

# Poly(A) binding protein nuclear 1 levels affect alternative polyadenylation

Eleonora de Klerk<sup>1</sup>, Andrea Venema<sup>1</sup>, S. Yahya Anvar<sup>1</sup>, Jelle J. Goeman<sup>2</sup>, OuHua Hu<sup>1</sup>, Capucine Trollet<sup>3,4</sup>, George Dickson<sup>3</sup>, Johan T. den Dunnen<sup>1</sup>, Silvère M. van der Maarel<sup>1</sup>, Vered Raz<sup>1,\*</sup> and Peter A. C. 't Hoen<sup>1,\*</sup>

<sup>1</sup>Center for Human and Clinical Genetics, <sup>2</sup>Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, 2300 RC Leiden, The Netherlands, <sup>3</sup>School of Biological Sciences, Royal Holloway – University of London, Egham, Surrey, TW20 0EX, UK and <sup>4</sup>Thérapie des Maladies du Muscle Strié, Institut de Myologie, UM76, Université Pierre et Marie Curie, INSERM-U974, CNRS-UMR7215, Paris, France

Received March 6, 2012; Revised June 7, 2012; Accepted June 13, 2012

## ABSTRACT

The choice for a polyadenylation site determines the length of the 3'-untranslated region (3'-UTRs) of an mRNA. Inclusion or exclusion of regulatory sequences in the 3'-UTR may ultimately affect gene expression levels. Poly(A) binding protein nuclear 1 (PABPN1) is involved in polyadenylation of pre-mRNAs. An alanine repeat expansion in PABPN1 (exp-PABPN1) causes oculopharyngeal muscular dystrophy (OPMD). We hypothesized that previously observed disturbed gene expression patterns in OPMD muscles may have been the result of an effect of PABPN1 on alternative polyadenylation, influencing mRNA stability, localization and translation. A single molecule polyadenylation site sequencing method was developed to explore polyadenylation site usage on a genome-wide level in mice overexpressing exp-PABPN1. We identified 2012 transcripts with altered polyadenylation site usage. In the far majority, more proximal alternative polyadenylation sites were used, resulting in shorter 3'-UTRs. 3'-UTR shortening was generally associated with increased expression. Similar changes in polyadenylation site usage were observed after knockdown or overexpression of expanded but not wild-type PABPN1 in cultured myogenic cells. Our data indicate that PABPN1 is important for polyadenylation site selection and that reduced availability of functional PABPN1 in OPMD muscles results in use of alternative polyadenylation sites, leading to large-scale deregulation of gene expression.

## INTRODUCTION

Poly(A) binding protein nuclear 1 (PABPN1) is a ubiquitous protein involved in polyadenylation of pre-mRNAs (1–4). An expansion mutation in the polyalanine repeat in the N-terminus of PABPN1 causes oculopharyngeal muscular dystrophy (OPMD) (5). OPMD is an autosomal dominant, late-onset and progressive muscle disorder. The expanded PABPN1 (exp-PABPN1) accumulates in insoluble nuclear inclusions in affected muscles of OPMD patients (6). We have previously shown that the muscle mRNA expression profiles of OPMD patients and animal models are widely different from controls (7,8). However, it is not clear if this is directly related to the function of PABPN1 in polyadenylation.

Polyadenylation of mRNAs requires a range of multi-subunit protein complexes. The cleavage and polyadenylation specificity factor (CPSF), the cleavage stimulation factor (CstF)*n* and other proteins are involved in the endonucleolytic cleavage at the poly(A) cleavage site (polyadenylation site) preceding the addition of the poly(A) tail (9). Poly(A) polymerase (PAP), PABPN1 and CPSF are involved in the addition of the poly(A) tail itself (10,11). The assembly of the 3'-end processing machinery is directed by specific RNA sequences: the polyadenylation signal (consensus sequence AAUAAA, recognized by CPSF), the downstream sequence element (recognized by CstF (12,13)) and the upstream sequence element (14–16). So far, two major roles for PABPN1 in polyadenylation have been established. PABPN1 increases the processivity of PAP during the elongation of the tail (10,17), and it controls the length of the poly(A) tail to ~250 nucleotides (1–3). Polyadenylation at different positions in the mRNA increases the variety of transcripts (18). Polyadenylation sites within different exons or introns give rise to

\*To whom correspondence should be addressed. Tel: +31 71 5269488; Fax: +31 71 5268285; Email: v.raz@lumc.nl  
Correspondence may also be addressed to Peter A.C. 't Hoen. Tel: +31 71 5269421; Fax: +31 71 5268285; Email: p.a.c.hoen@lumc.nl

alternative 3'-terminal exons and transcripts coding for different protein isoforms (19). Polyadenylation sites located at different positions in the same 3'-untranslated region (3'-UTR) give rise to transcript variants that differ in the length of the 3'-UTR. Shortening or lengthening of the 3'-UTR may result in the loss or gain of regulatory elements, such as miRNA binding sites or binding sites for proteins that can stabilize or destabilize the transcript (20,21). This may affect mRNA stability and overall gene expression. Alternative polyadenylation is a common regulatory mechanism in various developmental and physiological processes such as the immune response (21–24), and it may also contribute to carcinogenesis (25).

To investigate the role of PABPN1 in alternative polyadenylation, we developed a single molecule sequencing approach for genome-wide detection of polyadenylation sites and studied alternative polyadenylation in A17.1 mice, which overexpress exp-PABPN1 in muscle (26). We further investigated the effects of mutation and modulation of PABPN1 expression levels on polyadenylation site selection in a myogenic cell model. We found that manipulation of PABPN1 expression levels lead to changes in polyadenylation site usage and that reduced PABPN1 levels lead to a general shortening of 3'-UTRs. We suggest an involvement of PABPN1 in polyadenylation site selection and a novel molecular mechanism for OPMD, where sequestering of exp-PABPN1 in insoluble inclusions interferes with normal polyadenylation and disrupts gene expression patterns.

## MATERIALS AND METHODS

### RNA isolation

Total RNA was extracted from quadriceps muscles of mice overexpressing the exp-PABPN1 (A17.1 mouse model) (26) and FBV mice using RNA Bee solution (Tel-Test, Bio-Connect) after homogenization of the tissue with glass beads (diameter: 1.0 mm) on the BeadBeater (BioSpec) according to the manufacturer's instructions. Quadriceps have been isolated from A17.1 and FVB mice aged 6 and 26 weeks ( $N = 3$  per group). RNA quality and concentration was determined on the Bioanalyzer (Agilent) with RNA 6000 Nano kit (RIN > 8).

### Sample preparation and polyadenylation site single molecule sequencing method

Poly(A)<sup>+</sup> RNAs were isolated from 2  $\mu$ g of total RNA using oligo(dT)<sub>25</sub> magnetic beads (Invitrogen) according to the manufacturer's instructions. First strand cDNA synthesis (SuperScript III, Invitrogen) was performed on the beads primed by oligo(dT)<sub>25</sub> following manufacturer's protocol. RNase H (Invitrogen) treatment and second strand synthesis were carried out at 16°C for 2.5 h. dsDNA was digested with NlaIII (New England Biolabs (NEB)) for 1 h at 37°C. During Poly(A)<sup>+</sup> RNAs capture, first and second strand cDNA synthesis, and dsDNA digestion, RNA and DNA molecules were washed as described in the Tag Profiling Sample Prep Kit (Illumina), using respectively GEX binding and washing

buffers, GEX cleaning solution and GEX buffer C and D. dsDNA was heat denatured at 95°C for 2 min in 100  $\mu$ l of elution buffer (10 mM Tris-Cl, pH 8.5), and the eluted second strand cDNA was precipitated with 0.1 volumes of sodium acetate (3 M, pH 4.8–5.2), 1  $\mu$ l of co-precipitant Pellet-Point (EMD4Biosciences) and 2.5 volumes of ethanol 100% overnight at ~20°C. Poly(A) tailing reaction and blocking reaction were performed using Terminal Transferase kit (NEB) with the following modifications. Poly(A) tailing reaction was carried out for 30 min at 42°C using 5 units of Terminal Transferase enzyme and 4  $\mu$ l of 50  $\mu$ M dATPs (Helicos PolyA tailing dATP) in presence of 2  $\mu$ l of 2.5 mM CoCl<sub>2</sub> and 0.2  $\mu$ l of BSA (NEB). Blocking reaction was performed with 0.5  $\mu$ l of 200  $\mu$ M biotinylated ddATP (PerkinElmer) at 37°C for 1 h, followed by 20 min at 70°C. Biotinylated ddATPs are used to measure the concentration of the samples by biotin-streptavidin approach (OptyHyb assay, Helicos). Seventy-five microliters of each sample at a concentration of 200 pM was directly hybridized to the flow cell and ssDNA was sequenced on the HeliScope platform according to the manufacturer's instructions. A total of 12 samples were sequenced in individual lanes.

### Data analysis

#### Microarray analysis

Differential expression analysis of A17.1 and wild-type mice was described previously (8). The expression profiles were generated on the Illumina Mouse Sentrix-6 v2 Beadchip platform. This platform generally contains one probe per transcript and for genes with multiple transcript variants, multiple probes may be present. Only genes with at least two probes were included. The number of genes with at least one significantly upregulated and at least one significantly downregulated probe was counted (false discovery rate (FDR) < 0.05). We evaluated two control mouse datasets profiled on the exact same Beadchip platform using the same analysis procedure (normalization, differential expression and probe annotation) (27,28).

#### Sequencing analysis

During the sequencing process on the HeliScope platform some nucleotides can remain unlabelled and appear as deletions. Therefore, standard alignment software is not suitable for Helicos data. Preprocessing of raw reads, including filtering and alignment to the mouse genome (mm9), was performed using the basic pipeline (version 1.1.498.63) from the Helicos, described at [http://open.helicosbio.com/helisphere\\_user\\_guide/ch04s07.html](http://open.helicosbio.com/helisphere_user_guide/ch04s07.html). We ran the alignment with a maximum of 100 possible mapping locations per read. A second filtering step after alignment was performed, filtering for the best location of each read and setting as threshold a minimum score of 4 and a minimum sequence length of 25 nucleotides. Downstream analysis was carried out using Custom Perl scripts to report the estimated number of tags in each region.

#### Polyadenylation site assignment and annotation

After alignment, we used the 5'-end of the reads to identify the position of the polyadenylation site and generated

wiggle files using a Custom Perl script. Polyadenylation sites within a distance of 10 nt were clustered together and assembled into regions. Clustering of regions containing polyadenylation sites (Poly(A) clusters) was carried out in repeating cycles until every region was separate by a gap of at least 10 nucleotides. Regions were annotated based on the ENSEMBL GENE 63 (Sanger, UK) database, NCBI mouse genome build 37. Further statistical analysis focused only on reads mapping to the expanded 3'-UTR, up to 2 kb downstream of the annotated 3'-UTR. To this end, we used a Custom R script to filter out reads that were not mapping within the expanded 3'-UTR.

#### **Differential expression analysis and functional annotation**

The statistical programming language R (version R 2.12.0) was used for analysis of differential expression between A17.1 and control mice. The analysis was performed using the R Bioconductor package edgeR (29) (version 2.0.4). A negative binomial model was fitted and a common dispersion was estimated for all the tags prior testing procedures. Exact *P*-values were computed using the exact test and adjusted for multiple testing according to Benjamini and Hochberg method (30). Only reads mapping in the expanded 3'-UTR of a transcript were used for expression analysis. Poly(A) clusters containing just one read were filtered out prior any statistical analysis to reduce noise. KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway analysis was performed using DAVID Functional Annotation Tool (31).

#### **Statistical model to identify transcripts with usage of alternative polyadenylation sites**

We fitted a logistic generalized linear mixed model to the counts for each polyadenylation site in a particular 3'-UTR using the R Bioconductor package lme4 (<http://cran.r-project.org/web/packages/lme4/index.html>). The counts  $n_{ij}$  for polyadenylation site  $j$  of mouse  $i$  were modeled as binomial with parameters  $N_i$  and  $p_{ij}$ . Here,  $N_i = \sum_{j=1}^k n_{ij}$ , where  $k$  is the number of polyadenylation sites for a particular 3'-UTR, and  $N_i$  is the total number of reads for all polyadenylation sites for that 3'-UTR in mouse  $i$ . The log odds of the parameter  $p_{ij}$  was modeled using fixed effects for polyadenylation site and OPMD status, with their interaction, combined with a random intercept and polyadenylation site 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-polyadenylation site interaction set to zero. Modeling the fraction for each polyadenylation site within the total counts of all polyadenylation sites of the same transcript allows assessment of changes in relative, rather than absolute frequency of the polyadenylation site.

#### **Sequence motif analysis**

We used DREME (32) to identify enriched sequence motifs. From the list of transcripts showing changes in polyadenylation site usage, we expanded the sequence of every polyadenylation site up to 50 nucleotides upstream (accounting for the genomic strand on which the

polyadenylation site was located), and grouped the sequences into two categories, one containing only the most distal (3') polyadenylation site and the other containing all the other more proximal (5') polyadenylation sites. All sequences were masked for repeats. We first ran DREME limiting the search to a maximum width of six bases for each motif, according to the length of known polyadenylation signals sequences with the full 3'-UTR sequences as background (18). Subsequently, a discriminative motif search was performed using the sequences upstream the distal polyadenylation site as negative background for the proximal polyadenylation site and vice versa.

#### **Cell culture**

Mouse myoblasts C2C12 were grown on collagen-coated plates in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum, 1% glucose and 2% glutamax (Invitrogen). Differentiation was induced by serum deprivation for 7 days, by culturing in DMEM supplemented with 2% fetal bovine serum, 1% glucose and 2% glutamax. Cells were grown under 10% CO<sub>2</sub>.

#### **Lentiviral transduction**

Overexpression of the human wild-type PABPN1 and the exp-PABPN1 in C2C12 cells was achieved by transduction of lentiviruses expressing CFP-Ala10-PABPN1 and YFP-Ala16-PABPN1, which were generated from expression vectors previously described (33). For the downregulation of the endogenous *Pabpn1*, C2C12 were transduced with Short Hairpin (shRNA)-expressing lentiviruses. The shRNA lentiviral plasmids (pLKO.1-puro) were obtained from Sigma-Aldrich (TRCN0000102536 and TRCN0000000121). Lentiviral particles were produced as previously reported (34). C2C12 myoblasts were plated in either a 6-well (100 000 cells/well, overexpression) or 12-well (35 000 cells/well, knock-down) plate, and transduction was carried out in the presence of polybrene (8 µg/ml) after an initial wash with EDTA (0.54 mM EDTA in PBS). The titers of the lentiviruses were determined by the p24 elisa kit (Retro-Tek, ZeptoMetrix Corp.). A virus titer in the range of 35 ng p24 per 10<sup>5</sup> cells was used for YFP-Ala16-PABPN1, a range of 35–70 ng p24 per 10<sup>5</sup> cells for CFP-Ala10-PABPN1, and a titer of 80 ng p24 per 10<sup>5</sup> cells was used for shRNA-TRCN0000102536 and shRNA-TRCN0000000121. Culture medium was replaced after 48 h by differentiation medium and RNA was extracted after 7 days from C2C12 myotubes transduced with CFP-Ala10-PABPN1 and YFP-Ala16-PABPN1. CMV-GFP vector was used as negative control. The levels of CFP and YFP mRNA in transduced cells were checked by quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR). C2C12 transduced with shRNA were treated with puromycin (40 ng/ml) in fresh proliferating medium after 48 h of transduction, and RNA was extracted after 24 h from myoblasts. Experiments were performed in three independent wells and repeated on different days.

### RNA isolation, reverse-transcription, qRT-PCR

RNA was extracted from C2C12 cells using the NucleoSpin RNA II kit (Macherey-Nagel) according to the manufacturer's instructions. cDNA was synthesized from 1 µg of RNA using BioScript MMLV Reverse Transcriptase (Bioline) with 40 ng of random hexamer and oligo(dT)<sub>18</sub> primers following manufacturer's instructions. qRT-PCR was performed on the LightCycler 480 (Roche) using 2X SensiMix reagent (Bioline). Each measurement was performed in duplicates or triplicates. mRNA expression levels and PCR efficiency were determined using the LinRegPCR program (35) v.11.1 according to the described method (36). Relative expression levels of *Pabpn1* were normalized to the geometric mean of two housekeeping genes, *Gapdh* and *Hprt*. Ratios between short and long variants were then calculated. Primers for qRT-PCR were designed using Primer3Plus program. Primers for poly(A) site switches validation in qRT-PCR were designed in the sequences proximal to the detected polyadenylation site of six candidate genes (*Arih1*, *Atp1b1*, *Psmid14*, *Psme3*, *Tmod1* and *Vldlr*). Primers for *Pabpn1* were designed to detect both the human exogenous and mouse endogenous mRNA. Primer sequences are listed in Supplementary Table S1.

### RNA immunoprecipitation

RNA immunoprecipitation (RIP) was performed using C2C12 myoblasts extracts. Experiments were performed in 6-well plates 90% confluent. Cells were trypsinized (0.05% Trypsin-EDTA (Invitrogen)) and cell pellets were recovered by centrifugation (10 min at 290 rcf). One-sixth of the cell pellet was used for total RNA extraction (input RNA) using NucleoSpin RNA II kit (Macherey-Nagel) according to the manufacturer's instructions. Five-sixth of the cell pellet were resuspended in 1 ml lysis buffer (100 mM KCl, 5 mM MgCl<sub>2</sub>, 10 mM HEPES (pH 7.0), 0.5% NP40, 1 mM DTT, 80 U RNase Inhibitor (Roche), Protease Inhibitor Cocktail (Roche)). The lysate was passed five times through a 29 G needle and incubated for 10 min on ice. The lysates was then clarified by centrifugation at 16 000 rcf for 5 min at 4°C and supernatant was recovered.

Protein concentration was determined using Bradford assay. Immunoprecipitation of PABPN1 was conducted with 700 µg of protein extract using the VHH-3F5 antibody (37) (overnight incubation at 4°C) and immunocomplexes were isolated with Protein A Sepharose beads (GE Healthcare) pre-coated with sperm-DNA. Following extensive washing with lysis buffer, RNA was isolated from the immunocomplexes using the NucleoSpin RNA II kit according to the manufacturer's instructions. To validate the RIP, a parallel immunoprecipitation was conducted with 150 µg of protein extract for western blot analysis. The immunocomplexes were heat denatured (95°C for 5 min) and resolved by sodium dodecyl sulphate-polyacrylamide gel electrophoresis on a 10% polyacrylamide gel followed by western blot. PABPN1 was detected with rabbit anti-PABPN1 (LSBio, 1:10 000) using goat anti-rabbit as secondary antibody (IRDye800CW, Licor, 1:10 000) and Tubulin was used as

loading control and negative control for immunoprecipitation (tubulin was detected with mouse monoclonal anti-tubulin, Sigma Aldrich, 1:2000, and goat anti-mouse IRDye680CW, Licor, 1:8000). After several washing, signals were visualized with the Odyssey Infrared Imaging System (LI-Cor Biosciences).

## RESULTS

### Microarray analysis show alternative polyadenylation events in A17.1 mice

In a previous microarray expression profile study comparing A17.1 mice with FVB parental mice (8), we noticed that probes in the same gene frequently detect discordant changes in expression. We determined that for ~8% of the genes which are represented by at least two probes on the microarray, opposite changes in expression levels were detected (Table 1). In contrast, two disease models and age-matched mouse datasets, where RNA was profiled on the same microarray platform, showed 10–20-fold fewer genes with probes showing opposite changes in expression direction (Table 1). This gave a first indication for frequent alternative polyadenylation events in A17.1 mice.

### Polyadenylation site single molecule sequencing

The Illumina microarray platform contains only a limited number of genes where alternative transcripts from the same gene are interrogated by different probes and is therefore not suited for comprehensive analysis of alternative polyadenylation.

To identify polyadenylation sites on a genome wide level in a largely unbiased way, we developed a polyadenylation site sequencing method based on the amplification-free HeliScope single molecule sequencer technology (Figure 1). To retain only the 3'-ends of the mRNAs and reduce internal priming events, poly(A)<sup>+</sup> RNAs were captured on oligo(dT) beads, followed by first and second strand cDNA synthesis on the beads and a restriction enzyme digestion. Double stranded molecules were denatured, and the poly(A) stretch downstream of the polyadenylation site made the cDNA molecules directly amenable for sequencing on the HeliScope flow cell. Sequencing started directly after the poly(A) tail, and thus at the polyadenylation site.

This method enables strand-specific mapping of alternative polyadenylation sites at single nucleotide resolution. In contrast to other recently developed methods for genome-wide assessment of alternative polyadenylation events (24,38), our method is amplification- and ligation-free, resulting in lower quantification bias.

### Alternative polyadenylation sites in mouse transcripts

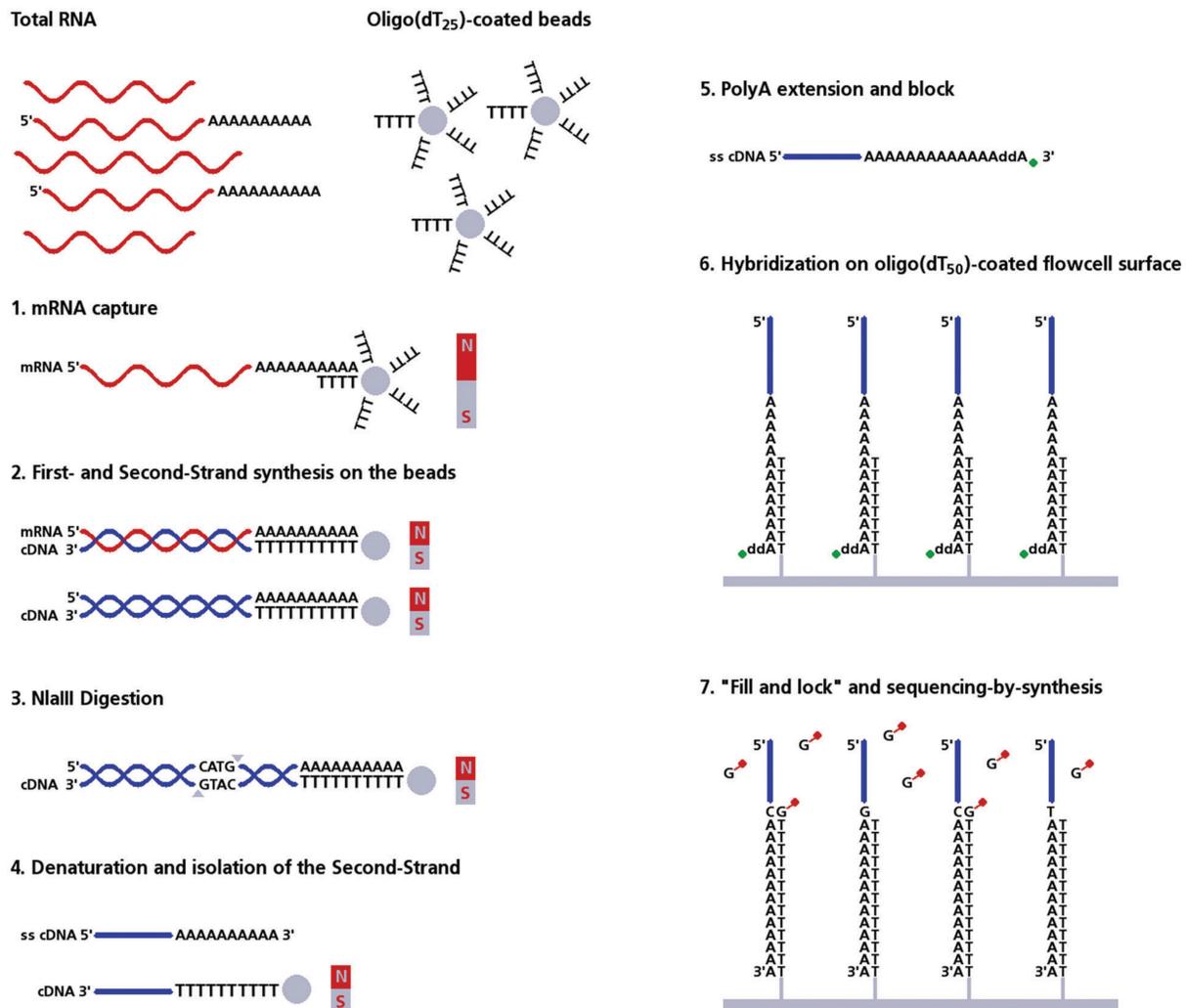
We used the single molecule sequencing method to investigate polyadenylation site usage in mouse muscles. Sequencing was performed on RNA isolated from quadriceps muscles of six individual A17.1 mice and six control mice (FVB). Mice of two different age groups were combined since there were no significant differences in

**Table 1.** Frequency of alternative polyadenylation events identified on microarrays

Mouse model	A17.1 PABPN1 overexpression	A17.1 PABPN1 overexpression	TNX <sup>a</sup> knockout	TBX3 <sup>b</sup> overexpression
Age (weeks)	6	26	32	8
Total number of deregulated genes	6012	4010	1111	1758
Number of deregulated genes with multiple probes	2819	1890	488	908
Number of deregulated genes with probes showing opposite direction in expression level	239	118	3	3
Percentage	8.48	6.24	0.61	0.33

<sup>a</sup>From (27).

<sup>b</sup>From (28).



**Figure 1.** Single molecule polyadenylation site sequencing method. The procedure of the method is detailed in the results and materials and methods sections. Red and grey boxes represent magnetic stands used to capture the oligo(dT) magnetic beads. Green anchors represent biotin labels, and red anchors represent fluorescent labels for nucleotide analogs.

polyadenylation site usage between young (6 weeks) and adult (26 weeks) mice. On average, we obtained 24 million reads per sample. An average of ~4.8 million aligned reads per sample remained after filtering for low quality reads inherent to single molecule sequencing, and after applying a stringent threshold on the confidence

with which the reads could be aligned to a position in the genome. Approximately 60% of those mapped to annotated 3'-UTRs or sequences up to 2 kb downstream of annotated 3'-UTRs (expanded 3'-UTR) and included many not yet annotated transcript ends. A summary of the sequencing results is shown in Supplementary Table S2.

To assign the location of every polyadenylation site, we considered the biological heterogeneity of the cleavage site (18,39). Therefore, we clustered together polyadenylation sites located within a window of 10 nt and continued with multiple rounds of clustering until all poly(A) clusters were separated by a gap of at least 10 nucleotides. The median width of the clusters of polyadenylation sites was 12 nucleotides, suggesting considerable variation in the exact position of the polyA site (Figure 2A). The width of the clusters is larger in case of mitochondrial RNAs, which may be polyadenylated at nearly any position in the transcript (Supplementary Figure S1) probably due to the polyadenylation-dependent degradation mechanism of mitochondrial RNAs (24,40). After removing noise by requiring a coverage of minimum two reads per polyadenylation site, we detected a total of 32 820 polyadenylation sites. 28 853 polyadenylation sites (88%) located in the expanded 3'-UTR of 11 529 different transcripts. Our analysis showed that 56% of the detected transcripts have multiple polyadenylation sites (Figure 2B).

### Widespread changes in polyadenylation site usage in A17.1 mice

A17.1 mice showed large differences in the relative polyadenylation site usage compared with control mice. An illustrative example of a transcript showing a change in polyadenylation site usage is given in Figure 3A. We detected two major polyadenylation sites in the 3'-UTR of *Psmc14*, a distal site at the annotated 3'-end of the transcript and a more proximal site. The distal and the proximal polyadenylation sites, giving rise to variants with long or short 3'-UTRs, can be observed in both FVB and A17.1 mice, but at different levels. In FVB mice, the majority of the reads mapped to the distal polyadenylation site, while in A17.1 mice, the majority mapped to the proximal polyadenylation site.

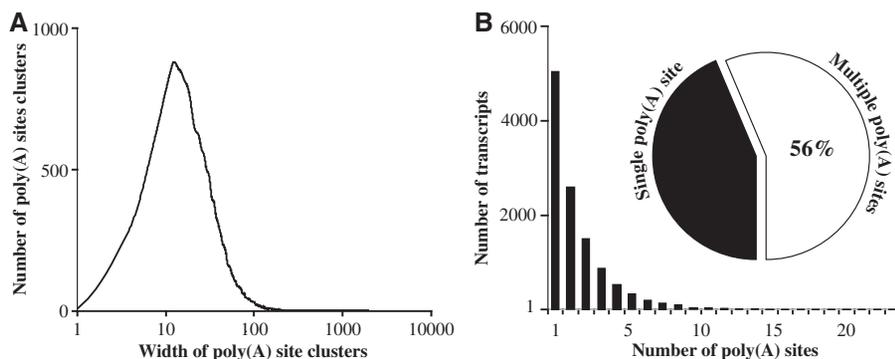
To analyse how widespread this type of switches in preferred polyadenylation site usage in A17.1 mice were, we evaluated the statistical significance of differences in the relative polyadenylation site usage using a generalized linear mixed model on the binomial distribution of the counts for each polyadenylation site in a transcript.

From 11 529 transcripts detected with our polyadenylation site sequencing method, 6506 transcripts contained at least two polyadenylation sites. Out of those, 31% showed significant differences (FDR < 0.05) in polyadenylation site usage between A17.1 and FVB mice (Supplementary Table S3). We observed a strong preference for the proximal polyadenylation site in A17.1 mice, because transcripts variants with proximal polyadenylation site were mainly upregulated and transcript variants with distal polyadenylation site were mainly downregulated in A17.1 mice (Figure 3B).

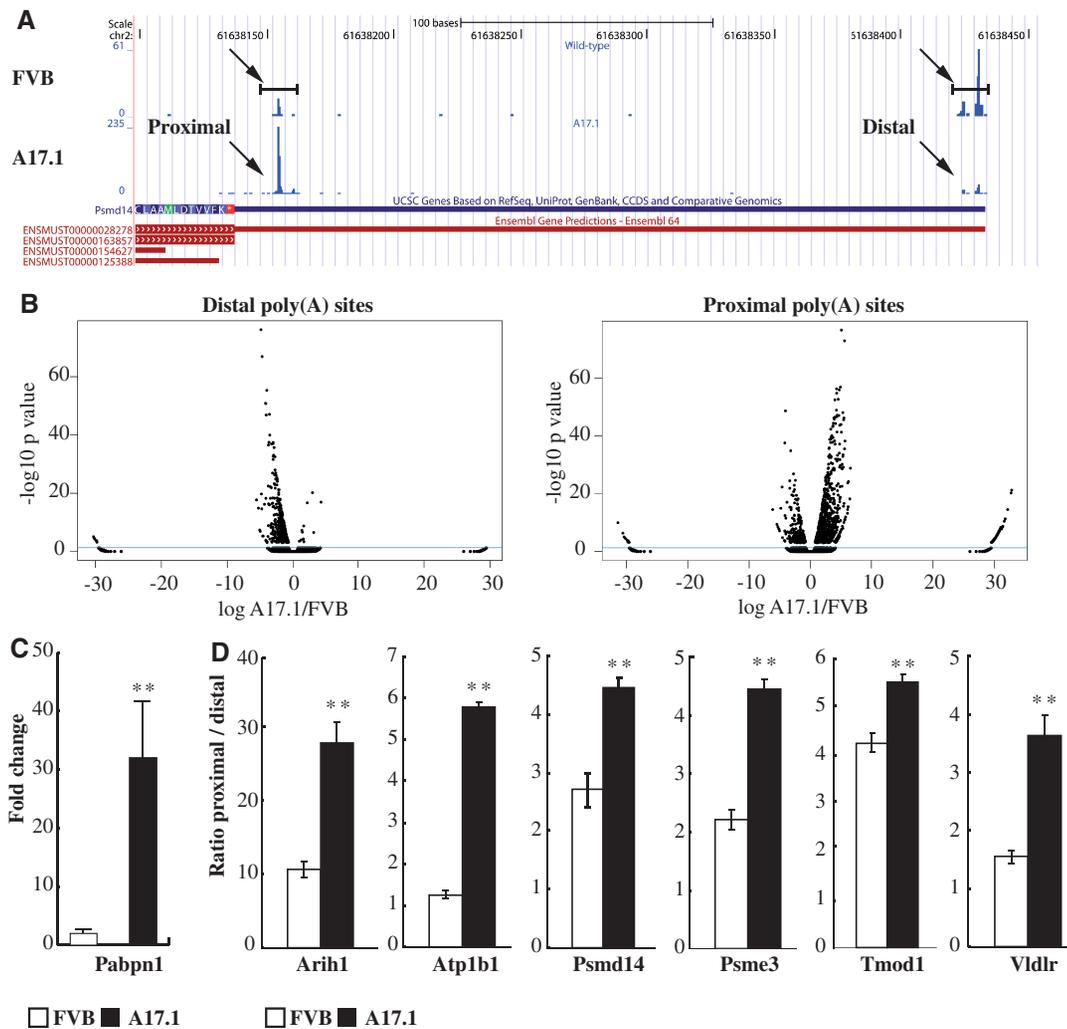
To validate the changes in relative polyadenylation site usage observed by sequencing analysis, qRT-PCR was performed using primer pairs designed just upstream of the detected polyadenylation site. We tested a panel of six genes (*Arih1*, *Atp1b1*, *Psmc14*, *Psme3*, *Tmod1* and *Vldlr*), identified from the transcriptome study in the mouse model (8) and from the cross-species study for OPMD (7). The ratio between the proximal PCR product (representing the shorter and longer isoforms) and the distal PCR product (representing only the level of the longer isoforms) was elevated for all six genes (Figure 3C and D), confirming the shortening of 3'-UTRs in A17.1 mice observed by polyadenylation site sequencing.

### Alteration of the length of the 3'-UTR results in disturbed gene expression patterns

A largely deregulated gene expression pattern in A17.1 mice has already been shown using microarray technology (8). Our polyadenylation site sequencing method can be used to identify differentially expressed transcripts, because every read uniquely identifies a transcript. To investigate the impact of changes in relative polyadenylation site usage on transcript levels, we performed a differential expression analysis using the R Bioconductor package edgeR (29) on the sum of all reads mapping to the expanded 3'-UTR of a transcript. With this procedure, we analysed the combined expression levels of short and long transcripts. From a total of 11 529 transcripts included in the analysis, 3441 were significantly deregulated (FDR < 0.05) (Supplementary Table S4).



**Figure 2.** Alternative polyadenylation analysis in mouse muscles. (A) Width of polyadenylation site clusters. The x-axis represents the width of the cluster on a logarithmic scale, the y-axis represents the number of polyadenylation site-clusters. (B) Bar graph showing the number of polyadenylation sites detected per transcript. Only polyadenylation sites mapping to the expanded 3'-UTR, and covered by at least two reads are shown. Pie chart shows the percentage of transcripts containing at least two polyadenylation sites.



**Figure 3.** Altered polyadenylation site usage in A17.1 mice. (A) A screenshot of USCS Genome Browser displaying polyadenylation sites in *Psdm14* gene. The y-axis represents the coverage of the peaks, corresponding to the number of reads mapping at each polyadenylation site. A control (FVB) and an A17.1 mouse are shown in independent traces. Below the coverage tracks, the four annotated transcripts are shown. The longest transcript (ENSMUST00000028278) contains a 3'-UTR of 297 nucleotides. There are two major polyadenylation sites in this region (indicated by arrows), a distal one (peak location chr2:61,638,431-61,638,432) at the annotated 3'-end of the transcript and a proximal one located 276 nucleotides upstream and just 23 nucleotides downstream of the stop codon (peak location chr2:61,638,155-61,638,156). (B) Volcano plots showing transcripts variants containing the distal polyadenylation site (left panel) or more proximal polyadenylation sites (right panel). The y-axis represents the  $-\log_{10}$  of the multiple testing adjusted *P*-values, while the x-axis represents the ratio of expression in A17.1 over wild-type mice on a logarithmic scale (base 2). The blue line represents a *P*-value threshold of 0.05. The outliers present in the graphs represent data points were the total counts in one of the two groups is zero. (C) *Pabpn1* expression level in quadriceps muscles of A17.1 mice and FVB mice, as measured by qRT-PCR with primers measuring both endogenous and exogenous *Pabpn1*. (D) The ratio of proximal PCR over distal PCR products in A17.1 mice (black bars) and FVB mice (white bars), as measured by qRT-PCR. Values are means  $\pm$  standard deviation for  $n = 6$  mice per group (\* $P < 0.05$ , \*\* $P < 0.01$ ).

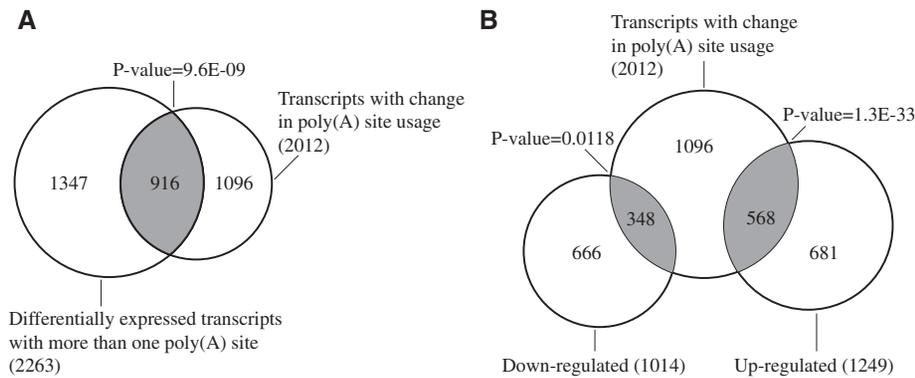
Approximately 60% of the deregulated transcripts were upregulated. The proteasome and ubiquitin-mediated proteolysis pathways were the most significantly deregulated KEGG pathways (Supplementary Table S4), confirming our previous microarray-based results (7,8,41).

To assess whether changes in polyadenylation site usage resulted in differences in transcript expression levels, we determined the overlap between deregulated transcripts in A17.1 mice and transcripts showing changes in relative polyadenylation site usage. Out of the 6506 transcripts with two or more polyadenylation sites, 2263 transcripts were differentially expressed between A17.1 and FVB mice (Supplementary Table S6), of which 1249 (55%) were

upregulated. The overlap between transcripts with differential polyadenylation site usage and those which were upregulated is highly significant (Figure 4A and B). This suggests that shortening of 3'-UTRs generally results in the loss of negative regulatory elements, such as miRNA binding sites, and higher transcript stability.

#### Alternative polyadenylation sites used in A17.1 mice contain primarily non-canonical polyadenylation signals

To obtain further insight into the mechanism leading to a preferential use of proximal polyadenylation sites in the A17.1 mice, we performed a sequence motif analysis to



**Figure 4.** Switches in polyadenylation site induce changes in gene expression. (A) Overlap between deregulated transcripts and transcripts with changes in polyadenylation site usage. (B) The overlap is shown for downregulated and upregulated transcripts separately. Indicated *P*-values were calculated with Fisher's test.

examine the sequences 50 nucleotides upstream of the distal and proximal polyadenylation sites. Most sequences with distal polyadenylation site contained one of the two canonical polyadenylation signals. The frequency of canonical polyadenylation signals in proximal sequences was lower (Table 2, 43% vs. 83%). We then performed a discriminative motif analysis using DREME (32), contrasting the motifs in sequences located upstream of the distal or the proximal polyadenylation site directly. The use of a discriminative approach enables the identification of motifs, which are enriched in only one of the two subsets of sequences. Distal polyadenylation site were enriched for the two canonical polyadenylation signals (Table 3). Proximal polyadenylation sites showed very moderate enrichment for nine hexamers, mainly GC-rich. These results indicate that the proximal polyadenylation sites preferentially used in A17.1 mice are predominantly non-canonical and do not contain a strong consensus sequence. We also performed a motif analysis on the sequences 50 nucleotides downstream of the detected polyadenylation site, but did not find enriched motifs there.

### PABPN1 levels affect 3'-end processing

To confirm that expression of the exp-PABPN1 alters polyadenylation site usage in muscle cells, we expressed the expanded human PABPN1 in C2C12 myoblasts by transduction of lentiviral particles containing YFP-Ala16-PABPN1. Myoblasts were then fused into myotubes. The overexpression level was assessed by measuring total (endogenous and exogenous) *Pabpn1* mRNA levels by qRT-PCR. *Pabpn1* was ~4-fold overexpressed (Figure 5A), much lower than in the A17.1 mice (Figure 3C). Overexpression at this level did not affect cell differentiation or fusion, as assessed by the expression levels of the myogenic markers *Myog*, *Tnnc1*, *Myh7* and *Myf5* (Supplementary Figure S2). We analysed proximal and distal polyadenylation site usage in the muscle cells with the same qRT-PCR assay as used for the A17.1 mice. The ratio between shorter and longer transcripts was significantly increased for five out of six tested genes (Figure 5B). Our *in vitro* data therefore

confirm the effect of overexpression of exp-PABPN1 on polyadenylation site usage observed in the A17.1 mouse model. The differences, however, were smaller than those found in the A17.1 mice, likely due to lower overexpression levels.

We next addressed whether overexpression of wild-type PABPN1 would also affect alternative polyadenylation. C2C12 myoblasts were transduced with lentiviral particles expressing the CFP-Ala10-PABPN1 construct. We did not succeed in obtaining 4-fold overexpression level of wild-type PABPN1. To be able to compare the effects of exp-PABPN1 and wild-type PABPN1, we overexpressed both forms 1.5-fold (Figure 5C). At this low level, overexpression of exp-PABPN1 caused significant changes in the length of 2 out of the 6 tested 3'-UTRs, whereas no significant increases were observed when transducing with wild-type PABPN1 (Figure 5D and Supplementary Figure S3A and B).

Following these findings, we asked whether reduction of *Pabpn1* expression would also affect polyadenylation site usage. PABPN1 knockdown by siRNAs is known to decrease myoblast differentiation *in vitro* (1). Thus, we downregulated endogenous *Pabpn1* in C2C12 myoblasts by lentiviral transduction with shRNAs. We used two different constructs which downregulated the expression of *Pabpn1* to 70% and 40% of control levels (Figure 5E). Interestingly, 40% downregulation of *Pabpn1* did not result in consistent alterations in polyadenylation site usage, whereas 70% downregulation resulted in shortening of the 3'-UTRs of 4 out of 6 genes tested (Figure 5F). These results suggest that polyadenylation site usage is affected by PABPN1 expression levels.

To support the notion that PABPN1 enhances the use of distal polyadenylation sites, we performed RIP experiments to investigate the binding abundance of transcripts with short and long 3'-UTRs to PABPN1. We investigated the same panel of genes previously used in Figure 5 and calculated the ratio between shorter and longer transcripts, comparing total RNA extracts and PABPN1 immunoprecipitates. Experiments were performed on C2C12 myoblasts. The proximal:distal ratio was significantly lower in PABPN1 immunoprecipitated-RNA

**Table 2.** Canonical polyadenylation signals in sequences upstream of polyadenylation sites

Number of sequences <sup>a</sup>	Total	With AATAAA	With ATTAATA
Distal	1917	1212 (63%)	397 (20%)
Proximal	6508	1855 (28%)	975 (15%)

<sup>a</sup>A DREME motif analysis was performed on sequences 50 nucleotides upstream the detected polyadenylation sites with the full 3'-UTR sequences as background.

**Table 3.** Discriminative motif analysis in sequences upstream of polyadenylation sites

Discriminative analysis <sup>(a)</sup>	Motif	E-values
Distal polyadenylation signals	AATAAA	3.60E-122
	ATTAATA	1.00E-03
Proximal polyadenylation signals	CCYTCY	7.10E-09
	CWGGYC	1.80E-06
	GARGAM	7.20E-05
	AGTGVC	1.20E-03
	GGTGMA	5.20E-03
	GCCCAC	1.10E-02
	GGGCTY	2.00E-02
	CYAGCA	3.60E-02
	GMACTA	3.70E-02

<sup>a</sup>A DREME discriminative motif analysis was performed to identify enriched motifs upstream of distal polyadenylation signals compared with proximal polyadenylation signals and enriched motifs upstream of proximal polyadenylation signals compared with distal polyadenylation signals.

compared with input RNA for five out of six genes (Figure 6A and B). This suggests that PABPN1 preferentially binds to transcripts with distal polyadenylation sites.

## DISCUSSION

Alternative polyadenylation is important to fine-tune gene expression levels, but it is currently unclear how the choice for alternative polyadenylation sites is regulated. The choice of polyadenylation site may depend on the proper orchestration of cleavage and polyadenylation factors (9,22,23,42-44), splicing (45-47) and general transcription factors (22). In this study, we provide evidence that PABPN1 regulates alternative polyadenylation in addition to its role in regulating poly(A) tail length (1-4).

To study alternative polyadenylation events on a genome wide level, we developed the polyadenylation site single molecule sequencing method. Our method is different from the recent SAPAS and PAS-Seq methods (24,38) because it is amplification and ligation free, allowing a more direct and therefore likely less biased, quantitative analysis of polyadenylation site usage. The noise in our data is low and internal priming events are rare as evident from the large majority of reads mapping to the 3'-UTR and having a canonical polyadenylation signal within a distance of 50 nucleotides. Moreover, our method has more power to detect alternative

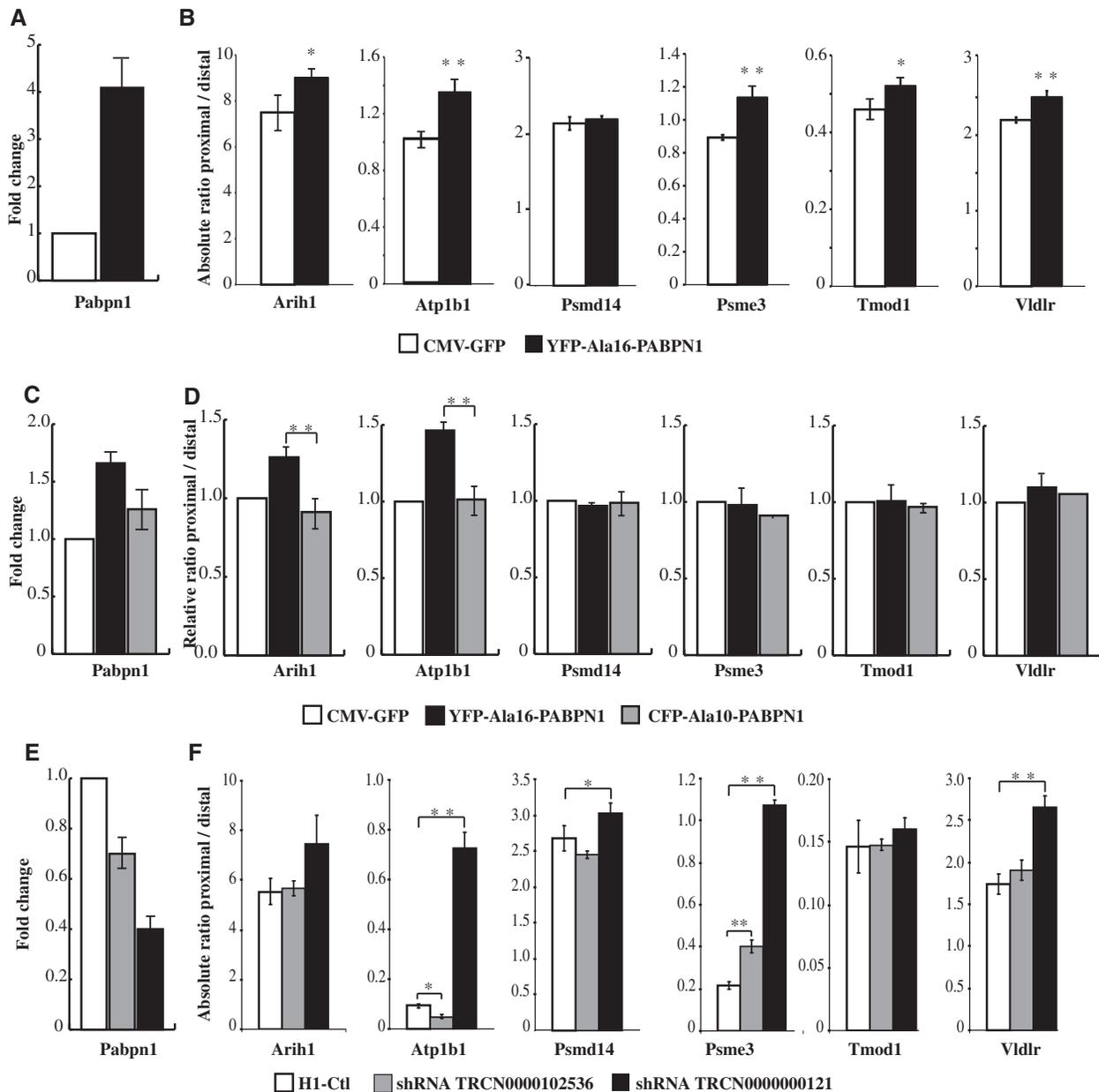
polyadenylation events than general RNA-seq technology (45), which usually comes with low coverage at the 3'-ends. A direct RNA sequencing method would also avoid a possible reverse transcription bias (48), but is not yet available for customers.

This study detected many not yet annotated polyadenylation sites and provides an extensive catalogue of alternative polyadenylation events in mouse skeletal muscle. Our results showed more widespread alternative polyadenylation in the mouse transcriptome compared to previous studies (18,24). This indicates that the number of alternative polyadenylation events identified depends very much on the technology used and suggests that alternative polyadenylation frequencies are high in most eukaryotes (49).

We investigated alternative polyadenylation events in A17.1 mice overexpressing exp-PABPN1. Our analysis showed a widespread change in relative polyadenylation site usage, with an increased use of proximal polyadenylation sites and thus a shortening of 3'-UTRs. Interestingly, we found that 45% of the transcripts showing changes in polyadenylation site usage were also deregulated. We found an increase in expression level of transcripts with shorter 3'-UTRs, which could be a result of increased stability of transcripts lacking certain miRNA binding sites or other destabilizing elements. Combined with a reduced potential for fine tuning of gene expression due to 3'-UTR shortening, this may contribute to the disturbed gene expression patterns observed in OPMD animal models and patient muscles (7,8).

Recently, Jenal *et al.* (50) published similar findings using the amplification-dependent Illumina sequencing technology. We both show an increase in proximal polyadenylation site usage after overexpression of Exp-PABPN1 or knock down of endogenous *Pabpn1*. In addition to the results presented by Jenal *et al.*, we directly compared the overexpression of wild-type and expPABPN1 in muscle cells. We demonstrated similar effects of the knockdown of endogenous *Pabpn1* and the overexpression of the expanded but not with wild-type PABPN1 in muscle cells. These results are in line with a reduction in the availability of functional PABPN1 in exp-PABPN1 expressing cells and OPMD muscle as a consequence of the higher aggregation potential of exp-PABPN1 compared with its wild-type counterpart, which will drain the nucleus from soluble PABPN1 (41). Importantly, OPMD muscles show myogenic defects (51) which are similar to knockdown of PABPN1 in mouse cells (1).

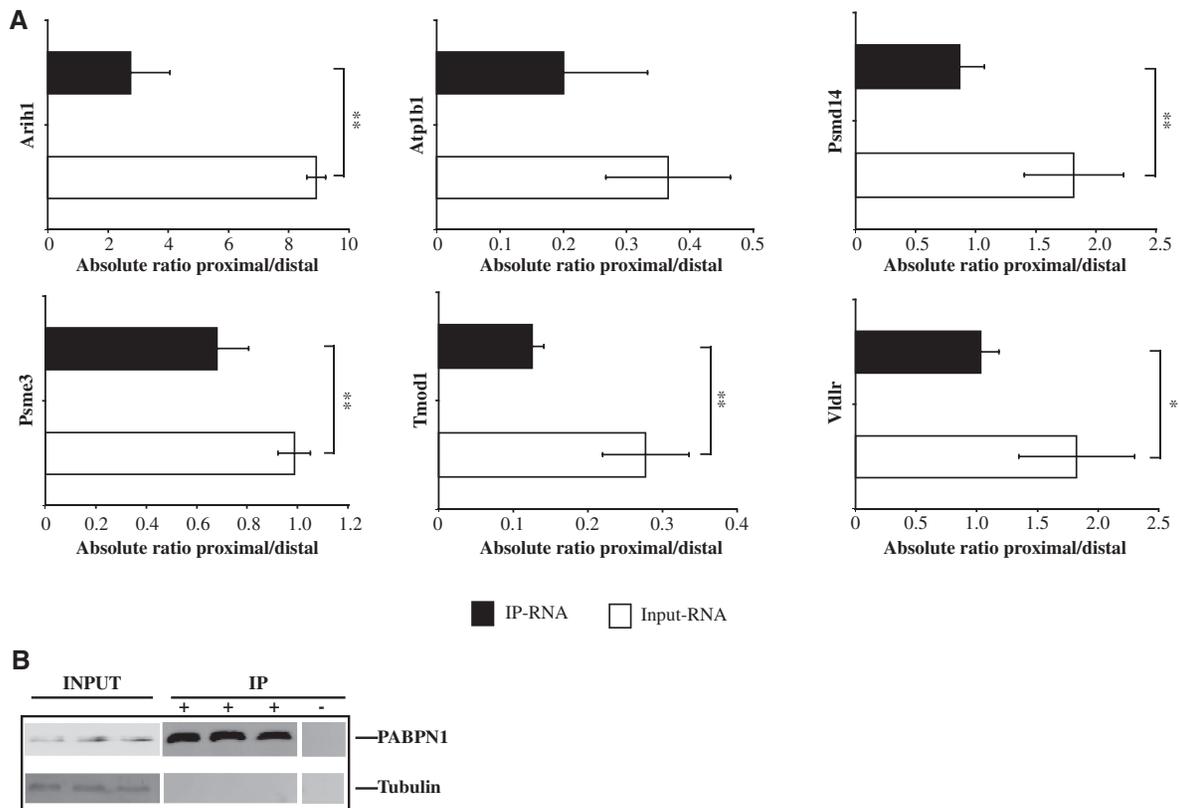
The exact molecular mechanism by which alteration in PABPN1 expression affects site selection is still not fully elucidated. Jena *et al.* excluded an effect of PABPN1 on the stability of long versus short transcripts and suggested that PABPN1 binds proximal polyadenylation signals, masking those sites and protecting them from cleavage. However, Jenal *et al.* only considered binding of PABPN1 to proximal polyadenylation sites. Our RNA-immunoprecipitation experiments provide evidences for a preferential or stronger binding of PABPN1 to transcripts with distal polyadenylation sites. This suggests that PABPN1 enhances the 3'-end processing at the



**Figure 5.** Modulation of PABPN1 expression levels induces changes in polyadenylation site usage in C2C12 cells. (A) Total (endogenous and exogenous) *Pabpn1* mRNA levels in C2C12 myotubes transduced with CMV-GFP (white) and YFP-Ala16-PABPN1 (black). (B) Ratio of proximal:distal PCR products, representing a combination of short and long 3'-UTRs respectively, in CMV-GFP (white) and YFP-Ala16-PABPN1 (black) transduced myotubes. (C) Total *Pabpn1* mRNA levels in C2C12 myotubes transduced with CMV-GFP (white), YFP-Ala16-PABPN1 (black) and CFP-Ala10-PABPN1 (grey). (D) Relative proximal:distal ratio between YFP-Ala16-PABPN1:CMV-GFP (black) and CFP-Ala10-PABPN1:CMV-GFP (grey), compared with the control CMV-GFP (white). (E) *Pabpn1* mRNA levels in C2C12 myoblasts transduced with TRCN0000102536 (grey) and TRCN0000000121 (black) shRNAs targeting *Pabpn1* and cells transduced with the control shRNA H1-Ctl (white). (F) Proximal:distal ratio for C2C12 myoblasts transduced with the two different sh-RNAs against *Pabpn1* (grey, black) and the control shRNA H1-Ctl (white). Values are means  $\pm$  standard deviation for 3 different wells. All experiments were repeated multiple times with similar results.

stronger, canonical sites but future studies with targeted mutagenesis should formally prove this. Moreover, Wahle (2) demonstrates a role for PABPN1 in conferring specificity to CPSF for canonical polyadenylation sites. Based on this, the observations in our paper, and the proven physical interaction between PABPN1 and RNA polymerase II (52), it may be postulated that PABPN1 and CPSF comigrate with the RNA polymerase II complex.

PABPN1 may then subsequently restrict the site in the RNA where CPSF dissociates from the RNA polymerase II complex to distal, canonical polyadenylation sites. Reduced soluble PABPN1 levels may alter the stoichiometry of the RNA polymerase II/CPSF/PABPN1 complex, leading to premature dissociation of CPSF at proximal, non-canonical poly(A) sites, resulting in general shortening of the 3'-UTRs of transcripts. An alternative



**Figure 6.** Transcripts with distal polyadenylation site are enriched in PABPN1 immunoprecipitated RNA. (A) The proximal:distal ratio for input RNA (white) and PABPN1-immunoprecipitated RNA (black) was determined by qRT-PCR using the same primer sets, as in Figure 5. Values are means  $\pm$  standard deviation for three independent experiments. (B) Western blot analysis shows levels of PABPN1 and tubulin, loading control, in the input. Levels of immunoprecipitated (IP) PABPN1 in C2C12 and tubulin, as negative control for immunoprecipitation, are shown. An IP without anti-PABPN1 antibody (–) was used as negative control.

mechanism that might be considered is that lower levels of PABPN1 directly or indirectly reduce the RNA Pol II-mediated transcriptional elongation rate, which has been shown to result in the preferred use of proximal poly(A) sites (53).

In any case, these alternative polyadenylation events resulting from reduced PABPN1 levels may affect mRNA stability and partly explain the observed aberrant muscle gene expression patterns and muscle weakness in OPMD animal models and patients (7).

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–6 and Supplementary Figures 1–3.

## ACKNOWLEDGEMENTS

The authors thank Henk Buermans (Leiden Genome Technology Center, The Netherlands) for the pre-processing of the sequencing data, Martijn Rabelink (Leiden University Medical Center, The Netherlands) for lentivirus production and Yavuz Ariyurek (Leiden Genome Technology Center, The Netherlands) for

critical advices on the developed poly(A) site sequencing method.

## FUNDING

The Netherlands Organisation for Scientific Research [NWO investment grant]; Center for Medical Systems Biology within the framework of the Netherlands Genomics Initiative (NGI)/NWO, the Association Française contre les Myopathies [15 123]; European Commission [TRI-EX QLG2-CT-2001-01673 and POLYALA LSHM-CT-2005-018675]; Muscular Dystrophy Association [68 015]. Funding for open access charge: Leiden University Medical Center.

*Conflict of interest statement.* None declared.

## REFERENCES

1. Apponi, L.H., Leung, S.W., Williams, K.R., Valentini, S.R., Corbett, A.H. and Pavlath, G.K. (2010) Loss of nuclear poly(A)-binding protein 1 causes defects in myogenesis and mRNA biogenesis. *Hum. Mol. Genet.*, **19**, 1058–1065.
2. Kuhn, U., Gundel, M., Knoth, A., Kerwitz, Y., Rudel, S. and Wahle, E. (2009) 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.
3. Wahle, E. (1995) Poly(A) tail length control is caused by termination of processive synthesis. *J. Biol. Chem.*, **270**, 2800–2808.
  4. Benoit, B., Mitou, G., Chartier, A., Temme, C., Zaessinger, S., Wahle, E., Busseau, I. and Simonelig, M. (2005) An essential cytoplasmic function for the nuclear poly(A) binding protein, PABP2, in poly(A) tail length control and early development in *Drosophila*. *Dev. Cell*, **9**, 511–522.
  5. Brais, B., Bouchard, J.P., Xie, Y.G., Rochefort, D.L., Chretien, N., Tome, F.M., Lafreniere, R.G., Rommens, J.M., Uyama, E., Nohira, O. *et al.* (1998) Short GCG expansions in the PABP2 gene cause oculopharyngeal muscular dystrophy. *Nat. Genet.*, **18**, 164–167.
  6. Calado, A., Tome, F.M., Brais, B., Rouleau, G.A., Kuhn, U., Wahle, E. and Carmo-Fonseca, M. (2000) 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.
  7. Anvar, S.Y., 't Hoen, P.A., Venema, A., van der Sluijs, B., van, E.B., Snoeck, M., Vissing, J., Trollet, C., Dickson, G., Chartier, A. *et al.* (2011) Deregulation of the ubiquitin-proteasome system is the predominant molecular pathology in OPMD animal models and patients. *Skelet. Muscle*, **1**, 15.
  8. Trollet, C., Anvar, S.Y., Venema, A., Hargreaves, I.P., Foster, K., Vignaud, A., Ferry, A., Negroni, E., Hourde, C., Baraibar, M.A. *et al.* (2010) 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.
  9. Shi, Y., Di Giammartino, D.C., Taylor, D., Sarkeshik, A., Rice, W.J., Yates, J.R. III, Frank, J. and Manley, J.L. (2009) Molecular architecture of the human pre-mRNA 3' processing complex. *Mol. Cell*, **33**, 365–376.
  10. Wahle, E. (1991) A novel poly(A)-binding protein acts as a specificity factor in the second phase of messenger RNA polyadenylation. *Cell*, **66**, 759–768.
  11. Wahle, E., Martin, G., Schiltz, E. and Keller, W. (1991) Isolation and expression of cDNA clones encoding mammalian poly(A) polymerase. *EMBO J.*, **10**, 4251–4257.
  12. Keller, W., Bienroth, S., Lang, K.M. and Christofori, G. (1991) Cleavage and polyadenylation factor CPF specifically interacts with the pre-mRNA 3' processing signal AAUAAA. *EMBO J.*, **10**, 4241–4249.
  13. MacDonald, C.C., Wilusz, J. and Shenk, T. (1994) The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location. *Mol. Cell Biol.*, **14**, 6647–6654.
  14. Danckwardt, S., Hentze, M.W. and Kulozik, A.E. (2008) 3' end mRNA processing: molecular mechanisms and implications for health and disease. *EMBO J.*, **27**, 482–498.
  15. Legendre, M. and Gautheret, D. (2003) Sequence determinants in human polyadenylation site selection. *BMC Genomics*, **4**, 7.
  16. Moreira, A., Takagaki, Y., Brackenridge, S., Wollerton, M., Manley, J.L. and Proudfoot, N.J. (1998) The upstream sequence element of the C2 complement poly(A) signal activates mRNA 3' end formation by two distinct mechanisms. *Genes Dev.*, **12**, 2522–2534.
  17. Kerwitz, Y., Kuhn, U., Lillie, H., Knoth, A., Scheuermann, T., Friedrich, H., Schwarz, E. and Wahle, E. (2003) Stimulation of poly(A) polymerase through a direct interaction with the nuclear poly(A) binding protein allosterically regulated by RNA. *EMBO J.*, **22**, 3705–3714.
  18. Tian, B., Hu, J., Zhang, H. and Lutz, C.S. (2005) A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res.*, **33**, 201–212.
  19. Tian, B., Pan, Z. and Lee, J.Y. (2007) Widespread mRNA polyadenylation events in introns indicate dynamic interplay between polyadenylation and splicing. *Genome Res.*, **17**, 156–165.
  20. Barreau, C., Paillard, L. and Osborne, H.B. (2005) AU-rich elements and associated factors: are there unifying principles? *Nucleic Acids Res.*, **33**, 7138–7150.
  21. Sandberg, R., Neilson, J.R., Sarma, A., Sharp, P.A. and Burge, C.B. (2008) Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science*, **320**, 1643–1647.
  22. Ji, Z., Lee, J.Y., Pan, Z., Jiang, B. and Tian, B. (2009) Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proc. Natl Acad. Sci. USA*, **106**, 7028–7033.
  23. Ji, Z. and Tian, B. (2009) Reprogramming of 3' untranslated regions of mRNAs by alternative polyadenylation in generation of pluripotent stem cells from different cell types. *PLoS One*, **4**, e8419.
  24. Shepard, P.J., Choi, E.A., Lu, J., Flanagan, L.A., Hertel, K.J. and Shi, Y. (2011) Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. *RNA*, **17**, 761–772.
  25. Mayr, C. and Bartel, D.P. (2009) Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell*, **138**, 673–684.
  26. Davies, J.E., Wang, L., Garcia-Oroz, L., Cook, L.J., Vacher, C., O'Donovan, D.G. and Rubinsztein, D.C. (2005) Doxycycline attenuates and delays toxicity of the oculopharyngeal muscular dystrophy mutation in transgenic mice. *Nat. Med.*, **11**, 672–677.
  27. Voermans, N.C., Verrijp, K., Eshuis, L., Balemans, M.C., Egging, D., Sterrenburg, E., van Rooij, J.A., van der Laak, J.A., Schalkwijk, J., van der Maarel, S.M. *et al.* (2011) Mild muscular features in tenascin-X knockout mice, a model of Ehlers-Danlos syndrome. *Connect. Tissue Res.*, **52**, 422–432.
  28. Singh, R., Hoogaars, W.M., Barnett, P., Grieskamp, T., Rana, M.S., Buermans, H., Farin, H.F., Petry, M., Heallen, T., Martin, J.F. *et al.* (2012) Tbx2 and Tbx3 induce atrioventricular myocardial development and endocardial cushion formation. *Cell Mol. Life Sci.*, **69**, 1377–1389.
  29. Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2010) EdgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.
  30. Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B (Methodol.)*, **57**, 289–300.
  31. Huang, da, W., Sherman, B.T. and Lempicki, R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.*, **4**, 44–57.
  32. Bailey, T.L. (2011) DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics*, **27**, 1653–1659.
  33. Raz, V., Abraham, T., van Zwet, E.W., Dirks, R.W., Tanke, H.J. and van der Maarel, S.M. (2011) Reversible aggregation of PABPN1 pre-inclusion structures. *Nucleus*, **2**, 208–218.
  34. Raz, V., Carlotti, F., Vermolen, B.J., van der Poel, E., Sloos, W.C., Knaan-Shanzer, S., de Vries, A.A., Hoeben, R.C., Young, I.T., Tanke, H.J. *et al.* (2006) Changes in lamina structure are followed by spatial reorganization of heterochromatic regions in caspase-8-activated human mesenchymal stem cells. *J. Cell Sci.*, **119**, 4247–4256.
  35. Ramakers, C., Ruijter, J.M., Deprez, R.H. and Moorman, A.F. (2003) Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci. Lett.*, **339**, 62–66.
  36. Ruijter, J.M., Ramakers, C., Hoogaars, W.M., Karlen, Y., Bakker, O., van den Hoff, M.J. and Moorman, A.F. (2009) Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. *Nucleic Acids Res.*, **37**, e45.
  37. Verheesen, P., de Kluijver, A., van Koningsbruggen, S., de Brij, M., de Haard, H.J., van Ommen, G.J., van der Maarel, S.M. and Verrips, C.T. (2006) Prevention of oculopharyngeal muscular dystrophy-associated aggregation of nuclear poly(A)-binding protein with a single-domain intracellular antibody. *Hum. Mol. Genet.*, **15**, 105–111.
  38. Fu, Y., Sun, Y., Li, Y., Li, J., Rao, X., Chen, C. and Xu, A. (2011) Differential genome-wide profiling of tandem 3' UTRs among human breast cancer and normal cells by high-throughput sequencing. *Genome Res.*, **21**, 741–747.
  39. Pauws, E., van Kampen, A.H., van de Graaf, S.A., de Vijlder, J.J. and Ris-Stalpers, C. (2001) Heterogeneity in polyadenylation cleavage sites in mammalian mRNA sequences: implications for SAGE analysis. *Nucleic Acids Res.*, **29**, 1690–1694.

40. Slomovic,S., Laufer,D., Geiger,D. and Schuster,G. (2005) Polyadenylation and degradation of human mitochondrial RNA: the prokaryotic past leaves its mark. *Mol. Cell. Biol.*, **25**, 6427–6435.
41. Raz,V., Routledge,S., Venema,A., Buijze,H., van der Wal,E., Anvar,S., Straasheijm,K.R., Klooster,R., Antoniou,M. and van der Maarel,S.M. (2011) Modeling oculopharyngeal muscular dystrophy in myotube cultures reveals reduced accumulation of soluble mutant PABPN1 protein. *Am. J. Pathol.*, **179**, 1988–2000.
42. Chuvpilo,S., Zimmer,M., Kerstan,A., Glockner,J., Avots,A., Escher,C., Fischer,C., Inashkina,I., Jankevics,E., Berberich-Siebelt,F. *et al.* (1999) Alternative polyadenylation events contribute to the induction of NF-ATc in effector T cells. *Immunity*, **10**, 261–269.
43. Shell,S.A., Hesse,C., Morris,S.M. Jr and Milcarek,C. (2005) Elevated levels of the 64-kDa cleavage stimulatory factor (CstF-64) in lipopolysaccharide-stimulated macrophages influence gene expression and induce alternative poly(A) site selection. *J. Biol. Chem.*, **280**, 39950–39961.
44. Takagaki,Y. and Manley,J.L. (1998) Levels of polyadenylation factor CstF-64 control IgM heavy chain mRNA accumulation and other events associated with B cell differentiation. *Mol. Cell*, **2**, 761–771.
45. Katz,Y., Wang,E.T., Airoidi,E.M. and Burge,C.B. (2010) Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods*, **7**, 1009–1015.
46. Danckwardt,S., Kaufmann,I., Gentzel,M., Foerstner,K.U., Gantzer,A.S., Gehring,N.H., Neu-Yilik,G., Bork,P., Keller,W., Wilm,M. *et al.* (2007) Splicing factors stimulate polyadenylation via USEs at non-canonical 3' end formation signals. *EMBO J.*, **26**, 2658–2669.
47. Licatalosi,D.D., Mele,A., Fak,J.J., Ule,J., Kayikci,M., Chi,S.W., Clark,T.A., Schweitzer,A.C., Blume,J.E., Wang,X. *et al.* (2008) HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature*, **456**, 464–469.
48. Ozsolak,F., Kapranov,P., Foissac,S., Kim,S.W., Fishilevich,E., Monaghan,A.P., John,B. and Milos,P.M. (2010) Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. *Cell*, **143**, 1018–1029.
49. Mangone,M., Manoharan,A.P., Thierry-Mieg,D., Thierry-Mieg,J., Han,T., Mackowiak,S.D., Mis,E., Zegar,C., Gutwein,M.R., Khivansara,V. *et al.* (2010) The landscape of *C. elegans* 3'UTRs. *Science*, **329**, 432–435.
50. Jenal,M., Elkon,R., Loayza-Puch,F., van,H.G., Kuhn,U., Menzies,F.M., Vrielink,J.A., Bos,A.J., Drost,J., Rooijers,K. *et al.* (2012) The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. *Cell*, **149**, 538–553.
51. Perie,S., Mamchaoui,K., Mouly,V., Blot,S., Bouazza,B., Thornell,L.E., St. Guily,J.L. and Butler-Browne,G. (2006) Premature proliferative arrest of cricopharyngeal myoblasts in oculo-pharyngeal muscular dystrophy: Therapeutic perspectives of autologous myoblast transplantation. *Neuromuscul. Disord.*, **16**, 770–781.
52. Bear,D.G., Fomproix,N., Soop,T., Bjorkroth,B., Masich,S. and Daneholt,B. (2003) Nuclear poly(A)-binding protein PABPN1 is associated with RNA polymerase II during transcription and accompanies the released transcript to the nuclear pore. *Exp. Cell Res.*, **286**, 332–344.
53. Pinto,P.A., Henriques,T., Freitas,M.O., Martins,T., Domingues,R.G., Wyrzykowska,P.S., Coelho,P.A., Carmo,A.M., Sunkel,C.E., Proudfoot,N.J. *et al.* (2011) RNA polymerase II kinetics in polo polyadenylation signal selection. *EMBO J.*, **30**, 2431–2444.