Distribution of alternative untranslated regions within the mRNA of the CELF1 splicing factor affects its expression

Arkadiusz Kajdasz, Daria Niewiadomska, Michal Sekrecki & Krzysztof Sobczak*

CUG-binding protein, ELAV-like Family Member 1 (CELF1) plays an important role during the development of different tissues, such as striated muscle and brain tissue. CELF1 is an RNA-binding protein that regulates RNA metabolism processes, e.g., alternative splicing, and antagonizes other RNA-binding proteins, such as Muscleblind-like proteins (MBNLs). Abnormal activity of both classes of proteins plays a crucial role in the pathogenesis of myotonic dystrophy type 1 (DM1), the most common form of muscular dystrophy in adults. In this work, we show that alternative splicing of exons forming both the 5′ and 3′ untranslated regions (UTRs) of CELF1 mRNA is efficiently regulated during development and tissue differentiation and is disrupted in skeletal muscles in the context of DM1. Alternative splicing of the CELF1 5′ UTR leads to translation of two potential protein isoforms that differ in the lengths of their N-terminal domains. We also show that the MBNL and CELF proteins regulate the distribution of mRNA splicing isoforms with different 5′ and 3′ UTRs and affect the CELF1 expression by changing its sensitivity to specific microRNAs or RNA-binding proteins. Together, our findings show the existence of different mechanisms of regulation of CELF1 expression through the distribution of various 5′ and 3′ UTR isoforms within CELF1 mRNA.

The strict control of every step of gene expression at the RNA level, such as transcription, mRNA maturation, transport, stability and translation, plays a crucial role during development, tissue differentiation and pathological processes. Many RNA-binding proteins (RBPs) regulate one or multiple of these steps through interactions with specific groups of transcripts. One such protein is CUG binding protein, ELAV-like Family member 1 (CELF1). CELF1 was isolated from HeLa cell extract as a protein that binds to CUG triplet repeats within RNA. It belongs to the family of CELF proteins, which consists of six members. CELFs are ubiquitously expressed and are important during cell cycle progression and tissue development. CELFs bind to many RNA targets and control RNA processing at different stages. Each CELF features three highly conserved RNA recognition motif (RRM) domains. Two of them are located on the N-terminus of the protein, and the third is located on the C-terminus. These domains are separated by divergent domains that are unique to each CELF member. CELF1 and CELF2 are members of one subfamily, while CELF3-6 belong to a second subfamily. Nuclear localization signals are located on the C-termini of all CELFs. CELF1-2 are common in most embryonic tissues, whereas CELF3-6 are present in adult tissues, mostly in the nervous system.

CELF1 may regulate RNA metabolism by binding to GU-rich elements (GREs) in both the nucleus and cytoplasm in three different ways: (1) through alternative splicing regulation by binding to pre-mRNAs; (2) through translation activation or inhibition by binding to the 5′ untranslated regions (5′UTRs) of many mRNAs (e.g., p21, C/EBPB, and MEF2A); and (3) through RNA decay by binding to 3′UTR sequences or mature microRNAs (miRNAs) and recruiting appropriate nucleases that deadenylate targeted transcripts. CELF1 is a critical splicing factor during muscle development. Its level is high in embryos and is significantly reduced in postnatal development, especially during the first days after birth, often without changes in CELF1 mRNA levels. The mechanism of these changes is still unknown.

CELF1 protein plays an important role in the pathogenesis of myotonic dystrophy type 1 (DM1). DM1 is an autosomal dominant multisystem disease, although the first symptoms usually involve skeletal and heart muscles. DM1 is caused by an expansion of trinucleotide CTG repeats located in the 3′ UTR of the Dystrophin gene.

Department of Gene Expression, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University Poznan, Uniwersytetu Poznanskiego 6, 61-614 Poznan, Poland. *Email: ksobczak@amu.edu.pl
Figure 1. Changes in the composition of CELF1 5’UTR and 3’UTR in DM1. (a) Distribution of CELF1 exons in DM1 based on the results of Affymetrix All-Exon Array experiments. The chart shows the hybridization signals of probe sets representing different CELF1 exons in non-DM (blue line and diamonds) and DM1 (red line and squares) skeletal muscle samples. Significant changes were observed for exons forming the 5’UTR and 3’UTRs but not for those forming the open reading frame (ORF; white background). The probes sets representing these regions are marked by colorful backgrounds and correspond to the exons represented by gray boxes in the cartoons (b and Fig. 2a). (b) The scheme of putative CELF1 5’UTR isoforms. Based on analysis of the 5’ ends of transcripts (FANTOM5 CAGE) and ENCODE Promoter-Associated Histone Mark (H3K4Me3) we distinguished two putative CELF1 transcription start sites (TSSs) in skeletal muscles (P1 and P2). CELF1 5’UTR may vary in length and exon composition due to alternative TSS usage and alternative splicing. Sequencing of RT-PCR products revealed five CELF1 5’UTR isoforms (i1–i5) transcribed from the first promoter (P1) and four isoforms (i1, i2, i3, and i5) transcribed from the second promoter (P2). Two translation start codons are marked (AUG1 and AUG2), and the values under the exon numbers (ex) indicate their lengths in nucleotides. Arrows show primer localizations. Lower boxes, UTR exons; upper boxes, ORF exons; lines, introns. (c) RT-PCR analysis of alternative splicing of CELF1 5’UTR in skeletal muscle samples of non-DM and DM1 patients (left panel: primers 5_F1 and 5_R; right panel: primers 5_F2 and 5_R). The splicing in this region of mRNAs transcribed from P1 (left panel) or P2 (right panel) is disrupted. In both cases, the shortest isoform (i1) lacking alternative cassette exons predominates (P = 0.02 for P1 and P = 0.003 for P2, t test), while isoform (i5) containing ex5 (AUG1) is less prevalent. Framed gels are presented. The unprocessed, full-sized images cannot be provided in Supplementary Information due to storage server damage. (d) Alternative splicing of CELF1 ex4 and ex5 in DM1 skeletal muscles (tibialis anterior) divided into the subgroups according to a normalized ankle dorsiflexion strength expressed in % (non-DM, n = 11; DM1 75–100, n = 7; DM1 50–75, n = 14; DM1 25–50, n = 9; DM1 0–25, n > 14) and based on RNA-Seq data. PSI, percent spliced in index. (e) Alternative splicing (3’S, 3’ splice site) and alternative polyadenylation sites (APAs) lead to the generation of at least ten potential CELF1 3’UTR isoforms (a, b, c, and d). Below the isoform structures, the miRNA target sites predicted for Homo sapiens and Mus musculus (based on TargetScan59, TarBase60, miRanda61, PicTar62, miRTarBase63 and the papers by Kalsotra et al.38 and Dong et al.41), the MBNL1 and Celf1 binding sites predicted based on crosslinking immunoprecipitation (CLIP)-Seq results, and the sequence conservation are shown. (f) RT-PCR analysis of CELF1 3’UTR isoforms distribution in DM1 skeletal muscles (left panel: primers 3_F1, 3_Ra and 3_Rcd; right panel: primers 3_F1, 3_Ra and 3_Rb). Framed gels are presented. The unprocessed, full-sized images cannot be provided in Supplementary Information due to storage server damage. (g) The expression of CELF1 is under the control of MBNLs. Knockdown of MBNL1 and MBNL2 has no effect on CELF1 mRNA level (normalized to GAPDH); although, decreases the inclusion of ex5 (real-time RT-PCR; primers 5_F3 and 5_R1, normalized to all transcripts) (siCtrl, n = 5; siMBNL1&2#1, n = 5). The distributions of CELF1 3’UTR isoforms are sensitive to MBNL protein level (primers 3_F1, 3_Ra and 3_Rcd) (n = 3); cropped gels are presented. CELF1 protein (normalized to GAPDH) is upregulated in MBNLs deficiency conditions (siCtrl, n = 6; siMBNL1&2#1, n = 6); cropped blots are presented. The efficiency of MBNL1 and MBNL2 knockdown is presented in Supplementary Fig. S2c. Statistical significance was calculated with unpaired, two-tailed t test (NS non-significant; *P < 0.05; **P < 0.01 and ***P < 0.001).

Myotonia. Protein Kinase gene (DMPK)17. Mutant mRNA containing hundreds or thousands CUG repeats is transcribed and undergoes proper maturation16. Expanded CUG repeat tract forms a stable RNA hairpin structure and sequesters Muscleblind Like proteins (MBNLs), which are regulators of alternative splicing and polyadenylation21–23. These complexes form within hyperphosphorylated by Protein Kinase C (PKC), leading to higher stability and relocation mainly to cell nuclei22. MBNL lack-of-function and CELF1 gain-of-function are the main causes of global abnormalities in alternative splicing and alternative polyadenylation22 in different tissues of DM1 patients. In some cases, these abnormalities are intensified because MBNLs and CELF1 mostly play opposing roles in splicing regulation of individual alternative exons8. In the muscles of adults with DM1, the splicing patterns of many mRNAs are similar to those observed by Protein Kinase C (PKC), leading to higher stability and relocation mainly to cell nuclei21. MBNL lack-of-function and CELF1 gain-of-function are the main causes of global abnormalities in alternative splicing and alternative polyadenylation22. CELF1 mRNA was one of the affected transcripts. The results of whole-transcriptome analysis suggested that both the 5’ and 3’ UTRs of this mRNA undergo abnormal expres-
Figure 2. Changes in the composition of CELF1 5′UTR and 3′UTR in different human and mouse tissues. (a) Alternative splicing of CELF1 ex5 in human skeletal muscles (n = 803), heart (n = 861) and brain (n = 2642) based on GTEx project RNA-Seq data. PSI, percent spliced in index, calculated according to the following equation, [exon in counts/(exon in counts + exon out counts)] × 100. (b) Splicing profiles of CELF1 5′UTR during skeletal muscle development (F; fetal, A; adult) and in adult heart and brain tissue (A; adult) (upper panel; primers 5_F1, 5_R, lower panel: primers 5_F2 and 5_R). The expression of i5-P1 (upper panel) and i5-P2 (lower panel) is the highest in adult skeletal muscles and is very low (i5-P1) or undetectable (i5-P2) in the adult brain. An unspecific RT-PCR product is marked with #; cropped gels are presented. The unprocessed, full-sized images cannot be provided in Supplementary Information due to storage server damage. (c) Upper panel, CELF1 3′UTR isoforms levels in human tissues. The screenshot from the UCSC genome browser showing GTEx project RNA-Seq data. Lower panel, CELF1 3′UTR isoforms levels in mouse tissues. The screenshot from the DM-Seq database31. (d) RT-PCR analysis of Celf1 3′UTR isoforms in mouse skeletal muscles (n = 4), heart (n = 4) and brain (n = 4) during the development in 1 and 90 days after birth (primers 3_F1, 3_Ra and 3_Rcd). Cropped gels are presented. (e) Real-time RT-PCR analysis of Celf1 3′UTR isoforms in mouse skeletal muscles (n = 4), heart (n = 4) and brain (n = 4) during the development in 1, 5, 14 and 90 days after birth. Isoforms c and d (primers 3_F1 and 3_Rcd) are downregulated in skeletal muscles, unchanged in heart and brain. Both isoforms a and b are upregulated in skeletal muscles (left panel) and downregulated during heart and brain development (middle and right panel, respectively) (primers 3_F1 and 3_Ra or 3_F1 and 3_R2a). All RT-qPCRs normalized to all Celf1 transcripts. Statistical significance was calculated with unpaired, two-tailed t test (NS non-significant; *P < 0.05; **P < 0.01 and ***P < 0.001).
appeared that the knockdown of MBNLs in HeLa cells caused significant upregulation of CELF1 (Fig. 1g, right panel and Supplementary Fig. S2d, right panel). These data suggest that MBNLs regulate the alternative splicing of the CELF1 5′UTR and 3′UTR and have an effect on the steady state level of CELF1 protein.

Previously, it has been shown that the expression of CELF1 occurs predominantly in the brain rather than in heart and skeletal muscles and that the levels of CELF1 protein are significantly decreased during muscle development.\textsuperscript{3,14,43} We decided to determine whether posttranscriptional changes in the sequence of the 5′UTR or 3′UTR of CELF1 occur during the development of mouse and human skeletal muscles, heart, and brain. It appeared that alternative splicing of the CELF1 5′UTR ex5 significantly differs in these tissues (Fig. 2a,b). The inclusion of CELF1 ex5, containing alternative AUG1 codon, is the highest in skeletal muscles, lower in heart and the lowest in the brain. The splicing profiles for the mRNAs transcribed from the two promoters in the same samples were slightly different. The longest isoforms, i5-P1 and i5-P2, predominated in skeletal muscles; their levels were much lower in the heart and brain (Fig. 2b). Therefore, it can be concluded that the 5′UTR isoform with exon 5 containing alternative AUG codon is favored in adult skeletal muscles and is reduced in fetal muscles and in DM1.

The distribution of CELF1 3′UTR isoforms also varied in different tissues (Fig. 2c,d). During mouse skeletal muscle development, the contribution of isoform a constantly increases but isoforms c and d are downregulated (Fig. 2d,e). During mouse heart and brain development, the percentage of isoform a is unchanged or significantly lowered, respectively. A similar pattern of the CELF1 3′UTR isoforms is observed in human tissues (Fig. 2c, upper panel).

We also investigated differences in the selection of three alternative polyadenylation sites (APAS1, APAS2 and APAS3) during tissue development. The percentage of transcripts terminated at APAS2 was decreased in mouse and human skeletal muscles during development, resulting in the formation of the longer CELF1 3′UTR isoform a2 (Fig. 2e, left panel). On the contrary, in mouse heart and brain, the CELF1 3′UTR isoform a2 was significantly decreased during the development (Fig. 2e, middle and right panels).

Based on obtained results, we hypothesized that due to sequence alterations, alternative 5′ and 3′UTRs of CELF1 may affect both the quantity and quality of CELF1 protein in different tissues and in DM1.

**Alternative splicing of the 5′UTR of CELF1 has no significant effect on the level and activity of CELF1.** Alternative splicing of CELF1 5′UTRs may affect protein production and selection of alternative translation start sites. To explain the functions of different CELF1 5′UTR isoforms, we formulated four hypotheses: (1) 5′UTRs affect protein production; (2) two CELF1 protein isoforms have different subcellular localization; (3) they are produced at different levels or (4) have different splicing activity. To investigate the first possibility, we designed two luciferase-expressing constructs containing the most common CELF1 5′UTRs differing in alternative exon compositions. The expression of shorter i1-P1 isoform results in slightly lower protein production than the longer i5-P1 isoform (Fig. 3a). As CELF1 protein can be translated from two start codons, AUG1 and AUG2 (Fig. 1b), we also investigated whether the activity of the two codons differed. We cloned both AUG codons, within contexts of natural flanking sequences, into the luciferase open reading frame. It appeared that both start codons had similar effects on protein production of the reporter (Fig. 3b). We also checked the effect of MBNL1 and CELF1 overexpression on tested reporter constructs as the level of these proteins are involved in muscle development and DM pathogenesis. They did not affect CELF1 protein production in context of different CELF1 5′UTR isoforms (Supplementary Fig. S3a). All these results suggest that the sequence of CELF1 5′UTR isoforms and the context of AUG codon have just marginal impact on CELF1 level.

The translation initiation codon AUG2 is present in the constitutive CELF1 ex6. Inclusion of the alternative ex5 introduces an additional initiation codon, AUG1. Therefore, inclusion of this alternative exon led to synthesis of longer CELF1 protein isoform with 27 extra amino acids at the N-terminus (Fig. 3c). Based on the above observation, we hypothesized that two CELF1 isoforms (AUG1 and AUG2), differing in the presence of an additional 27 AA sequence at N-terminus, have different subcellular localization or are produced on the different level and/or varied in the activity. To test these hypotheses, we generated constructs for overexpression of the two CELF1 variants. Both protein variants, CELF1-AUG1 and CELF1-AUG2, fused with GFP were localized mostly in cell nuclei (Fig. 3d). Then, we transfected COS-7 cells with different amounts (30 and 60 ng) of CELF1 constructs fused with small Myc-tag at C-termini to avoid the potential impact of GFP on their stability (Fig. 3e and Supplementary Fig. S3b). In cells transfected with the same amount of both constructs, the steady state level of CELF1-AUG1 and CELF1-AUG2 was the same. To further check the impact of two CELF1 isoforms on their function, we cotransfected COS-7 cells with different amounts (30–125 ng) of CELF1 expression constructs with NF1 and cTNT splicing minigenes, containing CELF1-dependent alternative exon (Fig. 3f). Previously it was shown that the CELF1 inhibits the inclusion of NF1 ex23a and promotes the inclusion of cTNT ex5 by binding to the upstream or downstream intronic regions of the alternative exons, respectively.\textsuperscript{\textsuperscript{44,45}} RT-PCR-based splicing assays showed small but statistically significant differences in splicing regulation by the CELF1 isoforms for the tested minigenes (Fig. 3f). However, observed differences are relatively small and could be caused by other experimental factors, therefore, their biological meaning is inconclusive.

We showed that alternative splicing of the CELF1 5′UTR causes the translation of two CELF1 protein isoforms differing in length by 27 amino acids at the N-terminus. However, different CELF1 5′UTRs have similar translation activity and both CELF1 proteins show no significant differences in their steady state level and activity. Therefore, the function of two CELF1 protein isoforms remains unknown.

**MiRNAs and CELFs regulate CELF1 3′UTR isoform distribution.** Sequences of 3′UTRs are very important in the regulation of mRNA stability, cellular localization and translation efficiency. Based on the analysis of RefSeq data, we could distinguish at least eleven CELF1 3′UTR isoforms that could be regulated by alter-
native splicing and alternative polyadenylation (Fig. 1e). In this study, we showed that the distribution of CELF1 3′UTR isoforms is regulated during skeletal muscle, heart and brain development and is significantly different in these tissues (Fig. 2c,d). Therefore, we hypothesized that different CELF1 3′UTR isoforms may affect the final levels of CELF1 protein. To test this hypothesis, we first determined whether the lengths and sequences of the CELF1 3′UTRs correlated with the translation efficiency in different cell lines. For this purpose, HeLa (epithelial, originate from human cervix), HepG2 (epithelial, originate from human liver) and COS-7 (fibroblasts, originate from African green monkey kidney) cells were transfected with luciferase reporter constructs containing one of three abundant CELF1 3′UTR variants: the short isoform (a1), which ended upstream of APAS2 (3′UTR-a1; approximately 500 nt); the longer isoform (a2), which ended upstream of APAS3 (3′UTR-a2; ~3000 nt); and isoform d, which ended upstream of APAS1 (3′UTR-d; ~2000 nt) (Fig. 1e). After 24 h, the activity of luciferase...
was measured. Unexpectedly, the effects of the studied 3′UTR variants were significantly different in the various cell lines, and we found no clear correlation between the lengths of CELF1 3′UTR isoforms and the protein production (Fig. 4a). The observed differences could be a result of variability in the distribution of miRNAs or RBPs in particular cell lines, which strongly affect mRNA stability and translation efficiency.

In the next step, we investigated whether miRNAs could modulate the levels of different CELF1 3′UTR isoforms. In silico prediction revealed several potential binding sites for miRNAs within the CELF1 3′UTR (Fig. 1e). Silencing of DROSHA and DGCR8, two main components of a microprocessor, caused a massive reduction in miRNAs. In cells with miRNA deficiency, the total level of CELF1 mRNA (Fig. 4b, left panel and Supplementary Fig. S4a, left panel) was not affected, but the level of isoform a2 was significantly increased, by ca. 20% (Fig. 4b, right panel and Supplementary Fig. S4a, right panel). These data suggest that longer CELF1 3′UTR isoform a2 is under negative control of miRNAs, which are important regulators of CELF1 expression depending on the distribution of alternative 3′UTRs in mRNA.

As MBNLs and CELF1 play important but opposing roles during skeletal muscle development, we decided to investigate the expression patterns of CELF1 3′UTR isoforms in cells with modulated level of MBNL (Fig. 1g) and CELF proteins (Fig. 4c and Supplementary Fig. S4c and d). Silencing and overexpression of CELF1 induced decreased and increased level of CELF1 3′UTR isoform a, respectively (Fig. 4c and Supplementary Fig. S4d). On the other hand, MBNL1 showed the opposite effect (Fig. 1g). Furthermore, overexpression of CELF2 in mouse heart increases the level of 3′UTR isoforms c and d based on RNA-Seq data31 (Fig. 4d).

Translation efficiency of CELF1 mRNA could be affected by both 5′UTR and 3′UTR. Therefore, we compared the level of endogenous Celf1 protein and Celf1 mRNA in mouse skeletal muscle, heart and brain tissue (Fig. 4c,f). The total level of Celf1 protein was much higher in brain than in striated muscles but the level of mRNA is clearly the lowest in heart in similar in skeletal muscles and brain (Fig. 4f and Supplementary Fig. S4e). These results suggest that the Celf1 protein is synthesized in all types of analyzed tissues, but its level does not depend solely on the level of mRNA. Its steady state level is the highest in tissue with the lowest contribution of mRNA isoforms containing ex5 (Fig. 2a) and 3′UTR variant a (Fig. 2b,d). The difference in Celf1 protein level in different tissues is reflected in splicing of CELF1-dependent alternative exons14. In tissues with a higher level of CELF1 protein, the CELF1-specific isoforms predominate (Fig. 4g); although, it should be underlined that many RBPs may contribute to splicing profile of analysed alternative exons.

We conclude, that the changes in the length and quality of CELF1 5′UTR and 3′UTR could simultaneously affect its translation and in consequence lead to differences in the steady state level of CELF1 in different tissues. We also showed that longer CELF1 3′UTR isoform a (a2) is under stronger pressure of miRNAs than shorter isoform a (a1). The length and sequence context of the CELF1 3′UTR depend on the activity of MBNLs and CELF1 itself, and these factors play the opposite role in this regulation. CELF activity inhibits the production of CELF1 3′UTR isoform a. Further investigations are needed to identify these miRNAs and their target sites in the CELF1 3′UTR.

Discussion

CELF1 is an RBP that plays crucial roles during development, tissue differentiation as well as pathological processes. Depending on the presence of binding sites within target RNAs, CELF1 may regulate alternative splicing, mRNA and miRNA degradation pathways or translation efficiency10,13,36. CELF1 is expressed in many tissues, especially the heart, skeletal muscles and brain. All CELF genes are highly expressed in developing fetal tissues. Interestingly, the level of CELF1 protein is significantly reduced postnatally without changes in the steady state level of CELF1 mRNA14. Changes in the activity of CELF1 protein may be regulated by transcriptional,
posttranscriptional or posttranslational mechanisms. For instance, expression of Celf1 can be controlled at the transcriptional level by activation of one of two promoters in myotubes because it is sensitive to muscle-specific drivers such as myogenin. On the other hand, translation of Celf1 could be repressed by miR-23, which partially explains the significant postnatal reduction in Celf1 levels in skeletal muscles. Regulation of the phosphorylation state of CELF1 is another posttranscriptional regulatory mechanism. Phosphorylation at the 3′ UTR of Celf1 mRNA by binding its 5′ UTR during neural development. Our data suggest that these isoforms have also slightly different splicing activity (Fig. 2f), while they are produced at the same level (Fig. 3e). It also was shown, that changes in subcellular localization of CELF1 has significant impact on muscles physiology. However, this work shows, that two Celf1 protein isoforms have mostly nuclear localization (Fig. 3d). One could also predict that these two isoforms are phosphorylated with different efficiencies or have different affinities for various RNA targets.

Our data also show that the length and sequence context of the Celf1 3′ UTR could be a result of alternative splicing and alternative polyadenylation (Figs. 1c, 2c). The distribution of different 3′ UTR isoforms strongly depends on tissue type and developmental stage. Distinct isoforms could be under the pressure of various miRNAs or RBPs whose concentrations are different in various cell types (Fig. 4). The impact of miR-23 or miR-322/503 on Celf1 expression was previously described; however, our work shows that in certain conditions, miRNA variants can gain or lose binding sites of these miRNAs. For example, in the brain, the main isoform d has neither miR-23- nor miR-322/503-sensitive sequences, but in skeletal muscles, miR-23- or miR-322/503-sensitive isoform a2 or a1, respectively, predominates (Fig. 2d,e). The impacts of other miRNAs potentially targeting Celf1 mRNA should be further examined.

Two RBP families, MBNLs and CELFs, can regulate the expression of several crucial tissue-specific factors at different stages of RNA metabolism, mostly through antagonism. It has been shown that both classes of proteins bind their mRNAs and regulate their stability. In this study, we showed that MBNLs can also control the expression of Celf1 by regulating the distribution of its 5′ UTR and 3′ UTR isoforms (Fig. 1g and Supplementary Fig. S2). Previous reports showed that knockdown of only one paralog of Mbnl genes is insufficient to increase Celf1 level. Mice with depletion of Mbnl1 or Mbnl2 manifest some DM-like phenotypes without significant effect on skeletal muscle weakness and heart arrhythmias. Lack of some DM-like symptoms and no change in Celf1 level in Mbnl1 knockout mouse could be partially explained by the elevated Mbnl2 level in this mouse model. While, a model with a decreased level of both Mbnl1 and Mbnl2 developed severe DM-like muscle defects and reflect much stronger splicing changes, similar to those observed in DM patients, suggesting a crucial, cooperative role of insufficiency of both Mbnl1 and Mbnl2 in the DM pathomechanism. We showed, that disruption of alternative Celf1 ex5 splicing and 3′ UTR isoforms in DM1 muscles (Fig. 1) caused by functional inaccessibility of Mbnl1 has a significant impact on Celf1 protein level in the affected tissue. Our findings suggest that an increased level of Celf1 protein in DM1 muscles could be caused not only by hyperphosphorylation of the protein but also by regulation of an alternative isoforms of Celf1 UTRs which favor the efficient production of a Celf1 protein.

RBP autoregulation through binding to its RNA is a common mechanism. More recently we presented, that autoregulation of Mbnl1 is driven by skipping of ex1 in Mbnl1 pre-mRNA, giving rise to a non-functional protein. In this work, we showed that Celf1 can also regulate its expression by modulating its 3′ UTR isoform distribution (Fig. 4c). These processes are also regulated by the other family member, CELF2 (Fig. 4d). Both proteins promote the formation of isoforms with proximal 3′ UTR sequences with different sensitivities to miRNAs and other RBPs.

In summary, changes in Celf1 protein levels without changes in Celf1 mRNA levels have been well documented thus far; however, in this work, we showed that both the length and sequence context of Celf1 5′ UTR and 3′ UTR change during ontogenesis, yielding numerous mRNA isoforms with different regulatory potential. Variation in the 5′ UTRs directly affects the level of two Celf1 protein isoforms, and they have slightly different splicing activity. Furthermore, changes in the 3′ UTR sequence may affect the quantity of Celf1 by altering the regulatory pressure of specific miRNAs and RBPs. Therefore, posttranscriptional modifications of Celf1 UTR sequences contribute to Celf1 activity under physiological and pathological conditions such as DM1, and modulation of alternative splicing of Celf1 can be considered as a potential therapeutic strategy.

Materials and methods
cDNA from human tissues. cDNA from non-DM and DM1 skeletal muscle samples were generously donated by Dr Charles Thornton (University of Rochester, NY, USA). cDNA panels from adult tissues (Human MTC Panel I, cat. no. 636742) and fetal tissues (Human Fetal MTC Panel, cat. no. 636747) were purchased from Clontech. According to the company description tissue samples were pooled from the following numbers of individuals: fetal skeletal muscle 13 spontaneously aborted male/female Caucasian fetuses, skeletal muscles 4
male/female Caucasians (ages 25–56), heart 3 male Caucasians (ages 24–41) and brain 8 male Caucasians (ages 43–65). Individual cDNA concentration was normalized using the following genes: alpha-tubulin, beta-actin, G3PDH and phospholipase A2.

RNA and proteins from mouse tissues. Skeletal muscles (tibialis anterior), hearts and brains were extracted from female C57BL/6J mice in 1, 5, 14 and 90 days after the birth. Mice were euthanized by carbon dioxide inhalation followed by cervical dislocation. All animals were held in the animal facility of the Center of Advanced Technologies, Adam Mickiewicz University, Poznan, Poland (Breeders Register no. 062). The study was approved by the National Ethics Committee for Animal Testing. The study was carried out in compliance with the ARRIVE guidelines. All tissues were obtained according to the National Ethics Committee for Animal Testing and Polish law guidelines and regulations.

Plasmids. Genetic constructs for expression of CELF1-AUG1 and CELF1-AUG2 were cloned by insertion of CELF1 sequences into a pEGFP-C1 backbone (Clontech). Both inserts, which were based on murine Celf1 open reading frames, were cloned between the Xhol and BamHI restriction sites. To avoid the potential impact of GFP on the CELF1 stability or activity we prepared constructs of CELF1 isoforms fused with the Myc-tag on their C-termini. For this purpose, pcDNA3.1 plasmid (Invitrogen) and inserts were amplified with primers pcDNA_CELF1_F and pcDNA_CELF1_R or CELF1_AUG1, CELF1_AUG2 and CELF1_R, respectively using CloneAmp HiFi PCR Premix (TakaRa Bio; 639298) and ligated with NEBuilder HiFi DNA Assembly Cloning Kit (New England Biolabs; E5520). To measure transfection and transcription efficiency we cloned a cassette containing an internal ribosome entry site (IRES) and GFP open reading frame between CELF1 open reading frame and pcDNA3.1 polyadenylation signal. This cassette was amplified from MIGR1 plasmid49 with IRES_F and IRES_R primers using CloneAmp HiFi PCR Premix (TakaRa Bio; 639298). pcDNA3.1 plasmid containing CELF1 open reading frames were amplified with pcDNA_IRES_F and pcDNA_IRES_R primers and ligated with IRES-GFP cassette using NEBuilder HiFi DNA Assembly Cloning Kit (New England Biolabs; E5520). The plasmid for MBNL1 overexpression has been described previously37. Plasmids for testing of the protein production of luciferase reporters with different CELF1 5′UTRs were generated based on a pmirGLO Dual-Luciferase miRNA Target Expression Vector (Promega, cat. no. E1330). New multiple cloning sites (with NdeI, SmaI/Xmal, and EcoRI restriction sites) were cloned between the PGK promoter and the luciferase open reading frame in the HindIII restriction site. This plasmid was named 5MCS-pmirGLO. CELF1 5′UTR isoforms sequences were amplified based on cDNA derived from human skeletal muscle, liver and brain tissues (Human MTC Panel I, Clontech, cat no. 636742). CELF1 5′UTR isoforms were cloned between the NdeI and EcoRI restriction sites. To analyze how the context of the AUG codon impacts CELF1 protein production, two plasmids were created. Two DNA fragments containing 94 bp with an AUG1 or an AUG2 codon were amplified and cloned upstream of luciferase 2 into the 5MCS-pmirGLO plasmid. Both inserts were cloned between the NdeI and EcoRI restriction sites. Plasmids used to study the translation efficiency of different CELF1 3′UTRs were constructed on the backbone of the original pmirGLO Dual-Luciferase miRNA Target Expression Vector. The sequences of CELF1 3′UTR isoforms a1 (473 bp) and d1 (1863 bp) were amplified from cDNA derived from human adult skeletal muscle, and liver tissues (Human MTC Panel I, Clontech, cat. no. 636742), whereas isoform a2 (2935 bp) was amplified from genomic DNA (isolated from HeLa cells; ATCC, CCL-2). Isoforms a1 and d1 were cloned between the Nhel and Xhol and between the Nhel and XbaI restriction sites, respectively. Due to the length of the a2 isoform, it was amplified in two parts. First, a shorter part (DNA fragment 1; 1196 bp) was amplified and ligated (at the Pmel and Nhel restriction sites) into the pmirGLO vector, then a longer part (DNA fragment 2; 1685 bp) was amplified and ligated (at the Nhel and Xhol restriction sites) downstream of the first part. NFl and cTNT splicing minigenes have been described previously34,35. CELF1 binds with the sequence of 87 nucleotides of the intronic region upstream to NFl alternative exon 23a or the sequence between 17 and 46 nucleotides of the intronic region downstream to cTNT alternative exon 5.

Cell cultures and transfection. HeLa (ATCC, CCL-2) and COS-7 (ATCC, CRL-1651) cells were cultured in DMEM (Lonza) supplemented with 10% fetal bovine serum (FBS, Sigma) and 1% antibiotic–antimycotic (Sigma). HepG2 (ATCC, HB-8065) cells were cultured in EMEM (Lonza) supplemented with 10% FBS (Sigma), 1% MEM Non-Essential Amino Acid Solution (Sigma) and 1% antibiotic–antimycotic (Sigma). The HSKM (Gibco, A11440) cell line was cultured in HAM-F10 medium (Sigma) supplemented with 20% FBS, 1% antibiotic–antimycotic (Sigma), 10 ng/mL Epidermal Growth Factor (EGF, Sigma), 0.4 µg/mL dexamethasone (Sigma) and 25 µg/mL insulin. The cell lines were cultured in an incubator in an atmosphere with proper humidity, 37 °C and 5% CO2. Subsets of HSKM and HepG2 cells were transfected with 50 nM control siRNA (siCtrl) duplexes (Future Synthesis). Other HSkM, HeLa and HepG2 cells were cotransfected with 25 nM MBLN1 siRNA (Future Synthesis) and 25 nM MBLN2 siRNA (RiboTask) (siMBLN1&2#1 and siMBLN1&2#2) or with 25 nM DROSHA siRNA and 25 nM DGCR8 siRNA (siDROSHA and siDGCR8) duplexes. HepG2 cells were transfected with 50 nM CELF1 siRNA (siCELF1#1 and siCELF1#2) duplex (Future Synthesis). All siRNA duplexes were delivered into cells using Lipofectamine 2000 or Lipofectamine 3000 (Invitrogen) according to the manufacturer’s protocol. The sequences of all siRNAs are shown in Supplementary Table S2.

All plasmids were delivered into cells with FuGENE HP (Roche) at a ratio of 1 µg of plasmid and 2 µl of FuGENE HP per 2 ml of culture medium. To assess the splicing activity of the CELF1 isoforms, COS-7 cells were seeded onto 12-well plates and cotransfected with 100 ng of Neuroglobin 1 (NF1) and cardiac troponin T (cTNT) minigenes and various amounts of CELF1-AUG1 or CELF1-AUG2 plasmids (30, 60 or 125 ng). To analyze the impact of MBNL1 or CELF1 on the distribution of CELF1 UTR isoforms, HeLa and HepG2 cells were transfected with 1 µg of MBNL1 or CELF1-AUG1 plasmids, respectively, in 6-well plates. For all luciferase
Luciferase activity assay. All measurements of luciferase activity were carried out after 24 h of transfection with a Dual-Luciferase Reporter Assay System (Promega, cat. no. E1910) according to the manufacturer's manual with Infinity 200 PRO (Tecan) equipment. These plasmids have a renilla luciferase open reading frame under a control of different promoter than luc2, which is used as transfection and loading control. Luc2 activity is shown as a ratio of signal from Firefly and Renilla.

RNA isolation and RT-PCR. Total RNA was isolated from the tissues or cell lines using TRI Reagent (Sigma) according to the manufacturer's protocol. 0.5–2 µg of total RNA was used in a reverse transcription reaction with a GoScript Reverse Transcription System (Promega) and random hexamers (Promega) or oligo-dT primers (Promega). All RT-PCRs were performed using GoTaq Flexi DNA Polymerase (Promega). The PCR products were separated on agarose gels with ethidium bromide (Sigma). The gels were visualized on G:BOX (Syngene) equipment, and the DNA band intensities were measured using Gene-Tools software (Syngene). For gels presented in Figs. 1c,f and 2b the unprocessed, full-sized images cannot be provided due to storage server damage. It was not possible to repeat these reactions due to limited amounts of tissue-derived cDNA; therefore, specificity of primers used in RT-PCR assays were additionally tested for commercially available RNA from human skeletal muscles what is shown in Supplementary Fig. S6. Nevertheless, specificity of all PCR products shown in Figs. 1c,f and 2b were validated by Sanger-sequencing. Real-time PCR was performed using SYBR Green PCR Master Mix (Thermo Fisher Scientific) and a CFX96 Touch Real-Time PCR Detection System (Bio-Rad). CELF1 expression levels were normalized to the levels of ACTB or GAPDH as a control. Expression levels of CELF1 5′UTR or 3′UTR exons were normalized to the levels of total CELF1. Fold changes in expression were calculated using the 2^ΔΔCt method. Localizations of all primers are presented in Fig. 1b, Supplementary Fig. S1.

Western blot analysis. Proteins were isolated using RIPA buffer (150 mM NaCl, 50 mM Tris–HCl pH 8.0, 1 mM EDTA, 0.5% NP-40, 0.5% Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS) supplemented with SIGMAFAST Protease Inhibitor Cocktail (Sigma). The lysates were sonicated at 4 °C and centrifuged at 18,000g for 10 min at 4 °C. The concentration of protein in mouse samples was measured using the Pierce BCA Protein Assay Kit (Thermo, cat. no. 23225) and 10 µg of protein were loaded per lane. Before electrophoresis, the samples were denatured at 95 °C for 5 min. The proteins were separated by electrophoresis in a 10% denaturing SDS-polyacrylamide gel and transferred onto a nitrocellulose membrane (Protran BA 85, Whatman) by wet electrophoretic transfer (1 h, 100 V, 4 °C). The membrane was blocked with 5% nonfat milk dissolved in PBS-T buffer (PBS, 0.1% Tween-20) for 1 h. PageRuler Prestained Protein Ladder (ThermoFisher, cat. no. 26617) was used as a molecular mass marker. Incubation with a primary antibody (Supplementary Table S3) was performed at 4 °C overnight. The membrane was washed three times with PBS-T buffer for 10 min. Incubation with a secondary antibody horseradish peroxidase conjugate was performed for 2 h at room temperature. The membrane was washed three times with PBS-T buffer for 10 min. For signal detection, Pierce ECL Plus Western Blotting Substrate (Thermo Scientific) was used. The western blots were visualized on G:BOX (Syngene) equipment, and the protein band intensities were measured using Gene-Tools software (Syngene). Total protein was stained after electrophoresis with SimplyBlue SafeStain (Invitrogen) and visualized on G:BOX (Syngene) equipment.

RNA-Seq data. Values of percentage splice junction index (PSI) of CELF1 ex4 and ex5 in DM1 skeletal muscles and heart were published previously. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from: the GTEx Portal on 06/01/2020 and/or dbGaP accession number phs000424.v8.p2 on 06/01/2020. Counts of exon-exon junctions in human tissues were provided by GTEx project (file named GTEx_Analysis_2017-06-05_v8_STARv2.5.3a_junctions.gct.gz). Based on these data PSI was calculated according to the following equation, \((\text{CELF1 ex5–ex6 junction counts}/(\text{CELF1 ex5–ex6 junction} + \text{CELF1 ex1–ex6 junction})) \times 100\). RNA-Seq data sets of human cell lines with MBNL1 or MBNL2 knock down and mouse heart with overexpression of Celf2 were published previously. Raw reads were aligned to the Ensembl human (GRCh38) or mouse (GRCm38) genomes with STAR (v2.7). Aligned reads were counted with featureCounts (v2.0.1). Alternative splicing events in human cell lines were calculated with RMASTS (v4.1). Level of Celf1 3′UTR isoform in a mouse heart after Celf2 overexpression were calculated as percentage of reads located in this isoform in comparison to all reads located in Celf1 3′UTR.

Statistical analysis. Statistical analysis was performed by an unpaired, two-tailed t test using Microsoft Excel (NS non-significant; \(P<0.05\); \(**P<0.01\) and \(***P<0.001\)). Statistical analysis was calculated using the mean from 3–6 biological replicates (\(n\) ± standard deviation (SD)). Violin plots were generated with R (v.4) and ggplot2 package. Exact \(P\)-values are listed in Supplementary Table S4.
38. Kalotra, A., Wang, K., Li, P. F. & Cooper, T. A. MicroRNAs coordinate an alternative splicing network during mouse postnatal heart development. *Genes Dev.* 24, 653–658. https://doi.org/10.1101/gad.194310 (2010).

39. Salisbury, E. et al. Ectopic expression of cyclin D3 corrects differentiation of DM1 myoblasts through activation of RNA CUG-binding protein, CUGBP1. *Exp. Cell Res.* 314, 2266–2278. https://doi.org/10.1016/j.yexcr.2008.04.018 (2008).

40. Cox, D. C., Guan, X., Xia, Z. & Cooper, T. A. Increased nuclear but not cytoplasmic activities of CELF1 protein leads to muscle wasting. *Hum. Mol. Genet.* 29, 1729–1744. https://doi.org/10.1093/hmg/ddaa095 (2020).

41. Dong, W. et al. miR-322/miR-503 clusters regulate defective myoblast differentiation in dystrophic mice by targeting Celf1. *Toxicol. Res. (Camb)* 10, 29–39. https://doi.org/10.1016/t-toxres.20095 (2021).

42. Wang, E. T. et al. Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. *Cell* 150, 710–724. https://doi.org/10.1016/j.cell.2012.06.041 (2012).

43. Masuda, A. et al. CUGBP1 and MBNL1 preferentially bind to 3' UTRs and facilitate mRNA decay. *Sci. Rep.* 2, 209. https://doi.org/10.1038/scr00209 (2012).

44. Kanadia, R. N. et al. A muscleblind knockout model for myotonic dystrophy. *Science* 302, 1978–1980. https://doi.org/10.1126/science.1088583 (2003).

45. Charizaris, K. et al. Muscleblind-like 2-mediated alternative splicing in the developing brain and dysregulation in myotonic dystrophy. *Neuron* 75, 437–450. https://doi.org/10.1016/j.neuron.2012.05.029 (2012).

46. Lee, K. Y. et al. Compound loss of muscleblind-like function in myotonic dystrophy. *EMBO Mol. Med.* 5, 1887–1900. https://doi.org/10.1002/emmm.201303275 (2013).

47. Konieczny, P., Stepniak-Konieczna, E., Taylor, K., Szajdjer, L. J. & Sobczak, K. Autoregulation of MBNL1 function by exon 1 exclusion from MBNL1 transcript. *Nucleic Acids Res.* 45, 1760–1773. https://doi.org/10.1093/nar/gkw1158 (2017).

48. Wojciechowska, M. et al. Quantitative methods to monitor RNA biomarkers in myotonic dystrophy. *Sci. Rep.* 8, 5885. https://doi.org/10.1038/s41598-018-24156-y (2018).

49. Pear, W. S. et al. Efficient and rapid induction of a chronic myelogenous leukemia-like myeloproliferative disease in mice receiving P210 bcr/abl-transduced bone marrow. *Blood* 92, 3780–3792 (1998).

50. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(−Delta Delta C(T)) method. *Methods* 25, 402–408. https://doi.org/10.1016/S1046-2024(01)00035-8 (2001).

51. Seachrist, D. D., Sznajder, L. J. & Ferrando, A. Quantitative methods to monitor RNA biomarkers in myotonic dystrophy. *Sci. Rep.* 3, 1639–1650. https://doi.org/10.1038/srep00318 (2013).

52. Dobin, A. et al. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–25. https://doi.org/10.1093/bioinformatics/bts635 (2013).

53. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. https://doi.org/10.1093/bioinformatics/btt656 (2014).

54. Shen, S. et al. rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci. USA.* 111, E5593–E5601. https://doi.org/10.1073/pnas.1419161111 (2014).

55. Forrest, A. R. et al. A promoter-level mammalian expression atlas. *Nature* 507, 462–470. https://doi.org/10.1038/nature13182 (2014).

56. Agarwal, V., Bell, G. W., Nam, J. W. & Bartel, D. P. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* 4, e05005. https://doi.org/10.7554/eLife.05005 (2015).

57. Karagkouni, D. et al. DIANA-TarBase v8: A decade-long collection of experimentally supported miRNA-gene interactions. *Nucleic Acids Res.* 46, D239–D245. https://doi.org/10.1093/nar/gkx1141 (2018).

58. John, B. et al. Human microRNA targets. *PLoS Biol.* 2, e363. https://doi.org/10.1371/journal.pbio.0020363 (2004).

59. Krek, A. et al. Combinatorial microRNA target predictions. *Nat. Genet.* 37, 495–500. https://doi.org/10.1038/ng1536 (2005).

60. Chou, C. H. et al. rMATBaseline update 2018: A resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res.* 46, D296–D302. https://doi.org/10.1093/nar/gkj067 (2018).

61. Sznajder, L. et al. Mechanistic determinants of MBNL activity. *Nucleic Acids Res.* 44, 10326–10342. https://doi.org/10.1093/nar/gkw915 (2016).

**Acknowledgements**

We thank Dr. Charles Thornton for sharing non-DM and DM1 skeletal muscle biopsies and Agnieszka Mykowska, Patryk Konieczny and Lukasz Szajdjer for sharing some biological materials used in this study. This work was supported by the Foundation for Polish Science-TEAM program cofinanced by the European Union within the European Regional Development Fund (TEAM POIR.04.04.00-00-5C0C/17-00 to KS), the National Science Centre of Poland (2020/38/A/NZ3/00498 to KS) and the Ministry of Science and Higher Education of the Republic of Poland, from the quality-promoting subsidy, under the Leading National Research Centre (KNOW) program for the years 2012–2017 (KNOW RNA Research Centre in Poznan; 01/KNOW2/2014). The authors also received financial support from the Initiative of Excellence–Research University (05/IDUB/2019/94) at Adam Mickiewicz University, Poznan, Poland.

**Author contributions**

A.K., K.S., Conceptualization; A.K., formal analysis; K.S., funding acquisition; A.K., D.N., M.S., investigation; A.K., K.S., methodology; A.K., D.N., M.S., visualization; A.K., D.N., M.S., K.S., review and editing; A.K., K.S., writing the paper; K.S., supervision. All authors reviewed the manuscript.

**Competing interests**

The authors declare no competing interests.

**Additional information**

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-03901-9.

Correspondence and requests for materials should be addressed to K.S.
