Neurobiology of Aging 35 (2014) 1514.e1–1514.e12

Contents lists available at ScienceDirect

Neurobiology of Aging

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

Assessment of common variability and expression quantitative trait loci for genome-wide associations for progressive supranuclear palsy

Raffaele Ferrari [a,b,c], Mina Ryten [c], Roberto Simone [b,c], Daniah Trabzuni [c,d], Naiya Nicolaou [b,c], Geshanthi Hondhamuni [b,c], Adaikalavan Ramasamy [c,e], Jana Vandrovcova [b,c], UK Brain Expression Consortium[c,e], Michael E. Weale [e], Andrew J. Lees [b], Parastoo Momeni [a], John Hardy [b,c], Rohan de Silva [b,c,*]

[a] Laboratory of Neurogenetics, Department of Internal Medicine, Texas Tech University Health Sciences Center, Lubbock, TX, USA
[b] Reta Lila Weston Institute, UCL Institute of Neurology, London, UK
[c] Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK
[d] Department of Genetics, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia
[e] Department of Medical and Molecular Genetics, King's College London, Guy's Hospital, London, UK

ARTICLE INFO

Article history:
Received 13 November 2013
Received in revised form 6 January 2014
Accepted 8 January 2014
Available online 13 January 2014

Keywords:
PSP
GWAS
Quantitative trait loci
Sequencing
Haplotype
Linkage disequilibrium
Missense variant

ABSTRACT

Progressive supranuclear palsy is a rare parkinsonian disorder with characteristic neurofibrillary pathology consisting of hyperphosphorylated tau protein. Common variation defining the microtubule associated protein tau gene (*MAPT*) H1 haplotype strongly contributes to disease risk. A recent genome-wide association study (GWAS) revealed 3 novel risk loci on chromosomes 1, 2, and 3 that primarily implicate *STX6*, *EIF2AK3*, and *MOBP*, respectively. Genetic associations, however, rarely lead to direct identification of the relevant functional allele. More often, they are in linkage disequilibrium with the causative polymorphism(s) that could be a coding change or affect gene expression regulatory motifs. To identify any such changes, we sequenced all coding exons of those genes directly implicated by the associations in progressive supranuclear palsy cases and analyzed regional gene expression data from control brains to identify expression quantitative trait loci within 1 Mb of the risk loci. Although we did not find any coding variants underlying the associations, GWAS-associated single-nucleotide polymorphisms at these loci are in complete linkage disequilibrium with haplotypes that completely overlap with the respective genes. Although implication of *EIF2AK3* and *MOBP* could not be fully assessed, we show that the GWAS single-nucleotide polymorphism rs1411478 (*STX6*) is a strong expression quantitative trait locus with significantly lower expression of *STX6* in white matter in carriers of the risk allele.

1. Introduction

Progressive supranuclear palsy (PSP) is a progressive neurodegenerative disorder characterized by early postural instability, supranuclear gaze palsy, and cognitive decline with age at onset typically between 60 and 65 years and 5.8–5.9 years median time from disease onset to death (Burn and Lees, 2002; Williams and Lees, 2009). PSP is a primary tauopathy with widespread tau pathology mainly defined by hyperphosphorylated tau protein, neurofibrillary tangles, neuropil threads and characteristic glial tau inclusions, astrocytic tufts, and oligodendroglial-coiled bodies in the white matter (Dickson et al., 2007).

By virtue of its pathology, it is not surprising that PSP risk has been consistently associated with the H1 haplotype of the tau gene (*MAPT*) (Baker et al., 1999; Conrad et al., 1997; Ezquerra et al., 1999) (for review, see Vandrovcova et al., 2010). While a handful of autosomal dominant PSP cases have been described (Rohrer et al., 2011; Rojo et al., 1999), the majority are sporadic, without evidence of familial clustering. Furthermore some of the *MAPT* mutations that cause familial frontotemporal dementia with parkinsonism linked to chromosome 17 (FTDP-17 and/or FTLD-tau) (Hutton et al., 1998; Spillantini et al., 1998) cause phenotypes resembling PSP (Choumert et al., 2012; Delisle et al., 1999; Morris et al., 2003; Pastor et al., 2001; Poorkaj et al., 2002; Rohrer et al., 2011; Ros et al., 2005b; Rossi et al., 2004; Spina et al., 2008; Stanford et al., 2000).

* Corresponding author at: Reta Lila Weston Institute and Department of Molecular Neuroscience, UCL Institute of Neurology, London, WC1N 3BG, UK. Tel.: +44 020 7679 4264; fax: +44 020 7278 4993.
E-mail address: r.desilva@ucl.ac.uk (R. de Silva).

0197-4580/$ – see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neurobiolaging.2014.01.010

Elucidation of the possible functional basis of the H1 haplotype association with PSP and corticobasal degeneration (CBD) has led to suggestions of allele-specific differences in transcription and alternative splicing of *MAPT*; both PSP and CBD tau pathology is predominantly 4-repeat tau (4R-tau), that is, consisting of tau protein isoforms with 4 microtubule-binding domains as a result of splicing of *MAPT* exon 10 (Caffrey et al., 2006, 2008; Chambers et al., 1999; Luk et al., 2010; Majounie et al., 2013; Takanashi et al., 2002; Trabzuni et al., 2012). The H1c sub-haplotype within the H1 clade, which drives the association with PSP and CBD (Pittman et al., 2005) was shown to be associated with increased transcription and exon 10 splicing and could thus form the basis of the 4R-tau-dominant pathology (Myers et al., 2007). Other associated loci have been suggested on chromosomes (Chr) 1q31.1 (Ros et al., 2005a) and 11p12 (Melquist et al., 2007) as well as a pV380L polymorphism in parkin (PARK2) (Ros et al., 2008), but these have not been replicated to date.

In 2011, an international consortium published a genome-wide association study (GWAS) that included a large majority of pathologically proven and clinical PSP cases available in Western Europe and the United States (Höglinger et al., 2011). This unequivocally revealed several new associations besides that with *MAPT* (Höglinger et al., 2011). The associated single-nucleotide polymorphisms (SNPs) clustered at 3 different loci on Chr1q25.3, Chr2p11.2, and Chr3p22.1; at the syntaxin 6 (*STX6*), eukaryotic translation initiation factor 2-alpha kinase 3 (*EIF2AK3*), and myelin-associated oligodendrocyte basic protein (*MOBP*) genes, respectively (Höglinger et al., 2011). In addition to the established role of tau, the genes implicated by these associations provide tantalizing suggestions about the cellular pathways and processes that could be affected in PSP, including intracellular vesicular trafficking, central nervous system (CNS) myelination and the endoplasmic reticulum (ER)-mediated cellular response to stress, and abnormally unfolded proteins (Höglinger et al., 2011).

As with all genetic association studies, the next challenge is to determine the functional underpinnings of these associations and how their allelic differences influence risk. In the first place, the strongest associations do not necessarily implicate the nearest gene, but could point to other genes in the vicinity that are in linkage disequilibrium (LD). On rare occasions, the associated SNP(s) are the functional variant(s) by missense change of protein coding sequences or, by influencing important gene regulatory motifs involved in expression or splicing. More frequently, associated SNPs reside within intronic or intergenic regions with the possibility that they act at the RNA level, thus affecting gene expression or they influence the variety of non-coding RNA genes such as micro-RNAs or their targets. Therefore, to delineate the associated region, mapping of LD with the associated SNPs enables us to short list the candidate genes implicated by the associations for more detailed sequence analysis for the identification of high-risk variants that could affect protein or gene function. In addition, the availability of array-based genome-wide gene expression data for various tissues and multiple brain regions (see Trabzuni et al., 2011) now enables expression quantitative trait locus (eQTL) analysis to identify expression and splicing changes that are influenced by individual SNP alleles, be they in cis within the same gene locus or neighboring and distal genes.

The first discovery stage of the PSP GWAS included 141 pathologically confirmed PSP cases from the Queen Square Brain Bank (Höglinger et al., 2011). In this study, we selected 84 of these cases to further investigate the associated loci by direct sequencing of the coding sequences of the implicated genes to identify further causative polymorphisms, including missense changes. Our primary aim was to determine if any coding variability underpinned the reported association. A secondary aim was to identify rarer variants that are key to contributing to disease risk. Furthermore, we carried out analysis of regional genome-wide eQTL analysis in genetically characterized control brains to determine if any of the associated SNPs cause allelic differences in gene expression that might also explain the association with disease.

2. Methods

2.1. Study population

Eighty-four pathologically confirmed PSP cases were screened in this study, all of which were included in the recent GWAS (Höglinger et al., 2011) and have genome-wide genotype data. They were all white, western European origin, and met modified NINDS possible or probable criteria (Litvan et al., 1996a) and diagnosis was confirmed pathologically using standardized criteria (Dickson et al., 2007; Litvan et al., 1996a, 1996b). All patients were collected under approved protocols followed by informed consent and this work was approved by the Joint Medical Ethics Committee of the National Hospital of Neurology and Neurosurgery, London.

2.2. Genetic analysis

Genomic DNA was extracted from dissected samples (100–200 mg) of human postmortem brain tissue using the Qiagen DNeasy Blood & Tissue Kit (Qiagen, Manchester, UK). Genotype data for the 3 GWAS-associated SNPs rs1411478 (Chr1q25.3; $P_{(joint)}$: 2.3×10^{-10}), rs7571971 (Chr2p11.2; $P_{(joint)}$: 3.2×10^{-13}), and rs1768208 (Chr3p22.1; $P_{(joint)}$: 1.0×10^{-16}) were derived from the PSP-GWAS data (Höglinger et al., 2011).

We sequenced all the coding exons and flanking intronic regions of the genes that were within the associated haplotype block (regions of LD of $r^2 > 0.8$ with the GWA associated SNPs). These were *STX6* (exons 2–8), trans-membrane protein 1 gene (*MR1*) (exons 2–7), *EIF2AK3* (exons 2–17), and *MOBP* (exons 3–5). Primers for polymerase chain reaction amplification of each exon and immediately flanking introns were designed using Primer3 (http://frodo.wi.mit.edu/primer3/) and are available upon request. Exploratory sequencing of purified polymerase chain reaction amplicons was carried out in a single direction and in both directions to confirm any identified variants. Sanger sequencing was carried out with the Big Dye Terminator kit (ABI, Foster City, CA, USA) following standard protocol as recommended by the manufacturer, and run on a 3730 DNA Analyzer (ABI), followed by analysis with Sequencher 4.9 software (Gene Codes Corporation, Ann Arbor, MI, USA). The sequences of the coding and flanking intronic regions were used to identify disease-associated haplotypes that are in LD with the most strongly associated SNPs from the GWAS. Chr1 (rs1411478 in intron 4 of *STX6*): common coding and noncoding variants in *STX6* and *MR1* were used to build a haplotype spanning ~70 Kbp. Chr2 (rs7571971 in intron 2 of *EIF2AK3*): common coding and noncoding variants in *EIF2AK3* were used to build a haplotype spanning ~42.6 Kbp. Chr3 (rs1768208 in intron 2 of *MOBP*): only noncoding variants in *MOBP* could be used to build a haplotype spanning ~32 Kbp. LD was analyzed using Haploview (www.broadinstitute.org/haploview) (Barrett et al., 2005) and SNAP, the SNP annotation and proxy search program (www.broadinstitute.org/mpg/snap/index.php) (Johnson et al., 2008) for HapMap Release 22 SNP data for CEU (Utah residents with Northern and Western European ancestry from the CEPH collection). The potential damaging effect of any novel missense