Dissection of *TAF1* neuronal splicing and implications for neurodegeneration in X-linked Dystonia-Parkinsonism

Supplementary information

Simona Capponi, PhD[1], Nadja Stöffler, MSc[1], Ellen B. Penney, MD, PhD[2], Karen Grütz, PhD[3], Sheikh Nizamuddin, PhD[1], Marit W. Vermunt, PhD[4], Bas Castelijns, MSc[4], Cara Fernandez-Cerado[5], G. Paul Legarda[5], M. Salvie Velasco-Andrada[5], Edwin L. Muñoz, MD[6], Mark A. Ang, MD[6], Cid Czarina E. Diesta, MD[7], Menno P. Creyghton, PhD[4], Christine Klein, MD[8], D. Cristopher Bragg, PhD[2], Peter De Rijk, PhD[9], H. Th. Marc Timmers, PhD[1]*

[1] German Cancer Consortium (DKTK) partner site Freiburg, German Cancer Research Center (DKFZ) and Department of Urology, Medical Center-University of Freiburg, Freiburg, Germany.

[2] The Collaborative Center for X-linked Dystonia Parkinsonism (CCXDP), Department of Neurology, Massachusetts General Hospital, Charlestown, Massachusetts, USA.

[3] Institute of Neurogenetics, University of Lübeck, Lübeck, Germany.

[4] Department of Developmental Biology, Erasmus University Medical Center, Rotterdam, The Netherlands.

[5] Sunshine Care Foundation, Roxas City, Philippines.

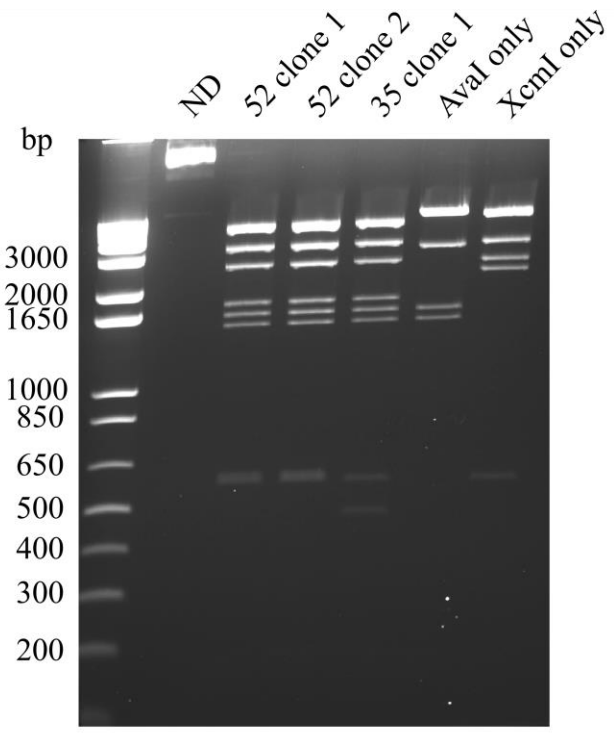
[6] Department of Pathology, College of Medicine, University of the Philippines, Manila, Philippines.

[7] Department of Neurosciences, Makati Medical Center, Makati City, Philippines.

[8] Institute of Neurogenetics and Department of Neurology, University of Lübeck, Lübeck, Germany.

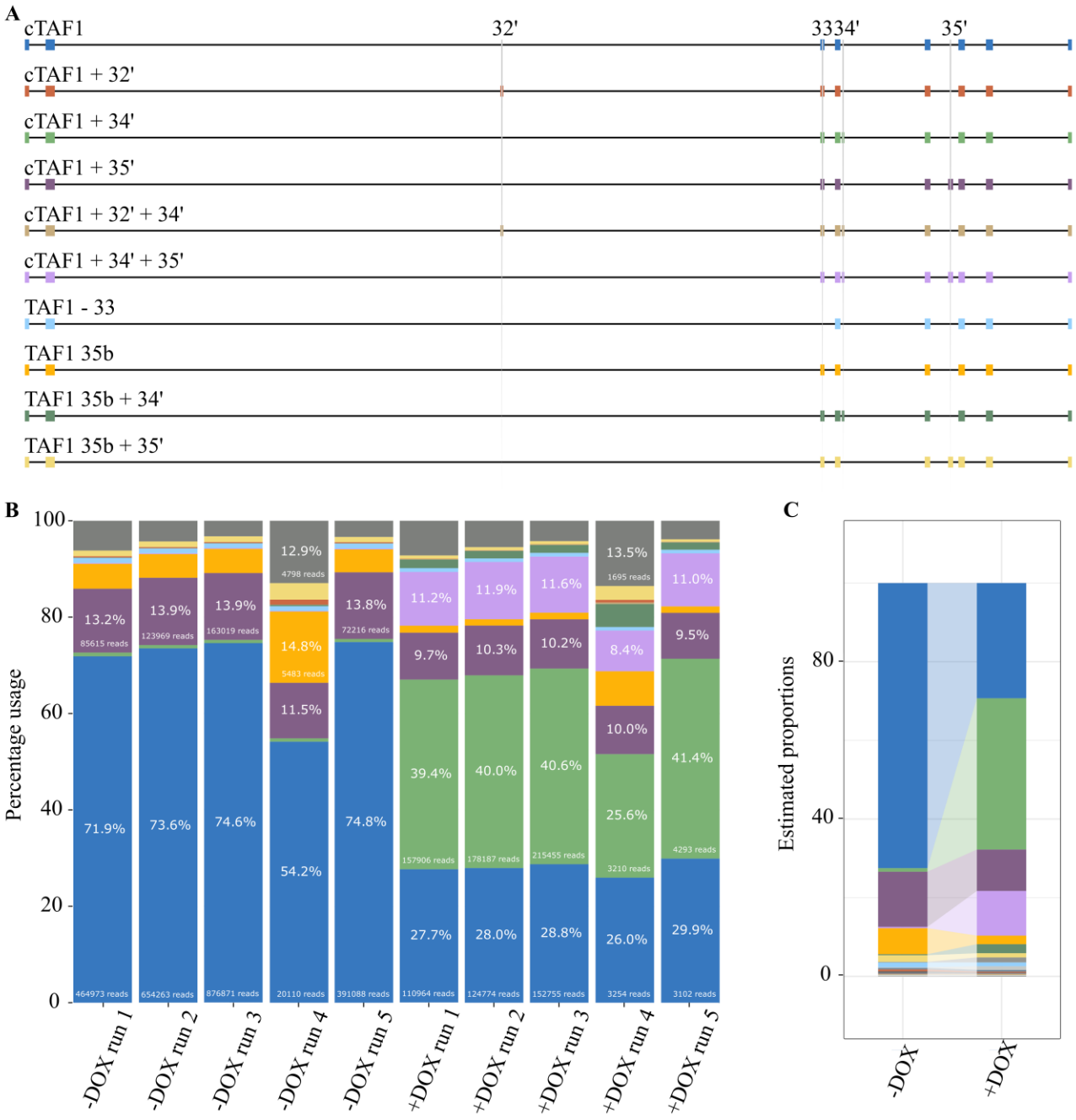
[9] Neuromics Support Facility, VIB Center for Molecular Neurology, VIB - University of Antwerp, Antwerp, Belgium.

* corresponding author:
H. Th. Marc Timmers, Ph.D
German Cancer Consortium (DKTK), partner site Freiburg
German Cancer Research Center (DKFZ) and
Department of Urology, Medical Center-University of Freiburg
Breisacher Str. 66
79106 Freiburg
Germany
phone: +49 761 270 63120/63121
mail: m.timmers@dkfz-heidelberg.de



Supplementary Fig. S1: Hexamer repeats length in pcDNA GFP-MG SVA

The hexamer repeat length of the XDP-SVA cloned into the pcDNA5 GFP-MG SVA was assessed using AvaI/XcmI (NEB) double digestion. Both enzymes have cutting sites flanking the hexamer repeat and the resulting band for a repeat length of 35 is 466 bp. The digestion of the pcDNA5 GFP-MG SVA containing 35 hexamers is shown in lane 5 (35 clone 1). As comparison, when the repeat size is increased to 52 (52 clone 1 and clone 2, lanes 3 and 4) the resulting band is 568 bp. Lane 1: 1 kb marker (Thermo Fisher Scientific), lane 2: ND=non-digested, lane 6 and lane 7: single digestion controls of 35 clone 1.



Supplementary Fig. S2: Validation of Nanopore sequencing analysis pipeline in HeLa cells expressing GFP-SRRM4

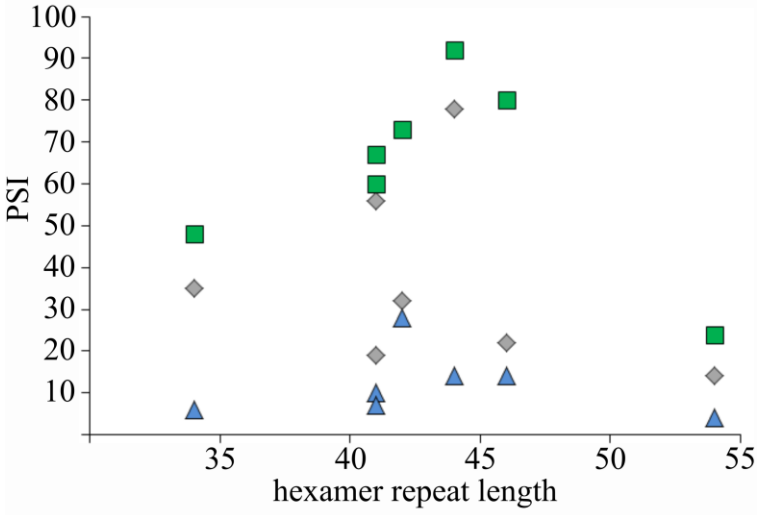
The computational analysis of the Nanopore sequencing data was validated using HeLa cells expressing DOX-inducible GFP-SRRM4 from a chromosomal locus. The same barcoded amplicons (-/+DOX) were included in five independent runs. The structure of the isoforms detected above 1% is depicted in panel A, while their distribution is detailed for each run in panel B including the number of supportive reads. Isoforms that do not reach the cut-off value of 1% are grouped and shown in grey in panel B. The quantification of the differential transcript usage (DTU) is depicted in panel C. DTU was upregulated in GFP-SRRM4 HeLa cells compared to non-induced for 34' containing transcripts (canonical +34', light green, p-value: 7.5e-195; canonical +34' +35', lilac, p-

value: 1.72e-50 and 35b +34', dark green, p-value: 5.24e-9). DTU was downregulated for canonical transcripts (canonical, dark blue, p-value: 5.31e-140; canonical +35', purple, p-value: 2.17e-10; 35b, orange, p-value: 3.01e-10).

A

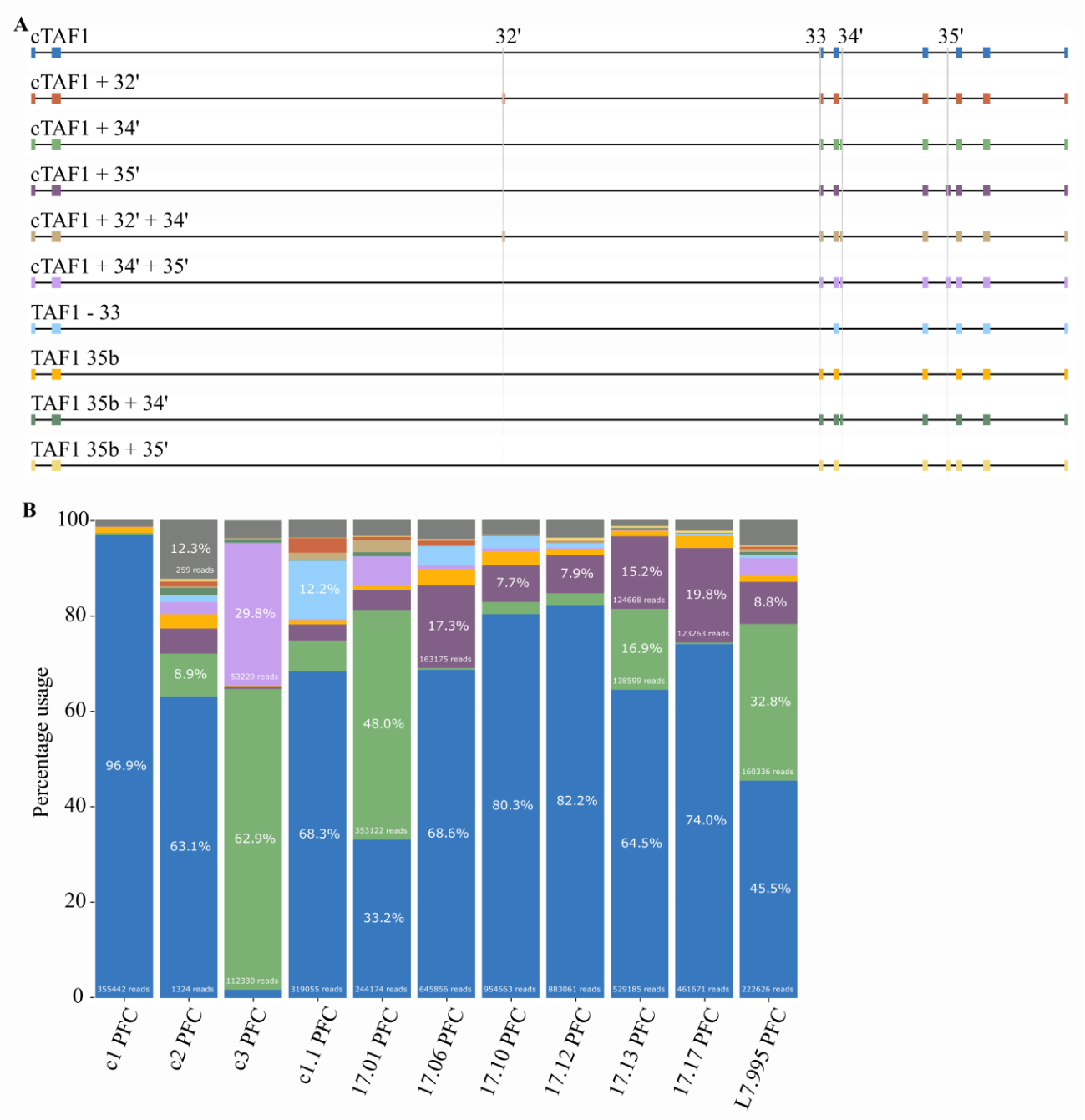
ID	repeat length	PSI		
		PFC	CB	AP
17.10	34	35	48	6
17.06	41	19	67	10
L7.995	41	56	60	7
17.12	42	32	73	28
17.01	44	78	92	14
17.13	46	22	80	14
17.17	54	14	24	4

B



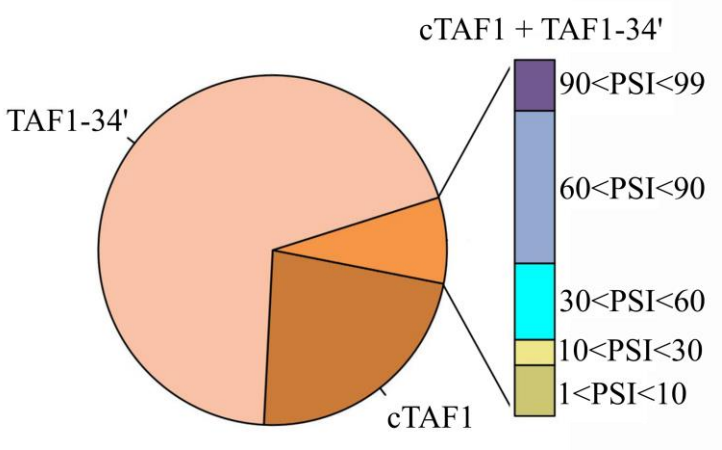
Supplementary Fig. S3: The hexamer repeat length within the XDP-SVA does not impact microexon 34' incorporation

The hexamer repeat length and the PSI of microexon 34' for the analysed XDP brain regions are detailed in panel A. ID: identification number for the analysed XDP brains, PFC: prefrontal cortex, CB: cerebellum: AP: anterior putamen. The correlation between microexon 34' PSI and hexamer length is plotted in panel B. PFC is represented in grey diamonds, CB in green squares and AP in blue triangles. Pearson's correlation and relative p-value were calculated for each set: PFC R^2=0.0798, p-value=0.5392; CB R^2=0.0773, p-value=0.5457; AP R^2=0.0089, p-value=0.8401.



Supplementary Fig. S4: Determination of *TAF1* mRNA isoforms definition for individual samples in PFC using Nanopore long-read sequencing

The structure of the isoforms detected above 1% is depicted in panel A. The percentage of the identified *TAF1* isoforms in PFC for each XDP and control individual is detailed in panel B, including the number of supportive reads.



Supplementary Fig. S5: Incorporation of *TAF1 34'* microexon in single cells from mouse cortex
Single cell transcriptome data from adult mouse primary visual cortex[1] was interrogated to explore the distribution of *TAF1* microexon 34' incorporation with vast-tools (https://github.com/vastgroup/vast-tools). In total 176 cells with at least 10 reads spanning microexon 34' were selected. Reads were aligned starting from the 5' splice site of exon 34 to detect the usage of the 3' splice site of exon 35 (cTAF1, n=40) or microexon 34' (TAF1-34', n=122). Reads spanning both exons were detected in 14 cells, indicating the simultaneous presence of *TAF1* isoforms including and excluding microexon 34' (cTAF1+TAF1-34'). PSI of microexon 34' in these 14 cells is detailed in the side bar.

References

1. Tasic, B., et al., *Adult mouse cortical cell taxonomy revealed by single cell transcriptomics.* Nat Neurosci, 2016. **19**(2): p. 335-46.

ID	gender	age at onset	age at death
17-01	M	*na*	46
17-06	M	40	46
17-10	M	58	67
17-12	M	*na*	43
17-13	M	33	42
17-17	M	27	43
L-10.322	M	50	57
L-7.995	M	31	35
c1	F	-	68
c2	F	-	90
c3	F	-	59

Supplementary table S1: Age and gender of XDP and control specimens.

Primer set	Application	Sequence
TAF1 34' cDNA-Fw	RT-PCR	TAGAAAGCCTGGACCCAATG
TAF1 34' cDNA-Rv	RT-PCR	GAGGCATCTCGAGACATACTGA
GFP Fw	RT-PCR	CAAAGACCCCAACGAGAAGC
TAF1 ex35 Rv	RT-PCR	CTGGAGTGGCACTGGGAATA
SRRM4 qPCR- Fw	RT-qPCR	CACAAGCGACGCAGGTCAT
SRRM4 qPCR- Rv	RT-qPCR	CGGTGGCGGTGAGACTTTC
hTBP-qPCR-FW	RT-qPCR	TGCACAGGAGCCAAGAGTGAA
hTBP-qPCR-RV	RT-qPCR	CACATCACAGCTCCCCACCA
hGAPDH-qPCR-FW	RT-qPCR	TGCACCACCAACTGCTTAGC
hGAPDH-qPCR-RV	RT-qPCR	GGCATGGACTGTGGTCATGAG
hHMBS-qPCR-FW	RT-qPCR	GGCAATGCGGCTGCAA
hHMBS-qPCR-RV	RT-qPCR	GGGTACCCACGCGAATCAC
hSDHA-qPCR-FW	RT-qPCR	TGGGAACAAGAGGGCATCTG
hSDHA-qPCR-RV	RT-qPCR	CCACCACTGCATCAAATTCATG
hHPRT1-qPCR-FW	RT-qPCR	TGACACTGGCAAAACAATGCA
hHPRT1-qPCR-RV	RT-qPCR	GGTCCTTTTCACCAGCAAGCT
hACTB fw	RT-qPCR	CACCATTGGCAATGAGCGGTTC
hACTB rv	RT-qPCR	AGGTCTTTGCGGATGTCCACGT
hTAF1ex30FW	Nanopore seq	TTTCTGTTGGTGCTGATATTGCGGCGTTTTCTTT CATTCTGG
hTAF1ex38RV	Nanopore seq	ACTTGCCTGTCGCTCTATCTTCTGCTGAAGCTTG TGTCTTGG

Supplementary table S2: Sequences of the primer sets used in the study.

transcript name with exons	CNTRL				XDP						
	c1 PFC	c2 PFC	c3 PFC	c4 PFC	17.01 PFC	17.06 PFC	17.10 PFC	17.12 PFC	17.13 PFC	17.17 PFC	L.7.995 PFC
canonical: 30a-31a-32-33-34-35-36-37	96.9	62.28	1.76	68.32	33.17	68.63	80.27	82.18	64.49	74.04	45.47
canonical +32'	0.11	1.03	0.03	3.1	0.7	1.11	0.1	0.19	0.07	0.08	0.45
canonical +34'	0.38	8.8	62.91	6.43	47.97	0.36	2.54	2.47	16.89	0.31	32.75
canonical +35'	0.02	5.17	0.56	3.39	4.22	17.34	7.71	7.93	15.19	19.77	8.79
canonical +32' +34'	0.01	0.28	0.07	1.58	2.51	0.02	0.02	0.16	0.02	0.01	0.63
canonical +34' +35'	0.01	2.54	29.81	0.17	6.08	1	0.7	0.18	0.26	0.09	3.5
canonical +32' +34' +35'	0	0.09	0.03	0.03	0.35	0.01	0.01	0.01	0	0	0.04
-33: 30a-31a-32-34-35-36-37-38	0.05	1.36	0.04	12.21	0.05	3.91	2.57	1.06	0.02	0.4	0.63
-33 +32'	0.05	0.14	0.02	0.06	0.03	0.04	0.06	0.06	0.05	0.05	0.06
-33 +34'	0	0.19	0.04	0.76	0.06	0.02	0.03	0.02	0.01	0	0.57
-33 +35'	0	0.28	0.02	0.38	0.04	1.44	0.09	0.05	0.01	0.9	0.06
35b: 30a-31a-32-33-34-35b-36-37	1.14	2.96	0.17	0.92	0.87	3.27	2.82	1.34	1.14	2.56	1.47
35b +34'	0.03	1.55	0.78	0.11	0.88	0.04	0.09	0.07	0.27	0.04	0.67
35b +35'	0.01	0.52	0.07	0.07	0.16	0.29	0.15	0.68	0.34	0.42	0.23
35b +34' +35'	0	0.24	0.36	0.01	0.22	0.03	0.02	0.02	0.02	0.01	0.09
canonical -32	0.03	0.14	0.01	0.04	0.03	0.05	0.19	0.01	0.01	0.02	0.07
canonical -35-36-37	0.01	0.14	0.01	0.01	0.02	0.02	0.02	2.25	0	0	0.05

Supplementary table S3: Sequences of the primer sets used in the study.

The table includes the percentage of the different isoforms identified with Nanopore sequencing (detected in more than 1% in at least one sample) for prefrontal cortex. The transcript names and the exonic compositions are listed in the first column. Percentages are listed for both control (CNTRL) and XDP samples.

Supplementary information: full size and uncropped blots and gels for Figure 1

Fig. 1B, PCR1

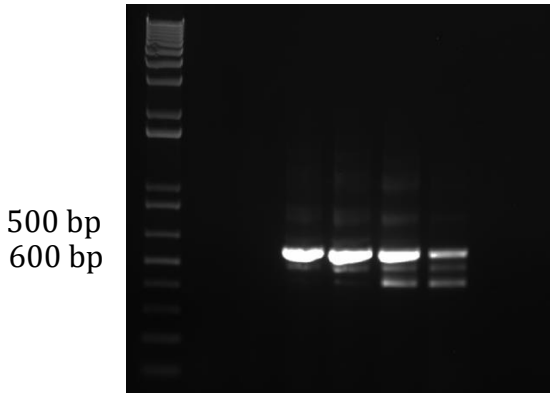


Fig. 1B, PCR2

The last two lanes represent a size control for the acrylamide run and are not included in the cropped panel. PCR bands at 150 bp represent canonical/34' heteroduplexes as demonstrated by their sensitivity to T7 endonuclease I digestion (data not shown).

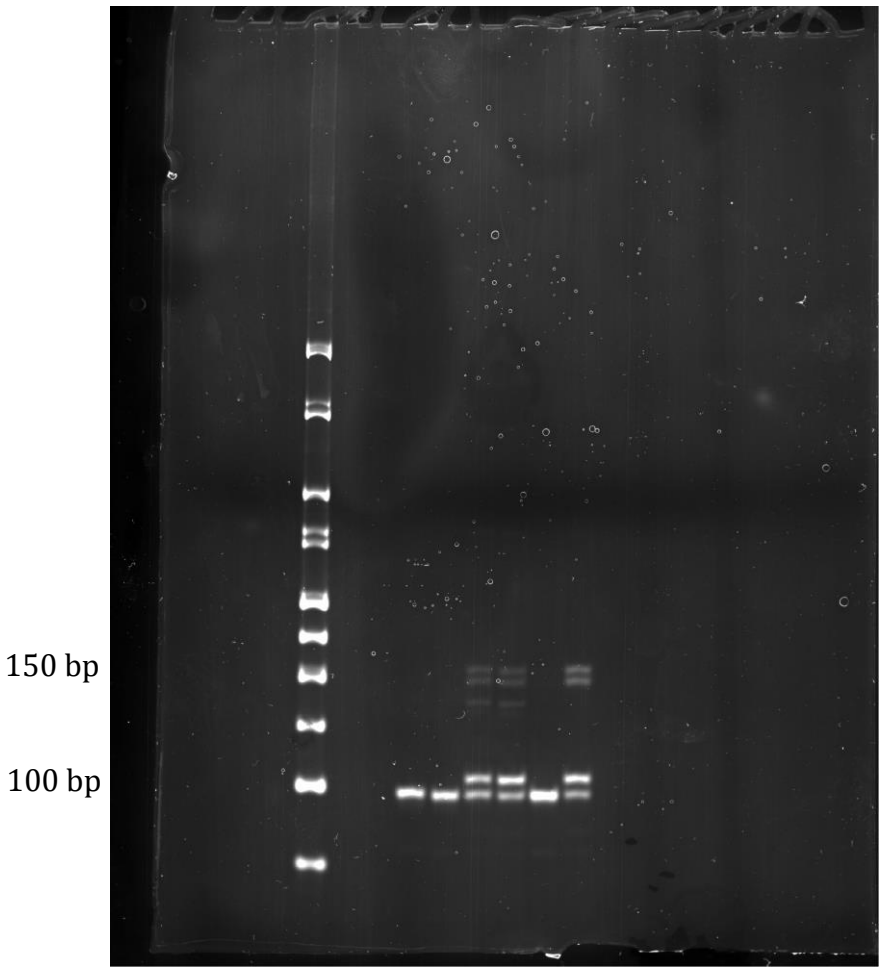


Fig.1C, α-SRRM4 immunoblot

The bands observed within the range of 100-135 kDa represent SRRM4 in different phosphorylated states, as previously proposed (Calarco et al., 2009). Bands with lower molecular weight represent aspecific bindings.

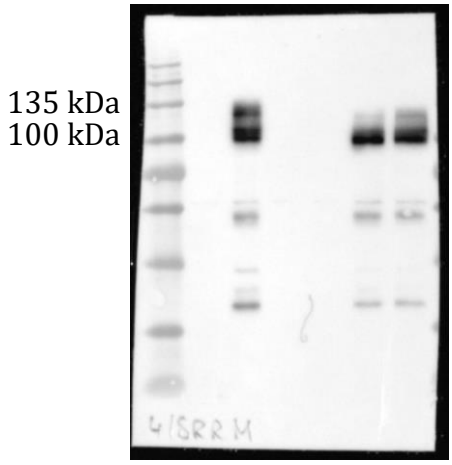


Fig.1C, α-cTAF1 immunoblot

The bands observed below 80 kDa and at 46 kDa represent alternative splicing products from the minigene processing. The GFP-fused minigene protein has a predicted molecular weight of 50 kDa and it is indicated by the star symbol. The bands migrating at 250 kDa represent the endogenous TAF1 protein.

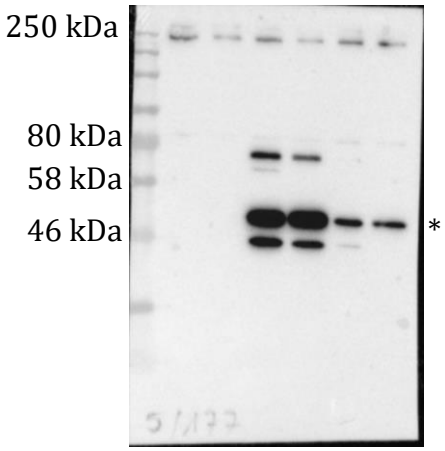


Fig.1C, α-TAF1-34' immunoblot

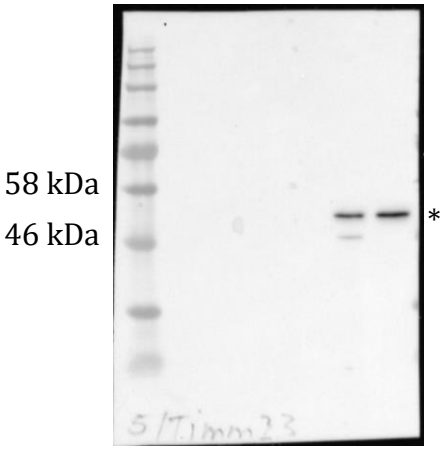


Fig.1C, α-vinculin immunoblot

The vinculin antibody has incubated on the membrane used to visualize the TAF1-34' antibody.

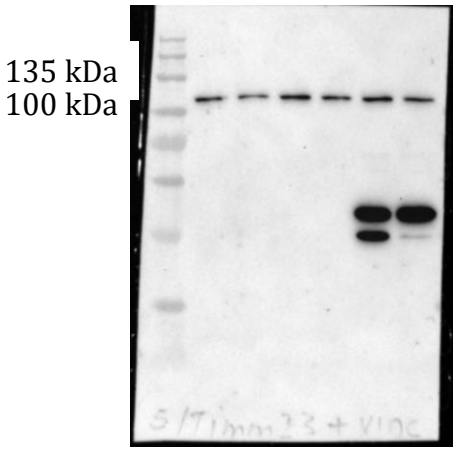
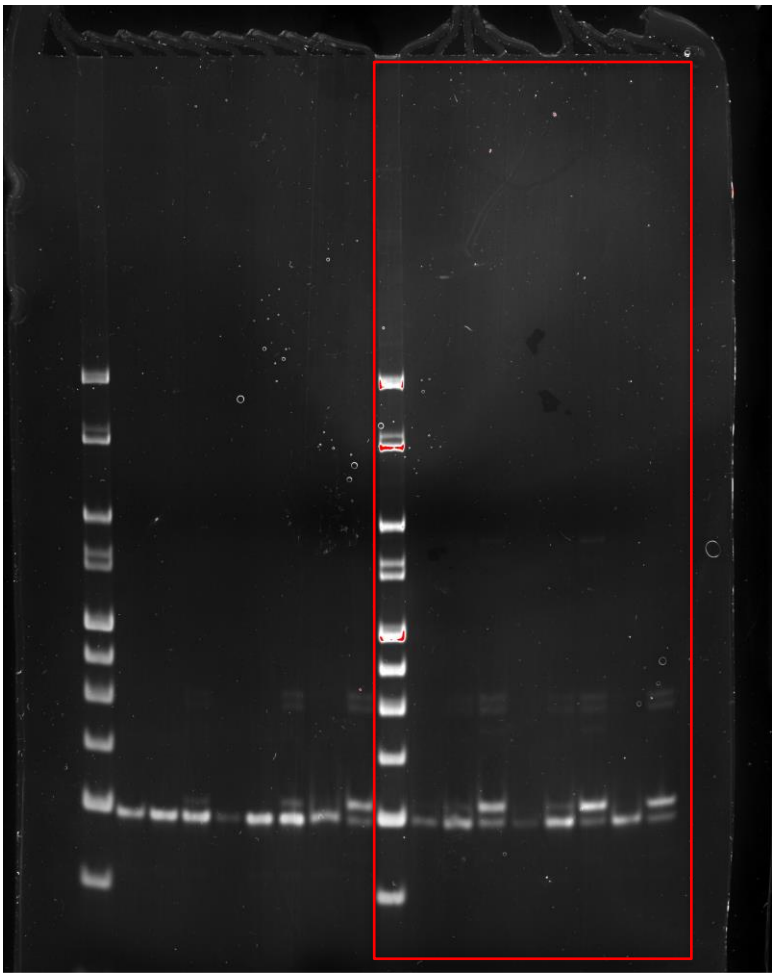


Fig.1D, N2a PCR2

The data depicted in Fig.1D is indicated by a red rectangle. Lanes 2 and 5 represent two internal control while lanes 8 and 9 represent a size control for the acrylamide run. These controls are not included in the cropped panel 1D. PCR bands at 150 bp represent canonical/34' heteroduplexes as demonstrated by their sensitivity to T7 endonuclease I digestion (data not shown).



Supplementary information: full size and uncropped blots and gels for Figure 2

In all four panels, the last two lanes represent a size control for the acrylamide run, not included in the cropped panels. PCR bands at 150 bp represent canonical/34' heteroduplexes as demonstrated by their sensitivity to T7 endonuclease I digestion (data not shown).

Fig. 2A, acrylamide run for control brains

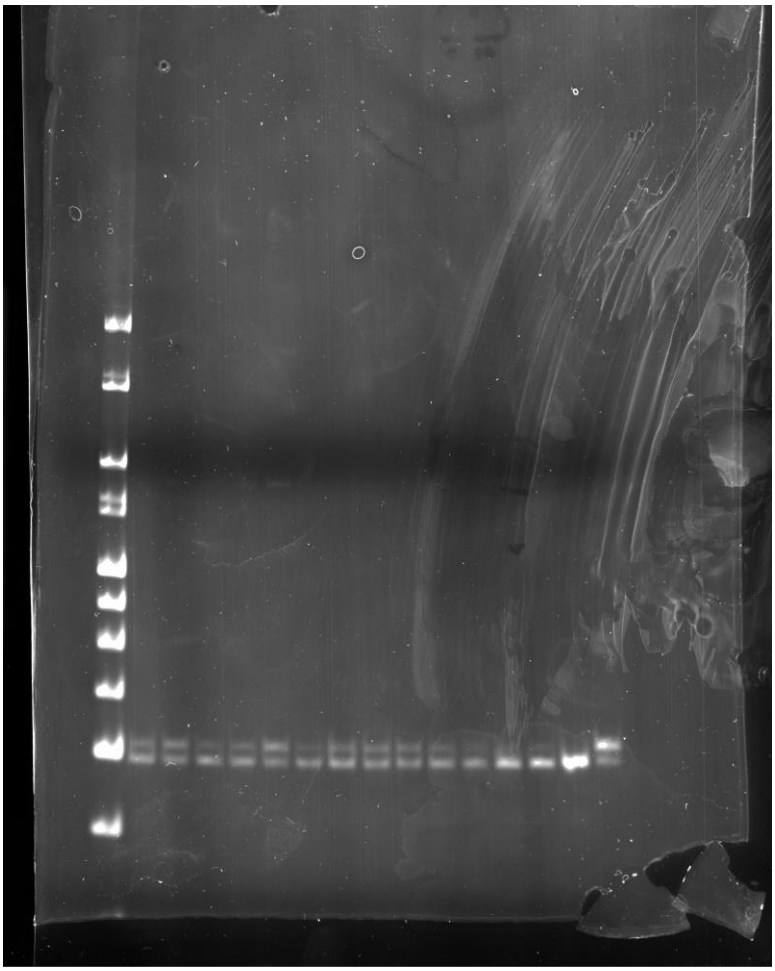


Fig. 2A, acrylamide run for XDP brains – set 1

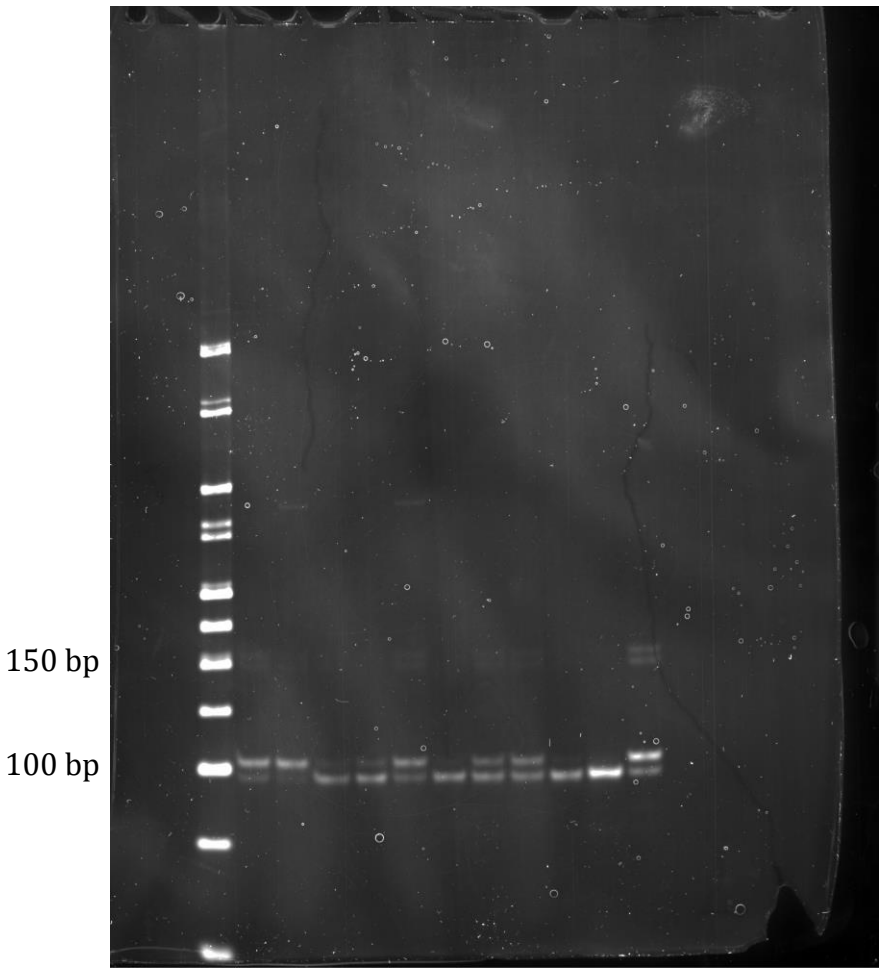


Fig. 2A, acrylamide run for XDP brains – set 2

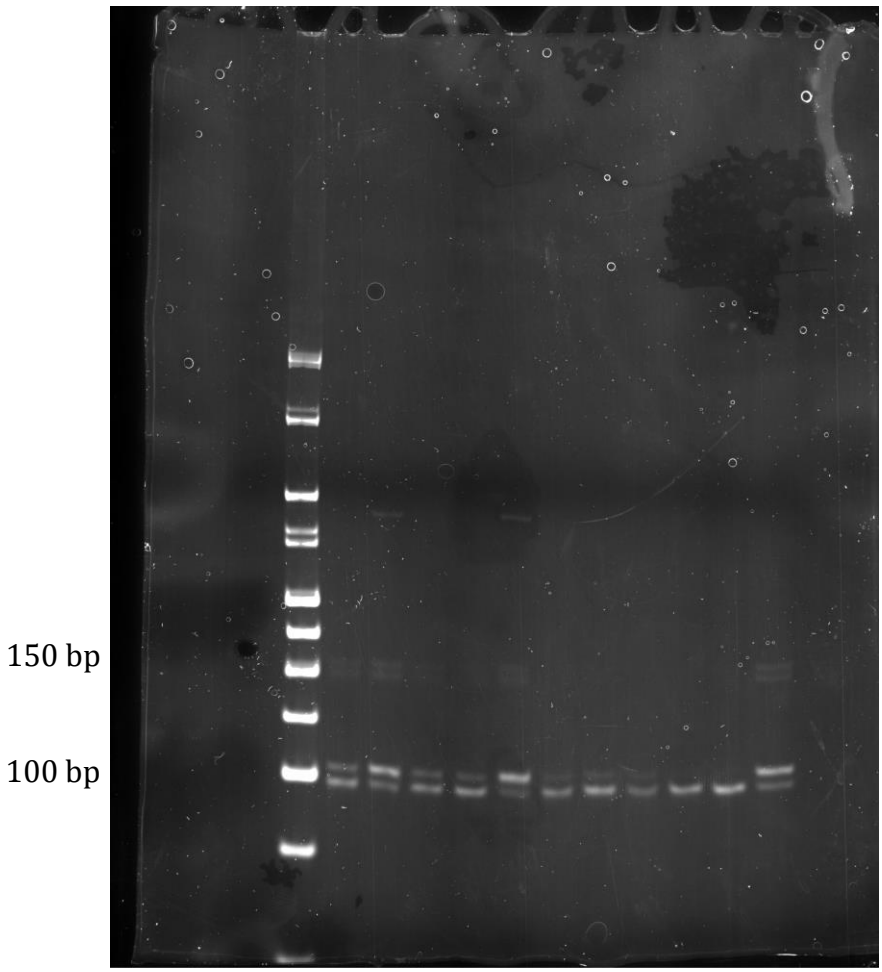


Fig. 2A, acrylamide run for XDP brains – set 3