Widespread readthrough events in plants reveal unprecedented plasticity of stop codons

Yuqian Zhang[1,2,6], Hehuan Li[1,6], Yanting Shen[3,6], Shunxi Wang[1], Lei Tian[1], Haoqiang Yin[1], Jiawei Shi[1], Anqi Xing[4], Jinghua Zhang[1], Usman Ali[1], Abdul Sami[1], Xueyan Chen[1], Chenxuan Gao[1], Yangtao Zhao[1], Yajing Lyu[1], Xiaoxu Wang[1], Yanhui Chen[1], Zhixi Tian[3,5,*], Shu-Biao Wu[2,*], Liuji Wu[1,2,*]

[1] National Key Laboratory of Wheat and Maize Crop Science, Collaborative Innovation Center of Henan Grain Crops, College of Agronomy, Henan Agricultural University, Zhengzhou, China.

[2] School of Environmental and Rural Science, University of New England, Armidale, NSW, Australia.

[3] State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China.

[4] Department of Plant and Environmental Sciences, Clemson University, Clemson, SC, USA.

[5] College of Advanced Agriculture Sciences, University of Chinese Academy of Sciences, Beijing, China.

[6] These authors contributed equally.

***Corresponding authors**

Zhixi Tian (zxtian@genetics.ac.cn), Shu-Biao Wu (shubiao.wu@une.edu.au), and Liuji Wu (wlj200120@163.com)

Abstract

Stop codon readthrough (SCR), the decoding of a stop codon as a sense codon by the ribosome, has important biological implications but remains largely uncharacterized in plants. Here, we identified 1,009 SCR events in two monocots (maize, rice) and two dicots (soybean, *Arabidopsis*) using a proteogenomic strategy with 80 customized databases. SCR transcripts were mostly significantly shorter and had fewer components than non-SCR transcripts in two monocot plants, although these differences were not as significant in the dicots. Mass spectrometry evidence revealed that all three stop codons involved in SCR events could be recoded as 20 standard amino acids, some of which were also supported by suppressor transfer RNA analysis. In addition, we observed multiple functional signals in the C-terminal extensions of 34 maize SCR proteins, and characterized the structural and subcellular localization changes in the extended protein of BASIC TRANSCRIPTION FACTOR 3. Overall, our study not only demonstrates that SCR events are widespread in plants but also reveals the unprecedented recoding plasticity of stop codons, which provides important new insights into the flexibility of genetic decoding.

Introduction

The rules of the genetic code were long considered universal dogma in all organisms after the standard genetic code was deciphered decades ago. However, exceptions to the rules have been increasingly observed in different species over the years, substantially altering the perception of the decoding rules from being static to dynamic[1,2]. Living organisms are now widely understood to employ various modes to expand the coding potential of their genome, such as alternative translation initiation[3], ribosome frameshifting[4], and translational bypassing[5]. Stop codon readthrough (SCR) is another important means of recoding that is classified as a non-canonical translational event[6,7]. SCR occurs when the ribosome reads a stop codon (UGA, UAG, or UAA) as a sense codon rather than the stop signal with the translation terminating at a later in-frame stop codon[7,8]. Such an event can generate a C-terminally extended protein isoform which may acquire novel biological functions, since the amino acid sequence appended at the C-terminus possibly alters the activity[9], stability[10], and/or subcellular localization[2,11] of the protein. Indeed, SCR proteins have been implicated in a variety of biological processes, such as the regulation of intercellular cytoplasm flow[12], neuronal integrity and reproductive health[13], cellular redox homeostasis[2,14], and angiogenesis[15,16].

Initially discovered in viruses as a strategy to expand protein-coding potential[17-19], SCR was subsequently identified in other organisms. In plant pathogens, *Phosphoglycerate kinase* (*PGK*) in *Ustilago maydis* and *Glyceraldehyde-3-phosphate dehydrogenase* (*GAPDH*) in *Aspergillus nidulans* are typical functional readthrough genes conserved in fungal species[2]. In the yeast *Saccharomyces cerevisiae*, the [PSI+] strain forms prion-like polymers of the abnormal eukaryotic release factor 3 (eRF3), which is implicated in translation termination and peptide release, resulting in a high level of readthrough across many genes[20], including *Phosphodiesterase 2* (*PDE2*), *Interacting with mpp10p 3* (*IMP3*), and *Bypass of stop codon 4* (*BSC4*)[10,21,22]. In *Drosophila melanogaster*, readthrough is estimated to affect over 300 genes[23,24], such as *kelch* (*kel*)[12,25], *Synapsin* (*Syn*)[26], *headcase* (*hdc*)[27], and *traffic jam* (*tj*)[13]. In

mammals, *Hemoglobin beta chain* (*HBB*)[28], *Myelin protein zero* (*MPZ*)[29], *Vascular endothelial growth factor-A* (*VEGF-A*)[15], *Mitochondrial carrier 2* (*MTCH2*)[30], *Argonaute 1* (*AGO1*)[31], *Opioid related nociceptin receptor 1* (*OPRL1*)[32], *Mitogen-activated protein kinase 10* (*MAPK10*)[32], *Aquaporin 4* (*AQP4*)[33], *Lactate dehydrogenase B* (*LDHB*)[34], *Malate dehydrogenase 1* (*MDH1*)[34], *Vitamin D receptor* (*VDR*)[35], and *Adenosylmethionine decarboxylase 1* (*AMD1*)[36,37] have been confirmed to undergo readthrough. However, SCR events in plants remain largely uncharacterized, with only one study reporting 144 readthrough genes in *Arabidopsis* detected with ribosome profiling[38]. Therefore, further exploration of SCR events with a more robust analysis is warranted in a wide range of plant species.

It is crucial to identify the amino acids decoded at the stop codon sites, as they can strongly affect the functionality of readthrough protein isoforms[11,21,39]. Previous studies have shown that UGA could be translated to arginine, asparagine, cysteine, glutamine, glycine, serine, threonine, tryptophan, and tyrosine[30,31,39-41]; UAA to alanine, glutamate, glutamine, lysine, serine, tryptophan, and tyrosine[39,42-45]; and UAG to alanine, arginine, aspartate, glutamate, glutamine, glycine, leucine, lysine, serine, threonine, tryptophan, tyrosine, and valine[30,39,46,47]. However, knowledge about the amino acids translated at stop codon sites in SCR events remains unknown in Plantae. In addition, the potential for stop codons to be recoded as amino acids other than those reported above has not been investigated yet.

Here, we employed a mass spectrometry (MS)-based proteogenomic strategy to study SCR events in plants and to identify the amino acids at the readthrough stop codon sites. We detected a total of 1,009 SCR events in the monocots maize and rice, and dicots soybean and *Arabidopsis*, and characterized their features. Interestingly, we discovered that the three stop codons could be recoded as all 20 standard amino acids, which reveals the unprecedented recoding plasticity of stop codons and the dynamic nature of genetic decoding. We also delineated the first comprehensive landscape of SCR events in plants at the protein level, uncovering a collection of previously unknown readthrough proteins to further explore their biological functions.

98 # Results

99 ## A proteogenomic strategy to identify SCR events in plants

100 The SCR event is a means of alternative translation by redefining UGA, UAG, or
101 UAA to specify amino acids, thus extending the translation at the C-terminus. The
102 most definitive evidence for SCR occurrence is the direct detection of translation at
103 the stop codon sites. However, SCR events are non-canonical and cannot be identified
104 by a typical proteomic strategy. Therefore, we implemented a proteogenomic strategy
105 integrating MS-based proteomics with 20 customized databases to identify SCR
106 events in maize at the protein level (Fig. 1). First, total protein was extracted from
107 maize leaves, followed by digestion, fractionation, and liquid chromatography-tandem
108 mass spectrometry (LC-MS/MS) analysis to obtain MS/MS datasets of peptides for
109 subsequent identification. We then constructed 20 customized databases with in-house
110 scripts based on the assumption that the annotated stop codons of genes were readable
111 as sense codons and could be decoded as one of the 20 standard amino acids. Each of
112 the 20 customized databases therefore contained the *in silico*-translated protein
113 isoforms whose annotated stop codon sites were replaced by one of the 20 standard
114 amino acids, and the protein sequences were extended to the subsequent in-frame stop
115 codon.

116 To capture SCR events in maize, we used the obtained MS/MS data to search
117 against the 20 SCR customized databases with MaxQuant search engine. This strategy
118 identified 708 unique "through" peptides spanning the stop codon site. For further
119 quality control of MS data, we next applied more stringent criteria for "through"
120 peptides to identify SCR events with high confidence (see Methods). The screening
121 steps resulted in much more reliable "through" peptides, which corresponded to 367
122 SCR events involving 1,063 proteins (Supplementary Table 1). All identified peptides
123 for SCR proteins are listed in Supplementary Table 2. Collectively, these results not
124 only demonstrate the abundance of readthrough events naturally occurring in maize,
125 but also uncover a previously unknown source of protein isoforms produced by SCR.

126 ## Verifying SCR events in maize

127 To verify the SCR events in maize identified above, we retrieved the available
128 ribosome profiling (Ribo-seq) datasets of 56 maize samples from NCBI to analyze the
129 SCR events at the mRNA level. Four Ribo-seq datasets with higher quality were
130 finally retained after further screening (see Methods). We then analyzed the three-
131 nucleotide (3-nt) periodicity of the Ribo-seq reads (30-mers) in the coding sequences
132 (CDSs), putative extensions, and distal 3′ untranslated regions (UTRs). The results
133 showed that the putative extensions exhibited a visible hallmark of 3-nt periodicity (P
134 < 0.05, Fig. 2a), i.e., the reading frame 0 occurred significantly higher than the other
135 two, which was similar to that of the CDSs (Fig. 2a). By contrast, such difference was
136 not observed in the distal 3′ UTRs (Fig. 2a). In addition, the average ribosome
137 footprint density (in RPKM) of putative extensions was significantly higher than that
138 of the distal 3′ UTRs in the four datasets ($P < 0.001$, Fig. 2b). These results suggest
139 that readthrough is a significant contributor to the ribosome footprint density in
140 putative extensions. After applying the conditions described in the Methods section
141 for the four Ribo-seq datasets to verify SCR, we found that 55 out of the 367 SCR
142 events identified by the proteogenomic approach were also corroborated by Ribo-seq
143 analysis (Fig. 2c and Supplementary Table 3).

144 We selected two SCR representative genes, *ZmH3C2* and *ZmCEQORH*, which
145 exerted functional roles in distinct biological processes related to chromatin assembly
146 and stress response, respectively, for further analysis. For *ZmH3C2*, which encodes
147 HISTONE H3.2, MS analysis identified the readthrough peptide AAAAGRPDRPWR
148 with its second alanine (A) translated at the UAG site (Fig. 2d and Supplementary Fig.
149 1a). The readthrough of this gene was also evidenced by ribosome footprints within
150 the extended region downstream of the annotated stop codon (Fig. 2e). For
151 *ZmCEQORH*, encoding CHLOROPLAST ENVELOPE QUINONE
152 OXIDOREDUCTASE, MS analysis identified the "through" peptide
153 AQEVWAKSIDGHATGKIVVEMGGAAEETR, in which the antepenultimate
154 glutamate (E) was incorporated at the UGA site (Fig. 2h and Supplementary Fig. 1b).

155 We also observed ribosome footprints beyond the annotated stop codon of the
156 corresponding transcript (Fig. 2i).

157 In addition, we performed *in vitro* translation experiments using wheat germ
158 extract to further demonstrate the readthrough of the above two genes. The results
159 showed that *ZmH3C2* and *ZmCEQORH* had significantly higher readthrough
160 activities compared to the SCR-negative gene *Zm00001d009138* ($P < 0.001$), as
161 demonstrated by the detection of luciferase activities. The SCR rates of *ZmH3C2* and
162 *ZmCEQORH* were 12% and 5%, respectively (Fig. 2f, j). It is known that the rate of
163 the basal translation error is lower than 0.1%[8], thus the results indicate that *ZmH3C2*
164 and *ZmCEQORH* underwent the programmed SCR events. Further MS analysis of the
165 readthrough products identified the "through" peptides AAAAGRPDRPWR and
166 GGAAEETR for the two genes, where the UAG and UGA were recoded as alanine
167 and glutamate (Fig. 2g, k), respectively. Taken together, our Ribo-seq analysis, *in vitro*
168 translation experiments, and MS analysis validated the SCR events identified using
169 proteogenomics in maize, and demonstrated the feasibility of the proteogenomic
170 method for large-scale identification of SCR events in plants.

171 Characteristics of SCR events in maize

172 To understand the characteristics of SCR events in maize, we first analyzed the stop
173 codon preferences. All three stop codons were read through among the 367 SCR
174 events: UGA was involved in 174 events (47.4%), followed by UAG in 102 events
175 (27.8%) and UAA in the remaining 91 events (24.8%) (Fig. 3a). We found that UGA
176 was the most frequent next in-frame stop codon following these readthrough stop
177 codons, accounting for 76 cases after the readthrough UGA, 46 cases after the
178 readthrough UAG, and 40 cases after the readthrough UAA (Fig. 3a). The genome-
179 wide distribution pattern showed that these SCR events were distributed
180 heterogeneously across each chromosome of maize, and many of the SCR events were
181 closer to the telomeres (Fig. 3b). In addition, a collection of 379 hotspot regions were
182 obtained, 244 of which were SCR hotspot regions while 135 were non-SCR hotspot
183 regions (Fig. 3b). These results revealed the wide distribution of SCR events in the

184 maize genome.

185 Furthermore, we compared the length and complexity of SCR and non-SCR
186 transcripts. Interestingly, we observed that the SCR primary transcripts were
187 significantly shorter than the non-SCR transcripts (Fig. 3c). The average lengths of
188 the elements from the primary transcripts, including coding sequences (CDSs), introns,
189 and 3′ UTRs, were also markedly shorter in SCR events than those in non-SCR events
190 (Fig. 3c). The 5′ UTR was also shorter in readthrough transcripts, albeit not
191 statistically significant. Consistent with the length of readthrough transcripts, their
192 corresponding proteins had notably shorter lengths with fewer amino acids
193 (Supplementary Fig. 2a) and lower molecular weight (Supplementary Fig. 2b).
194 Moreover, SCR primary transcripts had significantly fewer exons (Fig. 3d) and splice
195 variants (Fig. 3e) than those of non-SCR primary transcripts.

196 Earlier studies have shown that *cis*-acting elements such as nucleotides
197 immediately surrounding the readthrough stop codon and motifs downstream the stop
198 codon can influence readthrough[48-50]. We thus examined the local mRNA context of
199 SCR transcripts. An analysis of nucleotide frequencies in positions adjacent to the
200 readthrough stop codons showed a bias of A>U>C>G at the -2 position, A>G>C>U at
201 the -1 position, A>G>U>C at the +4 position, and U>C>G>A at the +5 position
202 (Supplementary Fig. 3a). The most frequent tetranucleotide pattern (stop codon plus
203 the following nucleotide) was UGA-U (Supplementary Fig. 3b). Extended regions
204 downstream of the annotated stop codons in the readthrough transcripts were enriched
205 with featured sequence motifs (Supplementary Dataset 1). Three motif examples,
206 ZNF263, ASCL1, and ETV6 are shown in Supplementary Fig. 3c. These motifs
207 enriched in extended regions likely exert effects on the particular facets of SCR by
208 interacting with specific proteins. In addition, extended regions of SCR transcripts
209 harbored remarkably lower free energy than their analogs of non-SCR transcripts
210 (Supplementary Fig. 3d), suggesting more stable secondary structures in the
211 extensions of the readthrough transcripts.

212 **Stop codons are recoded as 20 amino acids in maize SCR events**

213 Since translation of the stop codon is crucial for SCR events, we examined which
214 amino acids were translated at every readthrough stop codon site based on the MS
215 data. Quite unexpectedly, we found that, with the exception of UAA which was
216 recoded to 19 amino acids (except serine), UGA and UAG were translated to all 20
217 standard amino acids in maize SCR events (Fig. 4a). Among all the amino acids,
218 glutamate was the most recoded amino acid for UGA (in 17 SCR events, 8.9%),
219 methionine for UAG (in 17 cases, 15.6%), and aspartate for UAA (in 11 cases, 11.6%).
220 Previous literature has reported several amino acids that were able to be
221 translated at stop codon sites in different species (Fig. 4b and Supplementary Table 5),
222 prompting us to explore the reasons behind the recoding plasticity of the stop codons
223 in our results. Suppressor transfer RNA s (tRNAs) are specific tRNA species forming
224 classical base pairing with the stop codon, which can lead to the translation[24]. We thus
225 conducted a genome-wide search for such potential suppressor tRNAs capable of
226 reading stop codons in maize by running tRNAscan-SE[51] and ARAGORN[52]. As a
227 result, we acquired a collection of UGA-, UAG-, and UAA-recognizing tRNAs.
228 Multiple sequence alignments of the predicted suppressor tRNAs against the NCBI
229 Nucleotide (NT) database resulted in different groups of suppressor tRNAs well
230 aligned with the canonical tRNAs of a range of amino acids (Supplementary Table 6).
231 We identified six classes of UGA suppressor tRNAs (Supplementary Tables 5 and 6):
232 $tRNA^{Met}$(UCA), $tRNA^{Lys}$(UCA), $tRNA^{Val}$(UCA), $tRNA^{Gly}$(UCA), $tRNA^{Ile}$(UCA), and
233 $tRNA^{Thr}$(UCA). The predicted secondary structures of the first two UGA suppressor
234 tRNAs are shown in Fig. 4c I, II, and their nucleotide sequence alignments are
235 presented in Supplementary Fig. 4a, b. We also detected six classes of UAG
236 suppressor tRNAs (Supplementary Tables 5 and 6): $tRNA^{Glu}$(CUA) (Fig. 4c III and
237 Supplementary Fig. 5a), $tRNA^{Trp}$(CUA) (Fig. 4c IV and Supplementary Fig. 5b),
238 $tRNA^{Arg}$(CUA), $tRNA^{Asp}$(CUA), $tRNA^{Gly}$(CUA), and $tRNA^{Lys}$(CUA). Furthermore,
239 we identified eight classes of UAA suppressor tRNAs (Supplementary Tables 5 and 6):
240 $tRNA^{Lys}$(UUA) (Fig. 4c V and Supplementary Fig. 6a), $tRNA^{His}$(UUA) (Fig. 4c VI
241 and Supplementary Fig. 6b), $tRNA^{Ala}$(UUA), $tRNA^{Asn}$(UUA), $tRNA^{Gly}$(UUA),

242 $tRNA^{Phe}$(UUA), $tRNA^{Met}$(UUA), and $tRNA^{Thr}$(UUA). Collectively, the suppressor
243 tRNAs identified in maize provide potential reasons for the translation of at least
244 some amino acids at stop codon sites.

245 SCR events and recoding plasticity of stop codons in other plants

246 In light of the abundant SCR events identified in maize, we questioned if SCR also
247 occurred in other plant species. We thus followed the same proteogenomic method
248 with the respective 20 custom databases to explore SCR events in rice, soybean, and
249 *Arabidopsis*. Indeed, we detected abundant SCR events in these species with similar
250 features to those in maize. We identified 198 SCR events in rice (Supplementary
251 Table 7), 179 SCR events in soybean (Supplementary Table 9), and 265 SCR events in
252 *Arabidopsis* (Supplementary Table 11), corresponding to 269, 385, and 468 proteins,
253 respectively.

254 In rice, we observed that the stop codon usage SCR events was in the order of
255 UGA > UAA > UAG, and the next in-frame stop codon following the readthrough
256 stop codons also displayed a preference for UGA (Fig. 5a). For the dicots soybean and
257 *Arabidopsis*, the stop codon usages were the same in the order of UGA > UAA >
258 UAG in SCR events, while the most frequent next in-frame stop codon after the
259 readthrough stop codons was UGA for soybean and UAA for *Arabidopsis* (Fig. 5b, c).
260 Genome-wide distribution analysis showed that SCR events were distributed across
261 all chromosomes in their respective plant species (Fig. 5d, e, f). We also analyzed the
262 features of SCR and non-SCR transcripts in the three plants. In rice, the length of SCR
263 primary transcripts and transcript elements including CDSs, introns and the 5′ UTR
264 were shorter than those of non-SCRs ($P < 0.05$, Supplementary Fig. 7a, b). In addition,
265 the SCR primary transcripts tended to possess fewer exons (Supplementary Fig. 7c)
266 and had significantly fewer splice variants ($P < 0.001$, Supplementary Fig. 7d) than
267 non-SCR transcripts. In the dicots soybean and *Arabidopsis*, however, the SCR
268 transcript features were not apparent as compared to those in monocots. SCR
269 transcripts in soybean had significantly shorter UTR lengths and fewer exons than the
270 non-SCR transcripts ($P < 0.05$, Supplementary Fig. 8b, c) but no significance for

271 other feature comparison (Supplementary Fig. 8a, b, d). Similarly, SCR transcripts
272 from *Arabidopsis* showed no significant different from the non-SCR counterparts
273 except fewer splice variants (Supplementary Fig. 9a, b, c, d).

274 Analysis of the recoded amino acids at stop codon sites revealed that UGA
275 exhibited great plasticity to be recoded as the 20 standard amino acids in all the three
276 plant species (Fig. 5g, h, i). For UAG, it was recoded to 18 amino acids (except serine
277 and tryptophan) in rice (Fig. 5g), 17 amino acids (except cystine, aspartate, and
278 phenylalanine) in soybean (Fig. 5h), and 18 amino acids (except aspartate and
279 arginine) in *Arabidopsis* (Fig. 5i). Similarly, UAA was translated to 17, 16, and 20
280 amino acids in rice, soybean, and *Arabidopsis*, respectively (Fig. 5g, h, i). These
281 results further uncover the high recoding plasticity of stop codons in plant SCR events.

282 C-terminal extensions in maize SCR events contain multiple 283 functional signals

284 To examine the biological implications of SCR events, we searched the functional
285 signals in all C-terminal extensions of readthrough proteins in maize. As a result,
286 multiple functional signals were detected in 34 extensions, including the
287 transmembrane domains in 28 extensions, nuclear localization signals in two
288 extensions, and peroxisomal targeting signals (PTS1) in four extensions (Fig. 6a and
289 Supplementary Table 13). Among them, we selected BASIC TRANSCRIPTION
290 FACTOR 3 (ZmBTF3), which was predicted to possess a transmembrane domain in
291 its C-terminal extension (54 amino acids), for further analysis. First, we conducted the
292 *in vitro* translation experiment to validate the readthrough of this gene. We observed
293 high readthrough activity of *ZmBTF3* with an SCR rate of 15% (Fig. 6b), which
294 revealed that *ZmBTF3* was read through and generated a novel isoform (hereafter
295 termed ZmBTF3x). Next, we confirmed the expression of the fusion protein
296 ZmBTF3x-Fluc by immunoblotting using an anti-Fluc antibody (Fig. 6c). The results
297 showed that the molecular weight of the readthrough protein ZmBTF3x-Fluc was the
298 same as its positive control ZmBTF3x^{ACA}-Fluc and higher than that of ZmBTF3-Fluc
299 (Fig. 6c). By contrast, no visible band of readthrough products expressed from the

300 SCR-negative control *Zm00001d009138-TAG-ISR-Fluc* was observed in the
301 immunoblot (Fig. 6c). The fusion protein ZmBTF3x-Fluc was then subjected to mass
302 spectrometric analysis, resulting in the detection of different types of peptides labeled
303 as "before", "through", and "after", which further validated the production of the
304 readthrough protein isoform ZmBTF3x (Fig. 6d, top). The spectrum for "through"
305 peptide "EPEEKKESTFCCV" revealed that the stop codon UAA was translated as
306 threonine (T) (Fig. 6d, bottom).

307 The single transmembrane domain predicted in the C-terminus of ZmBTF3x was
308 found to contain 20 amino acid residues crossing the membrane (Fig. 6e). The
309 predicted 3D structures of ZmBTF3 and ZmBTF3x revealed an α-helix appended at
310 the ZmBTF3x C-terminus that was not present in ZmBTF3 (Fig. 6f). Furthermore, the
311 α-helix with a high pLDDT score and low PAE score corresponded exactly to the
312 region of the predicted transmembrane domain (Fig. 6f and Supplementary Fig. 10a,
313 b). To examine the biological significance of the C-terminal extension derived from
314 SCR of *ZmBTF3*, we constructed *35S:ZmBTF3-GFP* and *35S:ZmBTF3x-GFP* vectors
315 and individually transfected them in maize protoplasts for subcellular localization
316 analysis using confocal microscopy. The results showed that ZmBTF3 localized to the
317 nucleus and cytosol (Fig. 6g), consistent with the previous study reporting that
318 *Arabidopsis* BTF3 acted as the transcription factor and the β-subunit of nascent
319 polypeptide-associated complex (NAC)[53]. By contrast, the readthrough protein
320 ZmBTF3x localized to the plasma membrane and cytosol (Fig. 6g), confirming the
321 localization change between the isoforms ZmBTF3 and ZmBTF3x. In summary, these
322 results indicate that SCR can lead to the structural and subcellular localization change
323 of the readthrough protein isoform, which may confer new biological functions.

324 Discussion

325 SCR, classified as the recoding of stop codons, is a regulatory mode of translation that
326 could produce multiple protein isoforms from the same transcript[1,54]. SCR has been
327 reported in a variety of organisms with important biological implications. Recently,
328 Sahoo *et al*. have identified mRNAs of 144 genes that exhibited SCR in *Arabidopsis*

329 by analyzing Ribo-seq datasets[38]. However, a systematic study of SCR events in
330 plants at the protein level is still lacking. In the current study, we employed a high-
331 resolution MS-based proteogenomic strategy to analyze SCR events in different plants
332 and to identify the amino acids translated at stop codon sites. We established 20
333 customized databases for each of the four plant species with one of the 20 amino acids
334 defined at annotated stop codon sites in each database, which enabled detection of any
335 potential SCR events with all possible amino acids translated at stop codon sites.
336 Using this strategy, we detected 1,009 SCR events corresponding to a collection of
337 2,185 proteins in four plant species.

338 Our systematic analysis of SCR events revealed their similar characteristics
339 among the four plant species. Firstly, SCR events occur in all chromosomes of the
340 four species, and SCR events are prone to be dispersively distributed in the
341 chromosome, except in maize where the higher distribution of SCR was closer to
342 telomeres. Secondly, consistent with previous studies[6,23], UGA is the most preferential
343 stop codon for readthrough, and it also turned out to be the most preferred next in-
344 frame stop codon in SCR events of these species. In addition, SCR transcripts were
345 mostly significantly shorter and had fewer exons and splice variants than non-SCR
346 transcripts in two monocot plants, although these were not as significant as that in the
347 dicots. However, in other eukaryotes like *D. melanogaster* and *Anopheles gambiae*,
348 SCR transcripts were found to be much longer than the non-SCR transcripts, possibly
349 because the occurrence of SCR requires more regulatory elements embedded in their
350 longer transcripts[23,24]. Our findings suggest that the regulation of SCR in plants might
351 differ from that in animals.

352 The stop codons are able to be recoded as all 20 standard amino acids in the four
353 plant species in our study, which has not been reported elsewhere. This result suggests
354 a greater recoding potential of stop codons underestimated in the past and reveals a
355 more flexible mode of decoding. Since tRNA is the adapter molecule in decoding a
356 codon to an amino acid, we explored the reasons for the plasticity of stop codons by
357 tRNA analysis. Suppressor tRNAs, of which the anti-codon could fully form

358 canonical base-pairing with stop codons, have been reported to contribute to the
359 translation of stop codons[24,55,56]. Here, we detected an array of suppressor tRNAs
360 corresponding to different amino acids in maize, which may be the underlying reasons
361 for the translation of six amino acids at UGA, six amino acids at UAG, and eight
362 amino acids at UAA sites. In addition to suppressor tRNAs, near-cognate tRNAs of
363 stop codons are also an important reason for the translation due to the noncanonical
364 base pairing of one position between a stop codon and the anticodon of the tRNA[41,44].
365 The amino acids carried by these near-cognate tRNAs include arginine, serine,
366 tryptophan, cysteine, glycine, and leucine for UGA; lysine, glutamate, glutamine,
367 tryptophan, serine, leucine, and tyrosine for UAG; and lysine, glutamate, glutamine,
368 tyrosine, serine, and leucine for UAA[57,58]. However, we did not find suppressor
369 tRNAs for the rest small portion of the 20 amino acids recoded at stop codon sites in
370 our study, nor the tRNAs of these amino acids were near-cognate tRNAs (e.g., the
371 case of UAG recoded as alanine and UGA recoded as glutamate in Fig. 2). Indeed,
372 these amino acids are not exclusively observed in this study, as previous studies have
373 also reported some of their translation at stop codon sites[15,30,39,46]. These cases suggest
374 that our present understanding of stop codon reading by suppressor or near-cognate
375 tRNAs is still incomplete. Considering the sophisticated modifications and regulation
376 of the decoding adaptor tRNAs and multiple factors that influence SCR events[8,59], the
377 mechanism underlying SCR associated with different tRNAs warrants further study.

378 In summary, we implemented a proteogenomic strategy and identified abundant
379 SCR events across the genomes of four plant species. Stop codons in these SCR
380 events could be recoded as all 20 standard amino acids. We demonstrated that SCR
381 events are widespread in plant species and uncovered the unprecedented plasticity of
382 stop codons. Our findings provide evidence that genetic decoding is far more flexible
383 than has been recognized. We also identified a large volume of novel protein isoforms
384 derived from SCR, which not only expands the plant proteome but also lays a
385 foundation for further exploring the functionality of SCR products.

386 **Methods**

387 **Sample preparation**

388 Maize (*Zea mays,* B73 inbred line), rice (*Oryza sativa* L. ssp. *japonica* cv.
389 Nipponbare), *Arabidopsis thaliana* (Columbia-0), and soybean (*Glycine max* L.,
390 ZH13) plants were grown in a temperature- and light-controlled greenhouse. The
391 growth conditions for maize were a 15-h light/9-h dark photoperiod and a 28/25 °C
392 day/night air temperature, and samples were collected at the three-leaf stage. For rice,
393 the growth conditions were a 12-h light/12-h dark photoperiod and a 28/25 °C
394 day/night air temperature, and samples were collected on day 56. *Arabidopsis* plants
395 were grown in a greenhouse under a 16-h light/8-h dark photoperiod and a 22/21 °C
396 day/night air temperature, and samples were collected on day 21. Soybean plants were
397 grown in a greenhouse under a 16-h light/8-h dark photoperiod and a 22/21 °C
398 day/night air temperature, until the trifoliate leaves were fully expanded. Three
399 replicates of leaf samples for each of the plant species were harvested, frozen in liquid
400 nitrogen, and stored at –80 °C until analysis.

401 **Protein extraction**

402 Total proteins from each sample were extracted as previously reported[60]. Briefly, leaf
403 tissues (1 g) for each plant species were pulverized in liquid nitrogen, and the powder
404 was precipitated in 10% (w/v) trichloroacetic acid/acetone solution containing 65 mM
405 dithiothreitol (DTT) at –20 °C for 1 h. Then, the solution was centrifuged at 10,000×g
406 at 4 °C for 45 min and the supernatant was removed. The precipitate was then
407 vacuum-dried and solubilized in 1/10 volumes of SDT buffer (4% [w/v] SDS, 100
408 mM DTT, and 150 mM Tris-HCl, pH 8.0). After a 3-min incubation in boiling water,
409 the suspensions were ultrasonicated (80 w, ultrasonic 10 s/time, every 15 s, repeated
410 10 times) and incubated at 100 °C for 3 min. The extract was centrifuged at 12,000×g
411 at 25 °C for 10 min, and the total protein contents were quantified using BCA Protein
412 Assay Reagent (Promega Corporation). The supernatants were stored at –80 °C until
413 subsequent analysis.

414 **Protein digestion and peptide fractionation**

415 The total proteins (250 μg for each sample) were digested by FASP procedure using

416 trypsin (Promega Corporation) as described previously[60]. Then, the digested peptide
417 filtrate was fractionated by strong cation exchange (SCX) chromatography using an
418 AKTA Purifier system (GE Healthcare). The dried peptide mixture was reconstituted
419 and acidified with 2 ml buffer A (10 mM KH_2PO_4 in 25%[v/v] of ACN, pH = 2.7) and
420 loaded onto a PolySULFOETHYL 4.6×100 mm column (5 μm, 200 Å, PolyLC Inc).
421 The peptides were eluted at a flow rate of 1 ml/min buffer B (500 mM KCl, 10 mM
422 KH_2PO_4 in 25% of ACN, pH 2.7) in a gradient of 0-10% (v/v) for 2 min, 10-20% (v/v)
423 for 25 min, 20-45% (v/v) for 5 min, and 50-100% (v/v) for 5 min. The elution was
424 monitored by absorbance at 214 nm, and fractions were collected every 1 min. The 30
425 collected fractions were combined into 15 pools and desalted on C18 Cartridges
426 (Empore SPE Cartridges C18 standard density, bed I.D. 7 mm, volume 3 ml, Sigma-
427 Aldrich). Each fraction was concentrated by vacuum centrifugation and reconstituted
428 in 40 μl of 0.1% (v/v) trifluoroacetic acid. All samples were stored at –80 °C before
429 LC-MS/MS analysis.

430 **Liquid chromatography-tandem mass spectrometry analysis**

431 The samples were analyzed on a Q-Exactive HF-X mass spectrometer coupled to the
432 Easy nLC 1200 (Thermo Fisher Scientific). Peptide (2 μg) was loaded onto a 25 cm,
433 75 μm inner diameter C18-reversed phase column packed with 3 μm resin (EASY-
434 Column Capillary Columns, Thermo Fisher Scientific) in buffer A (2% [v/v]
435 acetonitrile and 0.1% [v/v] formic acid) and separated with a linear gradient of buffer
436 B (80% [v/v] acetonitrile and 0.1% formic acid) at a flow rate of 250 nl/min
437 controlled by IntelliFlow technology over 90 min. MS data were acquired using a
438 data-dependent top 10 method dynamically choosing the most abundant precursor
439 ions from the survey scan (m/z 300-1,800) for higher-energy collision dissociation
440 fragmentation (HCD). Determination of the target value was based on predictive
441 Automatic Gain Control (pAGC). The dynamic exclusion duration was 25 s. Survey
442 scans were acquired at a resolution of 70,000 (m/z 200), and the resolution for HCD
443 spectra was set to 17,500 (m/z 200). Normalized collision energy was 30 eV, and the
444 underfill ratio, which specifies the minimum percentage of the target value likely to

445 be reached at maximum fill time, was defined as 0.1%. The instrument was run with
446 peptide recognition mode enabled. MS experiments for each sample were conducted
447 in triplicate for each sample.

448 **Customized database construction**

449 The complete genome sequences of the four plants were downloaded from Ensembl
450 Plants (ftp://ftp.ensemblgenomes.org/pub/plants/release-41/fasta/zea_mays/dna/) for
451 maize, MSU Rice Genome Annotation Project
452 (http://rice.uga.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomol
453 ecules/version_7.0/all.dir/) for rice, the National Genomics Data Center
454 (ftp://download.big.ac.cn/gwh/Plants/Glycine_max_Gmax_ZH13_v2.0_GWHAAEV
455 00000000.1/GWHAAEV00000000.1.genome.fasta.gz) for soybean, and TAIR 10
456 (ftp://ftp.ensemblgenomes.org/pub/plants/release-
457 28/fasta/Arabidopsis_thaliana/dna/Arabidopsis_thaliana.
458 TAIR10.28.dna.toplevel.fa.gz) for *Arabidopsis* in FASTA format. The tailored
459 databases for SCR event identification were generated from the computational
460 translation of plant genome sequences to amino acid sequences by EMBOSS 6.6.0, as
461 previously described with modifications[46]. *In silico* translation followed the principle
462 that the annotated stop codon (UGA, UAG, or UAA) of a given gene was considered
463 a translatable codon to be designated as the symbol "X", representing any one of the
464 20 standard amino acids, and the elongation terminated at the next in-frame stop
465 codon in the 3′ UTR. The *in silico* translation was conducted by in-house scripts based
466 on Shell under a Linux environment to produce 20 customized databases for each
467 plant species. Each database has been in conjunction with a public plant protein
468 database for every plant species as follows: Ensembl Plants
469 (http://ftp.ebi.ac.uk/ensemblgenomes/pub/release-41/plants/fasta/zea_mays/pep/Zea_
470 mays.B73_RefGen_v4.pep.all.fa.gz) for maize, MSU Rice Genome Annotation
471 Project protein database (http://rice.uga.edu/pub/data/Eukaryotic_Projects/o_sativa
472 /annotation_dbs/pseudomolecules/version_7.0/all.dir/all.pep) for rice, the National
473 Genomics Data Center

474 (ftp://download.big.ac.cn/gwh/Plants/Glycine_max_Gmax_ZH13_v2.0_GWHAAEV
475 00000000.1/GWHAAEV00000000.1.Protein.faa.gz) for soybean, and Ensembl Plants
476 (ftp://ftp.ensemblgenomes.org/pub/plants/release-
477 51/fasta/Arabidopsis_thaliana/pep/Arabidopsis_thaliana.TAIR10.pep.all.fa.gz) for
478 *Arabidopsis*. The conjunction of the public protein database ensured the searching
479 precision and calculation of the false discovery rate (FDR) in the subsequent
480 searching stage.

481 **Identification of SCR events by MaxQuant**

482 The raw data were analyzed using MaxQuant software (v.1.6.1.10)[61]. The spectra
483 were searched against the 20 customized databases for each of the four plant species.
484 The initial search was set at a precursor mass window of 6 ppm. The search followed
485 an enzymatic cleavage rule of trypsin and allowed a maximum of two missed
486 cleavage sites and a mass tolerance of 20 ppm for fragment ions.
487 Carbamidomethylation of cysteines was defined as fixed modification, while
488 methionine oxidation was defined as variable modification for database searching.
489 The cutoff of global FDR for peptide and protein identification was set to 0.01.
490 Protein quantification was calculated by normalized spectral protein intensity.

491 All matched peptides were then relocated to the corresponding positions of
492 protein sequences in the reference databases and mapped to their genomic locations.
493 Every stop codon site was given an ID for genomic position recognition. The non-
494 redundant peptides uniquely mapped to the regions spanning the annotated stop codon
495 were labeled as "through", the peptides mapped to the coding sequence (CDS) region
496 were labeled as "before", and the peptides mapped to the 3′ UTR were labeled as
497 "after". We further implemented more stringent criteria to determine the “through”
498 peptides so as to identify SCR events. Only the unique “through” peptides that met the
499 following conditions were retained: (1) presented in at least two of the three replicates
500 in MS experiment; (2) at least two of the quality scores were higher than 20. Finally,
501 proteins sharing one stop codon site and containing a unique "through" peptide were
502 merged into one group and the group was defined as one SCR event.

503 **Ribosome profiling analysis**

504 We obtained ribosome profiling (Ribo-seq) datasets of 56 maize samples from
505 previous reports for SCR event analyzing[62-66]. The Ribo-seq data was analyzed
506 according to the methods described by Dunn et al. (2013)[11] and Sapkota et al. (2019)[6]
507 with modifications. Briefly, the ribosomal sequencing data were quality-controlled
508 using fastax (v.36.06) or cutadapt (v.1.13)[62] to remove low-quality bases, adaptor
509 sequences, and random primers. Reads of 25-35 nucleotides (nts) were retained by
510 seqkit (v.0.5.4) for subsequent analysis. The remaining reads were aligned to the
511 maize genome by STAR (v.2.7.3a)[67]. Ribosome-protected footprint alignments were
512 mapped to their estimated P-sites as follows: 12 nt were cut from each end of the
513 alignment, resulting in a fragment of $n \times$ nt in length ($n = l$-12×2). Each genomic
514 position covered by a nucleotide remaining in the truncated alignment was then
515 incremented by $1/n$ (count number). Then, the reads were mapped to the following
516 positions: 12 nts after the start codon, 15 nts before and 9 nts after the first stop codon,
517 and 15 nts after the second stop codon. This resulted in four regions remaining in the
518 transcripts: the 5′ UTR, CDS, readthrough region, and the distal 3′ UTR.

519 We then analyzed the length distribution of Ribo-seq reads in each dataset. Four
520 Ribo-seq datasets in which the read length was mainly distributed in 30-nt (30-mer)
521 and prepared with RNase I (little cutting bias) were finally retained for further SCR
522 validation. For analysis of 3-nt periodicity (phasing) of footprints, R package
523 ribosomeProfilingQC (v.1.10.0) was used and only 30-nt RPFs were counted, as
524 shown in maize with good-phased footprint population[62]. Reads were then counted in
525 all putative extended regions. Mean value of normalized RPKM (reads per kilobase
526 per million mapped reads) was calculated in R (v.3.2.5) in the four datasets. The
527 ribosome footprint density of putative extensions and the distal 3′ UTRs were
528 calculated in RPKM. Only extensions with RPKM > 1, read coverage > 10%, and no
529 overlap with any annotated 5′ UTRs, CDSs, and other non-coding RNAs were
530 retained and regarded as readthrough. Full data of samples that exhibited readthrough
531 in Ribo-seq analysis for each of the SCR events are shown in Supplementary Table 3.

532 **Distribution of readthrough events at the genomic level**

533 The genome-wide density of SCR and non-SCR events in maize chromosomes was
534 plotted using a sliding window of 20 Mb with 2-Mb steps according to the annotated
535 maize genome in Ensembl (https://plants.ensembl.org/index.html). We used "hotspot
536 region" to reflect the distribution density of SCR and non-SCR events. Hotspot
537 regions were defined as ≥ 5 SCR events for SCR hotspots and ≥ 335 for non-SCR
538 events in a window size of 20 Mb. The chromosomal distributions of rice SCR and
539 non-SCR events were drawn using a window size of 4 Mb with a step of 400 kb based
540 on the annotated rice genome from MSU (http://rice.uga.edu/). The same distribution
541 pattern was plotted with a window size of 10 Mb with a 1-Mb step based on the ZH13
542 reference genome (https://ngdc.cncb.ac.cn/gwh/Assembly/652/show) for soybean. For
543 *Arabidopsis*, the SCR and non-SCR genomic localizations were obtained by a 1.2-Mb
544 sliding window with a step size of 100 kb according to the annotated *Arabidopsis*
545 genome in TAIR (https://www.Arabidopsis.org/). Hotspot regions were defined as ≥3
546 for SCR events and ≥178 for non-SCR events in a window size of 4 Mb for rice, ≥2
547 for SCR events and ≥253 for non-SCR events in a window size of 10 Mb for soybean,
548 and ≥4 for SCR events and ≥157 for non-SCR events in a window size of 1.2 Mb for
549 *Arabidopsis*.

550 **Motif and minimum free energy analysis in readthrough regions of SCR**
551 **transcripts in maize**

552 Transcription factor-binding motifs in readthrough regions of maize SCR transcripts
553 were identified using the MEME-ChIP webserver (v.5.4.1)[68], and the annotations of
554 the obtained motifs were searched against the available databases for comparison
555 using Tomtom (v.5.0.5) with default parameters[69].

556 The minimum free energy of readthrough regions for maize SCR transcripts and
557 the comparable length between the first and downstream second stop codons for non-
558 readthrough transcripts were calculated using ViennaRNA Package (v.2.0) as
559 described[70].

560 **Identification and alignment of suppressor tRNAs in maize**

561 The complete maize genome sequences were downloaded from the Ensembl database
562 (ftp://ftp.ensemblgenomes.org/pub/plants/release-41/fasta/zea_mays/dna/). A genome-
563 wide search for suppressor tRNA genes was conducted using tRNAscan-SE (v.2.0)[51]
564 with default settings and ARAGORN[52] with both default settings and 0.95 default
565 settings as a lower scoring threshold. This analysis generated a collection of
566 suppressor/suppressor-like tRNAs recognizing the three stop codons from each search.
567 Comparison of the canonical tRNAs from different species with these predicted
568 suppressor tRNAs was performed using BLASTn (v.2.5.0+) against the NCBI
569 GenBank nucleotide (NCBI NT) database as previously described[55,56]. The potential
570 identity of the suppressor tRNAs corresponding to different amino acids could be
571 predicted from the high similarity of the sequence alignment between the suppressor
572 tRNAs and the canonical tRNAs of the standard amino acids.

573 **Motif and structure prediction of SCR proteins in maize**

574 C-terminal extensions at least 20 amino acids in length in maize SCR events were
575 scanned for transmembrane helices using TMHMM (v.2.0)[71] with default settings and
576 for nuclear localization signals using cNLS mapper[72]. C-terminal extensions at least
577 12 amino acids in length were searched for peroxisomal targeting signals using PTS1
578 Predictor[73] and prenylation signals using PrePS[74]. The 3D structures of ZmBTF3 and
579 ZmBTF3x were predicted using Alphafold[75].

580 **Plasmid construction**

581 Luciferase constructs for the luminescence-based SCR assay were generated in the
582 reconstructive pFN19K T7 SP6 flexi vector (Promega Corporation). The coding
583 sequences of the tested genes, *ZmH3C2*, *ZmCEQORH*, *ZmBTF3* and
584 *Zm00001d009138*, along with their canonical stop codons and inter-stop codon
585 regions (ISRs, the extension regions) were cloned upstream of the coding sequence of
586 the firefly luciferase gene (*Fluc*) without ATG between *EcoRI* and *KpnI* sites. A linker
587 sequence (5′ GGCGGCTCCGGCGGCTCCCTCGTGCTCGGG 3′) was added
588 upstream of the Fluc sequence to avoid potential distortions between the tested protein
589 and Fluc in the fusion protein[15,31,38]. Primer sequences are listed in Supplementary

590 Table 15.

591 For subcellular localization analysis, the ZmBTF3 and ZmBTF3x coding
592 sequences were introduced into the pAN580 vector containing the 35S promoter and
593 the coding sequence for GFP. GFP was fused to the C-terminus of ZmBTF3 and
594 ZmBTF3x. Primer sequences are listed in Supplementary Table 15.

595 ***In vitro* translation for SCR Assay**

596 For SCR assay, we used *in vitro* translation system based on Wheat Germ Extract
597 (Promega Corporation), which has been used for SCR validation as reported
598 previously[31,38]. The constructed plasmid DNA including the test gene and *Fluc* (no
599 ATG) was linearized using BamH I or XbaI enzyme, and 2 μg of the linearized DNA
600 was transcribed *in vitro* using T7 RNA polymerase (Promega Corporation). The
601 obtained RNA was then purified using GeneJET kit (Thermo Fisher Scientific). The
602 concentration and purity of the RNA were measured using Nanodrop 2000 (Thermo
603 Fisher Scientific). A total of 2 μg purified RNA was translated *in vitro* using the
604 Wheat Germ Extract at 25 °C for 2 h according to the manufacturer′s instructions.
605 Luciferase activity was then measured at room temperature (below 25 °C) using the
606 Luciferase Assay System (Promega Corporation) in a Glomax luminometer (Promega
607 Corporation) with a 10-s measurement read for luciferase activity.

608 **Immunoblotting and mass spectrometric analysis**

609 The proteins extracted from the *in vitro* protein expression system were separated by
610 10% (w/v) SDS-polyacrylamide gel electrophoresis (SDS-PAGE). One PAGE gel was
611 stained with Coomassie Brilliant Blue (CBB), while another was transferred to a
612 polyvinylidene difluoride membrane (0.22 μm, Millipore). Proteins were detected by
613 immunoblotting with anti-Fluc antibodies (1:10,000, ab185924, Abcam). The regions
614 of interest in the CBB staining gel were excised manually according to the
615 immunoblotting band, and then destained with 30% (v/v) ACN in 100 mM NH_4HCO_3
616 and dried in a vacuum centrifuge. The extracted proteins were reduced with
617 dithiothreitol (DTT) (10 mM DTT in 100 mM NH_4HCO_3) for 30 min at 56 °C, then
618 alkylated with iodoacetamide (200 mM IAA in 100 mM NH_4HCO_3) in the dark at

619 room temperature for 30 min. Gel pieces were briefly rinsed with 100 mM NH_4HCO_3
620 and ACN, respectively. The gel pieces were then divided equally into three parts and
621 separately digested overnight in 12.5 ng/μl trypsin, overnight in chymotrypsin in 25
622 mM NH_4HCO_3, and for 30 min in 12.5 ng/μl pepsin in PH3 HCl. The peptides were
623 extracted using 60% (v/v) ACN in 0.1% (v/v) trifluoroacetic acid (TFA) three times,
624 and the extracts were pooled and dried completely by a vacuum centrifuge.

625 LC-MS/MS was performed on a Q-Exactive HF-X mass spectrometer coupled to
626 Easy nLC 1200 (Thermo Fisher Scientific). The peptide mixture was loaded onto a
627 C18-reversed phase column (15 cm, 75 μm inner diameter) packed in-house with RP-
628 C18 (5 μm resin, Thermo Fisher Scientific) in buffer A (0.1% [v/v] formic acid in
629 HPLC-grade water) and separated with a linear gradient of buffer B (0.1% formic acid
630 in 84% [v/v] ACN) at a flow rate of 250 nl/min controlled by IntelliFlow technology
631 over 60 min. MS data were acquired using a data-dependent top-10 method
632 dynamically choosing the most abundant precursor ions from the survey scan (m/z
633 300-1,800) for HCD fragmentation. The target value was determined based on pAGC.
634 The dynamic exclusion duration was 20 s. Survey scans were acquired at a resolution
635 of 70,000 (m/z 200) and the resolution for HCD spectra was set to 17,500 (m/z 200).
636 Normalized collision energy was 27 eV and the underfill ratio, which specifies the
637 minimum percentage of the target value likely to be reached at maximum fill time,
638 was defined as 0.1%.

639 The MS data were then separately searched against the target protein sequences
640 (readthrough protein sequences of *H3C2*, *ZmCEQORH*, and *ZmBTF3*) using
641 MaxQuant software (v.1.6.1.10)[61]. An initial search was set with a precursor mass
642 window of 6 ppm. The search followed an enzymatic cleavage rule of "unspecified"
643 and a mass tolerance of 20 ppm for fragment ions. Carbamidomethylation of cysteines
644 was defined as fixed modification, while methionine oxidation was defined as
645 variable modification for database searching. The cutoff of the global FDR for peptide
646 and protein identification was set to 0.01.

647 **Subcellular localization of ZmBTF3-GFP and ZmBTF3x-GFP fusions.**

648 The GFP fusion plasmids were transformed into maize mesophyll protoplasts by
649 polyethylene glycol-mediated transformation. Protoplast preparation and transfection
650 were conducted as previously described[76]. A plasma membrane-localized marker
651 *35S:ZmPDCK7-mCherry* and a nucleus-localized marker *35S:AtHY5-mCherry* were
652 used for colocalization analyses[77,78]. Fluorescence signals were observed on a
653 confocal laser scanning microscope (LSM800, Zeiss).

654 Data availability

655 All data generated or analyzed in this study are included in this article or the
656 Supplementary Information and Tables. The raw datasets of proteomics have been
657 deposited at the ProteomeXchange under the accession ID PXD035958 and
658 PXD038762.

659 Code availability

660 Custom codes are available for research purposes from the corresponding authors
661 upon request.

662 **References**

663 1. Baranov, P. V., Atkins, J. F. & Yordanova, M. M. Augmented genetic decoding:
664 global, local and temporal alterations of decoding processes and codon
665 meaning. *Nat. Rev. Genet.* **16**, 517-529 (2015).

666 2. Freitag, J., Ast, J. & Bolker, M. Cryptic peroxisomal targeting via alternative
667 splicing and stop codon read-through in fungi. *Nature* **485**, 522-525 (2012).

668 3. Bazykin, G. A. & Kochetov, A. V. Alternative translation start sites are
669 conserved in eukaryotic genomes. *Nucleic Acids Res.* **39**, 567-577 (2011).

670 4. Jacks, T. & Varmus, H. E. Expression of the Rous sarcoma virus *pol* gene by
671 ribosomal frameshifting. *Science* **230**, 1237-1242 (1985).

672 5. Lang, B. F. et al. Massive programmed translational jumping in mitochondria.
673 *Proc. Natl Acad. Sci. USA* **111**, 5926-5931 (2014).

674 6. Sapkota, D. et al. Cell-type-specific profiling of alternative translation
675 Identifies regulated protein isoform variation in the mouse brain. *Cell Rep.* **26**,
676 594-607.e597 (2019).

677 7. Rodnina, M. V. et al. Translational recoding: canonical translation mechanisms
678 reinterpreted. *Nucleic Acids Res.* **48**, 1056-1067 (2020).

679 8. Palma, M. & Lejeune, F. Deciphering the molecular mechanism of stop codon
680 readthrough. *Biol. Rev. Camb. Philos. Soc.* **96**, 310-329 (2021).

681 9. Torabi, N. & Kruglyak, L. Genetic basis of hidden phenotypic variation
682 revealed by increased translational readthrough in yeast. *PLoS Genet.* **8**,
683 e1002546 (2012).

684 10. Namy, O., Duchateau-Nguyen, G. & Rousset, J. P. Translational readthrough
685 of the *PDE2* stop codon modulates cAMP levels in *Saccharomyces cerevisiae*.
686 *Mol. Microbiol.* **43**, 641-652 (2002).

687 11. Dunn, J. G., Foo, C. K., Belletier, N. G., Gavis, E. R. & Weissman, J. S.
688 Ribosome profiling reveals pervasive and regulated stop codon readthrough in
689 *Drosophila melanogaster*. *eLife* **2**, e01179 (2013).

690 12. Xue, F. & Cooley, L. *kelch* encodes a component of intercellular bridges in

691 *Drosophila* egg chambers. *Cell* **72**, 681-693 (1993).
692 13. Karki, P. et al. Tissue-specific regulation of translational readthrough tunes
693 functions of the traffic jam transcription factor. *Nucleic Acids Res.* (2021).
694 14. Schueren, F. et al. Peroxisomal lactate dehydrogenase is generated by
695 translational readthrough in mammals. *eLife* **3**, e03640 (2014).
696 15. Eswarappa, S. M. et al. Programmed translational readthrough generates
697 antiangiogenic VEGF-Ax. *Cell* **157**, 1605-1618 (2014).
698 16. Xin, H., Zhong, C., Nudleman, E. & Ferrara, N. Evidence for pro-angiogenic
699 functions of VEGF-Ax. *Cell* **167**, 275-284.e276 (2016).
700 17. Weiner, A. M. & Weber, K. Natural read-through at the UGA termination
701 signal of Q-beta coat protein cistron. *Nat. New Biol.* **234**, 206-209 (1971).
702 18. Pelham, H. R. Leaky UAG termination codon in tobacco mosaic virus RNA.
703 *Nature* **272**, 469-471 (1978).
704 19. Yoshinaka, Y., Katoh, I., Copeland, T. D. & Oroszlan, S. Murine leukemia
705 virus protease is encoded by the *gag-pol* gene and is synthesized through
706 suppression of an amber termination codon. *Proc. Natl Acad. Sci. USA* **82**,
707 1618-1622 (1985).
708 20. True, H. L. & Lindquist, S. L. A yeast prion provides a mechanism for genetic
709 variation and phenotypic diversity. *Nature* **407**, 477-483 (2000).
710 21. Namy, O. et al. Identification of stop codon readthrough genes in
711 *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **31**, 2289-2296 (2003).
712 22. Baudin-Baillieu, A. et al. Genome-wide translational changes induced by the
713 prion [PSI+]. *Cell Rep.* **8**, 439-448 (2014).
714 23. Jungreis, I. et al. Evidence of abundant stop codon readthrough in *Drosophila*
715 and other metazoa. *Genome Res.* **21**, 2096-2113 (2011).
716 24. Jungreis, I. et al. Evolutionary dynamics of abundant stop codon readthrough.
717 *Mol. Biol. Evol.* **33**, 3108-3132 (2016).
718 25. Hudson, A. M. et al. Tissue-specific dynamic codon redefinition in *Drosophila*.
719 *Proc. Natl Acad. Sci. USA* **118** (2021).

720 26. Klagges, B. R. et al. Invertebrate synapsins: a single gene codes for several
721 isoforms in Drosophila. *J Neurosci.* **16**, 3154-3165 (1996).
722 27. Steneberg, P. & Samakovlis, C. A novel stop codon readthrough mechanism
723 produces functional Headcase protein in *Drosophila* trachea. *EMBO Rep* 2,
724 593-597 (2001).
725 28. Geller, A. I. & Rich, A. A UGA termination suppression tRNATrp active in
726 rabbit reticulocytes. *Nature* **283**, 41-46 (1980).
727 29. Yamaguchi, Y. et al. L-MPZ, a novel isoform of myelin P0, is produced by
728 stop codon readthrough. *J. Biol. Chem.* **287**, 17765-17776 (2012).
729 30. Manjunath, L. E. et al. Stop codon read-through of mammalian *MTCH2*
730 leading to an unstable isoform regulates mitochondrial membrane potential. *J.*
731 *Biol. Chem.* **295**, 17009-17026 (2020).
732 31. Singh, A. et al. Let-7a-regulated translational readthrough of mammalian
733 *AGO1* generates a microRNA pathway inhibitor. *EMBO J.* **38**, e100727 (2019).
734 32. Loughran, G. et al. Evidence of efficient stop codon readthrough in four
735 mammalian genes. *Nucleic Acids Res.* **42**, 8928-8938 (2014).
736 33. De Bellis, M. et al. Translational readthrough generates new astrocyte AQP4
737 isoforms that modulate supramolecular clustering, glial endfeet localization,
738 and water transport. *Glia* **65**, 790-803 (2017).
739 34. Stiebler, A. C. et al. Ribosomal readthrough at a short UGA stop codon context
740 triggers dual localization of metabolic enzymes in Fungi and animals. *PLoS*
741 *Genet.* **10**, e1004685 (2014).
742 35. Loughran, G. et al. Stop codon readthrough generates a C-terminally extended
743 variant of the human vitamin D receptor with reduced calcitriol response. *J.*
744 *Biol. Chem.* **293**, 4434-4444 (2018).
745 36. Yordanova, M. M. et al. *AMD1* mRNA employs ribosome stalling as a
746 mechanism for molecular memory formation. *Nature* **553**, 356-360 (2018).
747 37. Yordanova, M. M., Loughran, G., Atkins, J. F. & Baranov, P. V. Stop codon
748 readthrough contexts influence reporter expression differentially depending on

749 the presence of an IRES. *Wellcome Open Res.* **5**, 221 (2020).
750 38. Sahoo, S. et al. Identification and functional characterization of mRNAs that
751 exhibit stop codon readthrough in *Arabidopsis thaliana*. *J. Biol. Chem.* **298**,
752 102173 (2022).
753 39. Blanchet, S., Cornu, D., Argentini, M. & Namy, O. New insights into the
754 incorporation of natural suppressor tRNAs at stop codons in *Saccharomyces*
755 *cerevisiae*. *Nucleic Acids Res.* **42**, 10061-10072 (2014).
756 40. Feng, Y. X., Copeland, T. D., Oroszlan, S., Rein, A. & Levin, J. G.
757 Identification of amino acids inserted during suppression of UAA and UGA
758 termination codons at the *gag-pol* junction of Moloney murine leukemia virus.
759 *Proc. Natl Acad. Sci. USA* **87**, 8860-8863 (1990).
760 41. Roy, B., Leszyk, J. D., Mangus, D. A. & Jacobson, A. Nonsense suppression
761 by near-cognate tRNAs employs alternative base pairing at codon positions 1
762 and 3. *Proc. Natl Acad. Sci. USA* **112**, 3038-3043 (2015).
763 42. Strigini, P. & Brickman, E. Analysis of specific misreading in *Escherichia coli*.
764 *J. Mol. Biol.* **75**, 659-672 (1973).
765 43. Sánchez-Silva, R., Villalobo, E., Morin, L. & Torres, A. A new noncanonical
766 nuclear genetic code: translation of UAA into glutamate. *Curr. Biol.* **13**, 442-
767 447 (2003).
768 44. Blanchet, S. et al. Deciphering the reading of the genetic code by near-cognate
769 tRNA. *Proc. Natl Acad. Sci. USA* **115**, 3018-3023 (2018).
770 45. Choi, S. et al. Proteogenomic Approach to UTR Peptide Identification. *J*
771 *Proteome Res.* **19**, 212-220 (2019).
772 46. Aerni, H. R., Shifman, M. A., Rogulina, S., O'Donoghue, P. & Rinehart, J.
773 Revealing the amino acid composition of proteins within an expanded genetic
774 code. *Nucleic Acids Res.* **43**, e8 (2015).
775 47. Valle, R. P., Morch, M. D. & Haenni, A. L. Novel amber suppressor tRNAs of
776 mammalian origin. *EMBO J.* **6**, 3049-3055 (1987).
777 48. Namy, O., Hatin, I. & Rousset, J. P. Impact of the six nucleotides downstream

778 of the stop codon on translation termination. *EMBO Rep.* **2**, 787-793 (2001).
779 49. Tork, S., Hatin, I., Rousset, J. P. & Fabret, C. The major 5' determinant in stop
780 codon read-through involves two adjacent adenines. *Nucleic Acids Res.* **32**,
781 415-421 (2004).
782 50. Mangkalaphiban, K. et al. Transcriptome-wide investigation of stop codon
783 readthrough in *Saccharomyces cerevisiae*. *PLoS Genet.* **17**, e1009538 (2021).
784 51. Chan, P. P., Lin, B. Y., Mak, A. J. & Lowe, T. M. tRNAscan-SE 2.0: improved
785 detection and functional classification of transfer RNA genes. *Nucleic Acids*
786 *Res.* **49**, 9077-9096 (2021).
787 52. Laslett, D. & Canback, B. ARAGORN, a program to detect tRNA genes and
788 tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* **32**, 11-16 (2004).
789 53. Ding, Y. et al. OST1-mediated BTF3L phosphorylation positively regulates
790 CBFs during plant cold responses. *EMBO J.* **37**, *e98228* (2018).
791 54. Atkins, J. F. & Baranov, P. V. The distinction between recoding and codon
792 reassignment. *Genetics* **185**, 1535-1536 (2010).
793 55. Ivanova, N. N. et al. Stop codon reassignments in the wild. *Science* **344**, 909-
794 913 (2014).
795 56. Swart, E. C., Serra, V., Petroni, G. & Nowacki, M. Genetic codes with no
796 dedicated stop codon: context-dependent translation termination. *Cell* **166**,
797 691-702 (2016).
798 57. Beier, H. & Grimm, M. Misreading of termination codons in eukaryotes by
799 natural nonsense suppressor tRNAs. *Nucleic Acids Res.* **29**, 4767-4782 (2001).
800 58. Dabrowski, M., Bukowy-Bieryllo, Z. & Zietkiewicz, E. Translational
801 readthrough potential of natural termination codons in eucaryotes--The impact
802 of RNA sequence. *RNA Biol.* **12**, 950-958 (2015).
803 59. Schimmel, P. The emerging complexity of the tRNA world: mammalian tRNAs
804 beyond protein synthesis. *Nat. Rev. Mol. Cell Biol.* **19**, 45-58 (2018).
805 60. Wiśniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample
806 preparation method for proteome analysis. *Nat. Methods* **6**, 359-362 (2009).

807 61. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for
808 mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **11**, 2301-2319
809 (2016).
810 62. Lei, L. et al. Ribosome profiling reveals dynamic translational landscape in
811 maize seedlings under drought stress. *Plant J.* **84**, 1206-1218 (2015).
812 63. Chotewutmontri, P. & Barkan, A. Dynamics of chloroplast translation during
813 chloroplast differentiation in maize. *PLoS Genet.* **12**, e1006106 (2016).
814 64. Chotewutmontri, P. & Barkan, A. Multilevel effects of light on ribosome
815 dynamics in chloroplasts program genome-wide and psbA-specific changes in
816 translation. *PLoS Genet.* **14**, e1007555 (2018).
817 65. Zoschke, R., Chotewutmontri, P. & Barkan, A. Translation and co-translational
818 membrane engagement of plastid-encoded chlorophyll-binding proteins are
819 not influenced by chlorophyll availability in maize. *Front. Plant. Sci.* **8**, 385
820 (2017).
821 66. Jiang, J. et al. An RNA chaperone-like protein plays critical roles in
822 chloroplast mRNA stability and translation in *Arabidopsis* and maize. *Plant*
823 *Cell* **31**, 1308-1327 (2019).
824 67. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**,
825 15-21 (2013).
826 68. Machanick, P. & Bailey, T. L. MEME-ChIP: motif analysis of large DNA
827 datasets. *Bioinformatics* **27**, 1696-1697 (2011).
828 69. Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L. & Noble, W. S.
829 Quantifying similarity between motifs. *Genome Biol.* **8**, R24 (2007).
830 70. Lorenz, R. et al. ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6**, 26 (2011).
831 71. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting
832 transmembrane protein topology with a hidden Markov model: application to
833 complete genomes. *J. Mol. Biol.* **305**, 567-580 (2001).
834 72. Kosugi, S., Hasebe, M., Tomita, M. & Yanagawa, H. Systematic identification
835 of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by

836 prediction of composite motifs. *Proc. Natl Acad. Sci. USA* **106**, 10171-10176
837 (2009).
838 73. Neuberger, G., Maurer-Stroh, S., Eisenhaber, B., Hartig, A. & Eisenhaber, F.
839 Prediction of peroxisomal targeting signal 1 containing proteins from amino
840 acid sequence. *J. Mol. Biol.* **328**, 581-592 (2003).
841 74. Maurer-Stroh, S. & Eisenhaber, F. Refinement and prediction of protein
842 prenylation motifs. *Genome Biol.* **6**, R55 (2005).
843 75. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold.
844 *Nature* **596**, 583-589 (2021).
845 76. Yoo, S. D., Cho, Y. H. & Sheen, J. *Arabidopsis* mesophyll protoplasts: a
846 versatile cell system for transient gene expression analysis. *Nat. Protoc.* **2**,
847 1565-1572 (2007).
848 77. Zhao, Y. et al. The calcium-dependent protein kinase ZmCDPK7 functions in
849 heat-stress tolerance in maize. *J. Integr. Plant. Biol.* **63**, 510-527 (2021).
850 78. Li, Y. et al. Plasma membrane-nucleo-cytoplasmic coordination of a receptor-
851 like cytoplasmic kinase promotes EDS1-dependent plant immunity. *Nat.*
852 *Plants* **8**, 802-816 (2022).

Acknowledgments

This study was supported by the National Natural Science Foundation of China (no. 32172073, U22A20474, 31872872), the National Key Research and Development Program of China (2022YFD1201802), and the Seed-Industrialized Development Program in Shandong Province (no. 2021LZGC003). We are grateful to Steven P. Briggs and David Backhouse for their constructive comments and suggestions. We also thank Dr. Anguo Sun and Dr. Yanwen Xiang for the technical assistance and Yingying Sha and Mengxia Niu for their assistance with the microscopic observations.

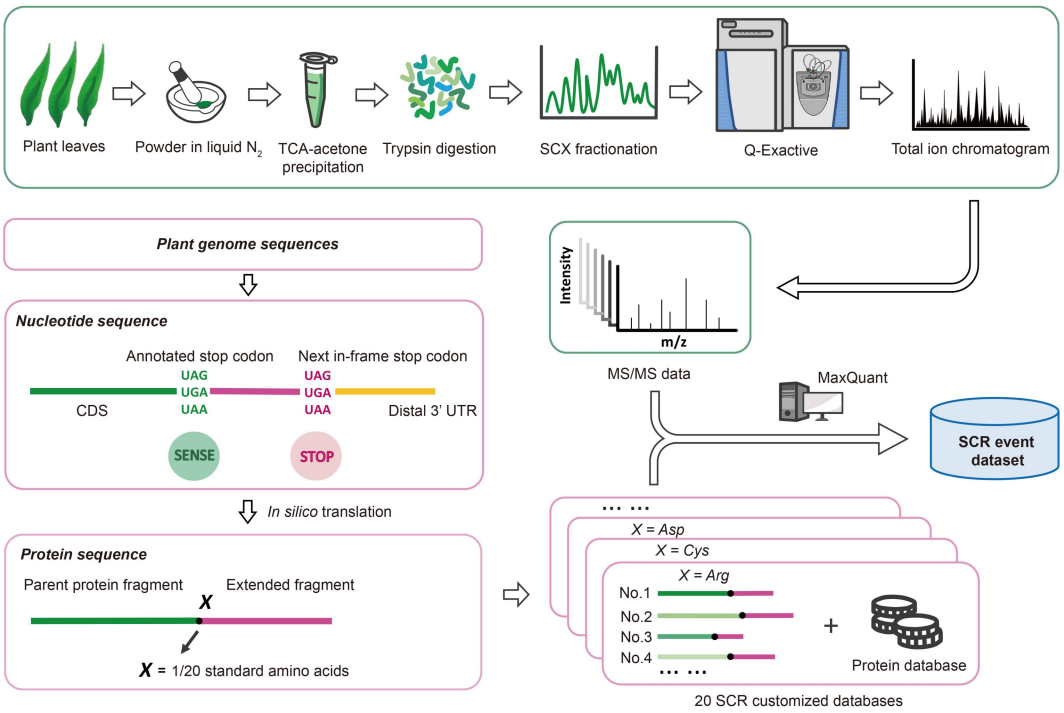
Author Contributions

L.W., S.-B.W., and Z.T. designed the project. Y. Zhang, Y.S., L.T., H.Y., A.X., J.Z., U.A., A.S., X.C. and Y.C. were involved in bioinformatics analysis. Y. Zhang, H.L., S.W. and J.S. conducted the experiments. C.G., Y. Zhao, Y. L. and X.W. collected the materials for mass spectrometry detection and subsequent experiments. Y. Zhang, H.L. and Y.S. wrote the manuscript. L.W., S.-B.W., and Z.T. contributed substantially to revisions. All authors read and approved the manuscript.

Competing interests

The authors declare no competing interests.

872 # Figures



873 **Fig. 1 | Proteogenomic workflow for identifying plant stop codon readthrough (SCR)**
874 **events.** The strategy consists of two procedures: shotgun proteomics experiment (green box)
875 and customized database construction (pink boxes). Total protein was extracted from plant
876 leaves and subjected to digestion, fragmentation, and LC-MS/MS analysis for identification
877 of spectra datasets. For customized database construction, the plant genome sequences were
878 downloaded from public databases and subjected to *in silico* translation. The annotated stop
879 codon sites were replaced with "X", representing one of the 20 standard amino acids during
880 translation each time. The translation process was assigned to terminate at the next in-frame
881 stop codon, producing 20 customized databases containing theoretical readthrough proteins
882 with the annotated stop sites corresponding to the 20 amino acids. The distal 3′ UTR
883 represents the portion of the 3′ UTR downstream of the putative readthrough extension.
884 MS/MS spectra datasets were searched against the 20 customized databases merged with the
885 public plant protein database by the MaxQuant. We defined the MS/MS-identified peptides
886 that exclusively span the annotated stop codon sites as "through" peptides, and classified
887 proteins sharing a unique stop codon site and containing a "through" peptide as one SCR
888 event.

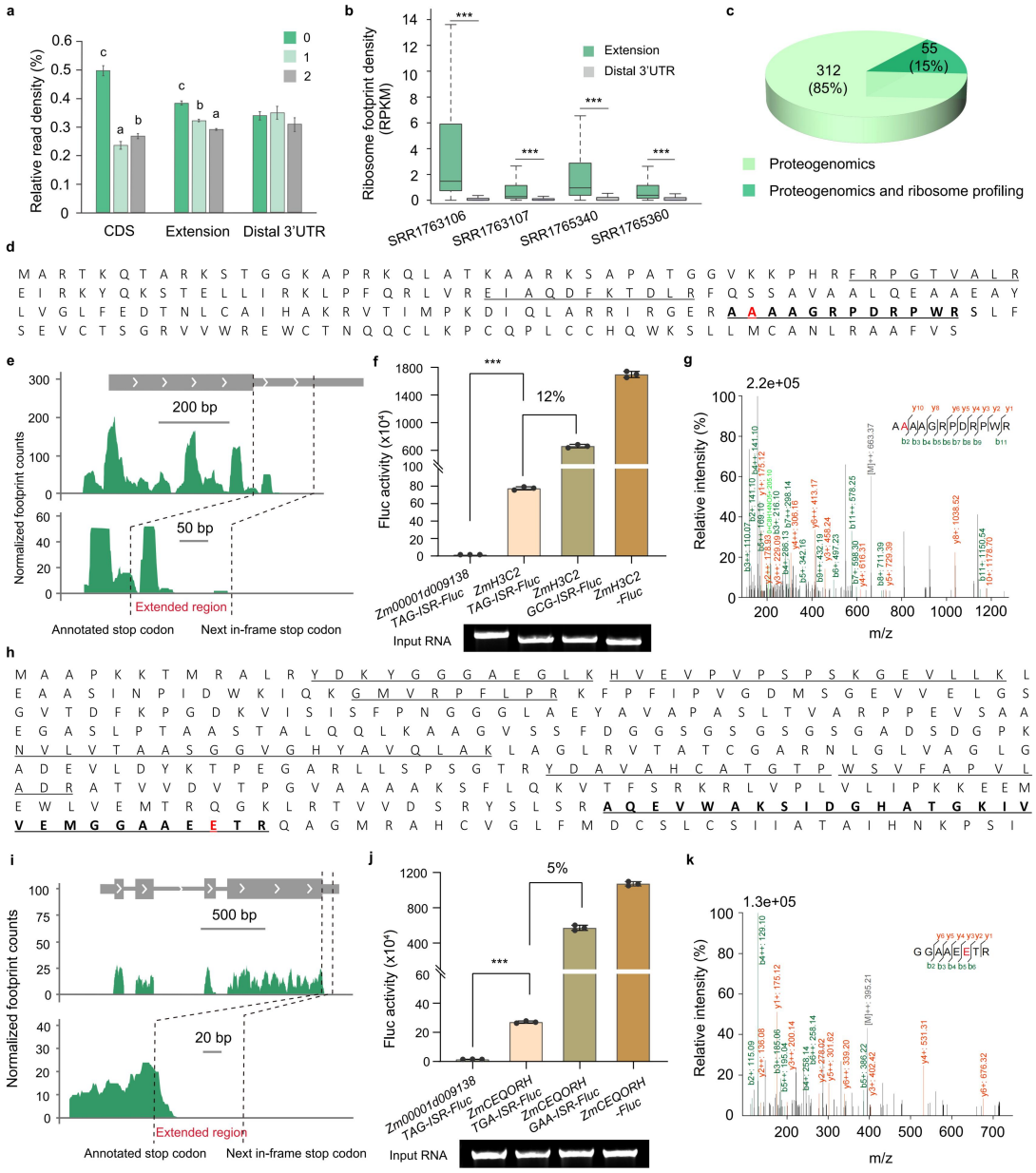


Fig. 2 | Verifying SCR events in maize. a Three-nucleotide periodicity of 30-mers in the CDSs, putative extensions, and the distal 3′ UTRs. Data are mean ± s.d. Different letters indicate significant differences at $P < 0.05$ by one-way ANOVA with Tukey's multiple comparisons test. Green, light green, and gray bars represent frame 0, 1, and 2, respectively. **b** Distribution of ribosome footprint density in putative extensions and the distal 3′ UTRs in the four Ribo-seq datasets. Distal 3′ UTR was defined as the region 40 codons after the next stop-codon or the region from the next stop-codon to the third stop codon within 40 codons. Box edges represent the 0.25 and 0.75 quantiles. The bold lines represent the median and whiskers represent 1.5×IQR. Statistical significance was determined using a two-sided Student's *t*-test,

898 ***$P < 0.001$. **c** The proportion of SCR events confirmed by ribosome profiling among those
899 identified by proteogenomics. **d-k** Validation of two readthrough examples, *ZmH3C2* (**d-g**)
900 and *ZmCEQORH* (**h-k**). **d, h** The complete protein sequences extended to the next in-frame
901 stop codon of the two genes showing the underlined peptides identified by mass spectrometry
902 analysis, and the readthrough peptides are in bold. The amino acids translated at the stop
903 codon sites in readthrough peptides are highlighted in red, which are alanine (A) at UAG (**d**)
904 and glutamate (E) at UGA (**h**). **e, i** Ribo-seq analysis revealing clear footprint signals in the C-
905 terminal extensions of the two transcripts. Dashed lines represent the annotated stop codon
906 and the next in-frame stop codon. **f, j** *In vitro* translation experiment showing readthrough
907 activities of *ZmH3C2* and *ZmCEQORH*. Plasmids containing test cDNA-stop codon-ISR-*Fluc*
908 were firstly transcribed *in vitro*, followed by *in vitro* translation using wheat germ extract.
909 Luciferase activities were used to represent the readthrough activities. Constructs with the
910 stop codon mutated to a sense codon were used as a positive control of the translation to
911 measure the efficiency of SCR. The non-SCR gene *Zm00001d009138* with its TAG and ISR
912 sequence fused with *Fluc* (*Zm00001d009138-TAG-ISR-Fluc*) was used as the SCR-negative
913 control, as it showed no SCR sign in either proteogenomics or Ribo-seq analysis. All samples
914 were tested in biological triplicates. Data are mean ± s.e.m (n=3). Asterisks indicate statistical
915 significance by two-sided Student's *t*-test (***$P < 0.001$). ISR, inter-stop codon region
916 (extension region). **g, k** MS/MS spectra of the identified readthrough peptides
917 AAAAGRPDRPWR for *ZmH3C2* (**g**) and GGAAEETR for *ZmCEQORH* **(k)** readthrough
918 products expressed from (**f**) and (**j**), respectively.

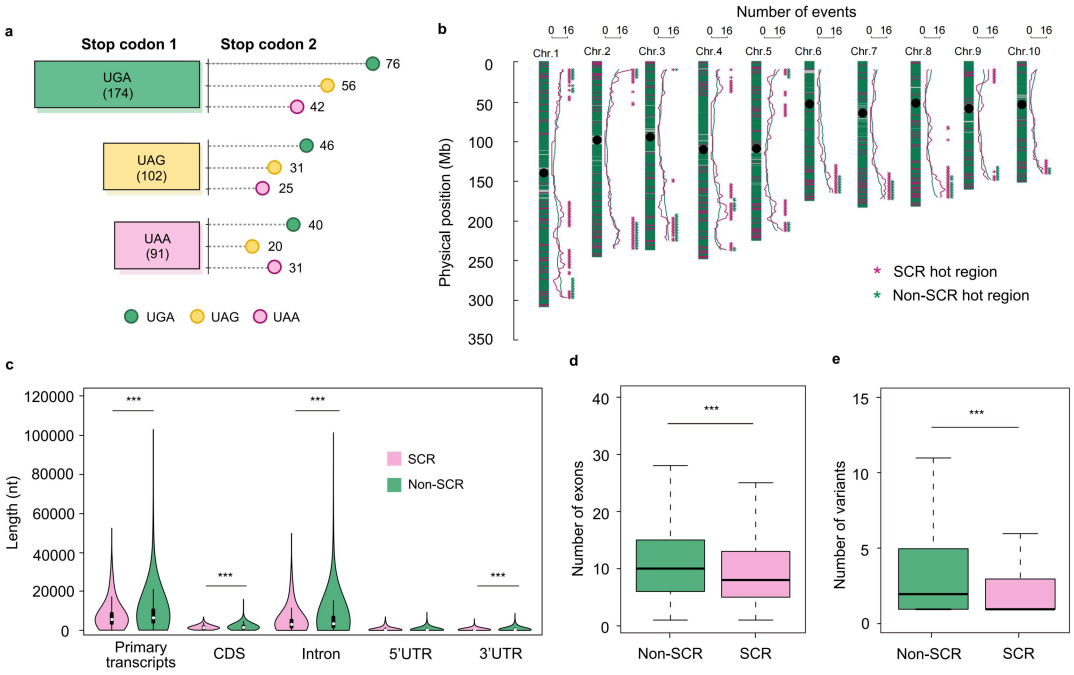


Fig. 3 | Characteristics of SCR events in maize. a Stop codon usage and the next stop codon preference in SCR events. Stop codon 1, readthrough stop codon; Stop codon 2, the next in-frame stop codon. **b** Genome-wide distribution of SCR (pink) and non-SCR (green) events in maize. For each chromosome, the distribution contains three columns. Left: SCR (pink) and non-SCR (green) events across the chromosomes. Black dots are centromeres. Middle: distribution patterns of SCR (pink) and non-SCR (green) events are shown by curves, shown as a 20-Mb sliding window with a 2-Mb step based on the B73 reference genome. Right: distribution of hotspot regions for SCR (pink) and non-SCR (green) events shown with asterisks. We defined hotspot regions as ≥5 SCR events and ≥335 non-SCR events in a 20-Mb window. **c** Violin plot describing the length distribution of primary transcripts and transcript elements in SCR and non-SCR events. The length of the primary transcript is the sum of CDSs, introns, and UTRs. CDS length represents the length of the open reading frame (ORF). Internal boxplots show the median values (white dots in the center) and the interquartile range (IQR; 0.25 and 0.75 quantiles) of the length distribution. **d** Number of exons in SCR and non-SCR primary transcripts. **e** Number of splice variants in SCR and non-SCR primary transcripts. For all boxplots, box edges represent the 0.25 and 0.75 quantiles. The bold lines represent the median values and the whiskers represent 1.5×IQR. ***$P < 0.001$ (two-sided Student's *t*-test).

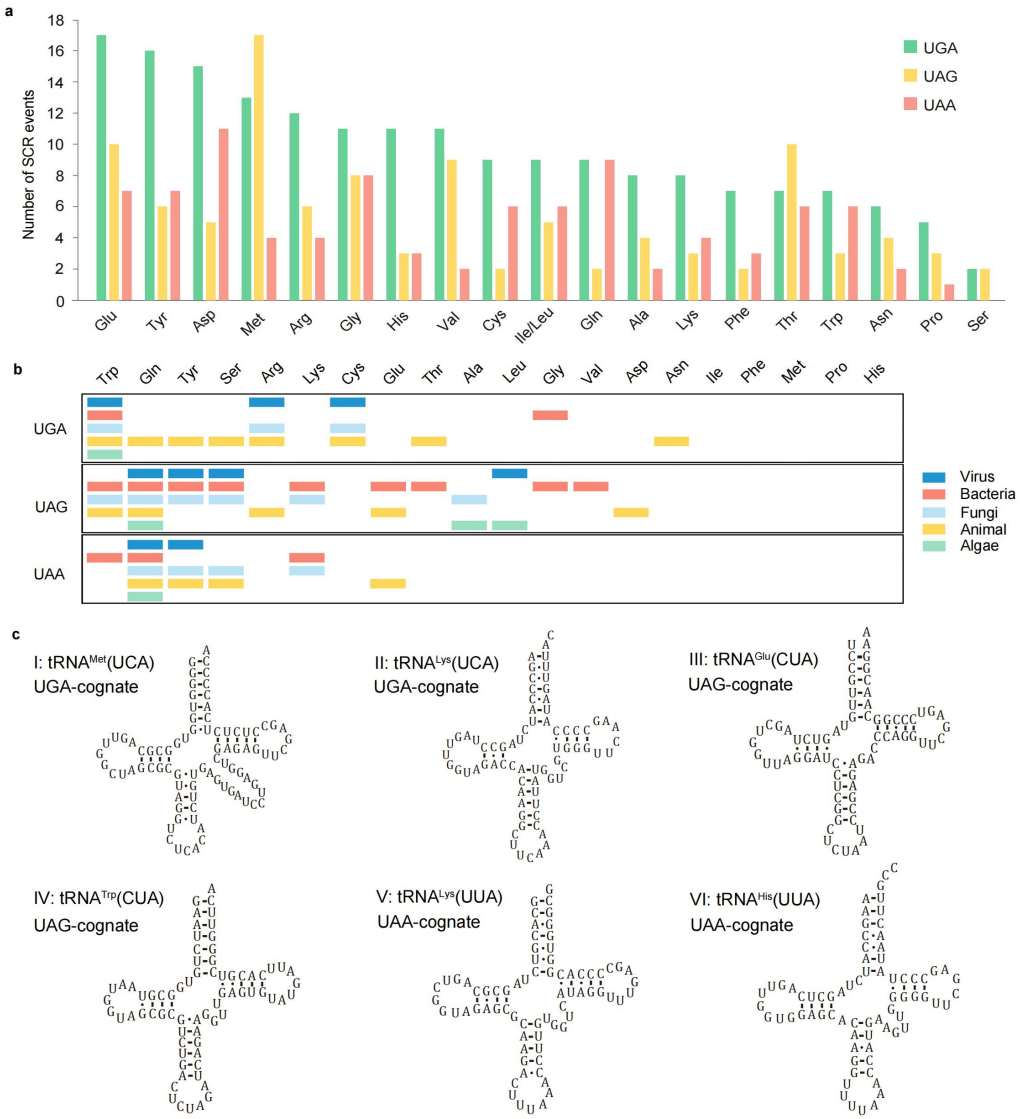


Fig. 4 | Stop codons are recoded as 20 amino acids in maize SCR events. a Frequency of amino acids specified by the three stop codons. The x-axis represents the 20 standard amino acids. The y-axis is the number of SCR events. **b** Previously reported amino acids translated at UGA, UAG, and UAA in different species. Blue, virus; orange, bacteria; light blue, fungi; yellow, animal; green, algae. **c** Examples of UGA-, UAG-, and UAA suppressor tRNAs identified by tRNAscan-SE and ARAGORN corresponding to amino acids shown at UGA, UAG, and UAA. The predicted secondary structures of the tRNAs are shown (I-VI): $tRNA^{Met}$ (UCA) and $tRNA^{Lys}$ (UCA) recognize UGA, $tRNA^{Glu}$ (CUA) and $tRNA^{Trp}$ (CUA) recognize UAG, and $tRNA^{Lys}$ (UUA) and $tRNA^{His}$ (UUA) recognize UAA.

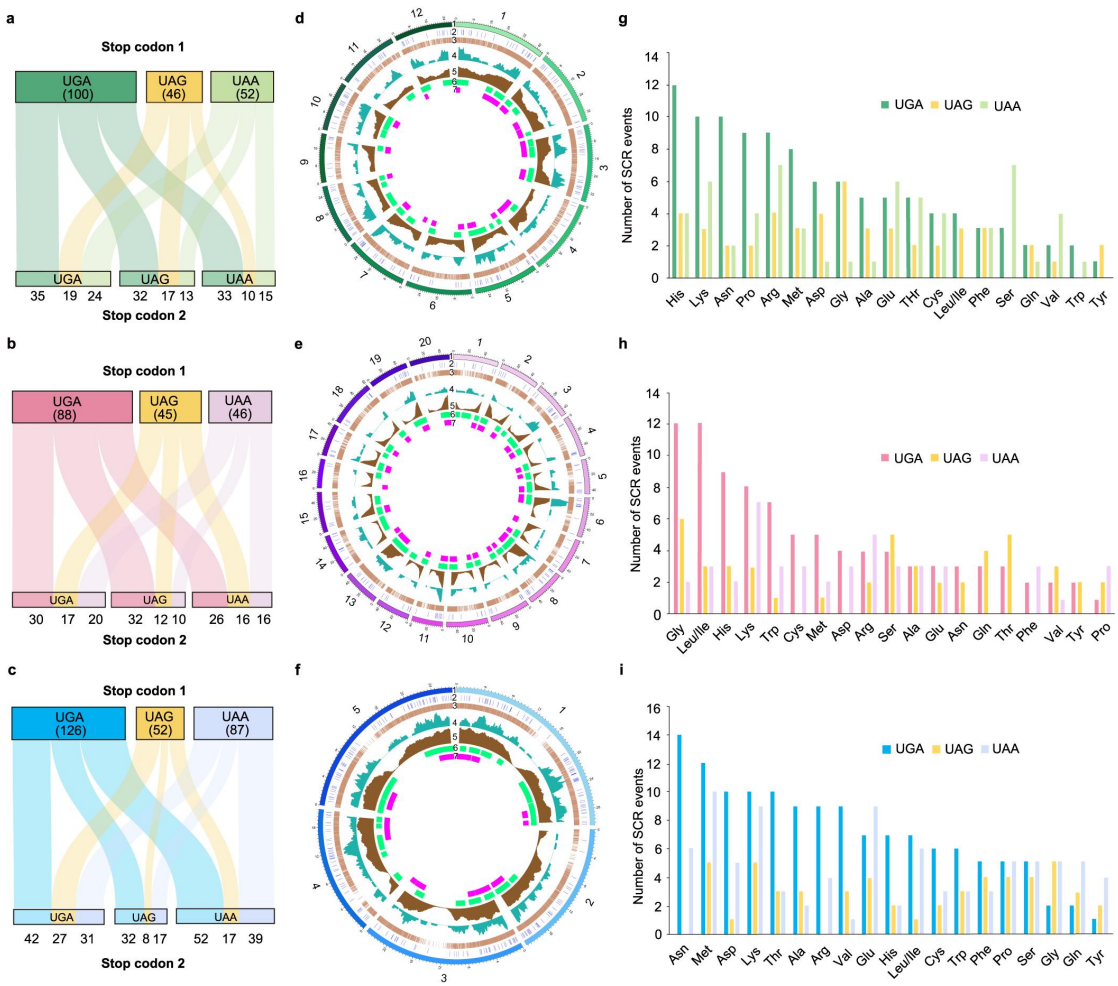


Fig. 5 | SCR events and recoding plasticity of stop codons in other plants. a, b, c Stop codon usage and the next stop codon preference of SCR events in rice (**a**), soybean (**b**), and *Arabidopsis* (**c**). Stop codon 1, readthrough stop codon; Stop codon 2, the next in-frame stop codon. **d, e, f** Circos plots showing the genome-wide distribution of SCR and non-SCR events in rice (**d**), soybean (**e**), and *Arabidopsis* (**f**). Each Circos plot consists of seven layers. 1, chromosomes; 2-3, distribution patterns of SCR (2) and non-SCR (3) events across the chromosomes; 4-5, distribution patterns of SCR (4) and non-SCR (5) events determined using a window size of 4 Mb with a step of 400 kb based on the Nipponbare reference genome for rice, a 10-Mb window with a 1-Mb step based on the ZH13 reference genome, and a 1.2-Mb window with a 100-kb step based on the Columbia-0 reference genome; 6-7, distribution of hotspot regions for SCR (6) and non-SCR (7) events. Hotspot regions were defined as ≥3 SCR events and ≥178 non-SCR events in a 4-Mb window for rice, ≥2 SCR events and ≥253 non-SCR events in a 10-Mb window for soybean, and ≥4 SCR events and ≥157 for non-SCR events in a 1.2-Mb window for *Arabidopsis*. **g, h, i** Frequency of amino acids specified by the

960 three stop codons in rice (**g**), soybean (**h**), and *Arabidopsis* (**i**) SCR events. The x-axis
961 represents the 20 standard amino acids. The y-axis is the number of SCR events.

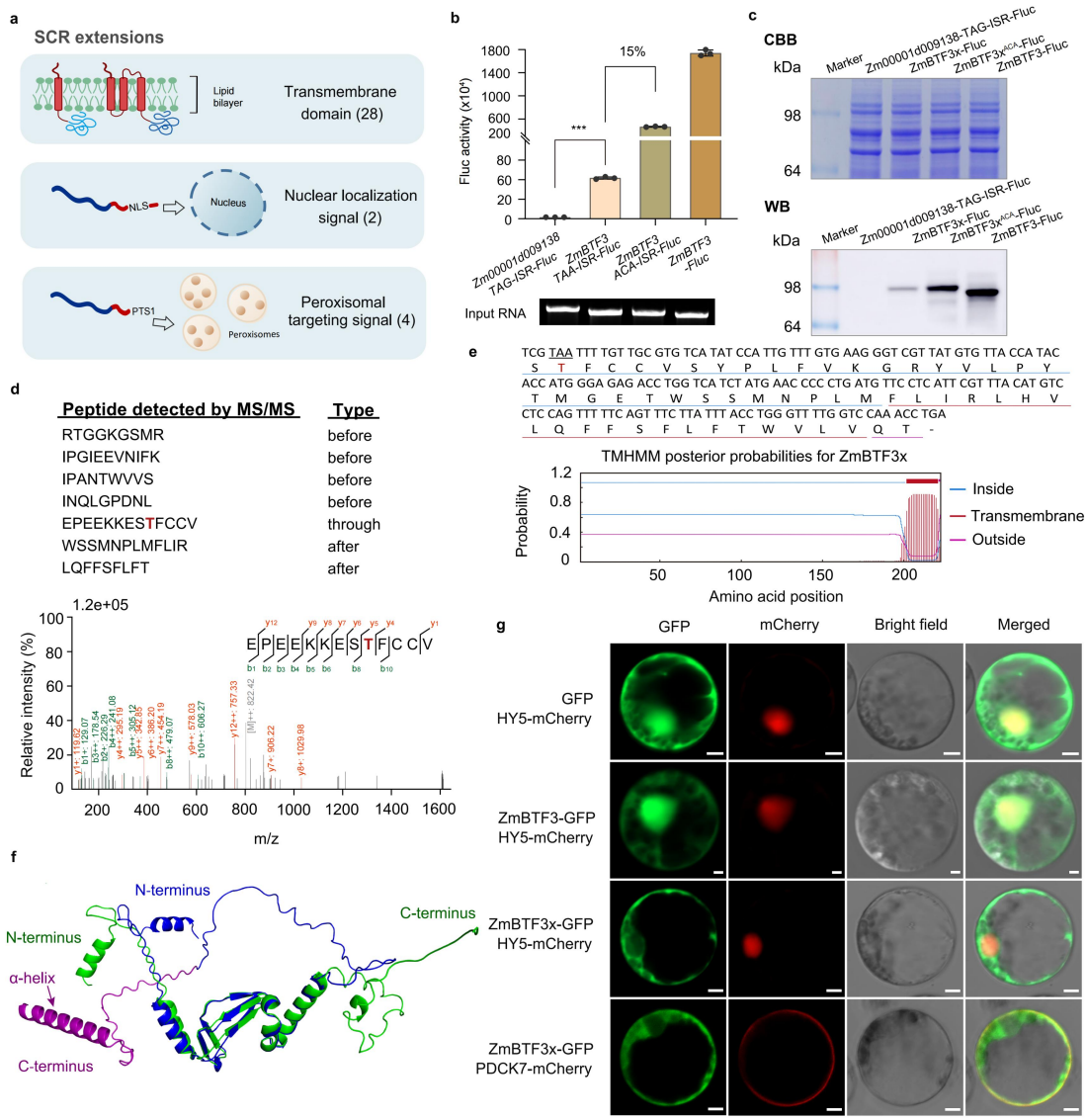


Fig. 6 | C-terminal extensions in maize SCR events contain multiple functional signals. a Multiple functional signals predicted in the extensions of SCR events. **b** Readthrough validation of *ZmBTF3* using the *in vitro* translation experiment. The plasmid containing *ZmBTF3-TAA-ISR-Fluc* was firstly transcribed *in vitro*, followed by *in vitro* translation using wheat germ extract. *ZmBTF3* with TAA mutated to ACA was used as the positive control of the translation to measure the efficiency of SCR. The *Zm00001d009138-TAG-ISR-Fluc* was used as the SCR-negative control. Luciferase activity reflects the readthrough activity. All experiments were performed in triplicates. Data are mean ± s.e.m (n=3). Statistical difference was calculated using a two-sided Student's *t*-test, ***$P < 0.001$. ISR, inter-stop codon region. **c** SDS-PAGE and immunoblot analysis of the ZmBTF3x-Fluc (ZmBTF3-TAA-ISR-Fluc), ZmBTF3x^{ACA}-Fluc (ZmBTF3-ACA-ISR-Fluc), and ZmBTF3-Fluc fusion proteins. Top:

973 Coomassie Brilliant Blue (CBB)-stained gel of the expressed products derived from (**b**).
974 Bottom: Immunoblot of the ZmBTF3x-Fluc, ZmBTF3x^{ACA}-Fluc, and ZmBTF3-Fluc fusion
975 proteins using an anti-Fluc antibody. The ZmBTF3x^{ACA}-Fluc was used as the positive control.
976 The Zm00001d009138-TAG-ISR-Fluc was used as the negative control. ISR, inter-stop codon
977 region. **d** MS/MS analysis of the digested peptides of the ZmBTF3x-Fluc fusion protein
978 excised from the CBB gel in (**c**). The sequences of MS-detected peptides, labeled as "before",
979 "through", and "after", are shown on the top. An MS/MS spectrum of the "through" peptide
980 EPEEKKESTFCCV is shown on the bottom. UAA was recoded as threonine (T) shown in red.
981 **e** Transmembrane domain prediction in the C-terminus of ZmBTF3x using the TMHMM
982 server. The blue, red, and magenta lines represent residues 1-201 inside the membrane, 202-
983 221 as a transmembrane helix, and 222-223 outside the membrane, respectively. The amino
984 acid sequence from the annotated stop codon to the next stop codon of BTF3x is shown on the
985 top. **f** AlphaFold structural models for ZmBTF3 and its readthrough protein ZmBTF3x. Green:
986 ZmBTF3; dark blue: parent protein region of the readthrough protein ZmBTF3x; magenta: C-
987 terminal extension region of ZmBTF3x. The predicted α-helix is shown at the C-terminus. **g**
988 Subcellular localization of ZmBTF3 and ZmBTF3x. Constructs encoding C-terminal GFP
989 fusions of ZmBTF3 or ZmBTF3x were transfected into maize protoplasts for localization
990 assays under a confocal microscope. GFP-vector transfected protoplasts were used as the
991 control. The experiments were performed three times with similar results. HY5-mCherry was
992 the nucleus marker and PDCK7-mCherry was the plasma membrane marker. Scale bars, 5 μm.

993 **Supplementary Information, Tables and Dataset have been submitted**
994 **as separate files.**
995
996 **Supplementary Information contains Supplementary Figures 1-10 and**
997 **References for Supplementary Table 5**
998
999 **Supplementary Table 1. SCR events identified in maize by proteogenomics.**
1000 **Supplementary Table 2. Peptides identified by mass spectrometry for SCR and**
1001 **parent proteins in maize.**
1002 **Supplementary Table 3. SCR events detected in maize by both proteogenomics**
1003 **and ribosome profiling.**
1004 **Supplementary Table 4. Non-SCR events identified in maize by proteogenomics.**
1005 **Supplementary Table 5. Standard amino acids translated by stop codons in the**
1006 **literature and their translation evidence in maize. I, UGA; II, UAG; III, UAA.**
1007 **UGA-, UAG-, and UAA suppressor tRNAs were predicted by tRNAscan-SE or**
1008 **ARAGORN.**
1009 **Supplementary Table 6. Predicted suppressor tRNAs in maize aligned with**
1010 **tRNAs in the NCBI NT database.**
1011 **Supplementary Table 7. SCR events identified in rice by proteogenomics.**
1012 **Supplementary Table 8. Non-SCR events identified in rice by proteogenomics.**
1013 **Supplementary Table 9. SCR events identified in soybean by proteogenomics.**
1014 **Supplementary Table 10. Non-SCR events identified in soybean by**
1015 **proteogenomics.**
1016 **Supplementary Table 11. SCR events identified in *Arabidopsis* by proteogenomics.**
1017 **Supplementary Table 12. Non-SCR events identified in *Arabidopsis* by**
1018 **proteogenomics.**
1019 **Supplementary Table 13. Functional signals predicted in C-terminal extensions**
1020 **of maize SCR events.**

1021 **Supplementary Table 14. Primers used in this study.**

1022

1023 **Supplementary Dataset 1: Motifs predicted by Tomtom in the C-terminal**
1024 **extensions of SCR transcripts in maize.**