nature genetics

Mutations in *SPG11*, encoding spatacsin, are a major cause of spastic paraplegia with thin corpus callosum

Giovanni Stevanin[1–3,17], Filippo M Santorelli[4,17], Hamid Azzedine[1,2,17], Paula Coutinho[5,6], Jacques Chomilier[7], Paola S Denora[4], Elodie Martin[1,2], Anne-Marie Ouvrard-Hernandez[8], Alessandra Tessa[4], Naïma Bouslam[1,2], Alexander Lossos[9], Perrine Charles[3], José L Loureiro[5,6], Nizar Elleuch[1,2], Christian Confavreux[10], Vítor T Cruz[6], Merle Ruberg[1,2], Eric Leguern[1–3], Djamel Grid[11], Meriem Tazir[12], Bertrand Fontaine[13–15], Alessandro Filla[16], Enrico Bertini[4], Alexandra Durr[1–3] & Alexis Brice[1–3,14,15]

Autosomal recessive hereditary spastic paraplegia (ARHSP) with thin corpus callosum (TCC) is a common and clinically distinct form of familial spastic paraplegia that is linked to the *SPG11* locus on chromosome 15 in most affected families. We analyzed 12 ARHSP-TCC families, refined the *SPG11* candidate interval and identified ten mutations in a previously unidentified gene expressed ubiquitously in the nervous system but most prominently in the cerebellum, cerebral cortex, hippocampus and pineal gland. The mutations were either nonsense or insertions and deletions leading to a frameshift, suggesting a loss-of-function mechanism. The identification of the function of the gene will provide insight into the mechanisms leading to the degeneration of the corticospinal tract and other brain structures in this frequent form of ARHSP.

ARHSP-TCC clinically manifests as spastic paraplegia, usually beginning during infancy or puberty (and up to 23 years), preceded by learning difficulties[1–3]. Cognitive impairment is first noticed in childhood and progresses insidiously to severe functional disability of a frontal type over a period of 10–20 years[2]. Some affected individuals develop a pseudobulbar involvement, with dysarthria, dysphagia and upper limb spasticity, associated with bladder dysfunction and signs of predominantly axonal, motor or sensorimotor peripheral neuropathy[3]. Positron emission tomography shows cortical and thalamic glucose hypometabolism. Magnetic resonance imaging shows TCC that predominates in the rostral third, with hyperintensities in periventricular white matter and cerebral cortical atrophy predominating in the frontal region.

The major ARHSP-TCC locus, *SPG11* (refs. 4–6), was recently restricted to the region between markers *D15S968* and *D15S132* (refs. 7–9). This disorder is found in several ethnic groups, particularly in countries around the Mediterranean basin and in Japan[5–9]. TCC is not restricted to SPG11, however[10–13].

In the current study, all available members of 12 ARHSP-TCC families were genotyped using 34 microsatellite markers in the candidate interval for *SPG11* and the adjacent and overlapping loci for *SPG21* (ref. 12) and agenesis of corpus callosum with polyneuropathy (ACCPN)[14]. Maximal positive multipoint LOD scores ranging from 0.60 to 3.85, which corresponded to the maximal expected values in the pedigrees, were obtained in ten families in the *SPG11* interval (**Supplementary Fig. 1** online). The combined multipoint LOD score reached the value of 17.32 in these families. Linkage was not conclusive in the two remaining kindreds. Haplotype reconstructions in two consanguineous families (FSP672 and FSP221) with strong evidence for linkage to *SPG11* (LOD scores of 2.60 and 3.85) further restricted the region most likely to contain the responsible gene to the 3.2-cM homozygous region between *D15S778* and *D15S659* (**Supplementary Figs. 1** and **2** online). This was the only region that segregated with the disease in a genome scan performed in the most informative family (FSP221).

This interval contains ~40 genes, according to the National Center for Biotechnology Information (NCBI) and Ensembl databases. We analyzed 18 genes (**Fig. 1**) in SPG11 index patients by direct sequencing of all exons and their splicing sites but did not find any disease-causing mutations in 17 of them. When we analyzed the *KIAA1840* gene, also known as *FLJ21439*, we found ten different mutations in 11 of these families (**Fig. 1b**). All mutations (four nonsense mutations,

[1]INSERM, UMR679, Federal Institute for Neuroscience Research, Pitié-Salpêtrière Hospital, Paris, France. [2]Université Pierre et Marie Curie-Paris 6, UMR679, Pitié-Salpêtrière Hospital, Paris, France. [3]Assistance Publique–Hôpitaux de Paris (AP-HP), Pitié-Salpêtrière Hospital, Department Cytogenetics and Genetics, Paris, France. [4]Unit of Molecular Medicine, IRCCS-Bambino Gesù Children's Hospital, Rome, Italy. [5]UnIGENe, University of Porto, Portugal. [6]Departamento de Neurologia, Hospital S. Sebastiao, Santa Maria da Feira, Portugal. [7]Institut de Minéralogie et de Physique des Milieux Condensés, Paris 6 and 7 Universities, Paris, France. [8]Department of Neurology, Centre Hospitalier Universitaire, Grenoble, France. [9]Department of Neurology, Agnes Ginges Center for Human Neurogenetics, Hadassah-Hebrew University Medical Center, Jerusalem, Israel. [10]Hôpital Neurologique et Neurochirurgical Pierre Wertheimer, Lyon, France. [11]Genethon, Evry, France. [12]Department of Neurology, Mustapha Hospital, Algiers, Algeria. [13]INSERM U546, Salpêtrière Hospital, Paris, France. [14]AP-HP, Pitié-Salpêtrière Hospital, Federation of Neurology, Paris, France. [15]Université Pierre et Marie Curie-Paris 6, Pitié-Salpêtrière Medical School, Paris, France. [16]Department of Neurological Sciences, Federico II University, Naples, Italy. [17]These authors contributed equally to this work. Correspondence should be addressed to G.S. (stevanin@ccr.jussieu.fr).

Received 26 July 2006; accepted 18 January 2007; published online 18 February 2007; doi:10.1038/ng1980

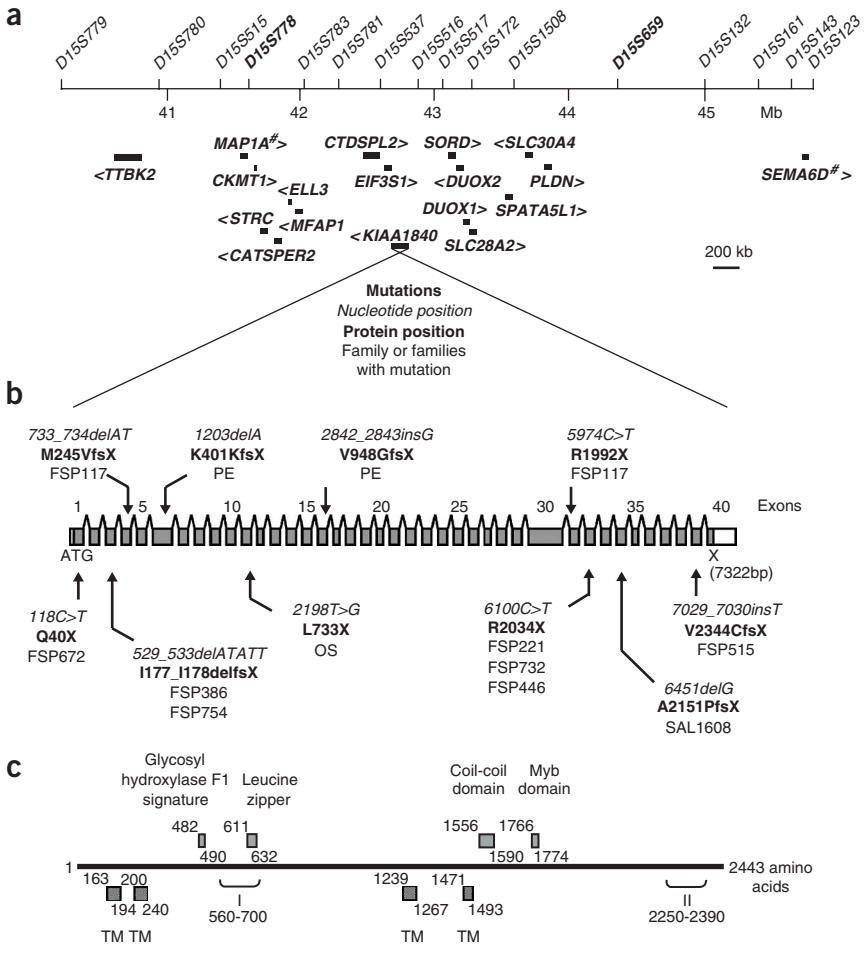

Figure 1 Critical region of *SPG11*. (**a**) Physical map of human chromosome 15q15–21 with selected genetic markers and the candidate genes that were sequenced. Distances in Mb are indicated relative to chromosome 15 according to the Ensembl database. Markers defining the reduced candidate interval are in bold. # indicates that these genes (*SEMA6D* and *MAP1A*) were analyzed in a previous study[8]. > and < indicate the orientation of the open reading frame (ORF) of each gene. (**b**) Exon-intron structure of the 101 kb of the *KIAA1840* gene, also known as *FLJ21439*, with positions of mutations identified in 11 SPG11 families. (**c**) Several of the putative functional domains (boxes) and their positions on the protein sequence. Regions I and II correspond to structurally similar domains based on their hydrophobicity status analyzed with DrawHCA software. Notably, this protein is assigned to the aromatic compound dioxygenase superfamily because of a 22% identity with the consensus sequence between residues 2104 and 2381 (ref. 24).

four small deletions and two small insertions) were in the coding sequence and were at the homozygous state, except in two kindreds in which affected individuals were compound heterozygous for two different variants in this gene. Mutations segregated with the disease in all families (**Fig. 2a**,**b**,**c**) and were not detected in a panel of at least 140 chromosomes from unrelated control individuals of European or North African origin. All were predicted to cause early protein truncation.

We found only two mutations in more than one pedigree (**Fig. 2a**,**b**). We identified a homozygous 6100C→T substitution (R2034X) in exon 32 in three consanguineous North African kindreds. The predicted protein—if not rapidly degraded—would be truncated by 27% of its normal size. A 5-bp deletion in exon 3 (529_533delATATT) leading to a frameshift that should cause early termination at amino acid residue 179 was homozygous in two Portuguese families. The absence of linkage disequilibrium in previous studies[8] might be explained by the diversity of the mutations identified, but similar haplotypes in Algerian and Portuguese families suggest regional founder effects (**Fig. 2a**,**b**). We did not find any mutations in the remaining family, but linkage was not conclusive in this family, either.

We examined a total of 22 individuals with mutations in the *SPG11* gene at a mean age of 24.8 ± 9.5 years (range, 12–49). Onset of the disease occurred between 2 and 23 years (mean age, 11.8 ± 5.5 years) and consisted of either spastic gait (57%, 12/21 patients) or cognitive impairment (19%, 4/21), sometimes diagnosed as mental retardation. After ~10 years of evolution, the full-blown clinical picture consisted of progressive and severe spastic paraplegia with distal wasting and cognitive problems (**Table 1**). In several cases ($n = 6$), cognitive dysfunction clearly worsened with disease progression. Cerebral imaging showed a TCC accompanied by periventricular white matter changes and cortical atrophy in the majority of affected individuals. Pseudobulbar dysarthria was frequent (54%, $n = 12$), and dystonic voice was noted in one individual. Notably, although electromyographic recordings were normal in a few affected individuals, peripheral neuropathy was frequent (72%, 13/18) and was mostly associated with pure motor changes. We occasionally observed additional signs, such as optic atrophy, retinitis pigmentosa, mild cerebellar signs, cataract and clinodactyly.

The human *SPG11* gene contains 40 exons on chromosome 15q21.1. The full-length 8-kb transcript encodes a predicted protein of 2,443 amino acids of unknown function, which we termed spatacsin, after 'spasticity with thin or atrophied corpus callosum syndrome protein'. The sequence of spatacsin has been strongly conserved throughout evolution, with orthologs in mammals and other vertebrates. Human *SPG11* shares 85% identity with the homologous protein in dog, 76% with mouse and 73% with rat. Less similarity was found with homologous proteins of smaller sizes in *Fugu rubripes* (44%), *Tetraodon nigroviridis* (39%) and *Drosophila melanogaster* (22%).

Neither the gene nor the protein it is predicted to encode in many species shows any significant sequence similarity to known cDNA or protein sequences. At the protein level, spatacsin has very few known domains, including four putative transmembrane signatures (**Fig. 1c**). The level of hydrophobicity (34.2%) over the entire sequence was typical of a globular protein, but a succession of globular domains is more likely because of its size. However, hydrophobic cluster analysis[15] identified only one small linker between amino acids 1410 and 1440. This analysis also confirmed one of the putative transmembrane regions at amino acids 200 to 240 that was found in four other vertebrates, but not in tetraodon or *D. melanogaster*, in which the

protein has a shortened N-terminal domain. A thorough search for putative duplications brought to light two structurally similar regions based on the hydrophobicity state (residues 560–700 and 2250–2390) in all vertebrate homologs of the protein. In the human protein, sequence identity between these two regions was 19% (**Supplementary Fig. 3** online).

Because there is no antibody against endogenous spatacsin as of yet, we examined its subcellular location by overexpression of a GFP fusion protein in COS-7 cells. The fusion protein had a diffuse and reticular cytosolic and perinuclear distribution and was sometimes also present in nuclei (**Supplementary Fig. 4** online). Fluorescence was still detectable after methanol fixation, suggesting that part of this protein is associated with membranes and/or the cytoskeleton. GFP-spatacsin slightly colocalized with mitochondria (erab) and endoplasmic reticulum (calnexin) but not with Golgi structures (giantin, alpha-COP), transport vesicles (Scamp1) or the cytoskeleton (alpha-tubulin). Examination of the endogenous protein with a specific antibody will refine the subcellular localization of this new and very large protein.

Previous expression profiling of the *SPG11* gene[16] showed low but ubiquitous expression in mouse tissues, including the brain and structures apparently not related to the phenotype—a profile reported in other neurodegenerative diseases[17]. We successfully amplified seven overlapping cDNA fragments from the *KIAA1840* mRNA extracted from human cerebral cortex and used them to probe human adult multiple-tissue RNA blots. At least three alternative transcripts were detected in all structures of adult brain. The full-length transcript (~8 kb) was most highly expressed in the cerebellum, the ~5-kb transcript in the cerebral cortex (**Fig. 3**). Using RT-PCR, we identified a 4.8-kb transcript composed of exons 28–40 with retention of intron 29 and parts of introns 27 and 28. Various sites of translation initiation are possible in exon 30 that would keep the protein sequence in the same phase as the full-length protein. We cannot exclude the possibility that other transcripts exist with the same size but a different composition.

When we investigated the temporal and regional expression of the rat *KIAA1840* mRNA by *in situ* hybridization in rat brain, it was undetectable in newborn rats (P1). However, we did detect it in the cerebellum from P6–P21. At adulthood (P68), it was expressed throughout the brain (**Fig. 4** and **Supplementary Figs. 5** and **6** online). Expression was low, but stronger signals were observed in the pineal gland, the edges of the lateral ventricles, the granular layer of the cerebellum and the hippocampus. In contrast to human RNA blots, we detected only weak expression in the cerebral cortex. Understanding the function of spatacsin in these structures would help to explain the major features of the disease phenotype: for example, expression in the hippocampus could be related to the cognitive impairment observed in affected individuals. In addition, it remains to be investigated whether the labeling of the edges of the lateral ventricles (the location of oligodendrocyte progenitors) is related to the white matter changes in affected individuals.

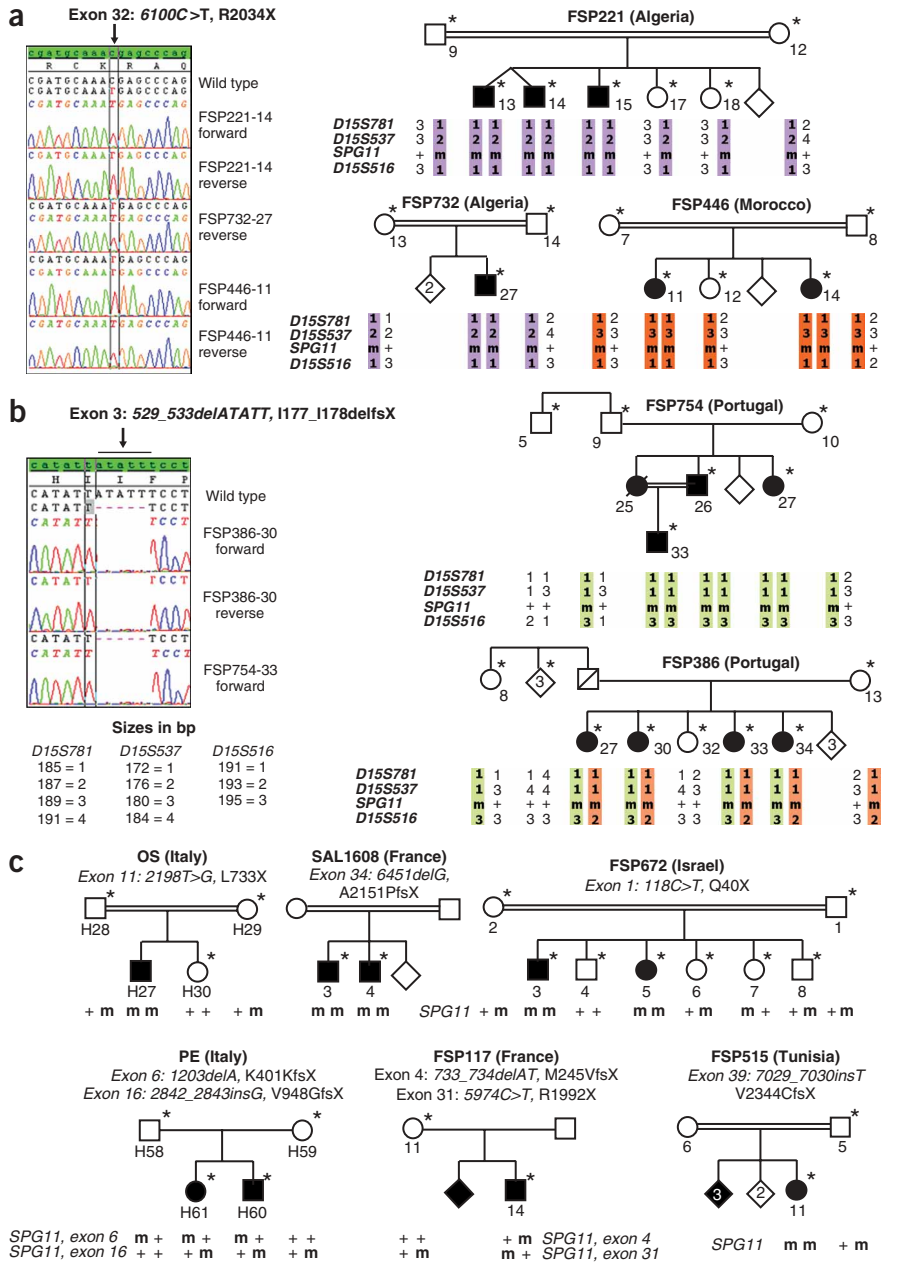

Figure 2 Pedigrees and segregation of the mutations detected in *KIAA1840*. Square symbols are men, and circles are women. The filled symbols are affected individuals. The numbers are an internal reference for each sampled individual. Asterisks indicate sampled subjects. m = mutation; + = wild type. Electropherograms are shown for the two recurrent mutations only. (**a**,**b**) Families with common origins sharing the same mutations. Haplotypes for three close microsatellites segregating with the mutations are highlighted. The correspondence between the numbering of alleles and their size in bp is indicated. (**c**) Nonrecurrent mutations.

Table 1 Clinical features of 22 SPG11 patients from 11 families

		Age (years)			Pyramidal signs									
Geographical origin	Individual (sex)	At onset of first sign(s)	At examination	Disease severity	UL reflexes	LL reflexes	Plantar reflex	Muscle wasting	Vibration sense diminished at ankles	Mental retardation/ cognitive decline	Sphincter disturbances	Cerebral MRI	ENMG	Other
France	1608-3 (M)	10 (spasticity)	25	WB	N	++	↑↑	Distal LL	No	MR	+	TCC, ventricular dilatation	N	Pigmentary retinitis
	1608-4 (M)	16 (spasticity)	30	WB	N	++	↑↑	Distal LL and UL	No	MR	+	TCC, global atrophy with ventricular dilatation	Motor axonal PNP	None
	117-14 (M)	13 (gait)	17	WB	+	++	↑↑	nd	No	LD	+	Mild ventricular dilatation, mild bulbopontine atrophy	N	Dysarthria, ptosis, delayed puberty
Algeria	732-27 (M)	9 (falls)	18	Moderate	nd	++	↑↑	No	No	MR	+	TCC, cerebellar vermian atrophy, periventricular WMH	Axonal PNP	Tremor, dysarthria
	221-13 (M)	12 (spasticity)	25	WB	++	++(– at ankles)	↑↑	Mild LL	No	CD	No	TCC, WMH	Axonal and demyelinating PNP	Dystonic voice, cataract
	221-14 (M)	23	25	Moderate	++	++	↑↑	No	No	Moderate MR (IQ 69)	No	TCC, WMH	Axonal and demyelinating PNP	Nystagmus, cerebellar clumsiness UL
	221-15 (M)	22 (spasticity)	23	Moderate	+	+	↑↑	No	No	Mild MR	No	TCC	Axonal and demyelinating PNP	None
Morocco	446-11 (F)	6 (MR), 16 (spasticity)	19	Moderate	+	++ (– at ankles)	↑↑	Mod. UL, LL	Abolished	Mild LD	No	TCC, cortical atrophy	Motor PNP	Hirsutism
	446-14 (F)	6 (MR)	20	Mild	N	++	↑↑	No	No	Mild MR	No	nd	nd	Clinodactyly
Portugal	386-27 (F)	14 (spasticity)	49	Severe, WB	N	++	↑↑	Mod. UL, LL	No	LD, CD	+	TCC, cortical atrophy, WMH	PNP	Optic atrophy, dysarthria
	386-30 (F)	14 (spasticity)	35	Severe, bedridden	N	++	↑↑	Mod. UL, LL	No	LD, CD	+	nd	PNP	Dysarthria, dysphagia, postural tremor
	386-33 (F)	13 (spasticity)	31	Severe, WB	++	++	↑↑	Mod. UL, LL	No	LD, CD	No	nd	PNP	Dysarthria
	386-34 (F)	14 (spasticity)	28	Severe, WB	++	++	↑↑	No	No	LD, CD	No	nd	No	Dysarthria
	754-26 (M)	14 (spasticity)	35	WB	++	++	↑↑	Mod. UL, LL	nd	Mild MR, PFD	No	TCC, cortical atrophy	PNP	Dysarthria
	754-27 (F)	17 (spasticity)	43	WB	++	++	↑↑	Mod. UL, LL	nd	Mild MR, PFD	No	nd	PNP	Dysphagia, dysarthria
	754-33 (M)	8 (spasticity)	19	Moderate	N	++	↑↑	No	nd	Mild MR, MFS	No	TCC	No	Dysarthria
Tunisia	515-11 (F)	10 (unsteadiness)	12	Moderate	N	++	→→	No	nd	MR, PCD	nd	TCC, Periventricular WMH	nd	None
Israel	672-3 (M)	14 (CD), 16 (gait)	22	Moderate	N	++	↑→	No	No	MR, PCD	No	TCC	Mild PNP	Saccadic ocular pursuit, dysarthria
	672-5 (F)	15 (CD) 17 gait	25	Moderate	++	++	↑↑	Yes	nd	MR, PCD	No	nd	nd	Dysarthria
Italy	OS-H27 (M)	2 (falls)	12	Severe	+	++	↑↑	No	No	Mild MR	++	TCC	No	Dysarthria
	PE-H60 (M)	3.5 (gait)	17	Moderate	N	++	↑↑	No	+	Moderate MR	+	TCC, Periventricular WMH	Motor axonal PNP	Tremor
	PE-H61 (F)	4 (spasticity)	16	Moderate	N	++	↑↑	No	+	Mild MR	+	TCC, Periventricular WMH	nd	None

M: male; F: female; N: normal; left and right plantar reflexes: ↑: extensor; →: indifferent. +: present; ++: increased; –: abolished/diminished; ENMG: electroneuromyogram; UL: upper limbs; LL: lower limbs; MR: mental retardation; PNP: peripheral polyneuropathy; TCC: thin corpus callosum; WMH: white matter hyperintensities; nd: not done; WB, wheelchair-bound; LD, learning difficulties; CD, cognitive decline; PFD, progressive frontal dementia; MFS, mild frontal signs; PCD, progressive cognitive decline; mod., moderate.

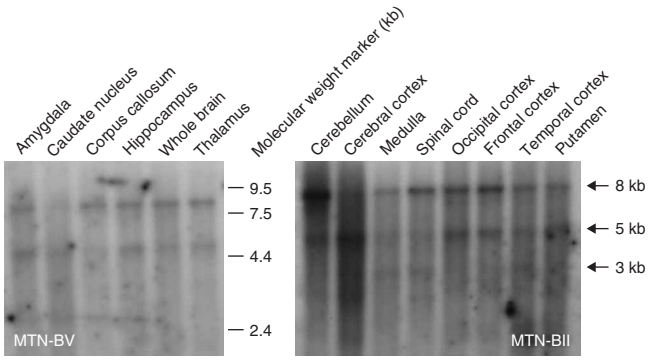

Figure 3 Expression profile of *KIAA1840* examined by RNA blot in human adult brain. The transcripts were present in all brain tissues. Note the 8-kb transcript more intensely expressed in the cerebellum, whereas the 5-kb transcript is mainly found in the cerebral cortex.

Several lines of evidence suggest that *KIAA1840* is responsible for ARHSP-TCC: (i) its location in the recently refined *SPG11* interval (present study and refs. 7–9), (ii) the identification of gene mutations that segregate in ten potentially linked families, (iii) the predicted deleterious effects of the mutations on the protein and (iv) the absence of the mutations in a panel of ethnically matched controls. Furthermore, *SPG11* seems to be the gene most frequently responsible for ARHSP-TCC. Only a single family (8% in our cohort) did not have a mutation in *SPG11*, indicating that there is at least one other responsible gene. On the other hand, whether the *SPG11* gene accounts for other clinical subtypes of ARHSP has yet to be determined.

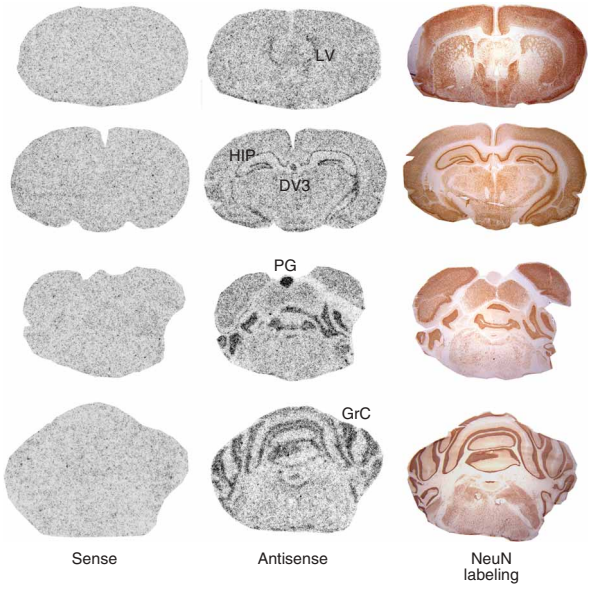

Figure 4 Spatial expression of rat *KIAA1840* in adult brain (P68) by *in situ* hybridization with a pool of three antisense probes or a pool of three sense probes. Note the labeling with the antisense probe in comparison with the Neu-N neuronal specific counterstaining in adjacent slices. No specific staining was observed with the sense probes. *KIAA1840* expression was low throughout the brain and was noted more particularly in the following structures: LV, lateral ventricles; HIP, hippocampus; PG, pineal gland; GrC, granular cell layer of the cerebellum; DV3, third ventricle. The same results were obtained using the pool of three probes or each probe independently.

Spastic paraplegias are believed to result from a dying back of axons. Mitochondrial metabolism, endosomal and *trans*-Golgi trafficking and axonal transport have been implicated in several HSPs[18]. Although the function of spatacsin remains unknown, the experimental evidence that it is expressed in all tissues and is highly conserved among species suggests that it has an essential biological function. The possible presence of at least one transmembrane domain suggests that spatacsin may be a receptor or transporter.

The identification of *SPG11* as the gene responsible for ARHSP-TCC will enable further studies on the natural history and clinical and biological characteristics of this disease. In addition, future studies on the normal function of spatacsin are likely to help to elucidate its pathophysiology.

METHODS

Affected individuals and families. We studied 53 individuals, 22 of whom were affected, from seven consanguineous and five non-consanguineous families of French ($n = 3$), Algerian ($n = 2$), Italian ($n = 2$), Portuguese ($n = 2$), Arab-Israeli ($n = 1$), Moroccan ($n = 1$) and Tunisian ($n = 1$) origin. All presented with a typical ARHSP-TCC phenotype defined by the combination of progressive spastic paraparesis associated with thin corpus callosum on cerebral magnetic resonance imaging[5]. Six SPG11 families had previously been reported with a limited set of analyzed markers[5,7,8] and six were new. They were selected from among 89 families with hereditary spastic paraparesis compatible with recessive transmission collected in our neurogenetic reference center in collaboration with the SPATAX network. We obtained blood samples from all affected individuals and their relatives after they had given their informed written consent in accordance with the local Paris-Necker Ethics Committee (**Supplementary Note**). Genomic DNA was extracted from leukocytes using standard procedures.

Linkage analyses. Genotypes at 34 microsatellites were determined by PCR with a fluorescently labeled primer on an ABI-3730 sequencer (Applied Biosystems) and GeneMapper 3.5 software (Applied Biosystems). Haplotypes were constructed manually. Allegro 1.2c[19] was used to calculate two-point and multipoint LOD scores between the disease phenotype and each of the markers or between the disease phenotype and the map of the markers, assuming complete penetrance, equal allele frequencies for the markers and a frequency of mutated alleles of 0.0005. Marker order and genetic distances were obtained from the Ensembl and Marshfield databases (see URLs section below). The genome scan in family FSP221 was performed using 400 microsatellites, regularly spaced on all chromosomes (ABI-Prism linkage mapping set v2; Applied Biosystems). We used 18 additional polymorphic markers to analyze and exclude three locations other than chromosome 15q with initial multipoint LOD scores of 2.2–2.5 (data not shown).

Mutation detection. Mutations in *SPG7* (ref. 20) were excluded in all families; *SLC12A6* (*ACCPN* locus) and *ACP33* (*SPG21* locus) were ruled out in families in which linkage extended to these regions on chromosome 15q[12,14].

Primers were designed using Oligo6 (MBI) to amplify all coding exons with at least 50 bp of intronic sequences of 16 genes in the *SPG11* candidate interval (**Supplementary Table 1**). PCR-amplified fragments of genomic DNA were purified using exonuclease 1 (New England Biolabs, 2 units per 5 µl PCR product) and antarctic phosphatase (New England Biolabs, 1 unit per 5 µl of PCR product) and sequenced by the fluorescent dideoxy-terminator method (BigDye v3.1, Applied Biosystems) on an automated ABI-3730 sequencer according to the manufacturer's recommendations. Sequences were aligned and compared with consensus sequences with SeqScape 2.5 software (Applied Biosystems). The positions of the mutations are indicated relative to the first ATG codon.

RNA blot analyses (human). Total RNA was extracted from the cerebral cortex of a healthy individual, post-mortem (Brain Bank of INSERM U679), using the RNeasy Mini kit (Qiagen). cDNAs were synthesized using random hexamers and the Termoscript reverse transcriptase as recommended by the supplier (Invitrogen). Seven 1.2-kb probes covering the *KIAA1840* cDNA were amplified by PCR at an annealing temperature of 60 °C (**Supplementary Table 1**).

A mixture of these probes were labeled with α-^{32}P by random priming (Prime-it II Random Primer Labeling Kit, Stratagene) and purified using ProbeQuant G-50 micro columns (Amersham Biosciences, specific activity $\geq 1 \times 10^9$ cpm/μg), were hybridized to human multiple-tissue RNA blots (Clontech) and washed as recommended and were then exposed to X-ray film for autoradiography.

Exon composition of the transcripts was determined by RT-PCR amplification on total RNA from human cerebral cortex using a series of primer pairs covering the whole cDNA and all introns (**Supplementary Table 1**). Alternative transcripts were purified from agarose gels after electrophoresis and were directly sequenced using appropriate primers.

***In situ* hybridization (rat).** Young (P1, P6, P15 and P21, $n = 1$ each) and adult (P68, 200 g, $n = 2$) Sprague-Dawley rats (Charles River) were killed by decapitation, and their brains were rapidly extracted and frozen in isopentane at –50 °C. Sections were cut every 600 μm on a cryostat (–20 °C) from the medulla to the striatum (+1.7 mm from bregma according to the rat brain coordinates of Paxinos & Watson), thaw-mounted on glass slides and stored at –80 °C. *KIAA1840* mRNA expression was analyzed with three antisense oligonucleotides recognizing exons 15, 29 and 32, designed with Helios ETC oligonucleotide design software (Helios Biosciences) from the mRNA sequence (XM-242139) of *Rattus norvegicus*. Each oligonucleotide or a mix of the three oligonucleotides gave identical results. A mix of three sense oligonucleotides was used as negative control.

In situ hybridization was performed as described in ref. 21. Briefly, oligonucleotides were labeled with [^{35}S]dATP using terminal transferase (Amersham Biosciences) to a specific activity of 5×10^8 dpm/μg. The day of the experiment, slides were fixed in 4% formaldehyde in PBS, washed with PBS, rinsed with water, dehydrated in 70% ethanol and air dried. Sections were then covered with 140 μl of hybridization medium (Helios Biosciences) containing 3×10^5 to 5×10^5 dpm of the labeled oligonucleotide mix. Slides were incubated overnight at 42 °C, washed and exposed to a BAS-SR Fujifilm Imaging Plate for 5–10 d. The plates were scanned with a Fujifilm BioImaging Analyzer BAS-5000 and analyzed with Multi Gauge Software (Fuji).

For double labeling experiments, brains were processed as for *in situ* hybridization. After the last wash step, sections were fixed in 4% paraformaldehyde in PBS, preincubated in PBS containing 6% goat serum and 0.1% triton, incubated in the same buffer with mouse antibodies against Neu-N (Chemicon International, 1:250), followed by incubation in the same buffer with biotinylated horse anti-mouse IgG antibodies and ABC reagents (Vector Laboratories). Labeling was captured by emulsion autoradiography.

Bioinformatics. Functional domains were searched for using bioinformatics tools available online at BABEL, Ressource Parisienne en Bioinformatique Structurale and PSORT (see URLs section below). Psi-blast was used to look for homologous proteins or peptides. Alignment of homologous proteins was performed using CLUSTALW. DomHCA was used for a bidimensional analysis of hydrophobicity[15,22]. CysState was used to predict the existence of disulfide bridges[23].

Overexpression studies. The *KIAA1840* cDNA from clone pf01011 (Kazusa DNA Research Institute) was excised from the pBluescript II SK(+) vector using *Xho*I and *Not*I restriction enzymes and was cloned in fusion with EGFP in a *Sal*I/*Bsp*20I–digested pEGFP-C1 vector (Clontech). The construction was verified by direct sequencing after ligation, transformation and plasmid extraction using standard procedures.

COS-7 cells were maintained in DMEM (Invitrogen) supplemented with 10% FBS, penicillin (100 UI/ml) and streptomycin (100 μg/ml). Cells were plated 24 h before transfection on cover slips coated with polyethylenimine and transfected with Lipofectamine-PLUS reagents according to the manufacturer's instructions (Invitrogen). For six-well plates, 1–2 μg of plasmid DNA was used per well. Cells were analyzed by immunofluorescence 48 h post-transfection. The spatacsin-EGFP fusion protein was observed directly after a 15-min fixation using either 4% formaldehyde, which preserves native structures through cross-linking, or ice-cold methanol, which precipitates proteins resulting in a loss of soluble proteins and increased detection of structure-bound proteins. Immunocytochemistry was performed using classical procedures with the following antibodies: mouse anti-alpha tubulin (1:400, Sigma), rabbit anti-giantin (1:2,000, abcam), rabbit anti-calnexin (1:400, Stressgen Bioreagents), mouse anti-erab (1:2,000, abcam), rabbit anti-scamp1 (1:400, Sigma) and rabbit anti-alpha-COP (1:1,000; Affinity Bioreagent). Cells were counterstained with DAPI (1 μg/ml, Sigma) and mounted with Fluoromount-G (Southern Biotech). Samples were observed with a Leica SP1 confocal microscope. Leica confocal software was used to acquire the images.

Accession codes. NCBI: *KIAA1840* homologs: human (NM_025137), dog (XP_544657), chicken (XP_413940.1), mouse (BAE27954), rat (XP_242139.3). Ensembl: *KIAA1840* homologs: *D. melanogaster* (CG13531), *Tetraodon nigroviridis* (GSTENG00003909001).

URLs. Ensembl: http://www.ensembl.org; Marshfield databases: http://research.marshfieldclinic.org/genetics; BABEL: http://babel.infobiogen.fr:1984/; Ressource Parisienne en Bioinformatique Structurale: http://bioserv.rpbs.jussieu.fr/RPBS; PSORT: http://psort.nibb.ac.jp/; Psi-blast: http://www.ncbi.nlm.nih.gov and CLUSTALW: http://www.ebi.ac.uk/clustalw/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

Most of the families were examined and sampled by neurologists participating in SPATAX (the European Network for Hereditary Spinocerebellar Degenerative Disorders). The authors wish to thank the family members for their participation as well as F. Durand-Dubief, C. Tallaksen, P. Ribai and SPATAX members for referring or examining several of the patients. We are also indebted to O. Corti, C. Depienne and S. Dumas for helpful discussions; N. Barton for critical reading of the manuscript and S. Forlani, I. Lagroua, L. Guennec, P. Ibanez, E. Denis and N. Benammar for their assistance. We also thank the DNA Bank of IFR-70 and the Brain Bank of INSERM U679 (E. Hirsch) for providing us with biological material and the Centre National de Génotypage for the genome scan in family FSP221. The pf01011 clone containing the *KIAA1840* full-length cDNA was provided by the Kazusa DNA Research Institute. This work was supported financially by grants from the French Rare Diseases Institute (to G.S. and A.D.), the Verum Foundation (to A.B.), the Italian Ministry of Health (to F.M.S. and E.B.), the Pierfranco and Luisa Mariani Foundation ONLUS (to F.M.S.), Telethon-Italia Foundation (grant number GGP06188 to F.M.S.), the association Strümpell-Lorrain (to the SPATAX network), the Portuguese Foundation for Science and Technology (to P.C., J.L.L. and V.T.C.) and the French National Agency for Research (to the SPATAX network). N.B. and N.E. received fellowships from the French Association for Friedreich Ataxia and the French association Connaître les Syndromes Cérébelleux, P.S.D. is a fellow of The Bambino Gesù Research Program and H.A. was the recipient of a fellowship from the French Association Against Myopathies.

AUTHOR CONTRIBUTIONS

Clinical data were acquired by F.M.S. P.C., A.M.O.H., A.L., P.C., J.L.L., C.C., V.T.C., D.G., M.T., B.F., A.F., E.B., E.L., A.D. and A.B.; F.M.S., N.B., A.T. and G.S. refined the candidate interval using additional markers and families. H.A., F.M.S., N.E., P.S.D. and G.S. analyzed the candidate genes and identified the mutations. E.M. and P.S.D. performed the overexpression studies. F.M.S., P.S.D. and G.S. analyzed the expression of the gene. Bioinformatics studies were performed by H.A., J.C. and G.S.; H.A., F.M.S., A.B., A.D., M.R. and G.S. wrote the paper. A.B., F.M.S. and G.S. supervised the work. Funding was obtained by A.B., F.M.S., E.B., P.C., A.D. and G.S.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Harding, A.E. Classification of the hereditary ataxias and paraplegias. *Lancet* **1**, 1151–1155 (1983).
2. Nakamura, A. *et al.* Familial spastic paraplegia with mental impairment and thin corpus callosum. *J. Neurol. Sci.* **131**, 35–42 (1995).
3. Winner, B. *et al.* Clinical progression and genetic analysis in hereditary spastic paraplegia with thin corpus callosum in spastic gait gene 11 (SPG11). *Arch. Neurol.* **61**, 117–121 (2004).
4. Martinez, M.F. *et al.* Genetic localization of a new locus for recessive familial spastic paraparesis to 15q13–15. *Neurology* **53**, 50–56 (1999).

5. Casali, C. *et al.* Clinical and genetic studies in hereditary spastic paraplegia with thin corpus callosum. *Neurology* **62**, 262–268 (2004).
6. Shibasaki, Y. *et al.* Linkage of autosomal recessive hereditary spastic paraplegia with mental impairment and thin corpus callosum to chromosome 15q13–15. *Ann. Neurol.* **48**, 108–112 (2000).
7. Lossos, A. *et al.* Hereditary spastic paraplegia with thin corpus callosum: reduction of the SPG11 interval and evidence for further genetic heterogeneity. *Arch. Neurol.* **63**, 756–760 (2006).
8. Stevanin, G. *et al.* Spastic paraplegia with thin corpus callosum: description of 20 new families, refinement of the SPG11 locus, candidate gene analysis and evidence of genetic heterogeneity. *Neurogenetics* **7**, 149–156 (2006).
9. Olmez, A. *et al.* Further clinical and genetic characterization of SPG11: hereditary spastic paraplegia with thin corpus callosum. *Neuropediatrics* **37**, 59–66 (2006).
10. Orlacchio, A. *et al.* Clinical and genetic study of a large SPG4 Italian family. *Mov. Disord.* **20**, 1055–1059 (2005).
11. Hughes, C.A. *et al.* SPG15, a new locus for autosomal recessive complicated HSP on chromosome 14q. *Neurology* **56**, 1230–1233 (2001).
12. Simpson, M.A. *et al.* Maspardin is mutated in mast syndrome, a complicated form of hereditary spastic paraplegia associated with dementia. *Am. J. Hum. Genet.* **73**, 1147–1156 (2003).
13. Al Yahyaee, S. *et al.* A novel locus for hereditary spastic paraplegia with thin corpus callosum and epilepsy. *Neurology* **66**, 1230–1234 (2006).
14. Howard, H.C. *et al.* The K-Cl cotransporter KCC3 is mutant in a severe peripheral neuropathy associated with agenesis of the corpus callosum. *Nat. Genet.* **32**, 384–392 (2002).
15. Callebaut, I. *et al.* Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. *Cell. Mol. Life Sci.* **53**, 621–645 (1997).
16. Nagase, T., Nakayama, M., Nakajima, D., Kikuno, R. & Ohara, O. Prediction of the coding sequences of unidentified human genes. XX. The complete sequences of 100 new cDNA clones from brain which code for large proteins *in vitro*. *DNA Res.* **8**, 85–95 (2001).
17. Paisan-Ruiz, C. *et al.* Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. *Neuron* **44**, 595–600 (2004).
18. Crosby, A.H. & Proukakis, C. Is the transportation highway the right road for hereditary spastic paraplegia? *Am. J. Hum. Genet.* **71**, 1009–1016 (2002).
19. Gudbjartsson, D.F., Jonasson, K., Frigge, M.L. & Kong, A. Allegro, a new computer program for multipoint linkage analysis. *Nat. Genet.* **25**, 12–13 (2000).
20. Elleuch, N. *et al.* Mutation analysis of the paraplegin gene (SPG7) in patients with hereditary spastic paraplegia. *Neurology* **66**, 654–659 (2006).
21. Moutsimilli, L. *et al.* Selective cortical VGLUT1 increase as a marker for antidepressant activity. *Neuropharmacology* **49**, 890–900 (2005).
22. Woodcock, S., Mornon, J.P. & Henrissat, B. Detection of secondary structure elements in proteins by hydrophobic cluster analysis. *Protein Eng.* **5**, 629–635 (1992).
23. Mucchielli-Giorgi, M.H., Hazout, S. & Tuffery, P. Predicting the disulfide bonding state of cysteines using protein descriptors. *Proteins* **46**, 243–249 (2002).
24. Kikuno, R., Nagase, T., Waki, M. & Ohara, O. HUGE: a database for human large proteins identified in the Kazusa cDNA sequencing project. *Nucleic Acids Res.* **30**, 166–168 (2002).