Supplementary materials to

Hypomorphic mutations in *POLR3A* are a frequent cause of sporadic and recessive spastic ataxia.

Table of Contents

I. Splicing effect of intronic mutations in *POLR3A*

Analysis of the c.1290-2A>G mutation affecting the highly-conserved splice acceptor sequence of exon 10 in patient F3-1 revealed an additional larger band (261bp) not observed in control brain cDNA (Supplementary Fig. 7A). Sequencing of the 261 bp amplicon confirmed a transcript including 25 base pairs of intron nine caused by cryptic splice site activation in intron nine predicting a frameshift and truncation of the protein after 26 amino acids. In patient F13-1, the c.2247+2T>G caused the loss of the canonical splice donor site after exon 16. Amplification of leukocytes cDNA from this patient revealed an aberrant of 574bp band (Supplementary Fig. 7B). Sanger sequencing of this aberrant amplicon revealed that exon 16 lacked the last 109 bases due to the activation of a cryptic splice site within this exon. At the protein level, the truncation of exon 16 predicted a frameshift resulting in a stop codon after two amino acids from the wrong frame (p.K713Kfs*3). In patient F15-1 (c.1048+5G>T), cDNA amplification showed a larger aberrant band of 781bp on the gel electrophoresis (Supplementary Fig. 7C). Sequencing of the larger band demonstrated the inclusion of sequence from intron seven due to inactivation of the canonical splice donor site in exon seven. The inclusion of intronic sequence in the transcript resulted in the frameshift and a premature termination codon after 26 extraneous amino acids (p.G350Gfs*27).

II. Protein structural modeling

Putative effects of amino acids changes on the crystal structure of human RPC1 are depicted in Fig. 3B and Supplementary Fig. 8A.

p.Cys109Ser: Cys109 belongs to a set of highly conserved (over 380 members of the evolutionary family of the RPC1) cysteines (109, 112, 156, and 159) that form a putative Zn^{2+} ion binding site. This binding site is conserved in the structural template (RNA Polymerase II from

Schizosaccharomyces pombe) and indeed, it was co-crystalized with a Zn^{2+} ion in that position. The mutation of a cysteine residue to serine may abolish the binding of Zn^{2+} ions.

p.Leu356Pro: The highly conserved Leu356 residue is located in an alpha helix at the interface between RPC1 and RPB2. The local fold of the helix could be modified by its mutation into a proline residue. Indeed, the latter can induce torsional angles not compatible with an alpha helix. This modification could alter the interaction between the two proteins.

p.Asp372Asn: Asp372 lies on the protein surface at the interface between RPC1 and one of its accessory proteins (RPABC5), where it putatively interacts with His65. This interaction forms part of a local network of charged residues that includes three arginine residues: 479, 487 and 469. The mutation into a neutral asparagine residue may affect the electrostatic equilibrium in the region.

p.Leu454Phe: Leu454 belongs to a hydrophobic region of the protein that includes the amino acids: Val80, Leu509, Ala478, and Leu390. Leu454 is highly conserved across the evolutionary family. The substitution with a bulkier aromatic phenylalanine residue may produce steric impediments with surrounding residues. The loss/gain of a phenyl group in pairwise interactions between hydrophobic side chains is related with a ~0.2 kcal/mol energy change, as measured in systematic studies(Zhu *et al.*, 1993), thus implying that the gain of the phenyl group of Phe may introduce rigidity in a region putatively involved in the interaction with other subunits (RPB2) as observed from the models.

p.Ala515Val: Ala515 is a highly conserved residue that forms part of the interface between RPC1 and the accessory protein RPB2. The loss/gain of a methyl group in pairwise interactions between hydrophobic side chains is related with a ~0.7 kcal/mol energy change, as measured in systematic

studies(Zhu *et al.*, 1993), thus implying that the gain of the methyl group of Val may destabilize the region and may affect the interaction between RPC1 and the accessory proteins.

p.Val1033Ala: Val1033 is a highly conserved residue that is involved in a network of hydrophobic interactions. The loss of the methyl group of Val may destabilize the region, probably changing the dynamic properties of RPC1.

p.Glu1261Lys: Glu1261 is a highly conserved negatively charged residue at the core of a network of salt bridges at the interface between RPC1 and the RPB5 protein. Moreover, Glu1261 putatively forms a salt bridge with a positively charged amino acid of the RPB5 accessory protein. A change in the charge of the amino acid from negative to positive due to exchange of a glycine residue by lysine may destabilize not only the local salt-bridge network but also the interaction of RPC1 with RPB5.

III. Additional clinical findings in identified POLR3A mutation carriers

Dental abnormalities: Seven cases (F1-3/-5/-7/-8, F2-1, F3-1, and F14-1) received a detailed dental and oro-facial examination that revealed mild abnormalities in most patients. With the exception of F1-8, all examined patients experienced an early loss of teeth following a reduction of periodontal bone anchorage due to severe and aggressive periodontal disease in young age. Furthermore, three family members of family F1 demonstrated developmental defects of dental and periodontal formation manifesting by a failure of eruption and hypoplasia of upper incisors. Interestingly, all patients showed subtle to mild signs of an oro-facial abnormality with tendency to class III malocclusion due to sagittal hyperplasia of the anterior maxillo-facial complex. An additional transversal maxillary deficit with cross bite tendency was observed in these subjects with the exception of F2-1. The same cases also showed subtle to mild signs of mandibular deviation and crepitation during mouth opening.

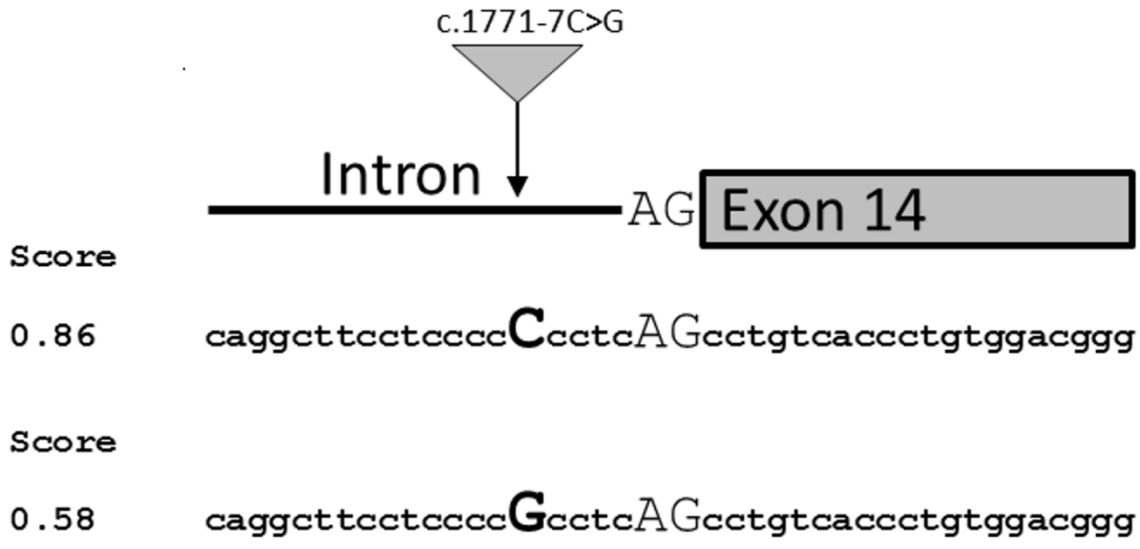
Cognitive assessments: Eight patients (F1-3/-5/-7/-8, F2-1, F3-1, F13-1, and F14-1) received a detailed neuropsychological evaluation including evaluation of verbal memory, visuo-spatial memory, short term memory, working memory, semantic and phonemic verbal fluency, sustained attention, interference, and fluid intelligence. Apart from a decreased digit span forward (short term memory) in the oldest patient (F3-1, 68 years) none of the patients revealed abnormalities in any of the test applied.

Imaging findings: Brain imaging showed in most cases cervical cord atrophy and in some slightly hypoplastic corpus callosum. However, we did not observe pattern previously described for hypomyelinating leukodystrophy, which is characterized by generalized supra- or infratentorial hypomyelination, i.e., T2-hyperintense cerebellar and cerebral white matter, T2 hypointensities (indicating relatively preserved myelination) of optic radiation/posterior limb of internal capsule, anterolateral part of the thalamus, globus pallidus and dentate nucleus.(Steenweg *et al.*, 2010; Takanashi *et al.*, 2014; Wolf *et al.*, 2014; La Piana *et al.*, 2016) Instead, in 11/12 patients receiving specific MRI including a 3D-fluid attenuated inversion recovery (FLAIR) MRI sequence, a striking bilateral hyperintensity along the entire superior cerebellar peduncles ranging from the cerebellar dentate nucleus to the midbrain just below the red nucleus was found. In each case the hyperintensity in the FLAIR images had a hypointense correlate in the T1 weighted images indicating secondary myelin degradation rather than hypomyelination (Fig. 4) (Schiffmann and van der Knaap, 2009). In the absence of other previously described POLR3A-associated cerebellar abnormalities (Steenweg *et al.*, 2010; Takanashi *et al.*, 2014; Wolf *et al.*, 2014; La Piana *et al.*, 2016), the MRI sign of degradation of the superior cerebellar peduncle has not been described in *POLR3A* mutations so far and may represent the structural correlate to the cerebellar symptoms developing in the disease course in our patient group.

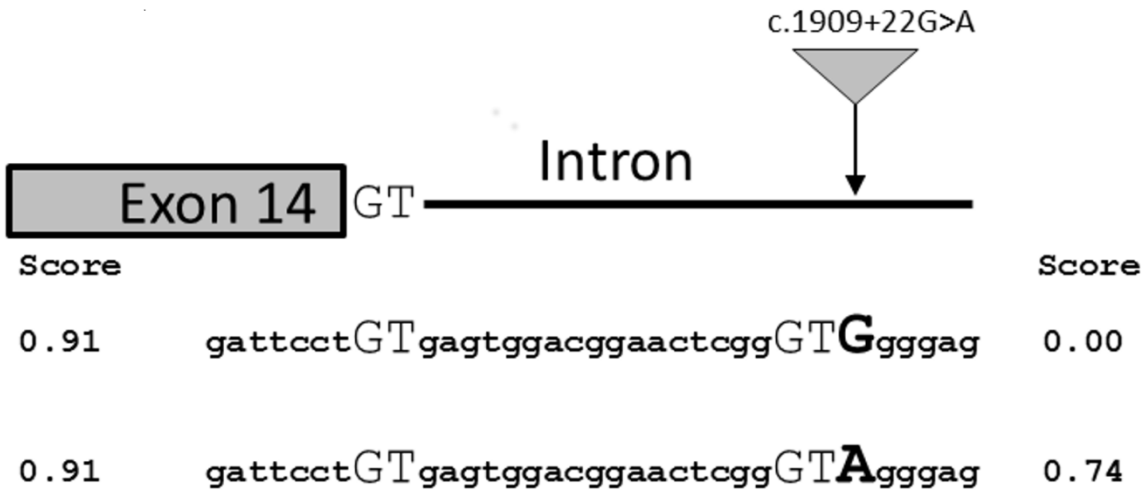
Phenotype characterization of homozygous c.1909+22G>A (F19-1) and homozygous c.3098T>C (F20-1) mutation carriers: Patient F19-1 is one of three children born to non-related parents of European-American descent. The variant was confirmed at heterozygous state in the unaffected mother; the father's DNA was not available for testing. Clinically signs and symptoms of a childhood-onset axonal neuropathy predominated. Marked gait ataxia was noted, the degree of which seemed disproportionate to the only mild peripheral sensory involvement of a loss of vibration sensation up to her ankles. Motor impairment was mild in the lower extremities. Patient F20-1 born from a consanguineous union had delayed motor milestones with clumsy walking only at age 2 years. A diagnosis of pure cerebellar ataxia was made and evolution was slow until the age of 20 years when he was evaluated for hypogonadotrophic hypogonadism, a diagnosis of Gordon Holmes syndrome was made at that point. IQ at that time was 100 (WAIS), MRI showed a marked inferior vermian-cerebellar hypoplasia with gracile aspect of pons and brainstem. Ataxia was progressive in his late twenties and early thirties with loss of ambulation occurring at age 34 years. Progressive cognitive decline became evident from age 33 years onward with behavioural disturbances, cognitive function was severely affected by age 39 years. Upon further clinical evaluation at that age, a severe cerebellar ataxia was noted with mild pyramidal tract signs without spasticity nor lower motor neuron or peripheral nerve involvement. Brain MRI at age 38 years showed pronounced central and cortical atrophy both supra- and infra-tentorial, a thin corpus, atrophy of pons and brainstem and diffuse periventricular white matter abnormalities. The patient died in a residential care facility at the age of 41 years.

IV. Supplemental Figures

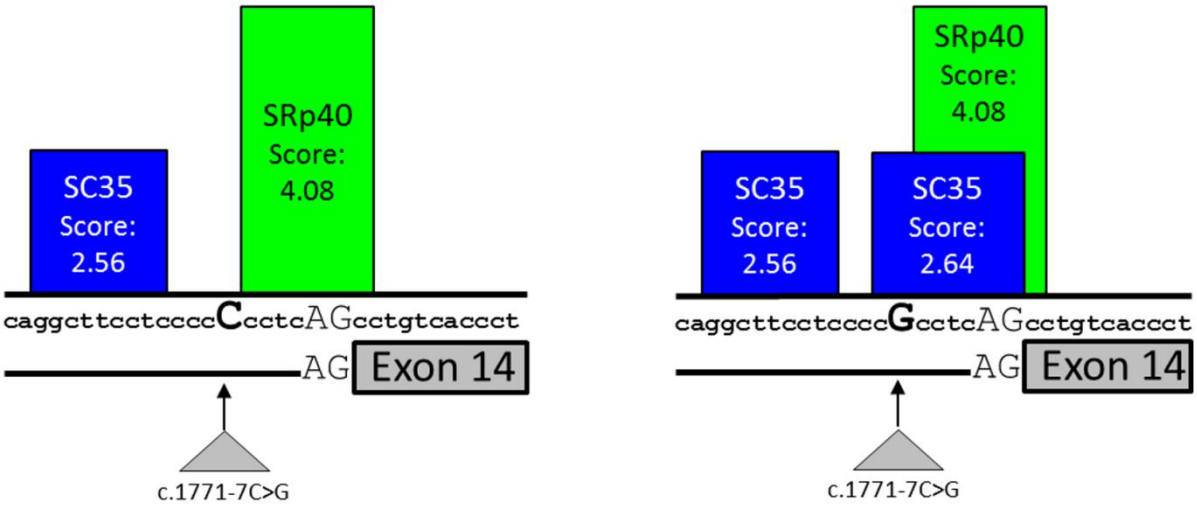
A



B

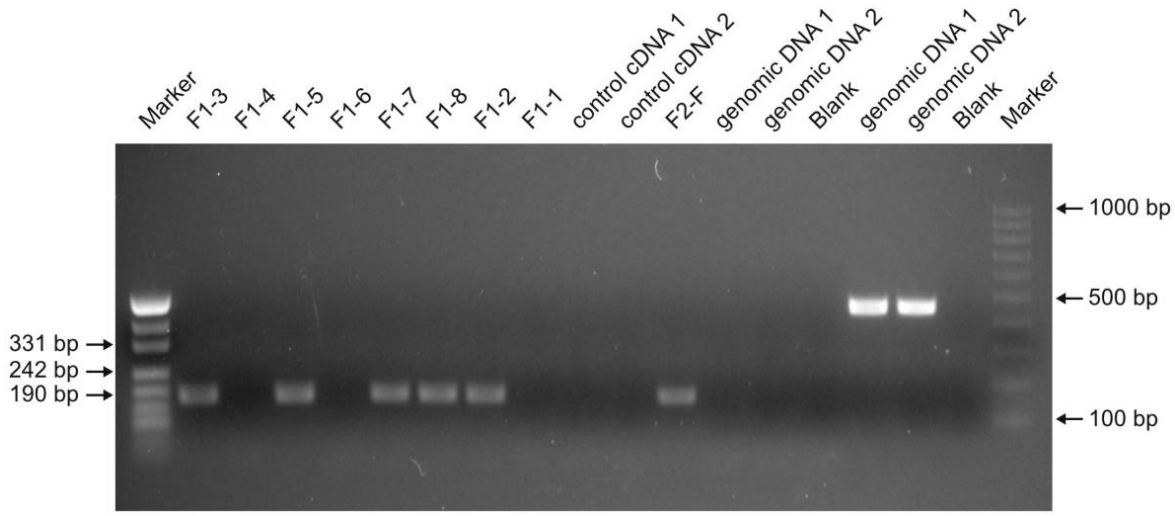


C



Supplementary Figure 1: Predicted effect on splicing due to c.1771-7C>G and c.1909+22G>A.

A. Predicted effect of c.1771-7C>G on acceptor site of exon 14. The rare G allele reduced the predicted splicing score of the acceptor site of exon 14. B. Predicted effect of c.1909+22G>A on donor site of exon 14. The rare A allele activates a cryptic splice site with a score close to the score of the native donor site of exon 14. C. The rare G allele in c.1771-7C>G creates a binding site for the splicing factor SC35. This binding site overlaps with the binding site for SRp40 that has been reported to be important in the splicing of exon 14.



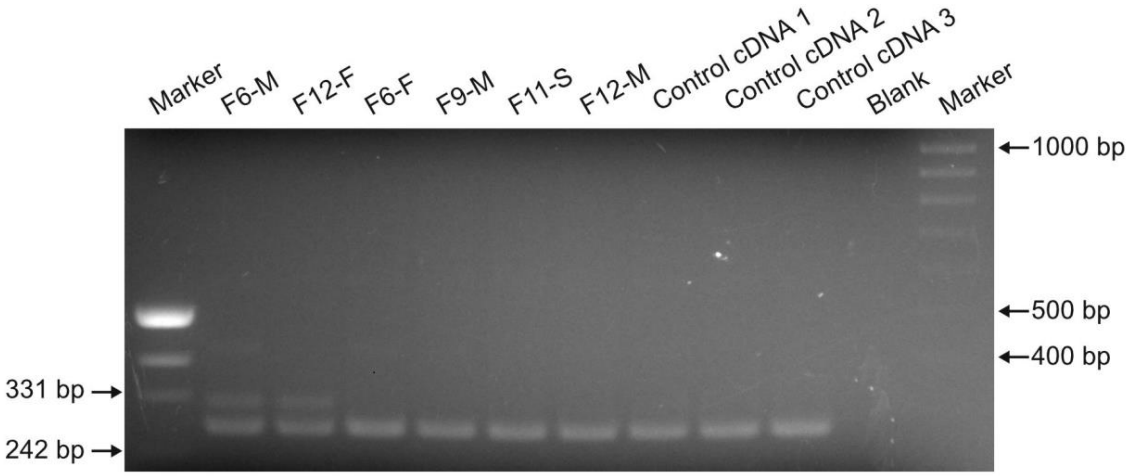
Supplementary Figure 2: Expression of the aberrant POLR3A transcript.

PCR amplification of the cDNA transcript specific to the aberrant splice variant produced by the rare allele A of c.1909+22G>A. Noncarriers of this allele (F1-4, F1-6, F1-1, and controls) as well as the genomic DNA do not amplify the expected 180bp PCR-product. Primers surrounding *POLR3A* exon 14 (471bp) were used to show good quality of the genomic DNA.

Marker	Position	F1-5		F4-1		F5-1		F6-1		F9-1		F10-1		F11-1		F12-1		F13-1		F14-1		F16-1		F16-2		F17-1		F18-1	
D10S1652	10:64,407,589-64,407,762	164	168	172	180	164	166	168	174	174	170	174	160	160	174	166	172	166	172	162	168	166	168	166	166	164	174	174	178
D10S537	10:72,395,349-72,395,474	126	130	132	134	126	132	130	132	122	130	122	134	128	130	130	136	130	136	122	130	132	136	122	132	122	136	126	130
D10S606	10:73,369,629-73,369,863	236	236	234	234	236	230	232	218	236	236	236	232	226	232	230	230	238	230	234	238	232	236	234	232	230	230	230	238
D10S605	10:79,251,427-79,251,530	104	106	96	106	104	96	96	106	100	102	100	96	98	96	88	100	96	100	96	104	100	102	102	100	96	96	102	104
	10:79,769,273C>T																												
D10S2327	10:80,712,049-80,712,248	197	197	197	209	209	213	209	209	197	197	197	217	209	217	209	209	197	197	217	213	197	217	217	197	197	209	197	213

Supplementary Figure 3: Haplotype analysis using short tandem repeats (STR) markers

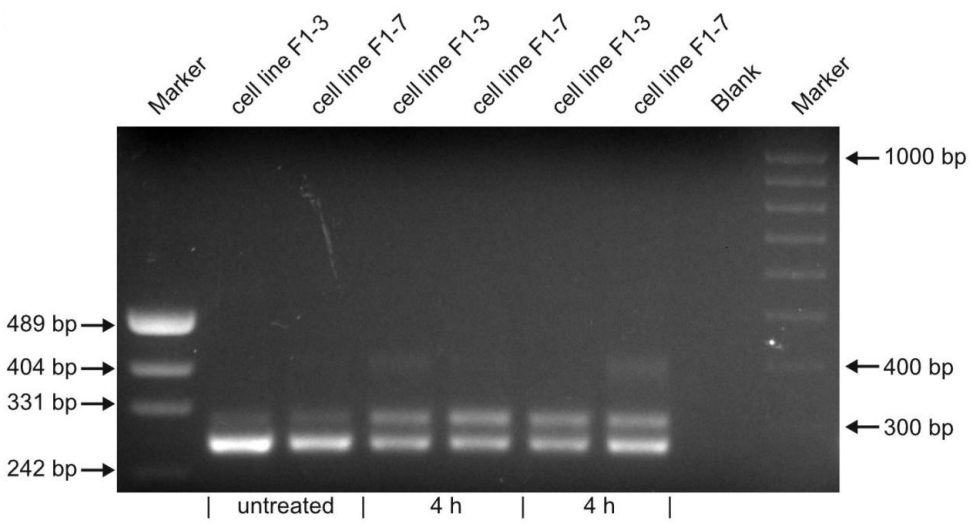
Haplotype analysis of the region surrounding c.1909+22G>A (10:79,769,273C>T) in patients presenting spastic ataxia phenotype and compound heterozygous mutations in *POLR3A* including c.1909+22G>A. Name of short tandem repeat (STR) markers is given on the left. Location of c.1909+22G>A is shown in grey. Genomic position for each STR marker is shown in second column. No common haplotype is observed between patients.



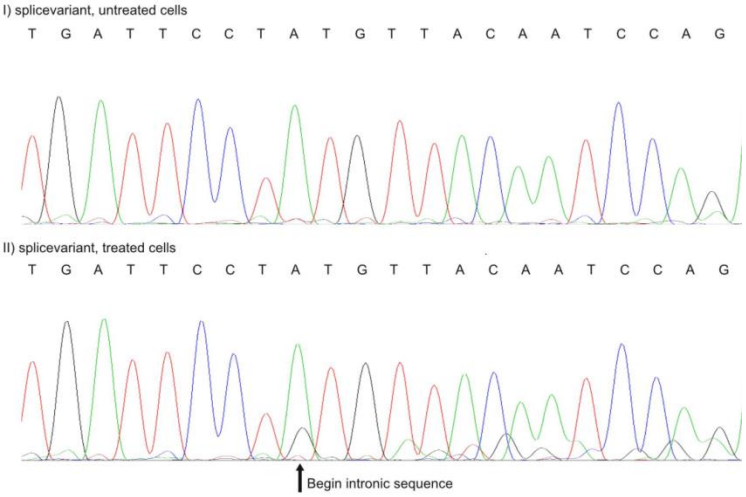
Supplementary Figure 4: Activation of cryptic splice site at c.1909+22G>A in healthy heterozygous carriers.

A primer pair flanking exon 14 was used to amplify leukocyte derived cDNA template from two healthy heterozygous carriers of the rare A allele of c.1909+22G>A (F6-M, F12-F), as well as different healthy non-carrier relatives of these two persons (F6-F, F9-M, F11-S, F12-M) and independent controls. The expected aberrant larger fragment (298bp) is seen in carriers of c.1909+22G>A, whereas this band is missing in all non-carriers of the splice site mutation.

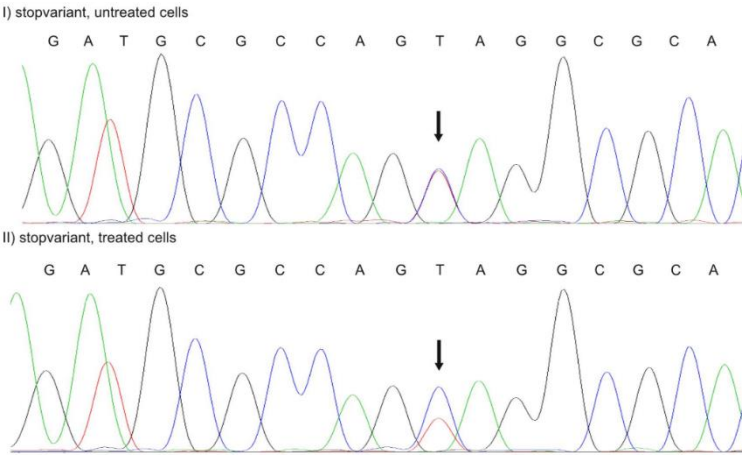
A


Marker
cell line F1-3
cell line F1-7
cell line F1-3
cell line F1-7
cell line F1-3
cell line F1-7
Blank
Marker
1000 bp
489 bp
404 bp
331 bp
242 bp
400 bp
300 bp
untreated
4 h
4 h

B


I) splicevariant, untreated cells
T G A T T C C T A T G T T A C A A T C C A G
II) splicevariant, treated cells
T G A T T C C T A T G T T A C A A T C C A G
Begin intronic sequence

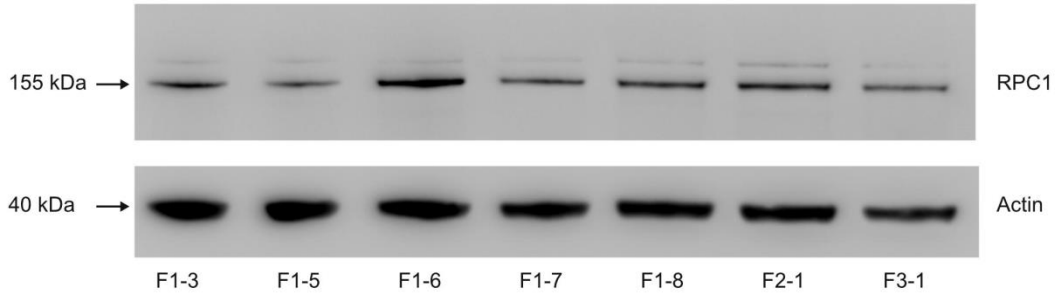
C


I) stopvariant, untreated cells
G A T G C G C C A G T A G G C G C A
II) stopvariant, treated cells
G A T G C G C C A G T A G G C G C A

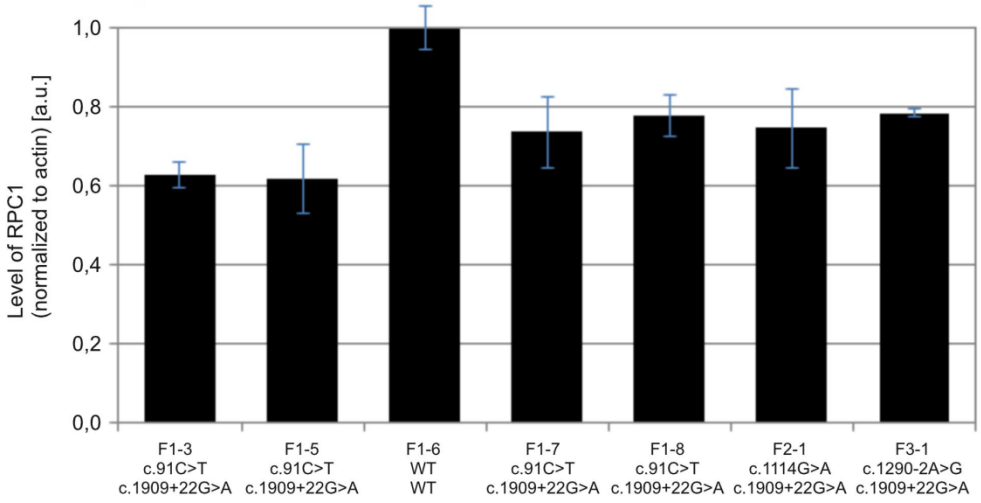
Supplementary Figure 5: Inhibition of nonsense-mediated mRNA decay (NMD) changes expression pattern of POLR3A mutants.

(A) Sanger sequencing of a PCR fragment amplified with primers flanking exon 14 before and after 4 hours treatment with cycloheximide. The aberrant splice transcript including intronic sequence becomes visible due to inhibition of NMD. (B) Sanger sequencing of an amplicon containing the variant c.91C>T before and after cycloheximide treatment. In the upper panel (I) the signal from the mutated T-allele has the same intensity as the wild type C-allele indicating escape of NMD of the truncating allele T. In the lower panel (II) the signal from mutated T-allele appears to be reduced, which probably due to stabilization of the aberrant splice variant at c.1909+22G>A.

A



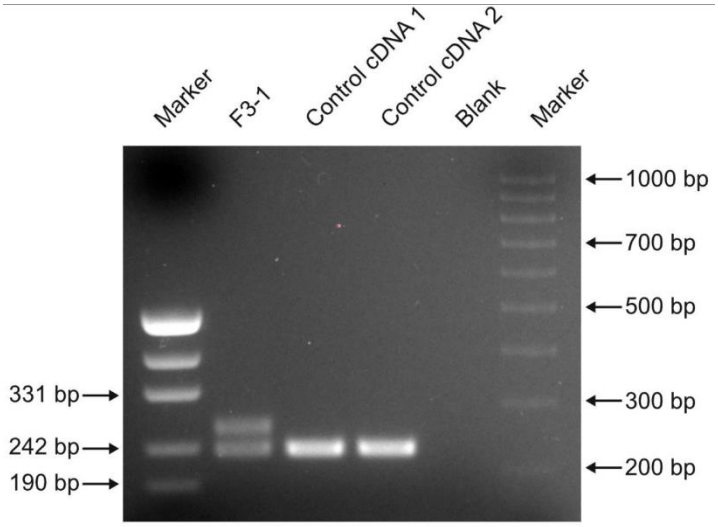
B



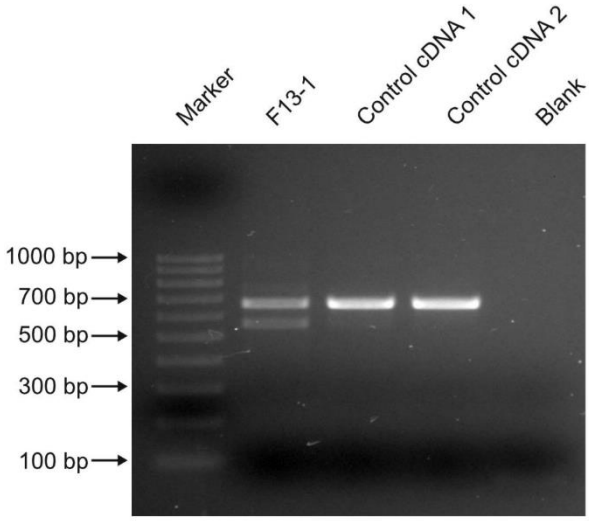
Supplementary Figure 6: Expression of RPC1 in Leukocytes.

Representative western blot **(A)** and quantification **(B)** of RPC1 expression. Analysis was done in members of family F1 and two sporadic cases. Western blots were performed three times. The protein expression levels were normalized to actin and expressed in arbitrary units (a.u.). All affected members of family F1 and both sporadic cases show reduced RPC1 levels compared to the healthy family member (F1-6).

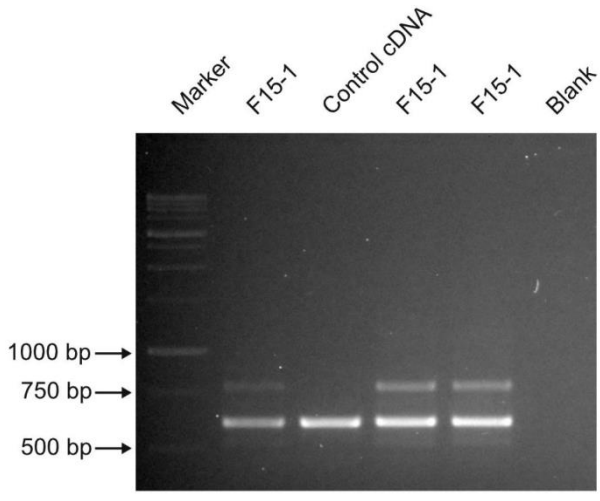
A


Marker
F3-1
Control cDNA 1
Control cDNA 2
Blank
Marker
1000 bp
700 bp
500 bp
300 bp
200 bp
331 bp
242 bp
190 bp

B

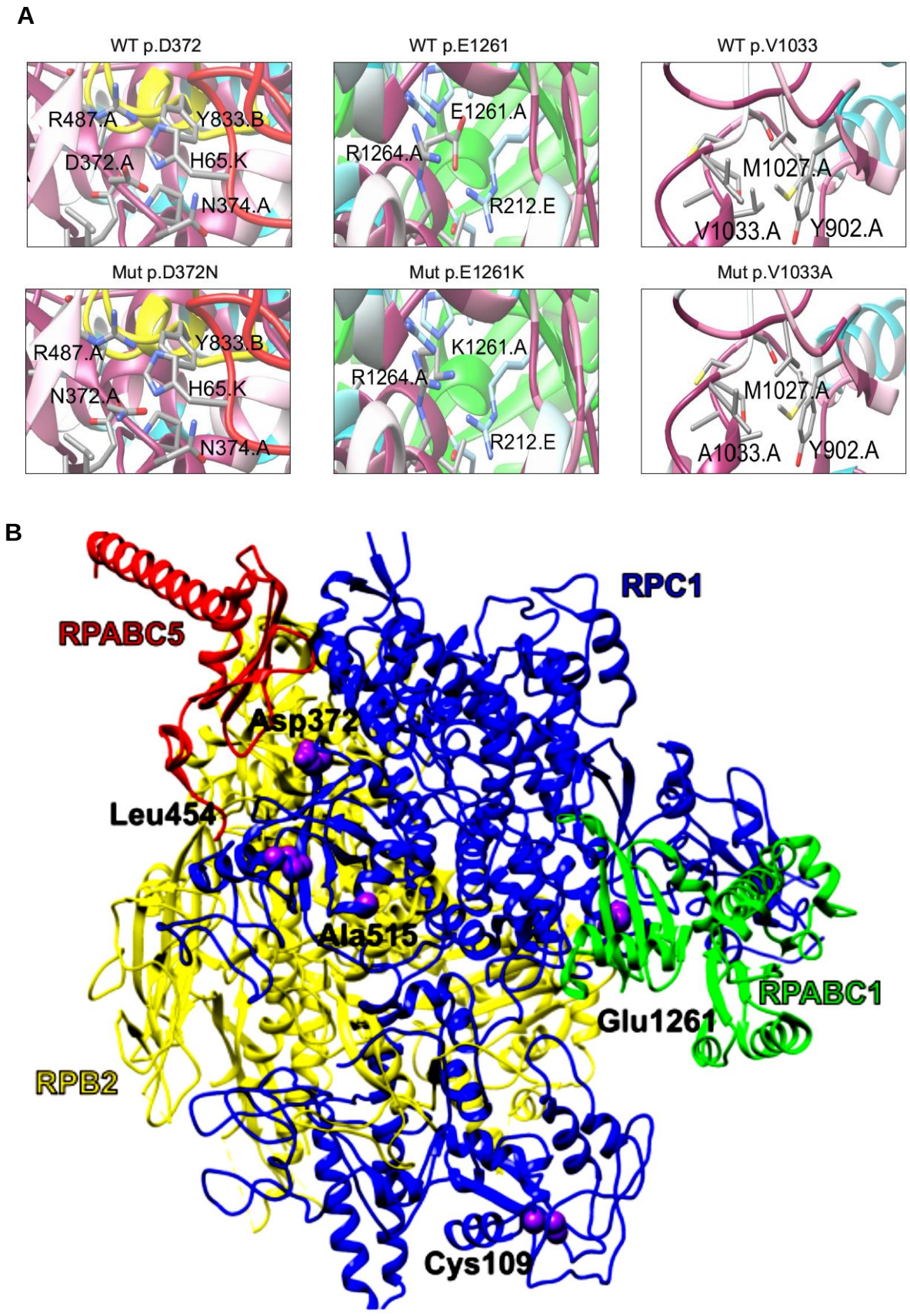

Marker
F13-1
Control cDNA 1
Control cDNA 2
Blank
1000 bp
700 bp
500 bp
300 bp
100 bp

C


Marker
F15-1
Control cDNA
F15-1
F15-1
Blank
1000 bp
750 bp
500 bp

Supplementary Figure 7: cDNA amplification to show the effect of splice site mutations.

(A) Amplification of leukocyte cDNA from individual F3-1 with the c.1290-2A>G and two control brain cDNAs. PCR product shows an additional aberrant band (261bp) for individual F3-1. Sanger sequencing of the longer aberrant band confirmed inclusion of 25 intronic bases of intron nine. **(B)** Amplification of leukocyte cDNA from individual F13-1 with the c.2247+2T>G and two control cDNAs. PCR products shows an additional shorter band of 574bp for individual F13-1. Sanger sequencing of this band revealed the missing of the last 109 bases of exon 16. **(C)** Amplification of leukocyte cDNA from individual F15-1 with the c.1048+5G>T and a control cDNA. PCR product shows an additional larger band (781bp) for individual F15-1. Sanger sequencing of the longer aberrant band demonstrated inclusion of sequence from intron seven.


A
WT p.D372
R487.A
Y833.B
D372.A
H65.K
N374.A
WT p.E1261
E1261.A
R1264.A
R212.E
WT p.V1033
M1027.A
V1033.A
Y902.A
Mut p.D372N
R487.A
Y833.B
N372.A
H65.K
N374.A
Mut p.E1261K
K1261.A
R1264.A
R212.E
Mut p.V1033A
M1027.A
A1033.A
Y902.A
B
RPC1
RPABC5
Asp372
Leu454
Ala515
RPABC1
Glu1261
RPB2
Cys109

Supplementary Figure 8: 3D modelling.

(A) Protein 3D representations of POLR3A missense mutations p.D372N, p.E1261K and.p.V1033A **(B)** Full length model of RPC1 (in blue) with its interacting accessory proteins. Those include: RPB2 subunit (in yellow), RPABC1 subunit (in green) and RPABC5 subunit (in red). Their interactions were assessed by homology (Sepehri et al.(Sepehri and Hernandez, 1997)). The localization of the *POLR3A* missense mutations is indicated as violet spheres.

V. Supplemental tables

Supplementary Table 1: Cohorts for POLR3A screening

Screening method	**HSP**	**ataxia / spastic ataxia**	**total**
Whole exome sequencing	256	189	445
Ataxia/HSP gene panel	94	74	168
Sanger sequencing	5	-	5
Total	**355**	**263**	**618**

HSP: Hereditary spastic paraplegia

Supplementary Table 2: Cohorts for POLR3A association study (c.1909+22G>A)

Cohort	**Definition**	**Number**	**Analysis method**	**Allele distribution**	**MAF (%)**
POLR3A spectrum cohort	ataxia, spastic paraplegia, peripheral neuropathy	1139 (including HSP n=482, ataxia n=213, CMT n=444)	WES (n=975) NGS gene panel (n=164)	C = 2252; T = 26	1.14
Disease controls	other neurological and non-neurological diseases	1448 (including 1112 probands with unrelated neurological phenotypes and 336 probands with non-neurological phenotypes)	WES (n=1448)	C = 2885; T = 11	0.38
Healthy controls	healthy elderly subjects	853	Custom-designed Taqman assay	C = 1700; T = 6	0.35

HSP: Hereditary spastic paraplegia; CMT: Charcot-Marie-Tooth; WES: Whole exome sequencing; NGS: Next generation sequencing;

MAF: Minor allele frequency.

Supplementary Table 3: Linkage regions in family F1

Chr	cytogenetic band	location (kb)	size (Mb)	flanking SNPs	overlapping known HSP loci
1	q23.2-32.1	160,237-201,605	41	rs7554737-rs10920222	SPG23
4	q33-q35.1	170,662-184,937	14	rs1047642-rs3811775	
10	q22.1-q24.1	74,480-98,699	24	rs7083190-rs4919060	SPG9, SPG27
11	q22.1-q22.3	98,661-110,364	12	rs12808518-rs12575639	
14	q11.2-q12	21,046-29,684	8.6	rs6576284-rs8006309	
15	q22.31-q25.1	66,529-81,499	15	rs1513933-rs11639206	
16	q21-q23.1	60,145-75,390	15	rs1423953-rs9934007	
20	q13.12-q13.2	42,585-52,378	10	rs6093902-rs11698898	
20	q13.33	58,724-63,025	4	rs1475030-rs7270744	

Genomic regions reaching the maximum parametric LOD score possible for this family (Z=1.33 at teta=0). A recessive model of inheritance, with a penetrance of 100% for homozygous mutation carriers and a disease allele frequency of 0.0001 was assumed to calculate expected two-point LOD score using the software FastSlink as implemented in EasyLinkage Plus v5.08 (Hoffmann et al. (Hoffmann and Lindner, 2005)). The suggestive linked regions encompass a total of 134 Mb of genomic space in which 2,616 kb correspond to coding exonic regions.

Supplementary Table 4: Candidate variants segregating in family F1 that localize within suggestive linkage regions

gene	genomic variant (hg19)	cDNA	protein effect	EVS6500 alleles	ExAC alleles	GERP	PhastCons	CADD score
HMCN1	chr1:185880722C>G	NM_031935:c.794-84C>G	Intronic	unknown	unknown	3.37	0.002	7.074
HMCN1	chr1:186010128G>A	NM_031935:c.6181-17G>A	Intronic	A=3/G=4403	C=4/A=363/G=122279	-3.2	0	0.766
POLR3A	chr10:79785941G>A	NM_007055:c.91C>T	Q31*	G=13006	A=1/C=2/G=122803	5.100	1.000	36.000
POLR3A	chr10:79769273C>T	NM_007055:c.1909+22G>A	cryptic splice site activation	T=36/C=12970	T=162/C=122804	3.810	0.007	7.377
ADCY4	chr14:24787778C>A	NM_139247:c.3082-4G>T	intronic	A=21/C=12985	G=1/A=216/C=122631	-1.670	0.346	5.683
ADCY4	chr14:24801966T>C	NM_139247:c.357+31A>G	intronic	C=22/T=11828	C=130/T=13230	-3.380	0.000	10.490

Filter criteria for WES data: homozygous or putative compound heterozygous variants, segregating with the phenotype under a recessive model, MAF < 0.05% (EVS6500, ExAC, in-house database). CADD scores are PHRED-scaled

VI. Supplementary References

Hoffmann K, Lindner TH. easyLINKAGE-Plus--automated linkage analyses using large-scale SNP data. Bioinformatics 2005; 21(17): 3565-7.

La Piana R, Cayami FK, Tran LT, Guerrero K, van Spaendonk R, Ounap K, *et al.* Diffuse hypomyelination is not obligate for POLR3-related disorders. Neurology 2016.

Schiffmann R, van der Knaap MS. Invited article: an MRI-based approach to the diagnosis of white matter disorders. Neurology 2009; 72(8): 750-9.

Sepehri S, Hernandez N. The largest subunit of human RNA polymerase III is closely related to the largest subunit of yeast and trypanosome RNA polymerase III. Genome research 1997; 7(10): 1006-19.

Steenweg ME, Vanderver A, Blaser S, Bizzi A, de Koning TJ, Mancini GM, *et al.* Magnetic resonance imaging pattern recognition in hypomyelinating disorders. Brain : a journal of neurology 2010; 133(10): 2971-82.

Takanashi J, Osaka H, Saitsu H, Sasaki M, Mori H, Shibayama H, *et al.* Different patterns of cerebellar abnormality and hypomyelination between POLR3A and POLR3B mutations. Brain Dev 2014; 36(3): 259-63.

Wolf NI, Vanderver A, van Spaendonk RM, Schiffmann R, Brais B, Bugiani M, *et al.* Clinical spectrum of 4H leukodystrophy caused by POLR3A and POLR3B mutations. Neurology 2014; 83(21): 1898-905.

Zhu BY, Zhou NE, Kay CM, Hodges RS. Packing and hydrophobicity effects on protein folding and stability: effects of beta-branched amino acids, valine and isoleucine, on the formation and stability of two-stranded alpha-helical coiled coils/leucine zippers. Protein Sci 1993; 2(3): 383-94.