Proposal: Strategic award application to establish an International Centre for Genomic Research in Neuromuscular Diseases

1. Summary of Mission, Vision and Objectives

Our Mission ***is to create a transcontinental genomics research and capacity building partnership between the UK and Official Development Assistance Lower and Middle Income Countries (ODA-LMICs) with an initial focus on India, Brazil, South Africa, Zambia & Turkey. We will discover new disease genes, understand comparative genetic architecture and explore disease mechanisms. We will increase the number of patients with an accurate genetic diagnosis, build trial ready cohorts and ultimately improve health outcomes for patients with neuromuscular diseases drawn from a population of over 1.5 billion.***

Our Vision ***a mature transcontinental academic partnership led by a group of outstanding clinical academics graduated from our training programme. They will be an enduring sustained legacy of expertise and will harness genomics to improve the lives and health outcomes of children and adults with neuromuscular diseases (NMDs) across the globe.***

Objectives

1. Establish an international fellowship training programme to generate a group of highly trained clinical academics who will be a legacy of leaders in personalised genomic medicine in the LMICs. These fellows will establish enduring mutually beneficial collaborative scientific links with the UK NMD clinical and genetic expertise based in Newcastle, Cambridge and UCL.
2. Establish a core international bioinformatics platform, pipeline and cloud-based clinical and genetic database. Openly share phenotypic and genetic data, and encourage other LMICs to join this collaborative programme.
3. Build ethnically diverse cohorts of children and adults with NMDs which we will deeply phenotype and comprehensively genotype.
4. Identify known and new disease genes, and assess comparative genetic architecture of NMDs across four continents. Use this knowledge to understand phenotypic variability, disease progression and disease mechanisms.
5. Increase the number of patients in LMICs with a precise genetic diagnosis enabling delivery of a personalised disease management plan, based on care guidelines to identify and manage complications using widely available and cheap interventions to improve health outcomes.
6. Enhance genetic diagnostics in UK-based Indian, Brazilian, Turkish & African NMD patients.
7. Build sustainability after five years, through retention of trained fellows in LMICs and through collaboration with host institutions, patient organisations, and local healthcare providers.

2. Importance of Neuromuscular Diseases

NMDs affect at least 17 million children and adults globally[1]. They cause either premature death, or are chronic diseases causing lifelong disability with economic impact. They include many different disorders affecting muscle and nerve function, and account for ~20% of all non-infectious neurological diseases. Examples include muscular dystrophies, congenital myopathies, neuropathies, motor neuron diseases, muscle channelopathies and mitochondrial diseases. The vast majority of NMDs are single gene disorders. Identifying genetic pathways and applying genetic testing has led to some of the most important advances in disease understanding alongside patient management plans and the development of new therapies[2-7]. Many of the key interventions involve the inexpensive practical applications of widely available medical technology (e.g. low-cost off-licence medication, targeted vaccination, cardiac monitoring and respiratory care), but their application is contingent on making a precise diagnosis[4,8]. For example, a precise genetic diagnosis can lead to a personalised and often simple management plan following best practice guidelines that includes basic screening for known complications (e.g. cardiac, respiratory, gastroenterological and metabolic), and often simple interventions that improve health outcomes - interventions that could be implemented easily in LMICs providing an accurate genetic diagnosis is made[8]. In the UK, a muscle biopsy has been the mainstay in the investigation algorithm in many patients, but this requires a specialist laboratory equipped for frozen section analysis with a growing panel of diagnostic antibodies. However, recent advances in genomics provide the opportunity to diagnose with high precision based on a DNA sample and clinical data collected remotely. *Our central objective is to build ethnically diverse cohorts of children and adults with NMDs and undertake genomic analysis to find known and identify new disease variants and*

genes. We will increase the number of patients with a precise genetic diagnosis to improve their care and increase knowledge on the comparative genetic architecture of NMDs across 4 continents. We will use this knowledge to understand phenotypic variability, disease progression and disease mechanisms in the world's largest NMD cohort of 15,000 probands. Importantly, we have access to several thousand ethnically matched control DNAs and will build this control resource further, complementing clinical and genetic data available through the Genomics England 100,000 genomes (GEL), NIHR BioResource and international collaborative resources.

Major benefits of this programme for partner LMICs:

1) Train a new generation of clinical academics with strong links to the UK NMD community
2) Develop cloud based capture systems for clinical and genetic data and bioinformatics support to discover new disease genes, and increase the number of patients with a precise diagnosis
3) Build trial ready cohorts in LMICs increasing the potential for clinical trial involvement
4) Potential for optimal patient management based on a precise genetic diagnosis

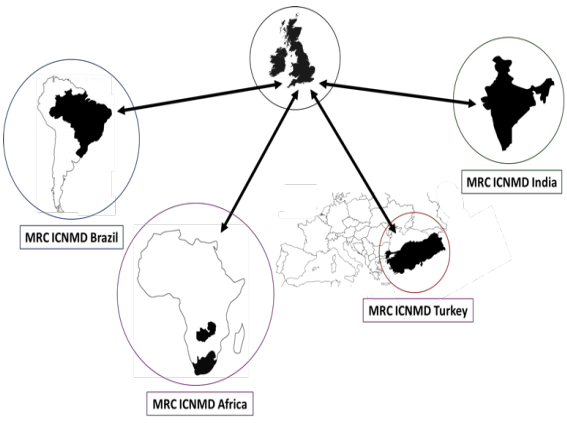
In addition, gene discoveries now provide an unprecedented scientific opportunity to use new advances in bioinformatics and disease modelling (such as induced pluripotent stem cells iPSC derived human cells) to deeply understand disease mechanisms, phenotypic variability, disease progression and therapy response. The opportunities for back translation from patient based gene discovery to understand mechanisms are unparalleled. Furthermore, dramatic advances in antisense genetic therapies are now showing efficacy in patient trials (some led by our PIs)[2,3]; and when coupled with rapid development in the viral vector preclinical gene therapy field and gene editing techniques, mean that NMDs are poised to fully harness genetic discoveries and make significant advances towards therapies.[2,3,5,6] Although these 'high-tech' solutions will not be first line therapy in the UK or LMICs for some time, directly involving LMICs in their development will enhance opportunities for global patient benefit during treatment development and by identifying a larger potential market for commercialisation, likely reduce overall, per-patient costs.

3. People and track record: why now and why us?

Over the last ten years our MRC Centre teams at UCL (Profs Hanna, Reilly, Greensmith, Houlden, Voit & Muntoni) and Newcastle University (Profs Straub, Lochmüller, Bushby, Horvath, Chinnery, Taylor, McFarland & Turnbull) have worked together effectively and gained major experience in building large, deeply phenotyped UK patient cohorts (>10,000), in building a national patient muscle cell biobank (10,000 lines), in gene discovery, in training clinical fellows, in "trial readiness", in delivering NMD experimental medicine and in patient care (see: www.cnmd.ac.uk review 8 year report). We discovered 41 disease genes, and we have increasing genomic data on UK patients with over 1,000 families submitted to the Genomics England 100,000 genomes project (GEL) and the NIHR Translational BioResource. This will provide a valuable substrate for our genomic database. We lead multiple investigator-led trials with full ownership of data generated, have extensive trials "know-how" with extensive industry network/partnerships[2-5](see: www.cnmd.ac.uk). We have undertaken over 100 paediatric and adult NMD natural history studies and interventional clinical trials (www.cnmd.ac.uk). Our work has led to recent European Medicines Agency and FDA licensing of drugs. The partnership working of our group linked to strong governance has underpinned our ability to connect, include and coordinate the national NMD community in full partnership with patient advocacy groups. Building on this track record, our ambition is to extend these advances globally with LMIC partners. **MRC strategic award support is an essential step to allow us to work as a team to maximise our synergies and address this ambitious but important challenge. This approach could serve as model for LMIC rare disease research more widely.** MRC strategic support will allow us to create a unique international resource and is an important opportunity for the UK to lead internationally. Expanding the NMD expertise of the London and Newcastle Centres to include Cambridge (Prof Chinnery) provides additional scientific and genetic "know-how" in complementary areas (e.g. Cambridge MRC Mitochondrial Biology Unit, MRC Biostatistics Unit, and the NIHR Translational BioResource). **Our International LMIC Partners**: We developed partnerships with five LMI countries across four continents; India, Brazil, South Africa, Zambia and Turkey. We held a two-day workshop in London in October 2017 to analyse opportunities/capabilities and develop a joint strategy (workshop appendix). The workshop was attended by all UK PIs and **all** five partner LMIC lead PIs. This enabled us to jointly undertake a detailed analysis of current clinical, genetic and training capabilities of each LMIC partner, and to assess the current and potential cohort sizes. Four consistent themes/gaps were apparent across all LMICS which this strategic award will specifically address; 1) the need to train clinical

academics who would pursue their careers in the LMIC, 2) LMIC centres serve extremely large populations of patients (at least 250 million) and there are large existing NMD clinical cohorts in India, South Africa, Turkey and Brazil (total ~15,000), but they need to be deeply phenotyped and genotyped, 3) patient enrolment into cohorts can be undertaken in LMIC NMD clinics provided there is manpower i.e. trained fellows, and 4) in LMICs where genetic analysis has commenced there is an important need for bioinformatics support and access to collaborations for functional studies. We aim to address these themes and maximise the opportunity to build the world's largest ethnically diverse cohorts through a focused strategy across LMIC partner sites (see letters of support for details and resource commitments of LMIC partners). The Brazilian centre is led by Prof Marques who previously completed his PhD with us in the UK, thus providing 'proof of principle' that our proposed model will be effective. The centres in India, Brazil, South Africa and Turkey have developed cohorts linked to NMD clinics, and have some access to local genetic analyses (PI, Prof Horvath, already has a Newton grant with our PI in Turkey (Oktay) on recessive developmental neurological diseases, and cohort numbers have rapidly increased showing we can deliver). The selection of our LMIC partners is based on verified large populations of NMD patients assessed at national clinical academic centres with large referral networks and the presence of a track record in genetics research albeit not fully connected with the clinical cohorts. In Zambia, we recognise a different challenge and opportunity where there is a single teaching hospital for the entire country in Lusaka. The NMD clinical service has recently been established our PI Dr Bearden who is now based in Zambia full time. He is a USA-trained Professor in Clinical Neurology and Neurogenetics with a co-appointment in Rochester Neurology Centre, New York. He is working with Profs Berch Griggs and Gretchen Birbeck (Zambia & Rochester) to develop neurology research in Zambia. We recently undertook a successful pilot study with Zambia confirming that we can transfer blood and extract good quality DNA for genetic analysis. Having a senior clinician scientist on the ground and proving we can extract DNA led us to conclude that partnering with Zambia has huge potential to access a very large population in one of the ODA lowest income category countries. We believe the range of centres we have selected offers major opportunities for cohort building, scientific advances, for local patient health benefits and for capacity building fellowships. By building a cloud-based deep phenotyping capture database and a bioinformatics platform that will enable data analysis drawn from four continental populations, we will have a significant opportunity to deeply understand the genetic architecture of NMDs. We will work with the Lead PIs in each LMIC and their senior clinical academic colleagues to train and develop fellows, which will advance clinical academic service development locally. For each partner LMIC, we agreed lead senior investigators who are all Co-PIs on this application. Each LMIC Co-PI has identified a senior team with significant expertise training fellows alongside expertise in preclinical science, genetics and clinical NMDs. CVs of the LMIC lead PIs with institutional letters of support see appendix. LMIC site lead senior investigators: **India: Prof K Thangaraj**, Head of Genetics, Centre for Cellular and Molecular Biology, Hyderabad. **Brazil: Prof W. Marques**, Head of Neuromuscular Disease Centre, São Paolo University. **South Africa: Prof FH van der Westhuizen**, Head, Centre for Human Metabolomics, North-Western University. **Zambia: Prof D. Bearden,** Neurology Dept, University Teaching Hospital, Lusaka, Zambia. **Turkey: Prof H. Topaloglu,** Head of NMD, Hacettepe University School of Medicine, Ankara & **Prof Y. Oktay,** Head of Genomics and Bioinformatics Services, International Biomedicine Genome Institute, Izmir (workshop photograph below; see appendix workshop report).

A 2 day international PI strategic award planning workshop was held in the MRC Centre for Neuromuscular Diseases at the UCL Institute of Neurology, October 2017. PIs from all international partner sites attended. Main photo, l-r: Rita Horvath, Rob Pitceathly, Henry Houlden, Berch Griggs (USA), Thomas Voit, Volker Straub, Wilson Marques Jr (Brazil), K Thangaraj (India), Yavuz Oktay (Turkey), Mike Hanna, David Bearden (Zambia), Francois van der Westhuiven (South Africa), Mary Reilly, Francesco Muntoni, Patrick Chinnery & Hanns Lochmüller. This workshop included detailed analysis of local clinical and genetic capabilities in each partner site, a detailed session on devising and sustaining the training and fellowship programme, and technical tours of UCL genetic labs and infrastructure.



Official Development Assistance (ODA) Compliance: We will address an important welfare related need in all partner countries; the significant disease burden of NMDs which are mainly genetic affecting ~17m people globally. These are either fatal or severely disabling disorders. In developed countries, recent genomic medicine advances have resulted in a significant increase in the proportion of patients with a confirmed genetic diagnosis. A genetic diagnosis has important clinical management consequences that directly benefit patients and families. Benefits include genetic counselling, prenatal/preimplantation genetic diagnosis, genotype-directed improved standards of care, and a reduction in avoidable predictable disease complications (e.g. cardiac, respiratory, gastrointestinal/nutritional and metabolic). In addition, for a small but increasing number of NMDs, there are now licensed drugs which are effective and improve quality of life and health outcomes; for example, in neuromuscular channelopathies (e.g. periodic paralysis, myotonias and congenital myasthenias)[4]. Furthermore, the first genetic therapy in Duchenne dystrophy has just been FDA approved[2,3]. Through establishing genetic diagnoses and finding new causative disease genes this initiative can have major impact by bringing these health benefits to patients in LMICs. Crucially, our fellowship and training programme will significantly upskill fellows and all participants in genomic medicine, delivering a lasting legacy to patients and LMIC health care systems and economies; including access to global registries and "trial ready" cohorts.

Key features of new international partnership with major impact potential

- MRC strategic award support is critical to catalyse this programme to generate a genetic discovery cohort of 15,000 patients and train a new generation of LMIC clinical academics
- Significant matched funding from host institutions £2,629,304 FEC
- Major added value from our connections with existing NIHR and MRC infrastructure
- New UCL, Newcastle, Cambridge alliance, places the UK as a leading international centre, with collectively one of the world's highest citations in NMD science

4. Scientific Plans

Overview: we will generate large cohorts of deeply phenotyped and genetically defined patients which will be an invaluable future research resource. These cohorts will be "trial-ready" for future local and international trials, which is a critical prerequisite for rare disease research. We will use our detailed "know-how" of the practicalities of building deeply phenotyped cohorts, including reliable data capture, acquired over 10 years as an MRC Centre to train a new generation of LMIC fellows. Approximately 20% of the ~20,000 genes that make up the human genome are expressed in muscle and nerve. We will discover new disease genes and variants, evaluate the genetic architecture of ethnic variation in clinical phenotypes, understand gene function and disease mechanisms and benefit patients LMICs and in the UK, providing a template for future expansion.

High level approach to achieve scientific objectives:

(i) Establish a common, easy to use platform for deep clinical phenotyping and sharing of genome data/results between all members of this project. We will adopt the pseudo-anonymised, secure and cloud-based RD-Connect platform (http://rd-connect.eu/), that uses PhenoTips (https://phenotips.org/) and Human Phenotype Ontology (HPO) (http://human-phenotype-ontology.github.io/) terminology. This freely available software platform will be used for each patient/family, entered on secure iPads in LMICs. This platform will allow future collaboration with European and international rare disease initiatives. We will also adopt OpenClinica to follow natural history of disease, and later therapeutic trials in LMICs. This approach has the added value of ensuring cross-compatibility with national (GEL, NIHR Translational BioResource) and international efforts (RD-Connect) in neuromuscular genomics.

(ii) Establish a common laboratory framework for high quality DNA extraction from blood and other samples such as saliva, muscle; logged and stored in the cloud database and following the guidelines of the UK accredited diagnostic laboratories.

(iii) Establish a common genetic screening platform across laboratories, capitalising on the local infrastructure, and based on a tiering system hierarchy that we developed and tested in the UK, accounting for the molecular test complexity, number of genes and analyses requirements.

(iv) Training in core bioinformatics skills, and where required, genetic laboratory techniques to ensure a high standard and an equal skill-set in each LMIC for clinical academics in training.

(v) Integrated genomic-phenotype cloud-based platform for effective interpretation and gene discovery.

(vi) Looking ahead, collaboration and sharing of all data and protocols will be important. Not only for the genetically-defined patient series, but also for genotypes and phenotypes of genetically negative cases where sharing with other groups such as European and International rare disease initiatives will only enhance results and lead to greater LMIC benefit.

A. Build ethnically diverse cohorts which we will deeply phenotype and genotype.

Ethical and relevant research approvals will be completed **prior to the award starting,** and all patients will be enrolled with informed consent. Deep phenotypic data and results of clinical investigations will be collected at source primarily by the clinical fellows following a period of dedicated UK training, and with support from the local research team each LMIC PI has assembled (see appendix host support). These data will be derived through targeted clinical examination to define Human Phenotype Ontology (HPO) terms specific for NMDs. We worked with Peter Robinson and the HPO team (HPO, P: JAX, Charité), to provide further granularity on specific NMD terms through our involvement in NeurOmics, RD-Connect, Solve-RD and the NIHR Rare Disease Translational Research Collaboration (2012-17).[9] HPO data will be entered for each family member through PhenoTips, which is a free open-source web-based application that allows standardised phenotyping using the HPO terms. Pseudo-anonymised data and clinical results will be stored locally as part of routine practice and also imputed directly on a secure laptop linked to PhenoTips onto the cloud-based RD-Connect platform. PhenoTips also includes a powerful pedigree editor, measurements and growth curves, fields to enter ORDO (ORPHAcodes) and OMIM numbers and the possibility to attach files such as external reports and images. PhenoTips can be used to record relevant information such as participant code, type of informed consent, gender, ethnicity, consanguinity, inheritance model, age of onset, evolution of the disease, and results from diagnostic and experimental or laboratory functional tests. We can also upload investigations undertaken as part of clinical practice, such as blood testing, electromyography/nerve conduction studies, imaging and videos of each patient. The online database will be encrypted allowing linked pseudo-anonymisation of the genetic/clinical data to the source files at the LMIC. Blood (and when possible other patient material) for DNA extraction will be taken from each affected proband, other affected and unaffected family members and logged onto PhenoTips. Modelling data from unsolved cases as an ontology will allow for having a complementary approach to match-making strategies. Indeed, match-making approaches allow discovery and pooling of similar unsolved cases characterised by their phenotypes and presumed pathogenic variants, with the aim to validate the causality for a known disease, and to discover new diseases. The UK cohort we previously built now exceeds 10,000 patients, principally (but not exclusively) of Northern European ancestry. This approach has allowed the separation of NMDs into more homogeneous phenotypic groups, creating a system to improve patient management. It has allowed us to identify new disease genes, make genotype-phenotype correlations, and correlate blood and imaging based biomarkers enabling the stratification of patients into 'trial ready' groups. For example, the successful trial of Mexiletine we led was only possible because we defined genetically stratified muscle channelopathy patients.[4] This provides class I evidence, and is now standard treatment using a cheap repurposed drug which could be available globally.

B. Gene discovery and genetic diagnosis to ultimately improve patient management

Identifying genetic pathways and applying genetic testing has led to some of the most important advances in disease understanding, development of therapies and patient management.[2-7,9-13] We will undertake genomic analysis in LMIC cohorts, capturing genetic variation in healthy and affected individuals, to find known and new disease genes and assess comparative genetic architecture across four continents. We will investigate phenotypic variability, disease progression and disease mechanisms in the world's largest NMD diversity cohort; target of 15,000 probands.

(i) Training the next generation of NM translational researchers in LMICs

We developed a successful PhD training program at Newcastle and UCL and we will extend this to Cambridge, integrating our fellows with the local MRC-funded Doctoral Training Programmes where available. LMIC clinical fellows will be incorporated into this programme and exposed to specialist NMD clinical practice with UK trainees (Professor Mary Reilly lead) enabling key phenotype training to ensure a standardised approach to clinical data collection. Training in bioinformatics, genetic techniques and understanding methods to functionally assess gene defects will be provided, although functional work will be beyond the scope of fellows in their time period. Each UK centre has an ISO accredited genetics laboratory providing diagnostic level interpretation

of genetic variants, and through GEL, we (Houlden, Chinnery) lead the Genomics Clinical Interpretation Partnership for neurological and neuromuscular diseases.

(ii) Establish genetic laboratories with a common platform in partner LMICs

The initial goal will be to ensure high quality DNA extraction principally from blood and genetic/genomic analysis within the existing laboratory facilities in each LMIC, and to share basic logistic processes (e.g. sample reception and logging, processing, databasing of results) as in our UK (ISO) accredited laboratories. This is based on our long track record in neurogenetics and recent knowledge gained through our direct involvement in the Genomics England 100,000 genomes project (Profs Houlden and Chinnery lead the Neurology Genomics England Clinical Interpretation Partnership (GeCIP)). The LMICs have all agreed to finance DNA extraction (or sample transmission to UK for Zambia until their team trained) locally as a host contribution to the programme. Currently genomic infrastructure in our partner LMICs varies, with Zambia being the least developed. To address this concern, we have undertaken a pilot study with Zambia and shown they are able to process, store and send biosamples to the UK where we have successfully extracted good quality DNA for whole genome sequencing. The other LMICs do have basic next generation sequencing technology, but little or no bioinformatic support. The development of these platforms, and training of staff in LMICs will form the foundation for consistent downstream genetic analysis. We will ensure that all are operating to the same common standard for blood and DNA protocols within 18 months of the start of the grant–by having LMIC lab staff attend host workshops and webinars prior to the award starting and harmonising with, and adopting the UK diagnostic labs' SOPs. All data will be owned by the host LMIC, but analysed as part of the collaboration.

(iii) Tiered approach to genetic analysis will be adopted (fig 1 and data management plan)

Tier 1: There is an important need to rapidly identify known genetic causes of NMD in LMICs using well-established technologies to avoid disease complications and manage symptoms, enable prevention approaches through counselling and family planning, and improve health outcomes. In the first instance we plan to use the Illumina Global Screening Array (GSA), that has multi-ethnic SNP content, and NMD gene mutation and copy number focused custom content will be added on. These arrays are comprehensive for known genes, cost effective, currently at ~£25 per sample (including running costs) and the addition of custom NMD disease focused content/SNPs at a cost of £4 per 3,000 SNPs per sample to include mutations such as common mitochondrial DNA mutations at nucleotide positions m3243, m8344, m8993, and identify deletions/duplications such as in HNPP/CMT1a and many other genes. The use of screening arrays offers a straightforward streamlined approach that can be adapted to LMICs. As DNA cannot be sent out of India, samples will be run in Hyderabad, while the other countries will initially have analysis done through our centre pipeline. The data handling is also easy to manage and store though a cloud-based system, and we have an effective bioinformatics analysis pipeline which is readily accessible to LMICs. All mutations identified will be confirmed with Sanger sequencing. We plan to initially select the most promising 7,000 samples for screening arrays, 2,000 from Brazil and India each, and 1,000 in South Africa, Zambia and Turkey, enabling rapid diagnosis of patients with known neuromuscular gene defects. In addition, we plan to run 200 elderly controls from each LMIC. Depending on the type of NMD, our pilot analysis predicts a ~25-40% pickup rate from the GSA arrays in these previously untested samples, with direct benefits to the genetic and clinical landscape, and will define the group for Tier 2. Some phenotypes will require specific laboratory screening, such as myotonic dystrophy and FSHD, and these will be carried out by the neuromuscular technician, working in the neurogenetics diagnostic lab, initially at the UK host.

Tier 2A: To overcome genetic heterogeneity in the LMICs, in the most likely genetic NMD cases, negative on GSA array, we will select 2,500 cases (500 from each LMIC) for Agilent focused exome sequencing. We have significant experience with the focused exome that is optimised for high and complete coverage (99% at >30-fold coverage) and contains over 6,000 genes including all NMD genes, genes in HGMD, OMIM, disease pathway genes and on ClinVar. The library enrichment and sequencing are performed in our UK host laboratories, and these techniques are transferable to LMICs with the exception of Zambia who will run samples in the UK. All Indian samples will be run in Hyderabad. We will also run focused exome on 200 elderly controls from each LMIC; the cost of each focused exome is around £115 per sample.

Tier 2B: In patients with a clinical indication of mitochondrial disease, we will first identify the common mutations on the screening array, and then deep-sequence the mitochondrial genome using a next generation sequencing approach and established bioinformatics which can detect

single nucleotide variants and deletions down to 1% heteroplasmy levels. Ideally this will be done on frozen muscle tissue, but where this is not possible we will use DNA from either urinary epithelial cells or blood. Sequence coverage of mitochondrial genome sequencing is very high at >4000x per sample to achieve effective heteroplasmy detection. We will sequence the mitochondrial genome in around 500 patients and 500 controls (100 controls from each LMIC).

Tier 3: Whole genome sequencing (WGS) is widely applicable to Mendelian disorders such as NMDs, and is particularly suited to identification of rare or private genetic mutations, including copy-number variants. It will increasingly become the gold-standard approach for the investigation of inherited disorders, but there are challenges that include expensive equipment, sophisticated computer infrastructure and data storage, although these are all being rapidly overcome in the UK by the work of GEL and the NIHR Translational BioResource. We will take full advantage of our links with GEL, GEL expertise, and the wealth of different population control data that this project holds. Selected negative samples after array and focused exome will undergo WGS. We are not seeking any funding in this application for WGS. Of the total of £350,000 UCL host support for consumables, £120,000 will be allocated for WGS, we will work with GEL to provide WGS analysis (current cost ~£616 per sample). We will apply for additional WGS funds e.g. H3Africa & Newton Fund initiatives.

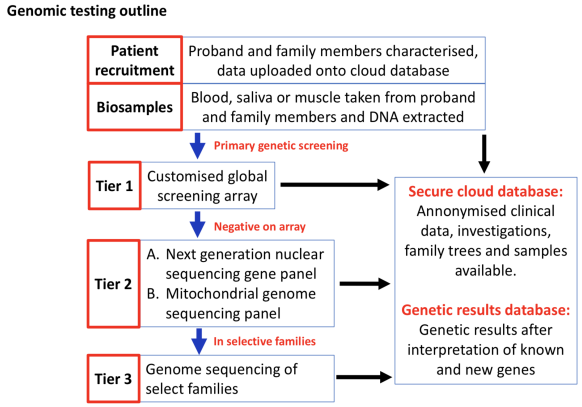


Figure 1: Sample flow and testing strategy; patients will be clinically characterised and data uploaded to the secure cloud database. DNA extracted, and Tier 1 customised global screening array will be run on each proband. Selected negative families undergo next generation focused exome or mitochondrial sequencing panels. In selected families, we will do genome sequencing (funded by host support, not MRC), important in solving undefined LMIC families.

C. Establish a core international bioinformatics pipeline, platform and linked database:
To gain the most translational benefit, genomic data should be interpreted in the light of the global genomic background and combined with clinical data. In addition to probands, relatives will be collected for segregation analyses and we will collect 1,000 new healthy elderly controls from LMICs (200 each LMIC). We will access control genomic data from the ExAC, gnomAD and H3Africa resources. Furthermore, we have agreements with GEL, RD-Connect and GENESIS (Zuchner-USA) to access control data (see letters of collaboration appendix). We will increase the control data in the relevant population as part of our sequencing strategy (i.e. the exome will generate control data from the non-relevant genes) and in addition we will run 1,000 arrays and 1,000 focused exomes (200 from each LMIC) as a direct control base. The bioinformatics core staff will be in the UK, and will support LMICs to establish the hardware and software infrastructure, as well as the collaboration with other local and distant research groups to share and enhance data. In the UK we have done extensive work on the development of SNP screening arrays, next generation gene panels and WGS as a technique that will deliver virtually all genetic abnormalities and be future-proof, but where the core need is bioinformatics combined with clinical interpretation.

The bioinformatics pipeline: In the analysis of screening arrays and next generation sequencing panels, each bioinformatician will focus on two core neuromuscular disease areas; Bioinformatician 1 (based at Newcastle University) will focus on muscular dystrophies and mitochondrial genome disorders, Bioinformatician 2 (based at UCL) will focus on inherited neuropathies and nuclear mitochondrial disorders, Bioinformatician 3 (based at Cambridge University) will focus on nuclear and mitochondrial genome disorders and Bioinformatician 4 (based at UCL) will focus on other rare neuromuscular disorders such congenital myopathies/dystrophies, channelopathies, myasthenia gravis, and familial ALS. The bioinformatics team will be led by Houlden and Chinnery, where the focus will be on comprehensive analysis and the development of these techniques in LMICs, working closely with fellows and their clinical teams. **The Infinium Global Screening Array (GSA)** v1.0 BeadChip combines multi-ethnic genome-wide content, curated clinical research variants, and quality control (QC) markers for precision medicine research. As part of the Erasmus MC consortium (http://www.glimdna.org/), we have access to the array at reduced cost with shared

protocols and control data. Although these arrays were designed based on European populations, Illumina have added significant multi-ethnic non-European content. The array was designed for global population coverage and a 640,000, genome wide background and 50,000 SNPs are in the exome and focused on the disease coverage. We also plan to add 3,000 custom SNPs focused on known NMD disease gene variants. The method of analysis is straightforward and involves DNA extraction, DNA labelling, probe hybridisation to patient DNA and SNP scanning. This will allow the identification of: 1) Known NMD gene mutations and genetic modifiers of disease, 2) Genetic deletions, insertions and duplications, 3) Pleotropic and extended genotype phenotype association, 4) Known pharmacogenomics modifiers such as SNPs associated with drug metabolism. In addition, we will work with the European Bioinformatics Institute (EBI) to continually improve analyses techniques as well as control resources, as discussed above. **Focused exome and genome sequencing**: here the sample, library and sequencing are straightforward. Bioinformatic analysis will involve a pipeline of processes from data filtering through to read alignment, SNP calling and annotation primarily using GATK. As discussed, the focused exome has high and complete coverage of over 6,000 clinically relevant genes and WGS covers the entire nuclear and mitochondrial genome to identify such defects as: 1) Known disease-causing and new mutations, small indels. 2) Mitochondrial genome variants and deletions at high coverage, 3) de-novo mutations as relatives will be collected, 4) Whether somatic mosaicism explains single cases. 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, applying the American College of Medical Genetics and MacArthur criteria. We will use state of the art Bayesian methods for rare pathogenic variant detection in collaboration with the NIHR Translational BioResource and MRC Biostatistics unit, and aim to take advantage of the GEL 'reading library' of >20,000 WGS where 1,500 are from NMD families to inform our analysis. We will use the control resources as discussed, also see Annexe.

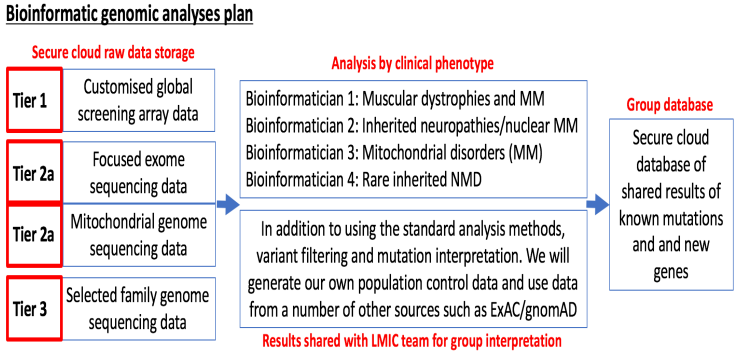


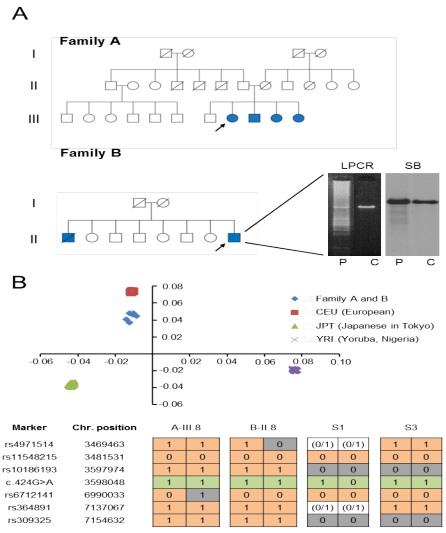
Figure 2: Bioinformatic analysis of data. Initially, Tier 1 customised global screening array data is run on each proband. Selected negative families are then run on the focused exome next, or on the mitochondrial sequencing panel; finally in selected families we will carry out genome sequencing (host funded). Rare inherited neuromuscular diseases (NMD), include familial ALS, channels, early onset myasthenia and congenital dystrophy/myopathies.

Data management plan and diagnostic genetics:
we are deeply committed to data sharing for the benefit of the global research community. Overall, demographic, phenotype, investigative and GSA array data will be stored using a secure cloud, with geographically local cloud infrastructure enabling cross-continent analysis together using the same pipeline. This format will aid training, deliver transparency, enable centralised harmonisation in a secure, fast, and cheap way. The WGS data will be stored and analysed in partnership with GEL, the NIHR Translational BioResource, and UK10K, and benefitting established databases such as ExAC and gnomAD. There will be clear, open access analysis for all participating LMIC clinicians and researchers. All research results will be shared with LMICs and diagnostically confirmed according to local clinical genetics guidelines. We have extensive experience in NHS diagnostic neurogenetics with our PIs leading UK diagnostic labs in UCL (Houlden) and Newcastle (Taylor) as well as our PI leadership of three NHS England nationally commissioned highly specialised NMD services (Muntoni, Hanna, Turnbull, Straub). Furthermore, our LMIC PIs in India, Turkey, South Africa and Brazil have leadership roles in diagnostic genetics labs. The extensive clinical and genomic data produced will have general benefit (genetic pipeline), importance for Health Data Research UK (HDR UK) and benefits for healthcare delivery.

D. Mutation pathogenicity, segregation and control samples (also see the Annexe, page 13)
We will use our standardised analysis and set of filters on the array and sequencing data to identify disease causing mutations with the same stringency that we use in our UK genetic testing network (GTN) of accredited diagnostic laboratories (https://ukgtn.nhs.uk/; https://www.eshg.org/;.[14] In order to understand the genetic causes of diseases in diverse populations, background control genotype

of countries and subgroup/language groups is essential. In each LMIC, the proband and other affected and unaffected family members, and unrelated elderly controls will also be examined and blood taken for DNA extraction. We have existing control DNAs in each LMIC which include >24,000 in India, >41,000 in South Africa, >9,000 in Turkey, >2000 in Brazil and control collection has just commenced in Zambia. These controls and family members will be used for segregation and also to screen specific mutations if required. We will run arrays and focused exomes on 1,000 LMIC controls we will collect as part of this programme and we also have access to exome or genome sequenced samples through collaborative population initiatives such as The Exome Aggregation Consortium (*ExAC*) (http://exac.broadinstitute.org/) and The Genome Aggregation Database (*gnomAD*) (http://gnomad.broadinstitute.org/) of exome and genome sequenced cases, as well as access to GEL, RD-CONNECT and GENESIS-(Zuchner USA). Together these databases have >120,000 exomes and genomes, from population regions (as opposed to countries), and from patients with non-Mendelian diseases, usually cardiovascular, lung or psychiatric diseases. Working with Rare Disease Connect (RD–Connect) (http://rd-connect.eu/), The H3Africa biobank (http://h3africa.org/) and The International Rare Diseases Research Consortium (IRDiRC) (http://www.irdirc.org/irdirc-related-calls/) we will have access to more specific, sequenced country, and ethnic or language subgroups of sequenced controls.

E. Increasing the number of diagnosed patients and identifying therapies from the discovery of new disease genes and pathways: In the UK, accurate genetic diagnosis and the formation of genetically defined cohorts have led to a step change in the management of NMDs. Each cohort is managed in a specialised clinic in a streamlined way to reduce complications, enrol into patient trials, genetically counsel families and improve health outcomes.[8]



Understanding gene function and downstream analysis: We are not seeking funding for the downstream functional analysis of genetic defects. Our PIs have major separately funded programmes of research that will add value and enable detailed functional analysis of new genes including a wide range of approaches including gene function/ expression modelling human iPSC lines/CRISPR/CAS9 system, fruit fly, zebrafish and animal model systems. We built an MRC biobank with 10,000 NMD patient muscle/fibroblast lines. We have well-validated in-vitro model systems 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. A recent example of the potential power of connecting with India was the discovery of *RNASEH1* mutations in two UK South Asian families **(Figure 3A, above left)**. Ribonuclease H1, encoded by *RNASEH1*, degrades RNA-DNA hybrids. We had already implicated the enzyme in the segregation of mitochondrial DNA, suggesting a new disease mechanism.[10] We linked with our partner Prof K Thangaraj and analysed 50 South Asian families with identical clinical phenotypes. We confirmed UK families were in the same ethnic group (**Figure 3B**, principal component analysis, top panel), but that the mutations had arisen independently (**Figure 3B**, haplotype analysis, bottom panel). Close links with Hyderabad were crucial to this discovery.[10] **Identifying therapies from the discovery of new disease genes and pathways:** Two recent examples where our gene discovery has linked to potential therapies; 1) we identified mutations in 2 subunits (*SPTLC1* and *SPTLC2*) of the enzyme serine palmitoyl transferase as the cause of a form of neuropathy. We identified a new disease mechanism i.e. neurotoxicity from deoxysphingolipids (DSBs) induced by the mutant protein, and a functional assay (DSBs plasma levels) was developed to validate pathogenicity of new *SPTCL1/2* variants, and we are now testing simple new therapies (serine) to reduce DSB levels.[12] 2) We discovered two riboflavin transporter genes that cause childhood motor neuron disease, we modeled genes in fruit flies and showed that supplementation with riboflavin leads to significant, often life-saving benefit to patients.[13]

5. Training Plans: LMIC Capacity Building Clinical Fellowship

Highly trained adult and paediatric NMD clinical academics are an essential element of an effective genomics medicine programme. The PIs have extensive experience in training clinical academics,

and in the last 10 years successfully trained 39 UK-based clinical fellows who have graduated PhD or MD. It was clear from the two-day workshop held with LMIC lead PIs that there is a significant shortage of programmes to generate such individuals in partner LMICs, but no shortage of enthusiastic capable trainees. Fellowships will be advertised competitively and drawn from clinical adult and paediatric neurologists who have completed or close to completing clinical training. All LMIC PIs confirmed such fellowships would be extremely popular, and is a clear current gap. The specific key **skills gaps** are in training in genomics, bioinformatics and in being able to apply this in the clinical context to enable personalised genomic based medicine. It was also apparent from the workshop that fellows trained with these skills would become an extremely valuable link between the genetics laboratory teams that are evolving in each LMIC, and large clinical services that are developing NMD subspecialty clinics. Each LMIC lead PI has established a team of senior clinical and academic colleagues who have committed to this training programme (see support letters), and each host has committed to employing the fellows at the end of this programme (see support letters), minimising the risk of a 'brain drain'.

We will produce a group of highly trained clinical academics who will pursue independent careers in the LMICs with an established UK support and collaboration network. We will also produce a smaller group of UK clinical academics with expertise in global NMD genomic medicine who will train alongside/partner LMIC colleagues, engendering a long-term UK/LMIC collaboration.

Training: Key Strategic Aims

- Develop and train a group of talented, enthusiastic and highly motivated NMD clinical academics who will become LMIC-based leaders and spearhead NMD genomic medicine and improved health outcomes in each LMIC over the next 5-10 years.
- Provide genomic research PhD training and highly specialist clinical training and mentorship provided by some of the world's leading NMD clinical academic centres based in the UK.
- Support individual PhD/MD research and training in a topic on genomic NMD medicine.
- Train a group of UK clinical academics with specific expertise in global NMD genomic medicine to provide a sustainable long-term UK/LMIC collaboration.

Our unique characteristics and track record add significant value to student training.

We developed and delivered highly successful clinical and non-clinical doctoral training programmes for UK-based trainees (see: www.cnmd.ac.uk). We took full advantage of our links with major academic Institutions [Institute of Neurology-UCL, Institute of Child Health-UCL, Newcastle University] and associated leading NHS Trusts, where clinical academic applicants in this proposal lead services assessing over 20,000 children and adults with NMDs annually. This includes the only NHS England Highly Specialised Commissioned Services for NMDs in the UK (muscle channelopathies, mitochondrial diseases and muscular dystrophies). Cambridge, led by Professor Chinnery, will now contribute significant additional clinical and scientific value to the proposed Fellowship programme through the MRC Mitochondrial Biology Unit and his co-leadership of the NIHR Translational BioResource for Common and Rare Diseases. The co-location and critical mass of basic scientists, clinical scientists and patient populations provides the perfect environment and expertise to support trainees both during their time in the UK and in their home LMIC. The LMIC partner centres all serve very large patient populations and have NMD clinical services at varying stages of development. Four of the LMIC Centres have started genetic research and diagnostic programmes, but there are clear bottlenecks in bioinformatics and in achieving a genetic diagnosis except in small proportions of patients. There are no clinical academic training programmes on NMD genomic medicine. Indeed, Prof Marques who leads the Centre in São Paolo, Brazil, undertook his PhD training with Professors Reilly and Hanna. **Brief operational delivery plan**: We will establish a bespoke 4-year international clinical research training fellowship programme only open to LMIC clinicians, and a matching 3-year research training fellowship for UK clinicians. There will be dedicated 0.3WTE clinical tutor at UCL to coordinate the entire fellowship programme. We propose 8 WTE four year LMIC fellowships and 3.0 WTE three year UK fellowships funded by the MRC. This will be matched by an additional three 3-year UK fellowships from the UK host universities and one LMIC fellow funded by Guarantors of Brain; therefore, we propose a total of 15 fellows (9 LMIC and 6 UK: Table 1). **The "brain drain" risk and mitigation:** the possibility that the programme may result in LMIC fellows pursuing their career in the UK cannot be completely excluded, but we agreed a clear mitigating strategy with LMIC PIs at our workshop to minimise this risk; 1) selection on the basis of a declared intention to pursue their career in the LMIC, 2) LMIC host support to employ LMIC fellows after the

four year programme, by which time a strong UK collaboration network will be established to support the fellows' onward research in their home country, 3) in the UK LMIC fellows will have observer status and will not have full GMC registration. Table 1 Distribution of LMIC/UK CRFs:

Clinical RF	LMIC					UK		
MRC	India	S. Africa	Turkey	Zambia	Brazil	UCL	NCL	CAM
11	2	2	2	1	1	1	1	1
Host Funded 4					1	1	1	1

We designed this model based on our planning workshop in London in October 2017. In four of the LMICs (India, South Africa, Brazil and Turkey), there is a clear need for at least two fellows (based partly on the geography of the centres) and in order to provide the critical clinical/laboratory interaction in those countries. In Zambia, there is no NMD genetic programme and one clinical fellow will be appointed to develop and phenotype the clinic cohorts and will link with the UK for genetic analysis and bioinformatics. All LMIC fellows will have a six-month induction training in the UK; the first month will be at UCL for all fellows and will be followed by five months in one of the three UK centres. This will be followed by two years in the LMIC, then a further 6 months in their UK centre and a final year in the LMIC. The six UK fellows will start at the same time and do the same first month induction in UCL. This will be followed by 18 months in one of the three UK institutes, then up to 6 months in a LMIC and the final year in the UK (Table 2).

Table 2: LMIC/UK MRC Clinical Fellowship programme

	No of CRFs	**Yr 1** M1-6	**Yr 1** M6-12 (M1 all at UCL)	**Yr 2** M1-6	**Yr 2** M6-12	**Yr 3** M1-6	**Yr 3** M6-12	**Yr 4** M1-6	**Yr 4** M6-12	**Yr 5** M1-6	**Yr 5** M6-12
LMIC	4	Appoint	UCL	LMIC	LMIC	LMIC	LMIC	UCL	LMIC	LMIC	
	3	Appoint	NCL	LMIC	LMIC	LMIC	LMIC	NCL	LMIC	LMIC	
	2	Appoint	CAM	LMIC	LMIC	LMIC	LMIC	CAM	LMIC	LMIC	
UK	2	Appoint	UCL	UCL	UCL	LMIC	UCL	UCL			
	2	Appoint	NCL	NCL	NCL	LMIC	NCL	NCL			
	2	Appoint	CAM	CAM	CAM	LMIC	NCL	CAM			

M = Month; CRF = Clinical Research Fellow, UCL = ION/ICH, NCL = Newcastle, CAM = Cambridge.

Each UK fellow will partner with LMIC fellows working on a similar condition. This joint training of LMIC / UK fellows is a major aim of the programme to engender future long-term collaborations between the UK and LMIC trainees. The programme is designed to provide a comprehensive induction training across NMDs including detailed training in genetics and genomics, clinical phenotyping, genetic diagnosis and research methodology. This will incorporate selected relevant modules (in person and web-based) of UCL's successful MSc in Neuromuscular Diseases course during the first 6 months. The fellows will be highly selected, talented trainees who already have trained in neurology (for 2-4 years), but during the entire four and three-year programmes, we will ensure further training and exposure to clinical neuromuscular diseases, neurophysiology, neuromuscular imaging and neuromuscular pathology. Each trainee will undertake a higher degree PhD/MD registered in UK or in LMIC. A wide variety of MD and PhD research projects will be offered on many different aspects of genetic/genomic NMDs, reflecting the extensive expertise of our PIs. Disciplines for PhD projects will be clinical genetics, bioinformatics, genomic NMD medicine and translation into diagnostics, and there will be four broad disease themes for which expertise is available across the three UK centres (UCL-ION: neuropathy/neuronopathy, mitochondrial disease; ion channel disease, UCL-ICH: muscular dystrophy; NCL mitochondrial disease and muscular dystrophy; Cambridge: mitochondrial disease). They will also have the opportunity to form collaborative links with basic laboratory projects available in London, Newcastle and Cambridge, spanning disease mechanisms and including cellular and animal models, IPS cell technology and genetic therapy development. **Trainee selection and supervision**: undertaken by a joint UK / LMIC PI panel with the expertise to identify the most talented trainees from diverse training systems. Professor Reilly, our training lead, has very extensive experience in developing and running the UK clinical PhD training programmes, and also developed and directs an NIH-funded clinical training programme across the UK and US with fellows appointed from the UK, US, Italy and Brazil. All trainees will have two lead supervisors (one UK-based and one LMIC-based). To ensure the trainees are fully supported and mentored there will be a UCL-based clinical tutor,

Dr Pitceathly (hon consultant academic level). **Evaluation and delivery**: 6-monthly trainee reports and logs to monitor progress. Output measured will include completion of higher degrees, abstracts and meeting presentations and publications. Post-doctoral career planning will be rigorously undertaken with each LMIC lead investigator ensuring robust post fellowship career.

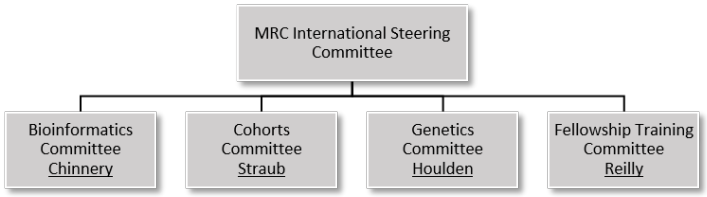
6. University and Other Commitments

UK hosts strongly support this proposal and provide £1,378,7609 in new directly incurred salary costs (FEC £2,279,304) for 3 bioinformaticians & 3 clinical fellows. UCL will also provide £350,000 consumables. LMIC hosts agreed to DNA extraction costs and each will provide one lab technician (except Zambia). Each LMIC formed senior clinical/genetic team to mentor fellows and committed to supporting fellows post-fellowship. Guarantors of Brain pledged £150k for 1 LMIC fellow. World Muscle Society; 3 x 1000 Euro LMIC scholarships per year for 5 years to attend annual congress.

7. Management

Prof Hanna will work with Co-PIs to provide strategic direction, and has a strong track record in his leadership of the MRC Centre across UCL & Newcastle. We employ a "flat" management structure enabling delivery of key governance and management functions including **oversight, strategy, operational, involvement, scientific and effective communications**. Prof Hanna will chair monthly international steering committees (Skype) with all UK & LMIC PIs. He will also convene monthly Skype meetings with each LMIC lead until systems and data flow are running. Subcommittees will ensure operational delivery in relation to cohort development, genomics, bioinformatics and training; accountable to steering committee according to "SMART" objectives. The full-time centre senior administrator (0.5 host funded) will be responsible for finance, communications, recruitment/HR, and will support the PI team to develop a clear PPI strategy. Scientific progress will be reviewed at the steering committee and by an external international science advisory board chaired by Professor Dame Kay Davies.

Indicative organogram of management structure:



8.Proposed outputs and outcomes

1. Build LMIC cohorts; our target is 15,000 LMIC patients over five years with minimum patient numbers (based on workshop scoping data): mitochondrial diseases ~4000, muscular dystrophies ~5000, inherited neuropathies ~7000, muscle channelopathies ~500, congenital myasthenic syndromes ~400, metabolic myopathies ~500, congenital myopathies ~600, motor neuron diseases ~700. 2. Genetic discovery: we discovered 41 new genes in last 5 years in UK cohorts; given rapid technological progress we anticipate ~100 new gene discoveries. We are in a strong position to add value, and will undertake at least 10 new gene functional studies in separately funded programmes. We will generate a valuable genetic data set allowing us to determine how genetic ethnic diversity determines disease course and phenotype. 3. Improved genetic diagnostic rates/health outcomes: Our PIs lead international efforts in defining standards/guidelines of care for patients with genetic diagnostic labs. We will increase the percentage of patients with a precise genetic diagnosis, who have had genetic counselling, who have had screening for recognised disease related complications, and who have started effective treatments (e.g. channelopathies). 4. Fellowship training: We will train 15 clinical fellows who will graduate PhD or MD.

References 1. Deenen JCW et al. *J. Neuromusc. Dis.* 2015;2;73-85. 2. Cirak S, *et al*. *Lancet* 2011; 378(9791):595-05. 3. Voit T, *et al*. *Lancet Neurol*; 2014;13:987-9. 4. Statland JM, *et al*. *JAMA*; 2012;308(13):1357-65. 5. Klopstock T, *et al. Brain* 2011;134:2677-2686. 6. Finkel RS, *et al.* N Engl J Med; 2017;2;377(18):1723-1732. 7. Mendell JR *et al.* N Engl J Med; 2017; 2; 377(18):1713-1722. 8. Bushby K, *et al*. *Lancet Neurol;* 2010; 9(2):177-89. 9. Köhler S, *et al. Nucleic Acids Research;* 2017 4; 45(D1):D865-D876 10. Akman G, *et al. Proc Natl Acad Sci U S A*; 2016;26;113(30), E4276-85 11. Bugiardini E, *et al. Neurol Genet*; 2017;2;3(3):e149.12. Penno A, *et al. J Biol Chem.* 2010; 285:11178-75. 13. Foley AR, *et al*. *Brain.* 2014;137(Pt 1);44-56 14. Keogh MJ, et al. *Genome Research* 2017;27(1):165-173.

Annexe: reproducibility and statistical design

1. Experimental approach to address objectives. Primary and secondary experimental outcomes: The primary objective is to identify known and new disease genes for NMDs across four continents, and improve patient diagnosis and management. Secondary outcomes will be increased understanding of phenotype variability, disease progression, gene function and disease mechanisms. **We will use a tiered approach for genetic analysis (see figure on sample flow below) to achieve the following outcomes: Tier 1** we will run 7,000 probands on the GSA screening array, designed with additional custom content to investigate known NMD genes, such as common mitochondrial DNA mutations at nucleotide positions 3243, 8344, 8993, CNVs, deletions/duplications such as in HNPP/CMT1a and many other genes. Depending on the type of NMD, our pilot analysis predicts a ~25-40% hit rate from the GSA arrays in these previously untested samples. In patients negative on the GSA array we will use **Tier 2A:** to select 2,500 cases (500 from each LMIC) for Agilent focused exome V2 sequencing. We have significant experience with the focused exome that is optimised for high and complete coverage and contains over 6,000 genes. We expect a pickup rate of around 35% of cases with the focused exome. **Tier 2B:** will be used in patients with clinical evidence of mitochondrial disease; using this next generation sequencing approach and established bioinformatics which can detect single nucleotide variants and large deletions down to 1% heteroplasmy levels, at very high >4000x coverage with a pickup rate of ~20%. **Tier 3:** In 200 probands, negative on array and focused exome, we will do WGS. This will be funded through host support and expected to yield ~25% of the cases analysed.

2. Data analysis, model used and controls Our bioinformatics pipeline uses a standardised approach to analyse the array data and filtering of sequencing variants, and this will enable us to identify disease causing mutations with high stringency https://ukgtn.nhs.uk/; https://www.eshg.org/. **In order to fully interpret the genetic causes of diseases in diverse populations, understanding the background control genotype is essential.** In each LMIC, the proband and other affected and unaffected family members, and unrelated elderly controls will also be examined and blood taken for DNA extraction. In addition, we will run 200 elderly controls from each LMIC (1,000 samples) on the GSA array and the focused exome. 500 control samples will have their mitochondrial genome sequenced. Our analysed LMIC controls will allow us to quickly filter variants with frequency over 1%. To complement this we also have access to large amounts of sequencing data (120,000 exomes/genomes-see earlier) through collaborative population initiatives such as *ExAC* and gnomAD resources. For example, the available African ancestry reference data from these resources, including data from the gnomAD consortium, currently have exome or genome data from over 12,000 individuals of African ancestry, and these numbers will increase over the course of this project. In addition, working with our collaborators in RD–Connect (http://rd-connect.eu/), The H3Africa biobank (http://h3africa.org/) and The International Rare Diseases Research Consortium (IRDiRC) (http://www.irdirc.org/irdirc-related-calls/) we will have future access to further more specific, and ethnic or language subgroups of sequenced controls. The controls and family members will also be used for mutation segregation and screening specific mutations if required.

3. Justification: The GSA array experimental approach is technically straightforward to establish in LMICs, has standard bioinformatics, and it is very cost-effective to analyse large numbers of samples. This will initially be carried out at Erasmus MC (except India) as they are the most cost-effective, with bioinformatic analysis from our team. The focused exome requires more advanced laboratory and bioinformatics but gives high coverage of the clinically important genes at a competitive cost. Whole genome sequencing is the gold standard for investigating genetic disease, but cost, equipment and bioinformatic analysis are challenging, although we plan to do a small number of probands, UK host-funded, to initiate this approach for the future LMIC projects.