An evolutionarily conserved mechanism for control the translation of long proteins

Xueao Zheng[1,2,#], Hanchen Chen[1,2,#], Zhiping Deng[3], Yujing Wu[1,2], Chong Wu[1,2], Linlin Zhong[4], Shunping Yan[1,2,5,*]

[1]Hubei Hongshan Laboratory, Wuhan, 430070, China

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

[3]State 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.

[4]Key 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: spyan@mail.hzau.edu.cn

SUMMARY

Length is one of the key features of mRNA, varying from dozens to thousands of nucleotides. However, the contribution of length to proteome and its underlying mechanism are largely unknown. Here we show that the translation of long mRNA is regulated by the evolutionarily conserved mcm^5s^2U modification of tRNA. Loss of mcm^5s^2U modification results in decreased expression of long proteins in Arabidopsis, budding yeast, and humans. Mechanistically, the mcm^5s^2U modification facilitates ribosome loading on long mRNAs. Interestingly, the mcm^5s^2U modification is dynamically modulated to maintain long protein homeostasis during stress responses and cancer metastasis. Furthermore, gene ontology analysis reveals that long proteins are enriched in various conserved biological processes including immunity and DNA repair. 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.

Keywords: tRNA modification, mRNA length, translation, immunity, DNA repair

Highlight

- mcm^5s^2U of tRNA is required for immunity and DNA repair
- mcm^5s^2U of tRNA is required for long protein expression
- mcm^5s^2U facilitate ribosome loading on long mRNAs
- Long proteins are involved in various conserved biological processes

INTRODUCTION

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

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

77 (mcm^5s^2U) is highly conserved in all eukaryotes. The mcm^5s^2U modification is
78 present in tRNA-Lys(UUU), tRNA-Gln(UUG), and tRNA-Glu(UUC) (Huang et
79 al., 2005; Lu et al., 2005; Sen and Ghosh, 1976). In budding yeast
80 (*Saccharomyces cerevisiae*), the 5-methoxycarbonylmethyl of uridine (mcm^5U)
81 was catalyzed by the Elongator Protein Complex (ELP) and Trm9/112 complex,
82 and the thiolation (s^2U) was mediated by the URM1 pathway involving URM1,
83 UBA4, NCS2, and NCS6 (Leidel et al., 2009; Nakai et al., 2004; Noma et al.,
84 2009; Zabel et al., 2008). Loss of the mcm^5s^2U modification causes ribosome
85 pausing at AAA and CAA codons (Nedialkova and Leidel, 2015; Ranjan and
86 Rodnina, 2017; Rezgui et al., 2013), which results in defective co-translational
87 folding of nascent peptides and protein aggregation, thereby eventually
88 disrupting proteome homeostasis (Johansson et al., 2008; Klassen et al., 2020;
89 Koplin et al., 2010; Laxman et al., 2013; Nedialkova and Leidel, 2015; Rezgui
90 et al., 2013; Tavares et al., 2021). In humans, loss of the mcm^5s^2U modification
91 causes numerous disorders including severe developmental defects,
92 neurological diseases, tumorigenesis, and cancer metastasis (Begley et al.,
93 2013; Pan, 2018; Shaheen et al., 2019; Simpson et al., 2009; Torres et al.,
94 2014; Waszak et al., 2020). In plants, loss of the mcm^5s^2U modification was
95 associated with developmental defects and hypersensitivity to heat stress
96 (Leiber et al., 2010; Nakai et al., 2019; Xu et al., 2020). In yeast, mcm^5s^2U
97 modification was reported to regulate cell cycle, DNA damage repair, and
98 abiotic stress responses (Bauer et al., 2012; Jablonowski et al., 2006; Klassen
99 et al., 2016; Leidel et al., 2009; Nedialkova and Leidel, 2015; Zinshteyn and
100 Gilbert, 2013). These studies demonstrated the importance of the mcm^5s^2U
101 modification in development and stress responses. However, it remains
102 unknown whether tRNA modification contributes to the length-dependent
103 translation.

104 As a sessile organism, plants are frequently infected by different pathogens,
105 which greatly affect plant growth and development and cause a tremendous
106 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 mcm^5s^2U modification dramatically compromises plant immunity in *Arabidopsis thaliana*. Further studies revealed that the translation of long protein decreases in the absence of the mcm^5s^2U modification, which is conserved in in other eukaryotes. We further showed that the mcm^5s^2U 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.

Results

The mcm^5s^2U modification is required for immune response in Arabidopsis

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

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

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

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

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

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

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

The mcm^5s^2U modification regulates protein expression in a length-dependent manner in Arabidopsis

The mcm^5s^2U modification is present in three specific tRNAs, tRNA-Lys (UUU), tRNA-Gln (UUG), and tRNA-Glu (UUC). Therefore, we hypothesized that the counts of their cognate codons (called s^2 codons in the following), AAA, CAA, and GAA, may affect the protein expression in *cgb* mutant. To test this hypothesis, we performed correlation analysis between protein expression changes (Log_2FoldChange) and counts of each codon (Figure 3A and Table S3). As expected, the counts of all three s^2 codons were negatively correlated with the protein expression changes, indicating that the genes with more s^2 codons were more likely to be down-regulated in *cgb* mutant (Figure 3B). Surprisingly, the majority of non-s^2 codons also showed a negative correlation (Figure 3B). One reasonable explanation is that the protein expression changes negatively correlated with mRNA lengths/protein length. In support of this notion, the counts of all 20 amino acids were negatively correlated with the protein expression changes (Figure 3C). Therefore, we hypothesized that the longer proteins were more likely to be down-regulated in *cgb* mutant. To test this hypothesis, we compared the length of DPs (n = 816) and UPs (n =789). As shown in Figure 3D, the length of DPs was significantly longer than that of

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

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

To test whether the mcm^5s^2U modification facilitates the expression of long proteins is conserved in other organisms, we analyzed public proteome data. It was reported that the budding yeast (*Saccharomyces cerevisiae*) *urm1* mutant lacks the mcm^5s^2 modification and changes the proteome profile (Rezgui et al., 2013). Compared with WT, 225 proteins were downregulated and 238 proteins were upregulated in *urm1* (p < 0.05, |Foldchange| > 1.2, Table S4). We performed correlation analysis between protein expression changes (Log_2FoldChange) and counts of each amino acid in each protein. Consistent with the results in Arabidopsis, the counts of most amino acids were negatively correlated with protein expression changes, suggesting that the expression of long proteins is downregulated in *urm1* (Figure 4A). The length of DPs was significantly longer than that of UPs (Figure 4B). The average length of DPs was 91 longer than that of UPs, which was unlikely to be obtained by random sampling (p < 2.22e−16, Figure 4C). In addition, we found the percentages of 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 mcm^5U 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 mcm^5s^2U modification is conserved in eukaryotes.

The mcm^5s^2U modification dynamically regulates protein expression

The above studies revealed that the mcm^5s^2U modification regulates proteome homeostasis. We next sought to study whether the mcm^5s^2U 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 mcm^5s^2U 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 mcm^5s^2U modification comprises the expression of long proteins.

Next, we wanted to know whether the increased level of the mcm^5s^2U modification promotes the expression of long proteins. Previously, it was reported that the proteins required for mcm^5s^2U were upregulated in cancers, leading to an increased level of mcm^5s^2U. They further found that mcm^5s^2U was required for cancer metastasis (Delaunay et al., 2016). Therefore, we investigated the relationship between protein length and protein expression changes using the proteome data of localized and metastatic prostate cancer tumors (Iglesias-Gato et al., 2018). Compared with the localized tumors, 393 proteins were downregulated and 784 proteins were upregulated in metastatic tumors (p < 0.05, |Foldchange| > 1.2, Table S7). In contrast to the above situations, correlation analysis revealed that the counts of most amino acids were positively correlated with protein expression changes, suggesting that the expression of long proteins is upregulated in metastatic tumors (Figure 5D). The length of UPs was significantly longer than that of DPs (Figure 5E). The average length of UPs was 189 longer than that of DPs, which was unlikely to be obtained by random sampling (p < 2.22e−16, Figure 5F). These data suggested that the increased level of the mcm^5s^2U modification indeed promotes the expression of long proteins.

The mcm^5s^2U modification regulates translation efficiency in a length-dependent manner

The finding that the mcm^5s^2U modification facilitates long protein expression suggested that the mcm^5s^2U modification may regulate the translation efficiency (TE) of long proteins. To test this hypothesis, we investigated the relationship between protein length and TE based on the public Ribo-seq data of mcm^5s^2U-deficiency mutants (Nedialkova and Leidel, 2015). TE is referred to the rate of protein production per mRNA and is represented by ribosome footprints per mRNA (Ingolia et al., 2009). Firstly, we analyzed the yeast *elp6 ncs2* mutant. Compared with WT, the TE of 152 genes were downregulated and 80 genes were upregulated in *elp6 ncs2* (p < 0.05, |Foldchange| > 1.5,

Table S8). We performed correlation analysis between the TE changes (Log_2FoldChange) 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 < 2.22e−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 mcm^5s^2U regulates TE in a length-dependent manner.

The mcm^5s^2U modification facilitates ribosome loading on long mRNA

To further investigate the effect of mcm^5s^2U 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.

Long proteins are involved in various important biological processes

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

this end, we carried out gene ontology (GO) analysis of long proteins in budding yeast, Arabidopsis, and humans. In each organism, we subjected the top 5% of long proteins to GO analysis (Figure 8A-8C, Table S10). Consistent with the role of the mcm^5s^2U modification in plant immunity, the immune response was significantly enriched in Arabidopsis (Figure 8A). Consistent with the role of the mcm^5s^2U modification in cancer metastasis, the WNT signaling pathway was significantly enriched in human (Figure 8B). Interestingly, Arabidopsis, budding yeast, and human shared many enriched GO terms including DNA methylation, cell cycle, DNA repair, RNA/protein transport, calcium ion transport, and vesicle trafficking (Figure 8D and Figure S4), indicating that the long proteins are involved in various conserved biological processes.

Notably, the GO terms related to DNA repair were enriched in all organisms, which indicated that the mcm^5s^2U modification plays important role in DNA repair. To test this possibility, we treated the *rol5* and *ctu2* mutants with bleomycin (BLM) and Camptothecin (CPT), which cause DNA double-strand breaks and inhibit root growth. Compared with WT, the root lengths of the *rol5* and *ctu2* mutants were similar on the control medium but were significantly shorter on the medium containing BLM and CPT, suggesting that ROL5 and CTU2 are required for efficient DNA repair in plants (Figure 8E-8G).

Discussion

Based on our data, we proposed a simplified model to illustrate the relationship between the mcm^5s^2U modification and protein expression (Figure 8H). When the mcm^5s^2U modification level is high, both the short and long proteins are translated efficiently to maintain proteome homeostasis. When the mcm^5s^2U modification level is low, the translation of long proteins reduces more dramatically than that of short proteins, resulting in proteome imbalance, which affects various important biological processes including immunity, DNA repair, 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 mcm^5s^2U 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 mcm^5s^2U modification facilitates ribosome loading on long mRNAs (Figure 7D-7F). It was shown that mcm^5s^2U 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 mcm^5s^2U modification, thereby inhibiting ribosome loading. In this sense, the mcm^5s^2U modification regulates both elongation and initiation events, especially for long mRNAs.

The mcm^5s^2U modification of tRNA is highly conserved in eukaryotes. The role of mcm^5s^2U 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 mcm^5s^2U modification, tRNA contains many other types of modifications such as the m^5C modification (Blaze et al., 2021). To test

whether the m^5C modification has similar functions in regulating the expression of long proteins, we analyzed the proteome data of the m^5C mutant. Indeed, we found that the expression of long proteins was also reduced in the m^5C 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 mcm^5s^2U 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 mcm^5s^2U 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 mcm^5s^2U modification may also contribute to human immunity, which is worthwhile testing in the future study.

STAR METHOD

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY

SUPPLEMENTAL INFORMATION

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

ACKNOWLEDGMENTS

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

College of Life Science and Technology, Huazhong Agricultural University.

AUTHOR CONTRIBUTIONS

S.Y., and X.Z. designed the project. H.C. and C.W. identified ROL5. H.C. found the interaction between ROL5 and CTU2. X.Z. and L.Z. performed quantification of mcm^5s^2U. H.C. and Y.W. performed pathogen infection assay. X.Z. performed RNA-seq assay. Z.D. and X.Z. performed TMT-based proteome assay. X.Z. performed bioinformatic analysis and found the relationship between length and translation. X.Z. and S.Y. wrote the manuscript with inputs from others.

DECLARATION OF INTERESTS

The authors declare no competing financial interests.

REFERENCES

Aylett, C.H.S., and Ban, N. (2017). Eukaryotic aspects of translation initiation brought into focus. Philosophical Transactions of the Royal Society B: Biological Sciences *372*, 1471–2970.

Bauer, F., Matsuyama, A., Candiracci, J., Dieu, M., Scheliga, J., Wolf, D.A., Yoshida, M., and Hermand, D. (2012). Translational control of cell division by elongator. Cell Reports *1*, 424–433.

Begley, U., Sosa, M.S., Avivar-Valderas, A., Patil, A., Endres, L., Estrada, Y., Chan, C.T.Y., Su, D., Dedon, P.C., Aguirre-Ghiso, J.A., et al. (2013). A human tRNA methyltransferase 9-like protein prevents tumour growth by regulating LIN9 and HIF1-α. EMBO Molecular Medicine *5*, 366–383.

Blaze, J., Navickas, A., Phillips, H.L., Heissel, S., Plaza-Jennings, A., Miglani, S., Asgharian, H., Foo, M., Katanski, C.D., Watkins, C.P., et al. (2021). Neuronal Nsun2 deficiency produces tRNA epitranscriptomic alterations and proteomic shifts impacting synaptic signaling and behavior. Nature Communications *12*, 4913.

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

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

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

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

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

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

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

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

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

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

Jablonowski, D., Zink, S., Mehlgarten, C., Daum, G., and Schaffrath, R. (2006). tRNAGlu wobble uridine methylation by Trm9 identifies Elongator's key role for zymocin-induced cell death in yeast. Molecular Microbiology *59*, 677–688.

Johansson, M.J.O., Esberg, A., Huang, B., Björk, G.R., and Byström, A.S. (2008). Eukaryotic wobble uridine modifications promote a functionally redundant decoding system. Molecular and Cellular Biology *28*, 3301–3312.

Klassen, R., Ciftci, A., Funk, J., Bruch, A., Butter, F., and Schaffrath, R. (2016). tRNA anticodon loop modifications ensure protein homeostasis and cell morphogenesis in yeast. Nucleic Acids Research *44*, 10946–10959.

Klassen, R., Bruch, A., and Schaffrath, R. (2020). Induction of protein aggregation and starvation response by tRNA modification defects. Current Genetics *66*, 1053–1057.

Koplin, A., Preissler, S., Llina, Y., Koch, M., Scior, A., Erhardt, M., and Deuerling, E. (2010). A dual function for chaperones SSB-RAC and the NAC nascent polypeptide-associated complex on ribosomes. Journal of Cell Biology *189*, 57–68.

Laxman, S., Sutter, B.M., Wu, X., Kumar, S., Guo, X., Trudgian, D.C., Mirzaei, H., and Tu, B.P. (2013). Sulfur amino acids regulate translational capacity and metabolic homeostasis through modulation of tRNA thiolation. Cell *154*, 416–429.

Leiber, R.M., John, F., Verhertbruggen, Y., Diet, A., Knox, J.P., and Ringli, C. (2010). The TOR pathway modulates the structure of cell walls in Arabidopsis. Plant Cell *22*, 1898–1908.

Leidel, S., Pedrioli, P.G.A., Bucher, T., Brost, R., Costanzo, M., Schmidt, A., Aebersold, R., Boone, C., Hofmann, K., and Peter, M. (2009). Ubiquitin-related modifier Urm1 acts as a sulphur carrier in thiolation of eukaryotic transfer RNA. Nature *458*, 228–232.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Xu, G., Greene, G.H., Yoo, H., Liu, L., Marqués, J., Motley, J., and Dong, X. (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

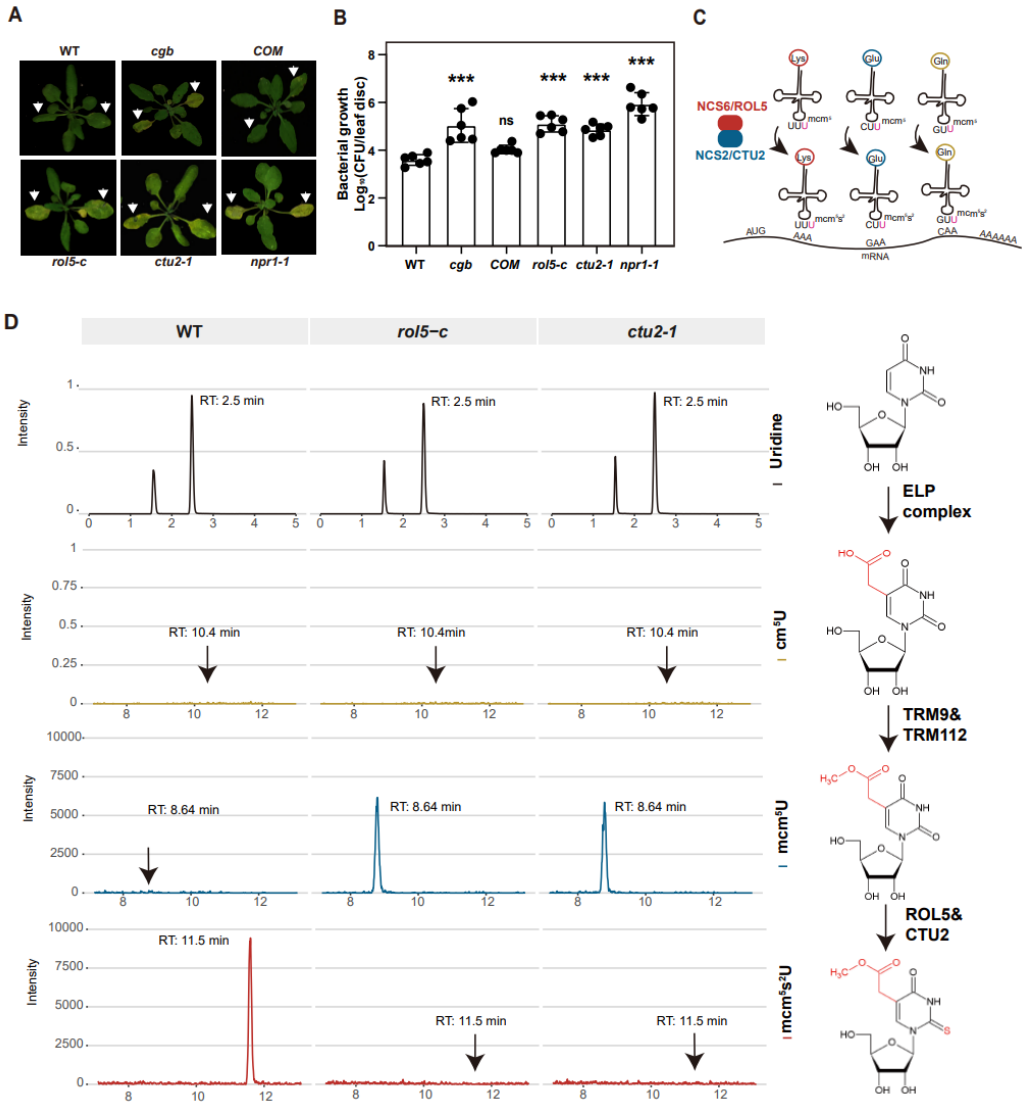


Figure 1. The mcm^{5}s^{2}U modification is required for plant immunity

(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.

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

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

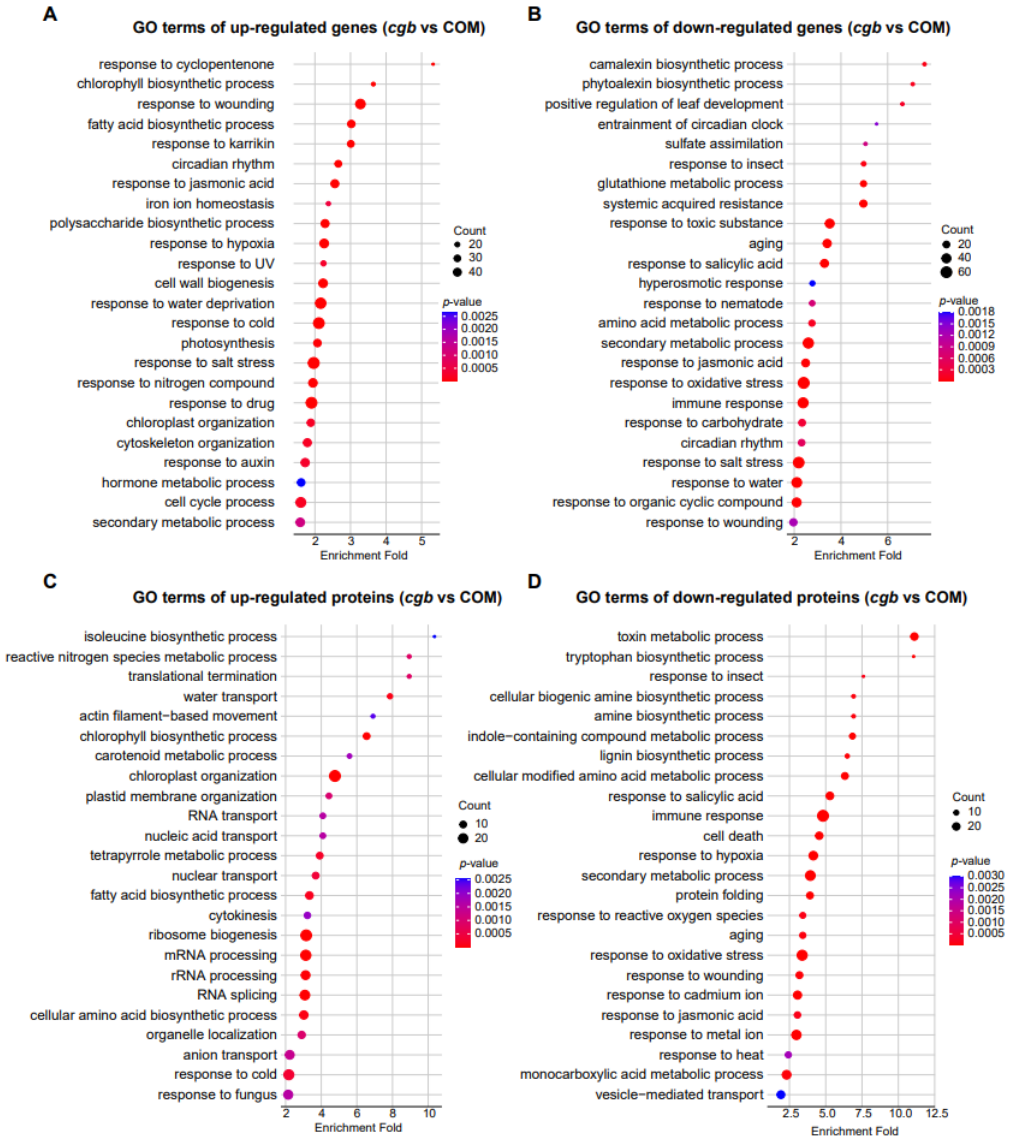
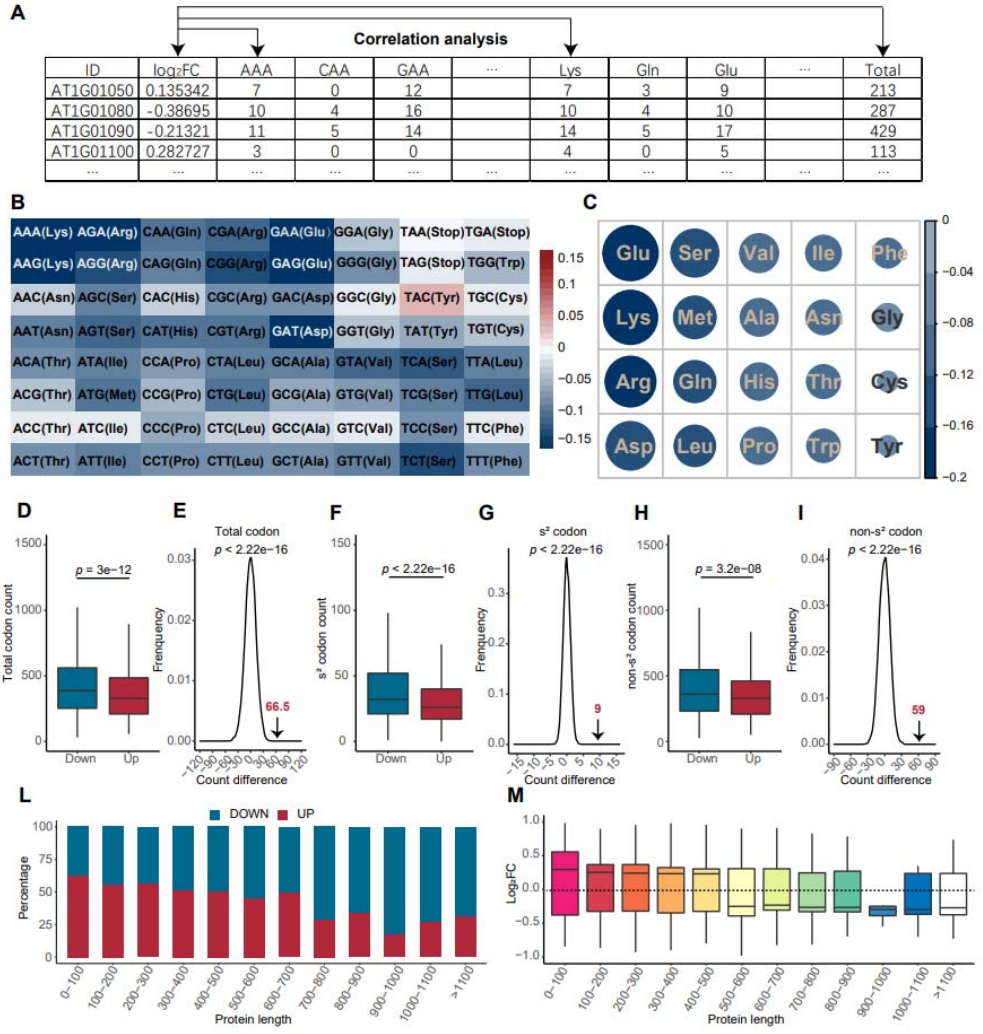


Figure 2. The mcm^{5}s^{2}U modification regulates immune genes expression in Arabidopsis.

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.



723

724 **Figure 3. The mcm^{5}s^{2}U modification regulates protein expression in a**
725 **length-dependent manner in Arabidopsis**

726 (A) A schematic diagram showing the correlation analysis between the protein
727 expression changes (Log_2FoldChange, *cgb* vs. WT) and the counts of each
728 codon or each amino acid. FC, fold change.

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

731 (D, F, and H) The boxplots showing total codon count (D), s^2 codon count (F),
732 and non-s^2 codon count (H) in downregulated (Down) and upregulated (Up)
733 proteins. The *p*-values were calculated using the two-sided Student's t-test.

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

count (E), s^2 codon count (G), and non-s^2 codon count (I) between the downregulated and upregulated proteins. The frequency of count difference based on 10,000 random samplings is shown. The observed count 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.

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

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

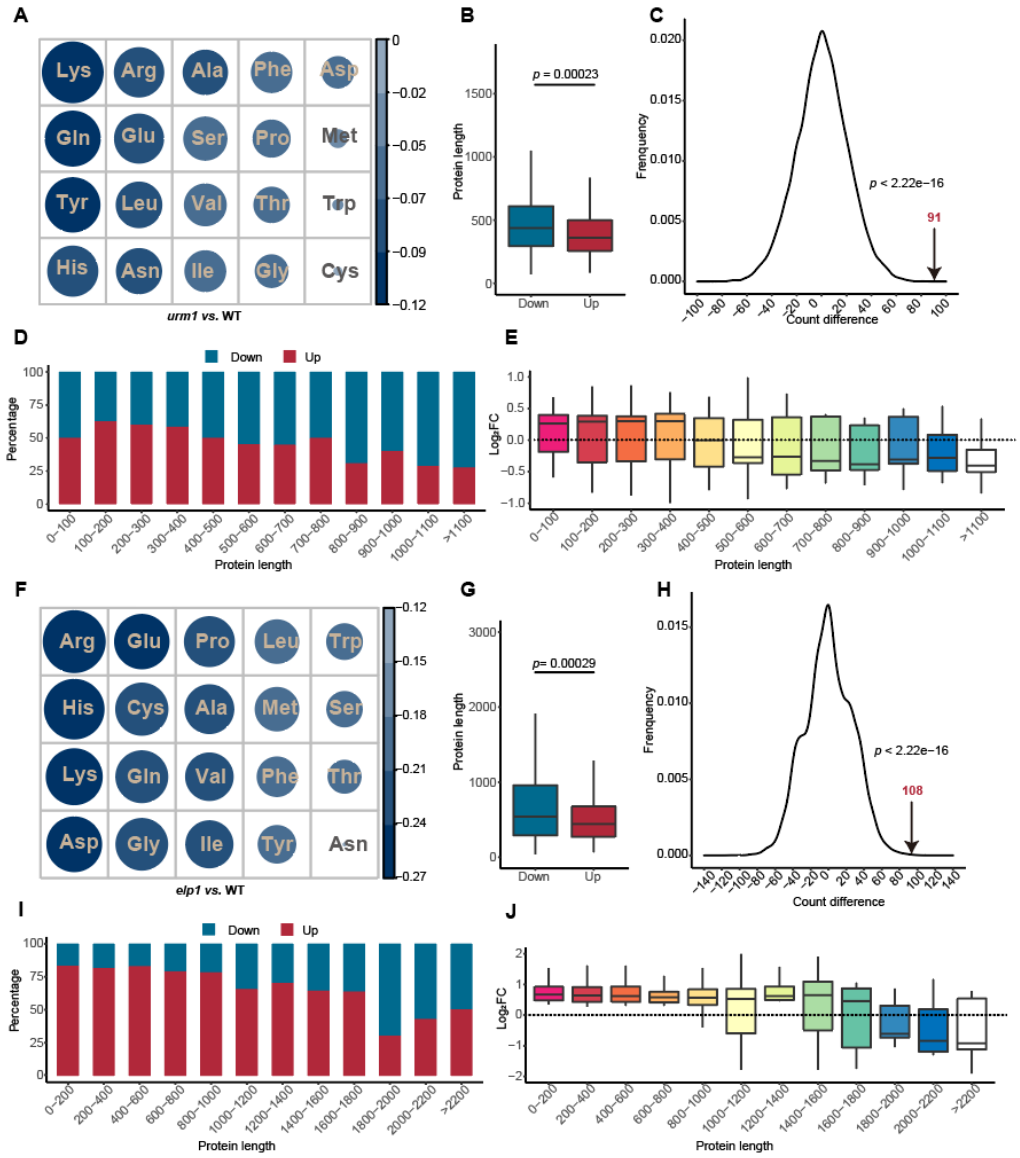


Figure 4. The mcm^{5}s^{2}U modification regulates protein expression in a length-dependent manner in budding yeast and human.

(A and F) The heatmap of the correlation coefficient between the protein expression changes (Log_2FoldChange, 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

Mann–Whitney U-test.

(C and H) Random sampling results showing the protein length difference between the downregulated and upregulated proteins in the budding yeast *urm1* mutant (G) or the human *elp1* mutant (H). 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*-value was calculated using a two-sided t-test.

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

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

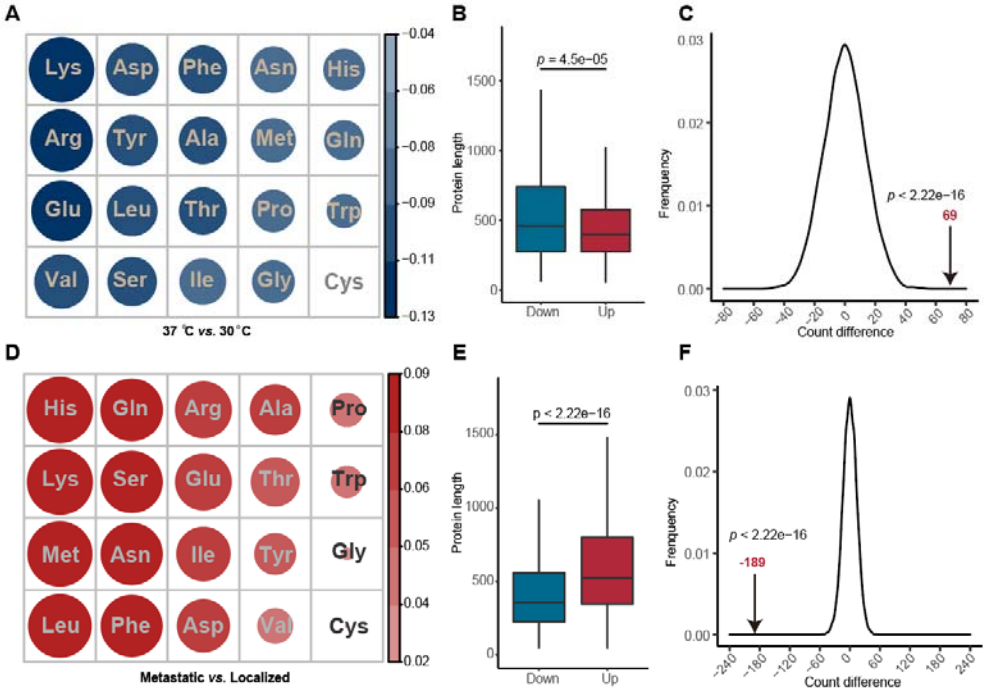


Figure 5. The mcm^{5}s^2 modification dynamically regulates the expression of long proteins.

(A and D) The heatmap of the correlation coefficient between the protein expression changes (Log_2FoldChange) and the counts of each amino acid. (A) In budding yeast (37℃ vs. 30℃). (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.

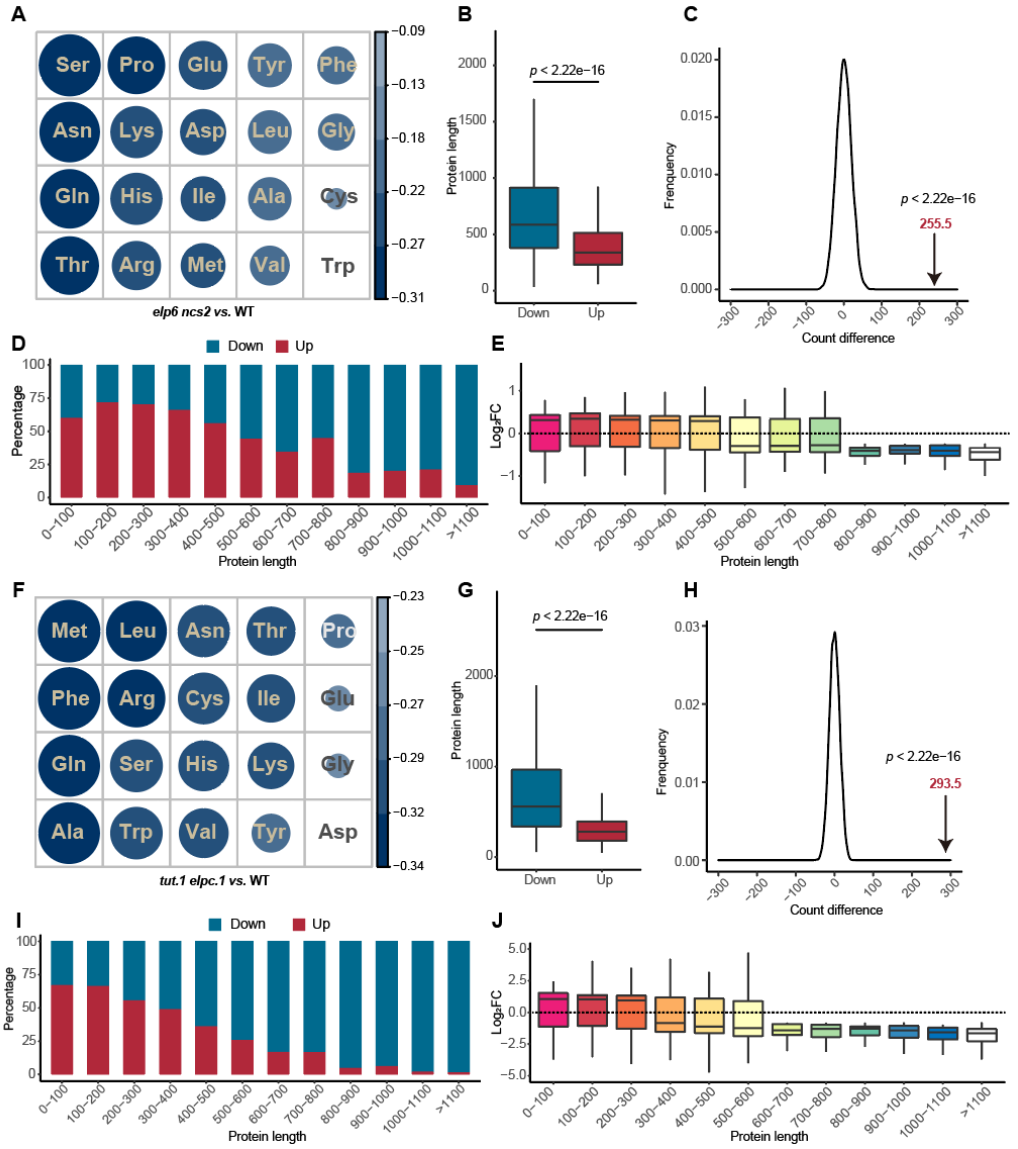


Figure 6. The mcm^{5}s^{2}U modification regulates translation efficiency in a length-dependent manner.

(A and F) The heatmap of the correlation coefficient between the translation efficiency changes (Log_2FoldChange, mutants vs. WT) and the counts of each amino acid in the budding yeast *elp6 ncs2* mutant (A) or the *C. elegant 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. elegant tut.1 elpc.1* mutant (G). The *p*-values were calculated using the two-sided Mann–Whitney U-test.

(C and H) Random sampling results showing the protein length difference between the downregulated and upregulated proteins in the budding yeast *elp6 ncs2* mutant (C) or the *C. elegant tut.1 elpc.1* mutant (H). The frequency of length different 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.

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

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

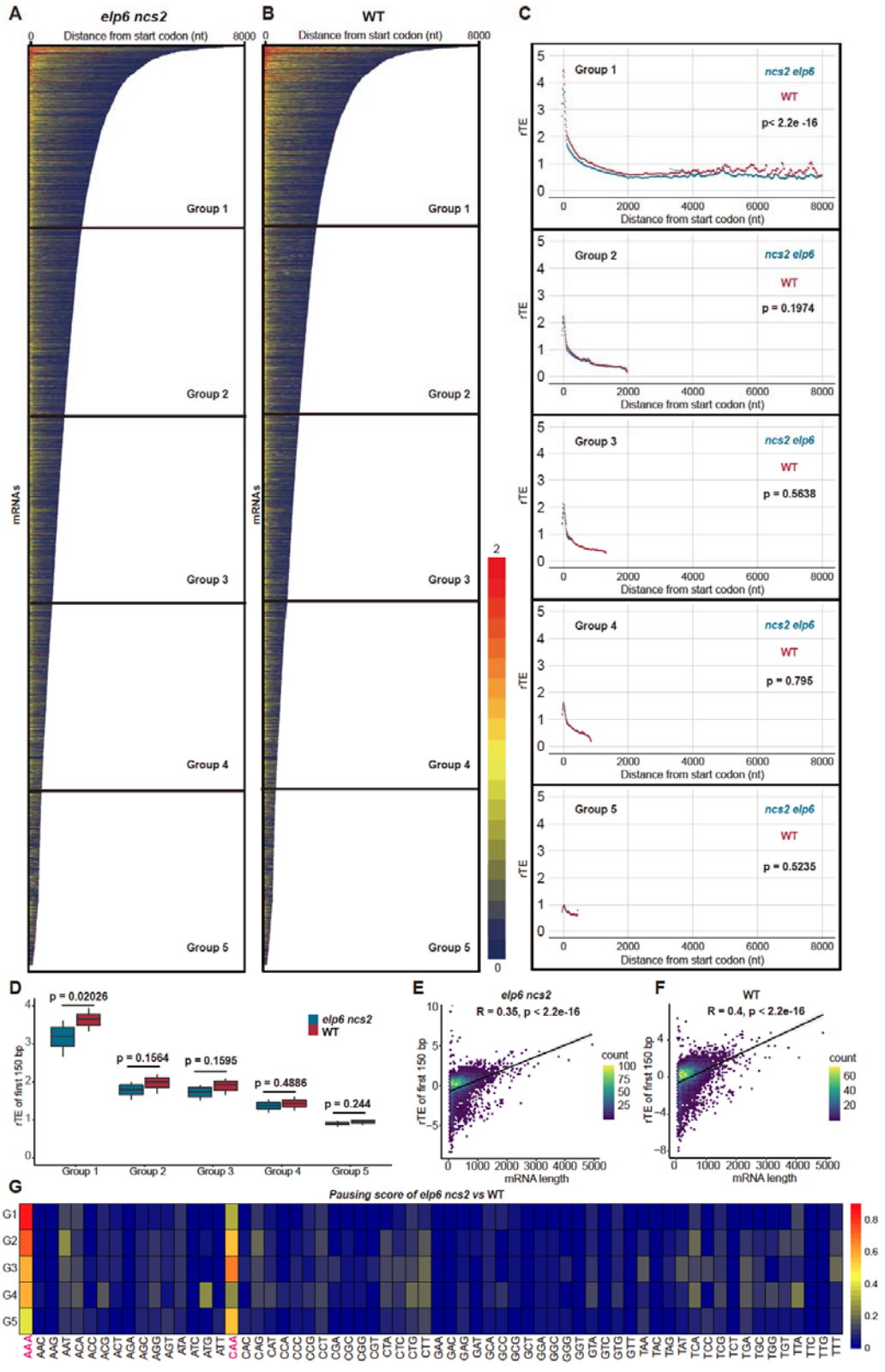


Figure 7. The mcm^{5}s^{2}U modification promotes elongation of long proteins

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

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

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

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

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

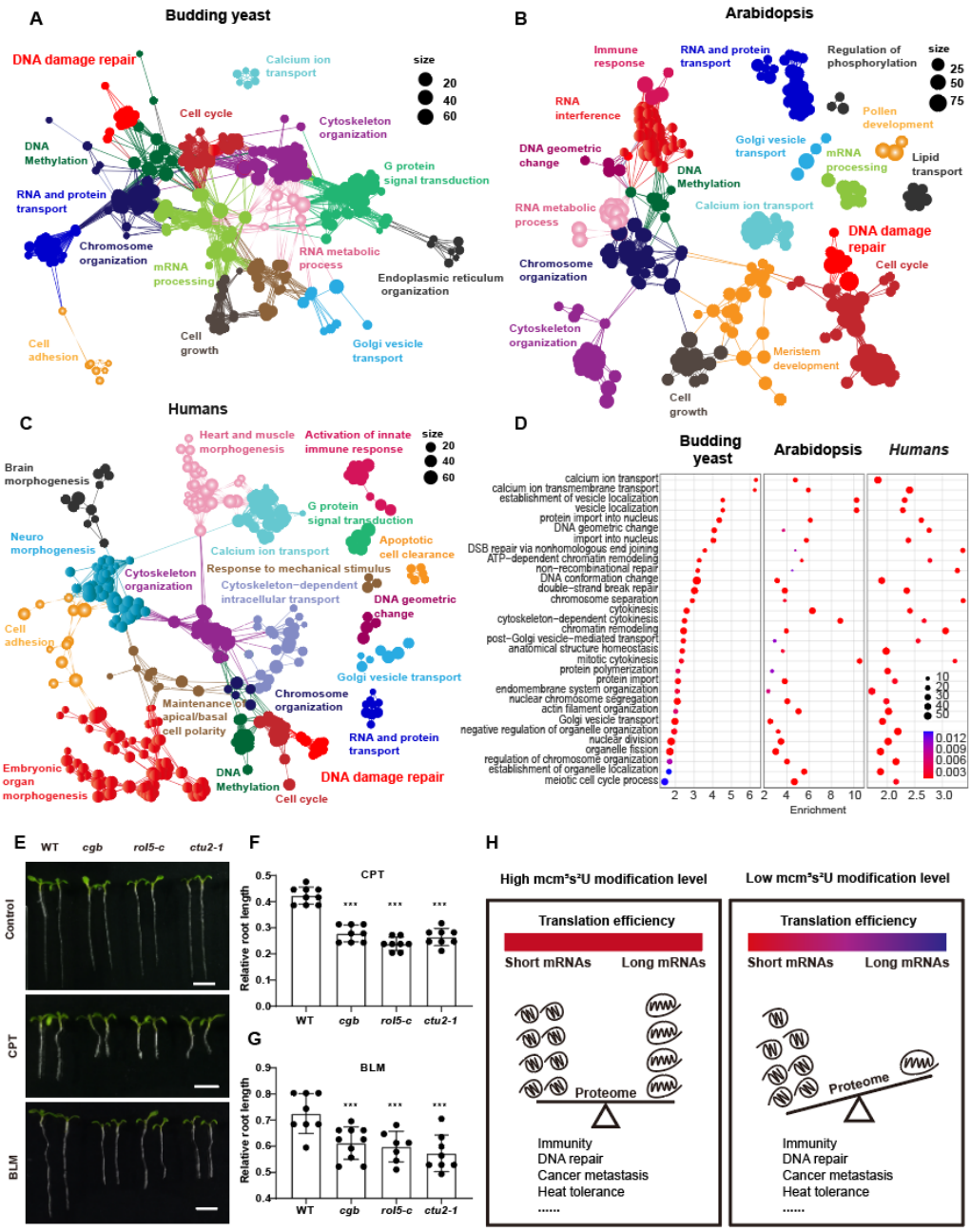


Figure 8. Long proteins are involved in important biological processes

(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

reagents bleomycin (BLM) or Camptothecin (CPT) than WT. Plants were grown vertically on a medium containing 2.5 mM BLM or 20 nM CPT. The photos (E) and relative root length (F, G) were shown. Data are represented as means ±SD (n = 8). Statistical significance was determined by one-tailed Student's t-test. ***, $p < 0.001$. Bar = 5 mm.

(H) A working model to illustrate the relationship between the mcm^5s^2U modification and protein expression.