



HnRNP K mislocalisation is a novel protein pathology of frontotemporal lobar degeneration and ageing and leads to cryptic splicing

Alexander Bampton^{1,2} · Ariana Gatt^{1,2} · Jack Humphrey³ · Sara Cappelli⁴ · Dipanjan Bhattacharya⁵ · Sandrine Foti^{1,2} · Anna-Leigh Brown⁸ · Yasmine Asi^{1,2} · Yi Hua Low^{1,2,6} · Marco Foiani^{5,7} · Towfique Raj³ · Emanuele Buratti⁴ · Pietro Fratta⁸ · Tammarn Lashley^{1,2}

Received: 5 April 2021 / Revised: 28 June 2021 / Accepted: 29 June 2021 / Published online: 18 July 2021
© The Author(s) 2021

Abstract

Heterogeneous nuclear ribonucleoproteins (HnRNPs) are a group of ubiquitously expressed RNA-binding proteins implicated in the regulation of all aspects of nucleic acid metabolism. HnRNP K is a member of this highly versatile hnRNP family. Pathological redistribution of hnRNP K to the cytoplasm has been linked to the pathogenesis of several malignancies but, until now, has been underexplored in the context of neurodegenerative disease. Here we show hnRNP K mislocalisation in pyramidal neurons of the frontal cortex to be a novel neuropathological feature that is associated with both frontotemporal lobar degeneration and ageing. HnRNP K mislocalisation is mutually exclusive to TDP-43 and tau pathological inclusions in neurons and was not observed to colocalise with mitochondrial, autophagosomal or stress granule markers. De-repression of cryptic exons in RNA targets following TDP-43 nuclear depletion is an emerging mechanism of potential neurotoxicity in frontotemporal lobar degeneration and the mechanistically overlapping disorder amyotrophic lateral sclerosis. We silenced hnRNP K in neuronal cells to identify the transcriptomic consequences of hnRNP K nuclear depletion. Intriguingly, by performing RNA-seq analysis we find that depletion of hnRNP K induces 101 novel cryptic exon events. We validated cryptic exon inclusion in an SH-SY5Y hnRNP K knockdown and in FTLN brain exhibiting hnRNP K nuclear depletion. We, therefore, present evidence for hnRNP K mislocalisation to be associated with FTLN and for this to induce widespread changes in splicing.

Keywords hnRNP K · Frontotemporal lobar degeneration · Frontotemporal dementia · RNA · Cryptic exons · Ageing

Alexander Bampton, Ariana Gatt and Jack Humphrey are Joint first authors.

Pietro Fratta and Tammarn Lashley are joint last authors.

✉ Pietro Fratta
p.fratta@ucl.ac.uk

✉ Tammarn Lashley
T.Lashley@ucl.ac.uk

¹ The Queen Square Brain Bank for Neurological Disorders, Department of Clinical and Movement Neuroscience, UCL Queen Square Institute of Neurology, London, UK

² Department of Neurodegenerative Diseases, UCL Queen Square Institute of Neurology, London, UK

³ Icahn School of Medicine at Mount Sinai, New York, NY, USA

⁴ International Centre for Genetic Engineering and Biotechnology (ICGEB), Padriciano 99, 34149 Trieste, Italy

⁵ The Firc Institute of Molecular Oncology Foundation (IFOM), Milan, Italy

⁶ Duke-NUS Medical School, Singapore, Singapore

⁷ University of Milan, Milan, Italy

⁸ Department of Neuromuscular Diseases, UCL Queen Square Institute of Neurology, London, UK

Introduction

Frontotemporal lobar degeneration (FTLD) describes a group of pathologically heterogeneous disorders all characterised by the selective and progressive atrophy of both frontal and temporal lobes [27, 66]. The ensuing neurocognitive syndrome broadly defined as frontotemporal dementia (FTD) is the second most common form of presenile dementia [8, 68]. FTD is itself an umbrella clinical term encompassing three distinct syndromes; behavioural variant FTD (bvFTD) and two language variants, progressive non-fluent aphasia (PNFA) and semantic dementia (SD) [12]. The current neuropathological classification of FTLDs recognises five major subgroups; three of which are characterised by specific proteinaceous inclusions principally containing TDP-43 (FTLD-TDP), FUS (FTLD-FUS) and Tau (FTLD-Tau) proteins [36, 38, 45, 53].

Two of these proteins, transactive response DNA-binding protein (TDP-43) and fused in sarcoma (FUS), are members of the diverse family of heterogeneous nuclear ribonucleoproteins (hnRNPs). hnRNPs are a group of ubiquitously expressed RNA-binding proteins (RBPs) implicated in the regulation of all aspects of nucleic acid metabolism including alternative splicing, mRNA stability, nucleo-cytoplasmic shuttling and translation [17]. They form highly dynamic co-operative complexes with RNA and other RBPs. The remodelling and restructuring of such complexes enables hnRNPs to bind and modulate an array of RNA regulatory machinery [23].

Disruption of RNA processing has been shown to be a central pathogenic mechanism in FTD and the overlapping motor disorder amyotrophic lateral sclerosis (ALS) [4]. Indeed, numerous transcriptomic and bioinformatic studies of TDP-43 and FUS depletion/mutation models have identified both gain and loss of splicing function phenotypes [7, 21, 29, 30, 49, 58, 67]. Recent studies have established that TDP-43 nuclear depletion induces the inclusion of normally repressed 'cryptic exons' into target transcripts [40, 48]. Cryptic exon inclusion can trigger premature polyadenylation leading to a sharp decline in protein expression, as in the case of *STMN2*/stathmin-2 [34, 48, 57]. Additionally, cryptic exon inclusion or constitutive exon skipping can trigger nonsense-mediated decay, or introduce evolutionarily untested maladaptive alterations to protein structure [30]. However, cryptic exon repression is not unique to TDP-43 and may indeed be an integral regulatory role of many RBPs to suppress the inclusion of intronic sequences by the spliceosome, as has been found for hnRNP C, RBM17, PTBP1, hnRNP L, SFPQ and MATR3 [2, 39, 47, 61, 69], but intriguingly, not for FUS [29, 30]. Understanding the transcriptional consequences of wider hnRNP dysfunction in the pathogenesis of FTLD is, therefore, imperative.

One of the most abundantly expressed proteins in the hnRNP family is hnRNP K. hnRNP K shares many structural characteristics typical of the hnRNP family including a nuclear localisation signal, a nuclear shuttling domain and three K homology (KH) domains which serve as nucleic acid recognition motifs [11, 50]. Unique to hnRNP K is its K interactive (KI) domain which facilitates its interaction with a multitude of protein targets within a vast interactome network [6, 55]. hnRNP K is widely expressed in multiple brain regions [63, 64] and has been specifically implicated in the post-transcriptional regulation of genes essential for maintaining ATP levels during cellular stress [22], axogenesis [41], myelination [37] and synaptic plasticity [19]. To date, research into the potential roles of aberrant hnRNP K expression in disease has been largely confined to cancer biology, where its overexpression and abnormal cytoplasmic accumulation have been correlated with advanced disease status and poor prognosis in several malignancies [5, 9, 10, 28].

Here we demonstrate, for the first time, hnRNP K mislocalisation from the nucleus to the cytoplasm as a frequent neuropathological feature in pyramidal neurons of the frontal cortex across the FTLD spectrum. We also present evidence for a degree of hnRNP K mislocalisation being a feature of the normal ageing process. We characterise this potentially novel pathological event with immunofluorescence to observe the spatial relationship of mislocalised hnRNP K puncta with the hallmark pathological inclusions of FTLD, markers of autophagy and mitochondria. We confirm hnRNP K mislocalisation to be a mutually exclusive pathology from TDP-43 or tau positive inclusions that are observed in different neuronal subtypes. We also utilise an hnRNP K knockdown cell model to predict the impact of hnRNP K depletion on the transcriptome. Analysis of cell-extracted RNA confirms the presence of cryptic exons specific to siRNA-induced nuclear loss of hnRNP K. Finally, we validate the presence of a cryptic exon inclusion event in postmortem human brain tissue, providing evidence for cryptic exon inclusion as a potentially direct functional consequence of hnRNP K nuclear depletion in FTLD neurons.

Materials and methods

Cases

Brains were donated to the Queen Square Brain Bank (QSBB) for neurological disorders (UCL Queen Square Institute of Neurology) and the Medical Research Council (MRC) Edinburgh Brain & Tissue Bank. All tissue samples were donated with the full informed consent. Accompanying clinical and demographic data of all cases used in this study were stored electronically in compliance with the 1998

data protection act and are summarised in Table 1. Ethical approval for the study was obtained from the NHS research ethics committee (NEC) and in accordance with the human tissue authority's (HTA's) code of practice and standards under licence number 12198.

All cases were diagnosed pathologically according to current consensus criteria [35, 45]. The cohort included pathologically diagnosed cases of FTLD-TDP A ($n=28$), FTLD-TDP B ($n=3$), FTLD-TDP C ($n=12$), FTLD-TDP D ($n=2$), FTLD-Tau ($n=5$), FTLD-ni ($n=2$), FTLD-FUS ($n=2$), ALS ($n=7$) and neurologically normal controls ($n=35$).

Immunohistochemistry

Slides with 8 μm mounted tissue sections were incubated at 60 °C overnight. Sections were deparaffinised in Xylene and rehydrated in decreasing grades of alcohol. Slides were incubated in methanol/hydrogen peroxide (0.3%) solution for 10 min to block endogenous peroxidase activity. For heat-induced antigen retrieval, slides were then transferred to a boiling solution of 0.1 M citrate buffer (pH 6.0) and pressure cooked at maximum pressure for 10 min. Slides were then incubated in 10% non-fat milk for 30 min at room temperature to block non-specific binding. Sections were incubated with primary antibody for 1 h at room temperature. The antibodies used for IHC in this study were both mouse-derived monoclonal hnRNP K antibodies (Abcam ab23644, 1:1000 and Bio-Rad MCA2622, 1:100). After three 5-min washes in tris-buffered saline with tween (TBS-T); slides were incubated for 45 min in 200 μl of biotinylated goat anti-mouse IgG secondary antibody (Vector Laboratories BA 9200, 1:200). Slides were washed as before and then incubated in pre-conjugated Strept(avidin)–Biotin Complex (ABC; DAKO) for signal amplification. The slides were then washed for a final time before being submerged in 3,3'-Diaminobenzidine (DAB) chromogen and then counterstained in Mayer's haematoxylin (BDH). Finally, slides were dehydrated in increasing grades of alcohol (70, 90 and 100% IMS), cleared in xylene and mounted.

Quantitative pathological assessment of hnRNP K localisation

hnRNP K-stained frontal and temporal tissue sections were scanned using a Olympus VS120 slide scanner at 20 \times magnification. The region of interest (ROI) was digitally marked, cropped and extracted using the Olympus VS desktop software to minimise file size. Extractions were consistently performed in the grey matter region of the second frontal gyrus for frontal lobe sections. Extracted ROI images were launched in ImageJ (v1.41) and a macro was used to generate the maximum number of random, non-overlapping,

1000 \times 1000 pixel of 0.345 mm² sized sample images (<300) from each ROI. Across the cohort, the mean number of images analysed was 129 or 44.5 mm² of analysed tissue per ROI.

Neurons with normal hnRNP K staining, as defined by a predominantly nuclear localisation of the protein, were detected and quantified by a supervised machine learning algorithm. We implemented a MATLAB-based program to generate the training data sets, which consisted of 250 images of frontal cortex neurons from different brain samples. AlexNet convolutional neural network with Adam optimizer [33] was used for training of the data sets and utilised a region based convolutional neural network (R-CNN) for identification of normally stained neurons. A manual estimation of the algorithm's accuracy found the detection rate to be consistently above 80% with minimal false-positives.

Neurons with abnormal hnRNP K pathology, as defined by nuclear clearance of hnRNP K and its subsequent mislocalisation to the cytoplasm, were counted manually within each randomly generated image, blinded to disease status. The total degree of hnRNP K mislocalisation in each case was given as the proportion (%) of all images with at least one abnormal neuron counted within it.

Double-label immunofluorescence

For double immunofluorescence, tissue sections were dewaxed, pre-treated and blocked as before. Sections were then either simultaneously or sequentially co-stained with mouse-derived hnRNP K (Abcam ab23644, 1:1000) and a second primary antibody. hnRNP K staining was amplified by incubation with a biotinylated IgG secondary antibody (DAKO/Vector laboratories, 1:200) prior to a 30 min incubation with ABC at room temperature as previously described for IHC. Antibody binding was visualised using a TSA Cyanine 3 amplification kit (Perkin-Elmer) which was applied to sections for 20 min at room temperature. After TBS-T washing, sections were incubated with species-appropriate Alexa Fluor 568 secondary antibodies (Invitrogen, 1:1000) for 2 h at room temperature to visualise the second (non-hnRNP K) antibody. Primary and secondary antibodies used for double immunofluorescence labelling as well as their respective conditions of incubation with anti-hnRNP K are detailed in Supplementary Table 1 (Online resource 1). Sections were washed a final three times in TBS-T with the second wash incorporating a 10-min incubation with 4',6-diamino-2-phenylindole (DAPI, Invitrogen, 1:1000) nuclei counterstain. Slides were mounted using Vectashield anti-fade mounting medium (Vector Laboratories).

Cross-reactivity was controlled for by the addition of two control sections stained as above with the individual omission of each primary antibody. Representative fluorescent images were captured at 20 \times , 40 \times or 63 \times magnification

Table 1 Cohort and clinical demographics

Case No	Clinical diagnosis	Pathological diagnosis	Age at disease onset	Age at death	Disease duration	Gender	Brain weight (g)	Mutations	Post mortem delay (h)
1*	AD	FTLD-TDP A	64	73	9	M	1252	C9orf72	61.1
2	CBD	FTLD-TDP A	51	61	10	M	1065		35.3
3	FTD	FTLD-TDP A	59	65	6	M	1176	C9orf72	30.0
4*	PNFA	FTLD-TDP A	66	72	6	M	1274		68.2
5	FTD	FTLD-TDP A	57	60	3	M	1673		40.4
6*	bvFTD	FTLD-TDP A	58	66	8	F	850	C9orf72	107.1
7	Control	FTLD-TDP A	75	79	4	M	–		10.0
8	FTD	FTLD-TDP A	52	58	6	M	1303	C9orf72	49.8
9	MND	FTLD-TDP A	47	53	6	M	1390		33.7
10*	FTD	FTLD-TDP A	53	63	10	M	955	C9orf72	77.3
11*	bvFTD	FTLD-TDP A	62	68	6	F	–	GRN (Q130fs)	99.8
12	MND	FTLD-TDP A	67	69	2	M	1398		62.5
13	CBD	FTLD-TDP A	75	79	4	F	1119		36.3
14*	PSP	FTLD-TDP A	83	87	4	F	1226		68.9
15*	FTD	FTLD-TDP A	57	62	5	M	–		92.9
16*	bvFTD	FTLD-TDP A	62	72	10	M	1320	TBK1	97.4
17	bvFTD	FTLD-TDP A	57	63	6	F	851		85.3
18	PNFA	FTLD-TDP A	57	62	5	F	981	C9orf72	63.1
19*	bvFTD	FTLD-TDP A	53	61	8	M	994	GRN (C31fs)	72.6
20*	FTD-MND	FTLD-TDP A	58	67	9	F	1000	GRN + C9orf72	115.0
21	FTD	FTLD-TDP A	62	68	6	M	1371	C9orf72	99.0
22*	bvFTD	FTLD-TDP A	43	45	2	M	1015	C9orf72	25.9
23	PNFA	FTLD-TDP A	56	67	11	F	789	C9orf72	85.6
24	FTD	FTLD-TDP A	59	71	12	F	1014	TBK1	76.0
25	bvFTD	FTLD-TDP A	49	55	6	M	974	GRN (C31fs)	29.3
26*	FTD-MND	FTLD-TDP A	66	74	8	F	782	C9orf72	85.8
27	FTD	FTLD-TDP A	54	60	6	M	1350	C9orf72	32.3
28	FTD-MND	FTLD-TDP A	66	71	5	M	1431	C9orf72	51.9
FTLD-TDP A Summary (<i>n</i> = 28)			59.6	66.1	6.5	18(M):10(F)	1142		64.0
29*	FTD-MND	FTLD-TDP B	63	67	4	F	1232		45.5
30	MND	FTLD-TDP B	67	69.0	2	M	1300		70.2
31	FTD	FTLD-TDP B	63	83	20	F	970		45.1
FTLD-TDP B Summary (<i>n</i> = 3)			64.3	73.0	8.7	1(M):2(F)	1167		53.6
32	FTD	FTLD-TDP C	52	65	13	F	899		27.7
33	bvFTD	FTLD-TDP C	64	66	2	F	1186	C9orf72	94.1
34	SD	FTLD-TDP C	73	83	10	M	1167		59.8
35	SD	FTLD-TDP C	58	72	14	F	972		31.2
36*	SD	FTLD-TDP C	67	76	9	M	1086		39.5
37	SD	FTLD-TDP C	58	73	15	F	976		37.9
38	SD	FTLD-TDP C	59	73	14	F	936		83.7
39	SD	FTLD-TDP C	64	78	14	M	1110		26.8
40	SD	FTLD-TDP C	64	74	10	M	1230		19.0
41	SD	FTLD-TDP C	50	65	15	M	1057		51.8
42	FTD	FTLD-TDP C	61	66	5	M	1117		70.8
43	PNFA	FTLD-TDP C	77	80	3	F	1502		26.3
FTLD-TDP C Summary (<i>n</i> = 12)			62.3	72.6	10.3	6(M):6(F)	1103		47.4
44	IBMPFD	FTLD-TDP D	–	48	–	F	1210	VCP	53.5
45	IBMPFD	FTLD-TDP D	53	71	18	M	1363		54.1
FTLD-TDP D Summary (<i>n</i> = 2)			53.0	59.5	18.0	1(M):1(F)	1287		53.8
46	bvFTD	FTLD-Tau	55	66	11	M	1208	MAPT (R406W)	60.8
47	PSP	FTLD-Tau	54	58	4	M	1225		78.3

Table 1 (continued)

Case No	Clinical diagnosis	Pathological diagnosis	Age at disease onset	Age at death	Disease duration	Gender	Brain weight (g)	Mutations	Post mortem delay (h)
48	bvFTD	FTLD-Tau	59	66	7	M	1399	MAPT (10+16)	58.2
49	bvFTD	FTLD-Tau	68	74	6	M	1048	MAPT (K280del)	125.0
50	bvFTD	FTLD-Tau	45	51	6	M	1046	MAPT (10+16)	52.6
FTLD-Tau Summary (n=5)			56.2	63.0	6.8	5(M):0(F)	1185		75.0
51	ALS	FTLD-FUS	44	46	2	M	1570		96.0
52	ALS	FTLD-FUS	69	72	3	F	1268		93.0
FTLD-FUS Summary (n=2)			56.5	59.0	2.5	1(M):1(F)	1419		94.5
53*	CBD	FTLD-Ni	48	54	6	F	1106		106.3
54	bvFTD	FTLD-Ni	50	57	7	M	1444		44.0
FTLD-Ni Summary (n=2)			49.0	55.5	6.5	1(M):1(F)	1275		75.2
55	ALS	ALS	55	59	4	M	875		36.0
56	ALS	ALS	63	66	3	F	1316		66.0
57	ALS	ALS	58	63	5	F	1228	C9orf72	66.5
58	ALS	ALS	40	53	13	F	1226		42.3
59	ALS	ALS	77	80	3	F	1086		5.0
60	ALS	ALS	80	84	4	M	1453		56.3
61	ALS	ALS	74	76	2	M	1138	C9orf72	26.4
ALS Summary (n=7)			63.9	68.7	4.9	3(M):4(F)	1189		42.6
62*	Control	Control	–	69	–	M	1435		171.0
63	Control	Control	–	67	–	M	1350		2.5
64	Control	Control	–	73	–	F	1214		24.0
65*	Control	Control	–	88	–	M	1077		16.3
66*	Control	Control	–	79	–	F	1288		88.8
67*	Control	Control	–	86	–	F	1234		120.0
68	Control	Control	–	93	–	F	1128		29.7
69	Control	Control	–	83	–	F	1263		99.0
70*	Control	Control	–	68	–	F	1330		45.1
71	Control	Control	–	70	–	M	1544		53.5
72	Control	Control	–	92	–	M	1213		46.3
73	Control	Control	–	96	–	F	1032		60.0
74	Control	Control	–	91	–	F	1130		71.8
75	Control	Control	–	84	–	F	–		40.6
76	Control	Control	–	80	–	M	–		11.5
77*	MSA	Control	–	83	–	M	1244		105.5
78	Control	Control	–	94	–	F	–		27.0
79	Control	Control	–	29	–	M	1590		44.0
80	Control	Control	–	25	–	M	1640		53.0
81	Control	Control	–	30	–	M	1670		71.0
82	Control	Control	–	25	–	M	1500		81.0
83	Control	Control	–	28	–	M	1330		38.0
84	Control	Control	–	34	–	M	1530		99.0
85	Control	Control	–	39	–	M	1360		76.0
86	Control	Control	–	37	–	F	1360		126.0
87	Control	Control	–	39	–	M	1470		86.0
88	Control	Control	–	46	–	M	1380		76.0
89	Control	Control	–	46	–	F	1400		99.0
90	Control	Control	–	48	–	M	1480		58.0
91	Control	Control	–	50	–	M	1350		122.0
92	Control	Control	–	57	–	M	1600		70.0
93	Control	Control	–	58	–	M	1650		96.0
94	Control	Control	–	53	–	F	1420		107.0

Table 1 (continued)

Case No	Clinical diagnosis	Pathological diagnosis	Age at disease onset	Age at death	Disease duration	Gender	Brain weight (g)	Mutations	Post mortem delay (h)
95	Control	Control	–	57	–	F	1320		73.0
96	Control	Control	–	51	–	M	1460		52.0
Control Summary (n = 35)			–	61.4	–	21(M):14(F)	1375		69.7

*RNA extracted from frozen frontal cortex tissue for cryptic exon validation; AD Alzheimer's disease; *aFTLD-U* atypical FTLD with ubiquitin inclusions; *ALS* Amyotrophic lateral sclerosis; *CBD* Corticobasal degeneration; *FTD* Frontotemporal dementia; *FTLD* Frontotemporal lobar degeneration; *FTLD-Ni* FTLD with no (known) inclusions; *bvFTD* Behavioural variant FTD; *IBMPFD* Inclusion body myopathy with early onset Paget disease; *PNFA* Progressive non-fluent primary aphasia variant FTD; *LPA* Logopenic aphasia; *MSA* Multiple system atrophy; *NIFID* Neuronal intermediate filament inclusion disease; *PPA* Primary progressive aphasia; *PSP* Progressive supranuclear palsy; *SD* Semantic dementia variant FTD; *SVD* Small vessel disease; Mutations: *C9orf72* Chromosome 9 open reading frame 72, *GRN* Progranulin, *MAPT* microtubule-associated protein tau, *TBK1* TANK-binding kinase 1, *VCP* Valosin-containing protein

using a Leica DM5500B fluorescence microscope and Z-stacks were subjected to a blind 3D deconvolution. Antibody staining was identified and imaged using the appropriate fluorescent channels, and colocalisation was confirmed or refuted on the combined, maximum-projected images.

Cell culture

Human neuroblastoma SH-SY5Y cells (ATCC) (Invitrogen, ThermoFisher) were cultured in DMEM: F-12, 1:1 Mixture (SigmaAldrich) supplemented with 15% fetal bovine serum (Gibco-BRL, Life Technologies Inc.), 1% antibiotic–antimycotic-stabilised suspension (SigmaAldrich) and 1% MEM Non-essential Amino Acid solution without L-glutamine (SigmaAldrich) at 37 °C in an incubator with a humidified atmosphere of 5% CO₂. SH-SY5Y cells were used up to 20 passages.

Generation of SH-SY5Y hnRNP K knockdown line

SH-SY5Y cells (80 × 10⁴) were seeded in 6-well plates and transfected with a mix composed of 150 µl Opti-MEM (Life-Technologies), 3 µl of 40 µM siRNA and 9 µl Lipofectamine RNAiMAX (Invitrogen, ThermoFisher) for a final siRNA concentration of 80 nM. Gene knockdown was performed according to the manufacturer's instructions. The siRNA sense sequences used for transfections were fire-fly luciferase (control) 5'-uaaggcuaugaagagauac-3' and (hnRNP K) 5'-aaauuaaggcucuccguaca-3'. After 48 h, cells were harvested for further analysis. Three replicates were performed for each condition.

SH-SY5Y RNA extraction and sequencing

RNA was obtained using a miRNeasy kit (Qiagen), following the manufacturer's instructions. RNA was recovered in

20–30 µl of RNase-free water and quantified with a Biophotometer D30 (Eppendorf, Hamburg, Germany). 9 µl (concentration from 5 to 9 µg) were sent to Novogene for library preparation and sequencing using their NEB unstranded protocol.

Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads and then fragmented randomly by addition of a fragmentation buffer. First strand cDNA was synthesized using random hexamer primer and M-MuLV Reverse Transcriptase (RNase H-). Second strand cDNA synthesis was subsequently performed using DNA Polymerase I and RNase H. Double-stranded cDNA was purified using AMPure XP beads. Remaining overhangs of the purified double-stranded cDNA were converted into blunt ends via exonuclease/polymerase activities. After adenylation of 3' ends of DNA fragments, NEBNext Adaptor with hairpin loop structure was ligated to prepare for hybridization. To select cDNA fragments of preferentially 150~200 bp in length, the library fragments were purified with the AMPure XP system (Beckman Coulter, Beverly, USA). Finally, the final library was prepared by PCR amplification and purification of PCR products by AMPure XP beads. Paired-end 150 bp reads were sequenced on an Illumina NovaSeq and base-called with CASAVA.

Each sample was aligned to the human genome (build hg38) using STAR v2.7.2 [16] using GENCODE v30 [20] as the transcript reference. Unique alignment rates for each sample were all above 90%. Gene abundance was estimated using RSEM v1.3.1 [42] using the GENCODE v30 transcript set. Differential expression was assessed by inputting the RSEM transcript counts summed for each gene and running DESeq2 [44] with apeglm fold-change shrinkage [70]. Genes were considered differentially expressed at a FDR < 0.05 and a $|\log_2\text{FoldChange}| > 1$, although we note that hnRNP K itself fell just short of these cutoffs (\log_2 fold change = -0.91; adjusted *p* value = 1.3e-7). Full differential

expression results are available in Supplementary Table 2 (Online resource 1).

SH-SY5Y western blotting analysis

Total protein extracts were obtained by sonicating cells for 10 min and homogenising them at high power with a BioRuptor UCD-200 in a mild lysis buffer composed of 1 × Phosphate Saline Buffer (PBS) and 1 × Complete Protease Inhibitor Cocktail (Roche Diagnostics). Protein extract (10 µg) from each sample was resuspended in 1 × NuPage LDS Sample Buffer (Invitrogen) prepared with 2.5% beta-mercaptoethanol and boiled at 95 °C for 5 min. Lysates were resolved on NuPage 10% Bis–Tris gel (Invitrogen) and transferred to a Nitrocellulose Power Blotter Select Transfer Stack (Invitrogen). After blocking with 4% milk, blots were probed with antibodies: polyclonal rabbit a-hnRNP K 1:250 (SigmaAldrich) and polyclonal mouse a-GAPDH 1:1000 (Santa Cruz Biotechnology). After washing, blots were probed with HRP-conjugated secondary antibody (Dako) and developed with SuperSignal West Femto (ThermoFisher) or ECL Luminata Classico Western HRP substrate (Merck Millipore). Images were acquired and analysed using Alliance 9.7 Western Blot Imaging System (UVitec Limited). Statistical analysis of protein expression was carried out by an unpaired *t* test on three independent experiments.

Identification of cryptic exon targets from RNA-seq data

Differential splicing was assessed using Leafcutter, a tool optimised for finding novel splicing by clustering overlapping splice junction reads and comparing the contribution of each junction between conditions [43]. Briefly, splice junction reads were extracted from each alignment file using Regtools [18]. Junctions were clustered together with the following parameters: minimum proportional contribution to a cluster = 0.001, minimum read contribution to a cluster = 60, maximum intron length = 200,000 bp. Differential splicing was tested using a Dirichlet multinomial model comparing the proportions of each junction within a cluster between hnRNP K knockdown siRNA and scrambled siRNA control. After correcting for multiple testing using a Benjamini–Hochberg false discovery rate threshold of 5%, 8355 clusters were differentially spliced. We generated a custom script to identify novel cassette exons (see code availability), based on whether the cluster contained three splice junctions in the correct orientation, with two junctions flanking a central exon (inclusion junctions) and a third junction spanning the length of the cluster (skipping junction), and whether the inclusion and/or skipping junctions were annotated in GENCODE (v30). Any putative

cassette exon larger than 250 bp was removed. 2290 differentially spliced clusters were classified as cassette exons. Percent spliced in (PSI) of each cassette exon in each sample was calculated by taking the average read count of the two inclusion junctions and dividing it by the read count of the skipping junction. Direction of effect or delta PSI (dPSI) was calculated by taking the difference between the average PSI in hnRNP K knockdown and the controls. 1039 cassette exons had a |PSI| ≥ 10%. We then classified exons as either cryptic or skiptic by comparing the mean $PSI_{control}$ with the PSI_{hnRNPK} . Cryptic exons were initially classified based on novelty in annotation, but as the comprehensiveness of transcript annotation improves, the number of novel exons will shrink. Cassette exons with a PSI < 10% in controls and a dPSI > 10% were considered *cryptic exons*. 101 exons met these criteria, of which 49 were not annotated by GENCODE. Cassette exons with a PSI > 90% in controls and a dPSI < − 10% were considered *skiptic exons*. Overlaps between differential expression and differential splicing were computed by matching gene names. To compare genes containing differentially spliced cassette exons to a null, a set of genes with similar expression levels was selected using the mean DESeq2 $\log_{10}(\text{baseMean})$ plus or minus 1 standard deviation.

Phylogenetic *p* values (PhyloP) conservation scores [56] between the human (hg38) and 29 other mammalian genomes (27 primates, dog, mouse, and armadillo), were downloaded from the UCSC table browser [31]. Mean PhyloP scores for each cassette exon were calculated using the UCSCtools bigWigSummary function [32]. Full results for the differential splicing analyses (all exons at FDR < 0.05, without filtering an effect size) are available in Supplementary Table 3 (Online resource 1). Whether an exon is labelled as cryptic or skiptic is listed in the “exon_set” column. The locations of these exons are listed in hg38 coordinates in the “exon_coords_hg38”, whereas the coordinates for the containing introns are listed in “intron_coords_hg38”.

DNA FASTA sequences were created for each cassette exon as well as the flanking 100 bp upstream and downstream, with respect to the transcript strand, using BSgenome.Hsapiens.UCSC.hg38. For the three sets, the sequences from the included events were directly compared to the skipped events (and vice versa) using the differential enrichment mode of MEME [3], setting possible motifs to be within four and eight nucleotides. The full command for each run was `meme <target.FASTA> -dna -neg <background.FASTA> -objfun de -mod zoops -minw 4 -maxw 8 -wg 11 -ws 1 -markov_order 0 -maxiter 50 -distance 0.001 -brief 1000 -shuf 2 -csites 1000 -nmotifs 3 -seed 0 -wnsites 0.8`. This identified motifs that are enriched in one set compared to the other. Each comparison was run in both directions, resulting in six different MEME runs.

As a comparison to the MEME motifs, we counted how often a particular nucleotide sequence occurred one or more times in a set of sequences. We compared the proportions of each group of sequences with a χ^2 test of equal proportions. We performed this for all possible DNA kmers with lengths 2–4. All p values were adjusted for multiple testing using the Bonferroni correction. We report enrichment as the proportion of included exon sequences/proportion of skipped exon sequences. Full results are in Supplementary Tables 4, 5 and 6 (Online resource 1).

RNA extraction from human postmortem brain tissue

RNA was isolated from frontal cortex tissue of nine brains exhibiting normal hnRNP K mislocalisation (5 controls, 4 FTLD-TDPA) and 12 brains exhibiting mislocalised hnRNP K (8 FTLD-TDP A, 1 FTLD-TDP B, 1 FTLD-TDP C, 1 FTLD-ni and 1 control) as indicated in Table 1. 50 mg of flash-frozen tissue was homogenised in 700 μ L of Qiazol (Qiagen) using a TissueRuptor II (Qiagen). Chloroform was added and RNA subsequently extracted following the spin-column protocol from the miRNeasy kit with DNase digestion (Qiagen). RNA was eluted off the column in 50 μ L of RNase-free water. RNA quantity and quality were evaluated using a spectrophotometer.

PCR validation of cryptic exon targets

cDNA was synthesised from SH-SY5Y cell-derived and postmortem tissue-derived RNA using SuperScript IV VILO Master Mix with ezDNase enzyme step (ThermoFisher). Cryptic exons in predicted targets of hnRNP K were amplified using primer pairs that flank the cryptic exon, as well as a third primer which spans the cryptic exon. Primer sequences are presented in Supplementary Table 7 (Online resource 1). PCR for splicing events was conducted using 2 \times GoTaq PCR Master Mix (Promega) using the following touchdown thermal cycling conditions: 95 $^{\circ}$ C for 5 min, (95 $^{\circ}$ C for 30 s, 75 $^{\circ}$ C for 45 s (– 1 $^{\circ}$ C per cycle), 72 $^{\circ}$ C for 1 min) \times 15 cycles, (95 $^{\circ}$ C for 30 s, 60 $^{\circ}$ C for 45 s, 72 $^{\circ}$ C for 1 min) \times 20 cycles, 72 $^{\circ}$ C for 5 min. Products were electrophoresed and visualised on agarose gels with SybrSafe (ThermoFisher) or using D1000 ScreenTape and reagents (Agilent) on the 2200 TapeStation system (Agilent).

PCR products from the amplification of the *FAM160B2* cryptic event in human postmortem tissue-derived cDNA were gel excised and DNA extracted using the QIAquick Gel extraction Kit (Qiagen). Sanger sequencing was performed (Source Bioscience, Cambridge, UK) and products were aligned using SnapGene to validate that the amplified product corresponded to a cryptic splicing event.

Statistical analysis

For pathology, all generated data plots and accompanying statistical analyses were conducted using Graphpad Prism software (v7.00 for Windows). Data sets were subjected to the D'Agostino–Pearson test for normal variance which in turn guided the selection of further statistical tests for t test comparisons and clinical data correlation purposes. In all statistical comparisons, a corresponding p value of < 0.05 was considered statistically significant. The level of significance is demonstrated in figures as * for $p < 0.05$, ** for $p < 0.01$, *** for $p < 0.001$. Where appropriate, for all data-plots provide the corresponding statistical test, n value, p value and r value are detailed in the figure legend.

Results

HnRNP K is frequently mislocalised in the frontal cortex

Immunohistochemical staining of hnRNP K in the frontal cortex demonstrates normal neuronal localisation of the protein within the most superficial layers of the cortex (layers I–II). Neurons within these layers exhibited a predominantly nuclear localisation of the protein with uniform intensity irrespective of disease status (Fig. 1a, b). However, the larger pyramidal neurons of layers III and V revealed a remarkably different staining pattern in FTLD cases compared to age-matched controls, a potentially novel neuropathological event. In control brains, pyramidal neurons were characterised by a predominantly nuclear localisation of hnRNP K with only low-intensity perinuclear or cytoplasmic staining (Fig. 1a). By contrast, FTLD brains exhibited vast regions of severely abnormal hnRNP K mislocalisation that were occasionally observed across the entire pyramidal cell layers. Afflicted neurons were morphologically characterised by an almost complete depletion of nuclear hnRNP K and a concurrent punctate cytoplasmic accumulation of the protein that extended into the neurites (Fig. 1b).

HnRNP K is mislocalised across frontotemporal lobar degeneration subtypes

Upon closer examination we identified many examples of this widespread mislocalisation across the FTLD-TDP and FTLD-Tau pathological spectrum and in a rare FTLD-ni subject with no known pathological inclusions (Fig. 1c, Supplementary Fig. 1a–d, Online resource 2). To quantify this novel neuropathological event, we sought to obtain large numbers of randomly generated images (average of 129 images per case or 44.5 mm² of analysed tissue) that

had been generated from immunohistochemically stained frontal cortex.

Age-matched cases of ALS, FTL-D-TDP A, FTL-D-TDP C, FTL-D-Tau and control subjects were then selected for comparative analysis (Supplementary Fig. 1e, Online resource 2). Disease categories were compared according to their degree of normal hnRNP K staining by calculating each case's average number of normally stained neurons per mm² analysed. As expected, control cases had the greatest frequency of normal hnRNP K-stained neurons per image which was significant compared to the FTL-D-TDP A disease subtype ($p=0.013$) (Fig. 1d). This reflected control frontal cortex having a greater number of surviving neurons than FTL-D subjects.

The same groups were then compared on their degree of abnormal hnRNP K staining by calculating the proportion of all sampled images which contained at least one mislocalised neuron for each case. For example, a case with a mislocalisation score of 50% would mean that 50% of all the case's randomly sampled images contained at least one neuron with hnRNP K mislocalisation. By contrast to normal hnRNP K staining, FTL-D-TDP A and FTL-D-Tau disease groups exhibited significantly more mislocalisation of hnRNP K protein relative to controls ($p=0.004$, $p=0.002$) (Fig. 1e). There was no difference between controls and FTL-D-TDP C subjects. Of interest, FTL-D-Tau cases also exhibited greater hnRNP K mislocalisation than ALS subjects which were included as a disease control due to the relative sparing of the frontal cortex in ALS pathology ($p=0.04$).

Notably, we did not find hnRNP K mislocalisation to be a specific pathological feature of any one FTL-D subtype. Equally, not all individuals of any one subtype were vulnerable to hnRNP K mislocalisation. Indeed, the proportion of all neurons counted that were found to have mislocalised hnRNP K within the FTL-D-TDP A group alone ranged from 0% to as much as 20% making this a very frequent neuropathological feature in some cases (Supplementary Fig. 1f, Online resource 2). This prompted us to explore the relationship of other variables with hnRNP K mislocalisation score and in particular age at death.

hnRNP K mislocalisation is an age-related feature of neurodegenerative disease

We analysed 35 control subjects with an age at death ranging over 70 years from 25 to 96 as well as a combined FTL-D disease group (FTL-D-TDP and FTL-D-Tau subjects, $n=50$) which involved pooling the previously analysed FTL-D-TDP A, FTL-D-TDP C and FTL-D-Tau cohorts with the further addition of $n=3$ and $n=2$ FTL-D-TDP B and FTL-D-TDP D cases, respectively. Analysing control individuals in

isolation, age at death was found to strongly correlate with hnRNP K mislocalisation ($r=0.552$, $p=0.0006$) (Fig. 2a). By contrast hnRNP K mislocalisation in the FTL-D cohort was much more weakly associated with age at death ($r=0.201$, $p=0.162$ ns).

To better visualise and compare the different relationships between increasing age and hnRNP K pathology in control and FTL-D groups, we generated a cumulative frequency plot (Fig. 2b). We plotted the rolling total of hnRNP K mislocalisation for each cohort with ascending age and normalised this to the sum total of each group's total level of quantified mislocalisation. This illustrated the advanced nature of hnRNP K pathology onset in younger FTL-D individuals relative to controls; with the median amount of mislocalisation in the FTL-D's (reached at 68 years of age) being 18 years earlier than in controls (86 years) (Fig. 2b).

Mislocalised hnRNP K is distinct from TDP-43 and Tau inclusions

To confirm that neurons with hnRNP K pathology were distinct from those with TDP-43 immunoreactive inclusions, we performed double fluorescence to determine the spatial relationship of these two pathologies in FTL-D-TDP A. Neurons with TDP-43 inclusions that were predominantly found in cortical layer II displayed normal, nuclear-localised hnRNP K (Fig. 3a, b). However, we found that the cytoplasmic puncta of mislocalised hnRNP K in pyramidal neurons did not colocalise with cytoplasmic inclusions of TDP-43 in FTL-D-TDP A (Fig. 3c). Indeed, a similar double-negative result was obtained with antibodies against the classical inclusion marker SQSTM1/p62 (Fig. 3d, e). Hence cytoplasmic hnRNP K puncta are unlikely to be components of ubiquitinated inclusions.

Indeed, neurons with hnRNP K pathology were also found to be distinct from those with tau-positive inclusions within the frontal cortex of FTL-D-Tau. Tau inclusions were readily identifiable within the pyramidal cell layers, but again, were not found within neurons exhibiting hnRNP K mislocalisation (Supplementary Fig. 2a, b, Online resource 2). In further attempts to characterise the subcellular location of hnRNP K cytoplasmic puncta in mislocalised neurons, double fluorescence was also performed with mitochondrial marker voltage-dependent anion channel (VDAC), classical autophagosome marker LC3 and stress granule/RNA-binding protein GTPase-activating protein-binding protein 2 (G3BP2). VDAC, LC3 and G3BP2 staining was principally cytoplasmic but no marker was enriched at the site of hnRNP K puncta (Supplementary Fig. 3a–f, Online resource 2).

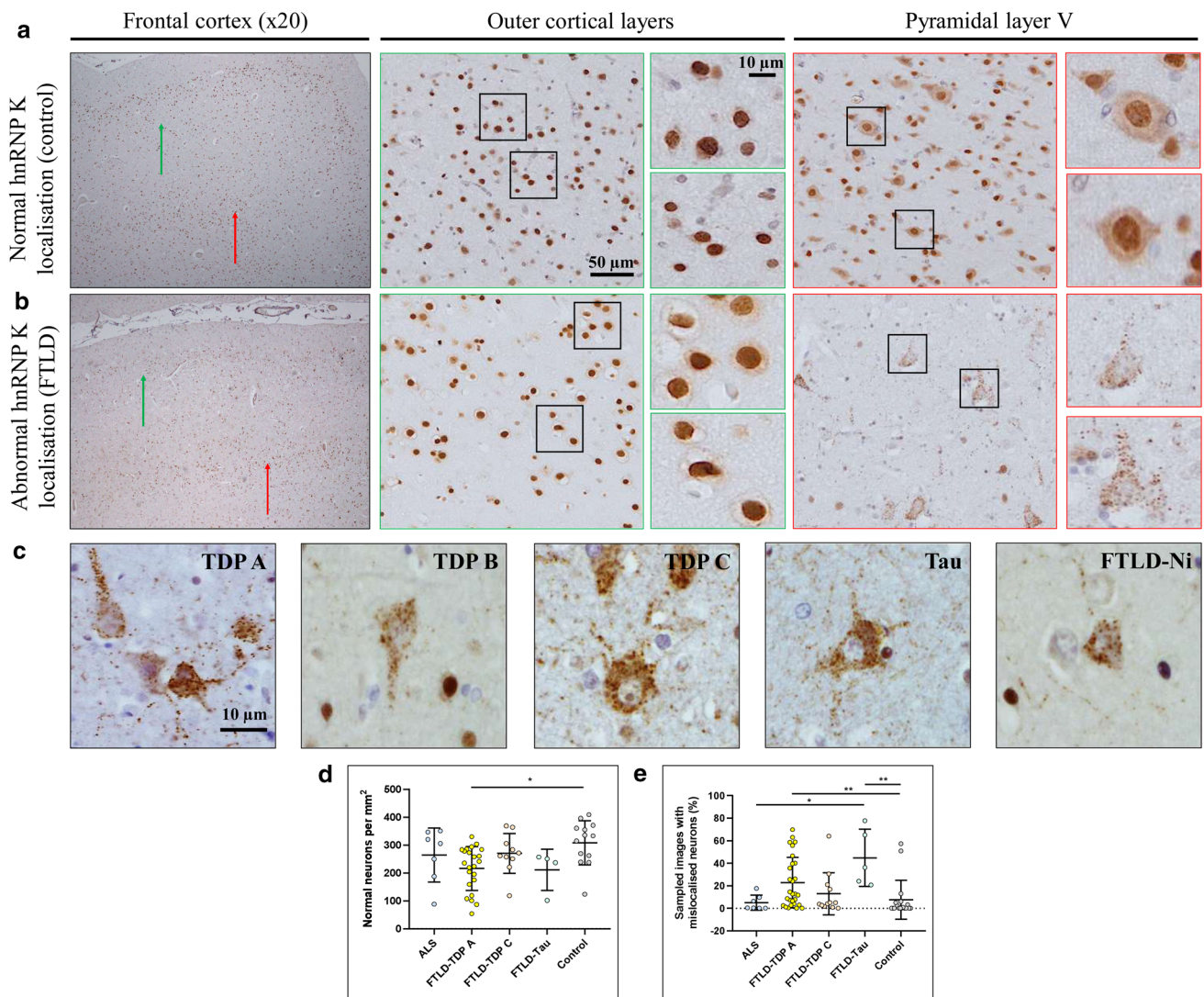


Fig. 1 Immunohistochemical staining and quantitation of hnRNP K neuronal pathology in FTLN and control subjects. **a** Normal nuclear localisation of hnRNP K localisation within neurons of outer and pyramidal cortical layers. **b** FTLN-TDP A case with normal neuronal staining within outer cortical layers and abnormal staining within pyramidal neurons as defined by hnRNP K nuclear depletion and accumulation of cytoplasmic puncta. **c** Examples of hnRNP K mislocalisation in pyramidal neurons of FTLN subtypes (from left to right): TDP A, TDP B, TDP C, Tau and FTLN-Ni (no inclusions).

d Age-matched controls ($n = 18$) showed significantly more normal hnRNP K-stained neurons per mm² region analysed than the largest disease group, FTLN-TDP A ($n = 28$) ($p = 0.013$; Tukey's test). **e** FTLN-TDP A and FTLN-Tau cases exhibited significantly more hnRNP K mislocalisation than controls expressed as proportion of sampled images with at least 1 (>0) abnormal neuron ($p = 0.004$, $p = 0.002$). FTLN-Tau cases also exhibited significantly more mislocalised hnRNP K than ALS cases using both metrics of abnormality ($p = 0.04$)

HnRNP K knockdown leads to widespread changes in gene expression and splicing

HnRNP K plays a major role in RNA processing in the nucleus. We hypothesised that loss of nuclear hnRNP K leads to aberrant processing of RNA targets,

including cryptic exon events. To study this, we knocked down hnRNP K using an siRNA in SH-SY5Y neuroblastoma cells, and performed RNA sequencing (three replicates of each condition). Knockdown was confirmed by western blot (Fig. 4a, b) and RT-PCR (Fig. 4c). Differential expression revealed widespread changes in the transcriptome caused by

hnRNP K knockdown, with 1284 genes upregulated and 428 downregulated (\log_2 fold change > 1 , FDR < 0.05) (Fig. 4d).

Differential splicing analysis found 1039 cassette exons with altered splicing in both directions (Fig. 5a) ($|\Delta\text{PSI}| > 10\%$, FDR < 0.05), suggesting that hnRNPK has a widespread role in maintaining both gene expression and splicing throughout the neuronal transcriptome. Despite discovering comparable numbers of genes and splicing events, we focussed our attention on splicing. This was due to both the previous evidence of hnRNP K directly binding mRNA and modulating splicing [65], and the possibility of observing hnRNP K-specific splicing events in human post-mortem brains.

Using GENCODE (v30) to annotate the introns used in each cassette exon splicing event, we discovered a large number of hnRNP K-associated cassette exons contained either novel inclusion or novel skipping junctions. Following our previous work on identifying cryptic exons associated with TDP-43 [21, 30, 57], we defined cryptic exons as cassette exons that were lowly included in control samples ($\text{PSI}_{\text{control}} < 10\%$) with a strong increase in inclusion upon hnRNP K knockdown ($\Delta\text{PSI} > 10\%$). We identified 101 exons that met our criteria, of which 49 were novel. The inverse of cryptic exons are skiptic exons, where normally constitutively included exons ($\text{PSI}_{\text{control}} > 90\%$) are skipped in response to perturbation of splicing factors ($\Delta\text{PSI} < -10\%$) [21]. We identified 261 skipped exons which met these criteria, 118 of which were novel skipping events.

The presence of cryptic or skiptic exons can lead to degradation of the host transcript by nonsense-mediated decay through frame-shifting or the introduction of a premature stop codon [30]. We compared the direction of differential gene expression with the genes found to harbour cryptic

and skiptic exons, as well as other differentially included and skipped exons. Non-canonically spliced genes had a bias towards downregulation, with 34 of the 46 differentially expressed cryptic exon containing genes (74%) downregulated, compared to 58% of a null set of genes matched for similar expression levels ($p = 9.7e-6$; Chi-squared test of equal proportion; Supplementary Fig. 4a, Online resource 2). However, only a small number of cryptic or skiptic exon-containing genes were strongly differentially expressed (\log_2 fold change > 1 ; Supplementary Fig. 4b). Cryptic exons, and novel exons in general, were less likely to be divisible by three compared to non-cryptic included and annotated exons, suggesting that many could cause frameshifting and protein degradation by nonsense-mediated decay (Supplementary Fig. 4c, d, Online resource 2).

The cryptic exons repressed by TDP-43 are largely non-conserved between species [30, 40], whereas the constitutive exons affected by skiptic splicing tend to be highly conserved [21]. We calculated mean PhyloP 30-way mammalian evolutionary conservation scores for each exon sequence (Fig. 5b). Cryptic exons were less conserved compared with the other included cassette exons ($p < 1e-16$, Wilcoxon test), whereas skiptic exons were slightly more conserved than the rest of the skipped exons ($p = 0.048$). Splitting each group by annotation showed the reduced conservation of cryptic exons was largely due to novel (non-annotated in GENCODE) exons (Supplementary Fig. 4c, Online resource 2).

We next investigated the sequence specificity of hnRNP K-associated splicing events. hnRNP K is known to have a binding preference for cytosine-rich sequences [46, 55, 60]. We took the genomic sequences for the exons strongly included and strongly skipped ($|\Delta\text{PSI}| > 10\%$) by hnRNP K knockdown and performed differential motif enrichment on the exon sequences, the upstream 100

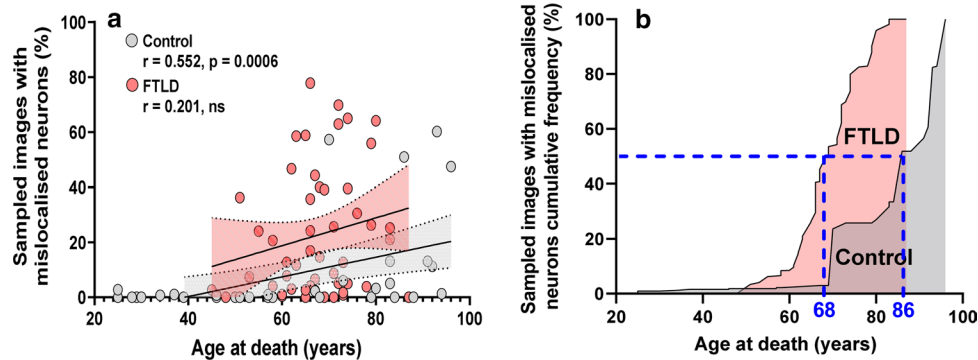


Fig. 2 HnRNP K mislocalisation is an age-related pathology that is advanced in FTLD. **a** HnRNP K mislocalisation in controls (in grey, $n = 35$) was found to positively correlate with age at death (Spearman's $r = 0.552$, $p = 0.0006$). HnRNP K mislocalisation in combined FTLD-TDP and FTLD-Tau subjects (in red, $n = 50$) was found to be only weakly associated with age at death (Spearman's $r = 0.201$,

ns). **b** Cumulative frequency plot of hnRNP K mislocalisation with ascending age normalised to total amount of quantified mislocalisation for each control/disease group. The intra-group median frequency of hnRNP K mislocalisation within the FTLD-TDP A group was 18 years in advance of the control cohort (dotted blue lines)

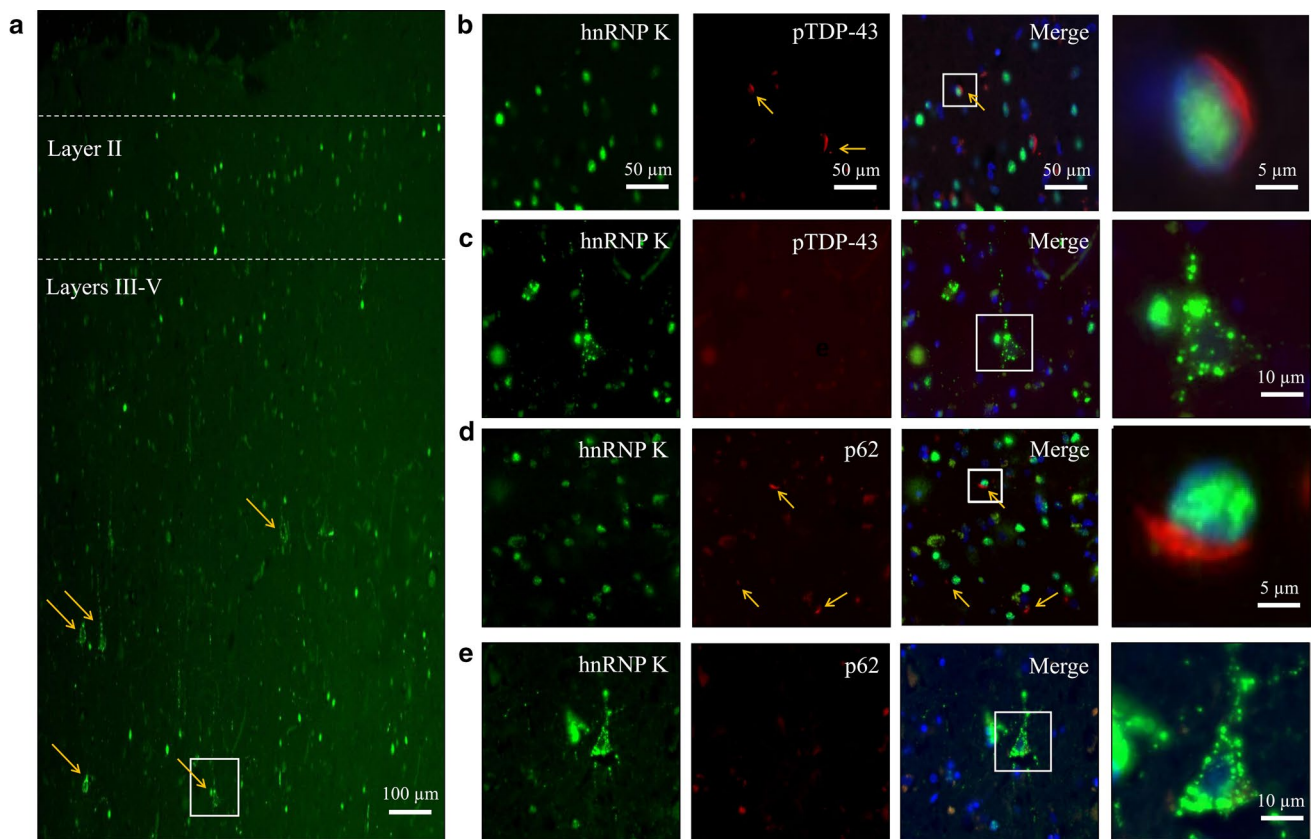


Fig. 3 Neurons that exhibit hnRNP K mislocalisation are independent of those that contain TDP-43 inclusions and hnRNP K pathology is p62-negative. **a** HnRNP K immunofluorescence of frontal cortex (10 ×) in FTLD-TDP A showing the relative anatomical positioning of normal (layer II) and abnormal (pyramidal layers III & V) hnRNP K localisation. Arrows point to neurons with hnRNP K mislocalisation. **b** Representative images of double-label immunofluorescence in layer II neurons exhibiting normal hnRNP K staining with TDP-43 inclusions. **c** Representative images of pyramidal layer V neurons

with mislocalised hnRNP K (as boxed in **a**) but no TDP-43 pathology, demonstrating that hnRNP K mislocalisation and TDP-43 pathologies do not co-occur in the same neurons. **d, e** Double-label immunofluorescence of hnRNP K and p62 in FTLD-TDP A cortical layers II and V showing neurons with p62-positive inclusions have normal nuclear localisation of hnRNP K (**d**) and that cytoplasmic puncta in pyramidal neurons with hnRNP K mislocalisation are p62-negative (**e**). Arrows point to TDP-43/p62-positive inclusions and scale bars are as indicated in the first row unless otherwise stated

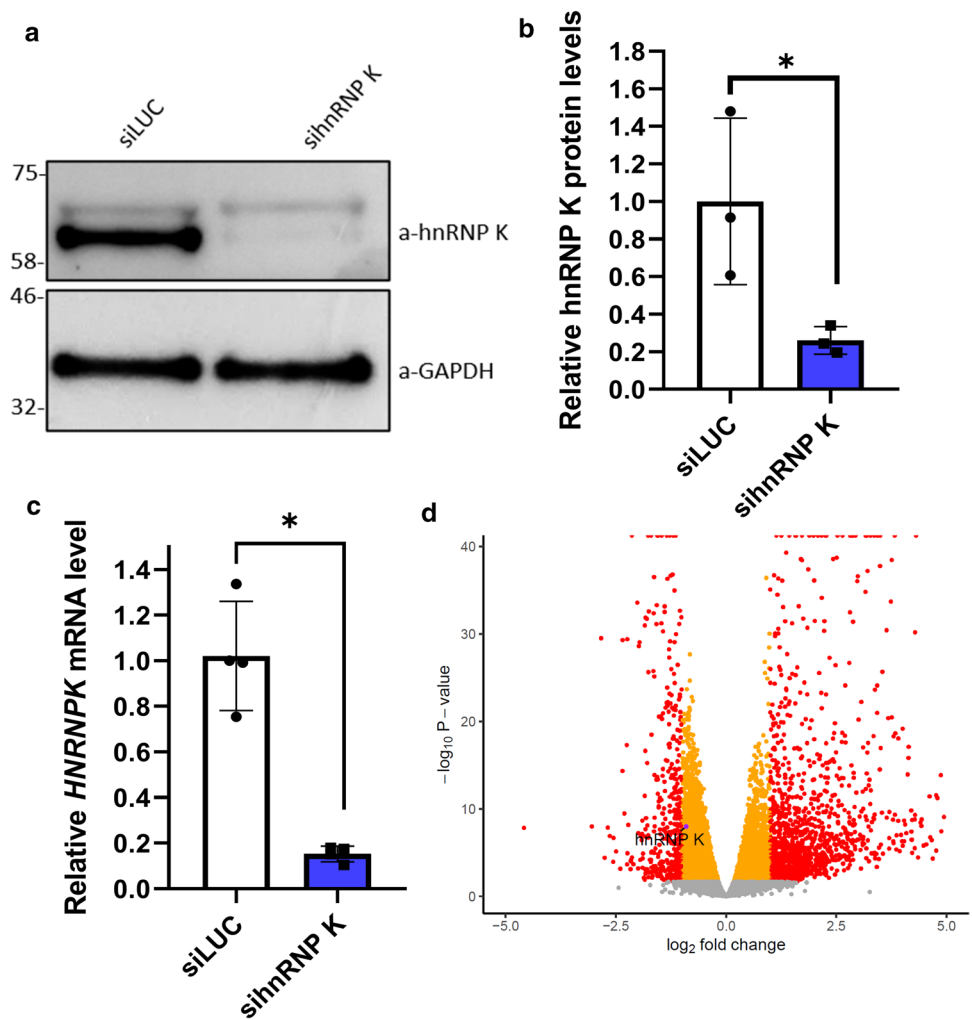
nucleotides and the downstream 100 nucleotides flanking either side (Fig. 5c). The C-rich motif *CCC[AT]* (square brackets refer to the presence of either nucleotide) was enriched in both the exon bodies and the 5' flanking sequences of included exons. We observed the *CCC[AT]* sequence within 77% of included exons but only 50% of skipped exons (OR = 1.55; $P = 1e-17$, Chi-squared). When restricting the analysis to only cryptic and skiptic exons this rises to 86% against 48% (OR = 1.8; $P = 7.7e-11$; Supplementary Fig. 3d, Online resource 2). Comparing the skipped exons to the included exons we identified the motif *[TA]G[GC]AG* present in 50% of included exons, but 69% of skipped exons (OR = 0.73; $P = 2.9e-9$; Chi-squared). *[AC]C[AT]G* was also identified in the 3' flanking sequences of the skipped exons. We verified these motifs by extracting all possible 4-mer nucleotides in the skipped and included exon sets and comparing their frequencies (Fig. 5d, Supplementary

Table 6, Online resource 1). C-rich 4-mers were enriched in the exon body, upstream 5' and downstream 3' flanking sequences of the included exons when compared to the skipped exons. Conversely, we observed a cluster of TG- and CG-rich sequences enriched in the skipped exon bodies sequences, with no 4-mers enriched in the flanking nucleotides of the skipped exons. This suggests that the exons differentially included under hnRNP K knockdown are directly bound by hnRNP K, whereas the skipped and skiptic exons may be regulated indirectly through other splicing factors.

HnRNP K mislocalisation in FTLD human brain is associated with the inclusion of cryptic exons

HnRNP K knockdown led to the inclusion of 49 cryptic exons which were identified as novel. We proceeded to validate five strong cryptic exon events in RNA targets

Fig. 4 siRNA knockdown of hnRNP K in SH-SY5Y cells leads to widespread gene expression changes. **a** siRNA knockdown of hnRNP K leads to a reduction in protein levels of hnRNP K when compared to siLUC (control) in SH-SY5Y cells. **b** Quantification of western blot analysis of hnRNP K protein levels relative to GAPDH. Protein levels of hnRNP K are reduced to 25% in sihnRNP K SH-SY5Y cells compared to control ($n=3$, $p=0.046$, unpaired t test). **c** siRNA knockdown of hnRNP K leads to a reduction in mRNA levels of hnRNP k when compared to siLUC (control) in SH-SY5Y cells. ($n=4$, $p=0.029$, Mann–Whitney test). **d** Differential gene expression of hnRNP K knockdown finds 1284 genes upregulated and 428 downregulated (FDR < 0.05, \log_2 fold change > 1, genes coloured red). HnRNP K is moderately downregulated (FDR < 0.05, \log_2 fold change = 0.91, equivalent to 53% reduction, orange points). Genes with $p < 1e-40$ are plotted at the edge of axis



for *HMBOX1*, *TMEM132A*, *CACTIN1*, *B4GALNT4* and *FAM160B2* (Fig. 6a). All five of these events had a deltaPSI > 25% in SH-SY5Y hnRNP K knockdown and are strongly expressed in human brain. A three-primer PCR protocol was designed to generate amplicons containing the cryptic exon. Cryptic exons within these five targets were significantly upregulated in hnRNP K knockdown (sihnRNP K) cells compared to control (siLUC) SH-SY5Y cells (*HMBOX1* ($p=0.003$), *TMEM132A* ($p=0.013$), *CACTIN1* ($p=0.004$), *B4GALNT4* ($p=0.033$) and *FAM160B2* ($p=0.008$), unpaired t test). (Fig. 6b, c). Therefore, hnRNP K knockdown leads to widespread cryptic exon incorporation. We proceeded to investigate whether hnRNP K mislocalisation observed in the frontal cortex of FTLN brains was associated with an increased inclusion of cryptic exons in hnRNP K targets. We found that inclusion of a cryptic exon in *FAM160B2* was significantly upregulated in cases exhibiting widespread hnRNP K mislocalisation ($n=12$, mean of 44% of sampled images containing at least one mislocalised neuron) compared to cases with predominantly nuclear hnRNP K ($n=9$, mean of 2% of sampled images with at least

one mislocalised neuron) (Mann–Whitney test, $p=0.019$) (Fig. 6d, e). This suggests that loss of nuclear hnRNP K in the frontal cortex is associated with the inclusion of cryptic exons in targets of hnRNP K.

Discussion

We have identified a novel pathological event in neurodegenerative disease characterised by the mislocalisation of hnRNP K from the nucleus to the cytoplasm within pyramidal neurons of the frontal cortex in a highly punctate manner. HnRNP K mislocalisation was more frequent in FTLN-TDP A and FTLN-Tau patient brains than in age-matched controls. Mislocalisation was also observed in some elderly controls and indeed was found to correlate with age at death. Hence, it is tempting to speculate its higher occurrence in FTLN may reflect an advanced-ageing phenotype.

Mislocalisation of hnRNP K occurs in neurons that are distinct from those which harbour proteinaceous TDP-43 and Tau inclusions that pathologically define FTLN subtypes.

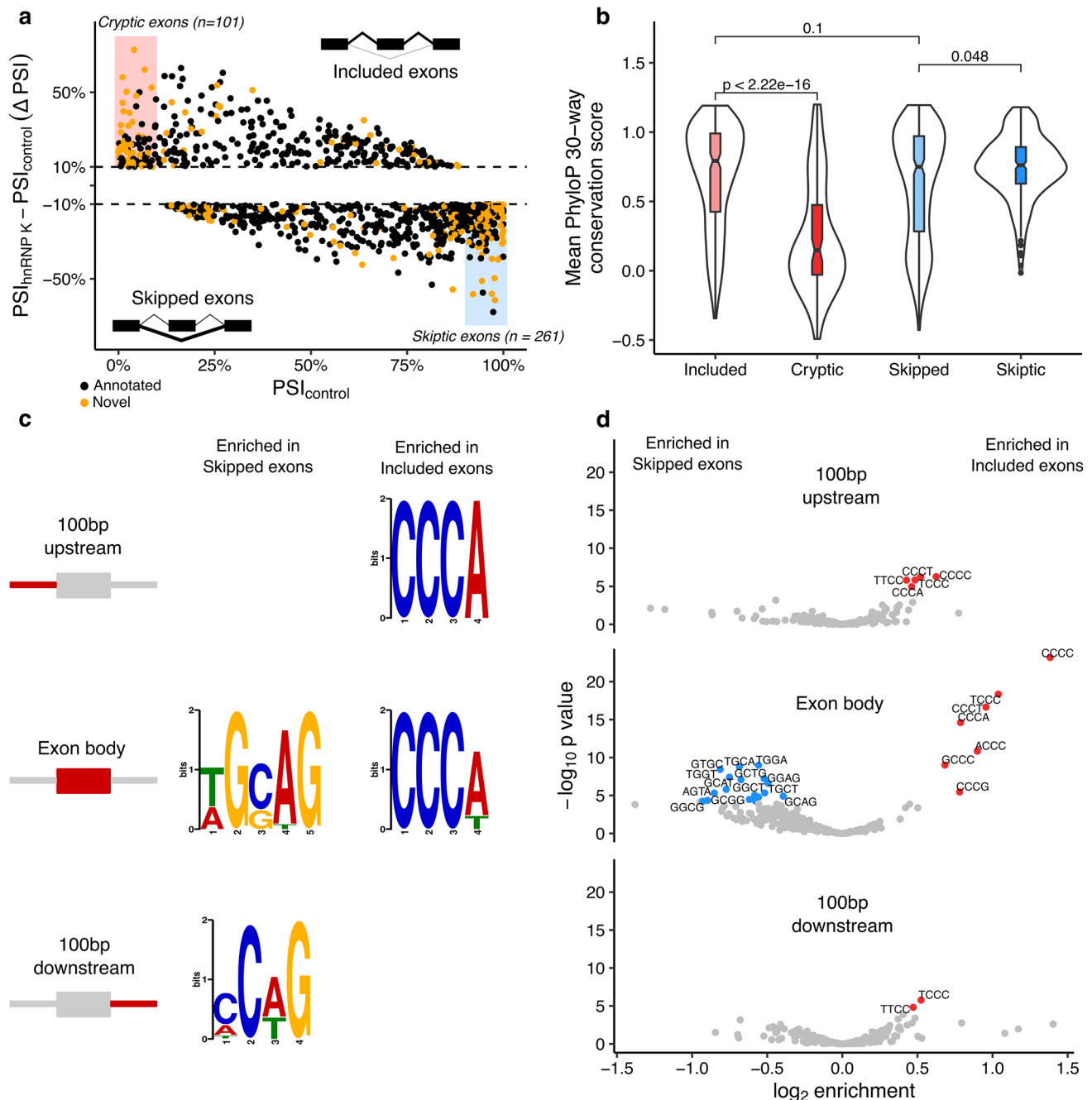


Fig. 5 HnRNP K depletion leads to widespread novel splicing. **a** All cassette exons identified with absolute relative inclusion (percentage spliced in, ΔPSI) $> 10\%$. Included exons have a positive ΔPSI and skipped exons have a negative ΔPSI . Exons lowly included in controls ($\text{PSI}_{\text{control}} < 10\%$) classified as cryptic exons (shaded red). Exons highly included in controls ($\text{PSI}_{\text{control}} > 90\%$) classified as skip-tic exons (shaded blue). **b** Mean phyloP mammalian conservation

scores for each class of exon. p values are from pairwise nonparametric Wilcoxon tests. **c** The top motifs identified by MEME when comparing the nucleotide sequences of all skipped exons (left column) and all included exons (right column) to the opposing set, split into the 100 bp upstream, the exon body and 100 bp downstream nucleotides. **d** Enrichments of 4-mers between the same sequence sets. p values are from a Chi-squared test of equal proportions

This is novel compared to previous work which has identified several other hnRNPs to be present within TDP-43 and FUS inclusions [15, 24]. There is also an abundance of research on hnRNPs co-depositing with C9orf72-associated

pathologies including RNA foci and dipeptide repeat proteins [13, 14, 25, 59]. Interestingly, cytoplasmic hnRNP K puncta within neurons were found to be p62/ubiquitin negative suggesting they are not typical of classical pathological

inclusions. We found no colocalisation between hnRNP K puncta and markers of mitochondria, autophagosomes and stress granules. Further investigations, including additional double immunofluorescence, co-immunoprecipitation and biochemical fractionation studies, will be required to characterise the cytoplasmic hnRNP K puncta we observe in pyramidal cells.

We have also identified hnRNP K to have a key physiological role in repressing cryptic splicing in neurons. Using our cryptic exon discovery pipeline, we have identified a large ($n = 101$) number of cryptic exons which are significantly upregulated within our hnRNP K knockdown (SH-SY5Y) cell model at the RNA-seq level, five of which we

validated in RNA by PCR. These cryptic exons resemble those found in TDP-43 knockdown experiments in that they are poorly conserved across evolution and likely contribute to reducing the expression of their host genes. Using motif analysis we have demonstrated that the widespread exon inclusion events are likely due to a loss of direct hnRNP K binding to C-rich RNA elements.

We have observed hnRNP K to be frequently and severely depleted from the nucleus of pyramidal neurons in many of our analysed FTLD brains and so hypothesised there to be an elevated degree of cryptic splicing within these cases. Indeed we validated one such event in *FAM160B2* in RNA derived from bulk frontal cortex homogenate. Validation of other

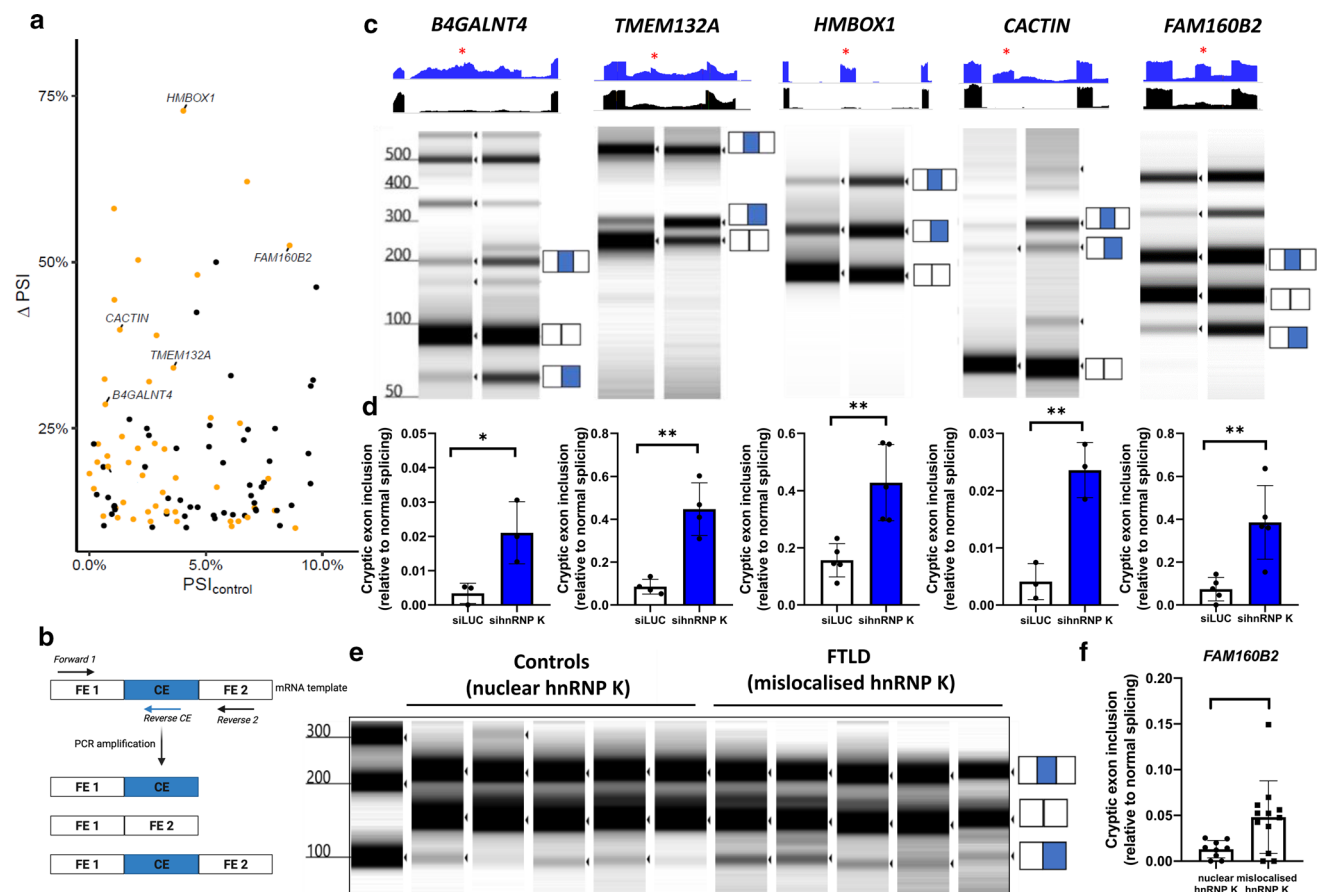


Fig. 6 HnRNP K-regulated cryptic exons can be identified in FTLD patient brains. **a** $PSI_{control}$ against Δ PSI for the 101 cryptic exons identified in SH-SY5Y cells. Novel exons are coloured orange. **b** Schematic diagram representing a 3 primer PCR method. A forward and reverse primer on the flanking exons (FE) as well as a reverse primer on the cryptic exon (CE) were used to generate the 3 PCR products shown. **c** Five novel cryptic events were validated in SH-SY5Y cells as a result of sihnRNP K knockdown. The red asterisk (*) in each IGV (integrated genome viewer) trace indicates the cryptic exon. The top trace (blue) corresponds to the sihnRNP K reads, whilst the bottom trace (black) corresponds to the control (siLUC). The presence of the cryptic exon was demonstrated using a 3 primer PCR utilizing primers that flank the cryptic exon (blue), as well as a

primer spanning the cryptic exon. **d** Quantification of inclusion of the cryptic exon relative to normal splicing for each of the five events. All five events were more present in sihnRNP K (blue bar) cells compared to siLUC (control, white bar) cells. (*B4GALNT4* ($p = 0.033$), *TMEM132A* ($p = 0.013$), *HMBOX1* ($p = 0.003$), *CACTIN1* ($p = 0.004$), and *FAM160B2* ($p = 0.008$), unpaired t test). **e** Representative image of products from a 3 primer PCR showing an increase in the inclusion of the cryptic exon within *FAM160B2* in FTLD cases with mislocalised hnRNP K. **f** Quantification of inclusion of the cryptic exon relative to normal splicing in *FAM160B2* in cases exhibiting nuclear hnRNP K and cases exhibiting mislocalised hnRNP K ($n = 9$ nuclear hnRNP K, $n = 12$ mislocalised hnRNP K, Mann–Whitney test, $*p = 0.019$)

cryptic events in addition to *FAM160B2* have been tried but without the success we have obtained for this specific splicing change. However, it must be pointed out that these analyses were performed in “bulk” tissue, and therefore, the sensitivity of the approach was distant from ideal. In the future, other types of approaches such as FISH or CISH techniques will be used to obtain a better appraisal of cryptic exon inclusion in the presence of hnRNP K nuclear depletion. Nonetheless, the specific detection of the *FAM160B2* in patient samples that exhibit hnRNP K mislocalisation served as a proof-of-principle that loss of nuclear hnRNP K in human brain can result in splicing changes. Hence, like TDP-43 and indeed several other RBPs, hnRNP K appears to have a crucial homeostatic role in maintaining transcriptional stability that may be compromised in FTLD.

Future work utilising single-cell approaches and alternative neuronal knockdown models of hnRNP K will be important to identify cryptic events that are mechanistically relevant to the neurodegeneration phenotype. Events that lead to functional depletion of proteins that enhance neuronal vulnerability (analogous to truncated stathmin-2 in TDP-43 depletion models) or even the elevation of novel protein isoforms that evade nonsense-mediated decay altogether will be of special interest [48, 57]. Indeed, investigating the potential interaction between hnRNP K and TDP-43 within FTLD-related RNA misprocessing may be of mechanistic interest in itself. The two RBPs have been found to colocalise within stress granules (SGs), with TDP-43 accumulation into SGs being dependent on prior phosphorylation of hnRNP K by cyclin-dependent kinase 2 [52]. Both proteins were also found to be robustly nuclear depleted within iPSC-derived motor neurons subjected to osmotic stress [26]. HnRNP K has also been identified as a key driver for nuclear retention of long non-coding RNA (lncRNA) Malat1 which, when disrupted, leads to increased Malat1-TDP-43 binding and an increased propensity for TDP-43 aggregation as has been observed in ALS patients [54, 62]. Therefore, protein levels of hnRNP K as well as its subcellular localisation and phosphorylation status may be mechanistically crucial in maintaining normal stress granule assembly/dissolution and TDP-43 proteostasis [52, 54]. In support of an interplay between these proteins, hnRNP K expression was found to be impaired in several TDP-43 mutant cell and murine models and was also found to be a modifier of TDP-43 pathology in *Drosophila* and cell-based models of TDP-43 overexpression/depletion [1, 51]. Potentially overlapping cryptic hits could illuminate converging pathways and networks of dysregulation in TDP-43 proteinopathies.

Further lines of enquiry may include those aimed at understanding the specific vulnerability of pyramidal

neurons. Laser capture microdissection and single-cell RNA-seq paradigms will prove powerful in delineating transcriptomic differences between cortical neuronal cell types. As is the case with several other hnRNPs, it will also be important to determine whether hnRNP K autoregulates its own expression and whether or not this mechanism is dysregulated in neurons with mislocalised protein. Establishing whether any sequence mutations or post-translational modifications of hnRNP K are more or less associated with mislocalisation will also be insightful in this regard.

To conclude, our discovery of hnRNP K mislocalisation in FTLD and induced cryptic exon inclusion following hnRNP K nuclear depletion adds to the growing body of research of RNA misprocessing in FTLD and ALS pathogenesis. HnRNP K joins other hnRNP proteins that have been found to be pathologically dysregulated in these disorders including TDP-43, FUS and hnRNP A1. It will also be important to establish whether hnRNP K mislocalisation and its impact on RNA splicing is relevant to other neurodegenerative diseases. A greater understanding of the pathomechanistic underpinnings and consequences of hnRNP K mislocalisation in neurodegeneration may have the potential to identify novel biomarkers and therapeutic targets.

Data and code availability

Raw RNA sequencing data, as well as processed gene expression and splice junction counts, have been uploaded to the Gene Expression Omnibus (GEO) with accession GSE171090. All code used to generate figures is hosted on github: <https://github.com/jackhump/hnRNPK/>. Script for classifying cassette exons from Leafcutter results: https://github.com/davidaknowles/leafcutter/blob/classify_clusters/leafviz/classify_clusters.R.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00401-021-02340-0>.

Acknowledgements AB is supported by the Wolfson Foundation and Eisai. AG is supported by Rosetree Trust. JH and TR are funded by grants from the US National Institutes of Health (NIH NIA R56-AG055824 and NIA U01-AG068880). SC and EB are supported by a grant from AriSLA (project PathensTDP). DB is funded by a Future and Emerging Technology (FET) Horizon 2020 grant (964800). SF is supported by the Alzheimer’s Society. A-LB is funded by the NIHR biomedical research centre (BRC) and the UK Medical Research Council (MRC). MF is supported by the Associazione Italiana per la Ricerca sul Cancro (AIRC). PF is funded by the MRC [MR/M008606/1 and MR/S006508/1], the UK Motor Neurone Disease Association, Rosetrees Trust and the UCLH NIHR Biomedical Research Centre. TL is supported by an Alzheimer’s Research UK senior fellowship. QSBB is supported by the Reta Lila Weston Institute for Neurological Studies.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Appocher C, Mohagheghi F, Cappelli S, Stuani C, Romano M, Feiguin F, Buratti E (2017) Major hnRNP proteins act as general TDP-43 functional modifiers both in *Drosophila* and human neuronal cells. *Nucleic Acids Res* 45:8026–8045. <https://doi.org/10.1093/nar/gkx477>
2. Attig J, de Los Mozos IR, Mozos I, Haberman N, Wang Z, Emmett W, Zarnack K, König J, Ule J (2016) Splicing repression allows the gradual emergence of new Alu-exons in primate evolution. *Elife*. <https://doi.org/10.7554/eLife.19545>
3. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* 37:W202–W208. <https://doi.org/10.1093/nar/gkp335>
4. Bampton A, Gittings LM, Fratta P, Lashley T, Gatt A (2020) The role of hnRNPs in frontotemporal dementia and amyotrophic lateral sclerosis. *Acta Neuropathol* 140:599–623. <https://doi.org/10.1007/s00401-020-02203-0>
5. Barboro P, Ferrari N, Balbi C (2014) Emerging roles of heterogeneous nuclear ribonucleoprotein K (hnRNP K) in cancer progression. *Cancer Lett* 352:152–159. <https://doi.org/10.1016/j.canlet.2014.06.019>
6. Bomsztyk K, Denisenko O, Ostrowski J (2004) hnRNP K: one protein multiple processes. *BioEssays* 26:629–638. <https://doi.org/10.1002/bies.20048>
7. Bose JK, Wang I-F, Hung L, Tarn W-Y, Shen C-KJ (2008) TDP-43 overexpression enhances exon 7 inclusion during the survival of motor neuron pre-mRNA splicing. *J Biol Chem* 283:28852–28859. <https://doi.org/10.1074/jbc.M805376200>
8. Bott NT, Radke A, Stephens ML, Kramer JH (2014) Frontotemporal dementia: diagnosis, deficits and management. *Neurodegener Dis Manag* 4:439–454. <https://doi.org/10.2217/nmt.14.34>
9. Carpenter B, McKay M, Dundas SR, Lawrie LC, Telfer C, Murray GI (2006) Heterogeneous nuclear ribonucleoprotein K is over expressed, aberrantly localised and is associated with poor prognosis in colorectal cancer. *Br J Cancer* 95:921–927. <https://doi.org/10.1038/sj.bjc.6603349>
10. Chen X, Gu P, Xie R, Han J, Liu H, Wang B, Xie W, Xie W, Zhong G, Chen C, Xie S, Jiang N, Lin T, Huang J (2017) Heterogeneous nuclear ribonucleoprotein K is associated with poor prognosis and regulates proliferation and apoptosis in bladder cancer. *J Cell Mol Med* 21:1266–1279. <https://doi.org/10.1111/jcmm.12999>
11. Chkheidze AN, Liebhaber SA (2003) A Novel Set of Nuclear Localization Signals Determine Distributions of the CP RNA-Binding Proteins. *Mol Cell Biol* 23:8405–8415. <https://doi.org/10.1128/MCB.23.23.8405-8415.2003>
12. Convery R, Mead S, Rohrer JD (2019) Review: Clinical, genetic and neuroimaging features of frontotemporal dementia. *Neuropathol Appl Neurobiol* 45:6–18. <https://doi.org/10.1111/nan.12535>
13. Cooper-Knock J, Walsh MJ, Higginbottom A, Robin Highley J, Dickman MJ, Edbauer D, Ince PG, Wharton SB, Wilson SA, Kirby J, Hautbergue GM, Shaw PJ (2014) Sequestration of multiple RNA recognition motif-containing proteins by C9orf72 repeat expansions. *Brain* 137:2040–2051. <https://doi.org/10.1093/brain/awu120>
14. Davidson YS, Flood L, Robinson AC, Nihei Y, Mori K, Rollinson S, Richardson A, Benson BC, Jones M, Snowden JS, Pickering-Brown S, Haass C, Lashley T, Mann DMA (2017) Heterogeneous ribonuclear protein A3 (hnRNP A3) is present in dipeptide repeat protein containing inclusions in Frontotemporal Lobar Degeneration and Motor Neurone disease associated with expansions in C9orf72 gene. *Acta Neuropathol Commun* 5:31. <https://doi.org/10.1186/s40478-017-0437-5>
15. Davidson YS, Robinson AC, Flood L, Rollinson S, Benson BC, Asi YT, Richardson A, Jones M, Snowden JS, Pickering-Brown S, Lashley T, Mann DMA (2017) Heterogeneous ribonuclear protein E2 (hnRNP E2) is associated with TDP-43-immunoreactive neurites in Semantic Dementia but not with other TDP-43 pathological subtypes of Frontotemporal Lobar Degeneration. *Acta Neuropathol Commun* 5:54. <https://doi.org/10.1186/s40478-017-0454-4>
16. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21. <https://doi.org/10.1093/bioinformatics/bts635>
17. Dreyfuss G, Matunis MJ, Piñol-Roma S, Burd CG (1993) hnRNP proteins and the biogenesis of mRNA. *Annu Rev Biochem* 62:289–321. <https://doi.org/10.1146/annurev.bi.62.070193.001445>
18. Feng Y-Y, Ramu A, Cotto KC, Skidmore ZL, Kunisaki J, Conrad DF, Lin Y, Chapman W, Uppaulri R, Govindan R, Griffith OL, Griffith M (2018) RegTools: Integrated analysis of genomic and transcriptomic data for discovery of splicing variants in cancer. *BioRxiv*. <https://doi.org/10.1101/436634>
19. Folci A, Mapelli L, Sassone J, Prestori F, D'Angelo E, Bassani S, Passafaro M (2014) Loss of hnRNP K impairs synaptic plasticity in hippocampal neurons. *J Neurosci* 34:9088–9095. <https://doi.org/10.1523/JNEUROSCI.0303-14.2014>
20. Frankish A, Diekhans M, Ferreira A-M, Johnson R, Jungreis I, Loveland J, Mudge JM, Sisu C, Wright J, Armstrong J, Barnes I, Berry A, Bignell A, Carbonell Sala S, Chrast J, Cunningham F, Di Domenico T, Donaldson S, Fiddes IT, García Girón C, Flicek P (2019) GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res* 47:D766–D773. <https://doi.org/10.1093/nar/gky955>
21. Fratta P, Sivakumar P, Humphrey J, Lo K, Ricketts T, Oliveira H, Brito-Armas JM, Kalmar B, Ule A, Yu Y, Birsa N, Bodo C, Collins T, Conicella AE, Mejia Maza A, Marrero-Gagliardi A, Stewart M, Mianne J, Corrochano S, Emmett W, Acevedo-Arozena A (2018) Mice with endogenous TDP-43 mutations exhibit gain of splicing function and characteristics of amyotrophic lateral sclerosis. *EMBO J*. <https://doi.org/10.15252/embj.201798684>
22. Fukuda T, Naiki T, Saito M, Irie K (2009) hnRNP K interacts with RNA binding motif protein 42 and functions in the maintenance of cellular ATP level during stress conditions. *Genes Cells* 14:113–128. <https://doi.org/10.1111/j.1365-2443.2008.01256.x>

23. Geuens T, Bouhy D, Timmerman V (2016) The hnRNP family: insights into their role in health and disease. *Hum Genet* 135:851–867. <https://doi.org/10.1007/s00439-016-1683-5>
24. Gittings LM, Foti SC, Benson BC, Gami-Patel P, Isaacs AM, Lashley T (2019) Heterogeneous nuclear ribonucleoproteins R and Q accumulate in pathological inclusions in FTLD-FUS. *Acta Neuropathol Commun* 7:18. <https://doi.org/10.1186/s40478-019-0673-y>
25. Haeusler AR, Donnelly CJ, Periz G, Simko EAJ, Shaw PG, Kim M-S, Maragakis NJ, Troncoso JC, Pandey A, Sattler R, Rothstein JD, Wang J (2014) C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* 507:195–200. <https://doi.org/10.1038/nature13124>
26. Harley J, Patani R (2020) Stress-specific spatiotemporal responses of RNA-binding proteins in human stem-cell-derived motor neurons. *Int J Mol Sci*. <https://doi.org/10.3390/ijms21218346>
27. Hodges JR, Piguet O (2018) Progress and challenges in frontotemporal dementia research: a 20-year review. *J Alzheimers Dis* 62:1467–1480. <https://doi.org/10.3233/JAD-171087>
28. Hornbaker MJ, Gallardo M, Zhang X, Ma H, Hu P, Khoury JD, Kornblau SM, Bueso-Ramos CE, Post SM (2016) hnRNP K overexpression drives AML progression by altering pathways critical for myeloid proliferation and differentiation. *Blood* 128:744–744. <https://doi.org/10.1182/blood.V128.22.744.744>
29. Humphrey J, Birsna N, Milioto C, McLaughlin M, Ule AM, Robaldo D, Eberle AB, Kräuchi R, Bentham M, Brown A-L, Jarvis S, Bodo C, Garone MG, Devoy A, Soraru G, Rosa A, Bozzoni I, Fisher EMC, Mühlemann O, Schiavo G, Fratta P (2020) FUS ALS-causative mutations impair FUS autoregulation and splicing factor networks through intron retention. *Nucleic Acids Res* 48:6889–6905. <https://doi.org/10.1093/nar/gkaa410>
30. Humphrey J, Emmett W, Fratta P, Isaacs AM, Plagnol V (2017) Quantitative analysis of cryptic splicing associated with TDP-43 depletion. *BMC Med Genomics* 10:38. <https://doi.org/10.1186/s12920-017-0274-1>
31. Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, Kent WJ (2004) The UCSC Table Browser data retrieval tool. *Nucleic Acids Res* 32:D493–D496. <https://doi.org/10.1093/nar/gkh103>
32. Kent WJ, Zweig AS, Barber G, Hinrichs AS, Karolchik D (2010) BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* 26:2204–2207. <https://doi.org/10.1093/bioinformatics/btq351>
33. Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. *arXiv:1412.6980v9*
34. Klim JR, Williams LA, Limone F, Guerra San Juan I, Davis-Dusenbery BN, Mordes DA, Burberry A, Steinbaugh MJ, Gamage KK, Kirchner R, Moccia R, Cassel SH, Chen K, Wainger BJ, Woolf CJ, Eggan K (2019) ALS-implicated protein TDP-43 sustains levels of STMN2, a mediator of motor neuron growth and repair. *Nat Neurosci* 22:167–179. <https://doi.org/10.1038/s41593-018-0300-4>
35. Kovacs GG (2019) Molecular pathology of neurodegenerative diseases: principles and practice. *J Clin Pathol* 72:725–735. <https://doi.org/10.1136/jclinpath-2019-205952>
36. Lashley T, Rohrer JD, Mead S, Revesz T (2015) Review: an update on clinical, genetic and pathological aspects of frontotemporal lobar degenerations. *Neuropathol Appl Neurobiol* 41:858–881. <https://doi.org/10.1111/nan.12250>
37. Laursen LS, Chan CW, Ffrench-Constant C (2011) Translation of myelin basic protein mRNA in oligodendrocytes is regulated by integrin activation and hnRNP-K. *J Cell Biol* 192:797–811. <https://doi.org/10.1083/jcb.201007014>
38. Lee EB, Porta S, Michael Baer G, Xu Y, Suh E, Kwong LK, Elman L, Grossman M, Lee VM-Y, Irwin DJ, Van Deerlin VM, Trojanowski JQ (2017) Expansion of the classification of FTLD-TDP: distinct pathology associated with rapidly progressive frontotemporal degeneration. *Acta Neuropathol* 134:65–78. <https://doi.org/10.1007/s00401-017-1679-9>
39. Ling JP, Chhabra R, Merran JD, Schaughency PM, Wheelan SJ, Corden JL, Wong PC (2016) PTBP1 and PTBP2 repress nonconserved cryptic exons. *Cell Rep* 17:104–113. <https://doi.org/10.1016/j.celrep.2016.08.071>
40. Ling JP, Pletnikova O, Troncoso JC, Wong PC (2015) TDP-43 repression of nonconserved cryptic exons is compromised in ALS-FTD. *Science* 349:650–655. <https://doi.org/10.1126/science.aab0983>
41. Liu Y, Szaro BG (2011) hnRNP K post-transcriptionally coregulates multiple cytoskeletal genes needed for axonogenesis. *Development* 138:3079–3090. <https://doi.org/10.1242/dev.066993>
42. Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323. <https://doi.org/10.1186/1471-2105-12-323>
43. Li YI, Knowles DA, Humphrey J, Barbeira AN, Dickinson SP, Im HK, Pritchard JK (2018) Annotation-free quantification of RNA splicing using LeafCutter. *Nat Genet* 50:151–158. <https://doi.org/10.1038/s41588-017-0004-9>
44. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:550. <https://doi.org/10.1186/s13059-014-0550-8>
45. Mackenzie IRA, Ansoorge O, Strong M, Bilbao J, Zinman L, Ang L-C, Baker M, Stewart H, Eisen A, Rademakers R, Neumann M (2011) Pathological heterogeneity in amyotrophic lateral sclerosis with FUS mutations: two distinct patterns correlating with disease severity and mutation. *Acta Neuropathol* 122:87–98. <https://doi.org/10.1007/s00401-011-0838-7>
46. Matunis MJ, Michael WM, Dreyfuss G (1992) Characterization and primary structure of the poly(C)-binding heterogeneous nuclear ribonucleoprotein complex K protein. *Mol Cell Biol* 12:164–171. <https://doi.org/10.1128/mcb.12.1.164>
47. McClory SP, Lynch KW, Ling JP (2018) HnRNP L represses cryptic exons. *RNA* 24:761–768. <https://doi.org/10.1261/rna.065508.117>
48. Melamed Z, López-Erauskín J, Baughn MW, Zhang O, Drenner K, Sun Y, Freyermuth F, McMahon MA, Beccari MS, Artates JW, Ohkubo T, Rodriguez M, Lin N, Wu D, Bennett CF, Rigo F, Da Cruz S, Ravits J, Lagier-Tourenne C, Cleveland DW (2019) Premature polyadenylation-mediated loss of stathmin-2 is a hallmark of TDP-43-dependent neurodegeneration. *Nat Neurosci* 22:180–190. <https://doi.org/10.1038/s41593-018-0293-z>
49. Mercado PA, Ayala YM, Romano M, Buratti E, Baralle FE (2005) Depletion of TDP 43 overrides the need for exonic and intronic splicing enhancers in the human apoA-II gene. *Nucleic Acids Res* 33:6000–6010. <https://doi.org/10.1093/nar/gki897>
50. Michael WM, Eder PS, Dreyfuss G (1997) The K nuclear shuttling domain: a novel signal for nuclear import and nuclear export in the hnRNP K protein. *EMBO J* 16:3587–3598. <https://doi.org/10.1093/emboj/16.12.3587>
51. Moujalled D, Grubman A, Acevedo K, Yang S, Ke YD, Moujalled DM, Duncan C, Caragounis A, Perera ND, Turner BJ, Prudencio M, Petrucelli L, Blair I, Ittner LM, Crouch PJ, Liddell JR, White AR (2017) TDP-43 mutations causing amyotrophic lateral sclerosis are associated with altered expression of RNA-binding protein hnRNP K and affect the Nrf2 antioxidant pathway. *Hum Mol Genet* 26:1732–1746. <https://doi.org/10.1093/hmg/ddx093>
52. Moujalled D, James JL, Yang S, Zhang K, Duncan C, Moujalled DM, Parker SJ, Caragounis A, Lidgerwood G, Turner BJ, Atkin JD, Grubman A, Liddell JR, Proepper C, Boeckers TM, Kanninen KM, Blair I, Crouch PJ, White AR (2015) Phosphorylation of hnRNP K by cyclin-dependent kinase 2 controls cytosolic accumulation of TDP-43. *Hum Mol Genet* 24:1655–1669. <https://doi.org/10.1093/hmg/ddu578>

53. Neumann M, Roeber S, Kretschmar HA, Rademakers R, Baker M, Mackenzie IRA (2009) Abundant FUS-immunoreactive pathology in neuronal intermediate filament inclusion disease. *Acta Neuropathol* 118:605–616. <https://doi.org/10.1007/s00401-009-0581-5>
54. Nguyen TM, Kabotyanski EB, Reineke LC, Shao J, Xiong F, Lee J-H, Dubrulle J, Johnson H, Stossi F, Tsui PS, Choi K-J, Ellis AG, Zhao N, Cao J, Adewunmi O, Ferreon JC, Ferreon ACM, Neilson JR, Mancini MA, Chen X, Rosen JM (2020) The SINEB1 element in the long non-coding RNA Malat1 is necessary for TDP-43 proteostasis. *Nucleic Acids Res* 48:2621–2642. <https://doi.org/10.1093/nar/gkz1176>
55. Paziewska A, Wyrwicz LS, Bujnicki JM, Bomszyk K, Ostrowski J (2004) Cooperative binding of the hnRNP K three KH domains to mRNA targets. *FEBS Lett* 577:134–140. <https://doi.org/10.1016/j.febslet.2004.08.086>
56. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A (2010) Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res* 20:110–121. <https://doi.org/10.1101/gr.097857.109>
57. Prudencio M, Humphrey J, Pickles S, Brown A-L, Hill SE, Kachergus JM, Shi J, Heckman MG, Spiegel MR, Cook C, Song Y, Yue M, Daugherty LM, Carlomagno Y, Jansen-West K, de Castro CF, DeTure M, Koga S, Wang Y-C, Sivakumar P, Petrucelli L (2020) Truncated stathmin-2 is a marker of TDP-43 pathology in frontotemporal dementia. *J Clin Invest* 130:6080–6092. <https://doi.org/10.1172/JCI139741>
58. Shiga A, Ishihara T, Miyashita A, Kuwabara M, Kato T, Watanabe N, Yamahira A, Kondo C, Yokoseki A, Takahashi M, Kuwano R, Kakita A, Nishizawa M, Takahashi H, Onodera O (2012) Alteration of POLDIP3 splicing associated with loss of function of TDP-43 in tissues affected with ALS. *PLoS ONE* 7:e43120. <https://doi.org/10.1371/journal.pone.0043120>
59. Suzuki H, Shibagaki Y, Hattori S, Matsuoka M (2019) C9-ALS/FTD-linked proline-arginine dipeptide repeat protein associates with paraspeckle components and increases paraspeckle formation. *Cell Death Dis* 10:746. <https://doi.org/10.1038/s41419-019-1983-5>
60. Swanson MS, Dreyfuss G (1988) Classification and purification of proteins of heterogeneous nuclear ribonucleoprotein particles by RNA-binding specificities. *Mol Cell Biol* 8:2237–2241. <https://doi.org/10.1128/mcb.8.5.2237>
61. Tan Q, Yalamanchili HK, Park J, De Maio A, Lu H-C, Wan Y-W, White JJ, Bondar VV, Sayegh LS, Liu X, Gao Y, Sillitoe RV, Orr HT, Liu Z, Zoghbi HY (2016) Extensive cryptic splicing upon loss of RBM17 and TDP43 in neurodegeneration models. *Hum Mol Genet* 25:5083–5093. <https://doi.org/10.1093/hmg/ddw337>
62. Tollervey JR, Wang Z, Hortobágyi T, Witten JT, Zarnack K, Kayikci M, Clark TA, Schweitzer AC, Rot G, Curk T, Zupan B, Rogelj B, Shaw CE, Ule J (2011) Analysis of alternative splicing associated with aging and neurodegeneration in the human brain. *Genome Res* 21:1572–1582. <https://doi.org/10.1101/gr.122226.111>
63. Trabzuni D, Ryten M, Walker R, Smith C, Imran S, Ramasamy A, Weale ME, Hardy J (2011) Quality control parameters on a large dataset of regionally dissected human control brains for whole genome expression studies. *J Neurochem* 119:275–282. <https://doi.org/10.1111/j.1471-4159.2011.07432.x>
64. Trabzuni D, Wray S, Vandrovicova J, Ramasamy A, Walker R, Smith C, Luk C, Gibbs JR, Dillman A, Hernandez DG, Arepalli S, Singleton AB, Cookson MR, Pittman AM, de Silva R, Weale ME, Hardy J, Ryten M (2012) MAPT expression and splicing is differentially regulated by brain region: relation to genotype and implication for tauopathies. *Hum Mol Genet* 21:4094–4103. <https://doi.org/10.1093/hmg/dds238>
65. Van Nostrand EL, Freese P, Pratt GA, Wang X, Wei X, Xiao R, Blue SM, Chen J-Y, Cody NAL, Dominguez D, Olson S, Sundararaman B, Zhan L, Bazile C, Bouvrette LPB, Bergalet J, Duff MO, Garcia KE, Gelboin-Burkhart C, Hochman M, Yeo GW (2020) A large-scale binding and functional map of human RNA-binding proteins. *Nature* 583:711–719. <https://doi.org/10.1038/s41586-020-2077-3>
66. Warren JD, Rohrer JD, Rossor MN (2013) Clinical review. Frontotemporal dementia. *BMJ* 347:f4827. <https://doi.org/10.1136/bmj.f4827>
67. White MA, Kim E, Duffy A, Adalbert R, Phillips BU, Peters OM, Stephenson J, Yang S, Massenzio F, Lin Z, Andrews S, Segonds-Pichon A, Metterville J, Saksida LM, Mead R, Ribchester RR, Barhomi Y, Serre T, Coleman MP, Fallon JR, Sreedharan J (2018) TDP-43 gains function due to perturbed autoregulation in a Tardbp knock-in mouse model of ALS-FTD. *Nat Neurosci* 21:552–563. <https://doi.org/10.1038/s41593-018-0113-5>
68. Young JJ, Lavakumar M, Tampi D, Balachandran S, Tampi RR (2018) Frontotemporal dementia: latest evidence and clinical implications. *Ther Adv Psychopharmacol* 8:33–48. <https://doi.org/10.1177/2045125317739818>
69. Zarnack K, König J, Tajnik M, Martincorena I, Eustermann S, Stévant I, Reyes A, Anders S, Luscombe NM, Ule J (2013) Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements. *Cell* 152:453–466. <https://doi.org/10.1016/j.cell.2012.12.023>
70. Zhu A, Ibrahim JG, Love MI (2019) Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics* 35:2084–2092. <https://doi.org/10.1093/bioinformatics/bty895>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.