

ARTICLE OPEN

An alanine expanded PABPN1 causes increased utilization of intronic polyadenylation sites

Tooba Abbassi-Daloui¹, Soheil Yousefi¹, Eleonora de Klerk^{2,3}, Laurens Grossouw¹, Muhammad Riaz¹, Peter A. C. 't Hoen¹ and Vered Raz¹

In eukaryote genomes, the polyadenylation site marks termination of mature RNA transcripts by a poly-adenine tail. The polyadenylation site is recognized by a dynamic protein complex, among which the poly-adenine-binding protein nuclear1 plays a key role. Reduced poly-adenine-binding protein nuclear1 levels are found in aged muscles and are even lower in oculopharyngeal muscular dystrophy patients. Oculopharyngeal muscular dystrophy is a rare, late onset autosomal dominant myopathy, and is caused by an alanine expansion mutation in poly-adenine-binding protein nuclear1. Mutant poly-adenine-binding protein nuclear1 forms insoluble nuclear aggregates leading to depletion of functional poly-adenine-binding protein nuclear1 levels. In oculopharyngeal muscular dystrophy models, increased utilization of proximal polyadenylation sites has been observed in tandem 3'-untranslated regions, and most often cause gene upregulation. However, global alterations in expression profiles can only partly be explained by polyadenylation site switches within the most distal 3'-untranslated region. Most poly-adenine signals are found at the distal 3'-untranslated region, but a significant part is also found in internal gene regions, like introns, exons, and internal 3'-untranslated regions. Here, we investigated poly-adenine-binding protein nuclear1's role in polyadenylation site utilization in internal gene regions. In the quadriceps muscle of oculopharyngeal muscular dystrophy mice expressing expPABPN1 we found significant polyadenylation site switches between gene regions in 17% of genes with polyadenylation site in multiple regions ($N = 574$; 5% False Discovery Rate). Polyadenylation site switches between gene regions were associated with differences in transcript expression levels and alterations in open reading frames. Transcripts ending at internal polyadenylation site were confirmed in tibialis anterior muscles from the same mice and in mouse muscle cell cultures overexpressing expPABPN1. The polyadenylation site switches were associated with nuclear accumulation of full-length transcripts. Our results provide further insights into the diverse roles of poly-adenine-binding protein nuclear1 in the post-transcriptional control of muscle gene expression and its relevance for oculopharyngeal muscular dystrophy pathology and muscle aging.

npj Aging and Mechanisms of Disease (2017)3:6; doi:10.1038/s41514-017-0007-x

INTRODUCTION

The maturation of a messenger RNA (mRNA) is a multistep process influencing the RNA landscape of a cell. Polyadenylation is an essential step in mRNA maturation, affecting mRNA stability, nuclear export, and translation efficiency.¹ Polyadenylation includes endonucleolytic cleavage of mRNA at the polyadenylation site (PAS) and is completed by addition of an adenine tail. The majority of genes have multiple PASs, generating distinct transcripts.² The majority of PASs are located in the distal 3'-untranslated region (3'-UTR) and alternative PAS in the distal 3'-UTR most often result in transcripts with a shorter 3'-UTR, suggesting to affect mRNA stability.³ Although PASs are most abundant in the distal 3'-UTR, polyadenylation within internal gene regions (introns, exons, and internal 3'-UTRs) were also found.^{2,4} Genes with PASs in internal gene regions could generate a wide range of alternative transcripts, among those with alternative coding sequence (CDS).⁵

Polyadenylation is regulated by a dynamic protein complex and the poly-adenine (poly(A))-binding protein nuclear 1 (PABPN1) plays a central role.⁶ PABPN1 is ubiquitously expressed and is primarily recognized for regulation of the poly(A) tail length.⁷

PABPN1 regulates PAS utilization in the distal 3'-UTR by suppressing proximal PASs.^{8,9} Reduced levels of functional PABPN1 are found in aged muscles and in muscles from oculopharyngeal muscular dystrophy (OPMD) patients.¹⁰ OPMD is caused by an alanine expansion at the N-terminus of PABPN1 (expPABPN1).¹¹ ExpPABPN1 forms insoluble aggregates in myonuclei,¹² which reduces levels of soluble PABPN1 (ref. 13). Reduced levels of functional PABPN1 cause switches in PAS utilization in the distal 3'-UTR, with most prominent switches from distal to proximal PAS 3'-UTR.^{8,9} Shorter transcripts lead to a loss of microRNA (miRNA)-binding sites, which could generally explain upregulation of shorter transcripts from proximal PAS.⁸ More recently, Li *et al.*,⁵ suggested a role for PABPN1 in internal PAS utilization. Yet, this was not investigated in the context of disease.

In this study, we investigated a role for expPABPN1 in PAS utilization in internal gene regions in muscles. We employed our previously generated genome-wide single molecule polyA-Seq data⁹ where dealt only with PAS switches within the same 3'-UTR region. Here, we extended PAS analysis across the genome and found internal PASs in a quarter of muscle genes. We show that changes in PAS utilization between gene regions occur in both upstream and downstream directions (Fig. 1). The expPABPN1-

¹Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands; ²Department of Microbiology and Immunology, University of California, San Francisco (UCSF), San Francisco, CA, USA and ³Diabetes Center, University of California, San Francisco (UCSF), San Francisco, CA, USA

Correspondence: Vered Raz (v.raz@lymc.nl)

Peter A. C. 't Hoen and Vered Raz are senior authors.

Received: 17 October 2016 Revised: 19 January 2017 Accepted: 21 February 2017

Published online: 07 April 2017

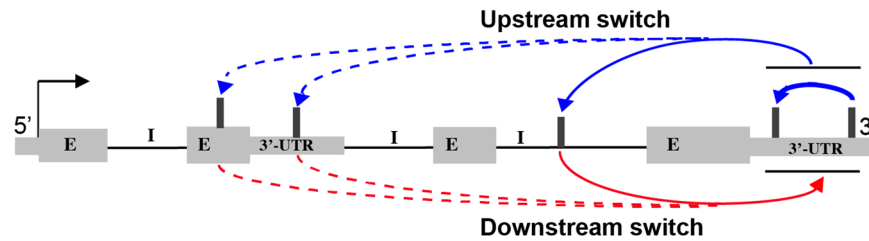


Fig. 1 Schematic representation of PAS switches. Switches in PAS utilization are shown with *arrows*: upstream switches are depicted in *blue* and downstream switches in *red*. An upstream switch is most often from distal 3'-UTR to any internal gene regions (intron, exon, and internal 3'-UTR), and a downstream switch is most often from any internal gene regions to distal 3'-UTR. Switches between tandem PASs in the distal 3'-UTR are also shown. The most prominent switches are depicted with *continues line*. PASs are depicted with *vertical bars*, the gene regions, exon (E), intron (I), and 3'-UTR are indicated. Illustrated are the most abundant switch possibilities

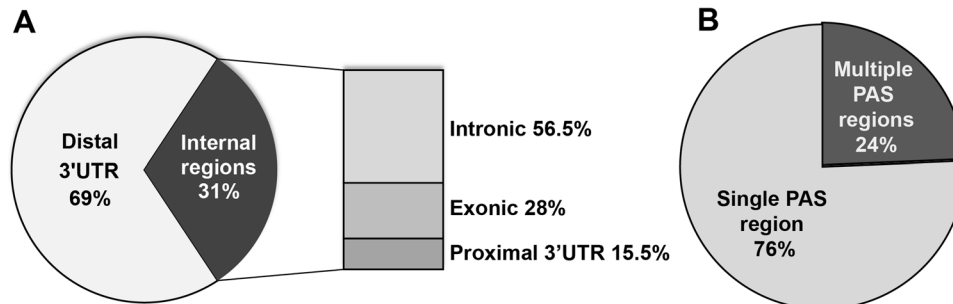


Fig. 2 Distribution of PAS locations. **a** Pie chart showing the percentage of PASs distribution in different regions of the gene (24,248; 6273; 3106, and 1724 PASs located at distal 3'-UTRs, introns, exons, and internal 3'-UTR, respectively). **b** Pie chart showing the percentage of genes containing PAS in single or multiple regions (3296 genes have multiple PAS regions and 10,355 genes have single PAS region)

induced PAS switches are associated with few non-canonical PAS motifs. We show that PAS switches in genes involved in muscle pathologies are affected by expPABPN1 expression. Our study suggests that the role of PABPN1 in regulation of mRNA processing is broad and can contribute to muscle weakness in OPMD via different mechanisms.

RESULTS

Polyadenylation within gene regions is affected by expPABPN1 in muscles

To investigate the effect of expPABPN1 on PAS usage in internal gene regions we employed our previously generated single molecule polyA sequencing data from the mouse quadriceps muscles of A17.1 mice expressing expPABPN1 and their matched wild-type FVB controls.⁹ Only PASs with coverage of at least two reads per PASs in at least 6 out of 12 samples were included in the analysis, leaving 35,351 PASs located in 13,685 genes. Confirming previous studies,^{8, 9} the majority of PAS (69%) mapped to the most distal 3'-UTR region, and the rest (31%; Fig. 2a) mapped into internal gene regions, including introns, exons, and internal 3'-UTRs. Three-thousand two-hundred and ninety six genes (24%) contained PASs in multiple regions (Fig. 2b), and those were used to investigate a role for expPABPN1 in PAS utilization in internal regions.

In our previous study, we reported the effect of expPABPN1 on PAS utilization in the distal 3'-UTR region (PASs in distal 3'-UTR). To investigate expPABPN1 effect on internal PAS utilization, PAS reads within a gene region were summed and the relative usage of the different PAS regions were assessed with a generalized linear mixed model on the binomial distribution of the counts for each region in a gene. Per gene, a PAS region that was more used in FVB control compared to A17.1 is dubbed as a default region, and a PAS region that was more used in A17.1 compared to FVB is referred to as a switch region.

The possible switches in PAS utilization could be from a default distal PAS to an upstream PAS (named here an upstream switch) or from a default PAS in internal region to a downstream PAS (named here a downstream switch) (Fig. 1). Examples of upstream and downstream switches are shown in Fig. 3. In *2300009A05Rik* (shortened here to *Rika05*), the default distal PAS was located in the distal 3'-UTR and a secondary PAS was found within the intron (Fig. 3a; FVB). The PAS within the intronic region was relatively more utilized in A17.1 compared to FVB (Fig. 3a), indicating an upstream switch. In *Arih1*, additionally to the previously described PAS switch within the 3'-UTR,⁹ a PAS was found within an intron, which was relatively more utilized in FVB. In A17.1 the PAS in the distal 3'-UTR was relatively more utilized, indicating a downstream switch (Fig. 3b).

From the 3296 genes with two or more PAS regions, 17% showed a significant switch in PAS utilization between regions ($N = 574$; false discovery rate (FDR) < 0.05; Fig. 3c). An upstream switch was found in 66% of those genes (Fig. 3c; list in Supplementary Table S2).

PASs within gene regions affect gene expression levels

Within the same 3'-UTR, switches from distal to proximal PAS predominantly affect mRNA stability leading to transcript upregulation.^{8, 9} Here, we investigated whether PAS utilization between gene regions also affects mRNA expression levels. For each gene fold changes were determined (Supplementary Table S2), and we found that gene dysregulation was significantly higher in genes with a PAS switch between gene regions compared to genes without a switch (50% and 28%, respectively; Fisher's exact test p -value = $8.2E10^{-12}$) (Fig. 4a). The majority of genes (66%) with PAS switch were found to be upregulated (Fig. 4a). We then examined whether PAS switch direction specifically affect gene dysregulation direction, and found in genes with a downstream switch a slightly higher proportion of dysregulated genes compared with an upstream switch (56.5% and 46%, respectively).

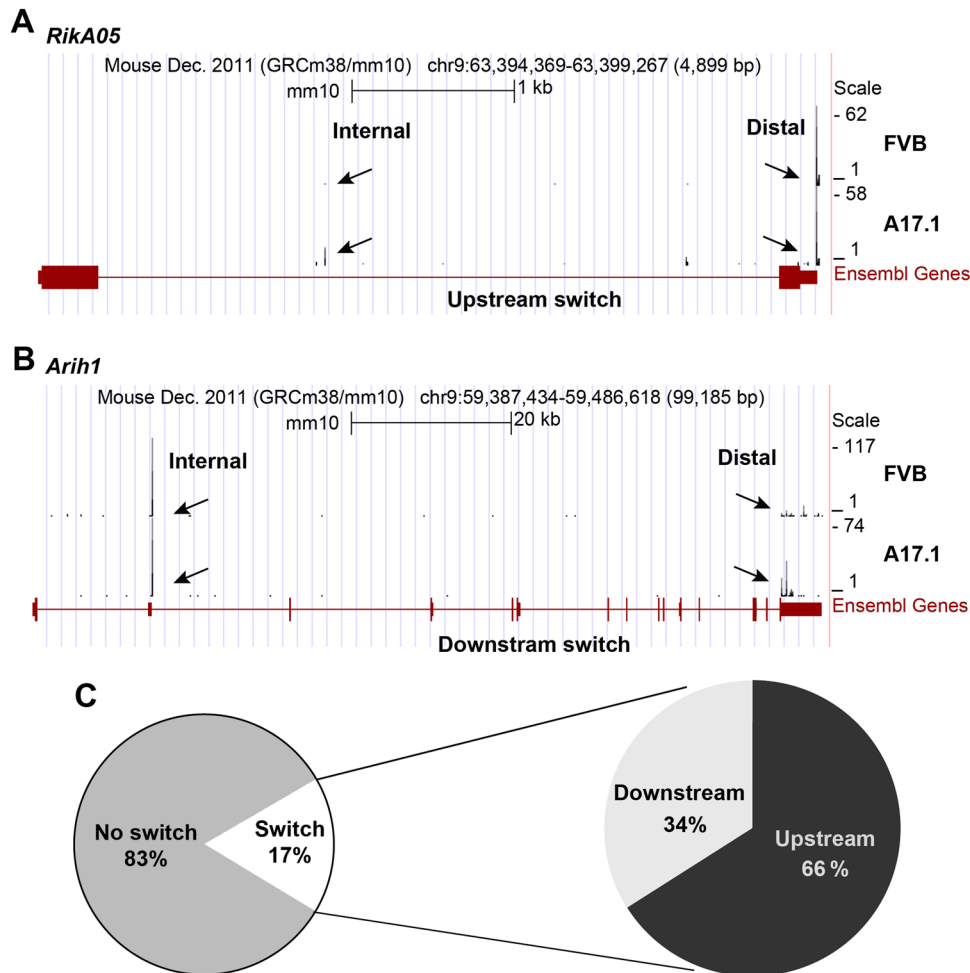


Fig. 3 Upstream and downstream switches in PAS regions. **a–b** USCS Genome Browser screenshots displaying PASs in *Rika05* (upstream switch) and *Arih1* (downstream switch) genes. The *y*-axis represents the coverage of the peaks, corresponding to the number of reads mapped at each PAS. One of the FVB control and A17.1 mice are depicted in independent traces. PASs are depicted by *arrows*; distal or internal PASs are indicated. The gene structure based on Ensembl annotation is presented in *red*. The amplified regions that were used for validation are indicated below the gene structure; **c** Pie chart on *left side* shows the percentage of genes with changes in PAS utilization region (567 genes) and the *right* pie chart shows the percentage of genes with downstream or upstream switches in PAS utilization (193 and 374 genes, respectively)

We further investigated whether switches to distinct gene regions may affect dysregulation direction. As expected, PAS switches to/from the distal 3'-UTR were predominant (78%) (Fig. 4b). In internal gene regions PAS switches from/to introns was the most abundant (Supplementary Table S3). The majority PAS switches were upstream (66%) compared with a downstream switch (34%). In both switch direction, the proportion of upregulated genes was higher compared with downregulation, but in the downstream switch relatively more genes had upregulation compared with the upstream switch (77% and 59%, respectively). Notably, an upstream switch from distal 3'-UTR to internal gene regions was not associated with a specific dysregulation direction (Fig. 4b). Thus, in contrast to upstream PAS switches within the 3'-UTR, which predominantly lead to increase in transcript levels, an upstream switch between 3'-UTR and introns or other gene regions does not specifically affect dysregulation direction.

To validate PAS switches between gene regions that were found in quadriceps muscles, reverse transcription quantitative PCR (RT-qPCR) was performed on RNA from tibialis anterior muscles, which are severely affected in A17.1 mouse.¹⁴ PAS switches between gene regions were validated in six genes containing one distinct PAS per gene region. *Arih1* represents a

downstream switch, whereas *Rika05*, *Vezf1*, *Lmna*, *Figf*, and *Ttn* genes represent an upstream switch (Fig. 5a). *Usp19* was selected as control as it possesses internal and distal PASs without switch in utilization in A17.1 (Fig. 5a). We first confirmed the expression of transcripts from internal PAS using a primer set specific to the internal 3'-UTR or intronic region (Supplementary Fig. S3). PAS usage in intronic regions resulted in alternative transcripts for *Rika05*, *Vezf1*, *Figf*, and *Arih1* transcripts (Supplementary Fig. S3). Alternative transcripts from internal 3'-UTR region were found for *Lmna* and *Usp19* (Supplementary Fig. S3). The intronic region near PAS in *Ttn* was not suitable for qRT-PCR, and the proximal primer set was, therefore, designed on the nearby exon (Supplementary Fig. S2). PAS switch was assessed by the ratio of PCR products from the internal region and the distal 3'-UTR. In accordance with the polyA-Seq data, we found a downstream switch between internal PAS and distal PAS for *Arih1* and an upstream switch for *Rika05*, *Lmna*, *Figf*, and *Ttn* (Fig. 5b). For *Vezf1* only a trend suggesting an upstream switch was found (Fig. 5b). As expected, no switch in PAS utilization was found for *Usp19* (Fig. 5b). For those seven genes we then also assessed whether PAS switch is accompanied by a change in expression levels. Overall, fold change direction was found to be consistent between RNA-seq and qRT-PCR for 6 out of the 7 genes (*Arih1*, *Rika05*, *Lmna*, *Figf*,

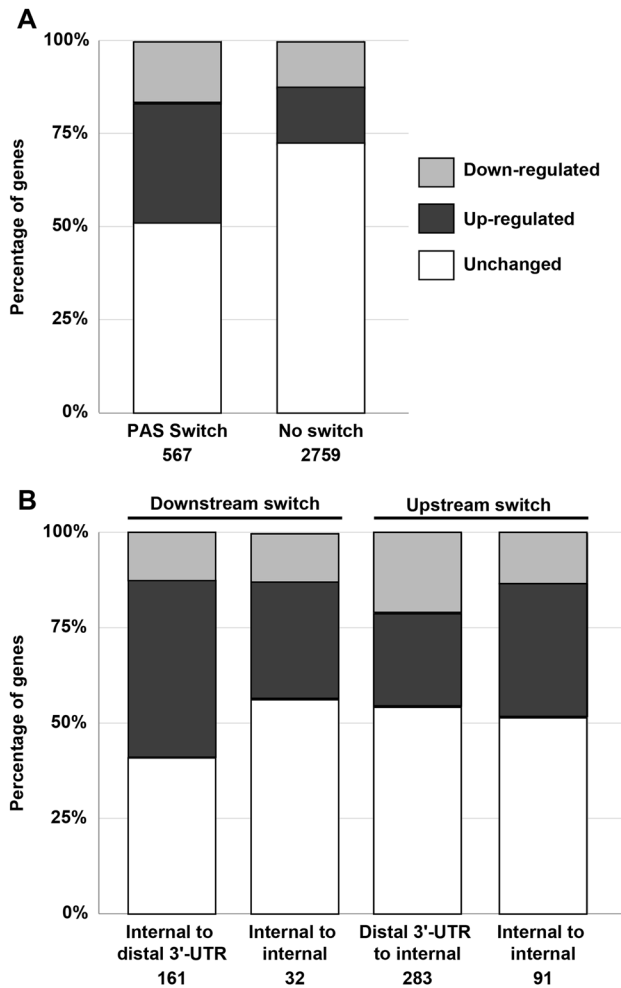


Fig. 4 Differential expression analysis in upstream and downstream switch groups. Bar chart shows differential expression analysis in genes with or without a significant switch in PAS utilization (a) and in genes with upstream or downstream switches (b). The number of genes for each category is indicated

Ttn, and *Usp19*; Fig. 5c, d). Only for *VeZF1* fold change was not validated using qRT-PCR (Fig. 5c, d). Among the five genes with a downstream switch in three genes upregulation and in two genes downregulation was found in A17.1.

To confirm a role of expPABPN1 in intronic PAS utilization, we performed qRT-PCR in the IM2 muscle cell culture model for OPMD. In this model, wild-type PABPN1 (WT) or expPABPN1 were stably overexpressed at similar levels.¹³ Significant switches between gene regions were found for 4 out of the 5 genes (Fig. 6a). Switch direction was similar between the muscle cell culture and muscles for *Rika05* and *VeZF1*, and *Arih1*, but the fold change was smaller in the cell culture model (Fig. 6a). In *Lmna* a downstream switch, instead of an upstream switch, was found (Fig. 6a). No PAS switch was found in *Figf*, and PAS switch also remained unchanged in *Usp19* (Fig. 6a). *Ttn* was excluded from this study, since the intronic region near PAS could not be amplified. Together those results further support a role for expPABPN1 in PAS switch between regions.

Nuclear export of transcripts is affected by expPABPN1

PABPN1 shuttles between the nuclear and cytoplasm, and expPABPN1 has a higher nuclear accumulation compared with its wild-type counterpart.¹⁵ A higher nuclear accumulation of expPABPN1 compared with WT PABPN1 was also found in the

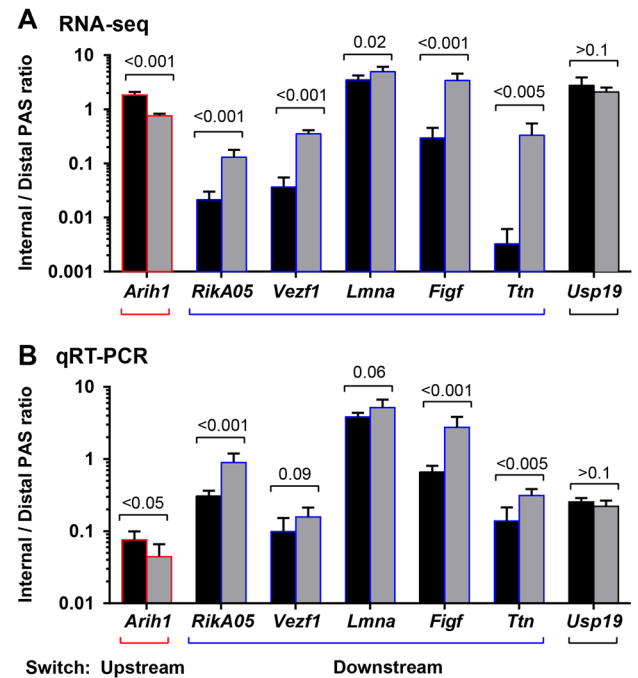


Fig. 5 PAS switch between regions in quadriceps and tibialis anterior muscles. **a** Bar chart shows the ratio between distal to internal PAS obtained from read counts from polyA-sequencing in quadriceps muscle. Expression values were obtained after TMM normalization. Averages and SD are from six mice. **b** Bar chart shows the ratio between distal to internal PAS obtained from qRT-PCR analysis in tibialis anterior muscle. Averages and SD are from six mice. *p*-values for the differences between A17.1 and WT tibialis anterior muscles were assessed with the Student's *t*-test. Red outlines depict a downstream switch, and blue outlines an upstream switch

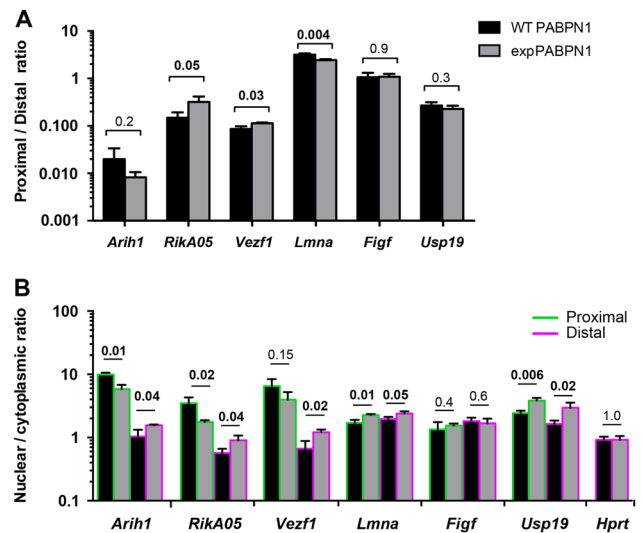


Fig. 6 The effect of expPABPN1 on PAS utilization in muscle cell culture. **a** Bar chart shows the ratio between distal and internal PAS from qRT-PCR analysis. Red outlines are depicted for a downstream switch, and blue outlines for an upstream switch and unchanged by a black line. **b** Bar chart shows changes in transcript abundance between nuclear and cytoplasmic fractions. Primer set amplifying intronic regions are depicted by green outlines and from primer set in the distal 3'-UTR by purple outlines. Averages and SD are from three biological replicates. Differences between WT-PABPN1 and expPABPN1 cultures were assessed with the Student's *t*-test. Significant *p*-values are in bold

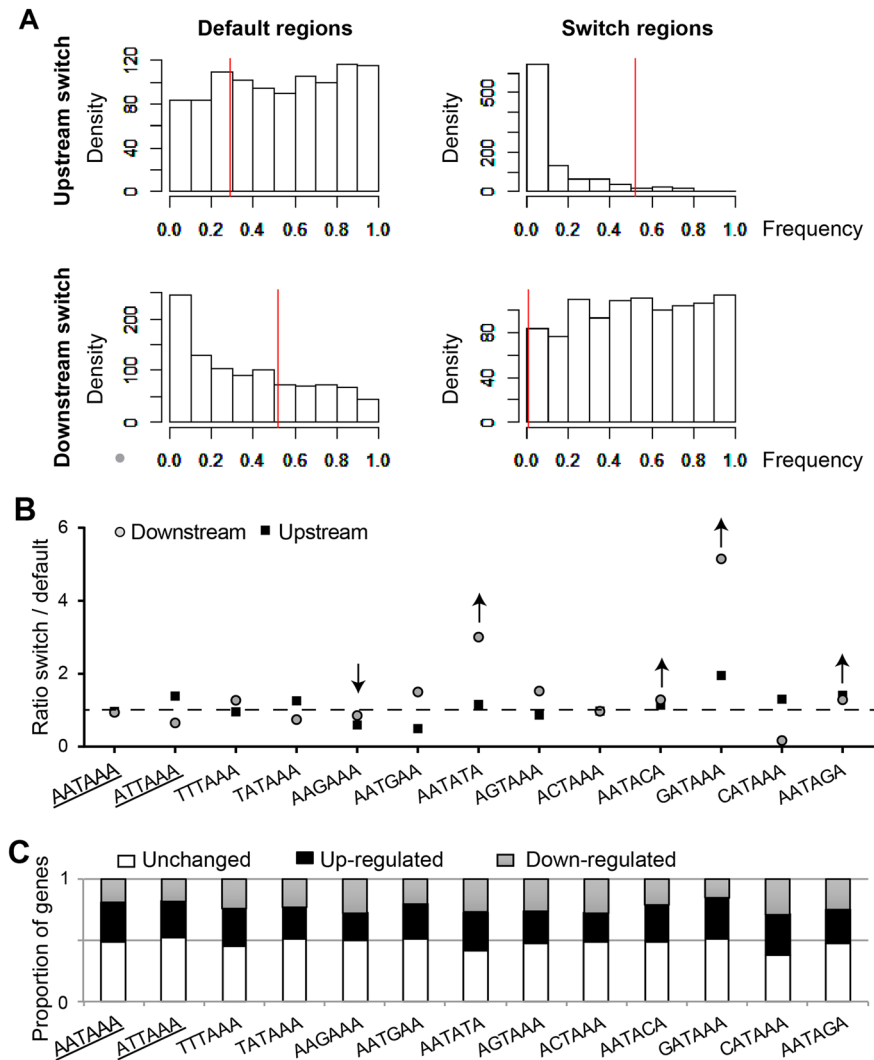


Fig. 7 Utilization of polyadenylation signal motifs in switch regions. **a** Density histograms of p -values of PAS motif distribution in switch or default PAS regions. Analysis was stratified for downstream and upstream switches. In each group motif frequency was compared to the frequency of all PASs, and statistical significance was assessed with Pearson's chi-square test. p -values obtained by comparing between motifs distribution in 1000 random subsets of PASs and motifs distribution of all PASs. The red line shows the calculated p -value for each group. **b** The ratio between switch (A17.1) and default (FVB) PAS utilization regions for 13 polyadenylation signal motifs. Numerical values are shown in Supplementary Table S4. Downstream or upstream switches are depicted by open circles or closed squares, respectively. Hashed line marks ratio = 1. Canonical motifs are underlined. Arrows mark motifs with consistent switches in A17.1 mice: up arrows represents more utilized in A17.1 and down arrow less utilized in A17.1. **c** Bar chart shows the proportion of genes with up/down or unchanged expression in each PAS motif

stable IM2 cell model.¹³ To assess whether subcellular distribution of transcripts from internal or distal 3'-UTR PAS is affected by nuclear retention of expPABPN1, the ratio between nuclear and cytosolic abundance of transcripts from distal or proximal PAS was investigated using qRT-PCR. RNA levels were normalized to *Hprt* whose levels were similar in both fractions and expression levels were unaffected by expPABPN1 expression (Fig. 6c). In 5 out of the 6 genes (*Arih1*, *Rika05*, *Vezf1*, *Lmna*, *Usp19*) the nuclear to cytoplasmic ratio was affected by expPABPN1 expression (Fig. 6c). Most consistently, the transcript from distal 3'-UTR showed a higher nuclear accumulation in expPABPN1 expressing cells compared with WT-PABPN1 cell cultures (Fig. 6c). Nuclear accumulation of transcripts from proximal PAS was reduced in expPABPN1 cell culture for *Arih1* and *Rika05* but was increased in *Lmna* and *Usp19* (Fig. 6c). High nuclear accumulation could suggest aberrant nuclear export of transcripts or higher mRNA decay in the cytoplasm. Our results suggest that nuclear accumulation of transcripts from distal PAS is associated with nuclear retention of expPABPN1. Nuclear accumulation of

transcripts from distal PAS was also confirmed in A17.1 tibialis anterior muscles (Supplementary Fig. S5).

Polyadenylation signal motifs utilization is associated with expPABPN1

Previously, we showed that in the distal 3'-UTR the non-canonical polyadenylation signals are more frequently utilized when expPABPN1 is expressed. The non-canonical polyadenylation signals are found near proximal PASs in the distal 3'-UTR, whereas the canonical polyadenylation signals (AATAAA and ATTAATA) are mostly present near distal PASs in the distal 3'-UTR.⁹ Here, we investigated whether any of the 2 canonical and 11 non-canonical motifs were associated with expPABPN1-mediated PAS switches. We counted the 13 polyadenylation signal motifs < 50 upstream of the identified PAS, and assessed their frequencies in gene regions demonstrating a PAS switch. The frequency of the motifs was investigated separately in upstream or downstream regions and default or switch regions and was compared with frequencies in

random selections from the identified PAS (Fig. 7a and Supplementary Table S4). Only the motif frequencies in PAS in downstream switch regions were found to be significantly different from the randomly selected PAS (p -value = 0.003; chi-square test; Fig. 7a).

To identify the specific polyadenylation signal motifs associated with expPABPN1 expression, we analysed the differences between the frequencies in the switch and default regions for each motif separately. In five non-canonical motifs, we found a consistent effect in both up- and downstream switches: AATATA, AATACA, GATAAA, and AATAGA motifs were more utilized in A17.1, whereas AAGAAA was less utilized in A17.1 (Fig. 7b). In six polyadenylation signal motifs switch to default PAS utilization was affected in A17.1, but this was not consistent between downstream and upstream switches (Fig. 7b). The CATAAA motif utilization was markedly decreased by downstream. PAS utilization in two motifs, including the major canonical motif AATAAA, was not affected in A17.1 (Fig. 7b). Together this suggests that few PAS motifs are associated with switch direction, and others are not.

We then also investigated whether motifs are associated with changes in expression. Gene dysregulation was found to be significantly higher across all motifs (p = 0.007; chi-square test). Higher proportion of upregulated genes was found for AATAAA, ATATAA, AATGAA, AATACA, and GATAAA motifs, whereas a higher proportion of downregulated genes was found dysregulation was found for AAGAAA and ACTAAA motifs (Fig. 7c). Overall, comparing between switch directions and dysregulation directions per motif a clear consensus was not found. This suggests that gene dysregulation is associated with PAS motifs but not with PAS switch direction.

DISCUSSION

Levels of functional PABPN1 are reduced in aged muscles and to a greater extent in OPMD muscles.¹⁰ The expanded PABPN1 is prone to aggregation leading to depletion of the soluble protein.¹³ Previous studies, including ours, demonstrated a regulatory role for PABPN1 in alternative PAS in the distal 3'-UTR.^{8, 9, 16} Although PASs are most abundant in the 3'-UTR, here we show that in skeletal muscles PAS reads in internal gene regions were found in one quarter of the genes. Most of the PASs in internal gene regions are mapped to introns. We demonstrate that expPABPN1 also alters PAS utilization between internal gene regions. This indicates that expPABPN1 effect on mRNA processing is not limited to the distal 3'-UTR, and implies that the spectrum of alternative transcripts in OPMD is broader than shortening of transcripts at the distal 3'-UTR. Our study demonstrates that expPABPN1-mediated switches in PAS utilization between gene regions affect not only expression levels in general but also the levels of transcripts variants with different open reading frames (ORFs). We suggest that PAS switches between gene regions potentially contribute to altered muscle homeostasis as a consequence of expPABPN1 expression.

Upstream PAS switches within the distal 3'-UTR result in shorter transcripts and generally upregulation of gene expression.^{8, 9} Transcripts from proximal PAS in the 3'-UTR could lose regulatory elements, like miRNA-binding sites, and lead to increase in transcript levels.^{8, 17, 18} Here, we show that downstream switches between gene regions are predominantly associated with gene upregulation. However, gene upregulation was mostly found in the downstream switches, whereas upstream switches were not associated with dysregulation direction. This suggests that PAS switches between gene regions may alter the nature of the coding sequence rather than expression levels. Since most PAS switches were found from/to introns, it is possible that those alternative transcripts could result in altered ORFs and/or non-sense transcripts.

Among the genes with PAS switch between gene regions we identified muscle genes that are associated with muscular disorders like *Dmd*, *Lmna*, *Ttn*, *Ano5*, and *Sgcb* (Supplementary Table S3). It is possible that altered transcript levels of those genes may contribute to muscle weakness in OPMD. Since PABPN1 levels are also reduced in muscles of the elderly,¹⁹ it is possible that those altered transcripts also contribute to sarcopenia. In agreement, complete knockdown of PABPN1 in cell culture also leads to internal PAS utilization.⁵ In addition to alterations in transcript levels, usage of PASs within internal gene regions could generate alternative transcripts with alternative coding sequences or be associated with lower levels of intact proteins and/or production of truncated proteins. This may also contribute to lower muscle strength in OPMD.

Increased internal PAS usage may be explained by changes in the rate of splicing, alternative or incomplete splicing (intron retention).²⁰ Alternative splicing is regulated by the spliceosome ribonucleoprotein complex.²¹ Although PABPN1 is not considered as a splicing factor, a regulatory role for PABPN1 in splicing was first suggested in non-muscle cells.²² More recently, it was demonstrated that splicing of intron 15 of Troponin T3 (*TNNT3*) is affected in OPMD muscle cells and can be regulated by PABPN1 expression levels.²³ In our study, we found an upstream switch in *Tnnt3* (Supplementary Fig. S4). This supports alternative splicing in *Tnnt3* in A17.1 quadriceps muscles. However, our PolyA-Seq data reveals internal PAS in several exons and introns in *Tnnt3*. A PAS was found in intron 15 in the A17.1, but the most significant switch was found in intron 14 (Supplementary Fig. S4). Additionally, we found intronic PAS in A17.1 of *Cln1* gene (Supplementary Fig. 4). *CLCN1* mis-splicing causes delayed muscle relaxation (myotonia) in Myotonic dystrophy 1 (ref. 24). This suggests that a switch in PAS utilization possibly cause altered polypeptides in genes regulating muscle contraction, and this can contribute to muscle weakness in OPMD.

Here, we show that transcripts from distal 3'-UTR are retained in the nuclei in expPABPN1 expressing cells, while transcripts from internal PAS have a higher cytoplasmic to nuclear ratio. Our observations are in agreement with a recent study showing nuclear retention of longer transcript isoforms, while the shorter transcript isoforms tend to be overrepresented in the cytoplasm.²⁵ Since expPABPN1 has a higher nuclear accumulation,²⁶ it is possible that the nuclear residence of expPABPN1 and expPABPN1 aggregation in the nucleus lead to nuclear retention of transcripts. Lower levels of full-length cytoplasmic transcripts could contribute to alterations in protein levels in the OPMD muscle.

In our previous study, we found that the two canonical polyadenylation signals are enriched in the distal PAS in the distal 3'-UTR, and those are less efficiently utilized in expPABPN1-expressing muscles. Instead the non-canonical motifs in tandem PAS sites in the distal 3'-UTR are utilized in OPMD conditions.⁹ Here, we extend this observation and show that a shift in PAS utilization is associated with increased utilization of non-canonical motifs, although the majority of transcripts from internal PAS are associated with canonical polyadenylation signal motifs, indicating that they represent *bona fide* PASs. The non-canonical AATATA, GATAAA, AATACA, and AATAGA motifs were more used in A17.1 mouse, whereas the non-canonical AAGAAA motif was found to be repressed in A17.1 switch regions. Those motifs were not specifically associated with a switch direction (upstream or downstream). This suggests that binding of polyadenylation factors to those motifs is increased in presence of low levels of PABPN1.

We show alterations in internal PAS utilization in OPMD mouse and cell models, and the results in muscles were partly replicated in the muscle cell culture. In muscles expPABPN1 is highly overexpressed (30-folds),¹⁴ whereas in the muscle cell culture the expPABPN1 overexpression is only 1.3-fold over endogenous PABPN1 (ref. 13). Additional differences, like different types of

stress to which the cells and the muscles are exposed, could contribute to the observed differences between the two models. These may signify a role for PABPN1 in coordinating the splicing and polyadenylation machineries, and these may add to the altered mRNA landscape in OPMD and muscle aging. In conditions with mutant PABPN1, internal gene PAS utilization affects transcript identity and its cytoplasmic abundance. The change in the transcriptome and nuclear export of transcripts could lead to widespread changes in the proteome. Alternative transcripts of contractile proteins, like *Tnnt3*, were indeed found in our switch analysis. Alteration in those proteins could contribute to muscle pathology.

MATERIAL AND METHODS

PAS single molecule sequencing method and sequencing

Helicos single-molecule PAS sequencing data were generated from quadriceps mouse muscles overexpressing the expPABPN1 (A17.1) (ref. 27) and control mice (FBV) aged 6 and 26 weeks ($N = 3$ per group) as previously described.⁹ No significant differences in PAS region usage were found between young (6 weeks) and adult (26 weeks) mice, therefore, the two age groups were combined ($N = 6$). PASs with low counts (<2 reads in <6 samples) were excluded. The genomic coordinates of mouse PASs (previously assigned and annotated as described in de Klerk, *et al.*⁹) were converted from genome assembly mm9 to mm10 using the LiftOver utility in UCSC (<http://genome.ucsc.edu>). Only 54% of the PASs were found in genes annotated in Ensembl version 84 (located in 3'-UTRs, exons, or introns).

Differential expression analysis of genes

Differential expression analysis between A17.1 and FVB control mice was carried out in R programming language (version 3.2.5). The analysis was performed on the sum of all reads mapping between the start and stop coordinates of a gene (on the same strand) using the R Bioconductor package edgeR²⁸ (version 3.14.0). A negative binomial model was fitted and a common and tag-wise dispersion was estimated for all tags prior testing procedures. p -values were computed using the exact test and adjusted for multiple testing using Benjamini and Hochberg's false discovery rate.²⁹ In our polyA-sequencing procedure, in contrast to regular RNA-sequencing, alterations in transcript length do not affect the number of counts. Gene dysregulation was determined by normalized counts per gene. Therefore, the definition of gene dysregulation (expression levels) is independent of transcript length.

Statistical model to identify genes with PAS utilization changes between gene regions

PAS counts within each gene region (intron, exon, 3'-UTR) were summed, and a logistic generalized linear mixed model was fitted to identify genes with a switch in PAS usage between different regions within a gene using the R Bioconductor package lme4 (ref. 30). The model for statistical evaluation of differential PAS utilization was described in.⁹ In brief, for each gene with multiple PAS regions, the counts n_{ij} for region j of mouse i were modeled as binomial using parameters N_i and p_{ij} . Here, $N_i = \sum_{j=1}^k n_{ij}$, k is the number of regions for a particular gene, and N_i is the total number of reads for all regions of a given gene in mouse i . The log odds of the parameter p_{ij} was modeled using fixed effects for PAS region and OPMD status, with their interaction, combined with a random intercept and PAS region effect within a mouse. We tested for the presence of an OPMD effect with a chi-squared likelihood ratio test, using as the null hypothesis the same model, but with the OPMD effect and the OPMD-PAS interaction set to zero. Significant

p -values (FDR < 0.05) indicate genes with a relative change in PAS usage between different regions in a gene.

To find the direction of the PAS switch, default and switch region were identified. After adding up the count for all PASs within a region, default and switch region were defined as regions with maximum $\sum \text{PAS}_{\text{FVB}} - \sum \text{PAS}_{\text{A17.1}}$ and minimum $\sum \text{PAS}_{\text{FVB}} - \sum \text{PAS}_{\text{A17.1}}$, respectively. With downstream switch we refer to genes where the default region is located 5' from the switch region and with upstream switch we refer to genes where the default region is located 3' from the switch region.

Cell culture, RNA isolation, and qRT-PCR

Stable cell cultures of immortalized mouse myoblasts (clone IM2) over-expressing wild-type (WT) PABPN1 or expPABPN1 are described in ref. 13. PABPN1 transgene expression is driven by the Desmin promoter, and was induced by switching proliferating cultures to fusion conditions using 4% Horse serum in Dulbecco's modified Eagle medium. Cells were seeded to 80% confluence and 2 days after incubation with 4% horse serum medium, cells were collected for RNA isolation.

Total RNA isolation from A17.1 and FBV tibialis anterior mouse muscles at 18 weeks was carried out as described in de Klerk, *et al.*⁹ cDNA was generated using oligo(dT) and random primer mix as described in de Klerk, *et al.*⁹ qRT-PCR was carried out with the LightCycler 480 (Roche) using 2× SensiMix reagent (Bioline). Primers for PAS switches validation in qRT-PCR were designed in the sequences upstream to the detected PAS (in both default and switch regions) of six candidate genes. Primers for qRT-PCR were designed using the Primer3Plus program (version: 2.4.0) (ref. 31). Primer sequences are listed in Supplementary Table S1. Relative expression levels of candidate genes were calculated after normalizing for *Hprt* housekeeping gene. Ratios between PCR products were calculated to verify a switch in PAS utilization. Fold change was calculated after normalization to *Hprt* FVB control.

Subcellular fractionation in cell culture was carried out as described in ref. 32 but RNasout® Ribonuclease Inhibitor (Promega) was added for all extraction and washing buffers. RNA was isolated from each fraction as described, using the phenol-chloroform method. Fractionation between the nuclear and cytoplasmic fractions was verified using western blotting, with Emerin (nuclear, 1:5000; anti-Rabbit EPR11071, AbCam) and Gapdh (cytosolic, 1:10,000; anti-mouse G8795, Sigma, Miss. USA). Quality of RNA from both fractions was assessed using the *Bioanalyzer* RNA 6000 Nano kit (Agilent Genomics). *Hprt*, whose expression remained unchanged in A17.1, was used for normalization.

Polyadenylation signal motifs analysis

Polyadenylation signal motif analysis was carried out for genes with a significant switch in PAS utilization. Every read in our data set marks a specific PAS, and the total number of reads associated with a specific PAS were used as input in the R Bioconductor package edgeR to identify the ones that were differentially utilized between A17.1 and WT mice (as previously described in gene expression analysis section). Low utilized PASs were defined as those PASs which were significantly less used in A17.1 compared with WT mice and high utilized PASs were those which were used significantly more in A17.1 than WT mice. Only PASs that were low utilized in default regions and high utilized in switch regions were considered in genes with upstream and downstream switches. Thirteen polyadenylation signal motifs were included in the analysis: 2 canonical and 11 non-canonical, and those motifs were counted within ~50 nt upstream of all PASs.² For 24,153 PASs (68% of all PASs) at least one of the 13 known motifs was found (39,389 motifs in total) and those were included in an analysis assessing an association with PAS switch. Frequency distribution of motifs in each group was compared with motif frequencies of all PASs using Pearson's chi-squared test. To assess whether the observed

significant differences differ from random, 1000 random subsets of PASs for each group were generated and the frequency distribution of motif in each subset was compared with motif frequencies of all PASs.

ACKNOWLEDGEMENTS

This study is supported by The Prinses Beatrix Spierfonds (grant nr.: W OR13-13).

COMPETING INTERESTS

The authors declare no competing interests.

REFERENCES

- Lee, J. Y., Ji, Z. & Tian, B. Phylogenetic analysis of mRNA polyadenylation sites reveals a role of transposable elements in evolution of the 3'-end of genes. *Nucleic Acids Res.* **36**, 5581–5590 (2008).
- Tian, B., Hu, J., Zhang, H. & Lutz, C. S. A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res.* **33**, 201–212, doi:10.1093/nar/gki158 (2005).
- Sandberg, R. et al. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science* **320**, 1643–1647, doi:10.1126/science.1155390 (2008).
- Proudfoot, N. J. Ending the message: poly(A) signals then and now. *Genes Dev.* **25**, 1770–1782, doi:10.1101/gad.17268411 (2011).
- Li, W. et al. Systematic profiling of poly(A)+transcripts modulated by core 3' end processing and splicing factors reveals regulatory rules of alternative cleavage and polyadenylation. *PLoS Genet.* **11**, e1005166, doi:10.1371/journal.pgen.1005166 (2015).
- Elkon, R., Ugalde, A. P. & Agami, R. Alternative cleavage and polyadenylation: extent, regulation and function. *Nat. Rev. Genet.* **14**, 496–506, doi:10.1038/nrg3482 (2013).
- Kühn, U. et al. Poly(A) tail length is controlled by the nuclear poly(A)-binding protein regulating the interaction between poly(A) polymerase and the cleavage and polyadenylation specificity factor. *J. Biol. Chem.* **284**, 22803–22814 (2009).
- Jenal, M. et al. The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. *Cell* **149**, 538–553, doi:10.1016/j.cell.2012.03.022 (2012).
- de Klerk, E. et al. Poly(A) binding protein nuclear 1 levels affect alternative polyadenylation. *Nucleic Acids Res.* **40**, 9089–9101, doi:10.1093/nar/gks655 (2012).
- Anvar, S. Y. et al. A decline in PABPN1 induces progressive muscle weakness in oculopharyngeal muscular dystrophy and in muscle aging. *Aging (Albany NY)* **5**, 412–426 (2013).
- Brais, B. et al. Short GCG expansions in the PABP2 gene cause oculopharyngeal muscular dystrophy. *Nat. Genet.* **18**, 164–167, doi:10.1038/ng0298-164 (1998).
- Calado, A. et al. Nuclear inclusions in oculopharyngeal muscular dystrophy consist of poly(A) binding protein 2 aggregates which sequester poly(A) RNA. *Hum. Mol. Genet.* **9**, 2321–2328 (2000).
- Raz, V. et al. Modeling oculopharyngeal muscular dystrophy in myotube cultures reveals reduced accumulation of soluble mutant PABPN1 protein. *Am. J. Pathol.* **179**, 1988–2000, doi:10.1016/j.ajpath.2011.06.044 (2011).
- Trollet, C. et al. Molecular and phenotypic characterization of a mouse model of oculopharyngeal muscular dystrophy reveals severe muscular atrophy restricted to fast glycolytic fibres. *Hum. Mol. Genet.* **19**, 2191–2207 (2010). doi:ddq098 [pii] 10.1093/hmg/ddq098.
- Abu-Baker, A. et al. Cytoplasmic targeting of mutant poly(A)-Binding protein nuclear 1 suppresses protein aggregation and toxicity in oculopharyngeal muscular dystrophy. *Traffic* **6**, 766–779, doi:10.1111/j.1600-0854.2005.00315.x (2005).
- Li, W. et al. Systematic profiling of Poly(A)+Transcripts modulated by core 3' end processing and splicing factors reveals regulatory rules of alternative cleavage and polyadenylation. *PLoS Genet.* **11**, e1005166, doi:10.1371/journal.pgen.1005166 (2015).
- Miura, P. et al. Brain-derived neurotrophic factor expression is repressed during myogenic differentiation by miR-206. *J. Neurochem.* **120**, 230–238, doi:10.1111/j.1471-4159.2011.07583.x (2012).
- Boutet, S. C. et al. Alternative polyadenylation mediates microRNA regulation of muscle stem cell function. *Cell Stem. Cell* **10**, 327–336, doi:10.1016/j.stem.2012.01.017 (2012).
- Raz, Y. & Raz, V. Oculopharyngeal muscular dystrophy as a paradigm for muscle aging. *Front. Aging Neurosci.* **6**, 317, doi:10.3389/fnagi.2014.00317 (2014).
- Braunschweig, U. et al. Dynamic integration of splicing within gene regulatory pathways. *Cell* **152**, 1252–1269, doi:10.1016/j.cell.2013.02.034 (2013).
- Wahl, M. C., Will, C. L. & Lührmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**, 701–718, doi:10.1111/j.1471-4159.2011.07583.x (2009).
- Muniz, L., Davidson, L. & West, S. Poly(A) polymerase and the nuclear poly(A) binding protein, PABPN1, coordinate the splicing and degradation of a subset of human Pre-mRNAs. *Mol. Cell. Biol.* **35**, 2218–2230, doi:10.1128/mcb.00123-15 (2015).
- Klein, P. et al. Nuclear poly(A)-binding protein aggregates misplace a pre-mRNA outside of SC35 speckle causing its abnormal splicing. *Nucleic Acids Res.* doi:10.1093/nar/gkw703 (2016).
- Oana, K. et al. Manumycin A corrects aberrant splicing of Clcn1 in myotonic dystrophy type 1 (DM1) mice. *Sci. Rep.* **3**, 2142, doi:10.1038/srep02142 (2013).
- Neve, J. et al. Subcellular RNA profiling links splicing and nuclear DICER1 to alternative cleavage and polyadenylation. *Genome Res.* **26**, 24–35, doi:10.1101/gr.193995.115 (2016).
- Uyama, E. et al. Nuclear accumulation of expanded PABP2 gene product in oculopharyngeal muscular dystrophy. *Muscle Nerve* **23**, 1549–1554 (2000).
- Davies, J. E. et al. Doxycycline attenuates and delays toxicity of the oculopharyngeal muscular dystrophy mutation in transgenic mice. *Nat. Med.* **11**, 672–677 (2005). http://www.nature.com/nm/journal/v11/n6/supinfo/nm1242_S1.html.
- Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
- Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B* **57**, 289–300 (1995).
- Douglas Bates, M. M., Bolker, Ben & Walker, Steve Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
- Untergasser, A. et al. Primer3—new capabilities and interfaces. *Nucleic Acids Res.* **40**, e115, doi:10.1093/nar/gks596 (2012).
- Dimauro, I., Pearson, T., Caporossi, D. & Jackson, M. J. A simple protocol for the subcellular fractionation of skeletal muscle cells and tissue. *BMC Res. Notes* **5**, 513, doi:10.1186/1756-0500-5-513 (2012).



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, and provide a link to the Creative Commons license. You do not have permission under this license to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2017

Supplementary Information accompanies the paper on the *npj Aging and Mechanisms of Disease* website (doi:10.1038/s41514-017-0007-x).