Human UFSP1 translated from an upstream near-cognate initiation codon functions as an active UFM1-specific protease

Received for publication, December 22, 2021, and in revised form, April 24, 2022 Published, Papers in Press, May 5, 2022,
https://doi.org/10.1016/j.jbc.2022.102016

Qian Liang, Yaqi Jin, Shiwen Xu, Junzhi Zhou, Jian Mao, Xiaohe Ma, Miao Wang, and Yu-Sheng Cong*

From the Key Laboratory of Aging and Cancer Biology of Zhejiang Province, School of Basic Medical Sciences, Hangzhou Normal University, Hangzhou, China

Edited by George DeMartino

Ubiquitin-fold modifier 1 (UFM1) is a recently identified ubiquitin-like posttranslational modification with important biological functions. However, the regulatory mechanisms governing UFM1 modification of target proteins (UFMylation) and the cellular processes controlled by UFMylation remain largely unknown. It has been previously shown that a UFM1-specific protease (UFSP2) mediates the maturation of the UFM1 precursor and drives the de-UFMylation reaction. Furthermore, it has long been thought that UFSP1, an ortholog of UFSP2, is inactive in many organisms, including human, because it lacks an apparent protease domain when translated from the canonical start codon (^{445}AUG). Here, we demonstrate using the combination of site-directed mutagenesis, CRISPR/Cas9–mediated genome editing, and mass spectrometry approaches that translation of human UFSP1 initiates from an upstream near-cognate codon, ^{217}CUG, *via* eukaryotic translation initiation factor eIF2A-mediated translational initiation rather than from the annotated ^{445}AUG, revealing the presence of a catalytic protease domain containing a Cys active site. Moreover, we show that both UFSP1 and UFSP2 mediate maturation of UFM1 and de-UFMylation of target proteins. This study demonstrates that human UFSP1 functions as an active UFM1-specific protease, thus contributing to our understanding of the UFMylation/de-UFMylation process.

The ubiquitin-fold modifier 1 (UFM1) system is a recently identified ubiquitin-like posttranslational modification with essential biological functions (1). Deficiency of this modification leads to embryonic lethality in mice and diseases in humans (2). UFM1 is present in nearly all eukaryotic organisms (except fungi) with a similar tertiary structure to ubiquitin. Similar to ubiquitination, the covalent conjugation of UFM1 (UFMylation) to target proteins involves a three-step enzymatic cascade catalyzed sequentially by UFM1-activating enzyme 5 (UBA5, E1), UFM1-conjugating enzyme 1 (UFC1, E2), and UFM1-specific ligase 1 (UFL1, E3) (2, 3). The UFMylation process is highly conserved in metazoans and plants, implicating its specific roles in multicellular organisms. In human, gene mutations in UFMylation components (UFM1, UFC1, UBA5, DDRGK1, or UFM1-specific protease [UFSP] 2) have been found to be associated with a variety of neurological disorders and skeletal abnormalities (4–7). Accumulating evidence suggests that UFMylation plays a critical role in diverse cellular processes, including erythrocyte differentiation during embryogenesis (8–10), endoplasmic reticulum (ER) homeostasis (11, 12), translational homeostasis (13, 14), DNA damage response, and cancer-related signaling pathways (15–18). So far, only one of the E1, E2, and E3 enzymes each of the UFMylation system have been identified, and a handful of substrates have been reported. The regulatory mechanisms governing UFM1 modification of target proteins and the cellular processes controlled by UFMylation remain largely unknown.

UFMylation is a reversible process because of UFSPs mediated de-UFMylation reaction (19). Although two UFSP genes (*UFSP1* and *UFSP2*) are present in the human genome, it has long been believed that human UFSP1 is inactive or nonfunctional because it lacks a specific protease domain as translated from the annotated ^{445}AUG (2, 17). Therefore, UFSP2 has been regarded as the only active protease that mediates UFM1 precursor (pro-UFM1) maturation and de-UFMylation in human cells. Intriguingly, several studies have shown that KO of *UFSP2* resulted in significantly increased protein UFMylation in human cells, indicating that other active UFSPs mediate pro-UFM1 maturation in human cells (13, 20).

Given that only *UFSP1* and *UFSP2* are present in the human genome, we are naturally concerned about the true identity of UFSP1 in the UFMylation/de-UFMylation process. It has been known for decades that translation can start from codons other than AUG, usually from near-cognate initiation codons, which have a sequence that differs from the AUG codon by one nucleotide (for example, CUG, GUG, and UUG) (21–23). Through sequence alignment, we identified a potential coding region of cysteine protease catalytic domain upstream of the canonical ORF in human UFSP1 gene, which may be derived from near-cognate CUG codons. In this study, we demonstrated that expression of human UFSP1 is initiated through eIF2A-mediated translational initiation, and human UFSP1 is a functional UFSP with distinguishing feature in protein UFMylation/de-UFMylation.

* For correspondence: Yu-Sheng Cong, yscong@hznu.edu.cn.

Results

Human UFSP1 is an active UFSP

To explore the possibility that the near-cognate start codon makes human UFSP1 acquired functional protease activity, we constructed two UFSP1 expression plasmids, one containing the canonical coding region (UFSP1-Short) and the other containing the full length of the complementary DNA (cDNA) with the 5′ untranslated region (5′ UTR) (UFSP1-Long). The canonical coding region of human UFSP1 is expected to generate the peptide with the molecular weight (MW) approximately 17 kDa. Intriguingly, we only detected a ∼23 kDa specific protein band in UFSP1-L construct transfected cells but no expected size of ∼17 kDa protein band was detected in UFSP1-S construct–transfected cells (Fig. 1*A*). Similarly, we constructed GFP-tag fused UFSP1-S or UFSP1-L vectors and only detected a prominent ∼50 kDa fusion protein band (GFP-tag, ∼27 kDa) in UFSP1-L-GFP construct–transfected cells (Fig. S1*A*). Meanwhile, the levels of UFMylation were significantly decreased by the expression of UFSP1-L or UFSP1-L-GFP, but not by that of UFSP1-S or UFSP1-S-GFP (Figs. 1*A* and S1*A*), and this decreased UFMylation associated with UFSP1-L expression is fully abrogated by the potential active site cysteine (^{349}TGC) to alanine (GCC) mutation (Fig. 1*B*). These results suggest that an upstream near-cognate codon, other than the annotated ^{445}AUG codon, was used for human UFSP1 translation initiation and for the production of an active UFSP.

During pro-UFM1 maturation, the C-terminal Ser-Cys dipeptide of pro-UFM1 (∼9.1 kDa) is cleaved by the UFSPs to expose its C-terminal conserved Gly residue (24). Matured UFM1 (∼8.9 kDa) is required for conjugation to its target proteins. To further confirm the functionality of human UFSP1 in pro-UFM1 maturation, we transfected a C-terminal hemagglutinin (HA)-tagged pro-UFM1 plasmid (pSG5-UFM1-HA) or control vector into HEK293T cells depleted *UFSP1*, *UFSP2*, or both *UFSP1* and *UFSP2*, respectively. We found that the patterns of protein UFMylation were significantly changed with distinctive features in *UFSP1*, *UFSP2*, or double KO cells (Figs. 1*C* and S1*B*). *UFSP2* KO resulted in considerably increased levels of protein UFMylation in cells (Fig. 1*C*, lines 3 and 7). In addition, the levels of protein UFMylation in *UFSP1* KO cells were slightly increased or comparable to those in the parental cells (Fig. 1*C*, lines 2 and 6). However, UFMylation was completely blocked in *UFSP1*/*UFSP2* double KO cells (Fig. 1*C*, lines 4 and 8). These results further demonstrate that human UFSP1 is an active UFSP with distinguished specificity in protein UFMylation/de-UFMylation; and UFSP1 and UFSP2 are the only UFSPs in human cells.

Human UFSP1 is translationally initiated from ^{217}CUG codon

In eukaryotic cells, CUG initiation codon is the most common among the various near-cognate codons identified (23, 25). We evaluated the 5′ UTR sequence of human UFSP1 mRNA for translation start site and found a total of seven CUG codons in-frame coding region containing the upstream catalytic domain. According to the MW, we have narrowed down the list to four candidate CUG codons, ^{190}CUG, ^{217}CUG, ^{226}CUG, and ^{247}CUG (Fig. 2*A*). To identify the true initiation site, we created a set of mutations on each of these four CUG codons (as CTG in plasmids) individually to determine which is required for human UFSP1 expression. Mutations at ^{217}CTG fully abrogate UFSP1 expression, whereas mutations at

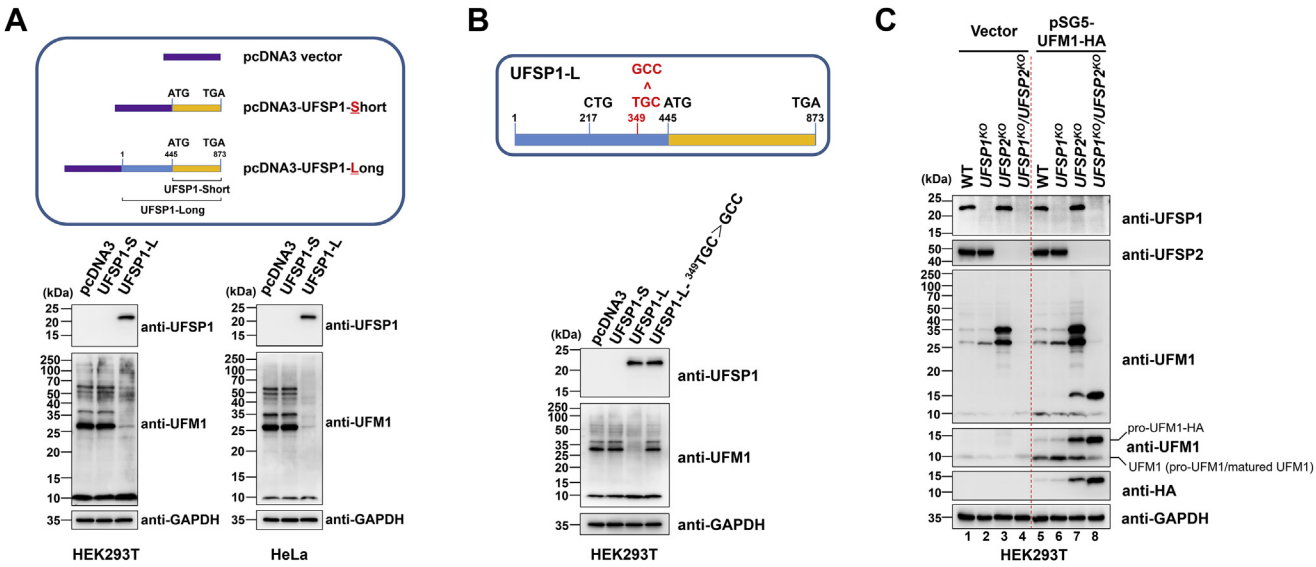


Figure 1. Human UFSP1 is an active UFM1-specific protease. *A*, de-UFMylation activity of the human UFSP1 in HEK293T cells (*left*) and HeLa cells (*right*). *B*, mutation of ^{349}TGC in human *UFSP1* gene fully abrogates its de-UFMylation function. Human UFSP1-L expression plasmid with ^{349}TGC to GCC mutation, UFSP1-L, UFSP1-S expression plasmid, and the control vector were introduced into HEK293T cells. *C*, human *UFSP1* and/or *UFSP2* KO changed the patterns of protein UFMylation and pro-UFM1 maturation. Human UFSP1 can cleave the pro-UFM1-HA (∼11.8 kDa) to produce the mature form UFM1 (∼8.9 kDa). *UFSP1* KO (*UFSP1*^{KO}), *UFSP2* KO (*UFSP2*^{KO}), and *UFSP1/UFSP2* double KO (*UFSP1*^{KO}/*UFSP2*^{KO}) cell lines were generated by CRISPR/Cas9–mediated genome editing and monoclonal screening in HEK293T cells. The C-terminal HA-tagged pro-UFM1 plasmid (pSG5-UFM1-HA) or control vector (pSG5-HA) were transfected into the WT or gene KO cells. After 48 h post-transfection, cell lysates were analyzed by Western blot analysis using the indicated antibodies. HA, hemagglutinin; UFM1, ubiquitin-fold modifier 1; UFSP, UFM1-specific protease.

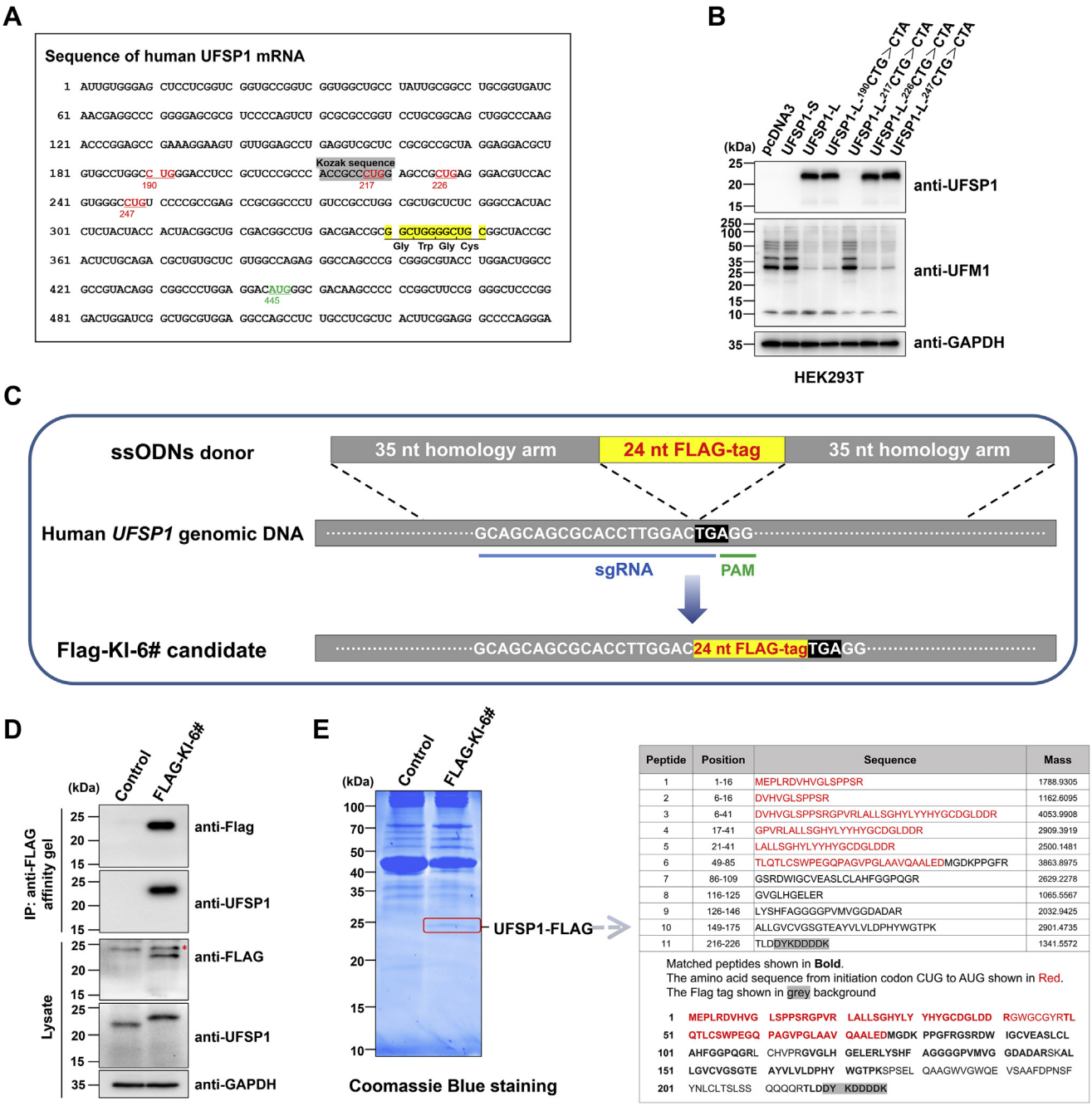


Figure 2. Human UFSP1 is translationally initiated from ^{217}CUG. *A*, four CUG near-cognate codon candidates were marked in *red* and selected for mutation analysis. Generally considered ^{445}AUG first start codon was labeled in *green*. The nucleotide sequences in *yellow background* is the potential Cys-box upstream of ^{445}AUG, the nucleotide sequences in *gray background* is a potential Kozak sequence. *B*, mutation of ^{217}CTG but not ^{190}CTG, ^{226}CTG, or ^{247}CTG to CTA eliminates the expression of human UFSP1. WT or mutant human UFSP1-L, UFSP1-S expression plasmids, and control vector were introduced into HEK293T cells; cell lysates were analyzed by Western blot with the indicated antibodies. *C*, CRISPR/Cas9 system mediated knock-in of the FLAG-tag at the C-terminal end of the *UFSP1* locus in HEK293T cells. The oligodeoxyribonucleotides (ssODNs) functioned as the linear donors and were used to introduce FLAG-tag insertion during homology-dependent repair (HDR) after the Cas9/sgRNA–mediated site-specific double-strand break (DSB). *D*, verification of expression of FLAG-tagged endogenous UFSP1 in FLAG-KI-6# clone. FLAG-KI-6# or parental HEK293T cells were lysed and immunoprecipitated with anti-FLAG M2 affinity gel for immunoblotting with FLAG or UFSP1 antibodies. *Red asterisk* indicates nonspecific band. *E*, identification of the translation initiation codon of human UFSP1 by mass spectrometry (MS). Cell lysates of FLAG-KI-6# or control cells were subjected to immunoprecipitation with anti-FLAG M2 affinity gel. The bound proteins were separated with SDS-PAGE and stained with Coomassie blue. The ~25 kDa specific band (in *red box*) was sliced and analyzed using MS, and six peptides in the UFSP1-FLAG knock-in sample that match the region between the potential ^{217}CTG start codon and the canonical ^{445}ATG codon (in *red font*) were identified. UFSP, UFM1-specific protease.

^{190}CTG, ^{226}CTG, or ^{247}CTG have no significant effect on UFSP1 expression (Fig. 2*B*). These results indicate that ^{217}CUG codon, but not the other three near-cognate codons, is the translation initiation site of human UFSP1.

In order to confirm the translation initiation site, we sought to obtain direct evidence using mass spectrometry (MS) for peptide sequencing (Fig. S2*A*). We generated a homozygous FLAG-tag knock-in (KI) cell line (FLAG-KI-6#) using CRISPR/

Cas9 system, in which a FLAG-tag coding sequence (CDS) was inserted at the C terminus of the human *UFSP1* locus (Figs. 2, *C* and *D* and S2*B*). The MS results of endogenous UFSP1-FLAG reveal a total of eleven peptides covering 75% of the predicted full-length human UFSP1, among them six peptides were covered more than 90% of the predicted N-terminal region (from ^{217}CUG to the canonical ^{445}AUG) (Fig. 2*E*). The MS data confirmed that ^{217}CUG is the human UFSP1 translation start site, which encodes the amino acid methionine (Met), but not leucine (Leu). CUG normally encodes for a Leu but when used as an alternative start site, CUG can encode for Met (22). The MS data with exogenous C-terminal HA-tagged UFSP1 (UFSP1-HA) showed a similar result (Fig. S2*C*). Theoretically, all potential UFSP1 isoforms with the same FLAG-tagged or HA-tagged C terminus were immunoprecipitated using anti-FLAG or anti-HA affinity gel. However, our data showed that only one tagged UFSP1 band was specifically detectable with a MW corresponding to the upstream near-cognate codon ^{217}CUG instead of annotated ^{445}AUG (Figs. 2, *D* and *E* and S2*C*), suggesting that UFSP1 expressed only the full-length enzyme-active form in human cells.

The 5′ UTR sequence is essential for human UFSP1 expression

While human UFSP1 translationally initiated from ^{217}CUG, we found the efficiency of exogenous expression was extremely low without the 5′ UTR sequence in the constructs. Thus, a certain 5′ UTR sequence may be required for the efficient translation of human UFSP1. The Kozak consensus sequence (RCC**AUG**G, where R is a purine) plays an important role in translation initiation (26, 27). In UFSP1 mRNA, we noted that ^{217}CUG is embedded in an optimal Kozak consensus sequence (ACCGCC**CUG**G), whereas no such motif exists surrounding the ^{445}AUG and other near-cognate codons (Fig. 2*A*). We constructed a series of UFSP1 plasmids with full-length or truncated 5′ UTR (Fig. 3*A*). After transfection, we found that the Kozak consensus sequence can only sustain basal UFSP1 expression, and efficient protein expression needs longer 5′ UTR sequence (Fig. 3*B*). We found that approximate 126 nt long sequence upstream of ^{217}CTG is sufficient for high efficiency of human UFSP1 translation as full-length 5′ UTR (Fig. 3*C*). In particular, the sequence between 97 and 126 nt upstream of ^{217}CTG seems to be required for efficient UFSP1 expression.

In order to ascertain the role of 97 to 126 nt for efficient UFSP1 expression, we constructed UFSP1-L-Random expression plasmids in which the 30 nt sequence between 97 and 126 nt were substituted by a 30 nt random sequence. After transfection, we found the efficiency of human UFSP1 expression with UFSP1-L-Random plasmids was extremely low as compared to that with WT UFSP1-L or 5′ UTR-UFSP1-126 (Fig. 3*D*). We found that the sequence between 97 and 126 nt contains an E-box motif (CAGCTG), and the expression of human UFSP1 was significantly inhibited by the E-box deletion or mutation (Fig. S3*A*). These results suggest that the 97 to 126 nt sequence plays an important role in human UFSP1 expression. In addition, we observed that this 97 to 126 nt sequence inserted upstream GFP recombinant construct is able to enhance the GFP expression (Fig. S3*B*). These data suggest that the specific sequence in 5′ UTR is essential for human UFSP1 expression.

eIF2A mediates ^{217}CUG initiation of human UFSP1 translation

The detailed mechanisms of non-AUG initiation are still unclear. It has been reported that eIF2A plays an important role in translational initiation at the CUG start codon, and some chemical compounds differentially regulate protein synthesis initiated at the AUG or CUG start codons (28, 29). For example, acriflavine selectively inhibits CUG initiation, whereas aurintricarboxylic acid (ATA) inhibits AUG initiation but enhances initiation at the CUG codon. We examined the expression levels of human UFSP1 in response to these chemical inhibitors in HEK293T and HeLa cells and found that acriflavine attenuated UFSP1 expression in a dose-dependent manner (Fig. 4, *A* and *B*), whereas ATA increased the expression of UFSP1 (Fig. 4, *C* and *D*). Whereas the expression of UFSP2 and global UFMylation were barely changed by the treatment with ATA or were inhibited by treatment with acriflavine only at high dosage (Fig. 4, *A*–*D*). Similarly, overexpression of FLAG-eIF2A increases the level of UFSP1 in both HeLa and HEK293T cells, but no changes in UFSP2 expression and in protein UFMylation were observed (Fig. 4, *E* and *F*). Moreover, eIF2A knockdown reduced human UFSP1 expression in both in HEK293T (Fig. 4, *G* and *H*) and HeLa (Fig. 4, *I* and *J*) cells. These results suggest the indispensable role of eIF2A in CUG translation initiation of human UFSP1.

Human UFSP1 functions in UFMylation/de-UFMylation process

To assess the function of UFSP1 in the maturation of the pro-UFM1 and de-UFMylation processes in comparison with UFSP2, we analyzed UFMylation/de-UFMylation of the known substrate activating signal cointegrator 1 (ASC-1) and showed that both human UFSP1 and UFSP2 can deconjugate the UFM1 from UFMylated ASC-1 (Fig. 5*A*). In addition, we purified human UFSP1 and pro-UFM1-HA from HEK293T cells and showed that human UFSP1 can cleave the pro-UFM1-HA to generate mature UFM1 in *in vitro* assay (Fig. 5*B*). We have initially attempted to purify human recombinant UFSP1 from *Escherichia coli*; however, human UFSP1 did not get expressed in *E. coli* (Fig. S4*A*). Using codon-optimized (CO) UFSP1-L plasmid with ^{217}ATG initiation codon, we have successfully purified the human recombinant UFSP1 protein from *E. coli* (Fig. S4*B*) and showed that human recombinant UFSP1 exhibits high enzymatic activity in maturation of pro-UFM1 (Fig. 5*C*). These results provide direct evidence that human UFSP1, like UFSP2, is an active UFM1-specific protease. In addition, our results showed that human UFSP1 exhibits higher catalytic activity than UFSP2 in de-UFMylation of ASC-1, overexpression of UFSP1 mediated stronger protein de-UFMylation than overexpression of UFSP2 (Fig. 5, *A* and *D*).

Both of UFSP1 and UFSP2 are widely expressed in human tissues, and UFSP1 is expressed at a very low level compared with UFSP2 (Fig. S5). Unlike predominantly ER membrane-localized

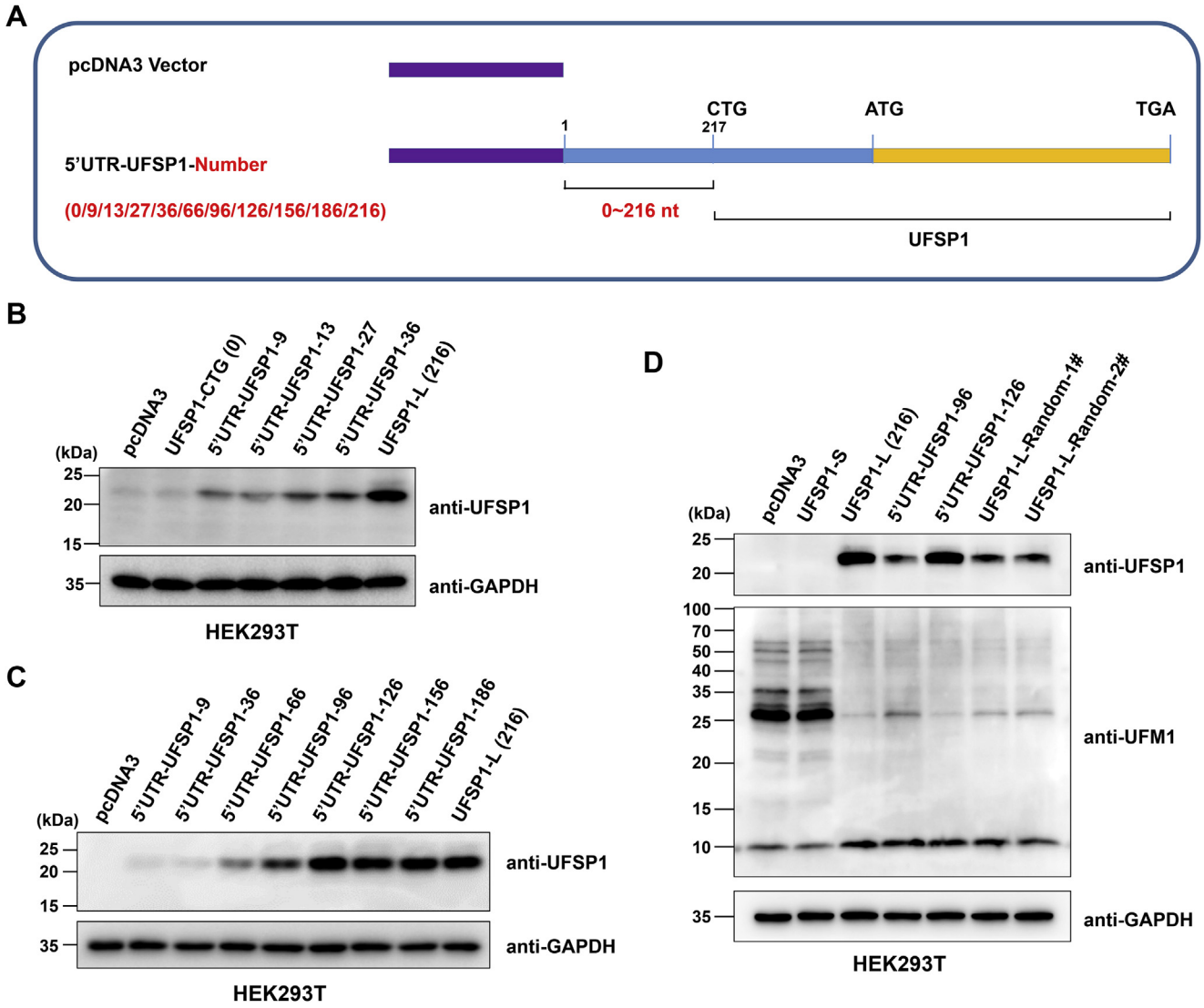


Figure 3. The 5′ UTR sequence is essential for human UFSP1 expression. *A*, schematic diagram of constructs of UFSP1 with different lengths of 5′ UTR upstream of the ^{217}CTG codon. *B* and *C*, 5′ UTR is essential for human UFSP1 expression. Human UFSP1 expression plasmids with different 5′ UTR lengths (0∼216 nt) upstream of ^{217}CTG codon and control vector were transfected into HEK293T cells, followed by Western blotting analysis. *D*, sequence of nucleotides 97 to 126 plays an important role in human UFSP1 expression. UFSP1-L-Random-1# and 2# plasmids, UFSP1 expression plasmids with different 5′ UTR lengths (96, 126, or 216 nt) upstream of ^{217}CTG codon, UFSP1-S, and control vector were transfected into HEK293T cells; cell lysates were analyzed by Western blotting analysis. UFSP, UFM1-specific protease.

UFSP2, we found that human UFSP1 is a cytosolic protein (Fig. 5*E*). Based on the differences of UFSP1 and UFSP2 in catalytic activity, expression levels, intracellular localization, and UFMylation pattern in *UFSP1* or *UFSP2* KO cells, we believe UFSP2 appears to be mainly involved in the deconjugation of UFM1, whereas UFSP1 is mainly involved in the maturation of pro-UFM1. Both of UFSP1 and UFSP2 together maintain a dynamic and reversible process of protein UFMylation/de-UFMylation in cells.

Discussion

In this study, we demonstrated that translation of human UFSP1 is initiated from the upstream near-cognate ^{217}CUG codon in 5′ UTR through eIF2A-mediated translational initiation, rather than the annotated ^{445}AUG codon, producing an active UFM1-specific protease. The approximate 126 nt long sequence upstream of ^{217}CUG plays an important role in the efficiency of UFSP1 expression. Like UFSP2, human UFSP1 functions in both maturation of the pro-UFM1 and de-UFMylation reaction. Unlike UFSP2, human UFSP1 expresses at low level in cytosol and may have different substrate specificity.

Most CUG-initiated proteins (such as PTEN, FGF2, and BiP) have the canonical AUG-initiated form besides one or two CUG-initiated forms (29–31), whereas human UFSP1 has only the ^{217}CUG-initiated form (∼23 kDa), but not the ^{445}AUG-initiated form (∼17 kDa). Even during exogenous expression, we can only detect a very small amount of ^{445}AUG-initiated form by overexposure, which is likely caused by the ribosomal leaky scanning mechanism during protein translational initiation (32). Thus, human UFSP1 is translated by using near-cognate codon ^{217}CUG, rather than annotated ^{445}AUG codon.

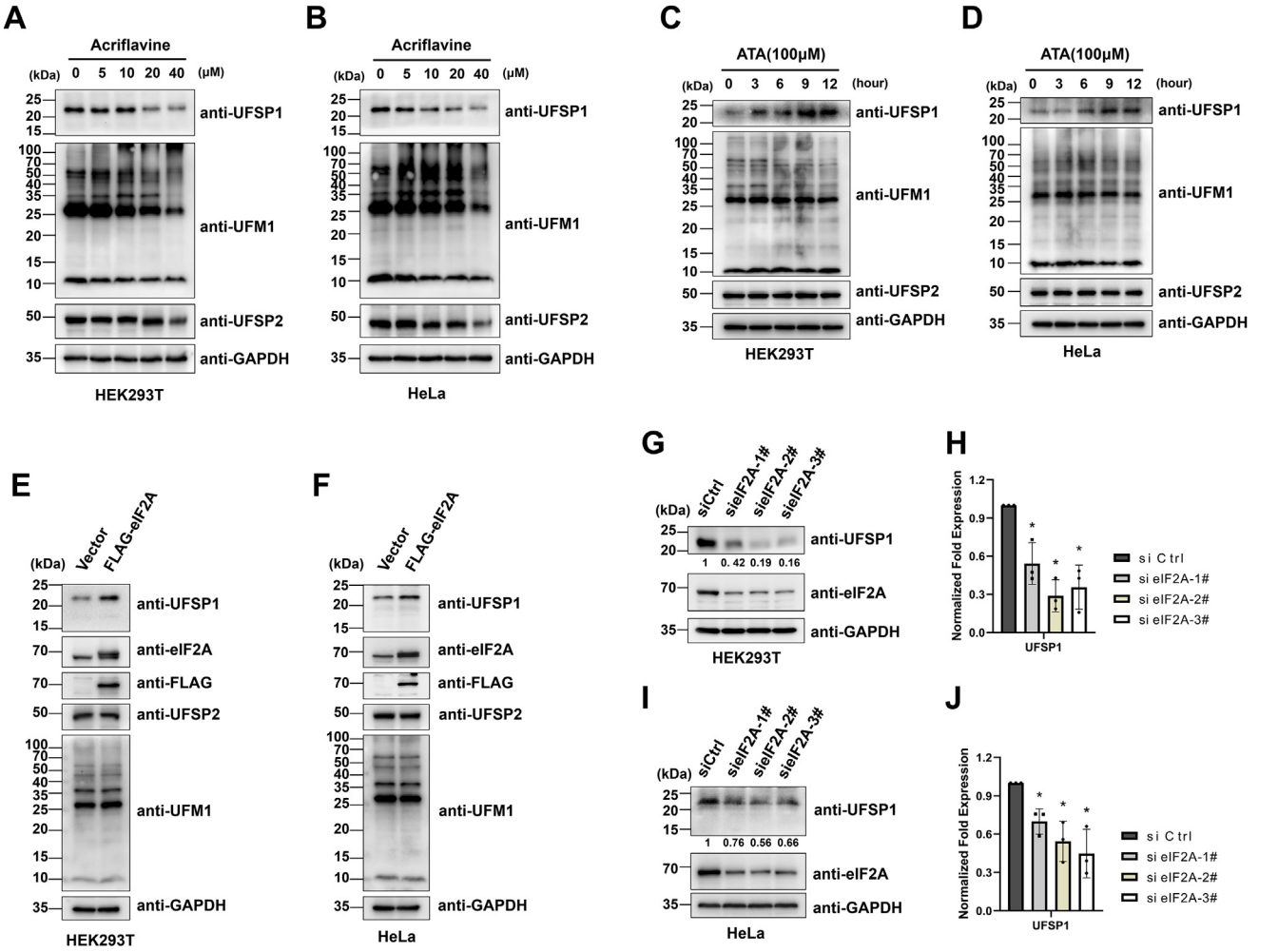


Figure 4. eIF2A mediates ^{217}CUG initiation of human UFSP1 translation. *A* and *B*, dose-dependent inhibition of human UFSP1 expression by acriflavine. HEK293T (*A*) and HeLa (*B*) cells were treated with different doses of acriflavine for 4 h, followed by Western blotting analysis. *C* and *D*, induction of human UFSP1 expression using ATA in a time-dependent manner. HEK293T (*C*) and HeLa (*D*) cells were treated with ATA (100 μM) for various periods of time, and the expression of UFSP1 and GAPDH were examined by Western blotting. *E* and *F*, eIF2A upregulates human UFSP1 expression. FLAG-tagged human eIF2A was overexpressed in HEK293T (*E*) and HeLa (*F*) cells, and UFSP1 expression was evaluated by Western blotting. eIF2A expression was verified by probing the same blot with anti-FLAG and anti-eIF2A antibodies. *G–J*, reduction of human UFSP1 expression in response to knockdown of eIF2A. HEK293T (*G* and *H*) and HeLa (*I* and *J*) cells were transfected with eIF2A siRNAs or control siRNAs. Cell lysates were analyzed using Western blotting. Densitometric analysis was performed using image processing software. Data are mean ± SD. Differences in means between two groups were analyzed using two-sided unpaired *t* test (*$p < 0.05$). ATA, aurintricarboxylic acid; UFSP, UFM1-specific protease.

Although a Kozak consensus sequence exits surrounding the ^{217}CUG, efficient protein expression (UFSP1 or GFP) still needs longer 5′ UTR sequence (Figs. 3*B* and S3*B*). The sequence in 97 to 126 nt upstream of the ^{217}CUG played a pivotal role in regulating UFSP1 expression efficiency. Previous studies showed a correlation between the efficiency of mRNA translation and secondary structure stability of local mRNA sequence near the initiation codon (33, 34). Therefore, it is possible that 5′ UTR sequence of human UFSP1 forms secondary structures (such as hairpin loops, bulges, and internal loops) necessary for the efficiency of translation initiation of UFSP1.

eIF2A plays an important role in translational initiation at non-AUG start codon. Like other CUG-initiated proteins, the translation initiation of human UFSP1 is mediated by eIF2A. Given that phosphorylation of eIF2α leads to a general inhibition of translation, and eIF2A is increased during multiple stress conditions (31), suggesting non-AUG translation is a common stress response mechanism. Although the detailed molecular mechanism remains unclear, UFMylation is closely related to ER homeostasis (8–12). Therefore, we propose that eIF2A mediated ^{217}CUG initiation of human UFSP1 expression (as well as UFSP1 in other primates) may provide a specific regulatory mechanism in ER homeostasis regulation.

Furthermore, sequence alignment showed that the Cys protease domain of UFSP1 is conserved among organisms from fruit flies to humans (Fig. S6*A*). However, the CUG translation initiation mechanism of UFSP1 only exists in primates but not in other mammals (Fig. S6*B* and Table S2), suggesting that specific regulatory mechanism is restricted to primates. Unlike the human *UFSP1*, mouse *UFSP1* has a canonical AUG start site upstream of Cys protease domain coding region and expressed as an active UFM1-specific protease with ∼23 kDa (19). Although two recent reports suggest

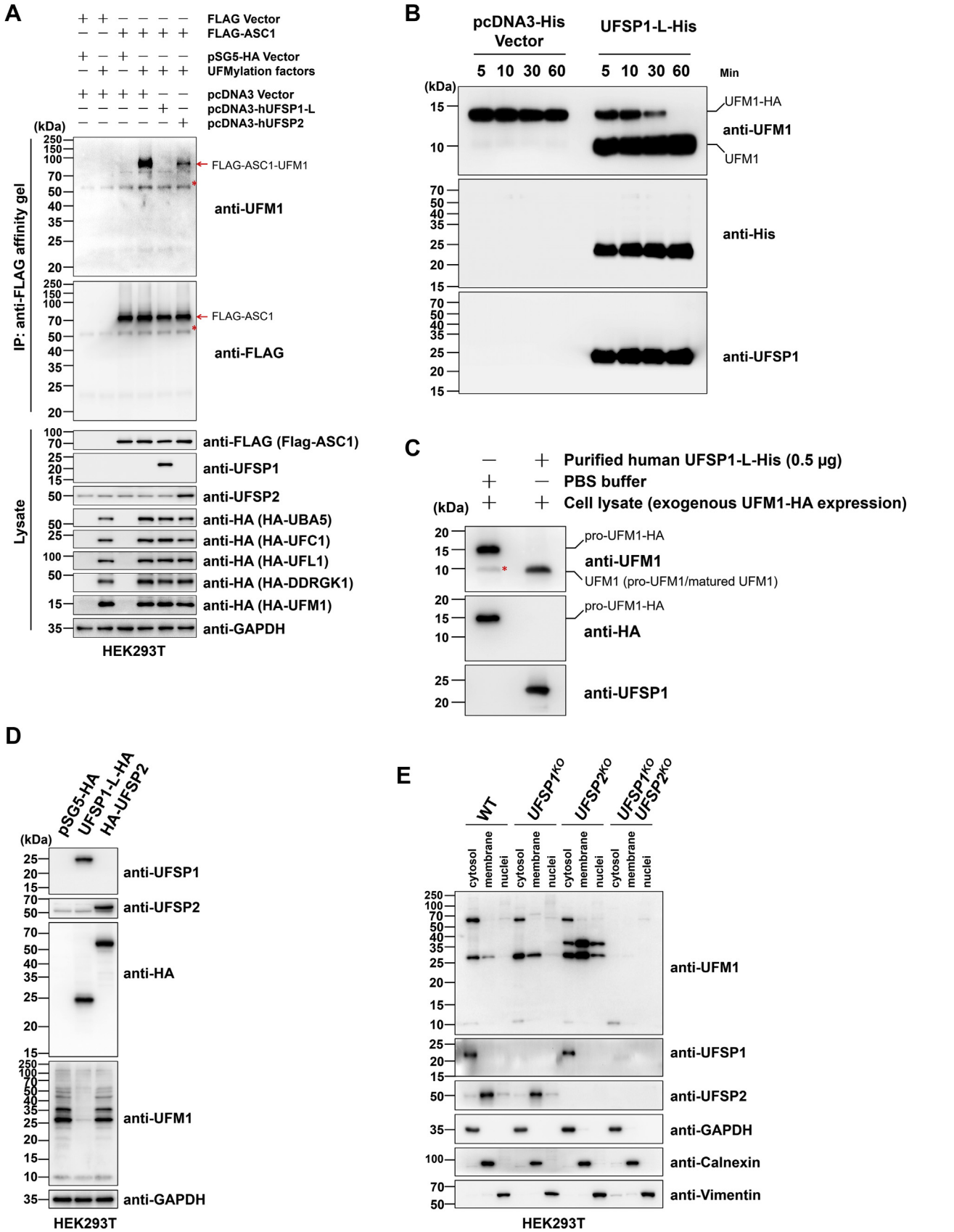


Figure 5. Human UFSP1 functions in UFMylation/de-UFMylation process. *A*, human UFSP1 can release UFM1 from UFMylated substrate ASC-1. HEK293T cells were transfected with the indicated plasmids, followed by UFMylation assay and immunoblotting with the indicated antibodies. *Red asterisks* indicate nonspecific band. *B*, human UFSP1 cleaves pro-UFM1 *in vitro*. C-terminal HA-tagged UFM1 was purified using anti-HA affinity gel. UFSP1-L-His or control vector were transfected in HEK293T cells. Cells lysates were subjected to immunoprecipitation with Ni-NTA Agarose. Amounts of UFSP1-His or His-tag

that mouse UFSP1 specifically expresses at the neuromuscular junction and functions as regulators of the acetylcholine receptor clustering, implicating that UFSP1 plays some role in the neuromuscular junction formation during development (35, 36), and the physiological roles of UFSP1 remain largely unknown.

In conclusion, we identified human UFSP1 as an active UFM1-specific protease participating in UFMylation/de-UFMylation process. These findings provide additional insights in our understanding of UFMylation/de-UFMylation process in cells.

Experimental procedures

Plasmids, siRNAs, antibodies, and chemicals

The 5′ UTR and CDS of human UFSP1 gene was amplified by PCR from HeLa cDNA. The fragment of UFSP1 canonical CDS from annotated ^{445}ATG codon to TGA stop codon is referred to UFSP1-**S**hort. The fragment that contains the canonical CDS and 5′ UTR sequence is referred to UFSP1-**L**ong. A series of truncated 5′ UTR sequence of human UFSP1 with UFSP1 or GFP (without ATG) CDS were constructed and named with the length of 5′ UTR. All of the amplified fragments were subcloned into the indicated vectors. Human UFSP1-L expression plasmids with specific point mutation were constructed using Q5 Site-Directed Mutagenesis Kit (New England Biolabs). UFSP1-L-Random expression plasmids in which the 30 nt sequence between 97 and 126 nt were substituted by a 30 nt random sequence. Prokaryotic expression plasmids of human UFSP1-L were subcloned in the WT or CO UFSP1-L cDNA with ^{217}CTG or ^{217}ATG initiation codon into pET-28a (+) vector using recombinant DNA technology, which destroyed the ATG initiation codon of the pET-28a (+) vector and seamlessly integrated 6 × His-tag at the C terminus. Human UFSP2 cDNA was subcloned into pSG5-HA and pcDNA3 (Invitrogen) vectors, respectively. Human UBA5 (E1), UFC1 (E2), UFL1 (E3), DDRGK1, and UFM1 cDNAs were subcloned into pSG5-HA vector. Human eIF2A and ASC-1 cDNAs were subcloned into p3 × FLAG-CMV (Sigma–Aldrich) vector. The HA tag in pSG5-UFM1-HA and pSG5-UFSP1-L-HA and 6 × His tag in UFSP1-L-His were all fused at the C-terminal of gene.

The single guide RNAs (sgRNAs) target human *UFSP1*, *UFSP2*, and *UFM1* genes as well as nontargeting control were synthesized and subcloned into the *BsmB*I site of lentiCRISPR vector (Addgene).

The siRNAs targeting human eIF2A and nontargeting control were purchased from GenePharma.

Antibodies against UFSP1, UFL1, DDRGK1 (Sigma–Aldrich), UFSP2, UFM1, UBA5, UFC1 (Abcam), HA, Calnexin, Vimentin (Cell Signaling Technology), eIF2A (Proteintech), FLAG, His (GenScript), GFP (Beyotime Biotechnology), and GAPDH (HuaBio) were used.

Acriflavine and ATA were purchased from Sigma–Aldrich. Puromycin was purchased from Selleck Chemicals.

More detailed information of PCR primers, sgRNAs and siRNAs, antibodies, and chemicals are described in Table S1.

Cell culture and transfection

HEK293T and HeLa cells were purchased from the American Type Culture Collection (ATCC) and maintained in Dulbecco's modified Eagle's medium (Biological Industries) supplemented with 10% fetal bovine serum (Gibco) and penicillin–streptomycin (Gibco). All cell lines were maintained at 37 °C in a 5% CO_2 humidified atmosphere. Plasmid transfection and RNA interference were carried out with Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol.

Lentivirus production and transduction

HEK293T cells were seeded at ∼40% confluence in 10 cm dishes the day before transfection. One hour prior to transfection, medium was removed and 10 ml of prewarmed reduced serum medium Opti-MEM was added to each dish. Transfection was performed using Lipofectamine 2000 (Invitrogen) to transduce the transfer vector lentiCRISPR-sgRNA with packaging plasmids psPAX2 and pMD2.G for production of lentiviral particles. The supernatant was collected, filtered through a 0.45 μm filter, and concentrated by passing through an ultrafiltration tube (Millipore). Concentrated supernatant was aliquoted, flash-frozen in liquid nitrogen, and stored at −80 °C.

For transduction, target cells (HEK293T cells or HeLa cells) were seeded in 12-well plates and allowed to adhere overnight. The concentrated supernatant and fresh medium were added to the target cells with 8 μg/ml polybrene. Cells were incubated with the virus-containing medium overnight, followed by replacement with fresh medium. After 48 h, puromycin (1 μg/ml for HEK293T cells or HeLa cells) was added to select stable cell lines.

CRISPR/Cas9 mediated KI of FLAG-tag at the C-terminal end of the UFSP1 locus

CRISPR/Cas9–mediated homology-dependent repair was performed as previously described (37). A DNA sequence encoding the FLAG tag (DYKDDDDK) was introduced at the C-terminal end of the UFSP1 locus followed by stop codon. sgRNA (UFSP1-sgRNA-5#) was designed to cut proximal to the stop codon. A single-stranded oligodeoxyribonucleotide (ssODN) was synthesized and used as donor template for

control was incubated with UFM1-HA for the indicated time at 37 °C. The mixtures were subjected to SDS-PAGE followed by Western blot analysis. *C*, prokaryotic expressed human UFSP1 has protease activity, which cleaves pro-UFM1 *in vitro*. The cell lysates of *UFSP1*^{KO}/*UFSP2*^{KO} HEK293T cells transfected with pSG5-UFM1-HA used as protease substrates. A total of 0.5 μg purified UFSP1-L-His protein or control buffer was incubated with UFM1-HA for the 20 min at 37 °C. The reaction mixtures were subjected to SDS-PAGE followed by Western blot analysis. *Red asterisk* indicates band of endogenous pro-UFM1 in *UFSP1*^{KO}/*UFSP2*^{KO} HEK293T cells. *D*, HEK293T cells were transfected with indicated expression plasmids (UFSP1-L with a C-terminal HA-tag, human UFSP2 with an N-terminal HA-tag, or control vector), followed by Western blotting analysis. *E*, WT, *UFSP1*^{KO}, *UFSP2*^{KO}, and *UFSP1*^{KO}/*UFSP2*^{KO} HEK293T cells were fractionated with sequential detergent extractions before Western blotting analysis with the indicated antibodies. ASC-1, activating signal cointegrator 1; HA, hemagglutinin; UFM1, ubiquitin-fold modifier 1; UFSP, UFM1-specific protease.

homology-dependent repair–mediated insertion of FLAG-tag. The ssODN is 94 nt, with two 35 nt flanking homology arms and a 24 nt FLAG-tag insertion (5′-GCCTTAGCTCCCAACAGCAGCAGCGCACCTTGGACGATTACAAGGACGACGATGACAAGTGAGGACGAAGTTACAGAACTGAGATTCTCGGGTC-3′).

HEK293T cells were seeded at 5×10^5 per well in a 6-well plate. After 24 h, cells were cotransfected with Cas9 vector expressing UFSP1-sgRNA-5# and ssODN donor template. Forty-eight hours after transfection, individual cells are sorted into 96-well plates and grown out as single-cell clones. The single-cell clones were lysed and genotyped by FLAG-tag specific PCR screen. For FLAG-tag–positive candidates, the second round of PCR was performed, and PCR products were directly cloned into a blunt-end vector. The homozygous KI clones were selected by Sanger sequencing.

Generation of CRISPR KO cell lines

For *UFSP1* or *UFSP2* single gene KO, HEK293T cells with stable expression of Cas9-sgRNA were generated by lentivirus infection. A population of UFSP1-sgRNA and UFSP2-sgRNA stable cells was obtained by selecting with 1 μg/ml puromycin. After selecting for 1 to 2 weeks, clones were isolated by limiting dilution and screened for KO by Western blotting analysis. *UFSP1* or *UFSP2* single-gene deficiency monoclonal cell lines were named $UFSP1^{KO}$ and $UFSP2^{KO}$, respectively. For *UFSP1* and *UFSP2* double gene KO, $UFSP2^{KO}$ cell line was infected with UFSP1-sgRNA lentivirus. Without puromycin selection, cells were single-cell seeded into 96-wells at 48 h after infection. Clones derived from single cells were screened by Western blotting analysis. Finally, the double gene KO cell line was named $UFSP1^{KO}/UFSP2^{KO}$.

Western blotting

For Western blotting assay, cells were lysed in radioimmunoprecipitation assay buffer (50 mM Tris–HCl pH 7.5, 150 mM NaCl, 1% nonidet P-40 [NP-40], 0.5% sodium deoxycholate, 1 mM EDTA, and 0.1% SDS) containing 1 × protease and phosphatase inhibitor cocktail (Roche). Protein concentrations were determined using Pierce BCA Protein Assay Kit (Thermo Scientific). Protein lysates were resolved by SDS-PAGE, transferred to polyvinylidene fluoride membranes (Millipore), blocked in 5% milk, and probed with primary antibody overnight at 4 °C, and then incubated with horseradish peroxidase–conjugated secondary antibodies. Western blots were visualized with Immobilon Western horseradish peroxidase substrate (Millipore).

In cell UFMylation and de-UFMylation assay

ASC-1 was identified as an UFMylation substrate (17). FLAG-tagged ASC-1 was constructed and used for the *in vivo* UFMylation and de-UFMylation assay. HEK293T cells were transfected with the appropriate constructs. After 48 h, cells were harvested and lysed by boiling in buffer (150 mM Tris–HCl pH 8.0, 5% SDS, and 30% glycerol) for 10 min. Cell lysates were diluted 20-fold with buffer A (50 mM Tris–HCl pH 8.0, 150 mM NaCl, 0.5% NP-40, and 2 mM N-ethylmaleimide) containing 1 × protease inhibitor cocktail, as described (17). After incubation with anti-FLAG M2 affinity gel (Sigma–Aldrich) overnight at 4 °C, the immunoprecipitates were resolved by SDS-PAGE followed by Western blotting analysis.

Prokaryotic expression and purification of human UFSP1

Prokaryotic expression plasmids of human UFSP1-L (pET-28a-UFSP1-^{217}CTG-His, pET-28a-UFSP1-^{217}ATG-His, pET-28a-UFSP1-^{217}CTG-His-CO, and pET-28a-UFSP1-^{217}ATG-His-CO) were transformed into *E. coli* BL21(DE3) cell for IPTG-induced expression. The positive clones were inoculated into LB medium (containing 50 μg/ml kanamycin) and cultured overnight. Then the bacteria solution was used to inoculate (1% v/v) 5 ml of fresh LB medium in 15 ml tubes and incubated at 37 °C to an A_{600nm} of 0.6. The expression of human UFSP1 was induced by the addition of IPTG (0.2 mM). After induction, the bacteria were collected, lysed, and centrifuged. The soluble (supernatant) and insoluble (pellet) fractions were subjected to SDS-PAGE, followed by Coomassie blue staining. Only pET-28a-UFSP1-^{217}ATG-His-CO plasmid transformed strain expressed human UFSP1 protein after IPTG induction. And, we used this strain for protein purification of UFSP1.

The IPTG concentration, temperature, and time of incubation with IPTG were optimized for human UFSP1 expression in *E. coli* BL21(DE3) cells. For UFSP1 protein purification, 200 ml bacteria solution was taken in 300 ml conical flask and incubated at 37 °C to an A_{600nm} of 0.6. After precooling, IPTG was added with a final concentration of 0.2 mM and induction for 16 to 20 h at 16 °C. UFSP1 protein was purified with Ni-nitrilotriacetic acid (NTA) Agarose (Qiagen) as described by the manufacturer. Purified proteins were concentrated and desalted using 10 kDa molecular mass cut-off centrifugal filter units (Amicon Ultra 15 ml). The proteins were then quantified, diluted, and stored at −80 °C in aliquots.

In vitro *enzymatic activity assay for UFSP1*

The human pro-UFM1 with a C-terminal HA-tag (pSG5-UFM1-HA) and UFSP1-L with a C-terminal 6 × His-tag (UFSP1-L-His) plasmids were constructed and transfected to HEK293T cells. For the purification of pro-UFM1-HA, cell lysates were incubated with anti-HA affinity gel (Sigma–Aldrich) overnight at 4 °C. The bead-bound pro-UFM1-HA protein was washed and eluted by addition of HA peptide (100 μg/ml). For the purification of UFSP1-His, cell lysates were subjected to Ni-NTA Agarose pulldown overnight at 4 °C. The Ni-NTA resin-bound UFSP1-His proteins were washed with wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 0.2% NP-40, adjust to pH 7.4) and eluted with elution buffer (50 mM NaH2PO4, 300 mM NaCl, 250 mM imidazole, 0.05% NP-40, adjust to pH 7.4). The imidazole was diluted by dialysis.

UFM1-processing activity was assayed by using pro-UFM1-HA as a substrate. Purified human UFSP1-His were incubated for different time points at 37 °C with purified pro-UFM1-HA in 10 mM Tris–HCl buffer (pH 8.0) containing 0.1 mM EDTA

and 0.1 mM DTT. The reaction products were subjected to SDS-PAGE and evaluated by Western blotting analysis with antibody recognizing UFSP1, His, and UFM1.

Cell fractionation

The same amount of HEK293T WT and gene KO cells were collected in cold PBS and pelleted by centrifuging at 1000*g*. Cytoplasmic, membrane/organelle, and nuclear/cytoskeletal fractions were separated by Cell Fractionation Kit (Cell Signaling Technology) according to the manufacturer's protocol. Equal volumes of the collected fractions were analyzed by SDS-PAGE and Western blotting analysis.

LC-MS/MS

For affinity enrichment of exogenously expressed UFSP1, HEK293T cells were transfected with the C-terminal HA-tagged UFSP1-L construct (UFSP1-L-HA). The cells were lysed with lysis buffer (50 mM Tris–HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, and 0.5% NP-40). UFSP1-HA protein was immunoprecipitated by incubation with anti-HA affinity gel overnight at 4 °C. For affinity enrichment of endogenously expressed UFSP1, C-terminal FLAG-tagged HEK293T cells (FLAG-KI-6#) were lysed and immunoprecipitated by incubation with anti-FLAG M2 affinity gel. All of the aforementioned eluates were resolved by SDS-PAGE, followed by Coomassie blue staining. Desired bands were excised from gels and dissected into small squared blocks. The proteins in excised gel pieces were subjected to in-gel tryptic digestion procedure that involves destaining, reduction, alkylation, digestion, and finally, extraction of peptides for LC-MS/MS analysis.

For LC-MS/MS analysis, the peptides were loaded onto a homemade reversed-phase analytical column and separated using a linearly programmed gradient mobile phase consisting of solvent A (0.1% formic acid and 2% acetonitrile in water) and solvent B (0.1% formic acid and 80% acetonitrile in water). The gradient was comprised of an increase from 8% to 35% solvent B over 60 min, 35% to 80% in 5 min, and then holding at 80% for the last 3 min, all at a constant flow rate of 300 nl/min on an EASY-nLC 1000 UPLC system. The peptides were subjected to MS on a Q Exactive HF Mass Spectrometer (Thermo Fisher Scientific). For data processing, the resulting MS/MS data were processed using Maxquant (Max-Planck-Institute of Biochemistry). Tandem mass spectra were searched against *Uniprot-Human*. The MS analysis was performed by Micrometer Biotech Company, and the lists of human UFSP1 peptide sequences identified are shown in Supplementary Data.

Statistics and reproducibility

The ChemiDoc MP Imaging system (Bio-Rad) was used to collect the Western blot images, and Image Lab 6.0.1 (Bio-Rad) and ImageJ 1.52a (National Institutes of Health) were used for image processing and densitometry quantifications. GraphPad Prism 5 (GraphPad Software) and Excel 2007 software (Microsoft) was used for all of the statistical analysis. Data are mean ± SD. Differences in means between two groups were analyzed using two-sided unpaired *t* test ($^*p < 0.05$). Each experiment was repeated independently with similar results.

Data availability

All data that support the findings of this study are available from the corresponding authors upon reasonable request. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium *via* the PRIDE partner repository with the dataset identifier PXD033083 (38).

Supporting information—This article contains supporting information.

Acknowledgments—We thank members of the Cong lab for helpful discussions. This work was supported by the National Natural Science Foundation of China (32170776, 31730020, 31900520) and the Hangzhou Science and Technology Bureau [20182014B01].

Author contributions—Q.L. and Y.-S.C. conceived and designed the experiments; Q.L. and Y.J. performed all of the experiments with help from S.X., J.Z., J.M., X.M and M.W.; Q.L., Y.J. and Y.-S.C. analyzed data; Q.L. and Y.-S.C. wrote the manuscript.

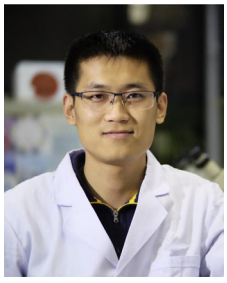
Conflict of interest—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: 5′ UTR, 5′ untranslated region; ASC-1, activating signal cointegrator 1; ATA, aurintricarboxylic acid; cDNA, complementary DNA; CDS, coding sequence; CO, codon-optimized; ER, endoplasmic reticulum; HA, hemagglutinin; KI, knockin; MS, mass spectrometry; MW, molecular weight; NP-40, nonidet P-40; NTA, nitrilotriacetic acid; sgRNA, single guide RNA; ssODN, oligodeoxyribonucleotide; UBA5, UFM1-activating enzyme 5; UFC1, UFM1-conjugating enzyme 1; UFL1, UFM1-specific ligase 1; UFM1, ubiquitin-fold modifier 1; UFSP, UFM1-specific protease.

References

1. Komatsu, M., Chiba, T., Tatsumi, K., Iemura, S. I., Tanida, I., Okazaki, N., *et al.* (2004) A novel protein-conjugating system for Ufm1, a ubiquitin-fold modifier. *EMBO J.* **23**, 1977–1986
2. Gerakis, Y., Quintero, M., Li, H., and Hetz, C. (2019) The UFMylation system in proteostasis and beyond. *Trends Cell Biol.* **29**, 974–986
3. Daniel, J., and Liebau, E. (2014) The ufm1 cascade. *Cells* **3**, 627–638
4. Nahorski, M. S., Maddirevula, S., Ishimura, R., Alsahli, S., Brady, A. F., Begemann, A., *et al.* (2018) Biallelic UFM1 and UFC1 mutations expand the essential role of ufmylation in brain development. *Brain* **141**, 1934–1945
5. Colin, E., Daniel, J., Ziegler, A., Wakim, J., Scrivo, A., Haack, T. B., *et al.* (2016) Biallelic variants in UBA5 reveal that disruption of the UFM1 cascade can result in early-onset encephalopathy. *Am. J. Hum. Genet.* **99**, 695–703
6. Egunsola, A. T., Bae, Y., Jiang, M.-M., Liu, D. S., Chen-Evenson, Y., Bertin, T., *et al.* (2017) Loss of DDRGK1 modulates SOX9 ubiquitination in spondyloepimetaphyseal dysplasia. *J. Clin. Invest.* **127**, 1475–1484
7. Di Rocco, M., Rusmini, M., Caroli, F., Madeo, A., Bertamino, M., Marre-Brunenghi, G., *et al.* (2018) Novel spondyloepimetaphyseal dysplasia due to UFSP2 gene mutation. *Clin. Genet.* **93**, 671–674
8. Tatsumi, K., Yamamoto-Mukai, H., Shimizu, R., Waguri, S., Sou, Y.-S., Sakamoto, A., *et al.* (2011) The Ufm1-activating enzyme Uba5 is indispensable for erythroid differentiation in mice. *Nat. Commun.* **2**, 1–7
9. Cai, Y., Pi, W., Sivaprakasam, S., Zhu, X., Zhang, M., Chen, J., *et al.* (2015) UFBP1, a key component of the Ufm1 conjugation system, is essential for

ufmylation-mediated regulation of erythroid development. *PLoS Genet.* **11**, e1005643
10. Zhang, M., Zhu, X., Zhang, Y., Cai, Y., Chen, J., Sivaprakasam, S., *et al.* (2015) RCAD/Ufl1, a Ufm1 E3 ligase, is essential for hematopoietic stem cell function and murine hematopoiesis. *Cell Death Differ.* **22**, 1922–1934
11. Liu, J., Wang, Y., Song, L., Zeng, L., Yi, W., Liu, T., *et al.* (2017) A critical role of DDRGK1 in endoplasmic reticulum homoeostasis via regulation of IRE1α stability. *Nat. Commun.* **8**, 1–12
12. Liang, J. R., Lingeman, E., Luong, T., Ahmed, S., Muhar, M., Nguyen, T., *et al.* (2020) A genome-wide ER-phagy screen highlights key roles of mitochondrial metabolism and ER-resident UFMylation. *Cell* **180**, 1160–1177.e20
13. Walczak, C. P., Leto, D. E., Zhang, L., Riepe, C., Muller, R. Y., DaRosa, P. A., *et al.* (2019) Ribosomal protein RPL26 is the principal target of UFMylation. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 1299–1308
14. Wang, L., Xu, Y., Rogers, H., Saidi, L., Noguchi, C. T., Li, H., *et al.* (2020) UFMylation of RPL26 links translocation-associated quality control to endoplasmic reticulum protein homeostasis. *Cell Res.* **30**, 5–20
15. Wang, Z., Gong, Y., Peng, B., Shi, R., Fan, D., Zhao, H., *et al.* (2019) MRE11 UFMylation promotes ATM activation. *Nucleic Acids Res.* **47**, 4124–4135
16. Qin, B., Yu, J., Nowsheen, S., Wang, M., Tu, X., Liu, T., *et al.* (2019) UFL1 promotes histone H4 ufmylation and ATM activation. *Nat. Commun.* **10**, 1–13
17. Yoo, H. M., Kang, S. H., Kim, J. Y., Lee, J. E., Seong, M. W., Lee, S. W., *et al.* (2014) Modification of ASC1 by UFM1 is crucial for ERα transactivation and breast cancer development. *Mol. Cell* **56**, 261–274
18. Liu, J., Guan, D., Dong, M., Yang, J., Wei, H., Liang, Q., *et al.* (2020) UFMylation maintains tumour suppressor p53 stability by antagonizing its ubiquitination. *Nat. Cell Biol.* **22**, 1056–1063
19. Kang, S. H., Kim, G. R., Seong, M., Baek, S. H., Seol, J. H., Bang, O. S., *et al.* (2007) Two novel ubiquitin-fold modifier 1 (Ufm1)-specific proteases, UfSP1 and UfSP2. *J. Biol. Chem.* **282**, 5256–5262
20. Ishimura, R., Obata, M., Kageyama, S., Daniel, J., Tanaka, K., and Komatsu, M. (2017) A novel approach to assess the ubiquitin-fold modifier 1-system in cells. *FEBS Lett.* **591**, 196–204
21. Peabody, D. S. (1989) Translation initiation at non-AUG triplets in mammalian cells. *J. Biol. Chem.* **264**, 5031–5035
22. Kearse, M. G., and Wilusz, J. E. (2017) Non-AUG translation: a new start for protein synthesis in eukaryotes. *Genes Dev.* **31**, 1717–1731
23. Ingolia, N. T., Lareau, L. F., and Weissman, J. S. (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**, 789–802
24. Yoo, H. M., Park, J. H., Jeon, Y. J., and Chung, C. H. (2015) Ubiquitin-fold modifier 1 acts as a positive regulator of breast cancer. *Front. Endocrinol. (Lausanne)* **6**, 36
25. Lee, S., Liu, B., Lee, S., Huang, S.-X., Shen, B., and Qian, S.-B. (2012) Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E2424–E2432
26. Kozak, M. (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283–292
27. Diaz de Arce, A. J., Noderer, W. L., and Wang, C. L. (2018) Complete motif analysis of sequence requirements for translation initiation at non-AUG start codons. *Nucleic Acids Res.* **46**, 985–994
28. Starck, S. R., Jiang, V., Pavon-Eternod, M., Prasad, S., McCarthy, B., Pan, T., *et al.* (2012) Leucine-tRNA initiates at CUG start codons for protein synthesis and presentation by MHC class I. *Science* **336**, 1719–1723
29. Liang, H., He, S., Yang, J., Jia, X., Wang, P., Chen, X., *et al.* (2014) PTENalpha, a PTEN isoform translated through alternative initiation, regulates mitochondrial function and energy metabolism. *Cell Metab.* **19**, 836–848
30. Vagner, S. P., Touriol, C., Galy, B., Audigier, S., Gensac, M.-C., Amalric, F., *et al.* (1996) Translation of CUG-but not AUG-initiated forms of human fibroblast growth factor 2 is activated in transformed and stressed cells. *J. Cell Biol.* **135**, 1391–1402
31. Starck, S. R., Tsai, J. C., Chen, K., Shodiya, M., Wang, L., Yahiro, K., *et al.* (2016) Translation from the 5′ untranslated region shapes the integrated stress response. *Science* **351**, aad3867
32. Kozak, M. (1995) Adherence to the first-AUG rule when a second AUG codon follows closely upon the first. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 2662–2666
33. Kudla, G., Murray, A. W., Tollervey, D., and Plotkin, J. B. (2009) Coding-sequence determinants of gene expression in Escherichia coli. *Science* **324**, 255–258
34. Kochetov, A. V., Palyanov, A., Titov, I. I., Grigorovich, D., Sarai, A., and Kolchanov, N. A. (2007) AUG_hairpin: prediction of a downstream secondary structure influencing the recognition of a translation start site. *BMC Bioinformatics* **8**, 1–7
35. Petrany, M. J., Swoboda, C. O., Sun, C., Chetal, K., Chen, X., Weirauch, M. T., *et al.* (2020) Single-nucleus RNA-seq identifies transcriptional heterogeneity in multinucleated skeletal myofibers. *Nat. Commun.* **11**, 1–12
36. Kim, M., Franke, V., Brandt, B., Lowenstein, E. D., Schöwel, V., Spuler, S., *et al.* (2020) Single-nucleus transcriptomics reveals functional compartmentalization in syncytial skeletal muscle cells. *Nat. Commun.* **11**, 1–14
37. Paix, A., Folkmann, A., Goldman, D. H., Kulaga, H., Grzelak, M. J., Rasoloson, D., *et al.* (2017) Precision genome editing using synthesis-dependent repair of Cas9-induced DNA breaks. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E10745–E10754
38. Perez-Riverol, Y., Bai, J., Bandla, C., Garcia-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S., *et al.* (2022) The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* **50**, D543–D552



Qian Liang earned his PhD from Northeast Normal University and conducted his postdoc training in Tsinghua University and University of Colorado Boulder. Dr Liang is currently a faculty member in Key Laboratory of Aging and Cancer Biology of Zhejiang Province, Hangzhou Normal University School of Basic Medical Sciences. His research is focusing on the functions and regulatory mechanisms of UFMylation, a recently identified ubiquitin-like posttranslational modification with important biological functions.