS will also allow detection of PGX variants that are emerging from large-scale programmes such as EpiPGX, www.epipgx.eu.

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

The research plan is based around the opportunities the genome sequencing will deliver to paroxysmal disorders with the availability of omic samples and phenotype data from this unique UK neurological cohort and the therapeutic potential that this work will bring. We would also like to highlight the collaboration with the paediatric GeCIP to share genetic results and the functional work as many of these disorders present in childhood.

Specific aims:
A) Developing a systematic phenotyping system that will be accepted by clinicians to capture deep clinical phenotypes and the results of investigations in patients with paroxysmal disorders. There will be significant overlap with the methods used to capture deep phenotypes in other conditions but there will also be additions needed for paroxysmal disorders, such as with severe headache phenotypes like trigeminal autonomic cephalalgias, and subtle variations in epileptic encephalopathy phenotypes.

B) To interpret the potentially pathogenic variants identified, we will use an approach that relies upon the investigation of variant properties and segregation with disease as well as our group's understanding of the pathophysiology of this group of diseases with ties to the strategic initiatives such as synaptopathies as discussed below.

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

D) With new genes, we will investigate the genome data in other GeCIPs (such as paediatric) and collaborate internationally to identify further patients, through well-established existing networks.

E) Functional effects. The level of investigation will depend on the likely pathogenicity of the variant and the putative gene function. In the case of potentially pathogenic mutations in channel genes or the transporting machinery as in the synapse Soluble (N-ethylmaleimide sensitive fusion proteins) Attachment protein Receptor (SNARE) we already have established collaborations with the UCL Kullmann laboratory and with the James Rothman (Yale/Cambridge/UCL) (synaptopathies collaboration). In the case of other channels such as in pain GeCIP sub-domain leads (Bennett/Woods) are experienced in working on this group of disorders and the channels involved.

F) Models of disease – In the short term we will create over-expression or loss of function mammalian cell lines with disease causing mutations. We also collaborate with those working on other models of disease, such as with Dictyostelium, Drosophila and zebrafish loss of function models. In addition the Bennet and Woods laboratory have established expertise in using iPSC models of painful channelopathies and mouse models.

G) Although the identification of therapies to modify disease causing pathways is likely to be beyond the scope of this research plan, there are a number of important examples of where the identification of a paroxysmal disease gene has led to therapeutic discovery and/or improvement in the management of patients. For example, in epilepsy there have been three main genetic discoveries that have influenced clinical care. Firstly, GLUT1 (glucose transporter 1) deficiency, which previously was diagnosed by the invasive procedure of a lumbar puncture, is now known

to be caused by mutations in *SLC2A1* gene. GLUT1 deficiency is associated with pharmacoresistant epilepsy as well as multiple other neurological conditions. However, patients may respond well to a ketogenic diet and early genetic diagnosis can therefore help with seizures and potentially forestall other complications such as developmental regression. Secondly, patients with Dravet syndrome due to *SCN1A* mutations do significantly better using non-sodium channel blocker antiepileptic drugs. Thirdly, the HLA-B*15:02 allele, which is more commonly found in the Han Chinese, can predispose to severe cutaneous drug reactions with carbamazepine, justifying potential screening for this allele in susceptible populations; the same is true for the HLA-A*31:01 variant in Europeans (Chung et al, 2014). In our research plan we plan to collaborate early with industry once we have identified disease genes, genetic risk and modifying factors. Given the time and expense of bringing new drugs to patients, this is the most efficient strategy. Genetic testing is also impacting on pain medicine. Finding mutations in the *SCN9a* gene encoding the voltage gated sodium channel $Na_v1.7$ in patients with erythromelalgia, paroxysmal extreme pain disorder or small fibre neuropathy has implications for treatment as specific mutations respond to carbamazepine or mexiletine. Specific blockers of $Na_v1.7$ are under development with clinical trials underway.

H) Training. We plan to train clinical fellows and non-clinical students of various degrees in genomic medicine, particularly focussed on paroxysmal disorders and the disease mechanisms.

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

We plan to work closely with the Paediatric GeCIP and a number of cross-cutting GeCIPs to help with the aims of this research plan, including; Education and Training, Electronic Records, Enabling Rare Disease Translational Genomics via Advanced Analytics and International Interoperability, Functional Cross Cutting and effects, Machine Learning, Quantitative Methods and Functional Genomics, Validation and Feedback.

We have also established international collaboration to replicate or repudiate identified variants and collaborate widely on a number of European initiatives such as the Epi4K initiative which will be important for variant interpretation.

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

We will work closely with the cross-cutting GeCIP training domain and the Neuro-GeCIP training research plan. It is crucial that trainees have specific training in paroxysmal disorders including clinical, neurophysiological, pathological and imaging (MRI studies). This will facilitate identification and stratification of patients and families but also validation of identified candidate variants through careful phenotype segregation studies. In addition, the GeCIP approach to research would be used to develop a common doctoral training program across sites.

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

Prof David Bennett (University of Oxford) and Prof Geoff Woods (University of Cambridge) experts in pain and channel genetics.
Dr Arjune Sen (University of Oxford), Prof Sanjay Sisodiya (UCL) and Prof Mike Johnson (Imperial College) are experts in epilepsy and genetics.

Dr Guy Leschziner (Guy's) is an expert in sleep disorders.
Prof Henry Houlden (UCL) is an expert in channels and paroxysmal movement disorders.

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

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

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

Clinicians and patients - Families; diagnosis, predictive, prenatal testing, avoidance of unnecessary test and an end to the diagnostic odyssey. NHS: cost savings.
Pharmaceutical companies - New disease pathways, targets, stratified patient groups, pre-clinical research models, clinical Trials.
Neurogenetic laboratories - New tests.

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?*

Potential new gene mutations and pathways may be patentable. Involving companies may be helpful to fund further work

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

Chung WH, Hung SI, Hong HS et al., Medical genetics: a marker for Stevens-Johnson Syndrome. Nature. 2004; 428:486 (this is the first paper to highlight link between HLA B*1502 and SJS).

Bagnall RD et al. Exome-based analysis of cardiac arrhythmia, respiratory control and epilepsy genes in sudden unexpected death in epilepsy. Ann Neurol. 2015 Dec 24. doi: 10.1002/ana.24596.

Leu C et al. Genome-wide Polygenic Burden of Rare Deleterious Variants in Sudden Unexpected Death in Epilepsy. EBioMedicine. 2015 Jul 10;2(9):1063-70.

International League Against Epilepsy Consortium on Complex Epilepsies. Genetic determinants of common epilepsies: a meta-analysis of genome-wide association studies. Lancet Neurol. 2014 Sep;13(9):893-903.

Johnson MR et al. Systems genetics identifies a convergent gene network for cognitive function and neurodevelopmental disease. Nature Neuroscience, in press. doi:10.1038/nn.4205

Detailed research plan

Full proposal (total max 1500 words per subdomain)	
Title *(max 150 characters)*	**Ataxia, chorea and other hyperkinetic movement disorders**

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

Ataxia, chorea and other hyperkinetic movement disorders are frequently seen in the neurology clinic but fall within the 'rare-disease' category with an overall frequency of around 5-10/100,000. Ataxia is genetically defined in around 60% of patients and Huntington's disease (HD) is diagnosed in around 80% of the choreiform conditions but nearly all are currently untreatable. Teams within GeCIP have contributed to the increase in molecular diagnosis in this group but many still do not have a genetic diagnosis. There are excellent UK and European collaborations for gene identification, and validation of potential pathogenic variants as well as novel assays for investigating the pathogenicity, understanding disease mechanisms and carrying out novel drug screening.

The previous contribution to research of next generation sequencing studies
Whole exome studies have had a major impact on the molecular diagnosis of movement disorders. This has led to the identification of a large number of genetic causes and mechanisms potentially amenable to treatment. However, this has also lead to the discovery that these diseases are extremely heterogeneous, for example in cerebellar ataxia there are over 60 genetic causes known (see references from Nemeth/Houlden/Chinnery). There are still a number of disease areas that have been poorly defined genetically, particularly in the congenital and late onset ataxic conditions and those with incomplete penetrance and/or phenotypic variability such as the late-onset chorea's or HD phenocopy disorders.

Important issues in ataxia, chorea and other hyperkinetic movement disorders
Ataxia, chorea and other hyperkinetic movement disorders may be caused by expanded repeats, most commonly CAG or polyglutamine expansions. The identification of unstable repeats has revealed a whole new mechanism of disease and opened up the biology of intronic repetitive elements. The use of genome sequencing, with larger unbiased reads will allow further repeat elements, many new disease genes and novel phenotypes to be identified, as evidenced by published work and also in the DDD program. This will support functional work to investigate new genes and the development of disease models, series of genetically defined patients to support clinical trials and the development of new genetic tests. In addition we will train the next generation of clinicians and scientists in the interpretation and application of genomic medicine. We already have extensive expertise in this field e.g. Nemeth research lab has translated a number of genes to the NHS diagnostic laboratory and these data have informed Genomics England for gene lists to use in filtering. The great advantage of the Neurology Genomics England Clinical Interpretation Partnership (Neuro-GeCIP) is the provision expertise and sample collection framework to facilitate the inclusion of the majority of undiagnosed cases and families in the genome sequencing.

The ataxia and choreiform disorders, where the primary genetic defect is known, show remarkable phenotypic variability and disease severity. This can be seen in the ataxia families causes by expanded spinocerebellar ataxia (SCA) repeats and in HD where genetic modifiers in the mismatch repair genes have been shown to account for a significant proportion of the age of

onset variability. These modifier genes are likely to be present in other genetic causes and will be unraveled by genome sequencing of larger cohorts of patients with an identified primary defect in combination with clinical data on disease severity (extreme phenotypes) and other omics data.

There are a number of reasons why whole genome sequencing should be applied to ataxia, chorea and other hyperkinetic movement disorders diseases, these include:

1. Exome sequencing has failed to identify a molecular basis of these diseases in a significant proportion of patients and families with a likely genetic cause.

2. The undefined families are mostly small or individual cases that will be hard to genetically dissect and require the accuracy of genome sequencing and use of all aspects of the GeCIP. This will allow the identification of single base changes and also structural variants such as deletions and *de novo* mutations that have previously been difficult to identify and reflect the lack of appropriate technology directed studies.

3. It is likely that the analysis of whole genome sequence will provide the first 'single-shot' comprehensive test for these disorders and the neuro-GeCIP will produce a framework for this future diagnostic process to be as streamlined and effective as possible.

4. The genome sequencing will form the platform to advance the understanding of disease pathogenesis and identify tractable disease pathways. New animal and human inducible pluripotent stem cell (iPSC) models combined with emergent technologies will enable deep phenotyping, and high throughput effective drug screening. This will allow us to take basic science discoveries to preclinical validation, capitalising on unrivalled experimental expertise available through the neuro-GeCIP.

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

Objectives:
1. Determine whether whole genome sequencing provide a near-comprehensive form of analysis to identify new disease genes, pathways and form the gold standard of diagnostic testing.
2. Use this genome sequencing as the core framework to drive integrated diagnostic, biomarker and therapeutic innovation.

Specific aims:
A) Working with our creation of resources for variant prioritisation research plan we will develop data capture and similarity scoring algorithms to help delineate the rare and severe ataxia/hyperkinetic disorders into homogeneous groups.

B) To interpret mutations we will use a number of approaches that include; (i) investigate variant properties (e.g. control frequency data, location and function of variants), (ii) segregation with the disease, (iii) functional prediction and investigation of putative protein changes, (iv) gene CNS expression and (v) investigation of possible disease associated repeat expansions.

C) Investigation of mutation mechanisms in instances such as *de-novo* variants (a proven cause in ataxias), potential splice site, structural variation and copy number variation.

D) With new genes we will investigate the genome data in other GeCIPs (such as paediatric) and

collaborate nationally and internationally to identify further patients and families. The comprehensive network that is represented by our GeCIP will provide cohesive collaborations for identifying additional cases. Some data is already available via research teams (Nemeth/Houlden/Chinnery/) and DDD and this will also inform further analysis of variants.

E) Functional effects. The level of investigation will depend on the likely pathogenicity of the variant and the putative gene function. Our GeCIP has already collated potential areas of expertise for investigating function and in highly likely pathogenic variants we will collaborate with other GeCIPs and groups to investigate new mechanisms and pathways.

F) Defining genetically defined series, deep phenotyping and genotype-phenotype correlations to define new disease associations and power clinical trials.

G) Models of disease – cellular, structural, invertebrate, animal and long-term iPSC derived neuronal models will be important for the future understanding of disease pathways but will require collaboration with other GeCIPs. We also have permission to deposit genetically defined ataxia fibroblasts to the Wellcome Trust funded Human Induced Pluripotent Stem Cell Initiative (HipSci) for reprogramming into iPSC lines.

H). Identifying therapies from pathway discovery and early industry collaboration.

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

We plan to work closely with a number of cross-cutting GeCIPs to help with the aims of this research plan, including;
Education and Training (Maxine Foster), Electronic Records (Harry Hemingway), Enabling Rare Disease Translational Genomics via Advanced Analytics and International Interoperability (Kate Bushby and Michael Simpson), Functional Cross Cutting (Colin Cooper and Gkikas Magiorkinis), Functional Effects (Ewan Birney), Machine Learning, Quantitative Methods and Functional Genomics (Martin Tobin), Validation and Feedback (Bill Newman).

We have also established international collaboration to replicate or repudiate identified variants with the major sequencing laboratories of Professors Andy Singleton, Alexis Brice, Stephan Zuchner and Peter Heutink (see letters attached).

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

We will work closely with the cross-cutting GeCIP training domain and the Neuro-GeCIP training research plan. It is crucial that trainees have specific training in deep ataxia and other hyperkinetic disorders phenotyping including clinical, neurophysiological, pathological and imaging (MRI studies). This will facilitate identification and stratification of patients and families but also validation of identified candidate variants through careful phenotype segregation studies.

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

Subdomain lead, Prof Andrea Nemeth,

Domain leads Prof Henry Houlden and Prof Patrick Chinnery.
All three individuals are experts in genetics and ataxia/chorea and have worked and collaborated closely together in the past.

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 over-arching aim of our GeCIP is to provide a comprehensive research system for new pathways and molecular diagnosis for patients with ataxia, chorea and hyperkinetic disorders. Results will be fed back where possible after validation and diagnostic confirmation through genetic and neurogenetics clinics to patients and families.

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

Clinicians and patients - Families; diagnosis, predictive, prenatal testing. NHS organisations and cost savings.
Practising clinicians - to help with diagnosis and interpretation.
Pharmaceutical companies - New disease pathways, targets, stratified patient groups, pre-clinical research models, clinical Trials.
Neurogenetics/Genetic laboratories – The development of new tests.

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?*

Potential new gene mutations and pathways may be patentable. Involving companies may be helpful to fund further work

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

Németh AH et al. Next generation sequencing for molecular diagnosis of neurological disorders using ataxias as a model. Brain. 2013 Oct;136(Pt 10):3106-18.

Parolin Schnekenberg R et al and Németh AH. De novo point mutations in patients diagnosed with ataxic cerebral palsy. Brain. 2015 Jul;138(Pt 7):1817-32.

Bettencourt C et al and Houlden H; United Kingdom Brain Expression Consortium. Insights from cerebellar transcriptomic analysis into the pathogenesis of ataxia. JAMA Neurol. 2014 Jul 1;71(7):831-9.

Detailed research plan

Full proposal (total max 1500 words per subdomain)

Title *(max 150 characters)*	**Neuromuscular Diseases**

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).*

Neuromuscular diseases (NMD) are untreatable muscle-wasting conditions that affect >100,000 people in the UK. They are genetic or acquired, span neonatal life to late age, and cause premature death or lifelong disability. For example, Duchenne dystrophy patients die in young adult life whereas most patients with inherited neuropathies have a normal lifespan but with significant morbidity. Disruption of neuromuscular systems biology underpins NMDs. The molecular diagnostic rate varies between these disorders, in proximal myopathy around 80% of the genes have been identified but in disorders such as Charcot Marie Tooth (CMT) disease type 2 only around 50% of patients have a genetic diagnosis.

The previous contribution to research of next generation sequencing studies

Whole exome studies have had a major impact on the molecular diagnosis of neuromuscular diseases. This has led to the identification of a large number of genetic causes and mechanisms potentially amenable to treatment (see references). In particular, muscle channelopathies and congenital myasthenic syndromes can be treated effectively with standard pharmacological agents once they are accurately diagnosed. However, this has also lead to the discovery that these diseases are extremely heterogeneous, for example in CMT where over 80 genetic causes are known. There are still a number of disease areas that have been poorly defined genetically, particularly in the adult-onset conditions and those with incomplete penetrance and/or phenotypic variability such as the myofibrillar and mitochondrial myopathies, CMT2 and even limb girdle muscular dystrophy where a number of genes have been identified but many patients are without a genetic diagnosis. Whole exome studies have also shown that genes that were thought to underlie a particular neurological syndrome e.g. REEP1 and Atlastin 1 for Hereditary Spastic Paraplegia can also be the cause of a pure hereditary neuropathy increasing the spectrum of potential candidate genes. Clinical phenotypes are redefined with overlapping symptomatology such as neuromuscular transmission defects combined with axonal neuropathies (SYT2) or muscular dystrophies (GMPPB) which are explained by converging molecular pathways.

The great advantage of the Neurology Genomics England Clinical Interpretation Partnership (Neuro-GeCIP) is the provision of expertise and a sample collection framework to facilitate the inclusion of the majority of undiagnosed cases and families in genome sequencing. A number of clinicians have specific clinics that cover these disorders and interact with patient societies to facilitate the enrolment of large numbers of well-characterised patients. This requires sharing of genomic datasets between groups and with the wider neuromuscular community in a standardized way and through common infrastructure. This careful phenotyping is not only important in selecting families but also in the validation of genetic variants identified.

Neuromuscular diseases have the benefit that the diseased tissue (muscle and nerve) can be directly studied in humans through electrophysiological and morphological methods, and rich resources of DNA, serum, tissue and cell samples exist through the MRC centre biobanks in London and Newcastle.

Important issues in neuromuscular diseases

The development of therapies in inherited neuromuscular diseases starts with and is constantly informed by the identification of the underlying causative genetic defect. An example of where this has already influenced therapy is in Familial Amyloid Polyneuropthy (FAP) due to mutations in the transthyretin gene (TTR). Studies mainly from the UK have shown that certain genotypes including the comment genotype seen in the UK, the Ala 60 mutation, do very badly with liver transplant as the cardiomyopathy deteriorates. Newer therapies recently introduced to stabilise TTR, including Tafamadis and Diflunasil, and therapies in clinical trials e.g Antissense oligonucleotide therapies are using this knowledge to explore mutation specific response to therapy. This is likely to apply to a range of other inherited neuromuscular diseases. Moreover, a mature clinical trial infrastructure for neuromuscular disorders exists across the partners having access to suitable patient cohort, natural history data and tested outcome measures, so that new targets and compounds – either via academic or through commercial partners – can be channeled seamlessly into clinical developments.

Finally there are a number of European and Worldwide networks that have been established to collaborate on the clinical, biomarker and therapeutic aspects of neuromuscular diseases. These will be important in proving disease genes in comparing genomic data, deep phenotyping and also in the joint development of therapies. These networks include ERN, ELIXIR, BIO-NMD, Optimistic, NeurOmics, TREAT-NMD and RD-Connect, with the latter 2 networks coordinated by Newcastle. To allow efficient linking and integration of the omics data generated, specific support (personnel and hardware) are required.

Even in a population as large as the UK's, confirmatory (second) families with a specific variant or gene in a rare neuromuscular disease may be absent, and systematic and computerized match-making capabilities need to be set up in conjunction with these wider international projects that focus on rare neuromuscular diseases.

Why genome sequencing, interpretation and functional investigation is required in neuromuscular diseases:

1. Exome sequencing has failed to identify a molecular basis of these diseases in a significant proportion of cases.

2. The undefined families are mostly small or individual cases that will be hard to genetically dissect and require the accuracy of genome sequencing and use of all aspects of the GeCIP. This will allow the identification of single base changes and also structural variants such as deletions and *de- novo* mutations that have previously been difficult to identify and reflect the lack of appropriate technology directed studies.

3. It is likely that the analysis of whole genome sequence will provide the first 'single-shot' comprehensive test for these disorders and the neuro-GeCIP will produce a framework for this future diagnostic process to be as streamlined and effective as possible.

4. For rare diseases having ethnically matched control groups and other collaborative cohorts from Europe and the USA that will be essential in proving the variants identified.

The genome sequencing will form the platform from which to accelerate neuromuscular disease diagnostics and to deeply understand the pathogenesis and identify tractable disease targets and pathways. New animal and human inducible pluripotent stem cell (iPSC) models combined with emergent technologies will enable deep phenotyping, and high throughput effective drug screening. This will allow us to take basic science discoveries to preclinical validation, capitalising on unrivalled experimental expertise available through the neuro-GeCIP.

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.*

Neuromuscular diseases objectives:
1. determine whether whole genome sequencing provides a near-comprehensive form of analysis to identify new disease genes, pathways and form the gold standard of diagnostic testing for neuromuscular diseases. This includes validation of variants and genes through advanced bioinformatics approaches, match-making, and model systems.
2. Use this genome sequencing as the core framework to drive an integrated diagnostic, biomarker and therapeutic innovation utilizing resources and research infrastructures already available to the neuromuscular field.

Specific aims:
A. For patients with neuromuscular diseases the genome sequencing will determine:
1) Whether whole genomic sequencing will identify known and new single nucleotide variants (SNVs), small indels or whole exon deletions
2) To determine whether *de novo* mutations cause novel de-novo dominant or recessive disorders.
3) Mutations will be interpreted with various levels of validation that range from level 1 (highly pathogenic) to level 3 (possible pathogenic) depending on the mutation type, change and gene.

B. To determine whether somatic mosaicism explains single cases based on the potential availability of other tissues. This will be possible in a number of trios with Neuromuscular diseases as through the UCL/Newcastle biobank we have taken frozen muscle, nerve, fibroblast and lymphoblastoid cell lines on over 2000 patients.

C. Systematic deep phenotyping of cases and efficient capture methods. This will involve customising the data capture and algorhythms from other disorders and tailoring them towards specific features in the rare disorders such as spasticity in HSP and calcification on CT in Fahr's syndrome. Phenotyping methods will be taken after collaboration with cross-cutting GeCIPs such as Electronic Records and Enabling Rare Disease Translational Genomics via Advanced Analytics and International Interoperability and from the neuro-GeCIP phenotyping research plan.

D. Understanding of gene function.
(i) Expression: the gene is expressed in tissues relevant to the disease of interest and/or is altered in expression in patients who have the disease.
(ii) Predicted protein function: the gene product interacts or performs a similar biochemical function with functional pathways implicated (genetically or biochemically) in the disease of interest.
(iii) Experimental functional effects: the variant significantly alters levels, splicing, normal biochemical function or cellular phenotype of the product of the affected gene. This is shown either in affected patient cells, or a well-validated in-vitro, ideally neuronal model system where rescue of the cellular phenotype is achieved by the wild-type gene or variant gene knockdown. A range of different biochemical and cellular approaches are available to validate pathogenicity of diseases at the genetic level.
(iv) Further models: the variant is consistent with the disease phenotype in fruit fly, zebrafish or mouse models and the cellular phenotype of human iPSC lines/CRISPR/CAS9 system.

E. Identifying therapies from pathways, this will be important and require early industry

collaboration. An example of where this has been successful in NMD is the identification of mutations of the first 2 subunits (SPTLC1 and SPTLC2) of the enzyme serine palmitoyl transferase as the cause of a rare form of hereditary sensory neuropathy (HSN1). This led to the identification of a new disease mechanism i.e. neurotoxicity from deoxysphingolipids (DSBs) produced when the mutant proteins are present. This has allowed the developmental of a functional assay (DSBs plasma levels) to validate the pathogenicity of new SPTCL1/2 variants (currently available as a research test) and also to identify potential new therapies such as serine treatment to reduce DSB levels (currently in pre-clinical trials in the UK). Other examples include the successful introduction of beta-adrenergic drugs for the treatment of newly identified neuromuscular transmission disorders (such as DOK7 defects), personalized medicine approaches for congenital myasthenic syndromes and muscular dystrophies according to the underlying genetic defect, and advanced therapies such as antisense oligonucleotides and gene replacement therapies under clinical development for muscular dystrophies and spinal muscular atrophies and repurposing of mexiletine and dichlorphenamide to treat myotonia and periodic paralysis respectively.

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

We plan to work closely with a number of cross-cutting GeCIPs to help with the aims of this research plan, including; Education and Training, Electronic Records, Enabling Rare Disease Translational Genomics via Advanced Analytics and International Interoperability, Functional Cross Cutting and effects, Machine Learning, Quantitative Methods and Functional Genomics, Validation and Feedback.

We have also established international collaboration to replicate or repudiate identified variants with Stephan Zuchner, Kym Boycott/Mike Brudno in Ottawa/Toronto, MYO-SEQ and the Broad Institute with Daniel MacArthur, NeurOmics group and RD-Connect where we are members.

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

We will work closely with the cross-cutting GeCIP training domain and the Neuro-GeCIP training research plan. It is crucial that trainees have specific training in deep neuromuscular phenotyping including clinical, neurophysiological, pathological and imaging (MRI studies). This will facilitate identification and stratification of patients and families but also validation of identified candidate variants through careful phenotype segregation studies. Training is also required for systematic collection and processing of samples for biobanking and biomarker studies, as well as the use of bioinformatics tools for variant annotation.

Training will also need to include whatever phenotype nomenclature is selected for the neuromuscular component of Neuro-GeCIP and international consortia e.g. HPO.

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

Professor Mike Hanna (Director of MRC Centre for Neuromuscular Diseases, UCL Institute of Neurology) and Prof Kate Bushby (Chair of Neuromuscular Genetics, Newcastle University).

Prof Mary Reilly (Peripheral Neuropathy, Professor of Neurology, UCL) and Prof Rita Horvath (Peripheral Neuropathy, Professor of Neurology, Newcastle University).

Prof Hanns Lochmuller (Chair of Experimental Myology, Newcastle University), Prof Volker Straub (Chair of Neuromuscular Genetics, Newcastle University) and Prof Francesco Muntoni (Chair of Paediatric Neurology, UCL Institute of Child Health).

Prof Henry Houlden and Prof Patrick Chinnery (domain co-leads) have particular interest and expertise in neuromuscular disease genetics.

All listed PIs are members of the MRC Centre for Neuromuscular Diseases and collaborate on all aspects of translational research in neuromuscular disorders.

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 over-arching aim of our GeCIP is to provide a comprehensive research system for new pathways and molecular diagnosis for patients with NMD leading to the identification of targets and pathways for biomarker discovery and therapies.

Beneficiaries. *How will the research benefit patients and healthcare institutions including the NHS, other researchers in the field? Are there other likely beneficiaries?*
Three groups of people will benefit from this project:

Patients: The ability to offer more accurate diagnosis will have an immediate benefit for patients. In some cases the elucidation of the genetic defect will offer the opportunity of already available, pharmacological treatment. In other cases, the genetic diagnosis will allow for participation in clinical research including treatment trials.

Practising clinicians; Novel variants in known and candidate genes are a significant problem for diagnostic genetic laboratories. Many novel variants ultimately turn out to be benign polymorphisms and much time and energy is spent examining relatives to determine whether the mutation segregates. The development of robust functional assays to determine the pathogenicity of novel variants will be of considerable benefit to diagnostic laboratories and to practising clinicians both within the NHS but also worldwide and may reduce the requirement for expensive and time consuming segregation studies.

Scientists: The identification of new genes and new diseases mechanisms will be of interest to neuromuscular scientists and will ultimately help therapy development. Also the use of a human iPSC derived motor neuron models will be of benefit to other researchers in the field of inherited neuromuscular diseases. It will provide further evidence of the feasibility of this model to identify the pathomechanisms of disease and for pre-clinical drug trials pf candidate therapies.

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?*

Potential new gene mutations and pathways may be patentable
Involving companies may be helpful to fund further work

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

Rossor AM, Evans MR, Reilly MM. A practical approach to the genetic neuropathies. Pract Neurol. 2015 Jun;15(3):187-98.

Suetterlin K, Männikkö R, Hanna MG. Muscle channelopathies: recent advances in genetics, pathophysiology and therapy. Curr Opin Neurol. 2014 Oct;27(5):583-90.

RD-Connect: an integrated platform connecting databases, registries, biobanks and clinical bioinformatics for rare disease research. J Gen Intern Med. 2014 Aug;29 Suppl 3:S780-7.

EXOSC8 mutations alter mRNA metabolism and cause hypomyelination with spinal muscular atrophy and cerebellar hypoplasia. Boczonadi V et al. Nat Commun. 2014 Jul 3;5:4287.

Affinity proteomics within rare diseases: a BIO-NMD study for blood biomarkers of muscular dystrophies. Ayoglu B et al. EMBO Mol Med. 2014 Jun 11;6(7):918-36.

Hexosamine biosynthetic pathway mutations cause neuromuscular transmission defect. Senderek J et al. Am J Hum Genet. 2011 Feb 11;88(2):162-72.

Detailed research plan

Full proposal (total max 1500 words per subdomain)

Title *(max 150 characters)*	**Mitochondrial Disorders**

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

Importance

Mitochondrial disorders have emerged as one of the largest groups of inherited neurometabolic disease. Epidemiological studies indicate that they affect approximately 1:5,000 of the population.[1] A molecular diagnosis is only possible in approximately half of these families. A proportion have mutations of mitochondrial DNA (mtDNA) responsible for their disorder. The remainder have a presumed nuclear genetic defect which can be autosomal dominant, autosomal recessive or X-linked.[2]

The previous contribution to research of next generation sequencing studies

Whole exome studies have had a major impact on the molecular diagnosis of mitochondrial disorders (eg Ref [3]). This has led to the identification of a wide array of new genetic causes for defects of the mitochondrial respiratory chain, revealing new mechanisms potentially amenable to treatment. However, conventional exome approaches have only identified molecular basis in approximately 60% of families, leaving a large minority without a genetic diagnosis.

Important issues in mitochondrial disorders

Understanding the molecular basis of a mitochondrial disorder has direct implications for the family, providing reliable prenatal diagnosis and genetic counselling.

In addition, it is well established that specific mitochondrial disorders have treatments that may be of immediate benefit to patients, perhaps the best example being Co-enzyme Q10 deficiency, and the possibility of enzyme replacement therapy in rare autosomal recessive forms that cannot be confidently identified based on phenotype alone.

Substantial phenotypic overlap between the different genetic causes means that a clinical diagnosis rarely points to a specific causative gene. With over 1500 known mitochondrial disorders this presents a major challenge in the clinic, which can only be circumvented comprehensively using a genome-wide approach.

Most mitochondrial disorders are associated with specific biochemical defects which can be measured in the research laboratory, thus providing a functional read-out which can be correlated with an underlying genetic abnormality. This is an important step in establishing cause and effect for novel gene defects.

Finally, it should be borne in mind that many mtDNA disorders also overlap with nuclear-genetic mitochondrial diseases. Many of these are diagnosed using conventional diagnostic approaches through NHS nationally commissioned services, but not all reach a diagnosis through this route. The analysis of off-target exome sequence reads, or whole genome sequence, will also provide a molecular diagnosis in mtDNA disorders.

Why genome sequencing, interpretation and functional investigation is required

There are four main reasons why whole genome sequencing should be applied to mitochondrial

disorders.

1. Exome sequencing has failed to identify a molecular basis of these diseases in a significant minority of cases (approximately 40% of nuclear genetic mitochondrial disorders).
2. There have been very few instances of *de novo* dominant mitochondrial disorders. This probably reflects a lack of appropriately conducted studies. The analysis of trios through the GeL project is highly likely to identify *de novo* dominant mitochondrial disorders based on well-established previous examples.
3. It is likely that the analysis of whole genome sequence will provide the first 'single-shot' comprehensive test for these disorders.
4. For mitochondrial DNA diseases it should always be remembered that mtDNA heteroplasmy complicates the situation. mtDNA deletions and specific point mutations may not be detected in blood using conventional approaches, although this has not been investigated comprehensively using deep whole genome 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.*

Aim: The over-arching aim is to determine whether whole genome sequencing provides a near-comprehensive diagnostic test for human mitochondrial disorders.

Specific aims:

A. For patients with established biochemical defects in the mitochondrial respiratory chain:

1) To determine whether stable exonic coverage provided by whole genomic sequencing will identify known and new single nucleotide variants (SNVs), small indels and whole exon deletions
2) To determine whether *de novo* mutations cause novel novo dominance or recessive mitochondrial disorders.
3) To identify novel and known mtDNA mutations through the genome sequence analysis.

B. Detection of mitochondrial disorders based on phenotypic similarity to known mitochondrial diseases.

In patients with no available biochemical analysis, to determine whether known or novel mutations affecting mitochondrial genes explain a clinical phenotype in patients with a broader area of neurogenetic disorders resembling mitochondrial diseases.

C. To determine whether somatic mosaicism (affecting the nuclear genome or mtDNA) provide an explanation for undiagnosed mitochondrial disorders.

D. Functional Investigation

A range of different biochemical and cellular approaches are available to validate pathogenicity of novel mitochondrial diseases at the genetic level. Cellular readouts include: conventional histology and histochemical and immuno-cytochemical approaches in solid tissue. Respiratory chain biochemistry in cultured cells including oxygen consumption using the Seahorse XF, mitochondrial morphology assays using quantitative network analysis.

Novel mutations in novel disease genes will be interrogated using the CRISPR/CAS9 system.

MtDNA mutations can be analysed using cybrid cell approaches.

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 mitochondrial community within England is well networked through ongoing research collaborations and close involvement with the NHS national commission diagnostic services. These link in turn to the 23 regional genetics disgnostic laboratories in the United Kingdom.

There are close links with the major European centres, and north American centres involved in molecular diagnostics, including international efforts to establish sequence databases (MISEQ-DR).

The GeCIP partners have close links with GSK, Astra Zeneca and Alexion Pharma.

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

We will work closely with the cross-cutting GeCIP training domain and the Neuro-GeCIP training research plan. The research project provides the opportunity to train laboratory, clinical research fellows and allied health care professionals. We will present projects for the MSc in Genomics coordinate by Health Education England, and additional PhD projects hosted by partner Universities.

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

Patrick F Chinnery FMed Sci is Professor of Neurology at the University of Cambridge where his laboratory is hosted within the MRC Mitochondrial Biology Unit. He has worked in the field of human mitochondrial disorders for 20 years, defining novel gene defects, mechanisms and developing treatments for these diseases.

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 over-arching aim of our GeCIP is to provide a comprehensive molecular diagnosis for patients with mitochondrial disorders. This will feed directly into national commissioned NHS services for the diagnosis of mitochondrial diseases, across the NHS through linkage with the 23 NHS diagnostic laboratories.

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

Predominantly through the rapid and cost-efficient diagnosis of mitochondrial disorders, leading to optimal family and patient management.

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 anticipate the identification of normal gene defects will reveal inter-related mechanisms potentially amenable to therapeutic manipulation. Members of the GeCIP already develop projects linked to earlier projects aimed at novel gene identification.

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

1. Gorman, G.S. *et al.* Prevalence of nuclear and mitochondrial DNA mutations related to adult mitochondrial disease. *Ann Neurol* **77**, 753-9 (2015).
2. Stewart, J.B. & Chinnery, P.F. The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. *Nature Reviews Genetics* (2015).
3. Taylor, R.W. *et al.* Use of whole-exome sequencing to determine the genetic basis of multiple mitochondrial respiratory chain complex deficiencies. *JAMA* **312**, 68-77 (2014).

Detailed research plan

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

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

Movement Disorders encompass a wide range of disorders from the common neurodegenerative disorder Parkinson's disease (prevalence 140/100,000) to much rare childhood onset dystonias and neurometabolic conditions (prevalence <10/100,000). Within the Movement Disorders sub-domain, we will focus on: i.) familial and early onset Parkinson's disease (PD); ii.) early-onset and familial dystonia and iii.) complex parkinsonism (including pallido-pyramidal syndromes).

PD: PD is heterogeneous spanning both "sporadic" late onset disease and early onset Mendelian versions of PD. From the 1990s forward autosomal recessive genes for early onset PD (parkin, PINK1 and DJ1) have been identified which have a common effect in regulating mitochondrial function [1].

Dystonia: An increasing number of genes for early onset dystonia/hyperkinetic movement disorders have been identified in individuals without a family history of disease, which includes both gene variants of reduced penetrance and *de novo* mutations.

Complex parkinsonism: Complex parkinsonism includes a series of "overlap" syndromes in which parkinsonism occurs with dementia, an eye movement disorder and/or a neuro-metabolic condition in some cases including neurodegeneration with brain iron deposition (NBIA). Most genes responsible for earlyonset complex parkinsonism are autosomal recessive.
Importantly, in each of these disease areas there are a substantial number of unexplained cases.

The previous contribution to research of next generation sequencing and GWAS studies
PD: There are 7 well-validated Mendelian genes for PD. However, many families and early onset cases with PD do not have a known genetic cause. In our cohort almost 90% of putative autosomal dominant families do not have a known pathogenic mutation. Similarly, around 90% of individuals with early onset PD do not have a known genetic course [2]. Conventional, singleton based whole exome sequencing has not made a significant impact on the identification of new Mendelian gene for PD so far, although this has proved to be an efficient technique for identifying variants in known genes. GWAS studies have identified 28 loci spanning 24 genes that contribute to the risk of sporadic PD, and indicated an overlap between mechanisms involved in sporadic and Mendelian PD.

Early onset dystonia: Several mendelian dystonia genes have been identified, DYT1 being the most important one. Some genes, such as ANO3, GNAL and CIZ1, have been discovered more recently for isolated dystonia, but they are probably not a common cause of classic cervical dystonia. Conversely, the phenotype associated with TUBB4A mutations expanded from that of isolated dystonia to a syndrome of hypomyelination with atrophy of the basal ganglia and cerebellum (H-ABC syndrome). Similarly, ATP1A3 mutations cause a wide phenotypic spectrum ranging from rapid-onset dystonia-parkinsonism to alternating hemiplegia of childhood. So far, GWAS studies have failed to identify a convincing association with any risk loci/SNP which could convincingly be replicated by others.

Complex parkinsonism:
The complex parkinsonian conditions include a range of "pallido-pyramidal" syndromes in which there is combined degeneration of basal ganglia and motor pathways, often with concomitant dementia, dystonia, basal ganglia iron deposition, leukodystrophy, eye movement disorders or ataxia. They are usually autosomal recessive disorders and an increasing number of genes have been identified.

Important issues in movement disorders
All of these diseases are phenotypically heterogeneous and gene identification will be accelerated and will benefit from the standardised national data collection that will be deployed in the 100K genome project.

PD: PD is particularly important because of the opportunity to use Mendelian forms of PD to elucidate mechanisms relevant to therapeutic development for a common sporadic neurodegenerative disease. The current state of PD research has been dramatically transformed by genetics highlighting the important of alpha-synuclein aggregation, mitochondrial function/mitophagy, LRRK2 kinase activity and lysosomal function [3]. We are actively engaged in developing therapies based on all of these areas partners. PD is a neurodegenerative disease and at some level there are likely to be overlapping mechanisms shared between PD, Alzheimer disease and motor neuron disease.

Dystonia: Dystonia is a "functional" disorder, that is, there are often major disabling disease features that appear related to disordered basal ganglia function without neurodegeneration or observable changes at post-mortem. Dystonia can be viewed as a disorder of neural plasticity which has the potential to provide major insights into basic neurobiology. The genetic mechanisms are particularly interesting – many of the genes (e.g. DYT1 for idiopathic torsion dystonia) show reduced penetrance indicating the importance of environmental or genetic co-factors. Some genes (e.g. PRRT2 for Paroxysmal Kinesigneic Dyskinesia) have a substantial rate of *de novo* mutations related to apparently "sporadic" disease.

Complex parkinsonism: Complex parkinsonism shares overlaps with many other neurodegenerative diseases and syndromes and there are a substantial number of undiagnosed cases. Complex parkinsonism occurs more commonly in individuals who have parental consanguinity.

Both dystonia and complex parkinsonism are often rare or very rare disorders and the 100K genomes project represents a tremendous opportunity to collate these rare cases, with standardised data from across the UK.

Why genome sequencing, interpretation and functional investigation isrequired
There are likely to be several reasons why whole exome sequencing has not yet been successful in identifying all of the Mendelian genes in movement disorders including: i.) the use of singleton cases rather than familial based sequencing leading to a dramatic loss of power in a heterogenous disorder and an inability to detect de novo mutations; ii.) failure to detect copy number variants (CNV); iii.) incomplete coverage of some genomic elements likely relevant to Mendelian disease (e.g. exon1, intronic regions that control splicing) and iv.) "lumping" together disease phenotypes likely represent different disorders. The 100K genome project represents a great opportunity to make a one-stop whole genome test for movement disorders available and to accelerate UK translational research.

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

Aims:

1) To develop a single test for Mendelian forms of PD, Complex parkinsonism and dystonia
2) To identify new disease genes and variants for PD, Complex parkinsonism and dystonia
3) To engage patients and families with Mendelian movement disorders in translational medicine

Specific aims:

A. Deep exploration of the contribution of known disease genes to undiagnosed movement disorders in the UK

We will use family /trio based whole genome sequencing to identify known and novel candidate variants in known disease genes for PD, complex parkinsonism and dystonia. There will be significant advantages in grouping these disorders together because of the potential clinical and functional overlap e.g. we recently identified variants in the GTP-CH1 dystonia gene as being relevant to typical late onset PD. This aspect of the work will reveal whether i.) comprehensive exonic coverage provided by whole genomic sequencing will identify known and new single nucleotide variants (SNVs), small indels and whole exon deletions relevant to disease, ii.) routine family based sequencing will be helpful in identifying *de novo* mutations causing disease, iii.) in concert with bioinformatics and functional analysis, routine family based sequencing will be useful in categorising variants of unknown significance. For example, we have identified >30 rare variants in the LRRK2 gene but currently only 5 are well established to be disease causing based on segregation in *bona fide* autosomal dominant families.

The major outcome measure in this phase will be the annotation and identification of new pathogenic variants in established disease genes. This will have major implications for genetic counselling and *in vitro* and animal based disease models

B Identification of new disease genes through integration of deep phenotype data and whole genome sequence

We believe that clustering together discrete disease groups using comprehensive 100 K phenotype data will enhance our power to discover new disease genes. This will require close collaboration with the phenotype sub-domain. This might follow both hypothesis led (e.g. clustering together individuals with basal ganglia iron deposition) and hypothesis free data driven approaches. Using the existing family based sequence data and the advantages of whole genome data listed above we will likely generate a list of candidate genes for specific disease sub-types. "Proof" of disease pathogenicity is likely to involve collaboration with functional, national and international partners who will be able corroborate potential novel disease genes through functional analysis and `genetic analysis.

The major outcome measure in this phase will be the annotation and identification of new disease genes. This will have major implications for genetic counselling and *in vitro* and animal based disease models. Furthermore, this may generate IP for 100K-genome in terms of future genetic diagnostic testing and the development of new therapeutic approaches.

C Engagement of the UK 100K genomes cohort in translational medicine

We believe that genetic sequence data will lead to specific therapies. We will endeavour to ensure that patients and families recruited through 100K genomes are engaged in translational medicine including: i.) Functional analysis – recruitment of patients to nationally available stem cell repositories to enable disease modelling and ii.) Therapeutic trials – recruitment of patients and families to studies aimed at targeting specific genetic pathways. We are currently engaged in translational therapeutic trial development related to LRRK2, GBA and parkin/PINK1.

The major outcome measure in this phase will be the numbers of patients recruited to functional translational and therapeutic translational studies.

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 Movement Disorders GeCIP partners have close ties with groups working on mitochondrial disease and movement disorders disease models together with links with a number of pharma companies including Edison, GSK, Eisai and Dextra l.

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

We will work closely with the cross-cutting GeCIP training domain and the Neuro-GeCIP training research plan. The research project provides the opportunity to train laboratory, clinical research fellows and allied health care professionals. We will present projects for the MSc in Genomics coordinate by Health Education England, and additional PhD projects hosted by partner Universities.

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

Prof Huw Morris PhD FRCP, Professor of Clinical Neuroscience, UCL - 15 years experience of research in Neurogenetics, PD and complex parkinsonism. Experienced in developing disease cohorts (PROBAND, PROSPECT, Parkinson's families project); characterisation of individuals with Mendelian disease (e.g. Early onset PD, myoclonus dystonia syndrome) and disease gene identification (e.g. C9orf72, commonest Mendelian gene for PD).

Prof Oliver Bandmann MD PhD FAAN, Professor of Movement Disorders Neurology, University of Sheffield – > 15 yr experience of research in Neurogenetics, PD and dystonia. Particular experience in the development of new disease models and drug development for stratified PD patients' cohorts.

Prof Nick Wood PhD FRCP Professor of Neurology, UCL - >20 years experience of leading Neurogenetics research. Has identified >5 Mendelian disease genes in PD and dystonia.

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 a comprehensive molecular diagnosis for patients with movement disorders. This will feed directly into national commissioned NHS services through liaison with the 23 NHS diagnostic laboratories.

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

PD and atypical parkinsonian disorders are progressive neurodegenerative disorders associated with high morbidity and resulting in major costs. Similarly, most dystonias cannot be treated

effectively with oral medication.
The development of a single test for Mendelian forms of PD, Complex parkinsonism and dystonia (aim 1) will greatly facilitate future routine genetic testing for these disorders in clinical practice. The identification of new disease genes and variants (aim 2) will lead to better insight into mechanisms leading to these conditions. Engagement with patients and families with Mendelian movement disorders in translational medicine (aim 3) will facilitate the subsequent formation of UK-wide, genetically stratified patient cohorts which will be particularly suitable for "personalized medicine" drug trials.

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

We do not currently have formal commercial partnerships in place but would look to develop these, following the Genomics England guidelines and policy as needed.

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

1. Cookson MR, Bandmann O. Parkinson's disease: insights from pathways. Hum Mol Genet. 2010 May;19(1):21–7.

2. Kilarski LL, Pearson JP, Newsway V, Majounie E, Knipe MDW, Misbahuddin A, et al. Systematic review and UK-based study of PARK2 (parkin), PINK1, PARK7 (DJ-1) and LRRK2 in early-onset Parkinson's disease. Mov Disord. 2012 Oct 6;27(12):1522–9.

3. Kumaran R, Cookson MR. Pathways to Parkinsonism Redux: convergent pathobiological mechanisms in genetics of Parkinson's disease. Hum Mol Genet. 2015 Oct 15;24(R1):R32–44.

Detailed research plan

Full proposal (total max 1500 words per subdomain)

Title *(max 150 characters)*	**Genome sequencing in very rare inherited neurological disorders: Identifying overlapping common pathological pathways**

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).*

Rare inherited neurological disorders such as hereditary spastic paraplegia (HSP), leukodystrophy, moyamoya disease and idiopathic basal ganglia calcification are separately rare with individual disease incidences of around 1-2 per 100,000 but altogether they account for a significant proportion of neurological disease. The molecular diagnostic rate varies between these disorders, in HSP around 60-70% of the genes have been identified but in disorders such as moyamoya disease the number of disease genes are few. No specific treatments are available in these conditions and identifying disease genes and pathways in these groups will not only allow an accurate diagnosis but also reveal important information on overlapping pathways in commoner neurological disorders.

The previous contribution to research of next generation sequencing studies
Whole exome studies have had a major impact on the molecular diagnosis of rare neurological disorders with the benefit of families to prove these genetic defects. This has led to the identification of a wide array of new genetic causes, revealing new mechanisms potentially amenable to treatment (see Novarino et al, 2014). However, conventional exome approaches have only identified molecular defects in approximately 60-70% of families at best, in the well characterised HSP series and only 5-10% in diseases such as moyamoya disease and idiopathic basal ganglia calcification, leaving a large number of rare disorders without a genetic diagnosis.

Important issues in very rare neurological disorders
Identifying the molecular basis of rare inherited neurological disorders will have important implications for the family. Many patients have struggled with the difficulty in coming to terms with their disorder and a disease gene will provide reliable predictive and prenatal diagnosis and genetic counselling for each kindred.

The great advantage of the Neurology Genomics England Clinical Interpretation Partnership (Neuro-GeCIP) is the national provision of expertise and a sample collection framework to facilitate the inclusion of the majority of undiagnosed cases and families in the genome sequencing. A number of clinicians have specific clinics that cover these disorders and interact with patient societies which will facilitate the enrolment of large numbers of well-characterised patients.

In the majority of rare inherited neurological disorders there are no specific treatments, only supportive case. The disease pathways identified in these rare neurological disorders may provide important therapeutic clues and overlap with other common neurological disorders such as the overlap between HSP and leukodystrophy with multiple sclerosis.

Inherited neurological disorders have specific clinical features that allow the categorisation of these disorder but there is substantial phenotypic overlap and heterogeneity between the different genetic causes means that a clinical diagnosis rarely points to a specific causative gene and the use of genome sequencing will be essential in identifying the exact genetic causes. The

detailed phenotyping and the availability of omic samples will be important in proving genetic defects and a functional readout to correlate with the genetic abnormalities, such as the IFN levels associated with a number of the idiopathic basal ganglia calcification disorders.

Finally there are a number of European and Worldwide networks that have been established to collaborate on the clinical, biomarker and therapeutic aspects of rare diseases. These will be important in proving disease genes in comparing genomic data, deep phenotyping and also in the joint development of therapies. These networks such as European Rare Disease Network (ERN), SPATAX and UK HSP network.

GeCIP members to help guide strategy and provide their expertise as part of a 'Rare Disease Clinical Data Working Group'**.**

Why genome sequencing, interpretation and functional investigation is required
There are a number of reasons why whole genome sequencing should be applied:

1. Exome sequencing has failed to identify a molecular basis of these diseases in a significant proportion of cases.

2. The undefined families are mostly small or individual cases that will be hard to genetically dissect and require the accuracy of genome sequencing and use of all aspects of the GeCIP.

3. There have been very few instances of disease causing structural variants such as deletions and *de-novo* changes causing these rare neurological disease that are juvenile or adult onset. This probably reflects a lack of appropriate technology and directed studies. The analysis of trios through the GEL project is highly likely to identify new structural and *de-novo* dominant defects as well as the expected AD or AR disease causing nucleotide changes.

4. It is likely that the analysis of whole genome sequence will provide the first 'single-shot' comprehensive test for these disorders and the neuro-GeCIP will produce a framework for this future diagnostic process to be as streamlined and effective as possible.

5. For rare diseases having ethnically matched control groups and collaborative cohorts will be essential in proving the variants identified from genome 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.*

Aim: The over-arching aim is to determine whether whole genome sequencing provide a near-comprehensive form of analysis to identify new disease genes, pathways and form the gold standard for NHS diagnostic testing of rare inherited neurological disorders.

Specific aims:

A. For patients with rare inherited neurological disorders, there will be various levels of variant validation:

1) To determine whether whole genomic sequencing will identify known and new single

nucleotide variants (SNVs), small indels or whole exon deletions

2) To determine whether *de novo* mutations cause novel de-novo dominant or recessive disorders.
3) Levels of pathogenicity will depend on the mutation type, variant change and gene and to help prioritisation we will overlap methods used here with the cross-cutting validation GeCIP, our prioritisation research plan and the neurogenetics diagnostic research plan.

B. To determine whether somatic mosaicism explains single cases based on the potential availability of other tissues.

C. Systematic deep phenotyping of cases and efficient capture methods. This will involve customising the data capture and algorhythms from other disorders and tailoring them towards specific features in the rare disorders such as spasticity in HSP and calcification on CT in Fahr's syndrome. Phenotyping methods will be taken after collaboration with cross-cutting GeCIPs such as Electronic Records (Harry Hemingway) and Enabling Rare Disease Translational Genomics via Advanced Analytics and International Interoperability and (Kate Bushby, Michael Simpson, Eamonn Sheridan and Eamonn Maher) and from the neuro-GeCIP phenotyping research plan.

D. Understanding of gene function.
(i) Expression: the gene is expressed in tissues relevant to the disease of interest and/or is altered in expression in patients who have the disease.
(ii) Predicted protein function: the gene product interacts or performs a similar biochemical function with functional pathways implicated (genetically or biochemically) in the disease of interest.
(iii) Experimental functional effects: the variant significantly alters levels, splicing, normal biochemical function or cellular phenotype of the product of the affected gene. This is shown either in affected patient cells, or a well-validated in-vitro, ideally neuronal model system where rescue of the cellular phenotype is achieved by the wild-type gene or variant gene knockdown. A range of different biochemical and cellular approaches are available to validate pathogenicity of diseases at the genetic level.
(iv) Further models: the variant is consistent with the disease phenotype in fruit fly, zebrafish or mouse models and the cellular phenotype of human iPSC lines/CRISPR/CAS9 system. We also have permission to deposit HSP genetically defined fibroblasts to the Wellcome Trust funded Human Induced Pluripotent Stem Cell Initiative (HipSci) for reprogramming into iPSC lines.

E. Identifying therapies from pathways. This will involve working closely with industry. The disease phenotypes and pathophysiology of some rare disorders overlaps with common conditions and this will potentially be attractive to industry to invest time and money. An example of this is in Prof David Rubinsztein's work on Huntington's disease (HD), in first identifying autophagy as an important pathway for clearing aggregate-prone proteins, then identifying drugs that upregulate autophagy signalling pathways as therapeutic candidates. One of these drugs, Rilmenidine, has beneficial effects in a mouse model of HD, has minimal side-effects and has recently successfully completed a safety trial in Huntington's disease patients (EudraCT 2009-018119-14) and maybe beneficial to other disorders caused by abnormalities in aggregate-prone proteins.

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

We plan to work closely with a number of cross-cutting GeCIPs to help with the aims of this

research plan. We have also established international collaboration to replicate or repudiate identified variants with the major sequencing laboratories of Professors Andy Singleton (NIH), Prof Stephan Zuchner, Prof Alexis Brice and Prof Peter Heutink (see letters).

We are also part of the HSP Alliance for Treatment, a European and Worldwide consortium funded by the FSP Society in the USA to share clinical phenotypes, sequencing data and functional work.

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

We will work closely with the cross-cutting GeCIP training domain and the Neuro-GeCIP training research plan. The research project provides the opportunity to train laboratory, clinical research fellows and allied health care professionals. We will present projects for the MSc in Genomics coordinate by Health Education England, and additional PhD projects hosted by partner Universities. In addition to the continuation of site-based training, the GeCIP would promote integration and cross-training, through workshops and other activities. In addition, the collaboration implicit in the GeCIP approach to research would be used to develop a common doctoral training program across sites.

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

Prof Tom Warner and Prof Evan Reid and interested in inherited rare disorders with a particular focus on hereditary spastic paraplegia.
Prof Henry Houlden and Prof Patrick Chinnery are interested in all aspect of rare neurological disease.
Dr Arianna Tucci (UCL) currently leads the UK HSP network and database and co-ordinates the sending of HSP genetically defined fibroblasts to the Wellcome Trust funded Human Induced Pluripotent Stem Cell Initiative (HipSci) for reprogramming into iPSC lines.

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

This GeCIP builds a closer working relationship between neurology clinicians, clinical geneticists and research genetics teams. This provides an immediate benefit to patients and families in enhancing the quality of information and advice during and after genetics testing, and in the interpretation of results. It will enable more sophisticated modelling of disease risk, explain phenotypic variation within and between families.

In the longer term, the GeCIP will lead to novel and effective therapies via the clinically validated model systems in preclinical pathways, from cell biology and high-throughput screening programs in early preclinical models through to optimised selection of candidate therapies and target populations in clinical trials for efficacy.

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

Clinician and patients - Families; diagnosis, predictive, prenatal testing. NHS organisations and

cost savings.
Pharmaceutical companies - New disease pathways, targets, stratified patient groups, pre-clinical research models, clinical Trials.
Neurogenetic laboratories - New tests.

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?*

Potential new gene mutations and pathways may be patentable. Involving industry will be a helpful to fund and accelerate further work

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

Novarino et al, Exome sequencing links corticospinal motor neuron disease to common neurodegenerative disorders. Science, 343 (2014), pp. 506–511

Detailed research plan

Full proposal (total max 1500 words per subdomain)

Title *(max 150 characters)*	**Understanding disease and deep phenotyping**

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).*

- To optimise clinical data and sample collection, clinical reporting and data validation and interpretation.
- To improve understanding of the implications of genomic findings and improve the accuracy and reliability of information fed back to patients. To add to the knowledge of the genetic basis of disease.
- To provide a sustainable thriving training environment.

Genetic neurological diseases are a real challenge for our society. Despite the fact that they are individually considered as rare, together they affect many, bringing a huge burden for patients and their families as well as to the healthcare system. Many of these conditions lead to premature death or chronic debilitation. They are currently incurable and because of their underlying genetic causes are associated with risk of recurrence for the families affected. Understanding and better describing neurological phenotypes is key to identifying patients and providing highly stratified cohorts in readiness for clinical trials.

Clinicians and scientists frequently have an expertise in many different genetic neurological conditions, and are involved in common research networks, databases, registries, and multicentre clinical trials. Extensive patient DNA and tissue collections are available, accelerating research and facilitating translational steps such as developing biomarkers and treatment concepts. This will be exploited within this GeCIP and the established networks will be used to define core phenotypic data sets to be collected across the diseases included in this GeCIP. In consultation with other disease-specific GeCIPs, we will ensure that areas of clinical overlap will be described in a harmonised way.

Linking phenotypic and long-term outcomes data with whole genome sequences will generate a uniquely rich and longitudinal dataset which will be invaluable to advancing understanding of these conditions.

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

The main aims of this sub-domain's research will be:

- To extend currently described phenotypes
- To better describe and understand different clinical manifestations of the same disease in order to inform the choice of inclusion/exclusion criteria and increase patient identification for clinical trials. Previous experience has shown the importance of understanding phenotypes to trial success.
- To enable global matchmaking of patients to research studies based on well-characterised

and harmonised phenotypic data.

- To enable biomarker discovery and validation – biomarkers may be specific even where a disease is not. Deep-phenotyping will allow biomarkers to be linked to specific aspects of a disease. These biomarkers may be relevant across several different conditions. Deep-phenotyping will also generate cohorts for biomarker analysis.
- To enable the identification of disease modifying factors – where a genetic cause is the same but the clinical profile is not, using deep-phenotyping data together with the genetic profile can shed light on reasons for differences in disease severity, age of onset and other differences in clinical features
- Understanding clinical overlap between different diseases with similar phenotypic features may point to opportunities to repurpose drugs from other disease areas. Well harmonised and systematic deep-phenotyping across diseases within neurology but also in other GeCIPs will indicate similarities in pathways which were not understood previously. To this end, we anticipate working with many other GeCIPs but in particular those of Hearing and Sight, Endocrine and metabolism, Paediatrics and Cardiovascular.
- Coordination of deep-phenotyping between disease-specific sub-groups – the harmonisation of how and what data is collected, recorded and shared is critical to success. This subdomain is keen to work with similar sub-domains in other GeCIPs that will capture patients with neurological features. We can also bring experience from previous projects where we have been responsible for deep-phenotyping and benefit from the 'lessons learned' there. These include:

  - the importance of pedigree drawing tools
  - availability of test results, including MRI scans for analysis
  - central coordination (of determining the agreed datasets, data-entry, sharing rules and trouble-shooting)
  - inclusion of mandatory data-items
  - We will use the deep-phenotyping data collection in NeurOmics (http://rd-neuromics.eu/) as a starting point for determining the core and compulsory data elements in the Neurology GeCIP.

- Informing disease-monitoring – by better and more detailed understanding of phenotypic profiles and the linking of these to other data, disease progression can be better anticipated, complications anticipated and the likelihood of treatment side-effects better assessed. This all leads to better patient treatment and care.
- Establishing links for the integration of patient data into patient registries through close collaboration with international platforms for rare diseases, like RD-Connect (http://rd-connect.eu/)

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.*

Industry:

Pfizer's generic rare-disease consortium – potential for repurposing of drugs

Genzyme and Ultragenyx – identification of patients with genetic neuromuscular diseases (GNE-myopathy, Pompe disease) from a cohort of 1000 patients with limb girdle weakness of unknown origin by exome sequencing (MYO-SEQ, http://myo-seq.org/)

Patient organisations – keep the patient community informed, ensure that the research priorities of the GeCIP meet those of patients, ensure that aspects of clinical profile important to patients are captured in this research.

Other projects:
Projects and participants of the Matchmaker Exchange initiative (http://www.matchmakerexchange.org/) : use of the rare genetic disease platform(s) to identify additional patients/families to confirm findings from the 100000 Genomes.
NeurOmics (http://rd-neuromics.eu/): sharing of best practice and experience in deep phenotyping across different neurological conditions; link with biomarker discovery work in this project and others – deep phenotyping will generate cohorts for biomarker analysis.
Patient registries: Update patient registries with data according with participant's wishes and consent
Care and Trial Site Registry (https://ctsr.uniklinik-freiburg.de/): Identification of centres with cohorts of patients with specific genetic neurological conditions.
Human Phenotype Ontology (HPO): to update and improve the terms available to describe clinical features. Also to extend what is covered by the HPO to include test results, description of MRI etc

Other disease specific GeCIPs:
Hearing and Sight
Endocrine and metabolism
Paediatrics
Cardiovascular

Cross-cutting GeCIPs:
Enabling Rare Disease Translational Genomics via Advanced Analytics and International Interoperability (ERDTG)
Ethics and Social Science
Education and training

The impact and success of this sub-domain will be maximised through data-sharing. This will be important to allow matchmaking and patient identification and the integration of data into registries. This will require that the necessary permissions are in place and this subdomain will therefore work with Prof Parker's Ethics and Social Science GeCIP to request that data-submitters follow the standards of consent required (and followed in projects such as NeurOmics and RD-Connect) which allow data-sharing across the rare-disease community.

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

We will work closely with the cross-cutting GeCIP training domain and the Neuro-GeCIP training research plan. Training in the use of OpenClinica to ensure consistent recording of phenotypic data. Expert workshops with HPO team to improve and extend HPO. This training will be relevant across other disease-specific GeCIPs and we will therefore work with the cross-cutting Education and Training GeCIP to explore rolling this training out across participating centres.

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

This sub-domain is led by Professor Straub who, along with a team at Newcastle, has extensive experience describing patients' clinical profile as well as in EU (FP6 and FP7) and industry funded projects where he has led tasks to establish deep-phenotyping practices and infrastructure (NeurOmics, MYO-SEQ, Seq-NMD). As part of this work the team at the John Walton Muscular Dystrophy Research Centre in Newcastle has coordinated training in the use of the Human

Phenotype Ontology and clinical databases in collaboration with RD-Connect (coordinated by Professor Hanns Lochmüller, also at Newcastle).

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.*

Clear phenotypes will allow a better understanding of likely disease course and therefore the anticipation of complications. Family-members at risk can also be monitored and offered early therapy. Meanwhile family-planning can be better informed. With a better described phenotype, patients are also more likely to be able to be enrolled in clinical trials.

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

The linking of clinical profiles with new diseases and new mutations will benefit diagnostic test choices and result in more patients with a genetic diagnosis.

Better and harmonised descriptions of patients' phenotypes will allow matchmaking to identify second cases to help confirm genetic diagnoses.

Establishing areas where phenotypes in different diseases overlap will lead to better understanding of similar pathways and suggest opportunities for repurposing licensed drugs.

Patients with a genetic diagnosis and well described phenotype will be more likely to have the opportunity to enrol in clinical trials.

By better anticipation of adverse reactions to treatments through identifying disease modifiers and genetic profiles associated with particular phenotypes or disease course, treatment can be better targeted to individual patients, maximising effectiveness and minimising side effects. This is of great benefit to patients and also to the cost burden of providing therapy for neurological conditions.

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?*

Well characterised patient cohorts with genetic diagnoses are an incentive to industry to do trials in these diseases.

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

1. Delude CM. Deep phenotyping: The details of disease. Nature. 2015 Nov 5;527(7576):S14-5.
2. Baynam G et al. Phenotyping: targeting genotype's rich cousin for diagnosis. J Paediatr Child Health. 2015 Apr;51(4):381-6.
3. Robinson PN. Deep phenotyping for precision medicine. Hum Mutat. 2012 May;33(5):777-80.
4. Lanktree MB et al. Phenomics: expanding the role of clinical evaluation in genomic studies. J Investig Med. 2010 Jun;58(5):700-6.
5. Tracy RP. Deep phenotyping': characterizing populations in the era of genomics and systems biology. Curr Opin Lipidol. 2008 Apr;19(2):151-7.

Detailed research plan

Full proposal (total max 1500 words per subdomain)

Title *(max 150 characters)*	**Transforming the diagnosis and understanding of neurological disorders through the creation of resources for variant prioritisation**

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).*

Over the past 5 years there has been a massive growth in genetic testing both in terms of the scope and the numbers of individuals offered genetic testing. This "genomic" revolution has had an immense impact on the diagnosis and understanding of neurological conditions. There is good reason to believe that whole genome sequencing (WGS) will not only increase discovery of pathogenic variants and genes, but will have a disproportionate impact on the understanding of neurological disorders as compared to other disease areas.

Firstly, existing RNAseq-based analyses of the human brain have demonstrated significant differences between the transcriptome as defined by current commonly-used annotation providers and empirically-based transcriptome definitions. In fact, a recent study demonstrated that >10% of transcribed regions in human brain were located within intergenic regions and that a high proportion of reads derived from cytoplasmic RNA extractions mapped to deep intronic regions. Furthermore, there is increasing evidence to show that lncRNAs expressed in the developing brain are better conserved evolutionarily than those expressed in other tissues. While these findings reflects the fact that existing genome annotation is largely based on "easy to access" human tissues, it means that WGS in neurological disease would be expected to identify significant numbers of pathogenic variants "missed" by exome sequencing.

Secondly, given the complexity of gene expression and processing within human brain, pathogenic variation within non-coding regulatory regions is likely to be more important in neurological diseases than other types of conditions. For example, it has recently been shown that unlike conventional single-step splicing of introns, intron removal is a multi-step process in some neuron-specific genes and that this depends on splice sites deep within long introns. This example emphasises the importance of regulatory sequences not normally captured by exome sequencing, but that nonetheless represent candidate sites for pathogenic variation.

Realising the potential of WGS for neurology will be challenging. The reliable identification of genuine disease-associated genetic variants, particularly in genomic regions currently annotated as intergenic or intronic, from amongst the broader background of variants present in all human genomes that are rare, but not actually pathogenic is a major concern. Addressing this issue requires a systematic approach to variant prioritisation and assignment of pathogenicity based on multiple different data sources.

We believe this should include the following:

i) Development and implementation of tools to improve the information content of phenotyping data models and enable accurate similarity scoring among patients
ii) Development of high confidence control WGS data sets for effective variant filtering, particularly in the context of adult-onset disorders
iii) Developing high quality annotation of genes (both protein-coding and non-coding),

isoforms and regulatory sites relevant to the human central nervous system across development

iv) Innovative use of patient samples, including iPSC-derived cells for disease modelling, to improve the quality of variant classification when insufficient statistical evidence is available or when this can provide biological insights

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.*

Given that approximately 20% of all diseases covered by the Rare diseases Project fall within the remit of the neurology GeCIP domain and that this domain currently has the highest levels of recruitment (even with the exclusion of patients recruited for investigation of intellectual disability), it is critical that the bioinformatics subdomain delivers on the promise of WGS in this patient group.

The research plans proposed are driven by the need to guard against falsely calling variants as pathogenic, while still enabling novel genes and regulatory positions to be identified. The research plans are focused on areas specifically required for the neurology domain and do not include analyses which while important are best delivered through cross-cutting GeCIPs, including improvements in variant calling (particularly structural variation) and detection of somatic mosaicism.

1. Development and implementation of tools to improve the information content of data neurology data models and enable accurate similarity scoring in the field

   The simplest and most reliable means of demonstrating pathogenicity of a novel gene remains statistically significant genetic association. However, this requires affected individuals with the disorder to be accurately described and matched. Given the complexity of neurological phenotyping recognising similarities in patient phenotypes can be difficult. Through initiatives such as NeurOmics (leading members of which are also key subdomain leads within the neurology GeCIP) and by working with the Quantitative Methods, Machine Learning and Functional Genomics GeCIP (QMMLFG) we will bring together clinical expertise and robust methodologies to maximise the value of phenotyping information.

2. Developing high quality control variant databases for filtering

   The identification of pathogenic variants in neurological disorders has been hindered by the absence of well-defined “control” genetic data against which to filter potentially novel variants. For some adult-onset conditions such as Parkinson’s diseases, which are characterised by a prolonged pre-symptomatic course when brain pathology is present, but the patient remains unaware, this is a significant issue. In such cases, accurately establishing “control” status may only be possible at post-mortem through detailed neuropathology or use of data generated from healthy centenarian control individuals. For this reason, we will develop high quality control databases for variant filtering by collating information from public initiatives, as well as working with GeCIP members. This includes collaboration with lead researchers at the UCL IoN and NIH presently undertaking WES/WGS of neuropathologically-confirmed control individuals. This will require collaboration with cross-cutting GeCIP to ensure that integration of sequencing data from multiple sources is performed in a reliable way which accounts for bias and uncertainty in

WES and WGS data.

3. Defining the transcribed portion of the human genome within the human nervous system

   An accurate map of transcription within the human nervous system is required in order to provide even the most basic annotation of novel variants identified in patient samples. We will re-analyse and collate publicly available RNAseq data generated from human CNS and PNS tissues, as well as relevant cell types. This includes data provided by major national and international consortia, including GTeX, NABEC and UKBEC. This analysis will be performed with diagnostics in mind to ensure that relatively conservative definitions of significant transcription are applied.

4. Annotating central and peripheral nervous system-specific regulatory positions within the human genome

   Within the last month alone 2 new initiatives have been announced expressly aimed at improving annotation of non-coding regions of the human genome in brain, namely PsychENCODE and BrainSeq. This type of information is also one of the key outcomes of brain-specific eQTL analyses being conducted by GTeX, UKBEC and NABEC, of which the latter two are collaborators. While such analyses have immense promise, they are currently in their infancy – the first data release from PsychENCODE will be in January 2016 and the release of brain-specific eQTL and transcriptomic data has lagged behind other tissues in the GTeX project. Given the anticipated growth in this type of data, collating data available directly to the GeCIP as well as through major public initiatives will form one research area.

5. Transcriptomic analysis of patient-derived samples

   While statistical proof of variant pathogenicity is the ideal, in some cases this may not be possible. In such circumstances RNA sequencing of patient-derived samples (blood and iPSC-derived cell types) and subsequent allele-specific expression may be a useful way to determine experimentally the impact of a potential pathogenic variant on gene expression. Since this is a genome-wide form of analysis, it will also provide a broader functional read-out of the impact of a potential pathogenic variant. Although this approach will not be appropriate in all cases, it has the major advantage in that it will require a single experimental and analytic pipeline. We will specifically work with the transcriptomics and RNA splicing (TRS), and Statistical Genomics cross-cutting GeCIP subdomains in this area.

6. Creation of an integrated web-based resource for variant interpretation

   Given the regional, cellular and molecular complexity of human brain, we believe there is a strong case for the creation of a tissue-specific resource for assessing the impact of genetic variation on human brain. At present no comprehensive resource of this kind exists. Instead potentially relevant data is located within multiple different repositories and sites, not all of which are publicly accessible or curated to the same degree. We intend to create an integrated resource which can be queried by gene or genomic position. We will ensure that it can be accessed in an automated manner for larger queries or integration with existing laboratory workflows as well as through a web-interface for less expert users. We believe this resource will not only benefit diagnostic laboratories, but will also be a valuable resource for the neuroscience research

community.

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

We will establish the following collaborations listed by type:

1. Collaborations with disease-specific GeCIPs
   - Hearing and sight
   - Endocrine and metabolic medicine

2. Collaborations with all cross-cutting GeCIPs, but in particular with:
   - Electronic Patient Records
   - Advanced Analytics
   - Quantitative Methods, Machine Learning and Functional Genomics

3. National/International consortia
   - International Parkinson's Disease Genomics Consortium
   - UK Brain Expression Consortium
   - North American Brain Expression Consortium

4. Disease-specific charities
   - Alzheimer's Research UK
   - Parkinson's Disease UK
   - Muscular Dystrophy UK

We will ensure that all collaborations are conducted according to the principles and established practice of Genomics England and the Global Alliance for Genomics and Health (GA4GH).

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

We will work closely with the cross-cutting GeCIP training domain and the Neuro-GeCIP training research plan. We will attract trainees from both clinical and laboratory settings, as well as from the emerging field of healthcare informatics. We envisage clinical trainee involvement particularly in the development and validation of phenotyping capture tools. We are particularly interested in encouraging trainees to develop bioinformatics skills through higher research degrees (MDs, MRes and PhDs).

We will also actively encourage applications from within the GeCIP membership for more senior post-doctoral training through existing schemes offered by the MRC, Wellcome Trust and NIHR as well as disease-specific funders.

Finally, we will develop short courses (also available online and potentially as part of MSc teaching) to develop an increasing awareness of the complexity of determining the pathogenicity of variants and the tools available to address this issue. These resources will be aimed at medically-qualified clinicians and nurse specialists, as well as laboratory staff.
In this way, we will contribute to the development of the next generation of neuroscientists and clinical specialists.

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

Dr Mina Ryten, UCL/KCL MRC clinician scientist (co-lead)
Prof Chris Ponting, University of Edinburgh (co-lead)
Prof Jernej Ule, UCL
Dr Alan Pittman, UCL research associate

These investigators lead well-established neurogenomics and neurotranscriptomics research programmes studying human samples. The group will organise their joint work using face-to-face and teleconference meetings and potentially a collaborative working tools, such as Slack (slack.com).

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 most important output of our research plan is an integrated tissue and cell-specific resource for the investigation of genetic variants relevant to patients with neurological disorders. While it will undoubtedly also be highly relevant to the neuroscience research community, we intend to design this resource with clinical diagnostics in mind.

The most important output of our research plan is an integrated tissue and cell-specific resource for the investigation of genetic variants relevant to patients with neurological disorders. While it will undoubtedly also be relevant to the neuroscience research community, we intend to design this resource with clinical diagnostics in mind.

We will involve clinical scientists already experienced in reporting to test this resource against existing systems for utility, as well as with regard to integration with standard reporting procedures/ systems. Furthermore, we will explicitly test in silico predictions of pathogenicity (both negative and positive) using cell model systems including iPSC-derived neuro-relevant cell types to ensure that the accuracy and rigour of this approach is formally tested.

Given that we are collaborating with the cross-cutting Advanced Analytics, QMMLFG and Validation and Feedback domains we envisage making test versions of all resources available at development stage so that patients can fully benefit from the collective expertise with the project.

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

The primary aim of the research plans proposed it to realise the potential of WGS in the diagnosis and understanding of neurogenetic disorders.

The most important beneficiaries of this research will be NHS patients. Through the creation of an integrated resource for variant interpretation we intend to improve the likelihood of reaching a genetic diagnosis efficiently, which will be of direct benefit to patients and could reduce costs within some NHS services.

As part of this project, we will use patient-derived samples including iPSC-derived cell types when

possible to experimentally test in silico predictions of pathogenicity. Such approaches represent a first step towards personalised medicine within the NHS with the aim of providing patients with the best quality treatment.

The resources generated will also be used by the neuroscience research community in order to: i) improve the understanding of the pathogenic processes underlying both known and newly discovered genes, and ii) improve the quality of in vitro model systems for neurological disorders. The latter is particularly important for the development and testing of novel treatments and would benefit drug companies.

Finally, the data generated will be of value to basic scientists interested in gene expression splicing, regulation and evolution.

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 do not have any commercial partners in place and do not envisagethe planned research to result in commercially exploitable results.

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

Akbarian S et al, The PsychENCODE project. Nat Neurosci. 2015 Nov 25; 18(12):1707-12.

Darmanis S et al, A survey of human brain transcriptome diversity at the single cell level. Proc Natl Acad Sci USA. 2015; 112(23):7285-90

Jaffe AE et al, Developmental regulation of human cortex transcription and its clinical relevance at single base resolution. Nat Neurosci. 2015 Jan; 18(1):154-61.

Macaulay IC et al, G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. Nat Methods. 2015 Jun; 12(6):519-22.

Melé M et al, The human transcriptome across tissues and individuals. Science. 2015 May 8; 348(6235):660-5.

Ramasamy A et al, Genetic variability in the regulation of gene expression in ten regions of the human brain. Nat Neurosci. 2014 Oct; 17(10):1418-28.

Detailed research plan

Full proposal (total max 1500 words per subdomain)

Title *(max 150 characters)*	**Training for the future in Neurogenetics**

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).*

A key aim of the 100K Genome project is to use genomic medicine to improve health outcomes across the UK, within the NHS. To do this we will need to train both current practicing physicians and professions allied to medicine (PAM), and the next generation of physicians and clinical specialists to ensure that they are able to interpret and explain individual patient level genomic data, and to use the results of the 100K-genome project to develop new and improved treatment. This training programme will need to include specialists from clinical genetics and adult and paediatric neurology. The training sub-domain of neurological disorders is clearly of central importance to the overall project.

Traditionally, genetics teaching has focussed on modes of inheritance and classical genetic diseases. To equip and prepare our current and future clinicians we need to provide training in the new areas that will be imminently relevant to clinical practice including for example: the consent process for clinical and research-based genetic testing, the correct procedures in relation to incidental genetic findings, the definition of pathogenic and non-pathogenic variants, *de novo* pathogenic mutations, age-dependant variation in penetrance, genome wide association study variants, genomic risk profiles and personalised medicine. In addition, we propose that training for all clinicians in genomics should include some familiarity with areas that lie firmly within clinical genetics such as predictive, pre-natal and pre-implantation genetic diagnosis.

We envisage that the training subdomain will lead in:

1) Developing the skills of **Current Physicians** in neurological disease (consultant neurologists, paediatricians, geneticists)
2) Developing the background level of expertise amongst **All Trainees** in neurology and paediatric neurology
3) Ensuring that there is a sub-set of **Neurogenetics Trainees** with specific expertise in neurogenetics, most likely developed in conjunction with MD and PhD programmes with a specific neurogenetics research project.

Each of these areas will involve the development of skills and expertise in genomics with specific focus on the 100K-genome project. Early engagement in training is important as this will help to enable rapid recruitment to the 100K project during the recruitment phase, and correct communication between clinicians, patients and families within the project, and in the follow-up and dissemination of results.

Although these aims are generic and apply to each disease focussed domain, and are also covered in a specific cross-cutting training domain, we believe that engagement with the system focussed medical specialties (neurology and paediatric neurology) is essential to enable the training of the next generation of clinicians and to ensure the maximum benefit for the UK from the 100K genomes project.

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

Aim:

1) **Current clinicians** – developing the skills of current clinicians, through development of online and meeting delivered education, with CPD accreditation

2) **All trainees -** ensuring that core genomic education content is included in the neurology and paediatric neurology curriculum, both through an online module and through course content included in regional "Calman" training days

3) **Neurogenetics trainees –** developing a register of current trainees carrying out specific research in neurogenetics; additionally, as a minimum funding a 100K neurogenetics clinical training fellow in each genomic medicine centre

Specific aims:

Although we have highlighted separate target groups for training and education there will be overlapping content, which might be accessed by all three groups. The proposed training in the genomics of neurological disease can be considered in terms of delivery, development, implementation and assessment/metrics. Importantly, our proposal for training includes both the delivery of teaching and learning content and the identification of individuals within genomics medicine centres who would be local "champions" for the 100K genomes project and for learning and training. In our view, training is about both information and knowledge, and about specific people who will develop the 100K genomes programme now and in the future.

A) **Delivery**

i.) Brief online core course

We would aim that this brief core content would include "need to know" for the practicing clinician. This would comprise a slide set, with accompanying summary document and interactive questions and answers for self directed learning. This would also contain pointers for further learning. Ideally this would be completed by all practicing clinicians and clinicians in training. Time commitment 3-4 hours.

ii.) Extended online course

This would comprise and in depth core content review with further detail on the 100K genome project. This would comprise a slide set, summary document, journal articles and book chapters for review, and an extended set of interactive questions and answers with a commentary on the learning component

Time commitment 30-40 hours

iii) Meeting delivered teaching

Items i.) and ii.) would be established as online content which could be completed over the internet. However, the brief online core course could also be delivered as an academic teaching session at an annual meeting (e.g. Association of British Neurologists, British Paediatric Neurology Association) or at a regional "Calman" day for trainees.

B) **Development**

The development of core content and extended content would contain speciality specific material but would also contain generic material, which would be applicable to all medical specialties. Excellent training material has already been developed, for example in consent in the 100K-genome project. Development of clinician orientated modules, integrated with CPD and training, would involve co-ordination between the disease specific GECIPs, the cross-cutting training GECIP, Health Education England, the regional Calman SpR training programme, Royal College specialty curricula and other sources of training such as the Paediatric Neurology distance learning

programme and the PHG foundation. Development of these modules would require funded administrative and academic support.

C) **Implementation**

We would propose identifying and supporting:

i.) **Senior local learning lead** – aligned to genomic medicine centres, these individuals might for example be the NHS consultant or clinical academic who convenes the regional Calman training day in neurogenetics.

ii.) **Junior local learning lead** - aligned to genomic medicine centres, these individuals would be specialist registrars with an interest in neurogenetics

iii.) **Neurogenetics research fellows** – these individuals would be engaged in neurogenetics research independent of the 100K genome project who might benefit from support and collaboration with the 100K genome structure. They would be encouraged to interact with the 100K genome project.

in addition we propose

iv.) 100K genome neurogenetics clinical research training fellowship programme – this would be a separate fully funded programme with clinical research fellows from each genomics medicine centre pursuing research aligned to the 100K genome project. This might be in clinical informatics, bioinformatics, functional analysis, or new variant identification and interpretation. This would be a prestigious "blue riband" programme, which would ensure that young clinicians were embedded in the development and dissemination of the 100K genomes programme.

D) **Assessment/Metrics**

i.) Number of individuals completing core and extended online training

ii.) Encouragement of CPD content in genomics in annual appraisal for consultants

iii.) Development of genomics content in the Neurology and Paediatric Neurology Core Curriculum

iv.) Development of assessment in genomic medicine in, for example the RCP Neurology exit exam.

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.*

As above.

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

As above

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

Prof Huw Morris, Professor of Clinical Neuroscience, UCL

PhD in Neurogenetics, 15 years experience of Neurology/Genetics clinics including a shared clinic with clinical geneticists and genetic nurse counsellors at University Hospital of Wales Cardiff. Chair of Clinical Research and Academic Committee of the ABN, Member of the Specialist Certificate Exam ("exit exam") Board in Neurology for Royal College of Physicians, Neurology Representative as "Clinical Champion" on the Genomics Mainstreaming Medicine Project organised by the PHG foundation. Instructor on the Wellcome Trust Molecular Neurology course. Organiser of the ABN clinical research training fellowship scheme.

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.* N/A
Beneficiaries. *How will the research benefit patients and healthcare institutions including the NHS, other researchers in the field? Are there other likely beneficiaries?* Investing in genomics training and education will benefit patients, the public and the wider 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?* N/A
References. *Provide key references related to the research you set out.* N/A

Detailed research plan

Full proposal (total max 1500 words per subdomain)

Title *(max 150 characters)*	**Developing a streamlined system to interpret and report genome sequencing results in the diagnostic laboratory**

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).*

An important part of the 100,000 genomes project is the implementation and long-term sustainability of genome sequencing as an NHS diagnostic test for the UK population. Transferring the experience, knowledge and techniques gained from the 100,000 genomes project into the diagnostic laboratory setting will be essential in establishing this important test into routine NHS diagnostic test.

Several of the larger diagnostic genetics laboratories around the UK have already implemented next generation sequencing gene panels into practice but the use of genome sequencing will be the ultimate genetic test. There will be a number of important tasks to perfect in the development of stringent diagnostic genome sequencing that range from data handling and long term storage of results, through to genomic coverage and variant interpretation. In addition, in the first instance genome sequencing will likely be combined with Sanger sequencing but as laboratories become more experienced and confident the use of Sanger sequencing will likely become redundant.

Most defects that are likely to be pathogenic will have the following properties found in a known disease-causing gene or potentially pathogenic non-coding or copy number variants close to these genes.

A). The variant causes a missense, nonsense, splice site or frameshift amino acid change or non-coding change that segregates with the disease and usually shows evolutionary conservation.
B). The mutation is consistent with the reported phenotype and is not present in controls. New phenotypes associated with disease gene will be assessed in each subdomain.
C). Predicted mechanistic effect: variant is found at the location within the protein predicted to cause functional disruption (such as active site or protein-binding region) or the variant causes a predicted abnormality in the RNA or the amount of RNA.
D). In new disease genes we would expect the disease gene to be expressed in the diseased tissue such as the CNS, PNS and muscle. In addition we would expect additional cellular functional work.

There is a GeCIP already dedicated to Validation and Feedback of genetic variants (lead Prof Bill Newman). We plan to work closely with this GeCIP and focus on the validation, development and transference of genome sequencing specifically applied to neurogenetic conditions. There will be common themes across disease areas but there will also be areas specific to neurogenetics such as the use of brain expression databases in the interpretation of variants, the development of long genome sequencing reads to identify expanded repeat diseases and mitochondrial genome sequencing. This research plan will also overlap with our research plan on transforming the diagnosis and understanding of neurological disorders through the creation of resources for variant prioritisation and the cross-cutting GeCIP on Quantitative Methods, Machine Learning, and Functional Genomics to develop statistical approaches for causal variant prioritisation, identification and *in-silico* functional validation.

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

The main task of this research plan will be on validating genetic proof of pathogenicity of the defects identified from the genome sequencing, focussing on neurogenetic conditions in which we have the most significant knowledge. We will also work to develop statistical approaches for improved causal variant prioritisation, identification and *in-silico* functional validation to stream line the reporting system of these mutations back to clinicians and patients in the NHS.

A). Current steps to establish the likely pathogenicity of a variant in a known gene. There are a number of steps that will benefit from streamlining and automating, current steps include: Software to help analyse and interpret mutations such as Alamut software to gather a majority of the information required to analyse a variant, including: Conservation at DNA and protein level, SIFT/PolyPhen/aGVGD predictions for missense variants, physiochemical difference between amino acids (Grantham score) for missense variants, in-silico splicing analysis. Importantly, this is not just performed for intronic variants, the potential for all variants to affect existing splice sites or activate cryptic sites is always considered.
If the variant has been reported in public databases of genetic variation.
The location of the variant in the protein is considered – whether it occurs in a known functional domain or in the vicinity of previously-reported mutations.

Once all of this information is collated a decision is made on how to classify the variant. We use a 5-point classification scheme, which will still be applicable to genome sequencing results:

- 1: Not pathogenic or of no clinical significance (do not report)
- 2: Likely not pathogenic or of little clinical significance
- 3: Uncertain significance
- 4: Likely pathogenic
- 5: Definitely pathogenic

B). Currently variants are diagnostically confirmed by Sanger sequencing: Download reference sequence and design PCR/sequencing primers, avoiding known polymorphisms and repeat motifs. Keeping records of all steps of primer design. A diagnostic report is written. We are planning to write a single report for GEL cases that includes the results of all members of the trio in the same document. This would include an interpretation section with relevant information/references to support the conclusion of pathogenicity of the variant(s) detected. This is checked and authorised to be issued by a senior HCPC-registered Clinical Scientist.

Developing systems for improved NHS diagnostic laboratory interpretation and reporting will be essential in streamlining the reporting of whole genome sequencing results. This will require significant collaboration with other research plans in the neuro-GeCIP and collaboration with cross-cutting GeCIPs.

C). The three most important areas to develop include:

1. Development of rapid, secure and long term storage solutions for next generation sequencing data in the NHS diagnostic laboratory setting.

2. Tools to improve annotation and the development of new programs and statistical methods to efficiently identify, and prioritise, likely pathogenic disease causing variants based on the known disease genes and phenotypes.

3. The integration of patient phenotype, details on likely pathogenic variants (classification 3 to 5) in the sample, genome data on other family members, past knowledge of the variant and gene in question and *in-silico* functional genomic data for variant and gene.

In the short term likely pathogenic genome sequencing mutations will likely be combined with Sanger sequencing but as laboratories become more Sanger sequencing will likely become redundant.

Using genome sequencing to develop future diagnostic areas
Important future areas to consider are the analysis of the mitochondrial genome from genome sequencing in blood DNA as compared to muscle DNA. Here we would carefully investigate the coverage and heteroplasmy as compared with samples that we have already Sanger sequenced before expanding the genome analysis approach. Expanded repeat disorders will be a significant challenge in the future. Currently reads are not long enough but this may be possible in the future.

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

We plan to establish collaborative links with other GeCIPs important to diagnostics such as: Education and Training (Maxine Foster), Enabling Rare Disease Translational Genomics via Advanced Analytics and International Interoperability (Kate Bushby and Michael Simpson), Quantitative Methods, Machine Learning, and Functional Genomics (Martin Tobin) and Validation and Feedback (Bill Newman).

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

Training of NHS technologists and clinical scientists in genomic medicine will be essential to the future. We will work closely with the cross-cutting GeCIP training domain and the Neuro-GeCIP training research plan. We will develop national workshops, short courses aimed at NHS technologists and scientists. An MSc programme in Genome Medicine has already been initiated at universities across UK and we will encourage PhD studentships and Fellowships for postdoctoral researchers and new NHS clinical scientist posts.

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

Dr James Polke, active head of diagnostics and next generation sequencing lead.
Prof Henry Houlden, Head of Neurogenetics Laboratory and Professor of Neurology and Prof Patrick Chinnery (Professor of Neurology) and experts in neurogenetics and diagnostics.

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 GeCIP with work with UK and international clinical genetics organisations to refine the strategy for feedback of genetics results flowing from research and clinical testing. This includes

public and patient engagement to develop the policies and methods for genetic feedback, minimising the risks of feedback and promoting the public understanding of the role of genes in disease.

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

Clinicians and patients - Families; diagnosis, predictive, prenatal testing. Other NHS organisations and cost savings.

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?*

Potential new gene test development may be patentable and involving companies may be helpful to fund further work

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

Neurogenetics Laboratory: Email: ucl-tr.NHNNgenetics@nhs.net and neurogeneticslab@uclh.nhs.uk