Neurobiology of Aging 51 (2017) 178.e11–178.e20

Contents lists available at ScienceDirect

Neurobiology of Aging

journal homepage: www.elsevier.com/locate/neuaging

Genetic epidemiology of motor neuron disease-associated variants in the Scottish population

Holly A. Black [a,1], Danielle J. Leighton [b,c,1], Elaine M. Cleary [b,c,d], Elaine Rose [a], Laura Stephenson [b,c], Shuna Colville [b,c], David Ross [a], Jon Warner [d], Mary Porteous [d], George H. Gorrie [b], Robert Swingler [b], David Goldstein [e], Matthew B. Harms [e], Peter Connick [c], Suvankar Pal [b,c], Timothy J. Aitman [a,*], Siddharthan Chandran [b,c,**]

[a] Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK
[b] The Euan MacDonald Centre for Motor Neurone Disease Research, University of Edinburgh, Edinburgh, UK
[c] Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK
[d] South East Scotland Genetics Service, Western General Hospital, Edinburgh, UK
[e] Institute for Genomic Medicine, Columbia University, New York, USA

ARTICLE INFO

Article history:
Received 7 October 2016
Received in revised form 24 November 2016
Accepted 13 December 2016
Available online 21 December 2016

Keywords:
Motor neuron disease
Amyotrophic lateral sclerosis
TBK1
NEK1

ABSTRACT

Genetic understanding of motor neuron disease (MND) has evolved greatly in the past 10 years, including the recent identification of association between MND and variants in *TBK1* and *NEK1*. Our aim was to determine the frequency of pathogenic variants in known MND genes and to assess whether variants in *TBK1* and *NEK1* contribute to the burden of MND in the Scottish population. *SOD1*, *TARDBP*, *OPTN*, *TBK1*, and *NEK1* were sequenced in 441 cases and 400 controls. In addition to 44 cases known to carry a *C9orf72* hexanucleotide repeat expansion, we identified 31 cases and 2 controls that carried a loss-of-function or pathogenic variant. Loss-of-function variants were found in *TBK1* in 3 cases and no controls and, separately, in *NEK1* in 3 cases and no controls. This study provides an accurate description of the genetic epidemiology of MND in Scotland and provides support for the contribution of both *TBK1* and *NEK1* to MND susceptibility in the Scottish population.

1. Introduction

Motor neuron disease (MND) is a rapidly progressive and fatal neurodegenerative disorder, characterized by loss of motor neuron function. Presentations can include limb onset, bulbar onset, or cognitive/behavioural disease, reflecting variable involvement of the upper motor neurons, lower motor neurons and frontotemporal cortex (Turner and Swash, 2015). Age and site of onset, rate of disease progression and clinical syndrome vary considerably between cases, presenting difficulties for diagnosis and disease management.

Phenotype data from Scottish MND cases are available through the Scottish Motor Neuron Disease Register (SMNDR), a prospective population-based record of all cases diagnosed with MND, which has been operational since 1989 (The Scottish Motor Neuron Disease Research Group, 1992). In Scotland, there is a predicted annual crude incidence of MND of 2.38 per 100,000 of the population (Forbes et al., 2007). Mean survival from symptom onset to death is 2.8 years (Forbes et al., 2007), albeit with a variable trajectory depending on the clinical syndrome.

Substantial progress in understanding the genetic landscape of MND has occurred over the last 10 years, including association with dominant variants at several genetic loci (Renton et al., 2014). Although the majority of cases of MND present without a family history (~90%), variants in the same genes are thought to contribute to the genetic etiology of both familial and apparently sporadic cases. In recent series, pathogenic variants in known genes have been found in around 70% of the cases with a family history and 10% of the cases with no family history (Renton et al., 2014). The main contributors in the UK and other European populations are expansions of an intronic hexanucleotide repeat in *C9orf72* and missense variants in *SOD1* and *TARDBP* (Abel, 2016; Abel et al., 2013; Renton et al., 2014). Although the incidence of MND-associated

* Corresponding author at: Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. Tel.: 0131 651 8021; fax: 0131 651 8800.
** Corresponding author at: The Euan MacDonald Centre for Motor Neurone Disease Research, University of Edinburgh, Edinburgh, UK. Tel.: 0131 465 9519; fax: 0131 465 9519.
E-mail addresses: tim.aitman@ed.ac.uk (T.J. Aitman), siddharthan.chandran@ed.ac.uk (S. Chandran).
[1] Joint first authors.

http://dx.doi.org/10.1016/j.neurobiolaging.2016.12.013

variants specifically in the Scottish population is unknown, screening of an unselected Scottish cohort for variants in *SOD1* identified a high frequency of the I114T variant, which was found in 9% cases (Jones et al., 1995), and more recently *C9orf72* expansions were found in 11% of Scottish cases (Cleary et al., 2016). Two genes, *TBK1* and *NEK1*, have recently been associated with MND, with each thought to account for a small proportion of cases, although they have not been screened in the Scottish MND population (Brenner et al., 2016; Cirulli et al., 2015; Freischmidt et al., 2015; Kenna et al., 2016).

In families where affected individuals carry the same MND-associated variant, for example *SOD1* I114T, the phenotype can be extremely variable (Lopate et al., 2010). This phenotypic heterogeneity, along with previous reports of pathogenic variants in more than one MND gene in the same affected individual (oligogenic cases; Bury et al., 2016; Cady et al., 2015; Chio et al., 2012b; Kenna et al., 2013; Lattante et al., 2012; van Blitterswijk et al., 2012), has led to a hypothesis that disease risk and subsequent phenotype is determined by a combination of genetic factors and modifiers, rather than a single genetic variant, in combination with environmental triggers, as described in the multistep hypothesis of MND (Al-Chalabi et al., 2014).

In this study, we report a multi-gene screen of MND cases and controls from the Scottish population. Our aim was to determine the contribution of variants in different genes to cases of MND in Scotland, to investigate whether variants in *TBK1* and *NEK1* contribute to the burden of cases and to assess the association of variants in MND genes with disease phenotype.

2. Materials and methods

2.1. MND case recruitment and control samples

MND cases were recruited through the SMNDR. The registration process achieved high ascertainment coverage (98% in 1989–98; Forbes et al., 2007), providing a cohort representative of the national MND population. Details of methodology of recruitment have been reported previously (The Scottish Motor Neuron Disease Research Group, 1992); latterly the El Escorial classification system was adopted (Brooks, 1994; Brooks et al., 2000). Recruited cases included individuals aged ≥16 with probable or definite amyotrophic lateral sclerosis and individuals with MND subtypes (progressive bulbar palsy, progressive muscular atrophy and primary lateral sclerosis). Individuals provided written consent for DNA extraction and genetic studies.

Four hundred forty-one samples, obtained from cases diagnosed with MND in Scotland in the years 1989–2014, were included in this study, which included 3 pairs of related individuals (2 brothers, 2 first cousins, and 2 first cousins once removed). Case records were examined for 7 phenotypic characteristics: sex, age at onset, age at diagnosis, time to diagnosis, duration of disease (until death or final data review [20th April 2016]), site of onset (bulbar or spinal), and family history of MND. Individuals were classified as having a family history of MND if a first, second, or third degree relative had been known to have MND, as for cases. The study did not include the presence of frontotemporal dementia (FTD) alone as a criterion for positive family history. Five individuals were lost to follow-up due to relocation from Scotland, and survival dates were censored to date of last contact. The cohort was screened for expansions of the *C9orf72* intronic hexanucleotide repeat, as described by Cleary et al. (2016). A subset of the cohort had been screened for variants in *SOD1* in several previous studies (Hayward et al., 1998; Jones et al., 1995; Swingler et al., 1995).

Five MND cases, each with a variant in 1 of the 5 genes sequenced, were included as positive controls; these were taken from the cohort described by Cirulli et al. (2015). Four hundred ethnicity and sex-matched healthy controls were selected from the Generation Scotland Donor DNA databank (Kerr et al., 2010). The selected controls were aged ≥56 at the time of collection (20% aged 50–56, 66% aged 60–65, 14% aged 66+); an older cohort was chosen to minimize inclusion of young subjects who could go onto develop MND later in life.

2.2. Targeted amplification and sequencing

Five genes were sequenced: *SOD1*, *TARDBP*, *OPTN*, *TBK1*, and *NEK1*; these will be referred to hereafter as the MND gene panel. The panel includes the recently associated *TBK1* and *NEK1*, alongside genes that are among the largest contributors to cases in UK and European populations (*SOD1*, *TARDBP*, and *OPTN*), after *C9orf72* (Abel, 2016; Abel et al., 2013; Renton et al., 2014).

Primers for 120 amplicons (Supplementary Table 1) were designed according to the Fluidigm Access Array protocol (Fluidigm). The primers amplified the coding regions of these 5 genes (Supplementary Table 2), excluding a total of 312bp of coding sequence across the 5 genes, due to primer design constraints. The Fluidigm Access Array was used for amplification, following the manufacturer's multiplex amplicon tagging protocol, with 2 modifications (Martyna Adamowicz-Brice, personal communication); an additional 1:1 AMPure XP (Agencourt) cleanup step was introduced post-harvesting and the subsequent dilution step was removed. The amplicon library was sequenced on an Illumina MiSeq with 2×150-bp reads. Each batch of 48 samples included at least one water blank as a negative control. The positive control samples were each run in 2 independent batches.

2.3. Read mapping, sample filtering, variant calling, and annotation

Quality control of the amplicon sequencing data was performed using FastQC (version 0.11.2) (Andrews, 2010). Primer sequences were removed from the 5′ ends of reads using the cutadapt tool (version 1.7.1) with the 'anchor' option (Martin, 2011). The maximum error rate for primer sequences was set to 10%. Reads were mapped to the human genome reference sequence hs37d5 using BWA MEM (version 0.7.10; Li, 2013). Picard (version 1.85; Broad Institute), and DepthOfCoverage GATK tool (version 3.3–0; McKenna et al., 2010) were used to collect alignment and amplicon coverage statistics. The median amplicon coverage of the negative control sample was used to determine the threshold for including samples from the same batch in further analysis. Only samples with a median coverage that was >10× that of the negative control were included. This removed 7 cases and 11 controls, leaving 434 cases (432 independent) and 389 controls for further analysis. The UnifiedGenotyper tool, as implemented in GATK version 2.6 (McKenna et al., 2010), was used for variant calling. ANNOVAR (version 2014 Nov 12) was used to provide functional annotation of the variants (Wang et al., 2010). Variants were annotated with reference to the transcripts listed in Supplementary Table 2.

2.4. Variant filtering and validation

Intronic and synonymous variants, variants with a population frequency greater than 1% in either the 1000 Genomes (October 2014 release; 1000 Genomes Project Consortium et al., 2015) or ExAC v0.2 (Lek et al., 2016) data sets and variants with a frequency >5% in our cohort were excluded from further analysis. Per-sample variant calls were filtered to exclude calls with a read depth <50 or an allele balance <0.3. Filtered variants were validated by Sanger sequencing. One sample failed to amplify using PCR at variant validation stage and was excluded from further analysis, leaving