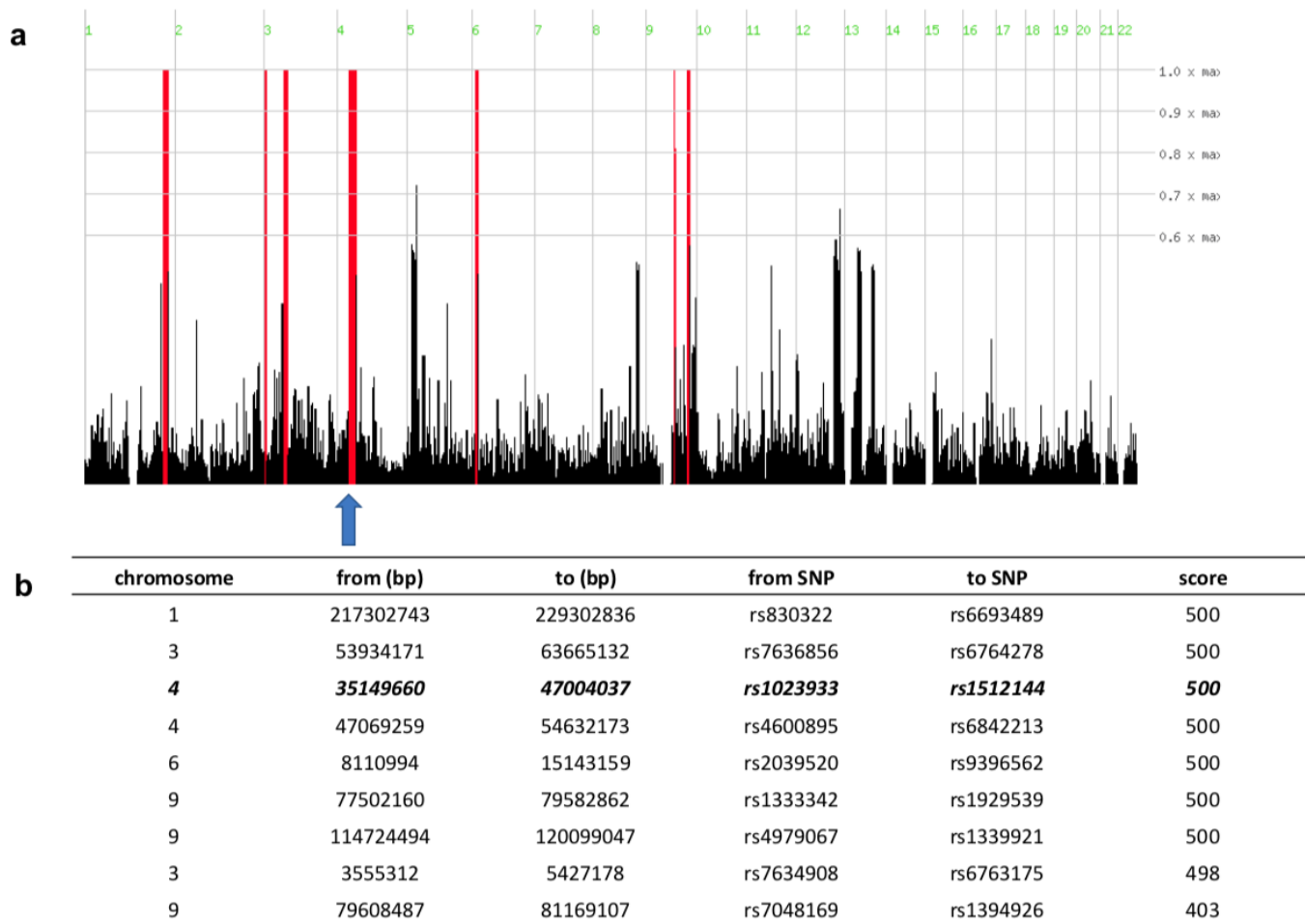


In the format provided by the authors and unedited.

Biallelic expansion of an intronic repeat in *RFC1* is a common cause of late-onset ataxia

Andrea Cortese^{1*}, Roberto Simone², Roisin Sullivan^{1,11}, Jana Vandrovцова^{1,11}, Huma Tariq¹, Wai Yan Yau¹, Jack Humphrey¹, Zane Jaunmuktane², Prasanth Sivakumar¹, James Polke³, Muhammad Ilyas⁴, Eloise Tribollet¹, Pedro J. Tomaselli⁵, Grazia Devigili⁶, Ilaria Callegari⁷, Maurizio Versino^{7,8}, Vincenzo Salpietro¹, Stephanie Efthymiou¹, Diego Kaski¹, Nick W. Wood¹, Nadja S. Andrade⁹, Elena Buglo¹⁰, Adriana Rebelo¹⁰, Alexander M. Rossor¹, Adolfo Bronstein², Pietro Fratta¹, Wilson J. Marques⁵, Stephan Züchner¹⁰, Mary M. Reilly^{1,12} and Henry Houlden^{1,3,12*}

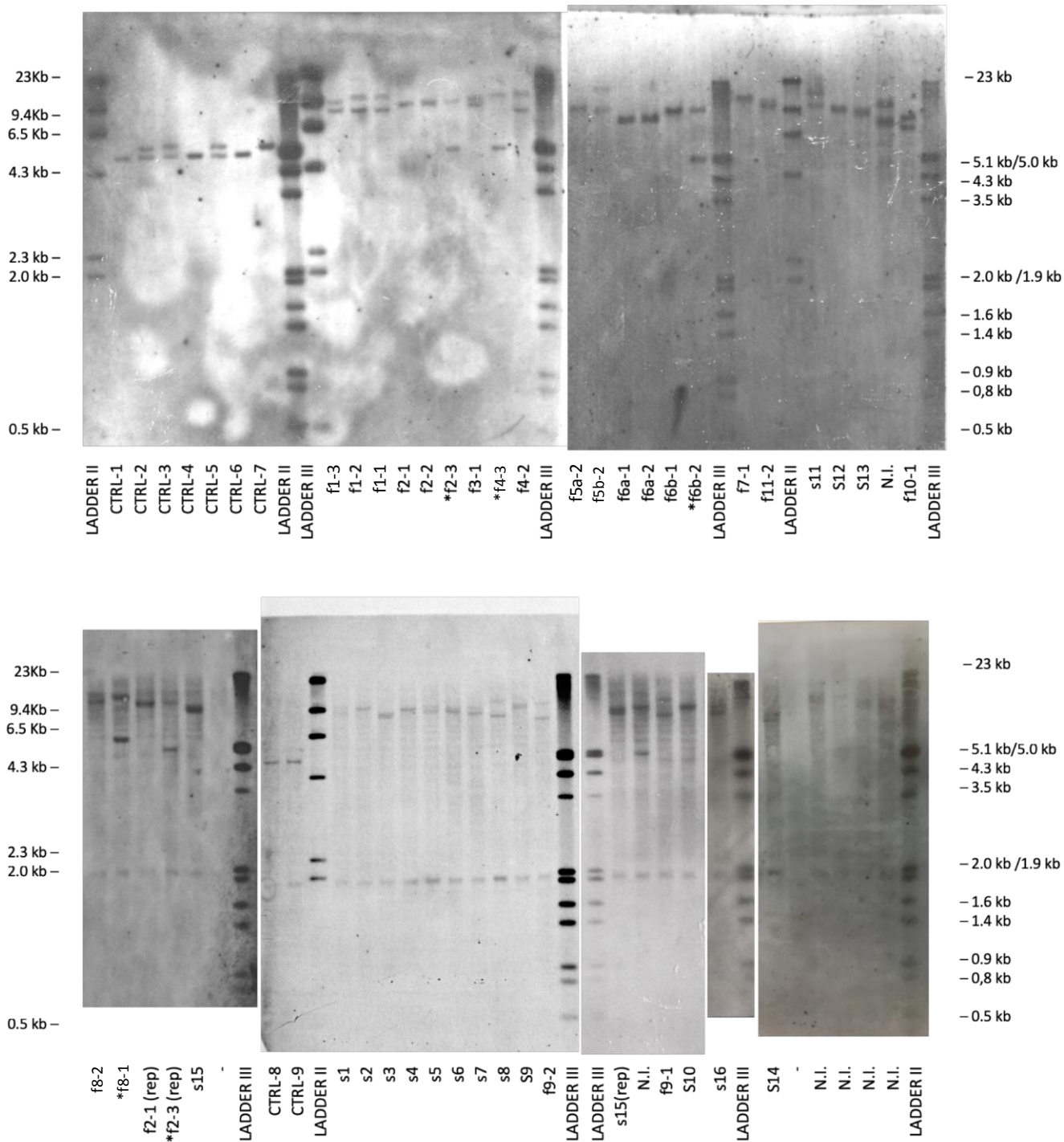
¹Department of Neuromuscular Disease, UCL Queen Square Institute of Neurology and The National Hospital for Neurology and Neurosurgery, London, UK. ²Department of Clinical and Movement Neurosciences, UCL Queen Square Institute of Neurology and The National Hospital for Neurology and Neurosurgery, London, UK. ³Neurogenetics Laboratory, UCL Queen Square Institute of Neurology and The National Hospital for Neurology and Neurosurgery, London, UK. ⁴Department of Biological Sciences, International Islamic University, Islamabad, Pakistan. ⁵Department of Neurology, School of Medicine at Ribeirão Preto, University of São Paulo, Ribeirão Preto, Brazil. ⁶UO Neurologia I, Fondazione IRCCS Istituto Neurologico 'Carlo Besta', Milan, Italy. ⁷IRCCS Mondino Foundation, Pavia, Italy. ⁸Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy. ⁹Department of Psychiatry and Behavioural Sciences, Center for Therapeutic Innovation, University of Miami Miller School of Medicine, Miami, FL, USA. ¹⁰Dr. John T. Macdonald Foundation Department of Human Genetics and John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL, USA. ¹¹These authors contributed equally: Roisin Sullivan, Jana Vandrovцова. ¹²These authors jointly supervised this project: Mary M. Reilly, Henry Houlden. *e-mail: andrea.cortese@ucl.ac.uk; h.houlden@ucl.ac.uk



Supplementary Figure 1

Homozygosity mapping in a consanguineous CANVAS family

(a,b) Homozygosity mapping plot (a) and genomic positions (b) defining the homozygous regions shared by Fam7-1 and Fam7-2. The blue arrow in a points to a ~12-Mb homozygous region on chromosome 4 (bold highlighted in b), which encompass the AAGGG repeat expansion locus.

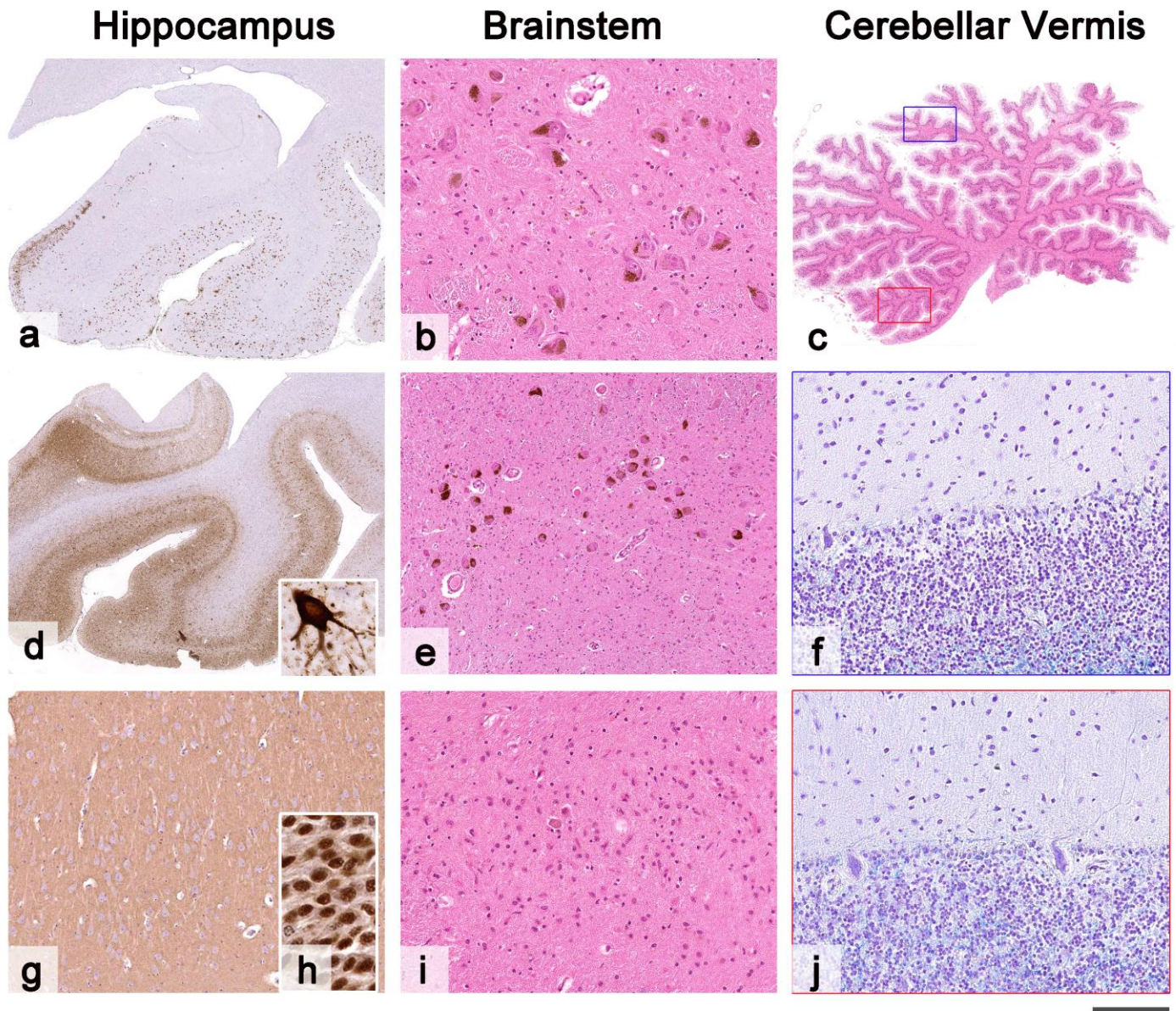


Supplementary Figure 2

Southern blots

Southern blotting of genomic DNA from 18 patients from 11 families and 16 sporadic cases. Patients show two discrete or overlapping bands of 7 to 15 kb. Unaffected siblings, indicated by “*”, carry an expanded allele and one allele in the normal range. In controls (CTRL), one 5-kb band corresponding to the expected size for reference allele (AAAAG)₁₁ is usually observed, but bands

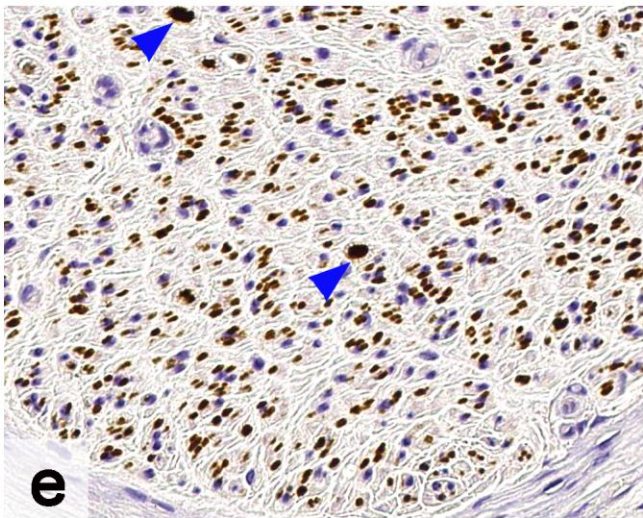
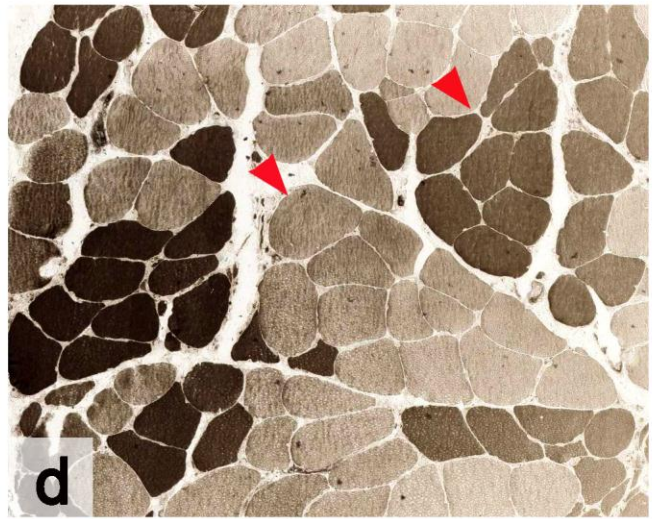
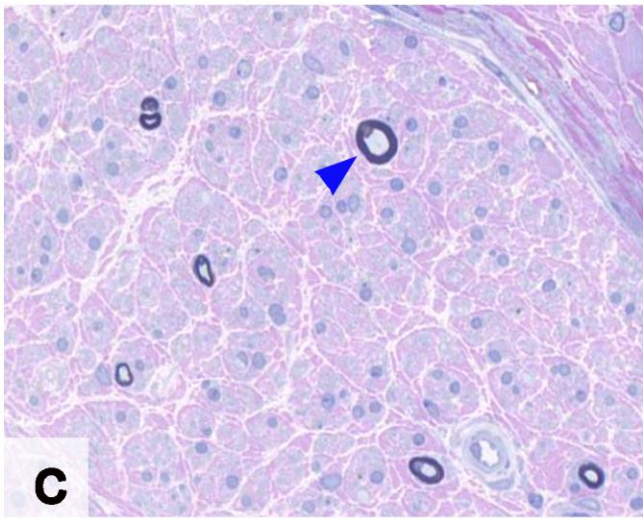
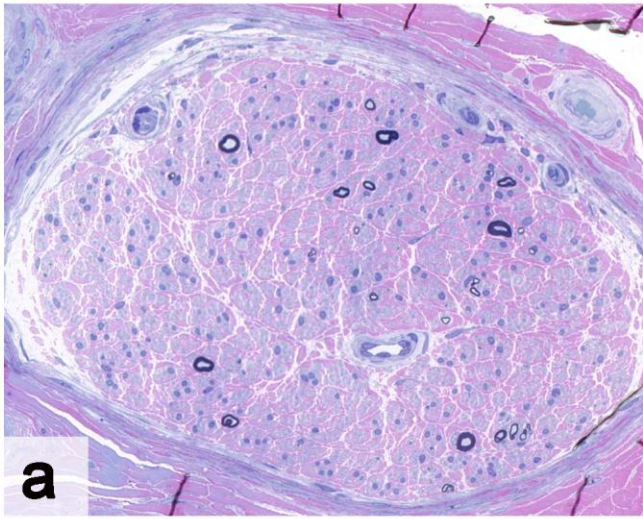
of increased size can also be seen. In some blots, an unspecific band at 2 kb is also observed. Ladder used are DIG-labelled DNA Molecular Weight Marker II (Roche) (LADDER II, left) containing 8 fragments with the following base pair lengths: 125 (not shown), 564, 2,027, 2,322, 4,361, 6,557, 9,416, and 23,130 bp and DIG-labelled DNA Molecular Weight Marker III (Roche) (LADDER III, right) containing 13 fragments with the following base pair lengths: 125 (not shown), 564, 831, 947, 1,375, 1,584, 1,904, 2,027, 3,530, 4,268, 4,973, 5,148, and 21,226 bp. N.I., sample not included in this study; rep, repeated sample.



Supplementary Figure 3

Morphological findings in the brain from patient with CANVAS

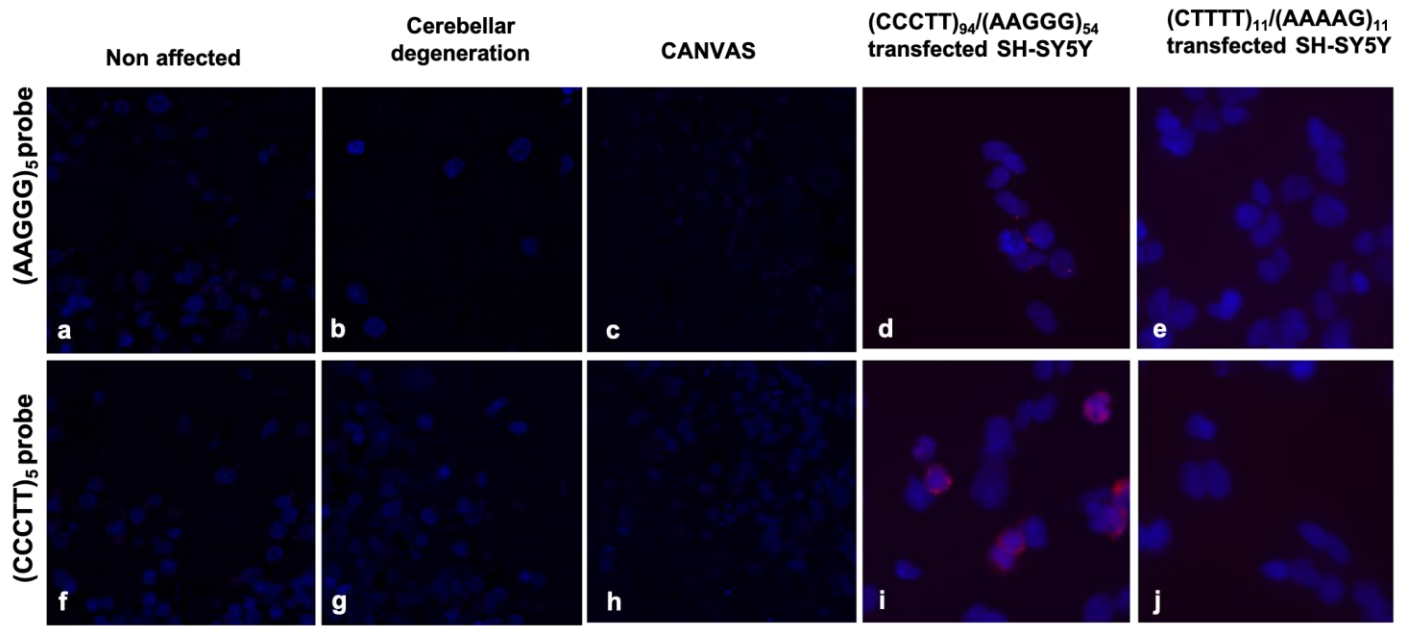
(a) Histological examination reveals some age-related changes in the medial temporal lobe with frequent diffuse parenchymal amyloid- β deposits (a) in the neocortex, subiculum, CA1 and caudate nucleus (Thal phase 3). Rare leptomenigeal amyloid angiopathy (not shown) is also seen in the cerebral hemisphere. Neurofibrillary tangle tau pathology (d) is restricted to the medial temporal lobe with the extent corresponding to Braak and Braak stage IV. There are no α -synuclein (g) or TDP43 (h) positive pathological inclusions in the medial temporal lobe. (b) In the brainstem, the locus coeruleus (b) and pontine base nuclei, and white matter tracts show no apparent neuronal loss. In the midbrain, the *substantia nigra* shows no significant reduction of pigmented neurones (e). In the inferior olivary nucleus, however, there is widespread depletion of neurones and marked chronic gliosis (i). (c) In the cerebellar hemisphere and cerebellar vermis, haematoxylin and eosin stained section shows prominent cortical atrophy with more prominent atrophy seen in the superior cerebellar vermis (c and f, blue square) when compared with the inferior cerebellar vermis (c and j, red square). There are no p62 positive pathological inclusions (f and j) in the cerebellum. Scale bar: 3 mm in a and d; 200 μ m in g and e; 40 μ m in h; 100 μ m in b, i, f and j; 35 mm in c. Stainings were carried out once on patients' samples with appropriate controls according to standard practice and histopathology procedures in an ISO15189 accredited laboratory.



Supplementary Figure 4

Morphological findings in the nerve and muscle biopsy from a representative patient with CANVAS

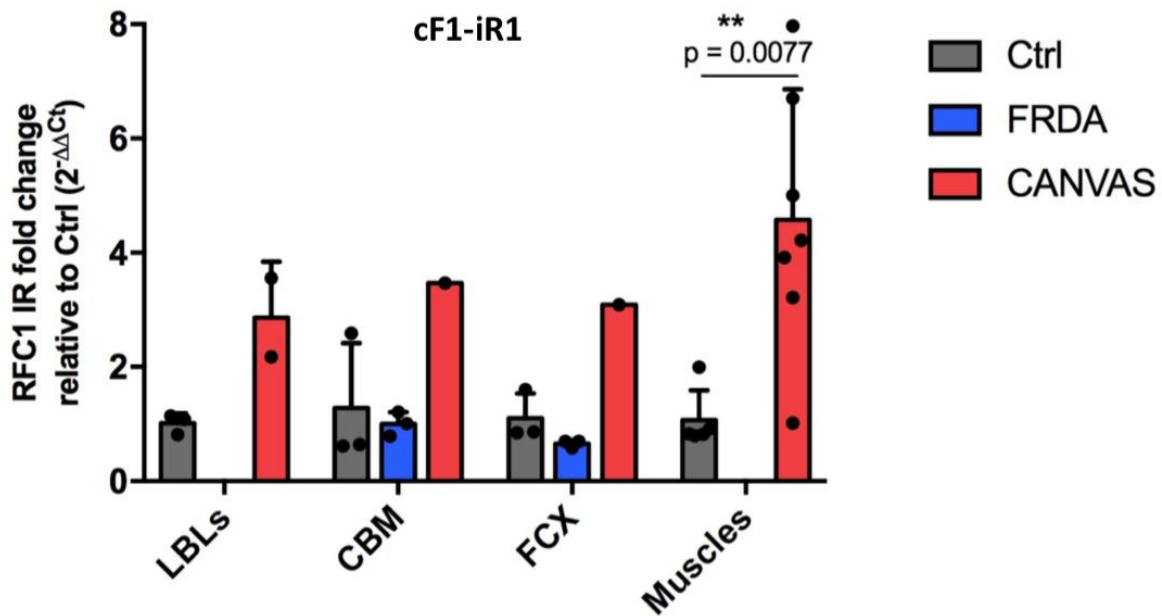
Sural nerve biopsy (**a**, **c**, and **e**). **a** and **c**, semi-thin resin section stained with methylene blue azure-basic fuchsin; **e**, immunostaining for neurofilament with SMI31. Quadriceps skeletal muscle biopsy (**b**, **d** and **f**). **b**, stained with haematoxylin and eosin; **d** and **f**, histochemical staining with ATP 4.2 (**d**) and ATP9.4 (**f**). The nerve biopsy demonstrates marked loss of large and small myelinated fibres (**a**, **c**, and **e**, blue arrowheads) with no active axonal degeneration or any signs of regeneration. The unmyelinated fibres (**e**) are comparably better preserved. The skeletal muscle biopsy reveals angulated fibres and particularly small atrophic fibres (**b**, yellow arrowhead). There is grouping of type 1 and type 2 fibres (**d** red arrows and **f**), in keeping with chronic denervation with reinnervation. Scale bar: 100 μ m in **a**, **d** and **f**; 50 μ m in **c**, **e** and **b**. Stainings were carried out once on patients' samples with appropriate controls according to standard practice and histopathology procedures in an ISO15189 accredited laboratory.



Supplementary Figure 5

RNA fluorescence *in situ* hybridization (FISH)

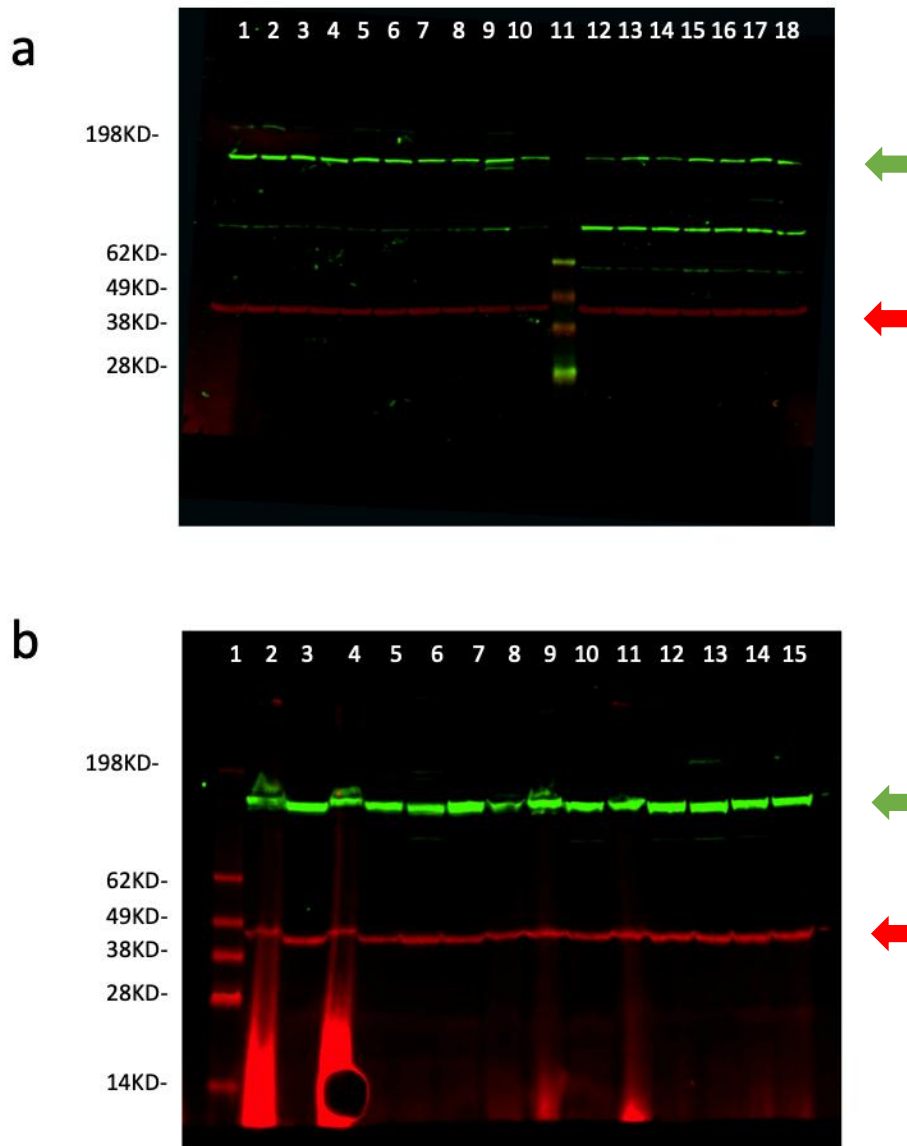
(a-e) No endogenous RNA foci were detected using oligonucleotides (AAGGG)₅ (a, b, c, d, e) or (TTCCC)₅ (f, g, h, i, j) probes targeting transcripts containing sense (TTCCC) or antisense (GGGAA) repeated units in unaffected individuals (a, f), pathological controls (b, g), or CANVAS patient (c, h). As a technical control, SH-SY5Y cells were transfected with plasmids expressing mutant sense (TTCCC)₉₄ (d), mutant anti-sense (AAGGG)₅₄ (i), wild-type sense (TTTTT)₁₁ (e) or wild-type anti-sense (AAAAG)₁₁ (j) and analyzed 24 h after transfection by RNA FISH. RNA foci were detected in d and i (red). Nuclei were stained with DAPI (blue). Scale bar: 20 μm in a, b, c, f, g, h; 10 μm in d, e, i, j. Experiments were repeated independently twice with similar results.



Supplementary Figure 6

***RFC1* intron 2 retention in multiple tissue types in CANVAS**

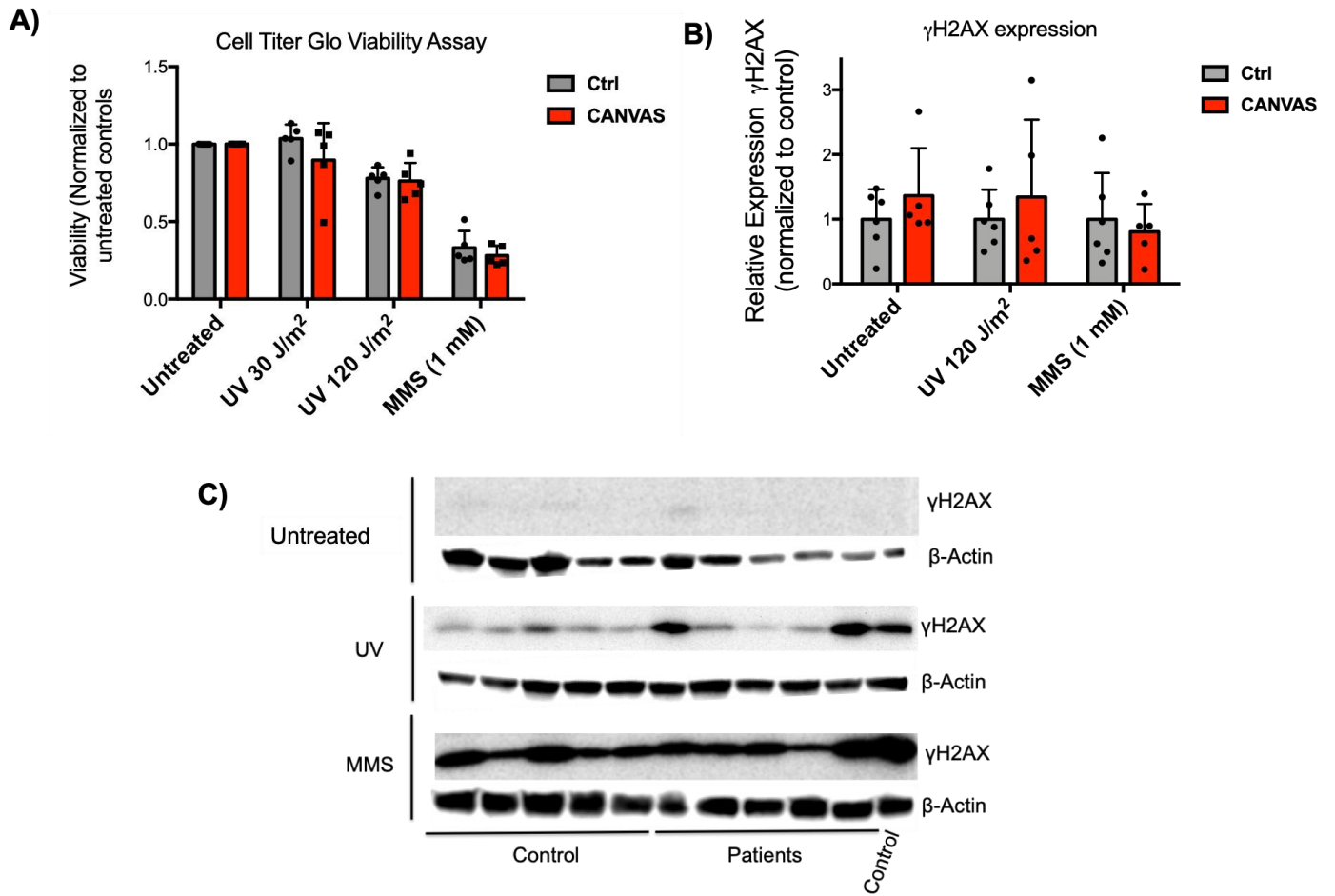
Expression levels of intron 2 in *RFC1* pre-mRNA as measured by qRT-PCR using primers cF1-iR1 in control ($n = 3$) and CANVAS ($n = 2$) lymphoblasts; control ($n = 3$), Friedreich's ataxia ($n = 3$) and CANVAS ($n = 1$) cerebellum and frontal cortex; control ($n = 5$) and CANVAS muscles ($n = 7$). Bar graphs show mean \pm s.d. and data distribution (black dots). Two-tailed t -test was performed to compare the expression level of intron 2 in *RFC1* pre-mRNA in patients versus healthy or disease controls. All experiments were repeated independently twice with similar results. CANVAS, cerebellar ataxia, neuropathy, vestibular areflexia syndrome; CBM, cerebellum; Ctrl, control; FCX, frontal cortex; FRDA, Friedreich's ataxia; IR, intron 2 retention; LBLs, lymphoblasts.



Supplementary Figure 7

Western blots for *RFC1*-encoded protein expression

(a,b) Uncropped gels showing *RFC1*-encoded protein levels as measured by Western blotting using the polyclonal antibody (GTX129291) and normalized to β -actin in (a) control ($n = 5$, lanes 1-5) and CANVAS ($n = 5$, lanes 6-10) fibroblasts, control ($n = 3$, lanes 12-14) and CANVAS ($n = 4$, lanes 15-18) lymphoblasts and (b) CANVAS ($n = 1$, lane 2), control ($n = 3$, lanes 3-5), Friedreich's ataxia ($n = 3$, lanes 6-8) vermis and CANVAS ($n = 1$, lane 9), control ($n = 3$, lanes 10-12) and Friedreich's ataxia ($n = 3$, lanes 13-15) frontal cortex. Green arrow indicates the expected 128 kDa band corresponding to *RFC1*-encoded protein, and the red arrow a 45 kDa band of β -actin. Ladder in lanes 11 (a) and 1 (b) is SeeBlue Plus2 (ThermoFisher).



Supplementary Figure 8

Response to DNA damage in CANVAS

(a) Cell viability after DNA damage inducing treatments of CANVAS ($n = 5$) and control ($n = 5$) fibroblasts. (b) Quantification of γ H2AX protein expression levels normalized to actin and plotted relative to controls. (c) Representative western blot of cells treated with different DNA damage inducing agents. Bar graphs show mean \pm s.d. and data distribution (dots). All experiments were repeated independently twice with similar results.

SUPPLEMENTARY TABLE 1. List of known HUGO Gene Nomenclature Committee (HGNC) genes present in the 1.7Mb region between markers rs6814637 and rs10008483 (chr4:38977921-40712231)

TMEM156
KLHL5
WDR19
RFC1
KLB
RPL9
LIAS
LOC401127
UGDH
UGDH-AS1
SMIM14
BC040333
UBE2K
PDS5A
LOC344967
N4BP2
RHOH
JN120858
CHRNA9
RBM47
MIR4802

SUPPLEMENTAL TABLE 2. Clinical features of patients with CANVAS and late-onset ataxia carrying the recessive AAGGG repeat expansion in *RFC1*

N	ID	gender	Age of onset	Age at examination	Symptom at onset	peripheral neuropathy	cerebellar dysfunction	bilateral vestibular areflexia	autonomic dysfunction	cough	SAPs UL	SAPs LL	CMAPs	cerebellar atrophy	number of repeats (smaller, larger allele)
1	Fam 1-1	male	49	69	unsteadiness	yes	yes	yes	no	no	absent	absent	normal	yes	880,1480
2	Fam 1-2	male	50	65	unsteadiness, pain unsteadiness in the dark, numb feet, pain	yes	yes	yes	no	no	absent	absent	normal	NA	880,1480
3	Fam 1-3	male	48	63	unsteadiness	yes	yes	yes	no	no	absent	absent	normal	no	880,1280
4	Fam 2-1	male	60	78	unsteadiness	yes	yes	no	no	yes	NA	NA	NA	NA	1200,1200
5	Fam 2-2	female	58	69	unsteadiness Unsteadiness, pain	yes	yes	yes	Yes (urinary urgency and retention)	yes	absent	absent	normal reduced in the lower limbs	yes	1200,1200
6	Fam 3-1	male	55	80	unsteadiness	yes	yes	yes	no	no	absent	absent	normal	yes	NA
7	Fam 3-2	female	55	78	unsteadiness	yes	yes	yes	no	no	absent	absent	normal reduced in the lower limbs	yes	1080,1480
8	Fam 4-2	male	66	77	unsteadiness unsteadiness, numb feet	yes	yes	yes	Yes (fecal incontinence)	no	absent	absent	normal	NA	880,1800
9	Fam 5a-1	male	59	65	unsteadiness	yes	yes	yes	no	no	absent	absent	normal	yes	NA
10	Fam 5a-2	female	53	54	unsteadiness, numb feet	yes	no	no	no	no	NA	NA	NA	no	1000,1000
11	Fam 5b-2	male	60	62	unsteadiness, falls	yes	yes	yes	no	no	absent	absent	normal	yes	1000,2000
12	Fam 6a-1	female	67	77	unsteadiness in the dark, pain numb hands, later unsteadiness and oscillopsia	yes	yes	yes	no	no	reduced	absent	normal	yes	600,700
13	Fam 6a-2	female	69	74	Dizziness. Later unsteadiness and falls	yes	yes	yes	no	yes	absent	absent	normal	yes	600,700
14	Fam 6b-1	male	40	70	Dizziness. Later unsteadiness and falls	yes	yes	yes	no	no	absent	absent	normal	yes	880,880
15	Fam 7-1	female	45	50	unsteadiness	yes	yes	yes	no	no	reduced	reduced	normal	yes	1200,1200
16	Fam 7-2	male	43	48	unsteadiness	yes	yes	yes	no	no	reduced	reduced	normal	yes	NA
17	Fam 8-2	female	NA	58	unsteadiness	yes	yes	no	no	no	absent	absent	normal	yes	NA
18	Fam 8-3	female	NA	52	unsteadiness unsteadiness in the dark, numb hands	yes	yes	no	Yes (abnormal pupil reactivity)	no	absent	absent	normal	yes	1400,1800
19	Fam 9-1	male	45	68	unsteadiness	yes	no	yes	Yes (erectile dysfunction)	yes	absent	reduced	normal	NA	880,1480
20	Fam 9-2	female	45	64	Numb feet	yes	no	yes	no	yes	reduced	reduced	normal	no	600,1200
21	Fam 10-2	female	55	57	unsteadiness	yes	no	no	no	no	reduced	reduced	normal	NA	400,800
22	Fam 10-1	male	50	57	unsteadiness	yes	no	no	no	yes	reduced	absent	normal	NA	NA
23	Fam 11-2	female	45	64	unsteadiness	yes	yes	yes	no	yes	absent	absent	normal	yes	880,1100
24	s1	male	65	72	unsteadiness	yes	yes	no	Yes (erectile dysfunction)	yes	absent	absent	normal	yes	800,1000

25	s2	female	48	65	unsteadiness, numb hands	yes	yes	no	Yes (constipation, urinary incontinence, impaired regulation of blood pressure)	yes	reduced	absent	normal	yes	1000,1000
26	s3	male	45	62	unsteadiness in the dark, numb feet	yes	no	yes	no	yes	absent	absent	reduced in the lower limbs patchy reduction of conduction velocities	NA	700,700
27	s4	female	62	72	unsteadiness, numb feet	yes	yes	no	Yes (urinary retention)	yes	absent	absent	normal	yes	880,880
28	s5	female	54	74	unsteadiness, pain	yes	no	no	Yes (urinary urgency)	yes	reduced	absent	normal	yes	880,1000
29	s6	male	49	55	unsteadiness, numbness	yes	no	no	Yes (constipation)	no	reduced	absent	normal reduced in the lower limbs	yes	880,1000
30	s7	female	48	62	numb hands	yes	no	yes	no	yes	absent	absent	normal	yes	800,1000
31	s8	female	60	82	numb feet	yes	yes	yes	no	no	absent	absent	normal	no	700,1300
32	s9	male	59	74	unsteadiness	yes	yes	yes	no	yes	absent	absent	normal reduced in the lower limbs	yes	1000,1100
33	s10	male	69	72	unsteadiness numb feet, dysesthesia	yes	yes	yes	no	yes	absent	absent	normal	yes	1150/1150
34	s11	male	45	55	NA	yes	no	no	no	no	NA	NA	NA	NA	1100,1800
35	s12	male	NA	NA	NA	yes	yes	yes	no	no	absent	absent	normal	yes	1000,1000
36	s13	female	45	65	numbness	yes	yes	no	Yes (urinary dysfunction) Yes (urinary urgency, gastro-intestinal dysmotility)	no	absent	absent	normal	no	800,900
37	s14	female	65	76	unsteadiness in the dark, pain	yes	yes	no	no	yes	absent	absent	normal	yes	900,1200
38	s15	male	65	71	unsteadiness in the dark	yes	yes	yes	no	no	absent	absent	normal	no	1100,1100
39	s16	female	69	73	unsteadiness, vertigo	yes	yes	yes	no	yes	absent	absent	normal	yes	880, 1100
40	s17	male	55	63	Unsteadiness, numb hands	yes	yes	no	no	no	absent	absent	normal	NA	NA
41	s18	male	50	60	unsteadiness, slurred speech	yes	yes	yes	no	no	absent	absent	normal	yes	NA
42	s19	male	45	55	unsteadiness in the dark	yes	yes	yes	no	no	absent	reduced	normal	yes	NA
43	s20	female	50	65	unsteadiness in the dark	yes	yes	yes	no	yes	absent	absent	normal	NA	NA
44	s21	female	NA	72	NA	yes	yes	yes	no	no	absent	absent	normal	yes	NA
45	s22	female	40	47	unsteadiness	yes	yes	no	no	no	absent	absent	normal	NA	NA
46	s23	female	70	76	unsteadiness	yes	yes	no	no	no	absent	absent	normal	yes	NA
47	s24	male	69	73	unsteadiness plodding/flatfooted, speech problems	yes	yes	no	no	no	absent	absent	normal	yes	NA
48	s25	male	35	40	age 35	yes	yes	no	no	yes	absent	absent	normal	yes	NA
49	s26	female	35	38	unsteadiness,	yes	yes	no	no	yes	reduced	absent	reduced in the	yes	NA

50	s27	female	53	54	vertigo and speech also early oscillopsia, unsteadiness, numbness	yes	yes	no	no	no	absent	absent	normal	yes	NA
51	s28	female	60	67	unsteadiness unsteadiness, dysarthria, numbness	yes	yes	no	no	no	absent	absent	reduced in the lower limbs	NA	
52	s29	male	54	64	unsteadiness	yes	yes	no	Yes (constipation, xerophthalmia)	no	absent	absent	normal	yes	NA
53	s30	female	65	73	unsteadiness	yes	yes	no	no	no	absent	absent	normal	NA	NA
54	s31	female	58	61	unsteadiness	yes	yes	no	no	no	NA	NA	NA	NA	NA
55	s32	female	50	64	unsteadiness in the dark, numb hands	yes	yes	yes	Yes (abnormal blood pressure control, abnormal bowel motility)	no	absent	absent	normal	yes	NA
56	s33	male	48	51	unsteadiness in the dark, pain	yes	no	no	no	yes	absent	absent	normal	no	NA

cMAP: compound motor action potential; LL lower limbs; SAP:sensory action potentials, UL upper limbs,

SUPPLEMENTAL TABLE 3. Primers sequences and thermocycling conditions

	Primers	Reagents	Thermocycling conditions
Probe generation	Fw: ATTAGGTGTCTGGTGAGGGC Rv: GAAGAATGGCCCCAAAAGCA	Faststart Master Mix 2X (Roche) Probe synthesis mix 20x (Roche) Primers 0.5 μM Plasmid DNA 50 ng	95°C 4 min [95°C 30 s 63°C 30 s 72°C 60 s] x 35 cycles
Short-range flanking PCR	Fw: TCAAGTGATACTCCAGCTACACCGTTGC Rv: GTGGGAGACAGGCCAATCACTTCAG	Faststart Master Mix 2X (Roche) Primers 0.5 μM gDNA 50 ng	72°C 5 min 95°C 4 min [95°C 30 s 59°C 30 s 72°C 60 s] x 35 cycles
Long-Range flanking PCR and cloning	Fw TCAAGTGATACTCCAGCTACACCGTTGC Rv GTGGGAGACAGGCCAATCACTTCAG	Phusion Flash High-Fidelity PCR Master Mix 2X (Thermo-Fisher) Primers 0.5 μM DMSO 3% gDNA 50 ng	72°C 5 min 98°C 3 min [98°C 10 s 65°C 15 s - Each cycle decreasing by 0.5°C 72°C 3 min] X18 cycles [98°C 10s 57°C 15s 72°C 3 min] X18 cycles
Repeat-primed PCR	Fw FAM-TCAAGTGATACTCCAGCTACACCGT Anchor CAGGAAACAGCTATGACC (AAAAG)₁₁ allele Rv1 CAGGAAACAGCTATGACCAACAGAGCAAGACTCTGT TTCAAAAAGAAAAGAAAAGAAAAGAAAA Rv2 CAGGAAACAGCTATGACCAACAGAGCAAGACTCTGT TTCAAAAAGAAAAGAAAAGAAAAGAAAA Rv3 CAGGAAACAGCTATGACCAACAGAGCAAGACTCTGT TTCAAAAAGAAAAGAAAAGAAAAGAAAA (AAAGG)_{exp} allele Rv1 CAGGAAACAGCTATGACCAACAGAGCAAGACTCTGT TTCAAAAAGGAAAGGAAAGGAAAGGAAA Rv2	Phusion Flash High-Fidelity PCR Master Mix 2X (Thermo-Fisher) Fw Primer 0.5 μM Anchor 0.5 μM Rv primers (Rv1:Rv2:Rv3=1:1:1) 0.05 μM DMSO 3% gDNA 50 ng	72°C 5 min 98°C 3 min [98°C 10s 65°C 15s 72°C 60s] x35 cycles 72°C 5 min

	<p>CAGGAAACAGCTATGACCAACAGAGCAAGACTCTGT TTCAAAAAGGAAAGGAAAGGAAAGGAAA</p> <p>Rv3 CAGGAAACAGCTATGACCAACAGAGCAAGACTCTGT TTCAAAAAGGAAAGGAAAGGAAAGGAAA</p> <p>(AAGGG)_{exp} allele</p> <p>Rv1 CAGGAAACAGCTATGACCAACAGAGCAAGACTCTGT TTCAAAAAGGGAAGGGAAGGGAAGGGAA</p> <p>Rv2 CAGGAAACAGCTATGACCAACAGAGCAAGACTCTGT TTCAAAAAGGGAAGGGAAGGGAAGGGAA</p> <p>Rv3 CAGGAAACAGCTATGACCAACAGAGCAAGACTCTGT TTCAAAAAGGGAAGGGAAGGGAAGGGAA</p>		
qRT-PCR	<p>cF1 (Fw) CATTCCGAAATTCTTTGGAGTA cR1 (Rw) CGGGAGCTATTTACCTTGAT cF2 (Fw) GGCAGTTGCATGAAGATGAAG cR2 (Rv) CCTTTCGAGCCTTTTTGGTC iR1 (Rv) TCAATGCAAAATTATACCCAGA</p>	<p>SYBR™ Green PCR Master Mix 2x Primers 0.4 μM cDNA from 500 ng RNA</p>	<p>95°C 10 min [95°C 15s 60°C 60s]x 40 cycles</p>
rs11096992 and rs2066790 genotyping	<p>rs11096992 Fw/sequencing TGGCTTAAATGATCTTTTCCCG Rv CACCAATAAACTTACACCCACA</p> <p>rs2066790 Fw CCTGAGGTGTGTGGCTTTAG Rv/sequencing TCAGGACTTACAGACTTTGGGA</p>	<p>Faststart Master Mix 2X (Roche) Primers 0.5 μM gDNA 50 ng</p>	<p>95°C 4 min [95°C 30 s 65°C 30 s - 72°C 45 s] X8 cycles</p> <p>95°C 30 s 65°C 30 s - Each cycle decreasing by 0.5°C 72°C 45 s] X16 cycles</p> <p>[95°C 30 s 55°C 30 s - 72°C 45 s] X16 cycles</p> <p>72°C 5 min</p>

PCR: polymerase chain reaction; qRT-PCR quantitative real-time PCR; Fw: forward; Rv: reverse, DMSO dimethyl sulfoxide; gDNA genomic DNA