Title: Gene-hunting in familial motor neuron disease (MND) using exon capture and high-throughput gene sequencing

Institution: MRC Centre for Neurodegeneration Research, Institute of Psychiatry, King's College London, SE5 8AF

Aims

1. To identify novel gene mutations and determine frequency in familial and sporadic MND.
2. To explore possible mechanisms of action of MND gene mutations in cellular models.

Background

At present the aetiology of the majority of MND cases is poorly understood and the strongest clues have come from genetics. Because MND is a late onset disorder with rapid disease progression, multiple DNA samples from affected members of the same family are rarely available for traditional genetic linkage studies. As a result gene discovery has been slow. Over the past decade we have built one of the largest and most successful laboratories researching the genetic and biological basis of the motor neuron disorders.

From 1993 to 2006 the only gene linked to "classical" MND was *SOD1*, which accounts for 20% of all familial cases. In the past three years we have described two further MND genes; TAR DNA binding protein (*TARDBP)* [1] and Fused in Sarcoma (*FUS)* [2]. Both genes encode for proteins involved in RNA processing and this has opened up a new chapter in MND research. In 2006 we identified a novel locus on chromosome 9p in a large Dutch Kindred with ALS-FTD [3]. In just the last month the mutation has been identified as a massive hexanucleotide (GGGGCC)n repeat expansion mutation within intron 1 of an uncharacterised gene named *C9orf72* [4,5]. We have confirmed this finding and can identify it in ~30% of familial and 7% of sporadic MND cases. Mutations in these four genes account for 50% of familial and 12% of sporadic MND cases, however the genes responsible for the remaining cases are unknown. New gene hunting strategies employing exon capture and high-throughput sequencing mean that we can dramatically accelerate the process of gene discovery.

Clinical and/or translational relevance of the research

Once a new gene for MND is discovered mutation screening can be established in accredited diagnostic genetic laboratories. Diagnostic testing can be made available for affected patients and predictive gene testing offered to at risk family members. Gene discovery will also advance the development and validation of cellular and animal models of disease, which will in turn accelerate the rate of drug screening. Each new gene will provide an additional piece of the molecular jig-saw puzzle that will enhance our understanding of disease mechanisms and advance drug discovery.

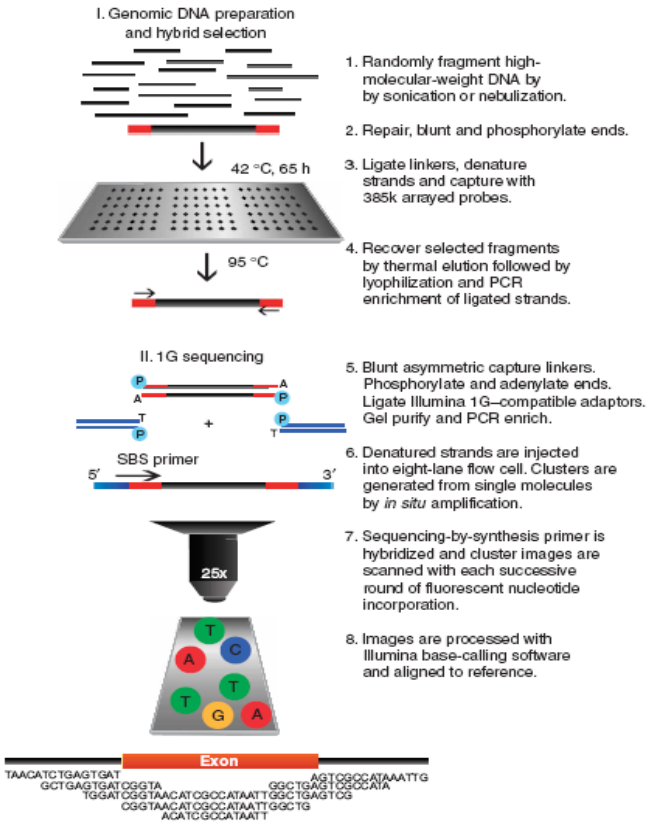
Experimental Plan

1. Case Ascertainment We have collected DNA samples from more than 200 familial and 1,200 sporadic MND cases in our own collection. In collaboration with groups in Sheffield and Birmingham we have collected a further 120 familial and 1,500 sporadic MND cases. With funding from the MNDA and Wellcome trust funded collection most of these cases have been banked as transformed lymphoblast cell lines. Through our International collaborators it is also possible to validate any findings in DNA from a cohort of ~800 familial and 5,000 sporadic MND cases.

2. Molecular Genetics Because the majority of neurodegenerative disorders are due to mutations in the 1% of genomic DNA that encodes proteins (exons) we intend to capture these exons using highly specific probes and sequence them using new high throughput technologies. We will begin by sequencing the exome of 200 familial MND cases.

Genomic DNA will be randomly fragmented by sonication to an average size of 300 base pairs and ligated with Illumina–compatible linkers. Fragmented DNA will then hybridized to exon tiling arrays, after which eluted material will be enriched by PCR and added to two lanes of an eight-chamber flow cell, and sequence clusters generated from single DNA molecules. For each base-incorporation cycle, an image will be read and a base called. The obtained sequence reads will be filtered for quality and mapped to the genome and aligned to reference sequences from NCBI and other databases. (Image from [6])

The 2.1m probe array (Nimblegen) targets 30Mb of genomic sequence and is capable of identifying 98% of the 180,000 target exons suitable for resequencing by the Illumina Hi-seq analyzer. This array will capture ~90% of all protein coding exons in which many pathogenic mutations may lie. DNA capture and deep resequencing has been used successfully by other groups to identify novel MND disease genes [7].



In a pilot study of 20 DNA samples we achieved on average 150 fold enrichment and have reads that pass stringent quality control for 98% of all exons. We use several Bioinformatic pipelines to detect insertions and deletions (PINDEL, DINDEL, BREAKDANCER) and will interrogate multiple SNP databases (dbSNP, 1,000 GENOME etc.) to identify those variants that are novel. We will filter these variants in each sample to prioritise those that are non-synonymous (i.e. are likely to change amino acids in the encoded protein). By sequentially combining the data from all 200 index cases we will be able to identify exons/genes with multiple novel variants and screen them *in silico* for pathogenicity using SIFT. Those that show the most promise will be validated in larger sample sets of familial and sporadic MND cohorts and are absent from a large cohort of controls (>1,000).

3. Cell Biology Having identified candidate variants we explore their pathogenicity by mutating tagged genes in mammalian expression vectors. We will transfect a range of neuronal cell lines (HEK293 and SH--SY5Y cells) and rodent primary cortical neurons to explore the effects of mutation on protein distribution, aggregation potential and toxicity (e.g. axonal transport, protein and organelle trafficking). We have extensive experience in these techniques and many robust cellular and biochemical assays [1,2,8,9,10].

4. Environment The PhD student would join a group of two post-doctoral scientists and another MND Association-funded PhD student (the David Heaton-Ellis Memorial Student) working on MND Genetics who have a substantial consumable budget. We have also been awarded a 5 year grant to model TDP-43 and FUS mutations in a range of cellular and animal models. Professor Shaw has successfully supervised 14 PhD students 12 as first supervisor and all were awarded their PhD's within 4 years. An academic committee monitors progress of the student and this involves written reports at key stages over the PhD. In addition, the student will present their data to other members of the MRC Centre for Neurodegeneration Research.

5. Timetable The PhD student will concentrate on the exome capture, sequencing and bioinformatic analysis for the first 18 months. Subsequent cell-based research will be dependent on the data arising from the exome capture, but it is anticipated that over the final 18 months, the

student will be engaged in genetic validation of variants and cellular expression studies of wild-type and mutant variants in the models outlined above.