Diminished miRNA activity is associated with aberrant cytoplasmic intron retention in ALS pathogenesis

Marija Petric-Howe[1,2], Hamish Crerar[1,2], Jacob Neeves[1,2], Giulia E. Tyzack[1,2], Rickie Patani[1,2,#] , Raphaëlle Luisier[5,#]

[1]The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK; [2]Department of Molecular Neuroscience, UCL Institute of Neurology, Queen Square, London, UK; [4]Idiap Research Institute, Genomics and Health Informatics, Martigny, Switzerland; [#]These authors contributed equally to this work.

Key words: Cytoplasmic intron retention, human stem cell model, amyotrophic lateral sclerosis, miRNA

Correspondence should be addressed to Raphaëlle Luisier (raphaelle.luisier@idiap.ch) and Rickie Patani (rickie.patani@ucl.ac.uk).

SUMMARY

Intron retention (IR) is now recognized as a dominant splicing event during motor neuron (MN) development, however the role and regulation of intron-retaining transcripts (IRTs) localized to the cytoplasm remain particularly understudied. By resolving the spatiotemporal dynamics of IR underlying distinct stages of MN lineage restriction, we identify a cytoplasmic group of IRTs that is not associated with reduced expression of their own genes but instead with an upregulation of predicted target genes of specific miRNAs, the motifs of which are enriched within the intronic sequences of this group. Next, we show that ALS-causing VCP mutations lead to a selective increase in IR of this particular class of introns. This in turn temporally coincides with an increase in the expression level of predicted target genes of these miRNAs, providing a potential mechanistic insight into ALS pathogenesis. Altogether, we propose a novel role for the cytoplasmic intronic sequences in regulating miRNA activity through miRNA sequestration, which potentially contributes to ALS pathogenesis.

INTRODUCTION

Intron retention, a mode of alternative splicing whereby one or more introns are retained within a mature polyadenylated mRNA, has been greatly understudied in mammalian systems and for a long time mostly considered as a product of inefficient or mis-splicing. With advances in detection strategies, IR became recognised as a more widespread and regulated process than previously thought, and the idea that IR could even functionally modulate cellular processes has come into focus, with its role(s) in cellular physiology beginning to unfold (1, 2).

Neural cells exhibit a higher proportion of retained introns compared to other cell types and there is an expanding body of evidence demonstrating a

2

functional role for intron retention (IR) both in neuronal development and homeostasis (1, 3–5). Transcripts that exhibit IR often remain in the nucleus, mostly considered to be a means of reducing the expression levels of transcripts not required for cellular physiology at a particular stage (2, 6, 7). Some of these transcripts would eventually be degraded by the nuclear exosome, while specific signals could stimulate splicing of the retained intron in others, resulting in export of the fully spliced mRNA into the cytoplasm and its subsequent translation (5). Indeed, nuclear detention of intron-retaining transcripts (IRTs) provides a powerful mechanism to hold gene expression in a suppressed but poised state that allows rapid protein production if and when an appropriate stimulus is received (4, 5, 8–10).

Although the stable cytoplasmic localisation of intronic sequences in neurons has been reported since 2013 (11), there has been limited investigation into the possible role of cytoplasmic IRTs. This has presumably been overlooked in part due to detection limitations, but also due to a notion that these transcripts would likely contain premature translation termination codons (PTCs) and as such, be degraded by nonsense mediated mRNA decay (NMD) (12). Whilst examples of IR coupled with NMD have been found to downregulate gene expression, such as in granulocyte development (13), these transcripts can encounter other fates in the cell (14). Indeed, one of the few studies focussing on cytoplasmic IR in neurons showed an 'addressing' function for intronic RNA sequences, determining the spatial localization of their host transcripts within cellular compartments such as dendrites (15).   Another speculated function of IR has arisen following the identification of miRNA binding motifs within the retained intronic sequences. This offers an intriguing route through which miRNA-directed degradation pathways might regulate abundance of IRTs; alternatively, the retained introns themselves may serve as miRNA sinks, or even encode novel miRNAs termed mirtrons (16, 17).

3

Altogether, despite advances in our understanding of IR in neuronal cells, much remains unanswered.

The importance of investigating the roles of IR has been further corroborated by studies that demonstrate its relevance across a diverse range of neurodegenerative diseases (18–21). One such example is amyotrophic lateral sclerosis (ALS) (20), a rapidly progressive and incurable disease, which leads to selective degeneration of motor neurons (MNs). ALS is characterised by protein inclusions and axonal degeneration, and is often associated with RNA processing defects. ALS-causing mutations occur in numerous genes encoding crucial regulators of RNA-processing, which are normally expressed throughout development. Despite the growing number of causative gene mutations being identified in ALS, the precise aetiology remains unknown and early molecular pathogenic events remain poorly understood. We previously made the novel discovery that aberrant IR is a widespread phenomenon in ALS (20), which was corroborated by subsequent studies (21, 22). Moreover, we went on to demonstrate aberrant cytoplasmic IR as a widespread molecular phenomenon in VCP-related ALS (23). We showed that ALS-related aberrant cytoplasmic IRTs have conspicuously high affinity for RNA binding proteins (RBPs), including those that are mislocalized in ALS and proposed that a subset of cytoplasmic intronic sequences serve as 'blueprints' for the hallmark protein mislocalization events in ALS (24, 25). This raises an exciting possibility that intronic RNA sequences play additional significant roles beyond their recognized nuclear function. Nevertheless, the role and physiological relevance of cytoplasmic IR during neuronal development and disease still remains largely unresolved.

Against this background we sought to characterise the spatiotemporal dynamics of IRTs by re-analysing RNA-seq data from nuclear and cytoplasmic fractions of patient-specific hiPSCs undergoing motor neurogenesis. We first show

that retained introns exhibit compartment-specific features including their dynamics, biological pathways, and molecular characteristics during this process. We reveal a specific class of retained introns in the cytoplasm that is not associated with gene expression changes but exhibits high miRNA binding potential, which is functionally validated by identifying an altered expression profile of the predicted miRNA target genes. We finally analyze this class of retained introns in stem cell models of familial ALS and find evidence for a functional depletion of specific miRNAs, possibly as a result of cytoplasmic intronic sequences-mediated sequestration, which has potential implications for ALS pathogenesis and the development of therapies in this devastating and incurable disease.

MATERIALS AND METHODS

Compliance with ethical standards

Informed consent was obtained from all patients and healthy controls in this study. Experimental protocols were all carried out according to approved regulations and guidelines by UCLH's National Hospital for Neurology and Neurosurgery and UCL's Institute of Neurology joint research ethics committee (09/0272).

RNA-sequencing data

We obtained paired-end polyA stranded RNAseq libraries prepared from fractionated nucleus and cytoplasm obtained from 6 distinct stages of motor neuron differentiation from control and VCP^{mu} samples (iPSC, and days 3, 7, 14, 21 and 35; **Supplementary Table S1**) from previously published study (**GSE152983**) (23). We also obtained paired-end RNA sequencing reads derived from one independent study on familial form of ALS caused by mutant SOD1 (n=5; 2 patient-derived SOD1A4V and 3 isogenic control MN samples where the mutation has been corrected; Hb9 FACS purified MNs, **GSE54409** (26).

5

Transcript and Gene expression analysis

Kallisto (27) was used to (1) build a transcript index from the Gencode hg38 release Homo sapiens transcriptome (-k 31), (2) pseudo-align the RNA-seq reads to the transcriptome and (3) quantify transcript abundances (-b 100 -s 50—rf-stranded). Subsequent analysis was performed with the R statistical package version 3.3.1 (2016) and Bioconductor libraries version 3.3 (R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2013). Kallisto outputs transcript abundance, and thus we calculated the abundance of genes by summing up the estimated raw count of the constituent isoforms to obtain a single value per gene. For a given sample, the histogram of log2 gene count is generally bimodal, with the modes corresponding to non-expressed and expressed genes. Reliably expressed genes/transcripts for each condition (VCP^{mu} or control at days 0, 3, 7, 14, 22 and 35 in each fraction) were next identified by fitting a two-component Gaussian mixture to the log2 estimated count gene/transcript data with R package mclust (28) ; a pseudocount of 1 was added before log2 transformation. A gene/transcript was considered to be reliably expressed in a given condition if the probability of it belonging to the non-expressed class was under 1% in each sample belonging to the condition. 18,834 genes and 102,047 transcripts were selected based on their detected expression in at least one of the 24 conditions (i.e. 6 different timepoints of lineage restriction for control and VCP^{mu} in nuclear and cytoplasm). Next we quantile normalized the columns of the count matrices with R package limma (29). For differential gene expression analysis we ran Sleuth (30).

Splicing analysis

The identification of all classes of alternative splicing (AS) events in motor neuron differentiation was performed with the Vertebrate Alternative Splicing and Transcription Tools (VAST-TOOLS) toolset, which works synergistically with the VastDB web server, a collection of species-specific alternative splicing library files (31). Paired-end stranded RNA-seq reads were first aligned with VAST-TOOLS against the Homo sapiens hg38, Hs2 assembly from VastDB with the scaffold annotation Ensembl v88. This contains 74030 exon skipping events, 153119 intron retention events, 474 microexon events, 20812 alternative 3' UTR events, and 15804 alternative 5' UTR events. We then merged files from identical samples but different lanes together and then performed differential splicing analysis over time either for the control or for the VCP mutant samples separately using the `vast-tools diff` command which takes into account the different biological replicates. We then imported the result tables into R. For an AS event to be considered differentially regulated between two conditions, we required a minimum average ΔPSI (between the paired replicates) of at least 15% and that the transcript targeted by the splicing event in question to be reliably expressed in all samples from the conditions compared i.e enough read coverage in all samples of interest.

Gene ontology enrichment analysis

GO enrichment analysis was performed using classic Fisher test with topGO Bioconductor package (32). Only GO terms containing at least 10 annotated genes were considered. A p-value of 0.05 was used as the level of significance. On the figures, top significant GO terms were manually selected by removing redundant GO terms and terms which contain fewer than 5 significant genes.

Mapping and analysis of CLIP data

To identify RBPs that bind to retained introns, we examined iCLIP data for 21 RBPs (33), and eCLIP data from K562 and HepG2 cells for 112 RBPs available from ENCODE (34, 35). Before mapping the reads, adapter sequences were removed using Cutadapt v1.9.dev1 and reads shorter than 18 nucleotides were dropped from the analysis. Reads were mapped with STAR v2.4.0i (36) to UCSC hg19/GRCh37 genome assembly. The results were lifted to hg38 using liftOver (37). To quantify binding to individual loci, only uniquely mapping reads were used.

Analysis of cis-acting features

MaxEntScan (38) was used to calculate maximum entropy scores for 9-bp 5' splice sites and 23-bp 3' splice sites. Intron lengths and GC content were calculated using the hg38 human genome assembly. The intronic enrichment for RBP binding site was obtained by computing the proportion of crosslink events mapping to retained intron compared to non-retained introns of the same genes, accounting for intron length. These were defined in relation to the acceptor and donor splice sites, namely the last 30 nucleotides (nts) of exonic sequence upstream of the 5' splice site (R1), the first 30nts of intronic sequence downstream of the 5' splice site (R2), the 30nts in the middle of the intron (R3), the last 30nts of intronic sequence upstream of the 3' splice site (R4), and the first 30ntsof exonic sequence downstream of the 3' splice site (R5). These regions were defined based on the past studies of the Nova RNA splicing map (39), which has been determined by the positioning of conserved YCAY clusters as well as by the binding sites identified by HITS-CLIP as reported in (40). The nucleotide-level evolutionary phastCons scores for multiple alignments of 99 vertebrate genomes to the human genome were obtained from UCSC (41, 42) and a median score was derived for each individual

intron and defined regions of interest. The RBP crosslink event enrichment scores in each region of interest or in each group of intron was obtained by dividing the fractions of introns in a given group over the fraction in the full list of introns that exhibit at least one crosslink event for a given RBP in the defined region or across the full intronic region.

Spatiotemporal taxonomisation of the retained introns

We performed singular value decomposition (SVD) on the PIR cytoplasmic versus nuclear values of 94,457 introns in n = 48 cytoplasmic samples and n=47 nuclear samples across the 5 distinct stages of motor neuron differentiation from healthy controls and VCP mutants. We analysed 94,457 introns out of the 153,119 annotated introns in VAST-TOOLS given their overlap with reliably expressed genes. We then selected the components maximally capturing variance in PIR. To visualize the right singular vectors $\{\vec{v}_k\}$, we plotted the PIR on the vertical axis as a function of the time corresponding to each sample on the horizontal axis and coloring all samples corresponding to healthy controls with filled circles, and those corresponding to VCP-mutants in empty circles. Next we identified introns whose PIR profiles correlated (Pearson correlation between individual intron PIR profile and right singular vectors) and contributed (projection of each individual intron PIR profile onto right singular vectors) most strongly (either positively or negatively) with the profile of the singular vectors. In order to identify representative introns for each singular vector, events were ranked according to both projection and correlation scores. The highest (most positive scores in both projection and correlation) and lowest (most negative scores in both correlation and projection) motifs were selected for each singular vector using K-mean clustering.

MiRNA *expression analysis*

Total RNA including small RNAs was extracted from "patterned" precursor motor neurons of five control and four mutant cell lines using mirVana™ miRNA Isolation Kit (ambion, life technologies). RNA quantification and its 260/280 ratio were assessed using the nanodrop. Poly(A) tailing and reverse transcription of mature miRNAs was performed using miRCURY LNA RT kit (QIAGEN), with 20 ng of total RNA as input. Reverse transcribed cDNA was quantified using miRCURY LNA SYBR green dye, specifically designed primers (appropriate miRCURY LNA miRNA PCR assays) and QuantStudio 6 Flex Real-Time PCR System (Applied Biosystems). Relative miRNA expression levels between control and mutant cells were quantified using ΔΔCT Method, with U6 snRNA as a reference gene for normalisation. Data was plotted using RStudio software.

RESULTS

Nuclear and cytoplasmic IR affect two functionally divergent mRNA subsets

We previously reported a transient IR programme early during human motor neurogenesis using whole-cell RNA-sequencing data (20). To further examine the spatiotemporal dynamics of IR during this process in healthy cells, we re-analysed high-throughput poly(A) RNA-seq data derived from nuclear and cytoplasmic fractions of human induced pluripotent stem cells (hiPSCs; day 0), neural precursors (NPCs; day 3 and day 7), 'patterned' ventral spinal motor neuron precursors (pMNs; day 14), post-mitotic but electrophysiologically immature motor neurons (MNs; day 22), and electrophysiologically active MNs (mMNs; day 35) (**Fig. 1A** and **Supplementary Table 1;** 47 nuclear and cytoplasmic samples from 6 time-points; 4 clones from 4 different healthy controls) (23, 43). Using the RNA-seq pipeline VAST-TOOLS (31), we identified 4,189 nuclear and 1,542 cytoplasmic

significant alternative splicing (AS) events over time (**Fig. 1B** and **Supplementary Fig. 1A**). In line with our previous study (20), IR was the predominant mode of splicing during neurodevelopment, accounting for 64% and 49% of the included AS events in the nucleus (638 events) and the cytoplasm (541 events) respectively, indicating that cytoplasmic IRTs are more abundant than previously recognized (**Fig. 1B**). Further examining the distributions of percent intron retention (PIR) during MN differentiation in the nucleus and the cytoplasm for 211,501 events revealed that IR exhibits distinct dynamics in the two compartments (**Fig. 1C**). In particular, the nuclear compartment exhibits the highest level of PIR at the hiPSC stage, while the cytoplasmic compartment exhibits the highest level of PIR at DIV=14, which is reminiscent of the early wave of IR we previously reported (20). Notably the cytoplasmic increase in PIR early during differentiation is likely explained by a change in the subcellular localisation of (some) IRTs rather than a modulation of the splicing given the coincident stable level of PIR in the nucleus. Genes related to RNA processing and splicing are among the most affected by IR (13, 44–48) and we previously showed that IR early during MN development specifically affects RNA processing related biological pathways (20). Here we find that cytoplasmic (but not nuclear) IR affects essential genes concerned with mRNA metabolism. In contrast we find that genes targeted by alternative exons (AltEx) are enriched in similar biological pathways in the nucleus and the cytoplasm as shown by Gene Ontology (GO) function analysis (**Fig. 1D**). These findings indicate that the previously reported wave of IR during MN differentiation, using whole-cell RNA-sequencing, likely reflected signals from cytoplasmic IRTs.

Prior studies reported specific features associated with retained introns including higher GC content, lower intron length and enrichment in RBP binding motifs compared to non-retained introns (1, 47, 49, 50). In line with these studies, we find that the PIR negatively correlates with the intron length and positively

correlates with the GC content and the enrichment in crosslink events for 131 RBPs for which CLIP data were available (33–35) both in nuclear and cytoplasmic compartments (**Figs. 1E, G, I** and **Supplementary Fig. 1F**). Additionally we find that i) retained introns are detected in genes containing fewer introns, and ii) the PIR positively correlates with the intronic sequence conservation score (**Supplementary Figs. 1B, D, F**). Surprisingly, however, by specifically comparing the cytoplasmic retained introns ($PIR_{NUCLEUS}$ > 20% and $PIR_{CYTOPLASM}$ > 15%; **Supplementary Table S2**) with the nuclearly detained retained introns ($PIR_{NUCLEUS}$ > 20% and $PIR_{CYTOPLASM}$ < 5%; **Supplementary Table S3**) we find that the retained introns that localise to the cytoplasm are on average longer, have lower GC content compared to their nuclear counterparts, have higher RBPs enrichment scores and are more evolutionarily conserved (**Figs. 1F, H, J** and **Supplementary Figs. 1C,E**). Altogether these results show that nuclear and cytoplasmic retained introns exhibit distinct features including their dynamics during human motor neurogenesis, their associated biological pathways, and their molecular characteristics.

A spatiotemporal taxonomy reveals cytoplasmic retained introns with distinct RBP binding profiles

Having established that nuclear and cytoplasmic IR affect two functionally divergent mRNA subsets, we next used singular value decomposition (SVD) analysis to categorize 94,457 analysed introns into nine groups based on their PIR spatiotemporal dynamics during MN differentiation (**Fig. 2A** and **Supplementary Tables S4-S12**). Of these, three categories are nuclearly-detained retained introns (termed N1-N3 hereafter) and the other six are cytoplasmic retained introns (termed C1-C6 hereafter), that exhibit the following compartment-specific PIR dynamics: stable nuclear (>15%) and low cytoplasmic (<10%) PIR over time (N1),

steady reduction in the nuclear PIR over time and stable low cytoplasmic PIR (N2), a transient increase in nuclear PIR and stable low cytoplasmic PIR (N3), steady reduction in both nuclear and cytoplasmic PIR over time (C1), steady increase in nuclear and cytoplasmic PIR over time (C2), increase in cytoplasmic PIR only in terminal differentiation (C3), consistently high nuclear and cytoplasmic PIR over time (C4), early transient increase in nuclear and cytoplasmic PIR (C5) and a late transient increase in nuclear and cytoplasmic PIR (C6).

Next looking at the percentage of introns per genes targeted by IR in each identified group revealed that IR during MN differentiation does not occur stochastically, but appears to target a specific set of introns in each gene (**Supplementary Fig. 2A**), indicating that additional layer(s) of specific regulation must underlie the regulation of these 9 distinct IR programmes as previously suggested (47). Previous studies suggest that cis-regulatory elements bound by trans-acting factors such as RBPs are likely to play a crucial role in regulating IR (50–52). Thus we next sought to test whether the 9 spatiotemporally distinct classes of IR we identified are associated with different combinations of trans-acting factors that could regulate them. We achieved this by using the publicly available CLIP data to evaluate the crosslink events for 131 RBPs mapping to 5 regions we defined in relation to the acceptor and donor splice sites, namely the last 30 nucleotides (nts) of exonic sequence upstream of the 5' splice site (R1), the first 30nts of intronic sequence downstream of the 5' splice site (R2), the 30nts in the middle of the intron (R3), the last 30nts of intronic sequence upstream of the 3' splice site (R4), and the first 30nts of exonic sequence downstream of the 3' splice site (R5) (**Fig. 2B**, *upper* and **Supplementary Tables S12–S17**). First, looking at the fractions of regions which are mapped by at least one crosslink event for each RBP, we find that the R1 and R5 exonic regions, sequences of which are the most evolutionarily conserved (**Supplementary Fig. 2B**), exhibit the highest frequency in

13

crosslink events across the 131 RBPs irrespective of the IR grouping, with the exception of the C1 group (**Fig. 2B**, *lower*). These results, which are in line with previous studies showing that the splicing machinery is more likely to form across the exonic regions than across the introns for similarly long introns (>250 nts), indicate that the 9 groups of introns bear similar chances of splicing complex formation with respect to their R1 and R5 exonic regions (53, 54). This is further supported by the finding that identically optimal splicing signals are detected among the 9 groups of introns, with the exception of the C4 group (**Supplementary Fig. 2C**) as failure in splice site recognition (50–52) or decreased expression levels of splicing factors (47) have also been proposed to underlie IR.

Although RBPs exhibit similarly high frequency of binding to the R1 and R5 exonic regions across the majority of the 9 groups of introns, we indeed noticed that the R2, R3 and R4 intronic regions display large variability in the percentages of crosslink events across the different spatiotemporal IR dynamics (**Fig. 2B**, *lower*). Next, looking at the enrichment of RBPs mapping to each of these regions further revealed that the C3, C4 and C5 groups of cytoplasmic retained introns is indeed specifically enriched in RBPs binding to the R2, R3 and R4 intronic regions as opposed to the R1 and R5 regions which display as much RBP binding as the full set of introns (**Fig. 2C**). One of the most enriched RBPs in the R2, R3 and R4 intronic regions is UPF1, an RNA helicase required for nonsense-mediated mRNA decay (NMD) in eukaryotes. UPF1 exhibits high binding occurrence across the five regions of interest for the {C4, C5, C6} groups as opposed to the other groups for which UPF1 is strictly enriched in R1 and R5 regions (**Supplementary Fig. 2D**). Different combinations of RBPs have been shown to coordinately regulate functionally coherent "networks" of exons and introns (55). Thus, using unsupervised hierarchical clustering of the 9 groups of introns based on their 131 RBP enrichment scores profiles we finally showed that {C3, C4, C5} form a coherent regulated group

of introns in respect to all selected regions except the R1 exonic region (**Fig. 2D**). Altogether, this analysis identifies a coherent regulated supergroup of retained introns which exhibits specific elements in their intronic regions that are not necessarily related to splicing efficiency, but rather may perform an additional role in the regulation of mRNA metabolism.

Identification of a cytoplasmic group of IRTs with a high capacity for miRNA sequestration

The finding that a coherent regulated supergroup of retained introns bears similar chances of splicing complex formation with respect to their R1 and R5 exonic regions but with a potential regulatory role in mRNA metabolism through RBP intronic sequence binding, prompted us to look at the occurrence of RBP crosslink events across the entire intronic region as opposed to the 5 predefined regions of interest. We showed that while the {C4, C5, C6} groups have a somewhat higher prevalence of crosslink events (**Fig. 3A**), the C5 group specifically has the highest fraction of introns with at least one crosslink event across the 131 studied RBPs compared to the full set of introns (**Fig. 3B**). Notably, the {C4, C5, C6} groups overall have a lower intron length compared to {C1, C2, C3}, and thus this result cannot be due to a bias in size. Of the 51 RBPs that exhibit >19% higher fraction of binding to the C5 group compared to the full set (Fisher enrichment P-value < 0.01; **Fig. 3C**) at least 9 are key regulators of mRNA transport such as UPF1 and IGF2BP1 (56–58), and 6 RBPs are involved in miRNA regulatory pathways such as DROSHA and PUM2 (59–61), UPF1 being involved in the two (62–64) (**Figs. 3C-E**). Noting that miRNA regulators are avidly binding the C5's intronic regions, we next looked at miRNA motif enrichment within the introns using HOMER (65), which revealed significant enrichment for 14 miRNA motifs (**Supplementary Fig. 3A**). C5 IR does not play a role in gene expression regulation, as revealed by the analysis of fold-changes over time

of the genes containing the retained introns (**Supplementary Fig. 3B**), and may thus serve other functions, particularly in the cytoplasm. Focusing on miR-4519, miR-1976, miR-4716-5p, miR-485-5p and miR-4267, the top five miRNAs motifs enriched in the C5 intronic sequences (**Fig. 3F**), we next sought to test whether changes in the C5 PIR over time might relate to changes in gene expression of specific miRNA predicted target genes, as a result of trapping/releasing these miRNAs. To this end, we examined the gene expression profile of their predicted target genes by combining two miRNA target prediction algorithms, TargetScan (66) and miRanda (67). Strikingly we find that the predicted target genes of miR-4519, miR-1976, and miR-4716-5p exhibit a reduction in expression from DIV=7 to DIV=14, which coincides with a reduction of the C5 introns PIR, while the predicted target genes of miR-485-5p and miR-4267 do not exhibit such trend (**Fig. 3G** and **Supplementary Figs. 5,6**). This result indicates that the decrease in IR from DIV=7 to DIV=14 correlates with an increase in miRNA activity, supporting the hypothesis that cytoplasmic retained introns reduce miRNA activity potentially by sequestering them, as previously shown for long non-coding RNA (lncRNA) (68).

VCP mutation-related transient accumulation of cytoplasmic IRTs correlates with reduced miRNA activity.

We previously demonstrated aberrant cytoplasmic IRTs in ALS-related VCP^{mu} samples during MN differentiation that exhibit high predicted binding affinity for RBPs (20, 23). We next sought to test whether VCP mutations affect any of the cytoplasmic groups of introns in particular. Examining the two most prominent classes of cytoplasmic IR dynamics during MN differentiation, as captured by right singular vectors of the SVD analysis performed on the cytoplasmic PIR values, confirmed prior findings that VCP mutations leads to exceptionally large IR perturbations at DIV=14 (**Supplementary Figs. 4A, B**) (23). Further comparing the

16

PIR distributions between control and VCP^{mu} samples in each of the six groups of cytoplasmic retained introns revealed that VCP mutations specifically impact two classes of events, namely C5, and to a much lesser extent C1 while the other groups remain unchanged (**Fig. 4A** and **Supplementary Fig. 4C**). Most notably VCP-driven changes in cytoplasmic IR are 1) unidirectional i.e. we only detect increases in IR in VCP^{mu} samples compared to control samples irrespective of PIR dynamics, and 2) the VCP mutation specifically affects groups of introns in which the PIR exhibits a large decrease from DIV=7 to DIV=14. This is in contrast to those groups of introns where the PIR *increases* from DIV=7 to DIV=14, such as C2 and C6, where we find similar increase in control and VCP^{mu} samples (**Supplementary Figs. 4C**). These results suggest that VCP mutations enhance the cytoplasmic stability of IRTs, rather than affecting nuclear export, which would equally impact C1, C2, C5 and C6.

Having found that VCP mutations lead to large perturbations of the C5 group of cytoplasmic IRTs at DIV=14, which we have shown above to associate with decreased activity of specific miRNAs, we next sought to test whether the increase in cytoplasmic IR of the C5 group in VCP mutants correlates with a reduction in miRNA activity by looking at the changes in gene expression of their predicted target genes between VCP and control samples. This analysis revealed that the increase in IR in VCP mutant cultures correlates with a decrease in miR-4519, miR-1976 and miR-4716-5p activities as predicted by the up-regulation of their respective target genes at DIV=14 (**Fig. 4B**). Additionally, these changes are not explained by a change in the expression levels of these miRNAs which are not significantly different between VCP and control cultures at this time point (**Supplementary Figure 4D**). Notably, the predicted activities of miR-485-5p and miR-4267, whose target gene expression profiles did not correlate with IR level in control samples over time, also do not correlate with increase in C5 IR in VCP

17

samples, thus supporting the hypothesis that the activities of the same miRNAs correlate to IR in both VCP and control samples over time (**Supplementary Fig. 6**).

Noting that we have studied a relatively rare form of familial ALS (fALS) caused by gene mutations in VCP (selected as it exhibits the pathological hallmark of TDP-43 nuclear-to-cytoplasmic mislocalisation), we next sought to understand the generalizability of the association between increase in IR in ALS samples and decrease in miRNA activity. To this end, we chose to study one of the most common forms of fALS (SOD1), which in contrast does not exhibit the pathological hallmark of TDP-43 nuclear-to-cytoplasmic mislocalisation. We first looked at the PIR of the C5 group in SOD1 mutant hiPSC-derived MNs, revealing a statistically significant increase in IR in SOD1 (**Fig. 4C**). Next, looking at the changes in gene expression of the miRNA predicted target genes between SOD1 mutant MNs and their isogenic controls, further showed a decrease in miR-4519, miR-1976 and miR-4716-5p predicted activities (**Fig. 4D**). These findings further substantiate the relevance of the correlation between increased IR and decreased miRNA activity. Altogether these findings support the hypothesis that the cytoplasmic pool of C5 introns leads to a reduction in miRNA activity, potentially through direct binding and sequestration, which may have important roles in ALS pathogenesis, and indeed implications for new therapeutic strategies.

DISCUSSION

Neuronal biology relies on complex regulation of gene expression and mRNA metabolism. Alternative splicing has been shown to play a key role in this process and IR is now recognized as the dominant mode of splicing during MN development (1, 20), including cytoplasmic IR, which we recently showed to affect >100 transcripts during neuronal development (23). Because nuclear IR has been the focus of most previous studies, the regulation and role of cytoplasmic IRTs remain

unclear. The objective of this study was twofold: to deepen our understanding of the role(s) of cytoplasmic IR in normal cellular physiology by resolving the spatiotemporal dynamics of IR underlying distinct stages of MN lineage restriction, and to decipher whether specific classes of IRTs become dysregulated in the context of disease by systematically examining the influence of ALS-causing VCP mutations on this process. In order to achieve this we re-analyzed nuclear and cytoplasmic RNA-sequencing data from a time course of patient-specific iPSCs differentiating into spinal MNs.

We first show that nuclear and cytoplasmic IR target distinct classes of mRNA associated with particular dynamics, biological pathways and molecular characteristics. Specifically, we find that the sequences of the retained introns that localise to the cytoplasm are evolutionarily more conserved and exhibit a higher capacity for RBP binding compared to the nuclearly detained introns. This argues against the hypothesis that cytoplasmic intron-containing pre-mRNAs simply 'leak' from the nucleus (69), which is also further excluded by polyA selection during library preparation, and suggests that 1) cytoplasmic localisation signals for these IRTs are contained in the intronic sequences, and 2) cytoplasmic IRTs likely serve a biological function that has yet to be discovered.

We next show that MN differentiation exhibits complex IR spatiotemporal dynamics captured by 9 distinct IR programmes, 3 which are nuclearly detained and 6 that localise to the cytoplasm. Given the time and cell compartment specificity of these programmes, they are expected to associate with distinct complex regulation. IR has been previously proposed to be the consequence of globally inefficient splicing (47, 70), that could be linked to several mechanisms including the occupancy of MeCP2 near the splice junction (71), the expression of PRMT5 (7), and relatively weak splice sites (1). Here we find that the 9 groups of introns exhibit similar 5' and 3' maximum entropy scores as well as similarly high RBP binding in their exonic regions juxtaposed to the splice sites where the splicing

19

machinery is more likely to form (53, 54) as opposed to the intronic regions. These findings indicate that an overall change in splicing efficiency during MN differentiation is unlikely to be the dominant regulatory factor for most of these IR programmes. Furthermore, IR during MN differentiation does not occur stochastically, but appears to target a specific set of introns in each gene, and thus an additional layer of specific regulation must underlie the regulation of these 9 distinct IR programmes as previously suggested (47). Indeed similar combinations of trans-acting factors are detected across 4 regions juxtaposed to the splices sites among 3 groups of cytoplasmic retained introns -{C4, C5, C6}- suggesting similar regulation. Additionally these three groups of introns exhibit avid RBP binding within their intronic regions juxtaposed to the splice sites when compared to the full set of analysed introns, indicating a potential regulatory mechanism in mRNA metabolism through intronic sequence binding for the {C4, C5, C6} groups of introns. Notably the full list of retained introns for each group together with the regional RBP enrichment is freely accessible as supplementary tables providing a rich resource for researchers across the disciplines of genomics and basic neuroscience.

Although the stable cytoplasmic localisation of intronic sequences in neurons has been recognized since 2013 (11), their role has remained poorly understood. One of the few studies focussing on cytoplasmic IR showed an 'addressing' function for intronic RNA sequences in determining their spatial localization within cellular compartments (15). Here we show that the avid RBP binding we previously observed in ALS-related aberrant cytoplasmic retained introns (23) is indeed specifically detected in one IR programme which exhibits a transient increase in the cytoplasm during MN differentiation, namely the C5 group. Notably this group of cytoplasmic intron, which is the most impacted by VCP mutations, exhibits the same PIR dynamics of the group of introns we previously showed to be impacted by VCP mutations using whole-cell RNA-sequencing (20). This suggests that VCP mutations

20

specifically affect the cytoplasmic stability of IRTs rather than leading to a reduction in splicing efficiency. The absence of correlation between the C5 PIR level and the gene expression dynamics during MN differentiation raises the possibility of new roles for intronic RNA sequences beyond a function in gene expression regulation, particularly in the cytoplasm. As previously proposed, cytoplasmic retained introns may act as RNA regulators in the homeostatic control of RBP localisation during development and disease (23), which may in turn lead to loss of function. For example some splicing factors that avidly bind the C5 intronic sequences may remain sequestered upon their nuclear export and cytoplasmic localisation, contributing to a transient reduction in splicing efficiency during MN differentiation. Another intriguing molecular characteristic of the C5 group of introns is the enrichment for several miRNA motifs across the full length of the intron, predicted activity for which negatively correlates with the PIR level of this intron group during MN differentiation. Thus the presence of long intronic sequences in the cytoplasm of neuronal cells may serve as a regulatory mechanism for miRNA functionality through their sequestration and downstream up-regulation of their target genes. Indeed previous studies speculated that stable intron-derived RNA sequences (sisRNA) (72) act as molecular sinks to sequester miRNA (72) and/or RBPs (73) leading to reduction in their activities and future studies will test whether sisRNA are derived from the cytoplasmic intronic sequences of the C5 group.

We previously demonstrated aberrant cytoplasmic IR in ALS-related VCP^{mu} samples during MN differentiation (23). Here we show that VCP mutations lead to an aberrant PIR increase specifically of the C5 group of introns. Furthermore we show that the aberrant increase in C5 PIR level at DIV=14 in ALS mutant cells correlates with a decrease in the predicted activities of miR-1976, miR-4519 and miR-4716-5p, motifs of which are enriched in the C5 intronic sequences. These findings were further generalized to SOD1-related ALS hiPSC-derived mutant MNs,

21

supporting the hypothesis of a functional depletion of specific miRNAs as a result of cytoplasmic intronic sequences-mediated sequestration in ALS cells. Notably a reduction in miR-1976 activity, motifs of which are detected in 76% of the C5 intronic regions, is expected to occur in some sporadic ALS patients due to a mutation (rs17162257) in its enhancer (74). Furthermore, several miRNAs, and their target genes, are recognized to be involved in the occurrence and pathophysiology of neurodegenerative diseases including ALS (75–77). Thus, here we propose that a group of intronic sequences which accumulate in the cytoplasm of VCP mutant cells, as previously shown (23), act as molecular sponges for miRNA, thus resulting in elevated expression of their target genes. Future work will be required to demonstrate the direct role of these cytoplasmic intronic sequences in regulating miRNA activity through sequestration.

In conclusion we propose that cytoplasmic retained introns function as RNA regulators in the homeostatic control of RBP localisation and miRNA activity during MN development and disease, which has potential implications for ALS pathogenesis and the development of therapies for this devastating and incurable disease.
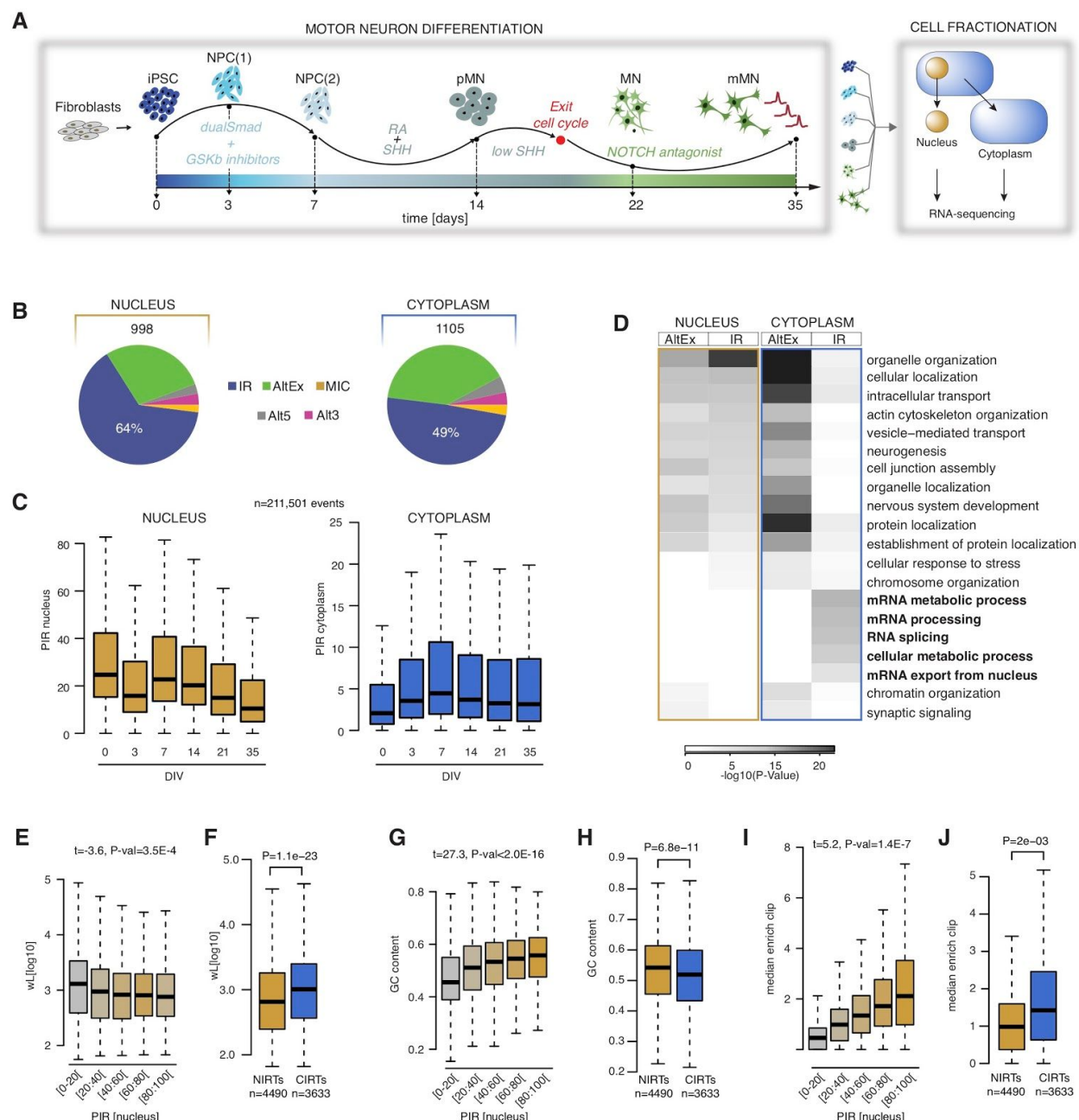
FIGURES AND TABLES

Figure 1 | Nuclear and cytoplasmic IR affect two distinct mRNA subsets.

A. Schematic depicting the iPSC differentiation strategy for motor neurogenesis. Arrows indicate sampling time-points in days when cells were fractionated into nuclear and cytoplasmic compartments prior to (polyA) RNA-sequencing. Four iPSC

23

lines were obtained from four different healthy controls. Induced-pluripotent stem cells (iPSC); neural precursors (NPC); "patterned" precursor motor neurons (ventral spinal cord; pMN); post-mitotic but electrophysiologically inactive motor neurons (MN); electrophysiologically active MNs (mMN). **B.** Pie charts representing proportions of included splicing events at defined stages during motor neurogenesis in nuclear (*left*) and cytoplasmic (*right*) fractions. Total number of events are indicated above the charts. Intron retention (IR); alternative exon (AltEx); microexons (MIC); alternative 5′ and 3′ UTR (Alt5 and Alt3). **C.** Comparison of the percent intron retention (PIR) during MN differentiation in nucleus (*left*) and cytoplasm (*right*) for 21,161 events that exhibit >10% PIR in at least 3 out of 47 nuclear samples. **D.** Heatmap of the GO biological functions enriched among the genes targeted by AltEx or IR in either the nucleus or the cytoplasm. P-values obtained by Fisher enrichment test. **E.** Analysis of the relationship between the PIR in the nucleus and the intron length. Retained introns are grouped in five categories of increasing level of retention in the nucleus as indicated on the *x*-axis. P-values obtained from analysis of variance comparing the full model of the logit of maximum IR across all nuclear samples according to the five characteristics with the reduced model removing the characteristic of interest. **F.** Comparison of intron length between nuclear and cytoplasmic retained introns. Nuclear retained introns are defined as intron exhibiting >20% IR in nuclear fraction and <5% IR in cytoplasmic fraction. Cytoplasmic retained introns are defined as intron exhibiting >20% IR in nuclear fraction and >15% IR in cytoplasmic fraction. P-values obtained from Mann-Withney test. **G.** Analysis of the relationship between the PIR in the nucleus and the GC content in %. Data shown as in (**E**). **H.** Comparison of GC content (%) between nuclear and cytoplasmic retained introns. **I.** Analysis of the relationship between the PIR in the nucleus and the median enrichment for RBP binding site compared to the non-retained introns of the same gene. Data shown as in (**E**). **J.** Comparison of median enrichment for RBP binding sites between nuclear

24

and cytoplasmic retained introns. For C, E-J-J: Data shown as box plots in which the centre line is the median, limits are the interquartile range and whiskers are the minimum and maximum.
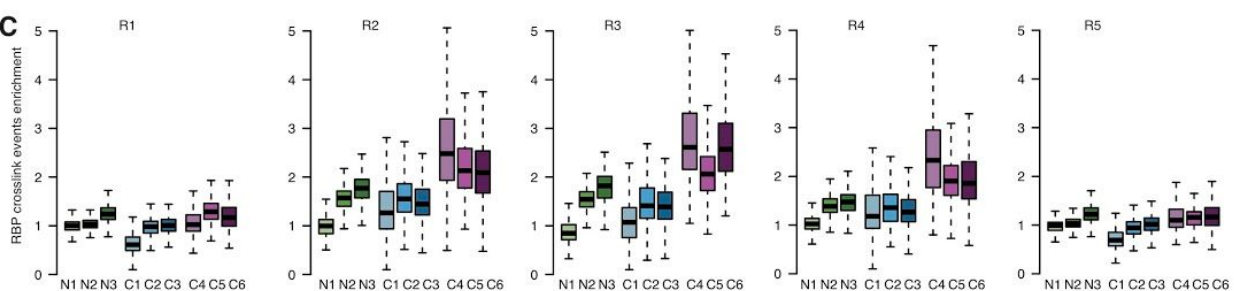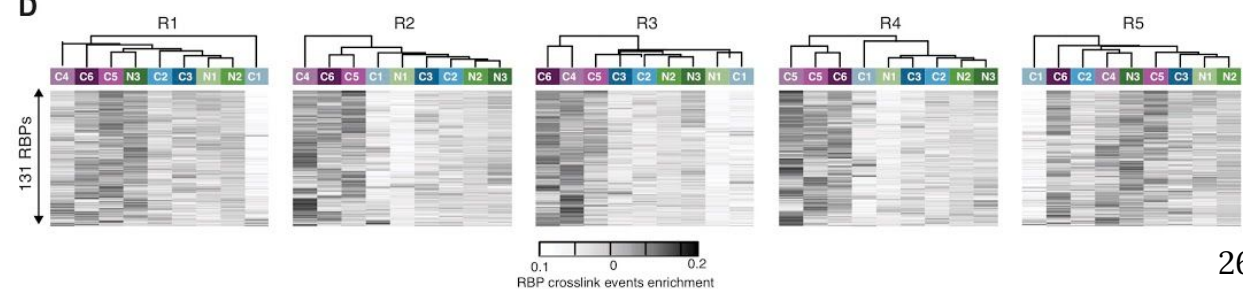
26
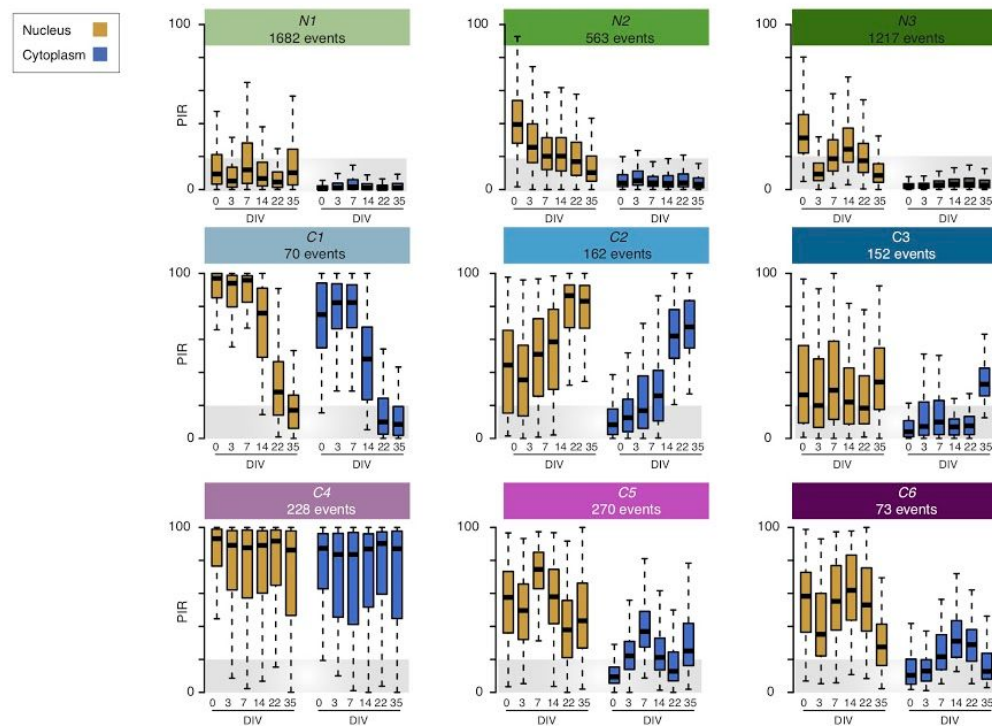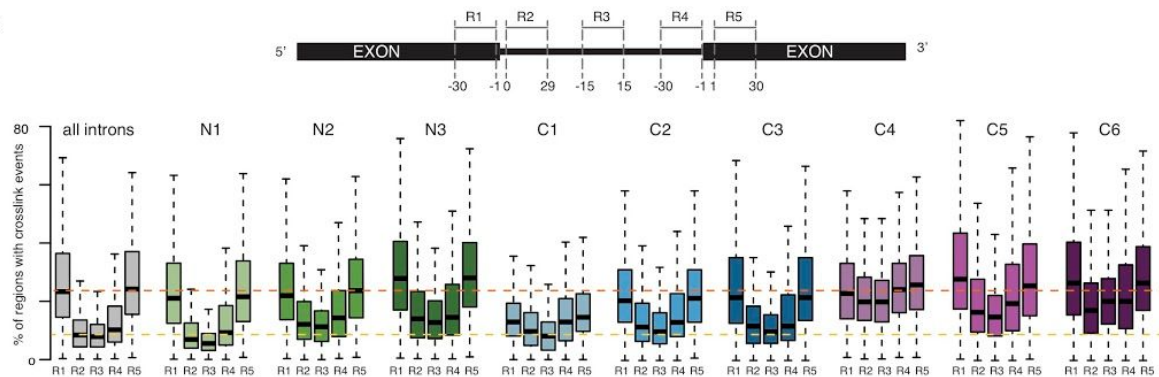
Figure 2 | A spatiotemporal taxonomy reveals cytoplasmic IRTs with distinct RBP binding profiles. A. Comparison of the nuclear and cytoplasmic percent intron retention (PIR) distributions for 9 groups of retained introns exhibiting distinct spatio-temporal dynamics during MN differentiation as identified using SVD (see Materials and Methods). N1, N2 and N3 contain introns primarily retained in the nuclear compartment while the remaining 6 groups contain introns with significant detection in the cytoplasm. Gold boxes = nucleus; blue boxes = cytoplasm. Grey area indicates the range of PIR values for which an intron is considered non-retained. **B.** (*Upper*) Schematic depicting the selected splicing regulatory regions juxtaposing the splice sites, namely the last 30 nucleotides (nts) of the upstream exon (R1), the first 30nts of 5' intron region (R2), 30nts in the middle of the intron (R3), the last 30nts of 3' intron region (R4), and the first 30nts of downstream exon (R5). (*Lower*) Distributions of the percentage of regions in each group of introns that are mapped by at least one crosslink event for each of the available 131 RBPs. **C.** Distribution of the enrichments in crosslink events in each of the selected regions R1, R2, R3, R4 and R5 for the available 131 RBPs across the 9 categories of introns. Enrichment is obtained by dividing the fraction of regions from the group of interest with at least one crosslink event with the fraction of regions from the complete set of introns (n=61872) with a crosslink event. **D.** Heatmaps of the enrichment scores of the crosslinking events for 131 RBPs in the R1, R2, R3, R4 and R5 regions for the 9 groups of introns hierarchically clustered using Manhattan distance and Ward clustering. Data shown as box plots in which the centre line is the median, limits are the interquartile range and whiskers are the minimum and maximum.
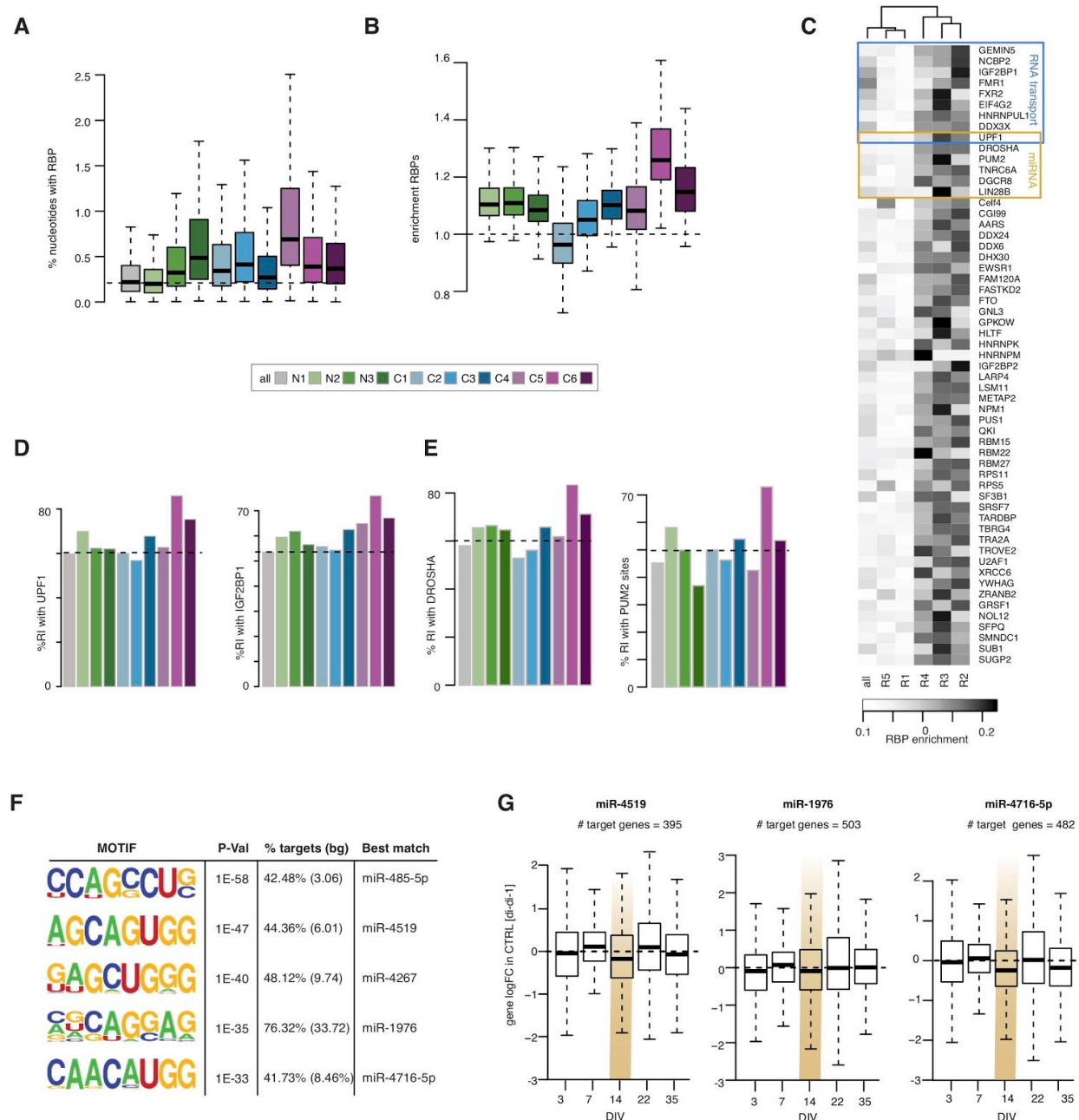
Figure 3 | Identification of a cytoplasmic group of retained introns with a high capacity for RBP and miRNA sequestration. **A.** Analysis of the percentage of nucleotides with cross-linking events for 131 RBPs across the entire retained intron for all 9 categories. **B.** Analysis of the enrichment in binding sites for 131 RBPs across the entire retained intron for all 9 categories. Enrichment is obtained by

dividing the fraction of retained introns in the category of interest with a CLIP binding with the fraction of retained introns in the complete set of introns (n=61872) with a crosslinking event. **C.** Heatmap of the enrichment score of crosslinking events in both the entire intron and each of the five regulatory regions of the 270 retained introns of the C5 category for 51 RBPs that exhibit a difference in the fraction of cross-linking events of more than 19% in the pool of C5 retained introns compared to the full set of introns. The blue box highlights RBPs involved in RNA transport and the gold box represents those involved in miRNA regulation. **D, E.** Percentage of retained introns with crosslink events for two RBPs involved in RNA transport (UPF1 and IGF2BP1), and two RBPs involved in miRNA regulatory pathway (DROSHA and PUM2). **F.** Five top motifs enriched in the 270 retained introns of the C5 category identified by HOMER (65). **G.** Distributions of the changes in nuclear expression over time of the control samples for the TargetScan (66) and miRDB (78) predicted target genes of miR-4519, miR-1976 and miR-4716-5p. Fold-changes over time obtained by comparing the log2 expression level at time of interest ($d_t = \{0,\ 3, 7, 14, 22, 35\}$) with the expression level at previous stage (d_{t-1}). Gold shaded area indicates the time-point where the largest changes in cytoplasmic IR are observed over time for the control samples. For A,B & G: Data shown as box plots in which the centre line is the median, limits are the interquartile range and whiskers are the minimum and maximum.
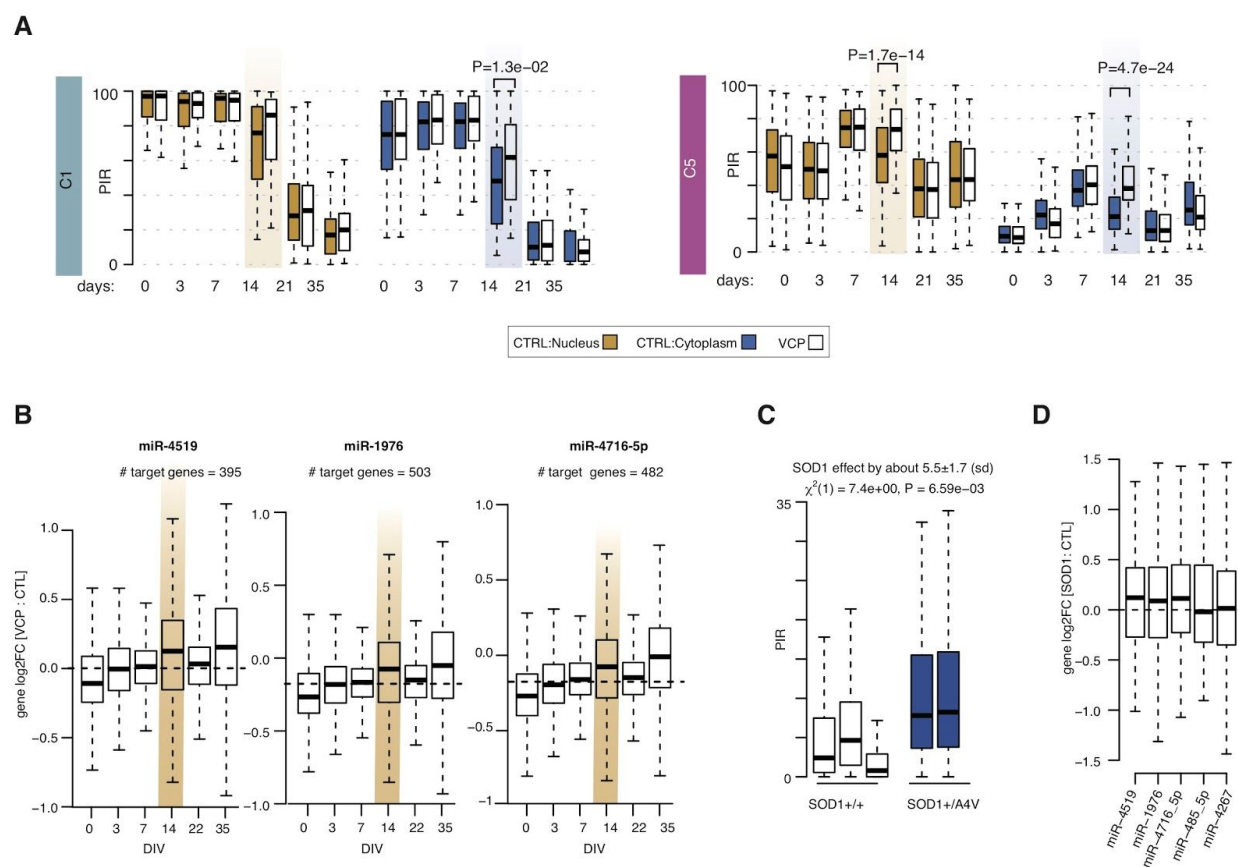
Figure 4 | ALS-related transient accumulation in cytoplasmic retained introns correlates with reduced miRNA activity. A. Comparison of the distributions of nuclear and cytoplasmic percent intron retention (PIR) between control (*colored* boxes) and VCP^*mu* (*white* boxes) samples during MN differentiation for the C1 and C5 groups of cytoplasmic retained introns. P-values obtained with two-sided Welch *t*-test. **B.** Distributions of the changes in nuclear expression between VCP^*mu* and control samples at each time-point (*right*) for the TargetScan (66) and miRDB (78) predicted target genes of miR-4519, miR-1976 and miR-4716-5p. Fold-changes obtained by comparing the log2 expression level at each time point ($d_t = \{0, 3, 7, 14, 22, 35\}$). Gold shaded area indicates the time-point where the largest changes in cytoplasmic IRTs are observed between the control and VCP mutant samples. **C.** Boxplots displaying the distribution of percentage retention for the 270

30

introns of the C5 category in control MNs (white box), and SOD1^{mu} MNs samples (*left*, blue bar) (79, 80). Mutant samples exhibit a systematically higher proportion of IR compared with controls. Linear mixed effects analysis of the relationship between the PIR for the 270 introns and SOD1 mutation to account for idiosyncratic variation due to patient differences: SOD1 mutation significantly increases IR by about 5.6% ± 3 (standard errors; $\chi^2(1)$ = 7.4, P = 6.5E-03). **D.** Distributions of the changes in expression between SOD1^{mu} and control samples for the predicted target genes of miR-4519, miR-1976, miR-4716-5p, miR-485-5p and miR-4267. For A-D: Data shown as box plots in which the centre line is the median, limits are the interquartile range and whiskers are the minimum and maximum.
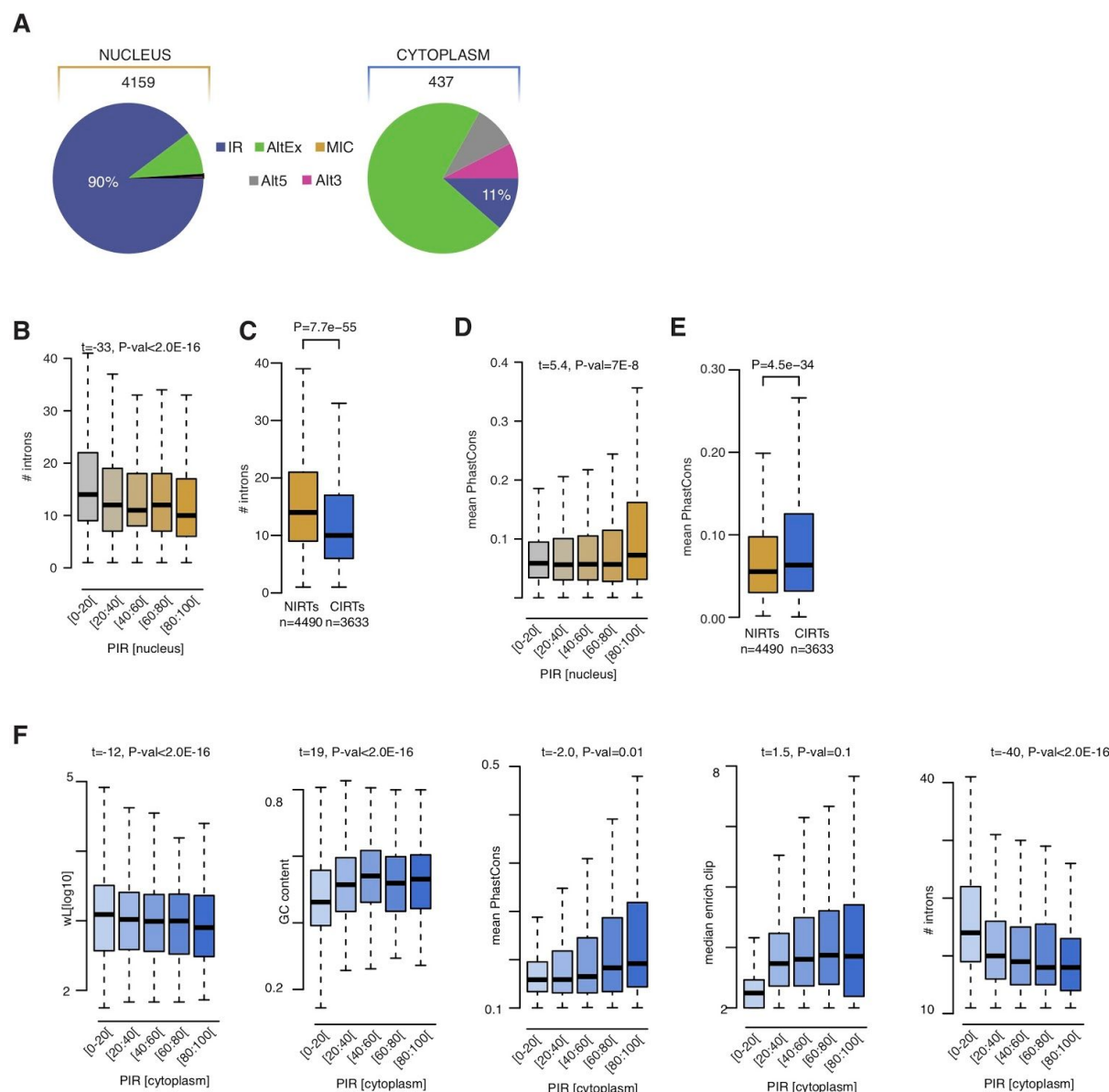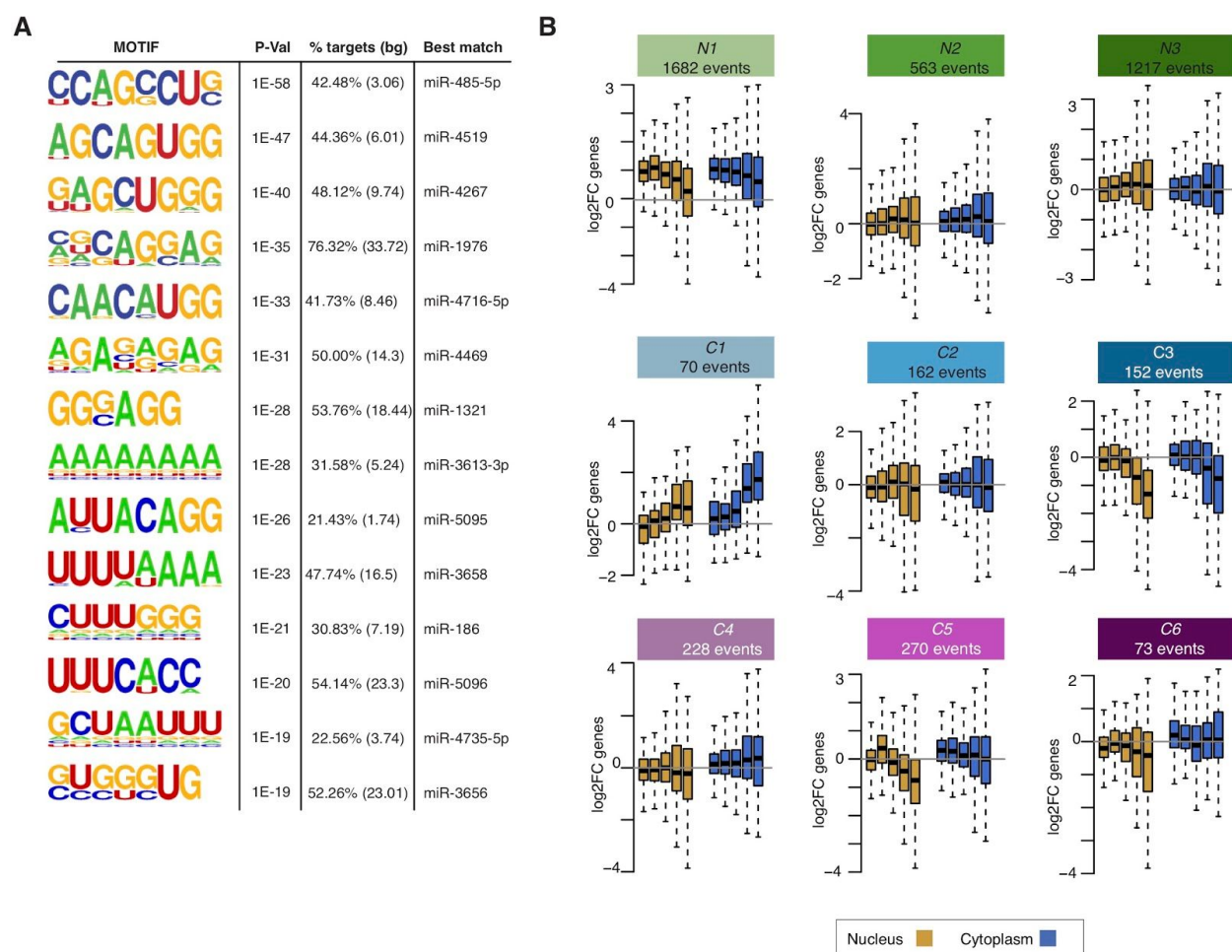
Supplementary Figure 1 | A. Pie charts representing proportions of skipped splicing events at defined stages during motor neurogenesis in nuclear (*left*) and cytoplasmic (*right*) fractions. Total number of events are indicated above the charts. Intron retention (IR); alternative exon (AltEx); microexons (MIC); alternative 5′ and 3′ UTR (Alt5 and Alt3). **B, D.** Analysis of the relationship between the percent intron retention (PIR) in the nucleus and the number of introns per gene (**A**), and

the retained intron average conservation scores (**D**). Retained introns are grouped in five categories of increasing level of retention in the nucleus as indicated on the *x*-axis. Data shown as in Fig. 1D. *P*-values obtained from analysis of variance comparing the full model of the logit of maximum IR across all nuclear samples according to the five characteristics with the reduced model removing the characteristic of interest. **C, E.** Comparison of number of introns per gene and the conservation scores between nuclear and cytoplasmic retained introns. Nuclear retained introns are defined as those exhibiting >20% IR in nuclear fraction and <5% IR in cytoplasmic fraction. Cytoplasmic retained introns are defined as those exhibiting >20% IR in nuclear fraction and >15% IR in cytoplasmic fraction. Data shown as in Fig. 1D. P-values obtained from Mann-Withney test. **F.** Analysis of the relationship between the percent intron retention (PIR) in the cytoplasm and the intron length, the GC content in %, the number of introns per gene, the retained intron average conservation scores and the median enrichment for RBP binding site compared to the non-retained introns of the same gene. Retained introns are grouped in five categories of increasing level of retention in the cytoplasm as indicated on the *x*-axis. Data shown as box plots in which the centre line is the median, limits are the interquartile range and whiskers are the minimum and maximum.
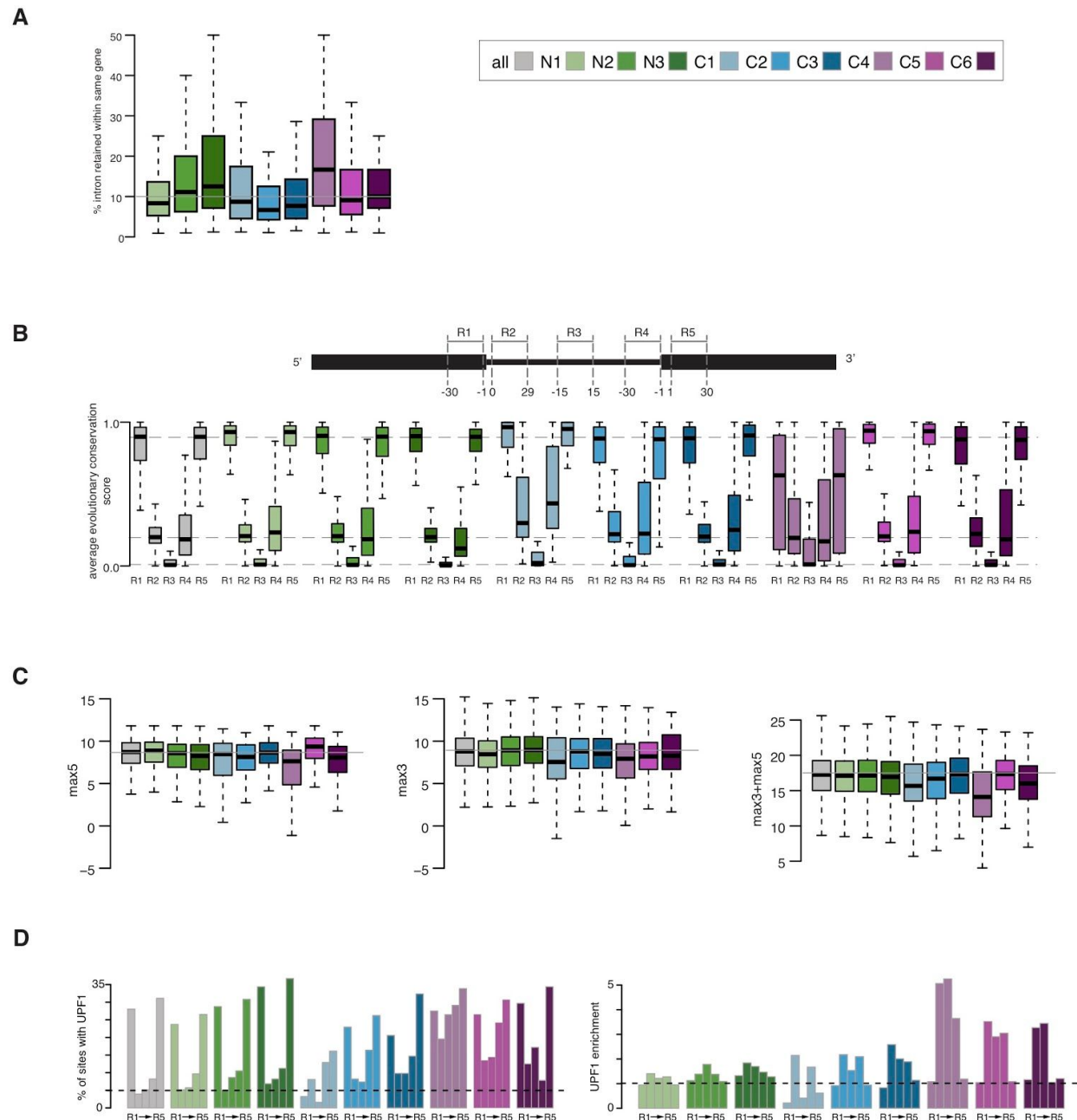
Supplementary Figure 2 | A. Percentage of retained introns per gene for the genes targeted by intron retention in each group. **B.** Distribution of the average evolutionary sequence conservation scores in the in the last 30nts of the upstream exon (R1), the first 30nts of 5' intron region (R2), the 30nts in the middle of the intron (R3), the last 30nts of 3' intron region (R4), and the first 30nts of downstream exon (R5) for 9 categories of introns for the 9 categories of introns. **C.** Distribution

of the maximum entropy scores for 9-bp 5′ splice sites and 23-bp 3′ splice sites for the 9 categories of intron as obtained from MaxEntScan (38). **D.** Percentage of introns with UPF1 regional cross-linking events (*left*) and UPF1 regional cross-linking enrichment (*right*) for each splicing regulatory regions R1, R2, R3, R4 and R5 in each group of introns. Dashed lines indicate the average percentage of all 61872 analysed introns with a CLIP binding (*left*) and the one-fold enrichment (*right*) in the intronic regulatory regions (R2, R3, R4).


A.

MOTIF	P-Val	% targets (bg)	Best match
CCAGCCUG	1E-58	42.48% (3.06)	miR-485-5p
AGCAGUGG	1E-47	44.36% (6.01)	miR-4519
GAGCUGGG	1E-40	48.12% (9.74)	miR-4267
CGCAGGAG	1E-35	76.32% (33.72)	miR-1976
CAACAUGG	1E-33	41.73% (8.46)	miR-4716-5p
AGAGAGAG	1E-31	50.00% (14.3)	miR-4469
GGGAGG	1E-28	53.76% (18.44)	miR-1321
AAAAAAAA	1E-28	31.58% (5.24)	miR-3613-3p
AUUACAGG	1E-26	21.43% (1.74)	miR-5095
UUUUAAAA	1E-23	47.74% (16.5)	miR-3658
CUUUGGG	1E-21	30.83% (7.19)	miR-186
UUUCACC	1E-20	54.14% (23.3)	miR-5096
GCUAAUUU	1E-19	22.56% (3.74)	miR-4735-5p
GUGGGUG	1E-19	52.26% (23.01)	miR-3656

Supplementary Figure 3 | A. 14 motifs enriched in the 270 retained introns of the C5 category identified by HOMER (65). **B.** Changes in gene expression over time in

35

the nucleus (*gold* boxes) and cytoplasm (*blue* boxes) for groups of genes containing the 9 different categories of retained introns. Fold-changes obtained by comparing the log2 expression level at time of interest ($d_i = \{3, 7, 14, 22, 35\}$) with the expression level at iPSC stage (d_0). Data shown as box plots in which the centre line is the median, limits are the interquartile range and whiskers are the minimum and maximum.
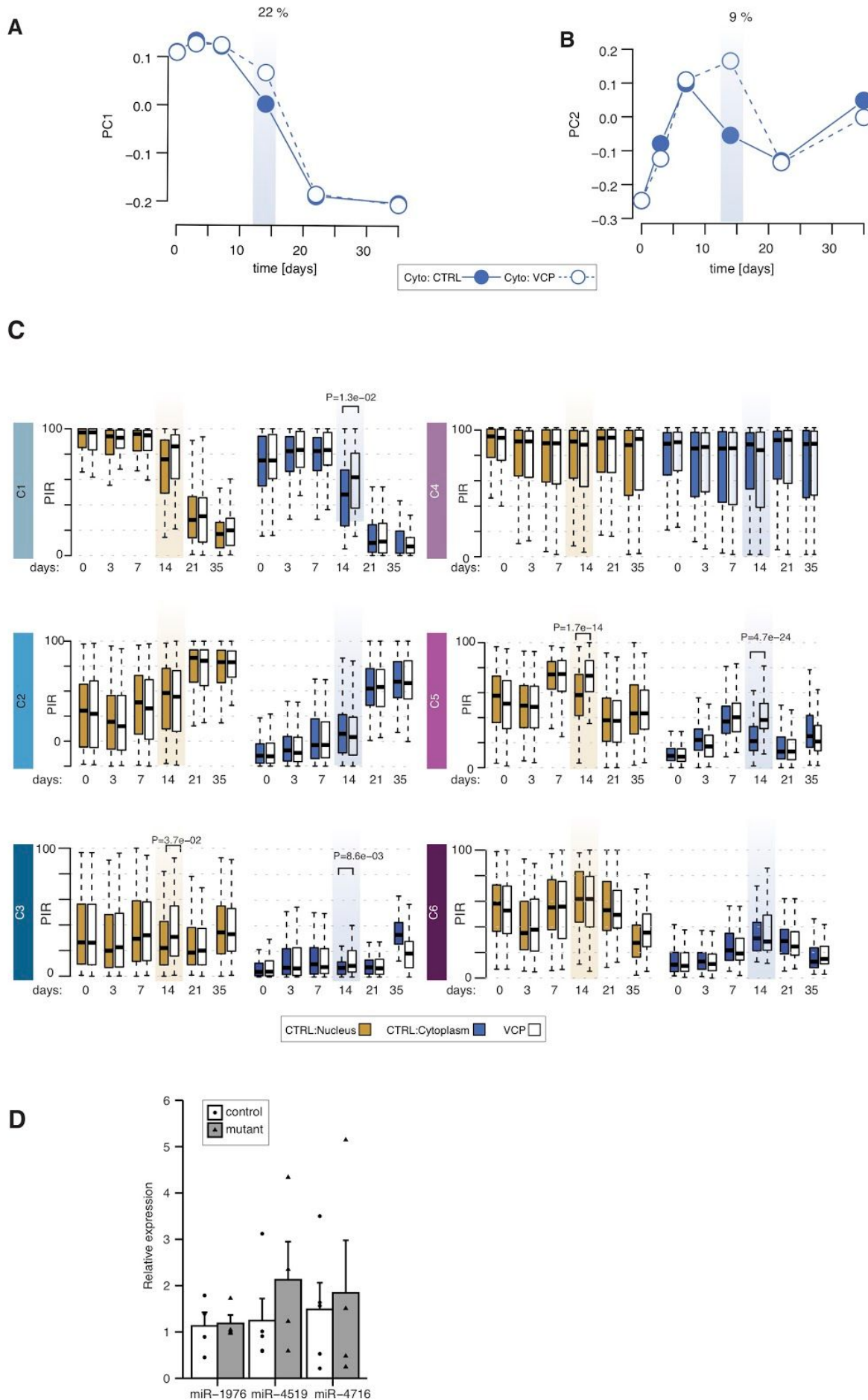
Supplementary Figure 4 | A, B. Singular value decomposition analysis of the PIR cytoplasmic values of 94,457 introns in n = 48 cytoplasmic samples. Line plots showing the PIR profiles of the first two singular vectors v_1 and v_2, capturing 22% and 9% of the variance in PIR respectively. Filled and empty data points indicate PIR values for the control and VCP^{mu} samples. **C.** Comparison of the distributions of nuclear and cytoplasmic PIR between control (*colored* boxes) and VCP^{mu} (*white* boxes) samples during MN differentiation for the 6 groups of cytoplasmic retained introns. P-values obtained with two-sided Welch t-test. Data shown as box plots in which the centre line is the median, limits are the interquartile range and whiskers are the minimum and maximum. **D.** MiRNA expression in "patterned" precursor motor neuron cells (DIV=14) – Relative expression of miR-1976, miR-4519 and miR-4716 in control (white) and mutant (gray) cells lines. Datapoints depicted as black circles in the controls and black triangles in mutant cells.
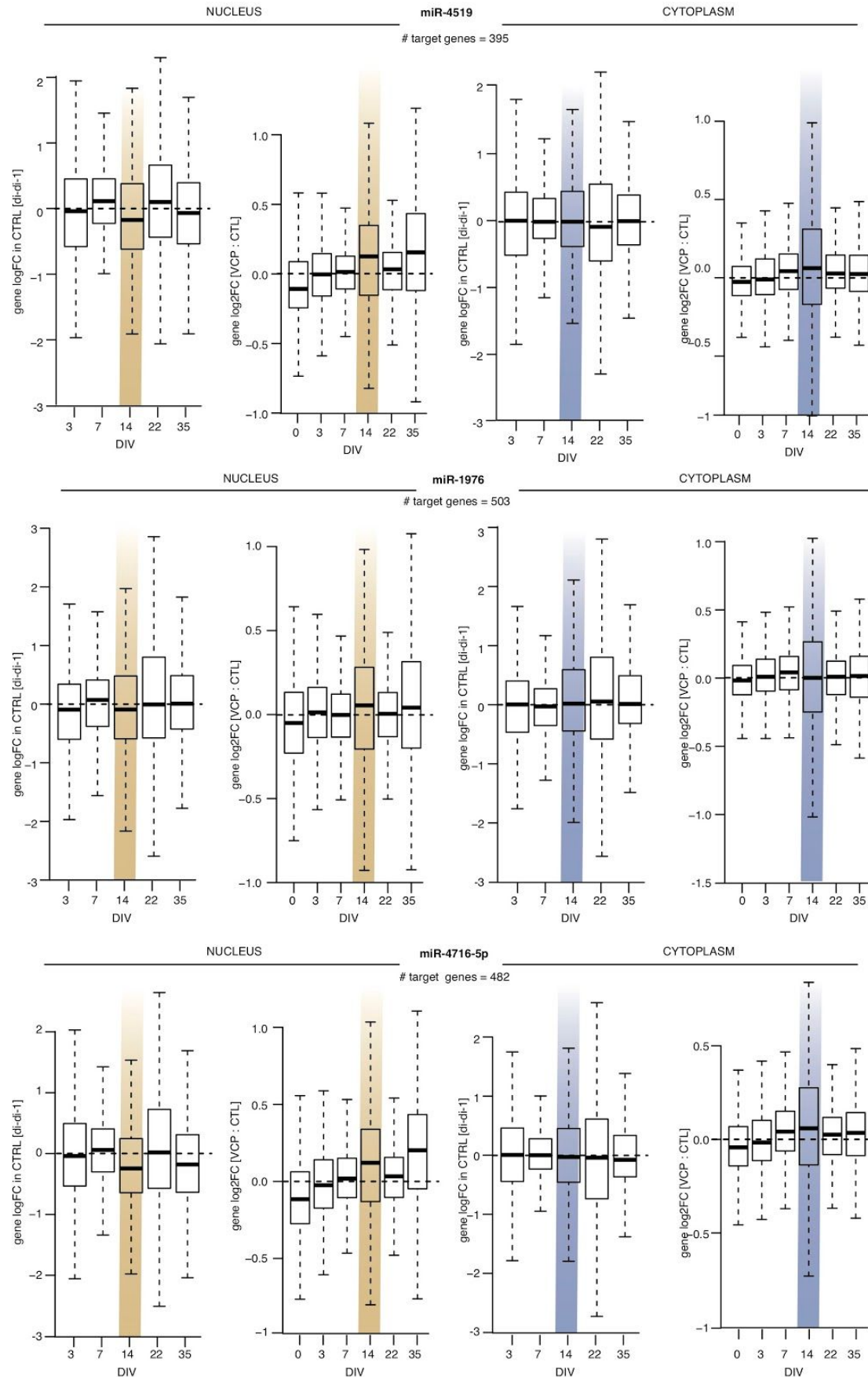
Supplementary Figure 5 | Distributions of the changes in nuclear and cytoplasmic expression over time of the control samples (*left*) and between VCP^{mu} and control samples at each time-point (*right*) for the 395 TargetScan (66) and miRDB (78) predicted target genes of miR-4519, miR-1976, and miR-4716-5p. Fold-changes over time obtained by comparing the log2 expression level at the time of interest ($d_t = \{0,\ 3, 7, 14, 22, 35\}$) with the expression level at previous stage (d_{t-1}). Gold shaded area indicates the time-point where the largest changes in cytoplasmic IR are observed either over time for the control samples or between control and mutant samples.
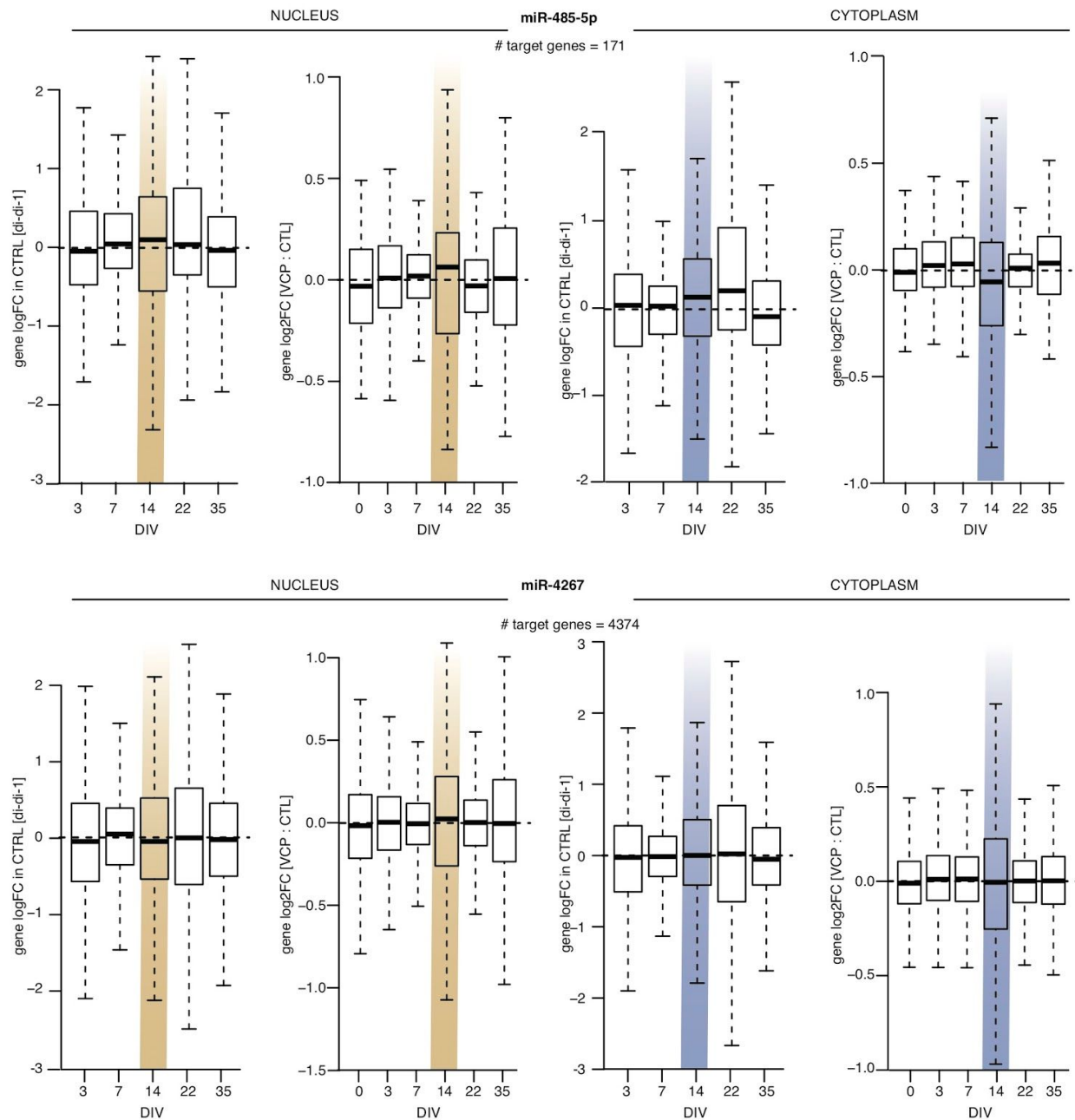
Supplementary Figure 6 | Same details and format as Supplementary Figure 5 but for miR-485-5p and miR-4267.

SUPPLEMENTARY TABLES 1-18 can be accessed here.

Table S1 | Description of the iPSC lines and RNA sequencing samples used in this study.

Table S2 | List of the 4490 nuclearly detained retained introns reported on Figs. 1F,H,J.

Table S3 | List of the 3633 cytoplasmic retained introns reported on Figs. 1F,H,J.

Tables S4-S12 | Lists of the 9 groups of retained introns (N1, N2 N3, C1, C2, C3, C4, C5, C6) associated with the distinct spatiotemporal dynamics during MN differentiation reported on Fig. 2A.

Tables S12-S17 | Frequency and enrichment in 133 RBPs crosslink events in 5 defined regions of interest (R1, R2, R3, R4, and R5).

AUTHOR CONTRIBUTIONS

Conceptualization, R.L., R.P.; Formal Analysis, R.L.; Investigation, R.L., M.P-H., H.C., J.N., G.E.T.; Writing − Original Draft, M.P-H., R.L., R.P.; Writing − Review & Editing, M.P-H., H.C., J.N., G.E.T., R.L., R.P; Resources, R.L., R.P.; Visualization, R.L.; Funding Acquisition, R.L., R.P.; Supervision, R.L, R.P.

REFERENCES

1. Braunschweig,U., Barbosa-Morais,N.L., Pan,Q., Nachman,E.N., Alipanahi,B., Gonatopoulos-Pournatzis,T., Frey,B., Irimia,M. and Blencowe,B.J. (2014) Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res.*, **24**, 1774–1786.

2. Boutz,P.L., Bhutkar,A. and Sharp,P.A. (2015) Detained introns are a novel, widespread class of post-transcriptionally spliced introns. *Genes Dev.*, **29**, 63–80.

3. Buckley,P.T., Lee,M.T., Sul,J.-Y., Miyashiro,K.Y., Bell,T.J., Fisher,S.A., Kim,J. and Eberwine,J. (2011) Cytoplasmic intron sequence-retaining transcripts can be dendritically targeted via ID element retrotransposons. *Neuron*, **69**, 877–884.

4. Yap,K., Lim,Z.Q., Khandelia,P., Friedman,B. and Makeyev,E.V. (2012) Coordinated regulation of neuronal mRNA steady-state levels through developmentally controlled intron retention. *Genes Dev.*, **26**, 1209–1223.

5. Mauger,O., Lemoine,F. and Scheiffele,P. (2016) Targeted Intron Retention and Excision for Rapid Gene Regulation in Response to Neuronal Activity. *Neuron*, **92**, 1266–1278.

6. Xu,Q., Walker,D., Bernardo,A., Brodbeck,J., Balestra,M.E. and Huang,Y. (2008) Intron-3 retention/splicing controls neuronal expression of apolipoprotein E in the CNS. *J. Neurosci.*, **28**, 1452–1459.

7. Braun,C.J., Stanciu,M., Boutz,P.L., Patterson,J.C., Calligaris,D., Higuchi,F., Neupane,R., Fenoglio,S., Cahill,D.P., Wakimoto,H., *et al.* (2017) Coordinated Splicing of Regulatory Detained Introns within Oncogenic Transcripts Creates an Exploitable Vulnerability in Malignant Glioma. *Cancer Cell*, **32**, 411–426.e11.

8. Ninomiya,K., Kataoka,N. and Hagiwara,M. (2011) Stress-responsive maturation of Clk1/4 pre-mRNAs promotes phosphorylation of SR splicing factor. *J. Cell Biol.*, **195**, 27–40.

9. Boothby,T.C., Zipper,R.S., van der Weele,C.M. and Wolniak,S.M. (2013) Removal of retained introns regulates translation in the rapidly developing gametophyte of Marsilea vestita. *Dev. Cell*, **24**, 517–529.

10. Kalyna,M., Simpson,C.G., Syed,N.H., Lewandowska,D., Marquez,Y., Kusenda,B., Marshall,J., Fuller,J., Cardle,L., McNicol,J., *et al.* (2012) Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in Arabidopsis. *Nucleic Acids Res.*, **40**, 2454–2469.

11. Khaladkar,M., Buckley,P.T., Lee,M.T., Francis,C., Eghbal,M.M., Chuong,T., Suresh,S., Kuhn,B., Eberwine,J. and Kim,J. (2013) Subcellular RNA sequencing reveals broad presence of cytoplasmic intron-sequence retaining transcripts in mouse and rat neurons. *PLoS One*, **8**, e76194.

12. Jaillon,O., Bouhouche,K., Gout,J.-F., Aury,J.-M., Noel,B., Saudemont,B., Nowacki,M., Serrano,V., Porcel,B.M., Ségurens,B., *et al.* (2008) Translational control of intron splicing in eukaryotes. *Nature*, **451**, 359–362.

13. Wong,J.J.-L., Ritchie,W., Ebner,O.A., Selbach,M., Wong,J.W.H., Huang,Y., Gao,D., Pinello,N., Gonzalez,M., Baidya,K., *et al.* (2013) Orchestrated intron retention regulates normal granulocyte differentiation. *Cell*, **154**, 583–595.

14. Vanichkina,D.P., Schmitz,U., Wong,J.J.-L. and Rasko,J.E.J. (2018) Challenges in defining the role of intron retention in normal biology and disease. *Semin. Cell Dev. Biol.*, **75**, 40–49.

15. Sharangdhar,T., Sugimoto,Y., Heraud-Farlow,J., Fernández-Moya,S.M., Ehses,J., de los Mozos,I.R., Ule,J. and Kiebler,M.A. (2017) A retained intron in the 3′- UTR of Calm3 mRNA mediates its Staufen2- and activity-dependent localization to neuronal dendrites. *EMBO reports*, **18**, 1762–1774.

16. Schmitz,U., Pinello,N., Jia,F., Alasmari,S., Ritchie,W., Keightley,M.-C., Shini,S., Lieschke,G.J., Wong,J.J.-L. and Rasko,J.E.J. (2017) Intron retention enhances gene regulatory complexity in vertebrates. *Genome Biol.*, **18**, 216.

17. Monteuuis,G., Wong,J.J.L., Bailey,C.G., Schmitz,U. and Rasko,J.E.J. (2019) The changing paradigm of intron retention: regulation, ramifications and recipes. *Nucleic Acids Res.*, **47**, 11497–11513.

18. Jangi,M., Fleet,C., Cullen,P., Gupta,S.V., Mekhoubad,S., Chiao,E., Allaire,N., Bennett,C.F., Rigo,F., Krainer,A.R., *et al.* (2017) SMN deficiency in severe models of spinal muscular atrophy causes widespread intron retention and DNA damage. *Proc. Natl. Acad. Sci. U. S. A.*, **114**, E2347–E2356.

19. Adusumalli,S., Ngian,Z.-K., Lin,W.-Q., Benoukraf,T. and Ong,C.-T. (2019) Increased intron retention is a post-transcriptional signature associated with progressive aging and Alzheimer's disease. *Aging Cell*, **18**, e12928.

20. Luisier,R., Tyzack,G.E., Hall,C.E., Mitchell,J.S., Devine,H., Taha,D.M., Malik,B., Meyer,I., Greensmith,L., Newcombe,J., *et al.* (2018) Intron retention and nuclear loss of SFPQ are molecular hallmarks of ALS. *Nat. Commun.*, **9**, 2010.

21. Wang,Q., Conlon,E.G., Manley,J.L. and Rio,D.C. (2020) Widespread intron retention impairs protein homeostasis in C9orf72 ALS brains. *Genome Res.*, **30**, 1705–1715.

22. Hogan,A.L., Grima,N., Fifita,J.A., McCann,E.P., Heng,B., Fat,S.C.M., Maharjan,R., Cain,A.K., Henden,L., Tarr,I., *et al.* (2020) SFPQ intron retention, reduced expression and aggregate formation in central nervous system tissue are pathological features of amyotrophic lateral sclerosis. *Cold Spring Harbor Laboratory*, 10.1101/2020.09.22.309062.

23. Tyzack,G.E., Neeves,J., Klein,P., Crerar,H., Ziff,O., Taha,D.M., Luisier,R., Luscombe,N.M. and Patani,R. (2020) An aberrant cytoplasmic intron retention programme is a blueprint for ALS-related RBP mislocalization. 10.1101/2020.07.20.211557.

24. Neumann,M., Sampathu,D.M., Kwong,L.K., Truax,A.C., Micsenyi,M.C., Chou,T.T., Bruce,J., Schuck,T., Grossman,M., Clark,C.M., *et al.* (2006) Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science*, **314**, 130–133.

25. Tyzack,G.E., Luisier,R., Taha,D.M., Neeves,J., Modic,M., Mitchell,J.S., Meyer,I., Greensmith,L., Newcombe,J., Ule,J., *et al.* (2019) Widespread FUS mislocalization is a molecular hallmark of amyotrophic lateral sclerosis. *Brain*, **142**, 2572–2580.

26. Kiskinis,E., Sandoe,J., Williams,L.A., Boulting,G.L., Moccia,R., Wainger,B.J., Han,S., Peng,T., Thams,S., Mikkilineni,S., *et al.* (2014) Pathways disrupted in human ALS motor neurons identified through genetic correction of mutant SOD1. *Cell Stem Cell*, **14**, 781–795.

27. Bray,N.L., Pimentel,H., Melsted,P. and Pachter,L. (2016) Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.*, **34**, 525–527.

28. Fraley,C. and Raftery,A.E. mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation.

29. Boldstad,B.M., Irizarry,R.A., Astrand,M. and Speed,T.P. (2003) A Comparison of Normalization Methods for High Density Oligonucleotide Array Data Based on Bias and Variance. *Bioinformatics*, **19**, 185–193.

30. Pimentel,H.J., Bray,N., Puente,S., Melsted,P. and Pachter,L. (2016) Differential analysis of RNA-Seq incorporating quantification uncertainty. bioRxiv.

31. Irimia,M., Weatheritt,R.J., Ellis,J.D., Parikshak,N.N., Gonatopoulos-Pournatzis,T., Babor,M., Quesnel-Vallières,M., Tapial,J., Raj,B., O'Hanlon,D., *et al.* (2014) A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell*, **159**, 1511–1523.

32. Alexa,A. and Rahnenfuhrer,J. (2016) topGO: Enrichment Analysis for Gene Ontology.

33. Attig,J., Agostini,F., Gooding,C., Chakrabarti,A.M., Singh,A., Haberman,N., Zagalak,J.A., Emmett,W., Smith,C.W.J., Luscombe,N.M., *et al.* (2018) Heteromeric RNP Assembly at LINEs Controls Lineage-Specific RNA Processing. *Cell*, **174**, 1067–1081.e17.

34. Sloan,C.A., Chan,E.T., Davidson,J.M., Malladi,V.S., Strattan,J.S., Hitz,B.C., Gabdank,I., Narayanan,A.K., Ho,M., Lee,B.T., *et al.* (2016) ENCODE data at the ENCODE portal. *Nucleic Acids Res.*, **44**, D726–32.

35. Van Nostrand,E.L., Freese,P., Pratt,G.A., Wang,X. and Wei,X. (2017) A large-scale binding and functional map of human RNA binding proteins. *bioRxiv*.

36. Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.

37. Hinrichs,A.S., Karolchik,D., Baertsch,R., Barber,G.P., Bejerano,G., Clawson,H., Diekhans,M., Furey,T.S., Harte,R.A., Hsu,F., *et al.* (2006) The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.*, **34**, D590–8.

38. Yeo,G. and Burge,C.B. (2004) Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.*, **11**, 377–394.

39. Ule,J., Stefani,G., Mele,A., Ruggiu,M., Wang,X., Taneri,B., Gaasterland,T., Blencowe,B.J. and Darnell,R.B. (2006) An RNA map predicting Nova-dependent splicing regulation. *Nature*, **444**, 580–586.

40. Cereda,M., Pozzoli,U., Rot,G., Juvan,P., Schweitzer,A., Clark,T. and Ule,J. (2014) RNAmotifs: prediction of multivalent RNA motifs that control alternative

splicing. *Genome Biol.*, **15**, R20.

41. Pollard,K.S., Hubisz,M.J., Rosenbloom,K.R. and Siepel,A. (2010) Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.*, **20**, 110–121.

42. Siepel,A., Bejerano,G., Pedersen,J.S., Hinrichs,A.S., Hou,M., Rosenbloom,K., Clawson,H., Spieth,J., Hillier,L.W., Richards,S., *et al.* (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.*, **15**, 1034–1050.

43. Hall,C.E., Yao,Z., Choi,M., Tyzack,G.E., Serio,A., Luisier,R., Harley,J., Preza,E., Arber,C., Crisp,S.J., *et al.* (2017) Progressive Motor Neuron Pathology and the Role of Astrocytes in a Human Stem Cell Model of VCP-Related ALS. *Cell Rep.*, **19**, 1739–1749.

44. Edwards,C.R., Ritchie,W., -L. Wong,J.J., Schmitz,U., Middleton,R., An,X., Mohandas,N., Rasko,J.E.J. and Blobel,G.A. (2016) A dynamic intron retention program in the mammalian megakaryocyte and erythrocyte lineages. *Blood*, **127**, e24–e34.

45. Pimentel,H., Parra,M., Gee,S.L., Mohandas,N., Pachter,L. and Conboy,J.G. (2016) A dynamic intron retention program enriched in RNA processing genes regulates gene expression during terminal erythropoiesis. *Nucleic Acids Res.*, **44**, 838–851.

46. Middleton,R., Gao,D., Thomas,A., Singh,B., Au,A., Wong,J.J.-L., Bomane,A., Cosson,B., Eyras,E., Rasko,J.E.J., *et al.* (2017) IRFinder: assessing the impact of intron retention on mammalian gene expression. *Genome Biol.*, **18**, 51.

47. Ullrich,S. and Guigó,R. (2020) Dynamic changes in intron retention are tightly associated with regulation of splicing factors and proliferative activity during B-cell development. *Nucleic Acids Res.*, **48**, 1327–1340.

48. Dvinge,H. and Bradley,R.K. (2015) Widespread intron retention diversifies most cancer transcriptomes. *Genome Med.*, **7**, 45.

49. Galante,P.A.F., Sakabe,N.J., Kirschbaum-Slager,N. and de Souza,S.J. (2004) Detection and evaluation of intron retention events in the human transcriptome. *RNA*, **10**, 757–765.

50. Sakabe,N.J. and de Souza,S.J. (2007) Sequence features responsible for intron retention in human. *BMC Genomics*, **8**, 59.

51. Dirksen,W.P., Sun,Q. and Rottman,F.M. (1995) Multiple splicing signals control alternative intron retention of bovine growth hormone pre-mRNA. *J. Biol. Chem.*, **270**, 5346–5352.

52. Lejeune,F., Cavaloc,Y. and Stevenin,J. (2001) Alternative splicing of intron 3 of the serine/arginine-rich protein 9G8 gene. Identification of flanking exonic splicing enhancers and involvement of 9G8 as a trans-acting factor. *J. Biol. Chem.*, **276**, 7850–7858.

53. Zhu,L., Zhang,Y., Zhang,W., Yang,S., Chen,J.-Q. and Tian,D. (2009) Patterns of exon-intron architecture variation of genes in eukaryotic genomes. *BMC Genomics*, **10**, 47.

54. Robberson,B.L., Cote,G.J. and Berget,S.M. (1990) Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol. Cell. Biol.*, **10**, 84–94.

55. Ule,J. and Blencowe,B.J. (2019) Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution. *Mol. Cell*, **76**, 329–345.

56. Graber,T.E., Freemantle,E., Anadolu,M.N., Hébert-Seropian,S., MacAdam,R.L., Shin,U., Hoang,H.-D., Alain,T., Lacaille,J.-C. and Sossin,W.S. (2017) UPF1 Governs Synaptic Plasticity through Association with a STAU2 RNA Granule. *J. Neurosci.*, **37**, 9116–9131.

57. Dalla Costa,I., Buchanan,C.N., Zdradzinski,M.D., Sahoo,P.K., Smith,T.P., Thames,E., Kar,A.N. and Twiss,J.L. (2020) The functional organization of axonal mRNA transport and translation. *Nat. Rev. Neurosci.*, 10.1038/s41583-020-00407-7.

58. Das,S., Singer,R.H. and Yoon,Y.J. (2019) The travels of mRNAs in neurons: do they know where they are going? *Curr. Opin. Neurobiol.*, **57**, 110–116.

59. Han,J., Lee,Y., Yeom,K.-H., Kim,Y.-K., Jin,H. and Kim,V.N. (2004) The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev.*, **18**, 3016–3027.

60. Hu,R., Zhu,X., Chen,C., Xu,R., Li,Y. and Xu,W. (2018) RNA-binding protein PUM 2 suppresses osteosarcoma progression via partly and competitively binding to STARD 13 3′ UTR with mi RNA s. *Cell Prolif.*, **51**, e12508.

61. Treiber,T., Treiber,N., Plessmann,U., Harlander,S., Daiß,J.-L., Eichner,N., Lehmann,G., Schall,K., Urlaub,H. and Meister,G. (2017) A Compendium of RNA-Binding Proteins that Regulate MicroRNA Biogenesis. *Mol. Cell*, **66**, 270–284.e13.

62. Elbarbary,R.A., Miyoshi,K., Hedaya,O., Myers,J.R. and Maquat,L.E. (2017) UPF1 helicase promotes TSN-mediated miRNA decay. *Genes Dev.*, **31**, 1483–1493.

63. Park,J., Seo,J.-W., Ahn,N., Park,S., Hwang,J. and Nam,J.-W. (2019) UPF1/SMG7-dependent microRNA-mediated gene regulation. *Nat. Commun.*, **10**, 4181.

64. Jin,H., Suh,M.R., Han,J., Yeom,K.-H., Lee,Y., Heo,I., Ha,M., Hyun,S. and Kim,V.N. (2009) Human UPF1 participates in small RNA-induced mRNA downregulation. *Mol. Cell. Biol.*, **29**, 5789–5799.

65. Heinz,S., Benner,C., Spann,N., Bertolino,E., Lin,Y.C., Laslo,P., Cheng,J.X., Murre,C., Singh,H. and Glass,C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*, **38**, 576–589.

66. Agarwal,V., Bell,G.W., Nam,J.-W. and Bartel,D.P. (2015) Predicting effective microRNA target sites in mammalian mRNAs. *Elife*, **4**.

67. John,B., Enright,A.J., Aravin,A., Tuschl,T., Sander,C. and Marks,D.S. (2004) Human MicroRNA targets. *PLoS Biol.*, **2**, e363.

68. Duss,O., Michel,E., Yulikov,M., Schubert,M., Jeschke,G. and Allain,F.H.-T. (2014) Structural basis of the non-coding RNA RsmZ acting as a protein sponge. *Nature*, **509**, 588–592.

69. Yoshimoto,R., Kaida,D., Furuno,M. and Burroughs,A.M. (2017) Global analysis of pre-mRNA subcellular localization following splicing inhibition by spliceostatin A. RNA.

70. Ninomiya,K., Adachi,S., Natsume,T., Iwakiri,J., Terai,G., Asai,K. and Hirose,T. LncRNA-dependent nuclear stress bodies promote intron retention through SR protein phosphorylation. 10.1101/681056.

71. Wong,J.J.-L., Gao,D., Nguyen,T.V., Kwok,C.-T., van Geldermalsen,M., Middleton,R., Pinello,N., Thoeng,A., Nagarajah,R., Holst,J., *et al.* (2017) Intron retention is regulated by altered MeCP2-mediated splicing factor recruitment. *Nat. Commun.*, **8**, 15134.

72. Moss,W.N. and Steitz,J.A. (2013) Genome-wide analyses of Epstein-Barr virus reveal conserved RNA structures and a novel stable intronic sequence RNA. *BMC Genomics*, **14**, 543.

73. Yin,Q.-F., Yang,L., Zhang,Y., Xiang,J.-F., Wu,Y.-W., Carmichael,G.G. and

Chen,L.-L. (2012) Long noncoding RNAs with snoRNA ends. *Mol. Cell*, **48**, 219–230.

74. Xie,T., Deng,L., Mei,P., Zhou,Y., Wang,B., Zhang,J., Lin,J., Wei,Y., Zhang,X. and Xu,R. (2014) A genome-wide association study combining pathway analysis for typical sporadic amyotrophic lateral sclerosis in Chinese Han populations. *Neurobiol. Aging*, **35**, 1778.e9–1778.e23.

75. Wang,L. and Zhang,L. (2020) MicroRNAs in amyotrophic lateral sclerosis: from pathogenetic involvement to diagnostic biomarker and therapeutic agent development. *Neurol. Sci.*, **41**, 3569–3577.

76. Eyileten,C., Sharif,L., Wicik,Z., Jakubik,D., Jarosz-Popek,J., Soplinska,A., Postula,M., Czlonkowska,A., Kaplon-Cieslicka,A. and Mirowska-Guzel,D. (2021) The Relation of the Brain-Derived Neurotrophic Factor with MicroRNAs in Neurodegenerative Diseases and Ischemic Stroke. *Mol. Neurobiol.*, **58**, 329–347.

77. Zuo,X., Zhou,J., Li,Y., Wu,K., Chen,Z., Luo,Z., Zhang,X., Liang,Y., Esteban,M.A., Zhou,Y., *et al.* (2021) TDP-43 aggregation induced by oxidative stress causes global mitochondrial imbalance in ALS. *Nat. Struct. Mol. Biol.*, 10.1038/s41594-020-00537-7.

78. Chen,Y. and Wang,X. (2020) miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res.*, **48**, D127–D131.

79. Kiskinis,E., Sandoe,J., Williams,L.A., Boulting,G.L., Moccia,R., Wainger,B.J., Han,S., Peng,T., Thams,S., Mikkilineni,S., *et al.* (2014) Pathways disrupted in human ALS motor neurons identified through genetic correction of mutant SOD1. *Cell Stem Cell*, **14**, 781–795.

80. Kapeli,K., Pratt,G.A., Vu,A.Q., Hutt,K.R., Martinez,F.J., Sundararaman,B., Batra,R., Freese,P., Lambert,N.J., Huelga,S.C., *et al.* (2016) Distinct and shared functions of ALS-associated proteins TDP-43, FUS and TAF15 revealed by multisystem analyses. *Nat. Commun.*, **7**, 12143.