An evolutionarily conserved mechanism for control the translation of long proteins

Xueao Zheng^{1,2,#}, Hanchen Chen^{1,2,#}, Zhiping Deng³, Yujing Wu^{1,2}, Chong Wu^{1,2}, Linlin Zhong⁴, Shunping Yan^{1,2,5,*}

¹Hubei Hongshan Laboratory, Wuhan, 430070, China

²College of Life Science and Technology, Huazhong Agricultural University, Wuhan, Hubei 430070, China

³State Key Laboratory for Managing Biotic and Chemical Threats to the Quality and Safety of Agro-products, Institute of Virology and Biotechnology, Zhejiang Academy of Agricultural Sciences, Hangzhou, Zhejiang 310021, China.

⁴Key Laboratory of Horticultural Plant Biology, Ministry of Education, College of Horticulture and Forestry Sciences, Huazhong Agricultural University, Wuhan 430070, China

These authors contribute equally.

⁵ Lead contact

* Correspondence: spy@mail.hzau.edu.cn

21 SUMMARY

22 Length is one of the key features of mRNA, varying from dozens to thousands
 23 of nucleotides. However, the contribution of length to proteome and its
 24 underlying mechanism are largely unknown. Here we show that the translation
 25 of long mRNA is regulated by the evolutionarily conserved mcm⁵s²U
 26 modification of tRNA. Loss of mcm⁵s²U modification results in decreased
 27 expression of long proteins in Arabidopsis, budding yeast, and humans.
 28 Mechanistically, the mcm⁵s²U modification facilitates ribosome loading on long
 29 mRNAs. Interestingly, the mcm⁵s²U modification is dynamically modulated to
 30 maintain long protein homeostasis during stress responses and cancer
 31 metastasis. Furthermore, gene ontology analysis reveals that long proteins are
 32 enriched in various conserved biological processes including immunity and
 33 DNA repair. Our study suggests that tRNA modification is a general
 34 mechanism for control the translation of long proteins and highlights the
 35 importance of mRNA length in shaping proteome.

36
 37 **Keywords:** tRNA modification, mRNA length, translation, immunity, DNA
 38 repair

40 Highlight

- 41 ● mcm⁵s²U of tRNA is required for immunity and DNA repair
- 42 ● mcm⁵s²U of tRNA is required for long protein expression
- 43 ● mcm⁵s²U facilitate ribosome loading on long mRNAs
- 44 ● Long proteins are involved in various conserved biological processes

47 INTRODUCTION

48 Translation is a complex process involving initiation, elongation, and
 49 termination. As the template of translation, the messenger RNA (mRNA) varies
 50 in several parameters including abundance, length, codon composition,
 51 primary structure, secondary structure, modification, and localization, among
 52 which the mRNA abundance was considered to be a major factor in
 53 determining the protein abundance (Buccitelli and Selbach, 2020; Ponnala et
 54 al., 2014; Schwanhüsser et al., 2011; Vogel and Marcotte, 2012). The primary
 55 and secondary structures of mRNA (i.e. 5' cap, polyA tail, upstream
 56 open-reading-frame (uORF), internal ribosome entry site (IRES),
 57 circularization, and hairpin) was believed to mainly regulate translation (Aylett
 58 and Ban, 2017; Chappell et al., 2000; Choe et al., 2018; Pestova et al., 2001;
 59 Sonenberg and Hinnebusch, 2009). Codon composition (i.e. codon usage, 5'
 60 first 50 codons, and adjacent codons) was reported to mainly regulate
 61 translation elongation (Gamble et al., 2016; Tuller et al., 2010). More and more
 62 studies also revealed that mRNA modification and mRNA localization are
 63 important factors to regulate translation (Boeynaems et al., 2018; Molliex et al.,
 64 2015). mRNA length varies considerably from dozens to thousands of
 65 nucleotides and thus is a basic characteristic of mRNA. However, much less is
 66 known about the relationship between mRNA length and translation.

67 During translation, the code information of mRNA is decoded by transfer
 68 RNA (tRNA) molecules, which carry different amino acids. In this sense, the
 69 tRNA molecules function as deliverers of the building blocks for translation.
 70 The decoding efficiency of tRNA is affected by tRNA abundance, tRNA
 71 modification, aminoacyl-tRNA synthetase, amino acid abundance, and
 72 elongation factors, among which tRNA modification is emerging as a key
 73 regulator during elongation (Rapino et al., 2017; Schaffrath and Leidel, 2017;
 74 Torres et al., 2014). Currently, more than 150 different tRNA modifications have
 75 been identified (Agris et al., 2018). Among them, the
 76 5-methoxycarbonylmethyl-2-thiouridine of uridine at wobble nucleotide

(mcm⁵s²U) is highly conserved in all eukaryotes. The mcm⁵s²U modification is present in tRNA-Lys(UUU), tRNA-Gln(UUG), and tRNA-Glu(UUC) (Huang et al., 2005; Lu et al., 2005; Sen and Ghosh, 1976). In budding yeast (*Saccharomyces cerevisiae*), the 5-methoxycarbonylmethyl of uridine (mcm⁵U) was catalyzed by the Elongator Protein Complex (ELP) and Trm9/112 complex, and the thiolation (s²U) was mediated by the URM1 pathway involving URM1, UBA4, NCS2, and NCS6 (Leidel et al., 2009; Nakai et al., 2004; Noma et al., 2009; Zabel et al., 2008). Loss of the mcm⁵s²U modification causes ribosome pausing at AAA and CAA codons (Nedialkova and Leidel, 2015; Ranjan and Rodnina, 2017; Rezgui et al., 2013), which results in defective co-translational folding of nascent peptides and protein aggregation, thereby eventually disrupting proteome homeostasis (Johansson et al., 2008; Klassen et al., 2020; Koplin et al., 2010; Laxman et al., 2013; Nedialkova and Leidel, 2015; Rezgui et al., 2013; Tavares et al., 2021). In humans, loss of the mcm⁵s²U modification causes numerous disorders including severe developmental defects, neurological diseases, tumorigenesis, and cancer metastasis (Begley et al., 2013; Pan, 2018; Shaheen et al., 2019; Simpson et al., 2009; Torres et al., 2014; Waszak et al., 2020). In plants, loss of the mcm⁵s²U modification was associated with developmental defects and hypersensitivity to heat stress (Leiber et al., 2010; Nakai et al., 2019; Xu et al., 2020). In yeast, mcm⁵s²U modification was reported to regulate cell cycle, DNA damage repair, and abiotic stress responses (Bauer et al., 2012; Jablonowski et al., 2006; Klassen et al., 2016; Leidel et al., 2009; Nedialkova and Leidel, 2015; Zinshteyn and Gilbert, 2013). These studies demonstrated the importance of the mcm⁵s²U modification in development and stress responses. However, it remains unknown whether tRNA modification contributes to the length-dependent translation.

As a sessile organism, plants are frequently infected by different pathogens, which greatly affect plant growth and development and cause a tremendous loss in agriculture. To defend against pathogens, plants are evolved with

sophisticated immune responses, which involve both transcriptional and translational reprogramming to regulate gene expression. Compared to transcriptional reprogramming, the mechanisms of translational reprogramming are far less understood. It was shown that the upstream open reading frames (uORF) regulate the translation of defense genes (Xu et al., 2017). Recently, it was reported that the 2'-O-ribose methylation of tRNA contributes to plant immunity, highlighting the importance of tRNA modification in plant immunity (Ramírez et al., 2018).

In this study, we found that loss of the mcm⁵s²U modification dramatically compromises plant immunity in *Arabidopsis thaliana*. Further studies revealed that the translation of long protein decreases in the absence of the mcm⁵s²U modification, which is conserved in other eukaryotes. We further showed that the mcm⁵s²U modification facilitates ribosome loading on long mRNAs, thereby promote the translation efficiency of long proteins. In addition, we found that long proteins are involved in various conserved biological processes including immunity and DNA repair. Altogether, our study suggests that tRNA modification is a general mechanism for control the translation of long proteins and highlights the importance of mRNA length in shaping proteome.

125

126 Results

127 The mcm⁵s²U modification is required for immune response in 128 Arabidopsis

129 In a study to test disease resistance of transgenic Arabidopsis, we found that
130 one line was very susceptible to *Pseudomonas syringae* pv. *maculicola* (Psm)
131 ES4326, resembling the disease phenotype of *npr1*, in which the master
132 immune regulator NPR1 was mutated (Figure 1A and 1B). We named this line
133 *cgb* (for chao gan bing; “supper susceptible to pathogens” in Chinese). The
134 disease susceptibility of *cgb* was likely due to T-DNA insertion rather than the
135 overexpression of transgene because only one line showed such phenotype.
136 In order to identify the causal gene of *cgb*, we sequenced its genome using the

137 next-generation sequencing technology, which revealed that there was a
138 T-DNA insertion in the third exon of *ROL5* (AT2G44270; Figure S1A and S1B).
139 The insertion was confirmed through genotyping analysis (Figure S1C). In *cgb*
140 mutant, the transcript of *ROL5* was undetectable, indicating that it was a
141 knock-out mutant (Figure S1E). To confirm that *ROL5* was the *CGB* gene, we
142 carried out a complementation experiment by transforming *ROL5* into *cgb*
143 mutant. As shown in Figures 1A and 1B, the disease phenotype of the
144 complementation line (*COM*) was similar to that of wild-type (WT). To further
145 confirm this, we generated another allele of *ROL5* mutant, *rol5-c*, using
146 CRISPR-Cas9 gene-editing approach (Wang et al., 2015). In *rol5-c*, there is a
147 2-bp deletion in the first exon, which caused frameshift. As expected, the *rol5-c*
148 mutant was as susceptible as *cgb* (Figure 1A and 1B). These data strongly
149 suggested that *ROL5* is required for plant immunity.

150 *ROL5* is a homolog of yeast NCS6, which forms a protein complex with
151 NCS2 to catalyze mcm^5s^2U34 (Figure 1C) (Leiber et al., 2010). The NCS2
152 homolog in Arabidopsis was CTU2 (Philipp et al., 2014). To test whether *ROL5*
153 interacts with CTU2, we first performed yeast-two-hybrid assays. As shown in
154 Figure S1D, only when *ROL5* and CTU2 were co-expressed, the yeasts could
155 grow on the selective medium, indicating that *ROL5* interacts with CTU2 in
156 yeast. To test whether they can interact *in vivo*, we carried out split luciferase
157 assays in *Nicotiana benthamiana*. *ROL5* was fused with the N-terminal half of
158 luciferase (nLUC) and CTU2 was fused with the C-terminal half of luciferase
159 (cLUC). An interaction between two proteins brings the two halves of luciferase
160 together, leading to enzymatic activity and production of luminescence that is
161 detectable using a hypersensitive CCD camera. As shown in Figure S1E, the
162 luminescence signal could be detected only when *ROL5*-nLUC and
163 cLUC-CTU2 were co-expressed. To test whether the interaction is direct, we
164 conducted pull-down assays. GST-CTU2 and *ROL5*-His proteins were
165 expressed in *Escherichia coli* and were purified using affinity resins. As shown
166 in Figure S1F, *ROL5*-His could be specifically pulled down by GST-CTU2, but

167 not the GST control, suggesting that ROL5 directly interacts with CTU2.

168 To confirm that ROL5/CTU2 is required for mcm^5s^2U , we measured the
169 mcm^5s^2U levels in WT, *rol5-c*, and *ctu2-1* (SALK_032692) using mass
170 spectrometry. In WT, mcm^5U was almost undetectable (Figure 1D), indicating
171 that mcm^5U is efficiently transformed into mcm^5s^2U in Arabidopsis. On the
172 contrary, mcm^5s^2U was undetectable while mcm^5U was high in the *rol5-c* and
173 *ctu2-1* mutants, suggesting that both ROL5 and CTU2 are required for
174 mcm^5s^2U . Similar to *rol5* mutant, the *ctu2-1* mutant was also susceptible to
175 pathogens (Figure 1A and 1B). These data revealed that ROL5 and CTU2 form
176 a complex to catalyze mcm^5s^2U modification, which is essential for plant
177 immunity.

178

179 **The mcm^5s^2U modification regulates the expression of immune proteins** 180 **in Arabidopsis**

181 To understand how mcm^5s^2U regulates plant immunity, we performed
182 transcriptome and proteome analysis using the *cgb* mutant and the
183 complementation line (COM). Each sample was divided into two parts, one for
184 transcriptome, and the other for proteome. RNA sequencing (RNA-seq)
185 approach and tandem mass tag (TMT)-based approach was used for
186 transcriptome and proteome analysis, respectively. Principal Component
187 Analysis (PCA) showed that the reproducibility of three biological replicates
188 was good (Figure S2A and S2B). Data analysis ($p < 0.05$, |Foldchange| > 1.5
189 for transcriptome, |Foldchange| > 1.2 for proteome) revealed 480 upregulated
190 genes (UGs), 697 downregulated genes (DGs), 789 upregulated proteins
191 (UPs), and 816 downregulated proteins (DPs) in the *cgb* mutant in comparison
192 with WT (Table S1). Venn diagram analysis showed that only 51 DGs
193 overlapped with DPs, and 72 UGs overlapped with UPs (Figure S2C). This
194 result indicated that the correlation between mRNA level and protein level is
195 low, suggesting there is a profound post-transcriptional regulation. To
196 investigate the relationship between the disease phenotype and gene

expression, we performed Gene Ontology (GO) analysis. Unexpectedly, immune response-related GO terms (response to salicylic acid, defense response to fungus, etc.) were significantly enriched in both UGs and Ups (Figure 2A, 2C, and Table S2). This was contradicted by previous studies showing that mutants (e.g. *snc1* and *cpr5*) with higher expression of immune genes were more resistant to pathogens (Zhou and Zhang, 2020). In DGs and DPs, some immune response-related GO terms including response to chitin and regulation of defense response were also significantly enriched (Figure 2B and 2D). This result suggested that the immune response-related gene expression is compromised in the *cgb* mutant.

207

208 **The mcm⁵s²U modification regulates protein expression in a** 209 **length-dependent manner in Arabidopsis**

210 The mcm⁵s²U modification is present in three specific tRNAs, tRNA-Lys (UUU),
211 tRNA-Gln (UUG), and tRNA-Glu (UUC). Therefore, we hypothesized that the
212 counts of their cognate codons (called s² codons in the following), AAA, CAA,
213 and GAA, may affect the protein expression in *cgb* mutant. To test this
214 hypothesis, we performed correlation analysis between protein expression
215 changes (Log₂FoldChange) and counts of each codon (Figure 3A and Table
216 S3). As expected, the counts of all three s² codons were negatively correlated
217 with the protein expression changes, indicating that the genes with more s²
218 codons were more likely to be down-regulated in *cgb* mutant (Figure 3B).
219 Surprisingly, the majority of non-s² codons also showed a negative correlation
220 (Figure 3B). One reasonable explanation is that the protein expression
221 changes negatively correlated with mRNA lengths/protein length. In support of
222 this notion, the counts of all 20 amino acids were negatively correlated with the
223 protein expression changes (Figure 3C). Therefore, we hypothesized that the
224 longer proteins were more likely to be down-regulated in *cgb* mutant. To test
225 this hypothesis, we compared the length of DPs (n = 816) and UPs (n = 789).
226 As shown in Figure 3D, the length of DPs was significantly longer than that of

227 UPs. The average length of DPs was 66.5 longer than that of UPs. To test
228 whether this difference was random, we performed 10,000 random samplings
229 of 816 (defined as UPs) or 789 (defined as UPs) proteins from the proteins
230 identified in our proteome analysis ($n = 7606$). The distribution frequency of
231 length difference between UPs and DPs for each sampling was plotted. As
232 shown in Figure 3E, the chance to obtain the length difference of 66.5 was
233 extremely low ($p < 2.2e-16$). The count difference between UPs and DPs was
234 also significant for both the s^2 codon and the non- s^2 codon (Figure 3F-3I). In
235 addition, we found that the percentages of DPs increased with protein length
236 (Figure 3L), and the proteins longer than 500 aa were more likely to be
237 downregulated (Figure 3M). These results suggested that mcm^5s^2U
238 modification facilitates the expression of long proteins in Arabidopsis.

239

240 **The mcm^5s^2U modification regulates protein expression in a** 241 **length-dependent manner in other organisms**

242 To test whether the mcm^5s^2U modification facilitates the expression of long
243 proteins is conserved in other organisms, we analyzed public proteome data. It
244 was reported that the budding yeast (*Saccharomyces cerevisiae*) *urm1* mutant
245 lacks the mcm^5s^2 modification and changes the proteome profile (Rezgui et al.,
246 2013). Compared with WT, 225 proteins were downregulated and 238 proteins
247 were upregulated in *urm1* ($p < 0.05$, $|\text{Foldchange}| > 1.2$, Table S4). We
248 performed correlation analysis between protein expression changes
249 ($\text{Log}_2\text{FoldChange}$) and counts of each amino acid in each protein. Consistent
250 with the results in Arabidopsis, the counts of most amino acids were negatively
251 correlated with protein expression changes, suggesting that the expression of
252 long proteins is downregulated in *urm1* (Figure 4A). The length of DPs was
253 significantly longer than that of UPs (Figure 4B). The average length of DPs
254 was 91 longer than that of UPs, which was unlikely to be obtained by random
255 sampling ($p < 2.22e-16$, Figure 4C). In addition, we found the percentages of
256 DPs increased with protein length (Figure 4D), and the proteins longer than

500 aa were more likely to be downregulated (Figure 4E), suggesting that protein expression changes were negatively correlated with protein length in the *urm1* mutant. ELP1 is required for mcm⁵U modification. The proteome profile of the human *elp1* mutant was reported (Waszak et al., 2020). Compared with WT, 199 proteins were downregulated and 765 proteins were upregulated in *elp1* ($p < 0.05$, |Foldchange| > 1.2, Table S5). Similar to budding yeast and Arabidopsis, protein expression changes were also negatively correlated with protein length in the *elp1* mutant (Figure 4F-4J). This evidence strongly suggested that length-dependent protein expression mediated by the mcm⁵s²U modification is conserved in eukaryotes.

The mcm⁵s²U modification dynamically regulates protein expression

The above studies revealed that the mcm⁵s²U modification regulates proteome homeostasis. We next sought to study whether the mcm⁵s²U modification regulates the dynamics of the proteome. Previously, it was shown that NCS2 and NCS6 were degraded through 26S proteasome after heat stress in yeast, leading to dynamic change of the mcm⁵s²U modification (Tyagi and Pedrioli, 2015). Therefore, we speculated that the long protein would be downregulated during stress. Thus, we investigated the relationship between protein length and the protein expression changes after heat stress in yeast. Compared with the control condition, 908 proteins were downregulated and 914 proteins were upregulated after heat stress ($p < 0.05$, |Foldchange| > 1.2, Table S6). Correlation analysis revealed that the counts of most amino acids were negatively correlated with protein expression changes, suggesting that the expression of long proteins is downregulated after heat stress (Figure 5A). The length of DPs was significantly longer than that of UPs (Figure 5B). The average length of DPs was 69 longer than that of UPs, which was unlikely to be obtained by random sampling ($p < 2.22e-16$, Figure 5C). These data suggested that decreased level of the mcm⁵s²U modification comprises the expression of long proteins.

287 Next, we wanted to know whether the increased level of the mcm⁵s²U
288 modification promotes the expression of long proteins. Previously, it was
289 reported that the proteins required for mcm⁵s²U were upregulated in cancers,
290 leading to an increased level of mcm⁵s²U. They further found that mcm⁵s²U
291 was required for cancer metastasis (Delaunay et al., 2016). Therefore, we
292 investigated the relationship between protein length and protein expression
293 changes using the proteome data of localized and metastatic prostate cancer
294 tumors (Iglesias-Gato et al., 2018). Compared with the localized tumors, 393
295 proteins were downregulated and 784 proteins were upregulated in metastatic
296 tumors ($p < 0.05$, |Foldchange| > 1.2, Table S7). In contrast to the above
297 situations, correlation analysis revealed that the counts of most amino acids
298 were positively correlated with protein expression changes, suggesting that the
299 expression of long proteins is upregulated in metastatic tumors (Figure 5D).
300 The length of UPs was significantly longer than that of DPs (Figure 5E). The
301 average length of UPs was 189 longer than that of DPs, which was unlikely to
302 be obtained by random sampling ($p < 2.22\text{e-}16$, Figure 5F). These data
303 suggested that the increased level of the mcm⁵s²U modification indeed
304 promotes the expression of long proteins.

305

306 **The mcm⁵s²U modification regulates translation efficiency in a** 307 **length-dependent manner**

308 The finding that the mcm⁵s²U modification facilitates long protein expression
309 suggested that the mcm⁵s²U modification may regulate the translation
310 efficiency (TE) of long proteins. To test this hypothesis, we investigated the
311 relationship between protein length and TE based on the public Ribo-seq data
312 of mcm⁵s²U-deficiency mutants (Nedialkova and Leidel, 2015). TE is referred
313 to the rate of protein production per mRNA and is represented by ribosome
314 footprints per mRNA (Ingolia et al., 2009). Firstly, we analyzed the yeast *elp6*
315 *ncs2* mutant. Compared with WT, the TE of 152 genes were downregulated
316 and 80 genes were upregulated in *elp6 ncs2* ($p < 0.05$, |Foldchange| > 1.5,

Table S8). We performed correlation analysis between the TE changes (Log₂FoldChange) and the counts of each amino acid in each protein, which revealed that the counts of most amino acids were negatively correlated with TE, suggesting that the translation of long protein is downregulated in *elp6 ncs2* (Figure 6A). The length of proteins with reduced TE was significantly longer than the proteins of enhanced TE (Figure 6B). On average, the length of proteins with reduced TE was 255.5 longer than proteins with enhanced TE. The random sampling analysis revealed that this length difference was highly significant ($p < 2.22 \times 10^{-16}$, Figure 6C). In addition, the percentages of proteins with reduced TE increased with protein length (Figure 6D), and the TE of proteins longer than 500 aa were more likely to be downregulated (Figure 6E). Next, we examined the Ribo-seq data of the *Caenorhabditis elegans tut.1 elpc.1* mutant. Compared to WT, the TEs of 1582 genes were downregulated and 865 genes were upregulated in *tut.1 elpc.1* ($p < 0.05$, |Foldchange| > 1.5, Table S9) in *tut.1 elpc.1*. The same analysis was performed and more striking results were obtained (Figure 6F-6J). The length of proteins with reduced TE was 293.5 longer than that with enhanced TE (Figure 6H). Among proteins longer than 800 aa, more than 90% showed reduced TE (Figure 6I). Furthermore, the TEs of proteins longer than 300 aa were more likely to be reduced (Figure 6J). This evidence strongly suggested that mcm⁵s²U regulates TE in a length-dependent manner.

The mcm⁵s²U modification facilitates ribosome loading on long mRNA

To further investigate the effect of mcm⁵s²U modification on TE of long mRNAs, we introduced the concept of regional translation efficiency (rTE) to investigate the details of translation, improving the resolution of TE analysis significantly. We calculated ribosome footprints per 30-bp windows of each mRNA (Figure S3) as rTE in the *elp6 ncs2* mutant and WT budding yeast. The mRNAs were ranked according to their lengths and were divided into 5 groups (Group 1-5). The heatmap showed the rTEs of each mRNA and the dot plot

showed the average of rTEs of each group (Figure 7A-7C). The overall rTE was similar between *elp6 ncs2* and WT (Figure 7A and 7B). However, the rTE of the *elp6 ncs2* mutant was significantly lower than WT in Group 1 ($p < 2.22e-16$, Figure 7C), which indicated the ribosome load of long mRNAs is significantly reduced in *elp6 ncs2*. Previous studies have shown that the first 30-50 codons are very important for ribosome loading in translation (Ingolia et al., 2009; Tuller et al., 2010). To further examine ribosome loading of mRNAs with different length, we analyzed the first 150-bp rTE of in different groups (Figure 7D). Interestingly, the first 150-bp rTE difference between the *elp6 ncs2* mutant and WT was more dramatical in Group 1 (the long mRNAs group) than in the other groups, suggested that the mcm^5s^2U modification facilitates ribosome loading on long mRNAs. Notably, the correlation analysis between the rTE of the first 150-bp and mRNA length revealed a positive correlation in both the *elp6 ncs2* mutant and WT (Figure 7E and 7F). The correlation in the *elp6 ncs2* mutant was a little bit weaker than that in WT. These results suggested that the mcm^5s^2U modification facilitates ribosome loading on long mRNAs.

It has been shown that loss of the mcm^5s^2U modification causes ribosome pausing (Nedialkova and Leidel, 2015; Ranjan and Rodnina, 2017; Rezgui et al., 2013). To test whether the pausing events are related to mRNA length, we determined the pausing score of each codon in different protein groups. Consistent with previous results, we found that the pausing events occur mainly at AAA and CAA (Figure 7G). Interestingly, we found that the pausing score of AAA, but not CAA, was positively correlated with the mRNA length, indicating that AAA plays a more important role in determining the translation of long mRNAs.

373

374 **Long proteins are involved in various important biological processes**

375 The relationship between mcm^5s^2U modification and long protein expression
376 prompted us to examine the roles of long proteins in different organisms. To

377 this end, we carried out gene ontology (GO) analysis of long proteins in
378 budding yeast, Arabidopsis, and humans. In each organism, we subjected the
379 top 5% of long proteins to GO analysis (Figure 8A-8C, Table S10). Consistent
380 with the role of the mcm⁵s²U modification in plant immunity, the immune
381 response was significantly enriched in Arabidopsis (Figure 8A). Consistent
382 with the role of the mcm⁵s²U modification in cancer metastasis, the WNT
383 signaling pathway was significantly enriched in human (Figure 8B).
384 Interestingly, Arabidopsis, budding yeast, and human shared many enriched
385 GO terms including DNA methylation, cell cycle, DNA repair, RNA/protein
386 transport, calcium ion transport, and vesicle trafficking (Figure 8D and Figure
387 S4), indicating that the long proteins are involved in various conserved
388 biological processes.

389 Notably, the GO terms related to DNA repair were enriched in all
390 organisms, which indicated that the mcm⁵s²U modification plays important role
391 in DNA repair. To test this possibility, we treated the *rol5* and *ctu2* mutants with
392 bleomycin (BLM) and Camptothecin (CPT), which cause DNA double-strand
393 breaks and inhibit root growth. Compared with WT, the root lengths of the *rol5*
394 and *ctu2* mutants were similar on the control medium but were significantly
395 shorter on the medium containing BLM and CPT, suggesting that ROL5 and
396 CTU2 are required for efficient DNA repair in plants (Figure 8E-8G).

397

398 Discussion

399 Based on our data, we proposed a simplified model to illustrate the relationship
400 between the mcm⁵s²U modification and protein expression (Figure 8H). When
401 the mcm⁵s²U modification level is high, both the short and long proteins are
402 translated efficiently to maintain proteome homeostasis. When the mcm⁵s²U
403 modification level is low, the translation of long proteins reduces more
404 dramatically than that of short proteins, resulting in proteome imbalance, which
405 affects various important biological processes including immunity, DNA repair,
406 cancer metastasis, and heat tolerance.

Length is one of the key features of mRNA. However, the relationship between mRNA length and translation is largely unknown. Our study found that the mcm⁵s²U modification regulates protein expression in a length-dependent manner, suggesting that mRNA length is an important factor in shaping proteome. Furthermore, we have shown that long proteins are involved in many important and conserved biological processes such as DNA repair and immunity (Figure 8A-8D and Table S10). In humans, long proteins are enriched in neurogenesis and cancer metastasis. And long proteins were upregulated in metastatic tumors compared with localized tumors (Figure 5D-5F), suggesting that protein length may be used as a hallmark of cancer malignancy. This evidence highlighted the importance of long proteins and indicated that the length of proteins is an important signature of proteomes. Therefore, our study revealed the previously underappreciated contribution of mRNA length in translation, which will encourage a new research direction.

Mechanistically, we surprisingly found that the mcm⁵s²U modification facilitates ribosome loading on long mRNAs (Figure 7D-7F). It was shown that mcm⁵s²U deficiency causes ribosome pausing (Nedialkova and Leidel, 2015; Ranjan and Rodnina, 2017; Rezgui et al., 2013), which may prolong the decoding time. Given that it takes more time for long mRNAs to be decoded than short mRNAs and the ribosome pausing is more server for long mRNAs (Figure 7G), the recycling time of ribosomes on long mRNA is significantly increased in the absence of the mcm⁵s²U modification, thereby inhibiting ribosome loading. In this sense, the mcm⁵s²U modification regulates both elongation and initiation events, especially for long mRNAs.

The mcm⁵s²U modification of tRNA is highly conserved in eukaryotes. The role of mcm⁵s²U modification in translation was reported in several studies (Johansson et al., 2008; Klassen et al., 2020; Koplin et al., 2010; Laxman et al., 2013; Nedialkova and Leidel, 2015; Rezgui et al., 2013; Tavares et al., 2021). In addition to the mcm⁵s²U modification, tRNA contains many other types of modifications such as the m⁵C modification (Blaze et al., 2021). To test

whether the m⁵C modification has similar functions in regulating the expression of long proteins, we analyzed the proteome data of the m⁵C mutant. Indeed, we found that the expression of long proteins was also reduced in the m⁵C mutant (Figure S5), suggesting that tRNA modification may be a general mechanism in controlling the translation of long proteins.

In this study, we introduced the concept of rTE (Figure 7 and Figure S3) to investigate the details of translation. Compared with traditional TE analysis, which is based on the ribosome footprints per mRNA, rTE improves the resolution of TE analysis significantly. The rTE analysis in this study help to reveal that the mcm⁵s²U modification facilitates ribosome loading on long mRNAs. We developed a script to analyze rTE and believed it will be very helpful for other translation studies.

Plants control immune responses precisely to ensure that they are not activated in the absence of pathogens but can be quickly and efficiently activated upon pathogen infection (Spoel and Dong, 2012; Yan and Dong, 2014; Zhang et al., 2020; Zhou and Zhang, 2020). Previous studies on plant immunity mainly focus on transcriptional regulation of gene expression. Here, we found that the mcm⁵s²U modification is required for plant immunity (Figure 1A and 1B) and regulates immune gene expression at the translational level. Therefore, our study revealed a new mechanism of plant immunity, providing a new strategy to enhance plant resistance to pathogens. Interestingly, GO analysis revealed that long proteins were also enriched in immune response in humans (Figure 8C), suggesting that the mcm⁵s²U modification may also contribute to human immunity, which is worthwhile testing in the future study.

461

462 **STAR METHOD**

463 Detailed methods are provided in the online version of this paper and include
464 the following:

- 465 ● KEY RESOURCES TABLE
- 466 ● RESOURCE AVAILABILITY

467 ✧ Lead Contact

468 ✧ Materials Availability

469 ✧ Data and Code Availability

470 ● EXPERIMENTAL MODEL AND SUBJECT DETAILS

471 ✧ Plant material and growth conditions

472 ✧ Strains and growth conditions

473 ● METHOD DETAILS

474 ✧ Yeast-two-hybrid assays

475 ✧ *In vitro* pull-down assays

476 ✧ Quantification of tRNA modifications

477 ✧ RNA and protein extraction for transcriptome and proteome analysis

478 ✧ Proteome analysis

479 ● QUANTIFICATION AND STATISTICAL ANALYSIS

480 ✧ Proteome data analysis

481 ✧ Transcriptome data analysis

482 ✧ Ribo-seq analysis

483 ✧ Calculation of codon or amino acid count

484 ✧ Random sampling analysis

485 ✧ Gene Ontology and functional category analyses

486 ✧ Accession Numbers

487

488 **SUPPLEMENTAL INFORMATION**

489 Supplemental Information includes 6 figures and 11 tables and can be found
490 with this article online.

491

492 **ACKNOWLEDGMENTS**

493 We are grateful to Dr. Zhipeng Zhou, Dr. Peng Chen, and Dr. Junkang He for
494 helpful suggestions and discussion. This work is supported by the National
495 Natural Science Foundation of China (31970311), Thousand Talents Plan of
496 China-Young Professionals, Longyun Program and BaiChuan Program from

497 College of Life Science and Technology, Huazhong Agricultural University.

498

499 **AUTHOR CONTRIBUTIONS**

500 S.Y., and X.Z. designed the project. H.C. and C.W. identified ROL5. H.C. found
501 the interaction between ROL5 and CTU2. X.Z. and L.Z. performed
502 quantification of mcm⁵s²U. H.C. and Y.W. performed pathogen infection assay.
503 X.Z. performed RNA-seq assay. Z.D. and X.Z. performed TMT-based
504 proteome assay. X.Z. performed bioinformatic analysis and found the
505 relationship between length and translation. X.Z. and S.Y. wrote the
506 manuscript with inputs from others.

507

508 **DECLARATION OF INTERESTS**

509 The authors declare no competing financial interests.

510

511 **REFERENCES**

512 Aylett, C.H.S., and Ban, N. (2017). Eukaryotic aspects of translation initiation
513 brought into focus. *Philosophical Transactions of the Royal Society B:*
514 *Biological Sciences* 372, 1471–2970.
515 Bauer, F., Matsuyama, A., Candiracci, J., Dieu, M., Scheliga, J., Wolf, D.A.,
516 Yoshida, M., and Hermand, D. (2012). Translational control of cell division by
517 elongator. *Cell Reports* 1, 424–433.
518 Begley, U., Sosa, M.S., Avivar-Valderas, A., Patil, A., Endres, L., Estrada, Y.,
519 Chan, C.T.Y., Su, D., Dedon, P.C., Aguirre-Ghiso, J.A., et al. (2013). A human
520 tRNA methyltransferase 9-like protein prevents tumour growth by regulating
521 LIN9 and HIF1- α . *EMBO Molecular Medicine* 5, 366–383.
522 Blaze, J., Navickas, A., Phillips, H.L., Heissel, S., Plaza-Jennings, A., Miglani,
523 S., Asgharian, H., Foo, M., Katanski, C.D., Watkins, C.P., et al. (2021).
524 Neuronal Nsun2 deficiency produces tRNA epitranscriptomic alterations and
525 proteomic shifts impacting synaptic signaling and behavior. *Nature*
526 *Communications* 12, 4913.

527 Boeynaems, S., Alberti, S., Fawzi, N.L., Mittag, T., Polymenidou, M., Rousseau,
528 F., Schymkowitz, J., Shorter, J., Wolozin, B., van den Bosch, L., et al. (2018).
529 Protein phase separation: A new phase in cell biology. *Trends in Cell Biology* 28,
530 420–435.

531 Buccitelli, C., and Selbach, M. (2020). mRNAs, proteins and the emerging
532 principles of gene expression control. *Nature Reviews Genetics* 21, 630–644.

533 Chappell, S.A., Edelman, G.M., and Mauro, V.P. (2000). A 9-nt segment of a
534 cellular mRNA can function as an internal ribosome entry site (IRES) and when
535 present in linked multiple copies greatly enhances IRES activity. *Proceedings of*
536 *the National Academy of Sciences of the United States of America* 97,
537 1536–1541.

538 Choe, J., Lin, S., Zhang, W., Liu, Q., Wang, L., Ramirez-Moya, J., Du, P., Kim,
539 W., Tang, S., Sliz, P., et al. (2018). mRNA circularization by METTL3–eIF3h
540 enhances translation and promotes oncogenesis. *Nature* 561, 556–560.

541 Delaunay, S., Rapino, F., Tharun, L., Zhou, Z., Heukamp, L., Termaat, M.,
542 Shostak, K., Klevernic, I., Florin, A., Desmecht, H., et al. (2016). Elp3 links
543 tRNA modification to IRES-dependent translation of LEF1 to sustain metastasis
544 in breast cancer. *Journal of Experimental Medicine* 213, 2503–2523.

545 Gamble, C.E., Brule, C.E., Dean, K.M., Fields, S., and Grayhack, E.J. (2016).
546 Adjacent codons act in concert to modulate translation efficiency in yeast. *Cell*
547 166, 679–690.

548 Huang, B., Johansson, M.J.O., and Bystrom, A.S. (2005). An early step in
549 wobble uridine tRNA modification requires the Elongator complex. *RNA* 11,
550 424–436.

551 Iglesias-Gato, D., Thysell, E., Tyanova, S., Crnalic, S., Santos, A., Lima, T.S.,
552 Geiger, T., Cox, J., Widmark, A., Bergh, A., et al. (2018). The proteome of
553 prostate cancer bone metastasis reveals heterogeneity with prognostic
554 implications. *Clinical Cancer Research* 24, 5433–5444.

555 Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., and Weissman, J.S. (2009).
556 Genome-wide analysis in vivo of translation with nucleotide resolution using

557 ribosome profiling. *Science* 324, 218–223.

558 Jablonowski, D., Zink, S., Mehlgarten, C., Daum, G., and Schaffrath, R. (2006).

559 tRNA^{Glu} wobble uridine methylation by Trm9 identifies Elongator's key role for

560 zymocin-induced cell death in yeast. *Molecular Microbiology* 59, 677–688.

561 Johansson, M.J.O., Esberg, A., Huang, B., Björk, G.R., and Byström, A.S.

562 (2008). Eukaryotic wobble uridine modifications promote a functionally

563 redundant decoding system. *Molecular and Cellular Biology* 28, 3301–3312.

564 Klassen, R., Ciftci, A., Funk, J., Bruch, A., Butter, F., and Schaffrath, R. (2016).

565 tRNA anticodon loop modifications ensure protein homeostasis and cell

566 morphogenesis in yeast. *Nucleic Acids Research* 44, 10946–10959.

567 Klassen, R., Bruch, A., and Schaffrath, R. (2020). Induction of protein

568 aggregation and starvation response by tRNA modification defects. *Current*

569 *Genetics* 66, 1053–1057.

570 Koplin, A., Preissler, S., Llina, Y., Koch, M., Scior, A., Erhardt, M., and Deuerling,

571 E. (2010). A dual function for chaperones SSB-RAC and the NAC nascent

572 polypeptide-associated complex on ribosomes. *Journal of Cell Biology* 189,

573 57–68.

574 Laxman, S., Sutter, B.M., Wu, X., Kumar, S., Guo, X., Trudgian, D.C., Mirzaei,

575 H., and Tu, B.P. (2013). Sulfur amino acids regulate translational capacity and

576 metabolic homeostasis through modulation of tRNA thiolation. *Cell* 154,

577 416–429.

578 Leiber, R.M., John, F., Verhertbruggen, Y., Diet, A., Knox, J.P., and Ringli, C.

579 (2010). The TOR pathway modulates the structure of cell walls in Arabidopsis.

580 *Plant Cell* 22, 1898–1908.

581 Leidel, S., Pedrioli, P.G.A., Bucher, T., Brost, R., Costanzo, M., Schmidt, A.,

582 Aebersold, R., Boone, C., Hofmann, K., and Peter, M. (2009). Ubiquitin-related

583 modifier Urm1 acts as a sulphur carrier in thiolation of eukaryotic transfer RNA.

584 *Nature* 458, 228–232.

585 Lu, J., Huang, B.O., Esberg, A., Johansson, M.J.O., and Byström, A.S. (2005).

586 The *Kluyveromyces lactis* γ -toxin targets tRNA anticodons. *RNA* 11,

1648–1654.

Molliex, A., Temirov, J., Lee, J., Coughlin, M., Kanagaraj, A.P., Kim, H.J., Mittag, T., and Taylor, J.P. (2015). Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* 163, 123–133.

Nakai, Y., Umeda, N., Suzuki, T., Nakai, M., Hayashi, H., Watanabe, K., and Kagamiyama, H. (2004). Yeast Nfs1p is involved in Thio-modification of both mitochondrial and cytoplasmic tRNAs. *Journal of Biological Chemistry* 279, 12363–12368.

Nakai, Y., Horiguchi, G., Iwabuchi, K., Harada, A., Nakai, M., Hara-Nishimura, I., and Yano, T. (2019). tRNA wobble modification affects leaf cell development in *Arabidopsis thaliana*. *Plant and Cell Physiology* 60, 2026–2039.

Nedialkova, D.D., and Leidel, S.A. (2015). Optimization of codon translation rates via tRNA modifications maintains proteome integrity. *Cell* 161, 1606–1618.

Noma, A., Sakaguchi, Y., and Suzuki, T. (2009). Mechanistic characterization of the sulfur-relay system for eukaryotic 2-thiouridine biogenesis at tRNA wobble positions. *Nucleic Acids Research* 37, 1335–1352.

Pan, T. (2018). Modifications and functional genomics of human transfer RNA. *Cell Research* 28, 395–404.

Pestova, T. v., Kolupaeva, V.G., Lomakin, I.B., Pilipenko, E. v., Shatsky, I.N., Agol, V.I., and Hellen, C.U.T. (2001). Molecular mechanisms of translation initiation in eukaryotes. *Proceedings of the National Academy of Sciences of the United States of America* 98, 7029–7036.

Philipp, M., John, F., and Ringli, C. (2014). The cytosolic thiouridylase CTU2 of *Arabidopsis thaliana* is essential for posttranscriptional thiolation of tRNAs and influences root development. *BMC Plant Biology* 14, 1–8.

Ponnala, L., Wang, Y., Sun, Q., and van Wijk, K.J. (2014). Correlation of mRNA and protein abundance in the developing maize leaf. *Plant Journal* 78, 424–440.

617 Ramírez, V., González, B., López, A., Castelló, M.J., Gil, M.J., Zheng, B., Chen,
618 P., and Vera, P. (2018). A 2'-O-methyltransferase responsible for transfer RNA
619 anticodon modification is pivotal for resistance to *Pseudomonas syringae*
620 DC3000 in *Arabidopsis*. *Molecular Plant-Microbe Interactions* 31, 1323–1336.

621 Ranjan, N., and Rodnina, M. v. (2017). Thio-modification of tRNA at the wobble
622 position as regulator of the kinetics of decoding and translocation on the
623 ribosome. *Journal of the American Chemical Society* 139, 5857–5864.

624 Rapino, F., Delaunay, S., Zhou, Z., Chariot, A., and Close, P. (2017). tRNA
625 modification: Is cancer having a wobble? *Trends in Cancer* 3, 249–252.

626 Rezgui, V.A.N., Tyagi, K., Ranjan, N., Konevega, A.L., Mittelstaet, J., Rodnina,
627 M. v., Peter, M., and Pedrioli, P.G.A. (2013). tRNA tKUUU, tQUUG, and tEUUC
628 wobble position modifications fine-tune protein translation by promoting
629 ribosome A-site binding. *Proceedings of the National Academy of Sciences of*
630 *the United States of America* 110, 12289–12294.

631 Schaffrath, R., and Leidel, S.A. (2017). Wobble uridine modifications—a reason
632 to live, a reason to die?! *RNA Biology* 14, 1209–1222.

633 Schwanhüsser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J.,
634 Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene
635 expression control. *Nature* 473, 337–342.

636 Sen, G., and Ghosh, H. (1976). Role of modified nucleosides in tRNA: Effect of
637 modification of the 2-thiouridine derivative located at the 5'-end of the anticodon
638 of yeast transfer RNA Lys. *Nucleic Acids Research* 3, 523–536.

639 Shaheen, R., Mark, P., Prevost, C.T., AlKindi, A., Alhag, A., Estwani, F.,
640 Al-Sheddi, T., Alobeid, E., Alenazi, M.M., Ewida, N., et al. (2019). Biallelic
641 variants in CTU2 cause DREAM-PL syndrome and impair thiolation of tRNA
642 wobble U34. *Human Mutation* 40, 2108–2120.

643 Simpson, C.L., Lemmens, R., Miskiewicz, K., Broom, W.J., Hansen, V.K., van
644 Vught, P.W.J., Landers, J.E., Sapp, P., van den Bosch, L., Knight, J., et al.
645 (2009). Variants of the elongator protein 3 (ELP3) gene are associated with
646 motor neuron degeneration. *Human Molecular Genetics* 18, 472–481.

647 Sonenberg, N., and Hinnebusch, A.G. (2009). Regulation of translation
648 initiation in eukaryotes: mechanisms and biological targets. *Cell* 136, 731–745.

649 Spoel, S.H., and Dong, X. (2012). How do plants achieve immunity? Defence
650 without specialized immune cells. *Nature Reviews Immunology* 12, 89–100.

651 Tavares, J.F., Davis, N.K., Poim, A., Reis, A., Kellner, S., Sousa, I., Soares,
652 A.R., Moura, G.M.R., Dedon, P.C., and Santos, M. (2021). tRNA-modifying
653 enzyme mutations induce codon-specific mistranslation and protein
654 aggregation in yeast. *RNA Biology* 18, 563–575.

655 Torres, A.G., Batlle, E., and Ribas de Pouplana, L. (2014). Role of tRNA
656 modifications in human diseases. *Trends in Molecular Medicine* 20, 306–314.

657 Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., Pan, T.,
658 Dahan, O., Furman, I., and Pilpel, Y. (2010). An evolutionarily conserved
659 mechanism for controlling the efficiency of protein translation. *Cell* 141,
660 344–354.

661 Tyagi, K., and Pedrioli, P.G.A. (2015). Protein degradation and dynamic tRNA
662 thiolation fine-tune translation at elevated temperatures. *Nucleic Acids*
663 *Research* 43, 4701–4712.

664 Vogel, C., and Marcotte, E.M. (2012). Insights into the regulation of protein
665 abundance from proteomic and transcriptomic analyses. *Nature Reviews*
666 *Genetics* 13, 227–232.

667 Wang, Z.-P., Xing, H.-L., Dong, L., Zhang, H.-Y., Han, C.-Y., Wang, X.-C., and
668 Chen, Q.-J. (2015). Egg cell-specific promoter-controlled CRISPR/Cas9
669 efficiently generates homozygous mutants for multiple target genes in
670 *Arabidopsis* in a single generation. *Genome Biology* 16, 144.

671 Waszak, S.M., Robinson, G.W., Gudenas, B.L., Smith, K.S., Forget, A., Kojic,
672 M., Garcia-Lopez, J., Hadley, J., Hamilton, K. v., Indersie, E., et al. (2020).
673 Germline Elongator mutations in Sonic Hedgehog medulloblastoma. *Nature*
674 580, 396–401.

675 Xu, G., Greene, G.H., Yoo, H., Liu, L., Marqués, J., Motley, J., and Dong, X.
676 (2017). Global translational reprogramming is a fundamental layer of immune

677 regulation in plants. *Nature* **545**, 487–490.

678 Xu, Y., Zhang, L., Ou, S., Wang, R., Wang, Y., Chu, C., and Yao, S. (2020).

679 Natural variations of SLG1 confer high-temperature tolerance in indica rice.

680 *Nature Communications* **11**, 5441.

681 Yan, S., and Dong, X. (2014). Perception of the plant immune signal salicylic

682 acid. *Current Opinion in Plant Biology* **20**, 64–68.

683 Zabel, R., Bär, C., Mehlgarten, C., and Schaffrath, R. (2008). Yeast α -tubulin

684 suppressor Ats1/Kti13 relates to the Elongator complex and interacts with

685 Elongator partner protein Kti11. *Molecular Microbiology* **69**, 175–187.

686 Zhang, J., Coaker, G., Zhou, J.M., and Dong, X. (2020). Plant immune

687 mechanisms: from reductionistic to holistic points of view. *Molecular Plant* **13**,

688 1358–1378.

689 Zhou, J.M., and Zhang, Y. (2020). Plant immunity: danger perception and

690 signaling. *Cell* **181**, 978–989.

691 Zinshteyn, B., and Gilbert, W. v. (2013). Loss of a conserved tRNA anticodon

692 modification perturbs cellular signaling. *PLoS Genetics* **9**, e1003675.

693

694

FIGURES & FIGURE LEGENDS

Figure 1. The mcm⁵s²U modification is required for plant immunity

(A and B) The *rol5* and *ctu2* mutants are more susceptible to the bacterial pathogen *Psm* ES4326 than wild-type (WT). (A) The photo of Arabidopsis 3 days post-infection. The arrows indicate the leaves inoculated with *Psm* ES4326. *cgb* and *rol5-c* are mutants defective in ROL5. *COM*, the complementation line of *cgb*. *npr1-1* serves as a positive control. (B) The growth of *Psm* ES4326. CFU, colony-forming unit. Error bars represent 95% confidence intervals (n = 6). Statistical significance was determined by two-tailed Student's t-test, ***, p < 0.001; ns, not significant.

706 (C) A schematic diagram showing the function of ROL5 and CTU2. The ROL5
 707 homolog NCS6 and the CTU2 homolog NCS2 form a complex to catalyze the
 708 mcm^5s^2U modification at wobble nucleotide of tRNA-Lys(UUU),
 709 tRNA-Gln(UUC), and tRNA-Glu(UUG), which pair with the AAA, GAA, and
 710 CAA codons in mRNA, respectively.

711 (D) The *rol5* and *ctu2* mutants lack the mcm^5s^2U modification. The levels of U,
 712 cm^5U , mcm^5U , and mcm^5s^2U are quantified through LC-MS analysis. The
 713 intensity and the retention time of each nucleotide are shown. The structure of
 714 each nucleotide and the catalyzing enzymes are shown on the right.

715

Figure 2. The mcm⁵s²U modification regulates immune genes expression in Arabidopsis.

Gene Ontology (GO) analysis of the differentially expressed genes or proteins in the *cgb* mutant. COM, the complementation line. The top 30 enriched GO terms are shown.

Figure 3. The mcm⁵s²U modification regulates protein expression in a length-dependent manner in Arabidopsis

(A) A schematic diagram showing the correlation analysis between the protein expression changes (Log₂FoldChange, *cgb* vs. WT) and the counts of each codon or each amino acid. FC, fold change.

(B and C) The heatmap of the correlation coefficient between the protein expression changes and the counts of each codon (B) or each amino acid (C).

(D, F, and H) The boxplots showing total codon count (D), s² codon count (F), and non-s² codon count (H) in downregulated (Down) and upregulated (Up) proteins. The *p*-values were calculated using the two-sided Student's t-test.

(E, G, and I) Random sampling results showing the difference of total codon

735 count (E), s^2 codon count (G), and non- s^2 codon count (I) between the
 736 downregulated and upregulated proteins. The frequency of count difference
 737 based on 10,000 random samplings is shown. The observed count difference
 738 (arrow) between the downregulated and upregulated proteins and their
 739 probability is shown. The p -values were calculated using two-tailed Student's
 740 t-test.

741 (J) The percentages of downregulated (Down) and upregulated (Up) proteins
 742 with different lengths.

743 (K) The boxplots showing the protein expression changes of the differentially
 744 expressed proteins (*cpg* vs. *COM*) with different lengths.

745

Figure 4. The mcm⁵s²U modification regulates protein expression in a length-dependent manner in budding yeast and human.

(A and F) The heatmap of the correlation coefficient between the protein expression changes (Log₂FoldChange, mutants vs. WT) and the counts of each amino acid in the budding yeast *urm1* mutant (A) or the human *elp1* mutant (F).

(B and G) The boxplots showing the protein length of downregulated (Down) and upregulated (Up) proteins in the budding yeast *urm1* mutant (B) or the human *elp1* mutant (G). The p -values were calculated using the two-sided

756 Mann–Whitney U-test.

757 (C and H) Random sampling results showing the protein length difference
758 between the downregulated and upregulated proteins in the budding yeast
759 *urm1* mutant (G) or the human *elp1* mutant (H). The frequency of length
760 difference based on 10,000 random samplings is shown. The observed length
761 difference (arrow) between the downregulated and upregulated proteins and
762 their probability is shown. The *p*-value was calculated using a two-sided t-test.

763 (D and I) The percentages of downregulated (Down) and upregulated (Up)
764 proteins with different lengths in the budding yeast *urm1* mutant (D) or the
765 human *elp1* mutant (I).

766 (E and J) The boxplots showing the protein expression changes of the
767 differentially expressed proteins with different lengths in the budding yeast
768 *urm1* mutant (E) or the human *elp1* mutant (J).

769

Figure 5. The mcm⁵s² modification dynamically regulates the expression of long proteins.

(A and D) The heatmap of the correlation coefficient between the protein expression changes (Log₂FoldChange) and the counts of each amino acid. (A) In budding yeast (37°C vs. 30°C). (D) In humans (Metastasized tumor vs. localized tumor).

(B and E) The boxplots showing the protein length of downregulated (Down) and upregulated (Up) proteins in budding yeast (B) and humans (E).

(C and F) Random sampling results showing the protein length difference between the downregulated and upregulated proteins in budding yeast (C) and humans (F). The frequency of length difference based on 10,000 random samplings is shown. The observed length difference (arrow) between the downregulated and upregulated proteins and their probability is shown. The *p*-values were calculated using two-tailed Student's t-test.

Figure 6. The mcm⁵s²U modification regulates translation efficiency in a length-dependent manner.

(A and F) The heatmap of the correlation coefficient between the translation efficiency changes (Log₂FoldChange, mutants vs. WT) and the counts of each amino acid in the budding yeast *elp6 ncs2* mutant (A) or the *C. elegans* *tut.1 elpc.1* mutant (F).

(B and G) The boxplots showing the protein length of downregulated (Down) and upregulated (Up) proteins in the budding yeast *elp6 ncs2* mutant (B) or the *C. elegans* *tut.1 elpc.1* mutant (G). The p-values were calculated using the two-sided Mann–Whitney U-test.

797 (C and H) Random sampling results showing the protein length difference
 798 between the downregulated and upregulated proteins in the budding yeast
 799 *elp6 ncs2* mutant (C) or the *C. elegans tut.1 elpc.1* mutant (H). The frequency
 800 of length different based on 10,000 random samplings is shown. The observed
 801 length difference (arrow) between the downregulated and upregulated proteins
 802 and their probability is shown. The *p*-values were calculated using two-tailed
 803 Student's t-test.

804 (D and I) The percentages of downregulated (Down) and upregulated (Up)
 805 proteins with different lengths in the budding yeast *elp6 ncs2* mutant (D) or the
 806 *C. elegans tut.1 elpc.1* mutant (I).

807 (E and J) The boxplots showing the protein expression changes of the
 808 differentially expressed proteins with different lengths in the budding yeast *elp6*
 809 *ncs2* mutant (E) or the *C. elegans tut.1 elpc.1* mutant (J).

810

Figure 7. The mcm⁵s²U modification promotes elongation of long proteins

814 (A and B) The heatmap showing the regional translation efficiency (rTE) in
 815 budding yeast *elp6 ncs2* mutant and WT. The ribosome footprints per 30-bp
 816 windows of each mRNA were calculated. The mRNAs were ranked according
 817 to their lengths and were divided into 5 groups.

818 (C) The dot plot showing the average of rTE in budding yeast *elp6 ncs2* mutant
 819 and WT. The *p*-values were calculated using the paired two-tailed Student's
 820 t-test.

821 (D) The boxplots showing the rTE of the first 150-bp in budding yeast *elp6*
 822 *ncs2* mutant and WT. The *p*-values were calculated using two-tailed Student's
 823 t-test.

824 (E and F) The correlation analysis between mRNA length and the rTE of the
 825 first 150-bp in budding yeast *elp6 ncs2* mutant and WT. The *p*-values were
 826 calculated using two-tailed Student's t-test.

827 (G) The heatmap showing the pausing scores of all codons in budding yeast
 828 (*elp6 ncs2* v.s. WT).

829

Figure 8. Long proteins are involved in important biological processes

(A-C) The enriched biological processes of long proteins in Arabidopsis (A), budding yeast (B), and humans (C). The top 5% long proteins in each organism are submitted for GO analysis.

(D) The bubble blot showing the 34 conserved enriched biological processes of long proteins in Arabidopsis, budding yeast, and humans.

(E-G) The *rol5* and *ctu2* mutants are more sensitive to DNA-damaging

838 reagents bleomycin (BLM) or Camptothecin (CPT) than WT. Plants were
 839 grown vertically on a medium containing 2.5 mM BLM or 20 nM CPT. The
 840 photos (E) and relative root length (F, G) were shown. Data are represented as
 841 means $\pm$ SD (n = 8). Statistical significance was determined by one-tailed
 842 Student's t-test. ***, p < 0.001. Bar = 5 mm.
 843 (H) A working model to illustrate the relationship between the mcm⁵s²U
 844 modification and protein expression.

845
 846