Upstream open reading frames buffer translational variability during *Drosophila* evolution and development

Yuanqiang Sun[1,#], Yuange Duan[1,#], Peixiang Gao[1], Chenlu Liu[1], Kaichun Jin[1], Shengqian Dou[2], Wenxiong Tang[1], Hong Zhang[3], & Jian Lu[1,4,5,*]

1. State Key Laboratory of Gene Function and Modulation Research, Center for Bioinformatics, School of Life Sciences, Peking University, Beijing 100871, China
2. Eye Institute of Shandong First Medical University, State Key Laboratory Cultivation Base, Shandong Provincial Key Laboratory of Ophthalmology, Qingdao 266071, China
3. College of Ecology, Lanzhou University, Lanzhou 730000, China
4. Beijing Advanced Center of RNA Biology (BEACON), Peking University, Beijing 100871, China
5. Southwest United Graduate School, Kunming 650092, China

# Equal contribution

* Correspondence:

Jian Lu

Email: luj@pku.edu.cn

Abstract

Protein abundance tends to be more evolutionarily conserved than mRNA levels both within and between species, yet the mechanisms underlying this phenomenon remain largely unknown. Upstream open reading frames (uORFs) are widespread *cis*-regulatory elements in eukaryotic genomes that regulate translation, but it remains unclear whether and how uORFs contribute to stabilizing protein levels. In this study, we performed ribosome translation simulations on mRNA to quantitatively assess the extent to which uORF translation influences the translational variability of downstream coding sequences (CDSs) across varying contexts. Our simulations revealed that uORF translation dampens CDS translational variability, with buffering capacity increasing in proportion to uORF efficiency, length, and number. We then compared the translatomes at different developmental stages of two *Drosophila* species, demonstrating that uORFs buffer mRNA translation fluctuations during both evolution and development. Experimentally, deleting a uORF in the *bicoid* (*bcd*) gene—a prominent example of translational buffering—resulted in extensive changes in gene expression and phenotypes in *Drosophila melanogaster*. Additionally, we observed uORF-mediated buffering between primates and within human populations. Together, our results reveal a novel regulatory mechanism by which uORFs stabilize gene translation during development and across evolutionary time.

Key words: Upstream open reading frames (uORFs), TASEP modeling, stabilizing selection, translational buffer, interspecific gene expression, canalization, *bicoid*

Introduction

Organisms have evolved various strategies for the spatiotemporal regulation of gene expression [1-3]. This is important because aberrant gene expression can result in phenotypic defects or diseases, while the variation and evolution of gene expression patterns frequently promote phenotypic diversification and adaptation [4]. Although variations in mRNA abundance are widely observed within or between species, protein abundance tends to show stronger evolutionary constraint [5,6], as observed in yeasts [7-9], primates [10,11], and other organisms [12-14]. Nevertheless, the molecular mechanisms by which the conservation of protein abundance across species is achieved are largely unknown [5,6].

Eukaryotic mRNA translation is a crucial step in gene expression and is highly regulated by multilayered mechanisms [15-17]. Upstream open reading frames (uORFs), which are short open reading frames in the 5-terminal untranslated regions (5' UTRs) of eukaryotic mRNAs, play crucial roles in regulating mRNA translation. Approximately 50% of eukaryotic genes contain uORFs [18], and their evolution has been tightly shaped by natural selection [19-23]. The functions of uORFs have been explored in various contexts, including development [23-30], disease [31-35], and stress responses [36-43]. The prevailing consensus is that uORFs typically repress downstream coding sequence (CDS) translation by sequestering ribosomes, a process influenced by factors such as uORF length, position, and sequence context [34,43-48]. However, under stress conditions, certain uORFs can facilitate CDS translation by promoting ribosome reinitiation, illustrating their context-dependent functions [42,43,48-52].

Gene expression noise, which arises from the inherent stochasticity of biological processes such as transcription and translation, is generally detrimental to organismal fitness [53] and is primarily determined at the translational level [54]. Recent studies suggest uORFs might play essential roles in buffering translational noise and stabilizing protein expression. For example, Wu et al. (2022) demonstrated that uORFs reduce protein production rates to stabilize TOC1 protein levels, ensuring precise circadian clock function in plants [55]. Similarly, Bottorff et al. (2022) used a human cell reporter system to show that a ribosome stall in cytomegaloviral *UL4* uORFs buffers against CDS translation reductions [56]. Under stress, translation initiation is typically downregulated, yet most human mRNAs resistant to this inhibition contain translated uORFs, with a single uORF often being sufficient for resistance [42]. The computational model of Initiation Complexes Interference with Elongating Ribosomes (ICIER) suggests that derepression of downstream translation is a general mechanism of uORF-mediated stress resistance [43]. Despite these findings, the current understanding of uORFs in stabilizing translation is limited to single-gene cases or stressed conditions. It remains unclear whether and how uORFs affect gene translation variability on a genome-wide scale during evolution and development, and whether the identified mechanisms are universal or vary substantially among different taxa. To address these questions, a combination of modeling, genome-wide analyses, and comparative studies across species is required.

In this study, we first adapted the ICIER framework [43] to simulate the translating ribosome on a mRNA to quantitatively measure the extent to which uORF translation reduces the translational variability of the downstream CDS under different translation contexts. We then compared the translatomes of two closely related *Drosophila* species, *D. melanogaster* and *D. simulans*, and further supported the notion that uORFs could buffer the fluctuations of CDS translation during the development and evolution of *Drosophila*. The patterns also reappeared among primates and human populations. We next knocked out the *bicoid* (*bcd*) uORF, a case showing the significant buffering effect in our data, and observed broad changes in the embryonic transcriptome and phenotypic defects in *D. melanogaster*. Together, our results demonstrate a novel role for uORFs by maintaining translation stabilization during *Drosophila* evolution and development.

Results

An extended ICIER model for quantifying uORF buffering in CDS translation

To quantitatively assess how uORF translation modulates the variability of downstream CDS translation, we adapted the ICIER model [43], originally grounded in the totally asymmetric simple exclusion process (TASEP). TASEP has been extensively utilized to model the stochastic nature of ribosome movement along mRNA, capturing the effects of ribosome traffic jams, where ribosomes may slow down or stall when a site ahead is occupied [57-60]. The ICIER model extends this by simulating the interplay between scanning (40S) and elongating (80S) ribosomes, particularly focusing on how uORFs impact the overall translation rate of the main CDS [43].

We extended the original ICIER model [43] with several major modifications (Figure 1A). First, while the original ICIER model only considered the scenario where the elongating ribosome (80S) causes downstream scanning ribosomes (40S) to dissociate from the mRNA when they move along the mRNA and collide, recent findings have shown that upstream dissociation can also play a critical role in uORF-mediated regulation [56]. To incorporate this, we accounted for more complex ribosome interactions, including three possible scenarios where the 80S collides with 40S: i) 80S only causes the downstream 40S to dissociate from the mRNA with a probability of K_{down} ("downstream dissociation", K_{down} ranging from 0 to 1), following the original ICIER model; ii) the 80S only causes the upstream 40S to dissociate from the mRNA with a probability of K_{up} ("upstream dissociation", $K_{up} > 0$ and $K_{down} = 0$); and iii) a combination of the downstream and upstream dissociation models ("double dissociation", $K_{up} > 0$ and $K_{down} > 0$).

Second, the original ICIER model only considered the ribosome collision and dissociation in the uORF and counted the 40S scanning ribosome escaping from the uORF as a proxy of the CDS translation rate. In our extended model, we also considered the ribosome collision and dissociation events in the CDS

112 downstream of the uORF. We recorded the number of 80S ribosomes that completed translation at the
113 stop codon of a CDS (N_{EC}) or uORF (N_{Eu}) during a given time interval, using these counts as proxies
114 to quantify the translation rate of CDS or uORF. These indices allowed us to directly and quantitively
115 measure the impact of uORFs on CDS translation.
116
117 Third, while the original ICIER model accounted for only a single uORF within an mRNA, we extended
118 it to consider two uORFs coexisting on an mRNA molecule, allowing us to explore the possible
119 combinatorial effects of uORF-mediated translation regulation in a more complex yet more common
120 scenario, as previous studies have shown that uORFs tend to be clustered within genes [18].
121
122 These extensions allow for a more comprehensive exploration of uORF-mediated translational
123 buffering, offering deeper insights into how these regulatory elements might stabilize protein synthesis
124 across varying translation contexts.
125
126 **uORF-mediated buffering of CDS translation across different parameter settings**
127 To systematically investigate the extent to which uORF translation modulates the variability of
128 downstream CDS translation, we conducted simulations across various parameter settings using three
129 dissociation models: upstream, downstream, and double dissociation, each reflecting different possible
130 interactions between scanning and elongating ribosomes on the mRNA. We considered a range of
131 parameters crucial to the translation process (Table S1), including the length of the 5' leader before the
132 uORF (fixed at 150 nucleotides), the length of the uORF itself (ranging from 2 to 100 codons), the
133 distance between the uORF stop codon and the CDS start codon (150 nucleotides), the length of the
134 CDS (500 codons), and the length of the 3' UTR (150 nucleotides). Additionally, we modeled the
135 probabilities associated with ribosome movement and initiation, such as the probability of a 40S
136 ribosome moving to the next nucleotide (ν_s = 0.3), the probability of an 80S ribosome moving to the
137 next position within the uORF (ν_{Eu} = 0.3) and within the CDS (ν_{EC} = 0.5), and the probability of
138 loading a new 40S ribosome at the 5' end of the mRNA (R_{in}). We also explored different probabilities
139 of translation initiation at both the uORF start codon (I_{uORF}) and the CDS start codon (I_{CDS}). Key
140 parameters were adapted from the original ICIER model [43], ensuring a robust basis for comparison
141 while allowing exploration of additional variables that influence uORF-mediated translational buffering.
142
143 In our simulations, R_{in} values were varied to simulate fluctuations in translational resources, such as
144 ribosome availability, that could arise from genetic differences or environmental changes during
145 evolution or development. By generating 1,000 R_{in} values following either uniform or exponential
146 distributions (ranging from 0 to 0.1, Figure S1A), we aimed to capture the natural variability in
147 ribosome loading rates that might occur across different species, individuals, or developmental stages.

These values were then fed into the simulation models to evaluate their impact on both the level and variability of translation rate for CDSs, with the number of ribosomes completing translation on CDSs (N_{EC}) in a given time interval used as a proxy for translational rate (Figure S1B).

Across all model settings, uORF translation ($I_{uORF} > 0$) consistently reduced CDS translation rate (N_{EC}) by about 30% to 80% as I_{uORF} increased from 0.1 to 0.5, compared to scenarios where the uORF was absent or untranslated ($I_{uORF} = 0$) (Figure S2 and S3). This confirms the inhibitory effect of uORFs on downstream CDS translation. Notably, the coefficient of variation (CV), a measurement of variability for translation rate (N_{EC}), was lower when the uORF was translated (Figure 1B). These CV values further decreased by approximately 10% to 25% as I_{uORF} increased from 0.1 to 0.5 (Figure 1B), and this buffering effect persisted across different parameter settings (Figure S4 and S5). Moreover, the CV of translation rate (N_{EC}) further decreased by about 6% to 30% as uORF length (L_{uORF}) increased from 2 to 100 codons (Figure 1C and S6-7). These simulation results indicate that uORF translation can reduce variability in downstream CDS translation, with the buffering capacity positively correlated with both uORF translation initiation efficiency and length under the applied simulation conditions.

To more quantitatively investigate the relationship between changes in translation rate in uORF (N_{EU}) and CDS (N_{EC}), for each parameter setting, we used the median value (R_{inm}) of the 1,000 R_{in} inputs as the baseline, calculating the corresponding N_{ECm} and N_{EUm}values. We then calculated the changes of N_{EC} (ΔN_{EC}) and N_{EU} (ΔN_{EU}) relative to N_{ECm} and N_{EUm} for each R_{in} value, respectively. Across different parameter settings, ΔN_{EU} was consistently and significantly positively correlated with ΔN_{EC} ($P < 0.001$, Spearman's correlation) (Figure S8-9), indicating that fluctuations of the ribosome loading rate influence the translation of both uORFs and CDSs in the same direction. Nevertheless, our simulations showed that variations in R_{in} led to a larger change in N_{EU} than in N_{EC}, as the median value of δ [defined by $\log_2(\Delta N_{EC}/\Delta N_{EU})$] was consistently less than 0 across various I_{uORF} and I_{CDS} combinations (Figures 1D and S10-11). This finding suggests that translational fluctuation in uORFs is greater than that in downstream CDSs, indicating that uORFs can buffer against upstream fluctuations.

Simulations of mRNAs with two uORFs revealed patterns consistent with a buffering role for uORFs (Figure S12). To compare the buffering effects of a single uORF versus two uORFs, we calculated the ratio of the CV of N_{EC} with two uORFs to that with a single uORF. A ratio less than 1 suggests that two uORFs provide greater buffering than a single uORF. For comparability, we examined the CV of N_{EC} where the I_{uORF} in the single-uORF model equals I_{uORF1} (I_{uORF} of the first uORF) in the two-uORF model, both ranging from 0 to 0.5. The CV ratio consistently remained below 1 across a range of I_{uORF2} (I_{uORF} of the second uORF) (Figure S13), indicating that two uORFs offer stronger buffering than a single uORF.

Collectively, these simulations collectively suggest: (1) uORF-mediated translational control buffers against CDS translation variability, (2) uORFs exhibit greater translational fluctuations than downstream CDSs, and (3) the buffering capacity of uORFs positively correlates with their translation initiation efficiency and length. Subsequently, we sought to validate these simulation results in a biological context by confirming the uORF-mediated buffering effect during organismal evolution and development, as both processes face environmental and/or genetic changes that frequently disturb mRNA translation.

Generating matched translatome data from two *Drosophila* species for comparative analysis

To validate the uORF-mediated translational buffering during *Drosophila* evolution, we performed a comparative analysis of the translatomes of two closely related *Drosophila* species, *D. melanogaster* and *D. simulans*, which diverged approximately 5.4 million years ago [61]. We generated high-throughput sequencing data for *D. simulans*, including transcriptome (mRNA-Seq) and translatome (Ribo-Seq) profiles from various developmental stages and tissues, including embryos at 0–2 h, 2–6 h, 6–12 h, and 12–24 h, third-instar larvae, P7–8 pupae, female and male bodies, and female and male heads. In total, we obtained approximately 786 million high-quality reads for *D. simulans* (Table S2). These datasets were designed to be directly comparable to the previously published *D. melanogaster* data [23], with identical embryonic stages and tissue types. For each sample, we followed established procedures [62-66] to calculate the translational efficiency (TE) for each feature (CDS or uORF). TE serves as a proxy for the translation rate at which ribosomes translate mRNA into proteins, typically quantified by comparing the density of ribosome-protected mRNA fragment (RPF) to the mRNA abundance for that feature (see **Materials and Methods**). This comprehensive comparative translatome analysis, utilizing matched developmental stages and tissues between the two species, allowed us to assess the uORF-mediated translational buffering effects during evolution.

Translational conservation and dominance of uORFs between *Drosophila* species

Given that the translation of uORFs is a crucial determinant of their functional impact, we first characterized the translational profiles of uORFs in *D. melanogaster* and *D. simulans* to explore the evolutionary roles of uORFs in translation regulation. We focused on canonical uORFs that initiate with an ATG start codon in the 5′ UTR and terminate with a stop codon (TAA, TAG, or TGA). Because the ATG start codon is the defining feature of a canonical uORF and tends to be more conserved than its downstream sequence [67], we defined uORF conservation based on the presence of the ATG start codon in the 5′ UTR of *D. melanogaster* and its orthologous positions in *D. simulans*, regardless of differences in the stop codon. Using this criterion, we identified 18,412 canonical uORFs with conserved start codons between the two species. Additionally, we identified 2,789 canonical uORFs specific to *D. melanogaster* and 2,440 canonical uORFs specific to *D. simulans*. The TE values of the conserved

221 uORFs were highly correlated between the two species across all developmental stages and tissues
222 examined, with Spearman correlation coefficients (*rho*) ranging from 0.478 to 0.573 (Figure 2A). In
223 contrast, TE of CDSs exhibited a significantly higher correlation between the two species in the
224 corresponding samples compared to that of uORFs, with Spearman's *rho* ranging from 0.588 to 0.806
225 (P = 0.002, Wilcoxon signed-rank test; Figure 2A). This observation is consistent with our simulation
226 results, which indicate that uORFs experience greater translational fluctuations than their downstream
227 CDSs. Notably, conserved uORFs exhibited significantly higher TEs compared to species-specific
228 uORFs in both species. The median TE of conserved uORFs was 1.62 times that of non-conserved
229 uORFs in *D. simulans*, while the corresponding ratio in *D. melanogaster* was 1.52 (Figure 2B).
230
231 In *D. melanogaster*, 7,259 (52.2%) genes had no uORFs, 2,687 (19.3%) had a single uORF, and 3,961
232 (28.5%) contained multiple uORFs. Among genes with multiple uORFs, we defined the uORF with the
233 highest TE as the dominant uORF for that gene, as TE is one of the most relevant metrics for assessing
234 uORF function [45,67]. The median TE of the dominant uORF was 4.84 times that of the second-highest
235 uORF within the same gene in *D. melanogaster*, and the corresponding ratio was 5.21 times in *D.*
236 *simulans* (Figure S14). To assess the consistency of this dominance across different tissues and
237 developmental stages, we identified 3,072 multiple-uORF genes in *D. melanogaster* with at least one
238 translated uORF (TE > 0.1 in at least five stages/tissues). Of these, 569 genes consistently used the
239 same dominant uORF across the measured samples, significantly higher than the number expected
240 under randomness (5 genes, 95% confidence interval: 1–10) based on shuffling the TEs of uORFs 1,000
241 times (Figure S15). This trend was also observed in *D. simulans* and persisted under different thresholds
242 for defining "translated uORFs" (Figure S15). These results suggest that genes with multiple uORFs
243 tend to retain the same dominant uORF across developmental stages, indicating that the dominant uORF
244 may serve as the key translational regulator of the downstream CDS. Moreover, we found that the
245 dominant uORFs showed a higher proportion of conserved uATGs than the other translated uORFs
246 (median proportion 82.5% versus 78.8%, P < 0.001) (Figure 2C). Additionally, the absolute values of
247 TE fold-change ($|\log_2$TE-FC$|$) between the two species were, on average, 23.2% smaller for dominant
248 uORFs than for other conserved uORFs (Figure 2D), suggesting that dominant uORFs are more likely
249 under stronger stabilizing selection. These findings suggest that, in genes with multiple uORFs, the
250 dominantly translated uORF may play a more important role in regulating CDS translation than the
251 other uORFs.
252
253 **uORFs buffer interspecific translational divergence of CDSs**
254 To investigate the relationship between translational changes in uORFs and their downstream CDSs
255 across species, we analyzed the translation efficiency (TE) of uORFs and corresponding downstream
256 CDSs in *D. melanogaster* and *D. simulans*. Consistent with our simulations that the translation of a
257 uORF is tightly linked to that of downstream CDS, uORFs exhibited a modest, yet statistically

258 significant, positive correlation with the TE of their downstream CDSs across all samples analyzed (P
259 < 0.001, Spearman's correlation) (Figure 3A). We then compared the interspecific TE change of a uORF
260 ($\beta_u = TE_{uORF,sim}/TE_{uORF,mel}$) with that of its corresponding CDS ($\beta_C = TE_{CDS,sim}/TE_{CDS,mel}$)
261 between *D. melanogaster* and *D. simulans*. We found β_u is significantly positively correlated with β_C
262 across all samples (P values < 0.001, Spearman's correlation) (Figure S16). These results align well
263 with our simulations, which showed that fluctuations in translational factors (such as ribosomes)
264 influence both uORF and CDS translation in the same direction (Figure S8–S9).
265
266 While the direction of TE changes for uORFs and CDSs tends to be consistent, our simulations suggest
267 that the magnitude of TE changes in CDSs is generally smaller than that in uORFs, due to the buffering
268 effect of uORF (Figure S10–S11). To validate this, we first identified uORFs and CDSs with significant
269 interspecific TE differences by assessing whether β_u or β_C significantly deviated from 1, using an
270 established statistical framework [23]. This analysis uncovered 1,151 to 4,189 CDSs with significant
271 interspecific TE changes (FDR < 0.05) (Table 1), with genes involved in development, morphogenesis,
272 and differentiation being significantly enriched during embryonic stages (Figure S17). Conversely,
273 genes related to metabolism, response to stimuli, and signaling were enriched in larval, pupal, and adult
274 stages (Figure S17). Additionally, we identified 144 to 1,193 uORFs with significant TE differences
275 between species, accounting for approximately 1–15% of expressed uORFs (Table 1). Note that due to
276 their shorter length and generally lower TE, uORFs had considerably lower read counts than CDSs,
277 limiting the statistical power to detect significant interspecific TE differences for uORFs. This trend
278 consistently holds whether analyzing all expressed uORFs (Figure S18A) or only highly expressed
279 genes (Figure S18B). Thus, the fewer uORFs showing significant TE divergence likely reflects lower
280 read counts and statistical sensitivity rather than reduced translational variability relative to CDSs. In
281 fact, the absolute values of $\log_2$(fold change) of TE for uORFs between *D. melanogaster* and *D.*
282 *simulans* were significantly greater than those observed for corresponding CDSs across all samples (P
283 < 0.001, Wilcoxon signed-rank test; Figure 3B), suggesting that the magnitude of TE changes in CDSs
284 is generally smaller than that in uORFs, due to the buffering effect of uORF.
285
286 We further quantitatively compared the magnitude of interspecific TE changes between uORFs and
287 their corresponding CDSs using a previous method [23]. We defined $\gamma = \beta_c/\beta_u$ for a uORF-CDS pair
288 within the same mRNA and tested whether γ was significantly different from 1 to identify pairs with
289 differential TE changes between uORFs and CDSs (Figure S19). When γ < 1 and β_u >1, or γ > 1 and
290 β_u <1, it indicates that the TE change for the CDS is smaller than that for the uORF (Figure S19).
291 Among CDS-uORF pairs where β_u > 1, nearly all (8–487) showed a significant γ < 1 in each sample,
292 except for one pair from the pupal stage where γ > 1 (Table 1). This suggests that the magnitude of TE
293 changes in CDSs was generally smaller than in uORFs when uORF TE increased, and vice versa when

uORF TE decreased (β_u < 1) (Table 1). These comparative translatome analyses indicate that uORFs buffered downstream CDS translation changes during the evolutionary divergence of *D. melanogaster* and *D. simulans*.

uORF buffering is influenced by its conservation, dominance, and length

To investigate how the conservation level and translation patterns of uORFs influence their buffering capacity on CDS translation, we categorized genes expressed in each pair of samples into three classes: Class I, genes with conserved uORFs that are dominantly translated (i.e., exhibiting the highest TE among all uORFs within the same gene) in both *Drosophila* species; Class II, genes with conserved uORFs that are translated in both species but not dominantly translated in at least one; and Class III, the remaining expressed genes. We then compared the absolute values of interspecific TE changes of the CDS ($|\beta_c|$) across these three categories. Significant differences in $|\beta_c|$ were observed, with a consistent hierarchy of Class I < II < III across all pairs of samples (Figure 3C and S20). On average, Class I genes exhibited an average of 8.18% and 23.8% lower $|\beta_c|$ values compared to Class II and Class III, respectively. This indicates that conserved and dominantly translated uORFs exert a stronger buffering effect on CDS translation.

To further validate the simulation results suggesting that longer uORFs have a stronger buffering effect (Figure 1C and S6–7), we divided genes expressed in each pair of samples into three groups: those without translated uORFs (No), those with short uORFs (short, total length below the median), and those with long uORFs (long, total length above the median). Consistently, longer uORFs were associated with stronger buffering effects on CDS translation across all pairs of samples (Figure 3D and S21). Specifically, genes with longer uORFs showed 12.7% and 26.5% lower $|\beta_c|$ values compared to genes with short uORFs or no uORFs, respectively.

Overall, these findings underscore that the buffering capability of a uORF is positively correlated with its conservation level, translation dominance, and length.

uORFs buffer translational fluctuations during *Drosophila* development

Gene expression undergoes dynamic changes during *Drosophila* development, with significant alterations in the translation program to meet developmental demands, including shifts in ribosome loading rates [24,68,69]. Therefore, we extended our analysis to investigate the role of uORFs in buffering these translational fluctuations, hypothesizing that uORFs could mitigate them during development. According to our hypothesis, if a gene has a translated uORF in *D. melanogaster* but not in its orthologous gene in *D. simulans*, then the translation of this gene is likely more stable across developmental stages in *D. melanogaster* than its ortholog in *D. simulans*, and vice versa.

To test this hypothesis, we compared the CV of CDS TE across 10 developmental stages in *D. melanogaster* and in *D. simulans*, respectively. Genes with translated uORFs (TE>0.1 in at least one sample) in *D. melanogaster* exhibited significantly 22.5% smaller CVs in *D. melanogaster* than their orthologs lacking these uORFs in *D. simulans*, indicating more stable translation (Figure 4A). Consistently, for genes with translated uORFs in *D. simulans* but not in *D. melanogaster*, the CVs were 13.3% lower in *D. simulans* compared to their orthologs lacking these uORFs in *D. melanogaster* (Figure 4B). Consistent results were observed when a uORF is required to be translated (TE > 0.1) in all 10 samples (Figure S22A), in 4 embryonic stages (Figure S22B), or in 6 stages including embryos, larva, and pupa (Figure S22C). Moreover, within each species, genes with translated uORFs also showed less variability in CDS TE across developmental stages compared to those without translated uORFs, with a 31.8% reduction in *D. melanogaster* and a 28.9% reduction in *D. simulans* (Figure 4C). This effect was consistent across different thresholds for defining "translated uORFs" (Figure S23).

These results suggest that uORFs function as translational buffers, reducing gene translation fluctuations during *Drosophila* development.

Knocking out the uORF of *bcd* increased *bcd* CDS translation in *D. melanogaster*

After verifying the uORFs-mediated translational buffering during *Drosophila* evolution and development, we next aimed to directly explore the biological function of these buffering-capable uORFs in vivo. We first applied stringent criteria to identify uORFs with significant buffering effects on CDS translation between *D. melanogaster* and *D. simulans*. Specifically, we looked for 1) uORFs with significant TE changes ($|log_2(\beta_u)| > 1.5$, adjusted $P < 0.05$), 2) negligible changes in its corresponding CDS translation ($|log_2(\beta_C)| < 0.05$, adjusted $P > 0.05$), and 3) a significant difference between the magnitude of these changes ($|\log_2(\gamma)| > 1.5$, adjusted $P < 0.05$). We identified 131 uORF-CDS pairs in 103 genes that meet these criteria in at least one stage/tissue (Table S3), with a majority of the genes (67%, 69 out of 103) from embryonic stages (Figure S24A), suggesting a crucial role for uORFs in maintaining translational stability during early development. Among these genes, one notable case is the *bicoid* (*bcd*) gene, a master regulator of anterior-posterior axis patterning during early embryogenesis [70-72]. The *bicoid* gene contains a 4-codon uORF (excluding the stop codon) in its 5' UTR, and branch length score (BLS) analysis [18] showed that the start codon (uATG) of the *bcd* uORF is highly conserved across the *Drosophila* phylogeny, with a BLS of 0.90 (on a scale from 0 to 1, where higher values indicate greater conservation) (Figure 5A and S24B). Ribo-Seq data revealed that in 0–2 h embryos, the TE of the *bcd* uORF varied more than threefold, while the TE of the CDS was virtually the same between *D. melanogaster* and *D. simulans* (Figure 5B and Table S3). This suggests that the *bcd* uORF significantly buffers translation during early development.

366

367 To investigate the regulatory role of the *bcd* uORF, we used CRISPR-Cas9 to knock out its start codon
368 in *D. melanogaster*, generating two mutant homozygotes (uKO1/uKO1 and uKO2/uKO2) with a
369 genetic background matched to that of the wild-type (WT) (Figures. 6A and S25). To determine whether
370 these mutations enhanced *bcd* CDS translation, we performed ribosome fractionation followed by qPCR
371 [73-75], comparing *bcd* mRNA levels in polysome and monosome fractions (P-to-M ratio) from 0–2 h
372 embryos of uKO1/uKO1, uKO2/uKO2, and WT flies (Figure 6B). A larger P-to-M ratio means more
373 mRNAs are enriched in the polysome fractions and bound by more ribosomes, thus indicative of higher
374 translation efficiency. P-to-M ratios were higher in the mutants compared to WT at 29°C (Figure 6C),
375 with a similar but less pronounced trend at 25°C, where the difference between uKO2/uKO2 and WT
376 was not statistically significant (Figure 6C). These findings, along with the known impact of
377 temperature on gene expression and phenotypic plasticity [76,77], suggest that the regulatory function of
378 the uORF and overall translation efficiency are temperature-sensitive, highlighting a complex interplay
379 between environmental conditions and gene regulation. To further confirm the *bcd* uORF's regulatory
380 function, we conducted dual-luciferase reporter assays. The 5' UTR from *bcd* mutant (uKO1 or uKO2)
381 or WT was cloned into a *Renilla* luciferase reporter construct (Figure 6D). Luciferase activity was
382 significantly higher in uKO1 and uKO2 compared to the WT 5' UTR, confirming the repressive role of
383 the *bcd* uORF.

384

385 ***bcd* uORF mutants show wide transcriptomic alteration during *Drosophila* embryogenesis**

386 Since Bcd regulates the expression of many zygotic genes [70-72], we anticipated that the increased
387 translation of *bcd* resulting from disrupting the uORF would influence *Drosophila* transcriptomes and
388 phenotypes. To verify this notion, we performed RNA sequencing on embryos from WT and
389 uKO2/uKO2 flies at four developmental stages (0–2 h, 2–6 h, 6–12 h, and 12–24 h) under both 25°C
390 and 29°C conditions, using two biological replicates (Table S4 and Figure S26). Differential expression
391 analysis revealed widespread alterations in gene expression between WT and mutant embryos, with the
392 number of differentially expressed genes (DEGs) increasing over developmental time and at higher
393 temperature (Figure 6E). At 25°C, we identified 674, 817, 1047, and 2,041 DEGs in 0–2 h, 2–6 h, 6–
394 12 h, and 12–24 h embryos, respectively; while at 29°C, we detected 3,884, 2,358, 2,901, and 4,164
395 DEGs in the corresponding stages (Figure 6E). The majority of DEGs were stage-specific, with only a
396 small fraction consistently differentially expressed across all four stages. Functional enrichment
397 analysis of the DEGs revealed distinct biological pathways affected at each stage, including cell
398 morphogenesis and pattern specification in 0–2 h embryos, metabolic processes and tissue development
399 in 2-6 h and 6-12 h embryos, and mitochondrial respiration in 12–24 h embryos (Figure 6F). Notably,
400 direct targets of Bcd [78] were significantly enriched among the DEGs in three out of the four stages, with
401 the exception of 2-6 h embryos (Figure S27). RT-qPCR validation of 20 target genes of Bcd confirmed
402 the reliability of the RNA-seq differential expression analysis (Figure S28). Together, these findings

403 demonstrate that disruption of the *bcd* uORF leads to widespread transcriptional changes during
404 *Drosophila* development, affecting processes ranging from embryogenesis to postembryonic
405 metabolism.
406
407 ***bcd* uORF mutants display decreased hatching rates and starvation resistance**
408 Given the widespread transcriptome alterations, we anticipated phenotypic abnormalities in the *bcd*
409 uORF mutants. As expected, both uKO1/uKO1 and uKO2/uKO2 mutants exhibited significantly lower
410 hatching rates compared to WT [P = 1.4×10^{-6} and 7.9×10^{-6}, respectively, Wilcoxon rank-sum test
411 (WRST); Figure 7A]. At 25°C, uKO1/uKO1 mutants produced fewer offspring than WT flies ($P < 0.001$,
412 WRST, Figure 7B). Given that *bcd* is a maternal gene, we expected reciprocal crosses between
413 uKO1/uKO1 mutants and WT flies to produce different outcomes. Indeed, crossing uKO1/uKO1 males
414 with WT females resulted in offspring numbers similar to those from WT crosses, while crossing
415 uKO1/uKO1 females with WT males yielded offspring numbers comparable to crosses between
416 uKO1/uKO1 mutants (Figure 7B). This verified that the reduction in offspring is due to maternal defects
417 in the uKO1/uKO1 mutants. Similar patterns were observed for uKO2/uKO2 mutants at 25°C (Figure
418 7B). Notably, the fecundity reduction in uORF-KO mutants was more pronounced at 29°C (Figure 7C).
419 Furthermore, crossing uKO1/uKO1 and uKO2/uKO2 mutants also produced significantly fewer
420 progeny than WT flies (Figure 7C), ruling out genetic background or off-target effects. Collectively,
421 these data demonstrate that disrupting the *bcd* uORF significantly impaired hatchability and fertility.
422
423 We also found that both uKO1/uKO1 and uKO2/uKO2 female mutants perished significantly faster
424 than WT flies under starvation conditions (Figure 7D). Males showed similar tendencies, although the
425 difference was not statistically significant for uKO1/uKO1 mutants (Figure 7D). These data suggest
426 that the knockout of *bcd* uORF diminished starvation resistance in adults, likely due to embryogenesis
427 abnormalities induced by the *bcd* uORF deletion, even in those that successfully developed to adulthood.
428
429 **Conservation of uORF-mediated translational buffering in primates**
430 To explore the generality of uORF-mediated translational buffering across evolutionary clades, we
431 analyzed previously published transcriptome and translatome data from three tissues (brain, liver, and
432 testis) in humans and macaques [79]. We identified 33,680 canonical uORFs in humans and 29,516 in
433 macaques, with 24,385 conserved between the two species. Despite the larger number of uORFs in
434 primates compared to *Drosophila* due to differences in genome size and gene number, the median TE
435 of conserved uORFs was 1.79 times that of non-conserved uORFs in humans, and the corresponding
436 ratio was 3.43 in macaques (Figures 8A and S29). TEs of uORFs were positively correlated between
437 humans and macaques across all tissues ($P < 0.001$, Figure 8B). Additionally, significant positive
438 correlations were observed between the TEs of uORFs and their corresponding coding sequences (CDSs)
439 in all tissues (Figure S30). Although interspecific TE divergence of uORFs (β_u) and CDSs (β_C) were

positively correlated ($P < 0.001$, Figure 8C), uORFs generally exhibited larger divergence (Table S5). Notably, longer uORFs showed stronger buffering effects on CDS translation, reducing interspecific TE divergence by 12.9% and 38.0% compared to genes with short or no uORFs (Figures 8D and S31).

As in *Drosophila*, we categorized expressed human genes into three classes based on uORF conservation and translation: Class I, genes with conserved and dominantly translated uORFs in both humans and macaques; Class II, genes with conserved uORFs translated in both species but not dominantly in at least one; and Class III, the remaining expressed genes (Figures 8E and S32). Consistent with findings in *Drosophila*, significant differences in $|\beta_c|$ between humans and macaques were observed in the order of Class I < II < III, with Class I genes showing 9.8% and 17.1% lower $|\beta_c|$ values compared to Class II and Class III genes, respectively (Figures 8E and S32). These similarities between primates and *Drosophila*—two clades that diverged over 700 million years ago [80] —suggest that uORF-mediated translational buffering is a widespread mechanism for stabilizing gene translation across evolutionary clades.

We also analyzed matched mRNA-Seq and Ribo-Seq data from 69 human lymphoblastoid cell lines [81,82] to test whether uORFs buffer against translational variability across different individuals within humans. Genes with translated uORFs exhibited, on average, 8.65% lower CVs in CDS TE across individuals than genes without translated uORFs (Figure 8F), with longer uORFs showing stronger buffering effects (Figure S33). Collectively, these findings suggest that uORFs play a crucial role in reducing translational variability both across evolutionary clades and within species.

Discussion

Translational control is vital for maintaining protein homeostasis and cellular activities [83,84]. However, the mechanisms underlying the conservation of protein abundance across species remain largely unknown [5,6]. In this study, we extended the ICIER model and conducted simulations to explore uORFs' regulatory roles in buffering translation during evolution. Our simulations demonstrated that uORF translation reduces variability in downstream CDS translation, with buffering capacity positively correlated with uORF translation efficiency, length, and number. Comparative translatome analyses across developmental stages of two *Drosophila* species provided evidence that uORFs mitigate interspecific differences in CDS translation. Similar patterns were observed between humans and macaques, and between human individuals, suggesting that uORF-mediated buffering is an evolutionarily conserved mechanism in animals. Additionally, in vivo experiments showed that knocking out the *bcd* uORF led to aberrant embryogenesis and altered starvation resistance, with more pronounced phenotypic effects at higher temperatures, supporting the role of uORFs in fine-tuning translation and phenotypic outcomes.

While the prevailing consensus on uORFs' functions is that they repress the translation of downstream CDS by sequestering ribosomes [34,43-48], recent studies have suggested uORFs might play essential roles in stabilizing protein expression [43,55,56]. While these studies have primarily focused on the role of uORFs in single-gene cases or under stress conditions, our work expands this knowledge by demonstrating that uORFs act as a general mechanism for buffering translational variability on a genome-wide scale across multiple species and developmental stages. Our findings also suggest that uORF-mediated translational buffering is not merely a response to environmental stress but an evolutionarily conserved mechanism integral to maintaining protein homeostasis across species. Organisms can buffer phenotypic variation against environmental or genotypic perturbations through "canalization" [85]. The discovery that uORFs are crucial for *Drosophila* development links them to canalization, providing evidence that uORFs reduce interspecific CDS translation differences and enhance phenotypic stability. While previous studies primarily focused on the stabilization of transcriptional levels [86-90] or protein levels [55,88,91-93], our study opens new avenues for exploring how organisms maintain phenotypic robustness at the translational level through molecular mechanisms like uORF-mediated buffering. This study lays the groundwork for future research into the evolutionary and functional roles of uORFs, particularly in the context of canalization and adaptive evolution.

Previous studies have shown that a significant fraction of fixed uORFs in the populations of *D. melanogaster* and humans were driven by positive Darwinian selection [63,67], suggesting active maintenance through adaptive evolution rather than purely neutral or deleterious processes. While uORFs have traditionally been recognized for their capacity to attenuate translation of downstream CDSs, accumulating evidence now underscores their critical role in stabilizing gene expression under fluctuating cellular and environmental conditions [43,55,56]. Whether the favored evolutionary selection of uORFs acts primarily through their role in translational repression or translational buffering remains a compelling yet unresolved question, as these two functions are inherently linked. Indeed, highly conserved uORFs tend to be translated at higher levels, resulting not only in stronger inhibition of CDS translation [34,45,67], but also in a more pronounced buffering effect, as demonstrated in this study. This buffering capacity of uORFs potentially provides selective advantages by reducing fluctuations in protein synthesis, thus minimizing gene-expression noise and enhancing cellular homeostasis. This suggests that selection may favor uORFs that contribute to translational robustness, a hypothesis supported by findings in yeast and mammals showing that uORFs are significantly enriched in stress-response genes and control the translation of certain master regulators of stress responses [41,42,94,95]. Our study suggests that translational buffering, rather than translational repression alone, can also drive evolutionary selection favoring uORFs, although it remains challenging to empirically disentangle these functions. Future comparative genomic analyses, coupled with experimental approaches such as

512 ribosome profiling and functional mutagenesis, will be crucial in elucidating the precise evolutionary
513 forces driving uORF conservation and adaptation.
514
515 The original ICIER model was devised to elucidate how a single uORF can confer resistance to global
516 translation inhibition on its host gene during stress conditions [43]. It proposed that an elongating ribosome
517 (80S) causes downstream 40S subunits to dissociate from the mRNA after collision. Our revised model
518 expands this scope to include bidirectional dissociation, where an 80S ribosome can release both
519 upstream and downstream 40S subunits, reflecting recent findings that 80S primarily causes upstream
520 40S dissociation during 40S/80S collisions [56]. Mechanistically, increasing the loading rate of 40S
521 ribosomes onto the mRNA (R_{in}) elevates the density of 40S and 80S ribosomes scanning the uORF,
522 thereby raising the likelihood of 40S-80S collisions and subsequent 40S dissociation. This leads to a
523 reduced flow of 40S ribosomes reaching the CDS relative to the initial increase in 40S at the 5' end of
524 the mRNA and uORF. Conversely, a decrease in 40S ribosome loading rate reduces the density of
525 scanning ribosomes, diminishing the probability of 40S-80S collisions and 40S dissociation, and
526 partially restoring the flow of 40S ribosomes to the CDS. Overall, the 40S-80S collision mechanism
527 and subsequent 40S dissociation at the uORF moderate downstream CDS translation less than changes
528 in ribosome loading at the 5' end of the mRNA or uORF. The revised model underscores the analogy of
529 uORFs as "molecular dams", adeptly modulating the stream of ribosomes and cushioning downstream
530 CDS translation against fluctuations (Figure 9A). Additionally, our study broadens its perspective to
531 account for the potential interactive effects of multiple uORFs within a gene, recognizing the prevalent
532 clustering of these regulatory elements and their combinatorial influence on translational control. In
533 essence, our study presents a nuanced and comprehensive expansion of the ICIER model, paving the
534 way for a deeper understanding of uORFs as pivotal elements in the maintenance of protein synthesis
535 stability under variable translational conditions. However, we realize the existence of alternative
536 mechanisms governing uORF function, such as those found in traditional repression models [55] or
537 ribosome queuing and reinitiation strategies [56]. As such, further computational studies are needed to
538 dissect and understand the full spectrum of uORF-mediated translational regulation.
539
540 Ribosome slowdown or stalling on mRNA due to rare codons [56,96-98] or nascent blocking peptides [99-102]
541 frequently triggers ribosome collisions genome-wide [103-105]. Such collisions, especially among
542 elongating 80S ribosomes, often activate ribosome quality control (RQC) pathways that recognize
543 collision interfaces on the 40S subunit, leading to ribosomal subunit dissociation and degradation [106-
544 108]. In mammals, ZNF598 specifically identifies collided ribosomes to initiate ubiquitin-dependent
545 protein and mRNA quality control pathways [109-113]. Analogously, yeast employs Hel2-mediated
546 ubiquitination of uS10, initiating dissociation via the RQC-trigger complex (RQT) [114]. Furthermore, the
547 human RQT (hRQT) complex recognizes ubiquitinated ribosomes and induces subunit dissociation
548 similarly to yeast RQT [115]. However, transient ribosome collisions can evade RQC by promoting

549 resumed elongation through mechanical force provided by trailing ribosomes, thereby mitigating
550 stalling [116]. Beyond 80S collisions, evidence increasingly highlights a distinct collision type involving
551 scanning 40S subunits or pre-initiation (43S) complexes. Recently, an initiation RQC pathway (iRQC)
552 targeting the small ribosomal subunit (40S) has been described, particularly involving collisions
553 between scanning 43S complexes or between stalled 43S and elongating 80S ribosomes (Figure 9B)
554 [117,118]. During iRQC, E3 ubiquitin ligase RNF10 ubiquitinates uS3 and uS5 proteins, resulting in 40S
555 degradation [118]. This mechanism aligns closely with our ICIER model, proposing collision-driven 43S
556 dissociation in the 5' UTRs. Future studies exploring these mechanisms in greater detail will clarify
557 how uORFs modulate translational regulation through buffering effects.
558
559 The *bcd* gene, crucial for *Drosophila* embryogenesis, has been well studied [71,72,119-123], but its uORF
560 remains underexplored. Our study showed that deleting the *bcd* uORF led to abnormal phenotypes and
561 reduced fitness, underscoring the importance of uORF-mediated translation. These effects were more
562 severe under heat stress (29℃) than at normal temperatures (25℃), suggesting uORFs play a vital role
563 in buffering phenotypic variation. We demonstrated that the *bcd* uORF represses CDS translation using
564 sucrose gradient fractionation followed by qPCR—an approach that directly measures translation
565 efficiency while minimizing confounding from RNA/protein degradation. However, detecting Bcd
566 protein levels with antibodies across developmental stages or conditions in the mutants and wild-type
567 controls would provide an even stronger validation of our model and should be explored in future
568 studies. Rescue experiments, typically used to confirm that phenotypic changes are due to specific
569 genetic edits, faced challenges in our study and previous uORF-KO experiments in animals [124,125] and
570 plants [126-129]. We altered the start codon of the *bcd* uORF to minimize effects on the 5' UTR, without
571 impacting the coding sequence, making traditional rescue experiments impractical. To control for
572 genetic background variations, we backcrossed two *bcd* uORF-KO mutant strains with w^{1118} flies for
573 nine generations, using w^{1118} as a control. Both uORF-KO mutants (uKO1/uKO1 and uKO2/uKO2)
574 consistently showed reduced progeny, ruling out background or off-target effects. Crosses between the
575 two mutants produced fewer offspring than WT crosses, confirming that the reduced progeny was due
576 to the uORF deletions, not genetic background. As *bcd* is maternally inherited, we expected defective
577 phenotypes only in offspring from mutant females crossed with wild-type males. This was confirmed:
578 uKO1/uKO1 males crossed with wild-type females had normal offspring, while uKO1/uKO1 females
579 crossed with wild-type males had reduced offspring, similar to mutant crosses. Similar results were
580 observed for uKO2/uKO2 mutants. These findings confirm that disrupting *bcd* uORFs significantly
581 reduces hatchability and fertility, effects not attributable to genetic background. Our results highlight
582 the critical role of *bcd* uORFs in development and fecundity, and provide a basis for future studies on
583 uORF regulation in other genes across species.
584

585 Taken together, this study demonstrates the role of uORF-mediated translational buffering in mitigating
586 variability in gene translation during species divergence and development, using large-scale
587 comparative transcriptome and translatome data. Our work addresses gaps in understanding the
588 stabilization of gene expression at the translational level and offers new insights into the evolutionary
589 and functional significance of uORFs.
590

591 **Materials and Methods**

592 **Modeling the uORF-mediated buffering effect on CDS translation**

593 We adapted the stochastic ICIER framework developed by Andreev *et al.* [43] based on the TASEP model,
594 with several major modifications. First, we concatenated a 500-codon CDS downstream of the uORF.
595 A leaky scanning ribosome from the uORF would initiate translation at the CDS start codon with a
596 probability of I_{CDS}. The probability of elongating ribosomes moving along the CDS (ν_{EC}= 0.5) in each
597 action was higher than that along uORFs (ν_{Eu}= 0.3) in the model settings, considering that uORFs
598 usually encode blocking peptides [99-102,130] or contain stalling codons [56,96,97]. Second, we recorded the
599 number of elongating ribosomes that completed translation at the stop codon of a CDS (N_{EC}) or uORF
600 (N_{Eu}) during a given time period and regarded them as proxies for quantifying the CDS or uORF TE.
601 Third, we considered three models for simulating the consequences of scanning ribosomes colliding
602 with elongating ribosomes mentioned above: a downstream dissociation model, an upstream
603 dissociation model, and a double dissociation model.

604

605 In detail, the mRNA molecule structure in the simulations consisted of five parts: a 5' leader before the
606 uORF (150 nucleotides), the uORF (ranging from 2 to 100 codons), a segment between the uORF and
607 CDS (150 nucleotides), the CDS (500 codons) and the 3' UTR (150 nucleotides). The R_{in} determined
608 the loading rate of the 40S scanning ribosomes on the mRNA 5'-terminus, and it roughly represented
609 the availability of *trans* translational resources in the cell. We generated two sets of R_{in} values (Figure
610 S1A) to simulate the variation in the availability of translational resources (40S scanning ribosomes,
611 etc.): i) 1000 values of R_{in} were generated from a random generator of a uniform distribution, U (0, 0.1);
612 ii) 1000 values were first generated from a random generator of an exponential distribution, E (1), and
613 then each of the 1000 values was divided by 70 to obtain 1000 values of R_{in}. I_{uORF} and I_{CDS} represent
614 the probability of translation at the start codon of a uORF or CDS, respectively. We used different
615 combinations of I_{uORF} and I_{CDS} to test how translational initiation strength influenced the buffering
616 effect of uORFs. ν_s determined the scanning rate of the 40S ribosome, and $\nu_{Eu}$and ν_{EC} determined the
617 elongation rate of the 80S ribosome at a uORF or CDS, respectively. All the parameters used in our
618 simulation are listed in Table S1.

619

620 The mRNA molecule was modeled as an array, and the value of each position in the array represented
621 the occupation status of ribosomes along the mRNA molecule. The process of translation was simulated
622 by a series of discrete actions referred to Andreev *et al.* [43]. Upon each action, there was a probability
623 that certain events would occur, including the addition of a 40S ribosome to the 5'-terminus, the
624 transformation of a 40S ribosome into an 80S ribosome at a CDS or uORF start codon, the movement
625 of a 40S ribosome or 80S ribosome to the next codon in the CDS or uORF, etc. More specific operations
626 in each action were described in Supplementary Materials and Methods.

After 1,000,000 actions, we recorded the number of 80S ribosomes that completed translation at the uORF (N_{Eu}) and CDS (N_{EC}) for each R_{in} input, which was used to represent the protein production rates (i.e. translation rate) of the uORF and CDS, respectively. To test the effects of different factors, we used various combinations of the parameters in Table S1 in the simulations. For a given distribution of R_{in} input (1000 values) and the dissociation model, we obtained the corresponding 1000 N_{EC} values and their CV values by calculating the ratio of the standard deviation to the mean of N_{EC}.

In the double-uORF simulation, we only adopted a uniform distribution of R_{in} input and the downstream dissociation model. Most parameters and simulation processes were similar to those in the single-uORF simulation, except that an additional uORF was introduced upstream of the CDS with an initiation probability of I_{uORF2}.

Annotation of uORFs in *D. melanogaster* and *D. simulans*

We downloaded the whole-genome sequence alignment (maf) of *D. melanogaster* (dm6) and 26 other insect species from the University of California Santa Cruz (UCSC) genome browser (genome.ucsc.edu) [131]. To improve the genome correspondence between *D. melanogaster* and *D. simulans*, we adopted a newly published reference-quality genome of *D. simulans* (NCBI, ASM438218v1) [132] to replace the original *D. simulans* genome in UCSC maf. We softly masked repetitive sequences in the new genome by RepeatMasker 4.1.1 (http://www.repeatmasker.org) and aligned it to dm6 with lastz [133] in runLastzChain.sh following UCSC guidelines. The lastz alignment parameters and the scoring matrix were the same as the parameters of dm6 and droSim1 in UCSC. Chained alignments were processed into nets by the chainNet and netSyntenic programs. The alignment was integrated into the multiple alignments of 27 species with multiz [134].

Based on the genome annotation of *D. melanogaster* (FlyBase r6.04, https://flybase.org/), we used the Galaxy platform to parse the multiple sequence alignments of 5' UTRs in *Drosophila*. The 5' UTR sequences of each annotated transcript from *D. melanogaster* and the corresponding sequences in *D. simulans* were extracted from the maf. The start codons of putative uORFs were identified by scanning all the ATG triplets (uATGs) within the 5' UTRs of *D. melanogaster* and *D. simulans*. uATGs that overlapped with any annotated CDS region were removed. The presence or absence of each uATG of *D. melanogaster* was determined at orthologous sites in *D. simulans* based on multiple genome alignments, and vice versa. The conserved uORF was defined as a uORF where its uATG was present in both *D. melanogaster* and the corresponding orthologous positions of *D. simulans*. For each protein-coding gene, we only considered the canonical transcript. For transcripts containing multiple uORFs,

662 we defined the uORF that showed the highest TE as the dominant uORF. The branch length score (BLS)
663 of the uATGs was calculated as we previously described [18].
664
665 **Fly materials and general raising conditions**
666 The *sim4* strain of *D. simulans* was used to generate all the libraries of *D. simulans* in this study. All
667 flies were raised on standard corn medium and grown in 12 h light: 12 h dark cycles at 25℃ for general
668 conditions or at 29℃ for specific experimental design. The samples of embryos at different stages, larva,
669 pupa, bodies and heads were collected following a previous protocol [23].
670
671 **Processing *Drosophila* mRNA-Seq and Ribo-Seq data**
672 Ribo-Seq and matched mRNA-Seq libraries for different developmental stages and tissues of *D.*
673 *simulans* were constructed as we previously described [23] and sequenced on an Illumina HiSeq-2500
674 sequencer (run type: single-end; read length: 50 nucleotides) according to the manufacturer's protocol.
675 The 3' adaptor sequences (TGGAATTCTCGGGTGCCAAGG) were trimmed using Cutadapt 3.0 with
676 default parameters [135], and the NGS reads were mapped to the genomes of yeast, *Wolbachia*, *Drosophila*
677 viruses and the sequences of tRNAs, ribosomal RNAs (rRNAs), small nuclear RNAs (snRNAs) or small
678 nucleolar RNAs (snoRNAs) of *D. melanogaster* (FlyBase r6.04) and *D. simulans* (FlyBase r1.3) using
679 Bowtie2 version 2.2.3 [136] with the parameters -p8 --local -k1. The mapped reads of these
680 genomes/sequences were further removed in the downstream analysis.
681
682 After filtering, the mRNA-Seq and Ribo-Seq reads were mapped to the reference genomes of *D.*
683 *melanogaster* (FlyBase, r6.04) and *D. simulans* (NCBI, ASM438218v1), respectively, using the Spliced
684 Transcripts Alignment to a Reference (STAR) algorithm [137]. For Ribo-Seq reads, we assigned a mapped
685 RPF (27-34 nucleotides in length) to its P-site using the psite script from Plastid [138]. The uniquely
686 mapped reads were extracted and then mapped to the CDS of *D. melanogaster* and *D. simulans*
687 respectively using STAR [137]. The P-sites of RPF or mRNA-Seq reads that overlapped with a CDS were
688 counted separately. The reads that were not mapped to CDSs were then mapped to the 5' UTRs of *D.*
689 *melanogaster* and *D. simulans*. The P-sites of RPF or mRNA-Seq reads that overlapped with uORFs
690 were counted separately.
691
692 The TE of a given uORF or CDS was calculated as the $RPKM_{P\text{-}site}/RPKM_{mRNA}$ ratio. For a few uORFs
693 with $RPKM_{mRNA} = 0$ but $RPKM_{P\text{-}site} > 0$, a pseudocount of 0.1 was added to both $RPKM_{mRNA}$ and
694 $RPKM_{P\text{-}site}$ to avoid dividing a positive value by zero. In each sample, expressed genes were defined as
695 the genes with a CDS $RPKM_{mRNA} > 0.1$ in both species, and translated uORFs were defined as uORFs
696 with a TE > 0.1.
697

Testing the statistical significance of the difference in the interspecific TE change between a uORF and its downstream CDS

For this analysis, we adopted the methods developed by Zhang *et al.* [23]. Briefly, for the samples of female bodies or male bodies of *D. melanogaster* with biological replicates, we obtained the $\log_2$(TE) and SE of the $\log_2$(TE) of a uORF or CDS by contrasting the RPF counts against mRNA-Seq read counts using DESeq2. Then, we fitted the SE values against the normalized mRNA counts and $\log_2$(TE) values using the gam function in the R package mgcv, with a log link to obtain the SE ~ mRNA counts + $\log_2$TE functions. For other samples without biological replicates, we estimated the SE of the $\log_2$(TE) for a feature (CDS or uORF) by applying the fitted functions obtained based on the biological replicates of female and male bodies to the observed mRNA counts and $\log_2$(TE) values. We identified uORFs whose TEs differed significantly between the paired samples of *D. melanogaster* and *D. simulans* by testing whether the value obtained for $log_2(\beta_u) = log_2(TE_{uORF,sim}) - log_2(TE_{uORF,mel})$ was significantly different from 0. Based on the SE of the $\log_2$(TE) derived as described above, the SE of the $\log_2(\beta_u)$ can be derived as follows:

$$\text{SE_log}_2(\beta_u) = \sqrt{(SE_log_2TE_{uORF,sim})^2 + (SE_log_2TE_{uORF,mel})^2}$$

As the Wald statistic, $\frac{log_2(\beta_u)}{\text{SE_log2}(\beta_u)}$, follows a standard normal distribution under the null hypothesis of $\log_2(\beta_u) = 0$, we calculated the *P* value as follows: $2 \cdot (1 - \phi(|\frac{log_2(\beta_u)}{\text{SE_log2}(\beta_u)}|))$. The identification of CDSs whose TEs differed significantly between the paired samples of *D. melanogaster* and *D. simulans* was similar to the procedures above.

Then, we defined $\gamma = \frac{\beta_c}{\beta_u}$ (where $\beta_c = TE_{CDS,sim}/TE_{CDS,mel}$) and tested whether $\log_2(\gamma)$ was significantly different from 0 to determine whether the magnitude of interspecific TE changes in CDSs and uORFs was significantly different. We obtained $\log_2TE_{uORF,sim}$, $\log_2TE_{uORF,mel}$, $\log_2TE_{CDS,sim}$, and $\log_2TE_{CDS,mel}$ as described above and estimated $SE_\log_2TE_{uORF,sim}$, $SE_\log_2TE_{uORF,mel}$, $SE_\log_2TE_{CDS,sim}$, and $SE_\log_2TE_{CDS,mel}$ based on the biological replicates of female and male bodies of *D. melanogaster*. Finally, $\log_2(\gamma)$ also follows a normal distribution with SE denoted as $SE_\log_2(\gamma)$ =

$$\sqrt{(SE_log_2TE_{uORF,sim})^2 + (SE_log_2TE_{uORF,mel})^2 + (SE_log_2TE_{CDS,sim})^2 + (SE_log_2TE_{CDS,mel})^2}$$

As the Wald statistic, $\frac{log_2(\gamma)}{\text{SE_log2}(\gamma)}$, follows a standard normal distribution under the null hypothesis that $\log_2(\beta_u) = 0$, we calculated the *P* value as follows: $2 \cdot (1 - \phi(|\frac{log_2(\gamma)}{\text{SE_log2}(\gamma)}|))$.

uORFs are shorter than CDSs and have substantially lower read counts, reducing statistical power when comparing TE changes. To mitigate this bias, we focused on highly expressed genes, defined as those

730 with both mRNA and RPF RPKM values above the 50th percentile in *D. melanogaster* and *D. simulans*.
731 uORFs and CDSs of these genes were used in the calculation of significant TE change.
732
733 **Knocking out a uORF with CRISPR-Cas9 technology in *D. melanogaster***
734 We searched for possible sgRNA target sites near the uATG start codon of a uORF using the Benchling
735 website (https://www.benchling.com/crispr/) to design optimal single guide RNA (sgRNA) sequences
736 with high specificity and low off-target effects. We then synthesized single-stranded complementary
737 DNAs (ssDNAs) and annealed them to obtain double-stranded DNA (dsDNA), which served as the
738 template for sgRNA expression. The template sequences of the sgRNA used for *bcd* uATG-KO are
739 listed in Table S6. The dsDNA was then ligated into the BbsI-digested pU6B vector. The pU6B-sgRNA
740 plasmid was purified and injected into the embryos of transgenic Cas9 flies collected within one hour
741 of laying at the Tsinghua Fly Center as described in Ni *et al.* [139]. The injected embryos were kept at
742 25°C and 60% humidity until adulthood (G0). The G0 adult flies that hatched from injected embryos
743 were individually crossed with other flies (*y sc v*) to increase the number of offspring. Then, the F1
744 progeny were crossed with flies carrying an appropriate balancer (*Dr, e/TM3, Sb*). After F2 spawning,
745 the F1 individuals were screened for mutations of interest by genotyping. The primers used for
746 genotyping are listed in Table S6. The F2 progeny whose parents showed positive genotyping results
747 were then screened for the *yellow*$^{-}$ gene to separate the chromosome carrying *nos-Cas9*. The screened
748 F2 males were crossed with the flies containing the same balancer as above. After F2 genotyping, the
749 progeny (F3) of positive F2 individuals showing the same mutation status were crossed individually to
750 generate homozygous mutants in the F4 generation. The original homozygous mutants were
751 sequentially outcrossed with w^{1118} flies for 9 generations to purify the genetic background (Figure S25B).
752
753 **Ribosome fraction analysis by sucrose gradient fractionation and RT–qPCR**
754 The 0-2 h embryos obtained from w^{1118}, uKO1/uKO1, and uKO2/uKO2 mutants raised at 25°C and
755 29°C were collected and homogenized in a Dounce homogenizer with lysis buffer [50 mM Tris pH 7.5,
756 150 mM NaCl, 5 mM $MgCl_2$, 1% Triton X-100, 2 mM dithiothreitol (DTT), 20 U/ml SuperaseIn
757 (Ambion), 0.5 tablets of proteinase inhibitor (Roche), 100 μg/ml emetine (Sigma Aldrich), and 50 μM
758 guanosine 5′-[β,γ-imido]triphosphate trisodium salt hydrate (GMP-PNP) (Sigma Aldrich)] at 4°C. The
759 lysates were clarified by centrifugation at 4°C and 20,000×g for 8 min, and the supernatants were
760 transferred to new 1.5 ml tubes. 10-45% sucrose gradients were prepared in buffer (250 mM NaCl, 50
761 mM Tris pH 7.5, 15 mM $MgCl_2$, 0.5 mM DTT, 12 U/ml RNaseOUT, 0.5 tablets of protease inhibitor,
762 and 20 μg/ml emetine) using a Gradient Master (Biocomp Instruments) in ULTRA-CLEAR Thinwall
763 Tubes (Beckman Coulter). A sample volume of up to 500 μl was applied to the top of each gradient.
764 After ultracentrifugation with a Hitachi P40ST rotor at 35,000 × rpm for 3 h at 4°C, the monosome and
765 polysome fractions were collected, flash-frozen in liquid nitrogen, and stored at -80°C until further use.
766

767 The RNA in the monosome and polysome fractions was extracted separately using TRIzol reagent (Life
768 Technologies, Inc.) and chloroform (Beijing Chemical Works) following the manufacturer's
769 instructions and were reverse transcribed into cDNA using the PrimeScript™ II 1st Strand cDNA
770 Synthesis Kit (Takara). RT–qPCR analysis of *bcd* cDNA and its targets was performed using
771 PowerUp™ SYBR™ Green Master Mix (Thermo Fisher) following the manufacturer's instructions
772 with *rp49* as an internal control. The primer sequences employed for RT-qPCR are listed in Table S6.
773 For each sample, the ratio of *bcd* mRNA abundance in the polysome fraction to that in the monosome
774 fraction was calculated as the P-to-M ratio. Six biological replicates were performed for each sample.
775
776 **Dual-luciferase reporter assays**
777 The wild-type (WT) 5' UTR of *bcd* was cloned from cDNA by PCR, and uATG mutations were
778 introduced into 5' UTR using specific amplification primers (listed in Table S6). The WT and mutated
779 5' UTR sequences were ligated into a linearized reporter plasmid (psiCHECK-2 vector, Promega). The
780 whole sequence of all the plasmids was validated by Sanger sequencing.
781
782 *Drosophila* S2 cells were cultured in Schneider's Insect Medium (Sigma) plus 10% (by volume) heat-
783 inactivated fetal bovine serum, 100 U/ml penicillin and 100 μg/ml streptomycin (Thermo Fisher) at 25℃
784 without CO_2 for 24 h to reach $1–2\times10^6$cells/ml before further treatments. Plasmid transfection was
785 conducted with Lipofectamine 3000 (L3000001, Thermo Fisher) according to the supplier's protocol.
786 The *Renilla* luciferase activity associated with WT or uORF-mutated 5' UTRs was measured according
787 to the manual of the Dual-Luciferase Reporter Assay System (Promega) 32 h after transfection and was
788 normalized to the activity of firefly luciferase.
789
790 **mRNA-Seq in the embryos of mutant and WT flies**
791 We collected 0-2 h, 2-6 h, 6-12 h, and 12-24 h embryos of w^{1118} and uKO2/uKO2 mutants raised at 25°C
792 and 29°C and conducted mRNA-Seq. Library construction and sequencing with PE150 were conducted
793 by Annoroad on the Illumina Nova6000 platform. Two biological replicates were sequenced for each
794 sample. The clean data were mapped to the reference genome of *D. melanogaster* (FlyBase, r6.04) using
795 STAR [137]. Reads mapped to the exons of each gene were tabulated with htseq-count [140]. The
796 differentially expressed genes were identified using DESeq2 [141]. The Gene Ontology (GO) analyses
797 were conducted using the "clusterProfiler" package in R [142].
798
799 **Measurement of embryo hatchability**
800 We collected embryos from WT and mutant flies and manually seeded them in vials containing standard
801 corn medium at a density of 30 embryos per vial and 20 vials per strain. The embryos were cultivated
802 under standard conditions (60% humidity, 12 h light: 12 h dark cycles at 25°C). We counted the number

of pupae in each vial after the completion of pupation and calculated the corresponding hatching rate = $number_{pupa}$ / 30

Quantification of offspring number per female fly

Newly hatched virgins were picked out and allowed to mature for two days in separate vials. They were then mated by placing one virgin female with three male flies for two days. After that, each female parent was transferred to a new vial to count the offspring number produced in each 10-day period at 25°C and 29°C. All assays were performed with 20 females per genotype.

Measurement of starvation resistance in adult flies

We selected 3- to 5-day-old adult males and females and placed them in the starvation medium (1.5% agar), with 10 flies per vial and 10 vials for both males and females from each strain. We made observations every 6 h or 12 h to count the number of deaths under starvation conditions until all flies had starved to death. The survival curves were plotted by the ggsurvplot package in R.

mRNA-Seq and Ribo-Seq data analysis in primates

The mRNA-Seq and Ribo-Seq data from the brains, livers, and testes of humans and macaques were downloaded from reference [79] with accession number E-MTAB-7247 (ArrayExpress). The mRNA-Seq and Ribo-Seq data of human lymphoblastoid cell lines (LCL) from Yoruba individuals were downloaded from references [81,82] under accession numbers GSE61742 and E-GEUV-1. The matched mRNA-Seq and Ribo-Seq libraries of 69 individuals were used.

The uORF annotation and downstream analysis procedures for the human and macaque data were similar to those applied in *Drosophila* as described above. The differential analysis of translational efficiency in humans and macaques was conducted by Xtail [143]. In each pair of human-macaque samples, expressed genes were defined as the genes with a CDS $RPKM_{mRNA} > 0.1$ in both species. The translated uORFs in a sample were defined as uORFs with a TE > 0.1. For the human cell line data, expressed genes were defined as genes with a mean CDS $RPKM_{mRNA} > 0.1$ across the cell lines, and translated uORFs were defined as uORFs with a mean TE > 0.1

Acknowledgments

We thank Drs. Wei Xie, Weiwei Zhai, and Xionglei He for constructive comments and suggestions. This work was supported by grants from the Ministry of Science and Technology of the People's Republic of China (2022YFE0132000), the Yunnan Provincial Science and Technology Project at Southwest United Graduate School (202302A0370006), the National Natural Science Foundation of China (32070597) and the Natural Science Foundation of Beijing (5212006). We thank the National

Center for Protein Sciences at Peking University for their technical assistance. Some of the analyses were performed on the High-Performance Computing Platform of the Center for Life Sciences.

Declaration of interests

The authors declare no competing interests.

Data Availability Statement

All deep-sequencing data generated in this study, including single-ended mRNA-Seq and Ribo-Seq data of 10 developmental stages and tissues of *Drosophila simulans* and paired-end mRNA-Seq data of 0–2 h, 2–6 h, 6–12 h, and 12–24 h *Drosophila melanogaster* embryos, were deposited in the China National Genomics Data Center Genome Sequence Archive (GSA) under accession numbers CRA003198, CRA007425, and CRA007426. The mRNA-Seq and Ribo-Seq data for the different developmental stages and tissues of *Drosophila melanogaster* were published in our previous paper [23] and were deposited in the Sequence Read Archive (SRA) under accession number SRP067542. All original code has been deposited on GitHub: https://github.com/lujlab/uORF_buffer; https://github.com/lujlab/Buffer_eLife2025.

857 **References**

858 1. Buccitelli, C., and Selbach, M. (2020). mRNAs, proteins and the emerging principles of gene
859 expression control. Nature Reviews Genetics *21*, 630-644. 10.1038/s41576-020-0258-4.
860 2. Lee, T.I., and Young, R.A. (2013). Transcriptional regulation and its misregulation in disease.
861 Cell *152*, 1237-1251. 10.1016/j.cell.2013.02.014.
862 3. MacNeil, L.T., and Walhout, A.J.M. (2011). Gene regulatory networks and the role of
863 robustness and stochasticity in the control of gene expression. Genome Research *21*, 645-657.
864 10.1101/gr.097378.109.
865 4. Hill, M.S., Vande Zande, P., and Wittkopp, P.J. (2021). Molecular and evolutionary processes
866 generating variation in gene expression. Nat Rev Genet *22*, 203-215. 10.1038/s41576-020-
867 00304-w.
868 5. Vogel, C. (2013). Protein Expression Under Pressure. Science *342*, 1052-1053.
869 10.1126/science.1247833.
870 6. Signor, S.A., and Nuzhdin, S.V. (2018). The Evolution of Gene Expression in cis and trans.
871 Trends Genet *34*, 532-544. 10.1016/j.tig.2018.03.007.
872 7. Artieri, C.G., and Fraser, H.B. (2014). Evolution at two levels of gene expression in yeast.
873 Genome Res *24*, 411-421. 10.1101/gr.165522.113.
874 8. McManus, C.J., May, G.E., Spealman, P., and Shteyman, A. (2014). Ribosome profiling reveals
875 post-transcriptional buffering of divergent gene expression in yeast. Genome Res *24*, 422-430.
876 10.1101/gr.164996.113.
877 9. Wang, Z., Sun, X., Zhao, Y., Guo, X., Jiang, H., Li, H., and Gu, Z. (2015). Evolution of gene
878 regulation during transcription and translation. Genome Biol Evol *7*, 1155-1167.
879 10.1093/gbe/evv059.
880 10. Khan, Z., Ford, M.J., Cusanovich, D.A., Mitrano, A., Pritchard, J.K., and Gilad, Y. (2013).
881 Primate transcript and protein expression levels evolve under compensatory selection pressures.
882 Science *342*, 1100-1104. 10.1126/science.1242379.
883 11. Wang, S.H., Hsiao, C.J., Khan, Z., and Pritchard, J.K. (2018). Post-translational buffering leads
884 to convergent protein expression levels between primates. Genome Biol *19*, 83.
885 10.1186/s13059-018-1451-z.
886 12. Schrimpf, S.P., Weiss, M., Reiter, L., Ahrens, C.H., Jovanovic, M., Malmström, J., Brunner, E.,
887 Mohanty, S., Lercher, M.J., Hunziker, P.E., et al. (2009). Comparative functional analysis of
888 the *Caenorhabditis elegans* and *Drosophila melanogaster* proteomes. PLoS Biol *7*, e48.
889 10.1371/journal.pbio.1000048.
890 13. Kusnadi, E.P., Timpone, C., Topisirovic, I., Larsson, O., and Furic, L. (2022). Regulation of
891 gene expression via translational buffering. Biochim Biophys Acta Mol Cell Res *1869*, 119140.
892 10.1016/j.bbamcr.2021.119140.
893 14. Laurent, J.M., Vogel, C., Kwon, T., Craig, S.A., Boutz, D.R., Huse, H.K., Nozue, K., Walia, H.,
894 Whiteley, M., Ronald, P.C., and Marcotte, E.M. (2010). Protein abundances are more conserved
895 than mRNA abundances across diverse taxa. Proteomics *10*, 4209-4212.
896 10.1002/pmic.201000327.
897 15. Teixeira, F.K., and Lehmann, R. (2019). Translational Control during Developmental
898 Transitions. Cold Spring Harb Perspect Biol *11*. 10.1101/cshperspect.a032987.
899 16. Jackson, R.J., Hellen, C.U., and Pestova, T.V. (2010). The mechanism of eukaryotic translation
900 initiation and principles of its regulation. Nat Rev Mol Cell Biol *11*, 113-127. 10.1038/nrm2838.
901 17. Sonenberg, N., and Hinnebusch, A.G. (2009). Regulation of translation initiation in eukaryotes:
902 mechanisms and biological targets. Cell *136*, 731-745. 10.1016/j.cell.2009.01.042.
903 18. Zhang, H., Wang, Y., Wu, X., Tang, X., Wu, C., and Lu, J. (2021). Determinants of genome-
904 wide distribution and evolution of uORFs in eukaryotes. Nature Communications *12*, 1076.
905 10.1038/s41467-021-21394-y.
906 19. Zhang, H., Wang, Y., and Lu, J. (2019). Function and Evolution of Upstream ORFs in
907 Eukaryotes. Trends Biochem Sci *44*, 782-794. 10.1016/j.tibs.2019.03.002.
908 20. Neafsey, D.E., and Galagan, J.E. (2007). Dual modes of natural selection on upstream open
909 reading frames. Mol Biol Evol *24*, 1744-1751. 10.1093/molbev/msm093.

21. Churbanov, A., Rogozin, I.B., Babenko, V.N., Ali, H., and Koonin, E.V. (2005). Evolutionary conservation suggests a regulatory function of AUG triplets in 5'-UTRs of eukaryotic genes. Nucleic Acids Res *33*, 5512-5520. 10.1093/nar/gki847.
22. Resch, A.M., Ogurtsov, A.Y., Rogozin, I.B., Shabalina, S.A., and Koonin, E.V. (2009). Evolution of alternative and constitutive regions of mammalian 5'UTRs. BMC Genomics *10*, 162. 10.1186/1471-2164-10-162.
23. Zhang, H., Dou, S., He, F., Luo, J., Wei, L., and Lu, J. (2018). Genome-wide maps of ribosomal occupancy provide insights into adaptive evolution and regulatory roles of uORFs during *Drosophila* development. PLoS biology *16*, e2003903.
24. Malzer, E., Szajewska-Skuta, M., Dalton, L.E., Thomas, S.E., Hu, N., Skaer, H., Lomas, D.A., Crowther, D.C., and Marciniak, S.J. (2013). Coordinate regulation of eIF2alpha phosphorylation by PPP1R15 and GCN2 is required during *Drosophila* development. J Cell Sci *126*, 1406-1415. 10.1242/jcs.117614.
25. Komonyi, O., Papai, G., Enunlu, I., Muratoglu, S., Pankotai, T., Kopitova, D., Maroy, P., Udvardy, A., and Boros, I. (2005). DTL, the *Drosophila* homolog of PIMT/Tgs1 nuclear receptor coactivator-interacting protein/RNA methyltransferase, has an essential role in development. J Biol Chem *280*, 12397-12404. 10.1074/jbc.M409251200.
26. Medenbach, J., Seiler, M., and Hentze, M.W. (2011). Translational control via protein-regulated upstream open reading frames. Cell *145*, 902-913. 10.1016/j.cell.2011.05.005.
27. Chen, J., Tresenrider, A., Chia, M., McSwiggen, D.T., Spedale, G., Jorgensen, V., Liao, H., van Werven, F.J., and Unal, E. (2017). Kinetochore inactivation by expression of a repressive mRNA. Elife *6*. 10.7554/eLife.27417.
28. Cheng, Z., Otto, G.M., Powers, E.N., Keskin, A., Mertins, P., Carr, S.A., Jovanovic, M., and Brar, G.A. (2018). Pervasive, Coordinated Protein-Level Changes Driven by Transcript Isoform Switching during Meiosis. Cell *172*, 910-923 e916. 10.1016/j.cell.2018.01.035.
29. Kurihara, Y., Makita, Y., Kawashima, M., Fujita, T., Iwasaki, S., and Matsui, M. (2018). Transcripts from downstream alternative transcription start sites evade uORF-mediated inhibition of gene expression in Arabidopsis. Proc Natl Acad Sci U S A *115*, 7831-7836. 10.1073/pnas.1804971115.
30. Yang, Y.F., Zhang, X., Ma, X., Zhao, T., Sun, Q., Huan, Q., Wu, S., Du, Z., and Qian, W. (2017). Trans-splicing enhances translational efficiency in C. elegans. Genome Res *27*, 1525-1535. 10.1101/gr.202150.115.
31. Wiestner, A., Schlemper, R.J., van der Maas, A.P., and Skoda, R.C. (1998). An activating splice donor mutation in the thrombopoietin gene causes hereditary thrombocythaemia. Nat Genet *18*, 49-52. 10.1038/ng0198-49.
32. Liu, L., Dilworth, D., Gao, L., Monzon, J., Summers, A., Lassam, N., and Hogg, D. (1999). Mutation of the CDKN2A 5' UTR creates an aberrant initiation codon and predisposes to melanoma. Nat Genet *21*, 128-132. 10.1038/5082.
33. Wen, Y., Liu, Y., Xu, Y., Zhao, Y., Hua, R., Wang, K., Sun, M., Li, Y., Yang, S., Zhang, X.J., et al. (2009). Loss-of-function mutations of an inhibitory upstream ORF in the human hairless transcript cause Marie Unna hereditary hypotrichosis. Nat Genet *41*, 228-233. 10.1038/ng.276.
34. Calvo, S.E., Pagliarini, D.J., and Mootha, V.K. (2009). Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. Proc Natl Acad Sci U S A *106*, 7507-7512. 10.1073/pnas.0810916106.
35. Lee, D.S.M., Park, J., Kromer, A., Baras, A., Rader, D.J., Ritchie, M.D., Ghanem, L.R., and Barash, Y. (2021). Disrupting upstream translation in mRNAs is associated with human disease. Nat Commun *12*, 1515. 10.1038/s41467-021-21812-1.
36. Costa-Mattioli, M., and Walter, P. (2020). The integrated stress response: From mechanism to disease. Science *368*. 10.1126/science.aat5314.
37. Hinnebusch, A.G. (2005). Translational regulation of GCN4 and the general amino acid control of yeast. Annu Rev Microbiol *59*, 407-450. 10.1146/annurev.micro.59.031805.133833.
38. Lu, P.D., Harding, H.P., and Ron, D. (2004). Translation reinitiation at alternative open reading frames regulates gene expression in an integrated stress response. J Cell Biol *167*, 27-33. 10.1083/jcb.200408003.

39. Bohlen, J., Harbrecht, L., Blanco, S., Clemm von Hohenberg, K., Fenzl, K., Kramer, G., Bukau, B., and Teleman, A.A. (2020). DENR promotes translation reinitiation via ribosome recycling to drive expression of oncogenes including ATF4. Nat Commun *11*, 4676. 10.1038/s41467-020-18452-2.
40. Vasudevan, D., Neuman, S.D., Yang, A., Lough, L., Brown, B., Bashirullah, A., Cardozo, T., and Ryoo, H.D. (2020). Translational induction of ATF4 during integrated stress response requires noncanonical initiation factors eIF2D and DENR. Nat Commun *11*, 4677. 10.1038/s41467-020-18453-1.
41. Vattem, K.M., and Wek, R.C. (2004). Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. Proc Natl Acad Sci U S A *101*, 11269-11274. 10.1073/pnas.0400541101.
42. Andreev, D.E., O'Connor, P.B., Fahey, C., Kenny, E.M., Terenin, I.M., Dmitriev, S.E., Cormican, P., Morris, D.W., Shatsky, I.N., and Baranov, P.V. (2015). Translation of 5' leaders is pervasive in genes resistant to eIF2 repression. Elife *4*, e03971. 10.7554/eLife.03971.
43. Andreev, D.E., Arnold, M., Kiniry, S.J., Loughran, G., Michel, A.M., Rachinskii, D., and Baranov, P.V. (2018). TASEP modelling provides a parsimonious explanation for the ability of a single uORF to derepress translation during the integrated stress response. Elife *7*. 10.7554/eLife.32563.
44. Kozak, M. (1989). The scanning model for translation: an update. J Cell Biol *108*, 229-241. 10.1083/jcb.108.2.229.
45. Johnstone, T.G., Bazzini, A.A., and Giraldez, A.J. (2016). Upstream ORFs are prevalent translational repressors in vertebrates. EMBO J *35*, 706-723. 10.15252/embj.201592759.
46. Hinnebusch, A.G., Ivanov, I.P., and Sonenberg, N. (2016). Translational control by 5'-untranslated regions of eukaryotic mRNAs. Science *352*, 1413-1416. 10.1126/science.aad9868.
47. Morris, D.R., and Geballe, A.P. (2000). Upstream open reading frames as regulators of mRNA translation. Mol Cell Biol *20*, 8635-8642. 10.1128/MCB.20.23.8635-8642.2000.
48. Young, S.K., and Wek, R.C. (2016). Upstream Open Reading Frames Differentially Regulate Gene-specific Translation in the Integrated Stress Response. J Biol Chem *291*, 16927-16935. 10.1074/jbc.R116.733899.
49. Dever, T.E., Feng, L., Wek, R.C., Cigan, A.M., Donahue, T.F., and Hinnebusch, A.G. (1992). Phosphorylation of initiation factor 2 alpha by protein kinase GCN2 mediates gene-specific translational control of GCN4 in yeast. Cell *68*, 585-596. 10.1016/0092-8674(92)90193-g.
50. Palam, L.R., Baird, T.D., and Wek, R.C. (2011). Phosphorylation of eIF2 facilitates ribosomal bypass of an inhibitory upstream ORF to enhance CHOP translation. J Biol Chem *286*, 10939-10949. 10.1074/jbc.M110.216093.
51. Baird, T.D., Palam, L.R., Fusakio, M.E., Willy, J.A., Davis, C.M., McClintick, J.N., Anthony, T.G., and Wek, R.C. (2014). Selective mRNA translation during eIF2 phosphorylation induces expression of IBTKalpha. Mol Biol Cell *25*, 1686-1697. 10.1091/mbc.E14-02-0704.
52. Chen, Y.J., Tan, B.C., Cheng, Y.Y., Chen, J.S., and Lee, S.C. (2010). Differential regulation of CHOP translation by phosphorylated eIF4E under stress conditions. Nucleic Acids Res *38*, 764-777. 10.1093/nar/gkp1034.
53. Fraser, H.B., Hirsh, A.E., Giaever, G., Kumm, J., and Eisen, M.B. (2004). Noise minimization in eukaryotic gene expression. PLoS Biol *2*, e137. 10.1371/journal.pbio.0020137.
54. Thattai, M., and van Oudenaarden, A. (2001). Intrinsic noise in gene regulatory networks. Proc Natl Acad Sci U S A *98*, 8614-8619. 10.1073/pnas.151588598.
55. Wu, H.W., Fajiculay, E., Wu, J.F., Yan, C.S., Hsu, C.P., and Wu, S.H. (2022). Noise reduction by upstream open reading frames. Nat Plants *8*, 474-480. 10.1038/s41477-022-01136-8.
56. Bottorff, T.A., Park, H., Geballe, A.P., and Subramaniam, A.R. (2022). Translational buffering by ribosome stalling in upstream open reading frames. PLoS Genet *18*, e1010460. 10.1371/journal.pgen.1010460.
57. Ciandrini, L., Stansfield, I., and Romano, M.C. (2010). Role of the particle's stepping cycle in an asymmetric exclusion process: a model of mRNA translation. Phys Rev E Stat Nonlin Soft Matter Phys *81*, 051904. 10.1103/PhysRevE.81.051904.

58. Reuveni, S., Meilijson, I., Kupiec, M., Ruppin, E., and Tuller, T. (2011). Genome-scale analysis of translation elongation with a ribosome flow model. PLoS Comput Biol *7*, e1002127. 10.1371/journal.pcbi.1002127.
59. von der Haar, T. (2012). Mathematical and Computational Modelling of Ribosomal Movement and Protein Synthesis: an overview. Comput Struct Biotechnol J *1*, e201204002. 10.5936/csbj.201204002.
60. Zhao, Y.B., and Krishnan, J. (2014). mRNA translation and protein synthesis: an analysis of different modelling methodologies and a new PBN based approach. BMC Syst Biol *8*, 25. 10.1186/1752-0509-8-25.
61. Tamura, K., Subramanian, S., and Kumar, S. (2004). Temporal Patterns of Fruit Fly (*Drosophila*) Evolution Revealed by Mutation Clocks. Molecular Biology and Evolution *21*, 36-44. 10.1093/molbev/msg236.
62. Ingolia, N.T., Ghaemmaghami, S., Newman, J.R., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218-223. 10.1126/science.1168978.
63. Zhang, H., Dou, S., He, F., Luo, J., Wei, L., and Lu, J. (2018). Genome-wide maps of ribosomal occupancy provide insights into adaptive evolution and regulatory roles of uORFs during *Drosophila* development. PLoS Biol *16*, e2003903. 10.1371/journal.pbio.2003903.
64. Chothani, S.P., Adami, E., Widjaja, A.A., Langley, S.R., Viswanathan, S., Pua, C.J., Zhihao, N.T., Harmston, N., D'Agostino, G., Whiffin, N., et al. (2022). A high-resolution map of human RNA translation. Mol Cell *82*, 2885-2899 e2888. 10.1016/j.molcel.2022.06.023.
65. Ingolia, N.T. (2016). Ribosome Footprint Profiling of Translation throughout the Genome. Cell *165*, 22-33. 10.1016/j.cell.2016.02.066.
66. Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell *147*, 789-802. 10.1016/j.cell.2011.10.002.
67. Zhang, H., Wang, Y., Wu, X., Tang, X., Wu, C., and Lu, J. (2021). Determinants of genome-wide distribution and evolution of uORFs in eukaryotes. Nat Commun *12*, 1076. 10.1038/s41467-021-21394-y.
68. Kronja, I., Yuan, B., Eichhorn, S.W., Dzeyk, K., Krijgsveld, J., Bartel, D.P., and Orr-Weaver, T.L. (2014). Widespread changes in the posttranscriptional landscape at the *Drosophila* oocyte-to-embryo transition. Cell Rep *7*, 1495-1508. 10.1016/j.celrep.2014.05.002.
69. Qin, X., Ahn, S., Speed, T.P., and Rubin, G.M. (2007). Global analyses of mRNA translational control during early *Drosophila* embryogenesis. Genome Biol *8*, R63. 10.1186/gb-2007-8-4-r63.
70. Berleth, T., Burri, M., Thoma, G., Bopp, D., Richstein, S., Frigerio, G., Noll, M., and Nusslein-Volhard, C. (1988). The role of localization of bicoid RNA in organizing the anterior pattern of the *Drosophila* embryo. EMBO J *7*, 1749-1756.
71. Driever, W., and Nusslein-Volhard, C. (1988). A gradient of bicoid protein in *Drosophila* embryos. Cell *54*, 83-93. 10.1016/0092-8674(88)90182-1.
72. Driever, W., and Nusslein-Volhard, C. (1988). The bicoid protein determines position in the *Drosophila* embryo in a concentration-dependent manner. Cell *54*, 95-104. 10.1016/0092-8674(88)90183-3.
73. Abdelmohsen, K., Panda, A.C., Kang, M.J., Guo, R., Kim, J., Grammatikakis, I., Yoon, J.H., Dudekula, D.B., Noh, J.H., Yang, X., et al. (2014). 7SL RNA represses p53 translation by competing with HuR. Nucleic Acids Res *42*, 10099-10111. 10.1093/nar/gku686.
74. Panda, A.C., Abdelmohsen, K., Martindale, J.L., Di Germanio, C., Yang, X., Grammatikakis, I., Noh, J.H., Zhang, Y., Lehrmann, E., Dudekula, D.B., et al. (2016). Novel RNA-binding activity of MYF5 enhances Ccnd1/Cyclin D1 mRNA translation during myogenesis. Nucleic Acids Res *44*, 2393-2408. 10.1093/nar/gkw023.
75. Panda, A.C., Martindale, J.L., and Gorospe, M. (2017). Polysome Fractionation to Analyze mRNA Distribution Profiles. Bio Protoc *7*. 10.21769/BioProtoc.2126.
76. Amourda, C., Chong, J., and Saunders, T.E. (2018). MicroRNAs buffer genetic variation at specific temperatures during embryonic development. bioRxiv, 444810. 10.1101/444810.

77. Lu, G., Zhao, Y., Chen, Q., Lin, P., Tang, T., Tang, Z., Liufu, Z., and Wu, C.-I. (2021). When development is constantly but weakly perturbed - Canalization by microRNAs. bioRxiv, 2021.2009.2004.458966. 10.1101/2021.09.04.458966.
78. Li, X.Y., MacArthur, S., Bourgon, R., Nix, D., Pollard, D.A., Iyer, V.N., Hechmer, A., Simirenko, L., Stapleton, M., Luengo Hendriks, C.L., et al. (2008). Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm. PLoS Biol *6*, e27. 10.1371/journal.pbio.0060027.
79. Wang, Z.Y., Leushkin, E., Liechti, A., Ovchinnikova, S., Mößinger, K., Brüning, T., Rummel, C., Grützner, F., Cardoso-Moreira, M., Janich, P., et al. (2020). Transcriptome and translatome co-evolution in mammals. Nature *588*, 642-647. 10.1038/s41586-020-2899-z.
80. Hedges, S.B., Dudley, J., and Kumar, S. (2006). TimeTree: a public knowledge-base of divergence times among organisms. Bioinformatics *22*, 2971-2972. 10.1093/bioinformatics/btl505.
81. Battle, A., Khan, Z., Wang, S.H., Mitrano, A., Ford, M.J., Pritchard, J.K., and Gilad, Y. (2015). Genomic variation. Impact of regulatory variation from RNA to protein. Science *347*, 664-667. 10.1126/science.1260793.
82. Lappalainen, T., Sammeth, M., Friedlander, M.R., t Hoen, P.A., Monlong, J., Rivas, M.A., Gonzalez-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P.G., et al. (2013). Transcriptome and genome sequencing uncovers functional variation in humans. Nature *501*, 506-511. 10.1038/nature12531.
83. Stein, K.C., and Frydman, J. (2019). The stop-and-go traffic regulating protein biogenesis: How translation kinetics controls proteostasis. Journal of Biological Chemistry *294*, 2076-2084.
84. Sherman, M.Y., and Qian, S.B. (2013). Less is more: improving proteostasis by translation slow down. Trends Biochem Sci *38*, 585-591. 10.1016/j.tibs.2013.09.003.
85. Waddington, C.H. (1942). CANALIZATION OF DEVELOPMENT AND THE INHERITANCE OF ACQUIRED CHARACTERS. Nature *150*, 563-565. 10.1038/150563a0.
86. Payne, J.L., and Wagner, A. (2015). Mechanisms of mutational robustness in transcriptional regulation. Front Genet *6*, 322. 10.3389/fgene.2015.00322.
87. Denby, C.M., Im, J.H., Yu, R.C., Pesce, C.G., and Brem, R.B. (2012). Negative feedback confers mutational robustness in yeast transcription factor regulation. Proc Natl Acad Sci U S A *109*, 3874-3878. 10.1073/pnas.1116360109.
88. Jarosz, D.F., Taipale, M., and Lindquist, S. (2010). Protein homeostasis and the phenotypic manifestation of genetic diversity: principles and mechanisms. Annu Rev Genet *44*, 189-216. 10.1146/annurev.genet.40.110405.090412.
89. Ebert, M.S., and Sharp, P.A. (2012). Roles for microRNAs in conferring robustness to biological processes. Cell *149*, 515-524. 10.1016/j.cell.2012.04.005.
90. Lu, G.A., Zhang, J., Zhao, Y., Chen, Q., Lin, P., Tang, T., Tang, Z., Wen, H., Liufu, Z., and Wu, C.I. (2023). Canalization of Phenotypes-When the Transcriptome is Constantly but Weakly Perturbed. Mol Biol Evol *40*. 10.1093/molbev/msad005.
91. Alon, U. (2007). Network motifs: theory and experimental approaches. Nat Rev Genet *8*, 450-461. 10.1038/nrg2102.
92. Somogyvari, M., Khatatneh, S., and Soti, C. (2022). Hsp90: From Cellular to Organismal Proteostasis. Cells *11*. 10.3390/cells11162479.
93. Zabinsky, R.A., Mason, G.A., Queitsch, C., and Jarosz, D.F. (2019). It's not magic - Hsp90 and its effects on genetic and epigenetic variation. Semin Cell Dev Biol *88*, 21-35. 10.1016/j.semcdb.2018.05.015.
94. Lawless, C., Pearson, R.D., Selley, J.N., Smirnova, J.B., Grant, C.M., Ashe, M.P., Pavitt, G.D., and Hubbard, S.J. (2009). Upstream sequence elements direct post-transcriptional regulation of gene expression under stress conditions in yeast. BMC Genomics *10*, 7. 10.1186/1471-2164-10-7.
95. Mueller, P.P., and Hinnebusch, A.G. (1986). Multiple upstream AUG codons mediate translational control of GCN4. Cell *45*, 201-207. 10.1016/0092-8674(86)90384-3.
96. Meijer, H.A., and Thomas, A.A. (2003). Ribosomes stalling on uORF1 in the Xenopus Cx41 5' UTR inhibit downstream translation initiation. Nucleic Acids Res *31*, 3174-3184. 10.1093/nar/gkg429.

97. Ivanov, I.P., Shin, B.S., Loughran, G., Tzani, I., Young-Baird, S.K., Cao, C., Atkins, J.F., and Dever, T.E. (2018). Polyamine Control of Translation Elongation Regulates Start Site Selection on Antizyme Inhibitor mRNA via Ribosome Queuing. Mol Cell *70*, 254-264 e256. 10.1016/j.molcel.2018.03.015.
98. Lin, Y., May, G.E., Kready, H., Nazzaro, L., Mao, M., Spealman, P., Creeger, Y., and McManus, C.J. (2019). Impacts of uORF codon identity and position on translation regulation. Nucleic Acids Res *47*, 9358-9367. 10.1093/nar/gkz681.
99. Wei, J., Wu, C., and Sachs, M.S. (2012). The arginine attenuator peptide interferes with the ribosome peptidyl transferase center. Mol Cell Biol *32*, 2396-2406. 10.1128/MCB.00136-12.
100. Ito, K., and Chiba, S. (2013). Arrest peptides: cis-acting modulators of translation. Annu Rev Biochem *82*, 171-202. 10.1146/annurev-biochem-080211-105026.
101. Vilela, C., and McCarthy, J.E. (2003). Regulation of fungal gene expression via short open reading frames in the mRNA 5'untranslated region. Mol Microbiol *49*, 859-867. 10.1046/j.1365-2958.2003.03622.x.
102. Lovett, P.S., and Rogers, E.J. (1996). Ribosome regulation by the nascent peptide. Microbiol Rev *60*, 366-385. 10.1128/mr.60.2.366-385.1996.
103. Zhao, T., Chen, Y.-M., Li, Y., Wang, J., Chen, S., Gao, N., and Qian, W. (2021). Disome-seq reveals widespread ribosome collisions that promote cotranslational protein folding. Genome Biology *22*, 16. 10.1186/s13059-020-02256-0.
104. Meydan, S., and Guydosh, N.R. (2020). Disome and Trisome Profiling Reveal Genome-wide Targets of Ribosome Quality Control. Mol Cell *79*, 588-602.e586.
105. Han, P., Shichino, Y., Schneider-Poetsch, T., Mito, M., Hashimoto, S., Udagawa, T., Kohno, K., Yoshida, M., Mishima, Y., Inada, T., and Iwasaki, S. (2020). Genome-wide Survey of Ribosome Collision. Cell Rep *31*, 107610. 10.1016/j.celrep.2020.107610.
106. Juszkiewicz, S., Slodkowicz, G., Lin, Z., Freire-Pritchett, P., Peak-Chew, S.Y., and Hegde, R.S. (2020). Ribosome collisions trigger cis-acting feedback inhibition of translation initiation. Elife *9*. 10.7554/eLife.60038.
107. Simms, C.L., Yan, L.L., Qiu, J.K., and Zaher, H.S. (2019). Ribosome Collisions Result in +1 Frameshifting in the Absence of No-Go Decay. Cell Reports *28*, 1679-1689.e1674.
108. Goldman, D.H., Livingston, N.M., Movsik, J., Wu, B., and Green, R. (2021). Live-cell imaging reveals kinetic determinants of quality control triggered by ribosome stalling. Mol Cell *81*, 1830-1840.e1838.
109. Sundaramoorthy, E., Leonard, M., Mak, R., Liao, J., Fulzele, A., and Bennett, E.J. (2017). ZNF598 and RACK1 Regulate Mammalian Ribosome-Associated Quality Control Function by Mediating Regulatory 40S Ribosomal Ubiquitylation. Mol Cell *65*, 751-760.e754. 10.1016/j.molcel.2016.12.026.
110. Juszkiewicz, S., Chandrasekaran, V., Lin, Z., Kraatz, S., Ramakrishnan, V., and Hegde, R.S. (2018). ZNF598 Is a Quality Control Sensor of Collided Ribosomes. Mol Cell *72*, 469-481.e467. 10.1016/j.molcel.2018.08.037.
111. Ugajin, N., Imami, K., Takada, H., Ishihama, Y., Chiba, S., and Mishima, Y. (2023). Znf598-mediated Rps10/eS10 ubiquitination contributes to the ribosome ubiquitination dynamics during zebrafish development. Rna *29*, 1910-1927. 10.1261/rna.079633.123.
112. Geng, J., Li, S., Li, Y., Wu, Z., Bhurtel, S., Rimal, S., Khan, D., Ohja, R., Brandman, O., and Lu, B. (2024). Stalled translation by mitochondrial stress upregulates a CNOT4-ZNF598 ribosomal quality control pathway important for tissue homeostasis. Nat Commun *15*, 1637. 10.1038/s41467-024-45525-3.
113. Narita, M., Denk, T., Matsuo, Y., Sugiyama, T., Kikuguchi, C., Ito, S., Sato, N., Suzuki, T., Hashimoto, S., Machova, I., et al. (2022). A distinct mammalian disome collision interface harbors K63-linked polyubiquitination of uS10 to trigger hRQT-mediated subunit dissociation. Nat Commun *13*, 6411. 10.1038/s41467-022-34097-9.
114. Best, K., Ikeuchi, K., Kater, L., Best, D., Musial, J., Matsuo, Y., Berninghausen, O., Becker, T., Inada, T., and Beckmann, R. (2023). Structural basis for clearing of ribosome collisions by the RQT complex. Nat Commun *14*, 921. 10.1038/s41467-023-36230-8.

115. Hashimoto, S., Sugiyama, T., Yamazaki, R., Nobuta, R., and Inada, T. (2020). Identification of a novel trigger complex that facilitates ribosome-associated quality control in mammalian cells. Sci Rep *10*, 3422. 10.1038/s41598-020-60241-w.
116. Madern, M.F., Yang, S., Witteveen, O., Segeren, H.A., Bauer, M., and Tanenbaum, M.E. (2025). Long-term imaging of individual ribosomes reveals ribosome cooperativity in mRNA translation. Cell. 10.1016/j.cell.2025.01.016.
117. Garshott, D.M., An, H., Sundaramoorthy, E., Leonard, M., Vicary, A., Harper, J.W., and Bennett, E.J. (2021). iRQC, a surveillance pathway for 40S ribosomal quality control during mRNA translation initiation. Cell Rep *36*, 109642. 10.1016/j.celrep.2021.109642.
118. Ford, P.W., Garshott, D.M., Narasimhan, M., Ge, X., Jordahl, E.M., Subramanya, S., and Bennett, E.J. (2025). RNF10 and RIOK3 facilitate 40S ribosomal subunit degradation upon 60S biogenesis disruption or amino acid starvation. Cell Rep *44*, 115371. 10.1016/j.celrep.2025.115371.
119. Cho, P.F., Poulin, F., Cho-Park, Y.A., Cho-Park, I.B., Chicoine, J.D., Lasko, P., and Sonenberg, N. (2005). A new paradigm for translational control: inhibition via 5'-3' mRNA tethering by Bicoid and the eIF4E cognate 4EHP. Cell *121*, 411-423. 10.1016/j.cell.2005.02.024.
120. Singh, A.P., Wu, P., Ryabichko, S., Raimundo, J., Swan, M., Wieschaus, E., Gregor, T., and Toettcher, J.E. (2022). Optogenetic control of the Bicoid morphogen reveals fast and slow modes of gap gene regulation. Cell Rep *38*, 110543. 10.1016/j.celrep.2022.110543.
121. Hannon, C.E., Blythe, S.A., and Wieschaus, E.F. (2017). Concentration dependent chromatin states induced by the bicoid morphogen gradient. Elife *6*. 10.7554/eLife.28275.
122. Struhl, G., Struhl, K., and Macdonald, P.M. (1989). The gradient morphogen bicoid is a concentration-dependent transcriptional activator. Cell *57*, 1259-1273. 10.1016/0092-8674(89)90062-7.
123. Dubnau, J., and Struhl, G. (1996). RNA recognition and translational regulation by a homeodomain protein. Nature *379*, 694-699. 10.1038/379694a0.
124. Wethmar, K., Begay, V., Smink, J.J., Zaragoza, K., Wiesenthal, V., Dorken, B., Calkhoven, C.F., and Leutz, A. (2010). C/EBPbetaDeltauORF mice--a genetic model for uORF-mediated translational control in mammals. Genes Dev *24*, 15-20. 10.1101/gad.557910.
125. Miyake, T., Inoue, Y., Shao, X., Seta, T., Aoki, Y., Nguyen Pham, K.T., Shichino, Y., Sasaki, J., Sasaki, T., Ikawa, M., et al. (2023). Minimal upstream open reading frame of Per2 mediates phase fitness of the circadian clock to day/night physiological body temperature rhythm. Cell Rep *42*, 112157. 10.1016/j.celrep.2023.112157.
126. Xing, S., Chen, K., Zhu, H., Zhang, R., Zhang, H., Li, B., and Gao, C. (2020). Fine-tuning sugar content in strawberry. Genome Biol *21*, 230. 10.1186/s13059-020-02146-5.
127. Zhang, H., Si, X., Ji, X., Fan, R., Liu, J., Chen, K., Wang, D., and Gao, C. (2018). Genome editing of upstream open reading frames enables translational control in plants. Nat Biotechnol *36*, 894-898. 10.1038/nbt.4202.
128. Si, X., Zhang, H., Wang, Y., Chen, K., and Gao, C. (2020). Manipulating gene translation in plants by CRISPR-Cas9-mediated genome editing of upstream open reading frames. Nat Protoc *15*, 338-363. 10.1038/s41596-019-0238-3.
129. Xue, C., Qiu, F., Wang, Y., Li, B., Zhao, K.T., Chen, K., and Gao, C. (2023). Tuning plant phenotypes by precise, graded downregulation of gene expression. Nat Biotechnol. 10.1038/s41587-023-01707-w.
130. Luo, Z., and Sachs, M.S. (1996). Role of an upstream open reading frame in mediating arginine-specific translational control in Neurospora crassa. J Bacteriol *178*, 2172-2177. 10.1128/jb.178.8.2172-2177.1996.
131. Rosenbloom, K.R., Armstrong, J., Barber, G.P., Casper, J., Clawson, H., Diekhans, M., Dreszer, T.R., Fujita, P.A., Guruvadoo, L., Haeussler, M., et al. (2015). The UCSC Genome Browser database: 2015 update. Nucleic Acids Res *43*, D670-681. 10.1093/nar/gku1177.
132. Chakraborty, M., Chang, C.H., Khost, D.E., Vedanayagam, J., Adrion, J.R., Liao, Y., Montooth, K.L., Meiklejohn, C.D., Larracuente, A.M., and Emerson, J.J. (2021). Evolution of genome structure in the *Drosophila simulans* species complex. Genome Res *31*, 380-396. 10.1101/gr.263442.120.

133. Chiaromonte, F., Yap, V.B., and Miller, W. (2002). Scoring pairwise genomic sequence alignments. Pac Symp Biocomput, 115-126. 10.1142/9789812799623_0012.
134. Blanchette, M., Kent, W.J., Riemer, C., Elnitski, L., Smit, A.F., Roskin, K.M., Baertsch, R., Rosenbloom, K., Clawson, H., Green, E.D., et al. (2004). Aligning multiple genomic sequences with the threaded blockset aligner. Genome Res *14*, 708-715. 10.1101/gr.1933104.
135. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. journal *17*, 10-12.
136. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359. 10.1038/nmeth.1923.
137. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.
138. Dunn, J.G., and Weissman, J.S. (2016). Plastid: nucleotide-resolution analysis of next-generation sequencing and genomics data. BMC Genomics *17*, 958. 10.1186/s12864-016-3278-x.
139. Ni, J.Q., Zhou, R., Czech, B., Liu, L.P., Holderbaum, L., Yang-Zhou, D., Shim, H.S., Tao, R., Handler, D., Karpowicz, P., et al. (2011). A genome-scale shRNA resource for transgenic RNAi in *Drosophila*. Nat Methods *8*, 405-407. 10.1038/nmeth.1592.
140. Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166-169. 10.1093/bioinformatics/btu638.
141. McCarthy, D.J., Chen, Y., and Smyth, G.K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res *40*, 4288-4297. 10.1093/nar/gks042.
142. Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS *16*, 284-287. 10.1089/omi.2011.0118.
143. Xiao, Z., Zou, Q., Liu, Y., and Yang, X. (2016). Genome-wide assessment of differential translations with ribosome profiling data. Nat Commun *7*, 11194. 10.1038/ncomms11194.

1262 **Main Tables and Figures**

1263 **Table 1.** Numbers of genes showing different magnitudes of TE changes between uORFs and CDS at
1264 the interspecific level

1265

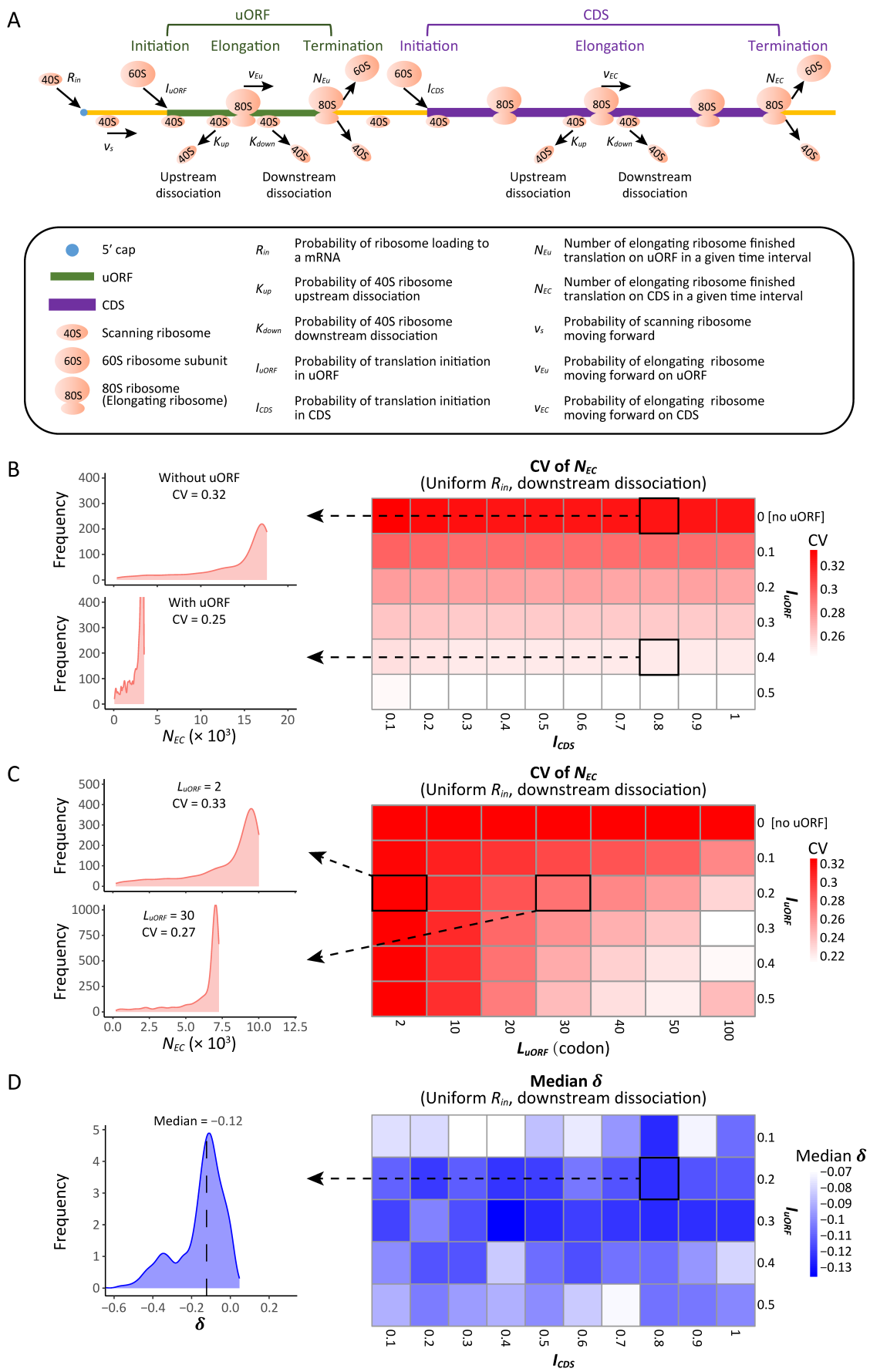
Sample types	# of expressed uORFs *	$\beta_u \neq 1$ (%) **	# of expressed CDSs *	$\beta_c \neq 1$ (%) **	uORF-CDS pairs with $\beta_u > 1$			uORF-CDS pairs with $\beta_u < 1$		
					Total	$\gamma > 1$	$\gamma < 1$	Total	$\gamma > 1$	$\gamma < 1$
0–2 h embryo	7,704	1,193 (15.49)	7,934	4,189 (52.80)	770	0	366	423	69	0
2–6 h embryo	7,822	567 (7.25)	9,795	3,063 (31.27)	249	0	63	318	135	0
6–12 h embryo	10,400	1,040 (10.00)	10,643	2,924 (27.47)	973	0	641	67	22	1
12–24 h embryo	11,365	535 (4.71)	11,158	3,537 (31.70)	454	0	234	81	8	2
Larva	10,008	464 (4.64)	11,831	3,554 (30.04)	110	0	26	354	177	0
Pupa	12,309	635 (5.16)	12,209	4,087 (33.48)	136	1	15	499	210	0
Male body	10,894	197 (1.81)	12,284	2,432 (19.80)	122	0	29	75	4	0
Male head	10,904	144 (1.32)	10,447	1,151 (11.02)	119	0	31	25	5	0
Female body	9,809	340 (3.47)	11,002	3,605 (32.77)	279	0	17	61	3	0
Female head	10,935	332 (3.04)	10,545	1,270 (12.04)	324	0	152	8	3	0

1266

1267 * Only uORFs and CDSs with an mRNA RPKM > 0.1 in both *D. melanogaster* and *D. simulans* were
1268 considered in each sample pair in the analysis.

1269 ** $\beta_u = TE_{uORF,sim}/TE_{uORF,mel}$ is the fold change of TE_{uORF} in *D. simulans* relative to *D. melanogaster*
1270 for each sample. $\beta_c = TE_{CDS,sim}/TE_{CDS,mel}$ is the fold change of TE_{CDS} in *D. simulans* relative to *D.*
1271 *melanogaster* for each sample. $\gamma = \beta_c/\beta_u$. For each CDS–uORF pair, either β_u>1 and $\gamma < 1$ or β_u<1
1272 and $\gamma > 1$ means that the magnitude of TE change is lower for a CDS than for a uORF. The statistical
1273 significance of β_c, β_u, and γ were determined according to an FDR < 0.05.

1274


A
uORF
CDS
Initiation
Elongation
Termination
Initiation
Elongation
Termination
Upstream dissociation
Downstream dissociation
Upstream dissociation
Downstream dissociation
5′ cap
uORF
CDS
Scanning ribosome
60S ribosome subunit
80S ribosome (Elongating ribosome)
Probability of ribosome loading to a mRNA
Probability of 40S ribosome upstream dissociation
Probability of 40S ribosome downstream dissociation
Probability of translation initiation in uORF
Probability of translation initiation in CDS
Number of elongating ribosome finished translation on uORF in a given time interval
Number of elongating ribosome finished translation on CDS in a given time interval
Probability of scanning ribosome moving forward
Probability of elongating ribosome moving forward on uORF
Probability of elongating ribosome moving forward on CDS
B
Without uORF
CV = 0.32
With uORF
CV = 0.25
Frequency
N_{EC} (× 10³)
CV of N_{EC}
(Uniform R_{in}, downstream dissociation)
0 [no uORF]
I_{uORF}
I_{CDS}
CV
C
L_{uORF} = 2
CV = 0.33
L_{uORF} = 30
CV = 0.27
L_{uORF} (codon)
D
Median = −0.12
δ
Median δ
(Uniform R_{in}, downstream dissociation)

1275

Figure 1. Modeling simulation of uORF-mediated translation buffering. (A) Model schema of the modified ICIER model (on the top); the parameters are listed in the box below the schema. (B) Heatmap showing the CVs of CDS translation rate (N_{EC}) under different I_{CDS} (x-axis) and I_{uORF} (y-axis) combinations with a uniform distribution of R_{in} input and the downstream dissociation model. The left panels elicited by the dotted lines from specific squares of right heatmap were two examples showing the distribution of N_{EC} under I_{CDS} = 0.8 & I_{uORF} = 0 (top panel, without uORF) and I_{CDS} = 0.8 & I_{uORF} = 0.4 (bottom panel, with uORF). (C) Heatmap showing CVs of CDS translation rate (N_{EC}) under different L_{uORF} (x-axis) and I_{uORF} (y-axis) combinations with a uniform distribution of R_{in} input and the downstream dissociation model. The left panels elicited by the dotted lines from specific squares of right heatmap were two examples showing the distribution of N_{EC} under L_{uORF} = 2 & I_{uORF} = 0.2 (top panel) and L_{uORF} = 30 & I_{uORF} = 0.2 (bottom panel). (D) Heatmap showing median δ $[log_2(\Delta N_{EC}/\Delta N_{EU1})]$ under different I_{CDS} (x-axis) and I_{uORF} (y-axis) combinations with a uniform distribution of R_{in} input and the downstream dissociation model. The left panel elicited by the dotted line from a specific square of right heatmap was an example showing the distribution of δ under I_{CDS} = 0.8 & I_{uORF} = 0.2. The vertical dashed line indicated the median value of δ.

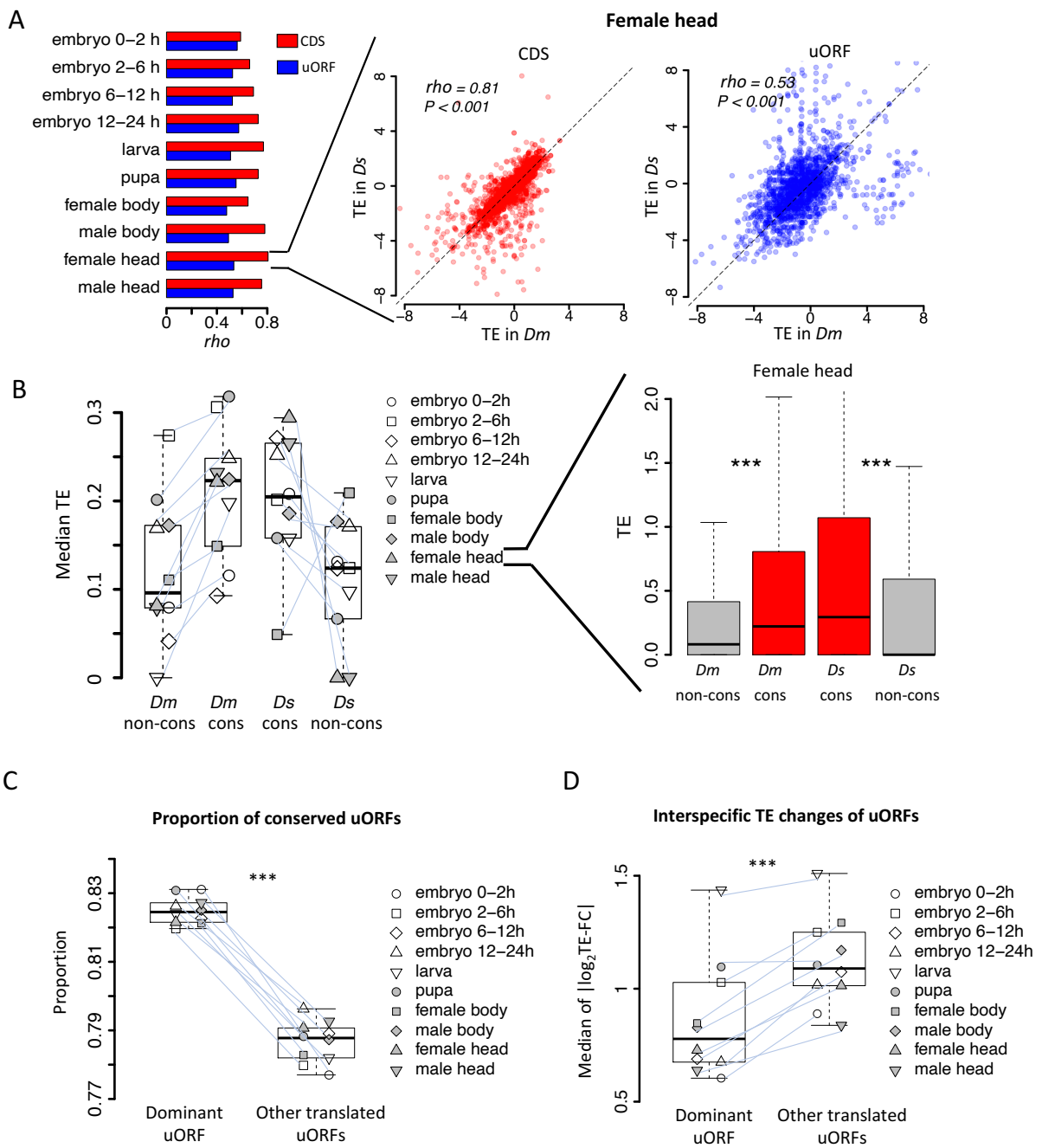


Figure 2. Conservation and translation of uORFs between *D. melanogaster* and *D. simulans*. (A) Spearman's correlation coefficients (*rho*, represented by the bars) of TEs between *Dm* (*D. melanogaster*) and *Ds* (*D. simulans*) for CDS (red) and uORFs (blue). All *P*-values for Spearman's correlation are less than 0.001. The *P* value for the comparison between *rho* values of CDSs TE and uORFs TE is 0.002 (Wilcoxon signed-rank test). Data for the female head sample is shown as an example in the right panel. The x- and y-axes represent the TEs in *Dm* and *Ds*. (B) The median of TE of conserved and species-specific uORFs in each sample. Each dot represents the median TE of a sample for a specific uORF class. Data from the female head sample is shown as an example in the right panel. *P* values were obtained from Wilcoxon rank-sum tests. ***, $P < 0.001$. (C) Fraction of conserved uORFs among

1302 dominant uORFs and other translated uORFs in each sample. The paired samples in *Dm* and *Ds* were
1303 linked together. The *P* value was obtained by the paired Wilcoxon signed-rank test. ***, $P < 0.001$. (D)
1304 Absolute values of the interspecific TE fold changes ($\log_2$TE-FC) of dominant uORFs and the other
1305 translated uORFs in each sample. The paired samples in *Dm* and *Ds* were linked together. The median
1306 value of each sample is shown. The *P* value was obtained via the paired Wilcoxon signed-rank test. ***,
1307 $P < 0.001$. Data from the female head sample were used as an example in the right panel.
1308

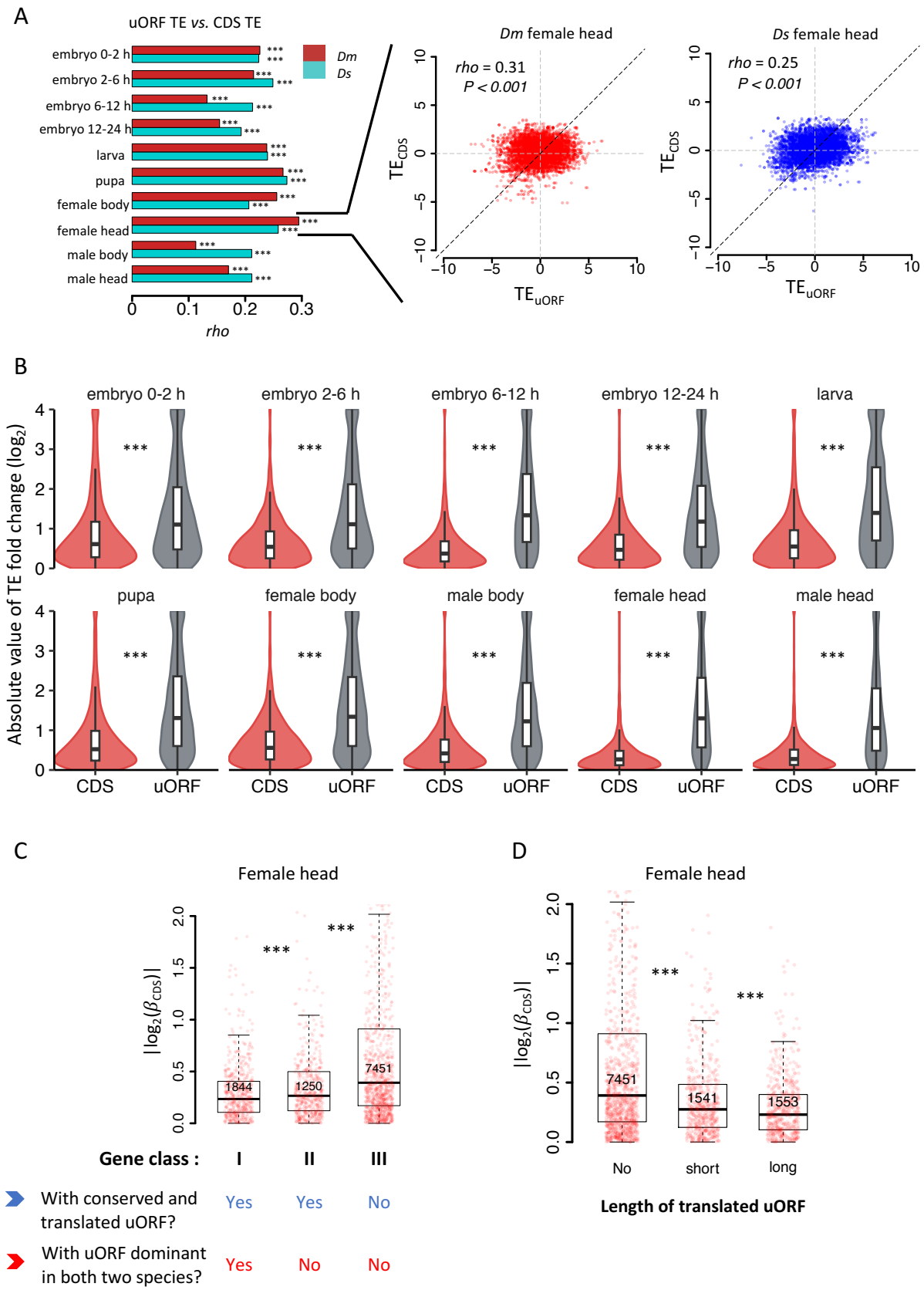
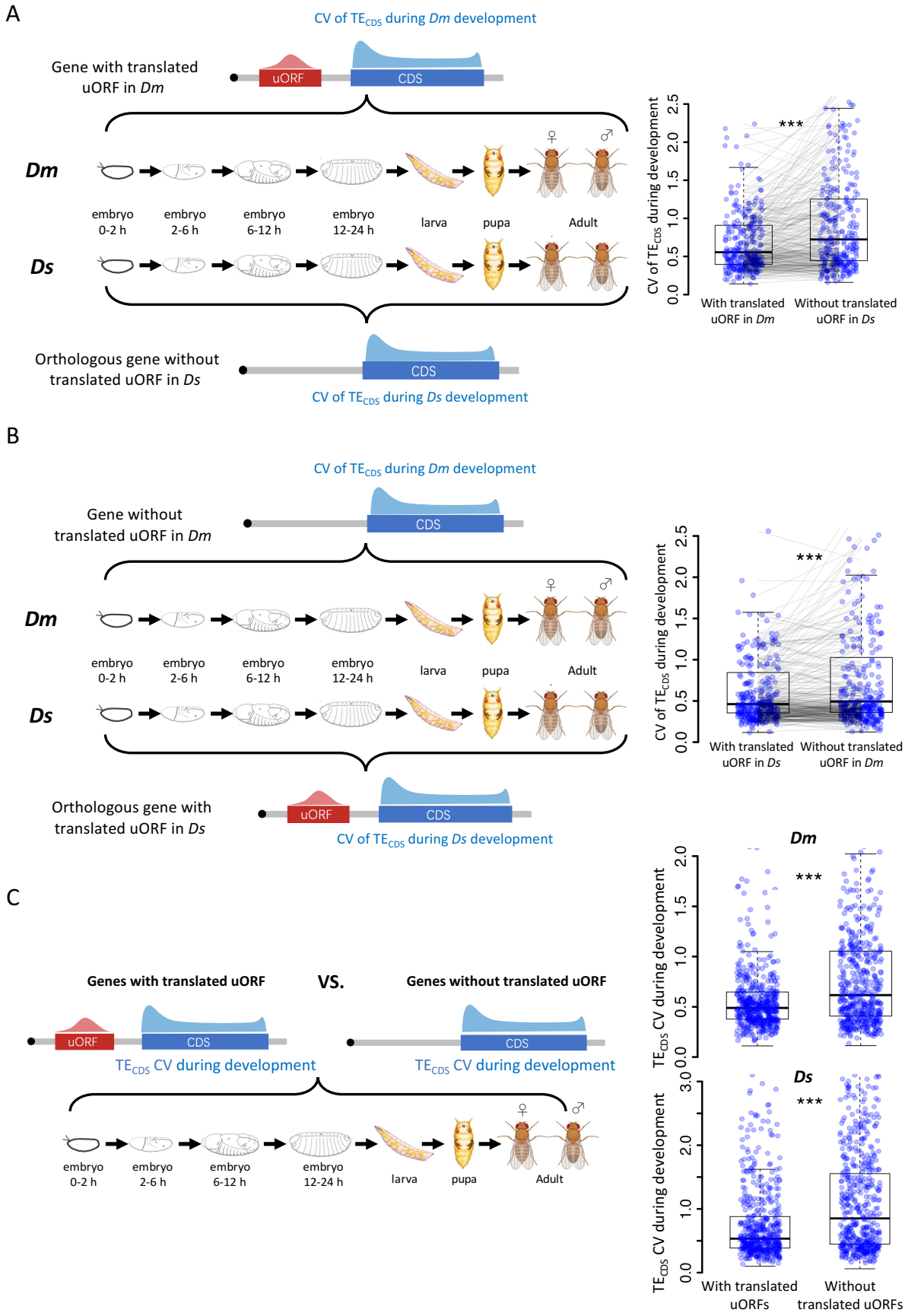


Figure 3. uORFs reduce CDS translational divergence between *D. melanogaster* and *D. simulans*. (A) The correlation of uORF TEs and the corresponding CDS TEs in 10 samples of *Dm* (*D. melanogaster*) and *Ds* (*D. simulans*). The bars represent Spearman's correlation coefficient (*rho*). In all

samples, we obtained both *P* values < 0.001. Data for the female head sample of *Dm* and *Ds* are shown as examples in the right panel. (B) The absolute values of interspecific TE changes for CDS and uORF in each sample between two species. For visualization purposes, all values greater than 4 were assigned a value of 4. ***, *P* < 0.001, Wilcoxon rank-sum test. (C) Genes expressed in female heads (mRNA RPKM > 0.1 in both species) were classified into three classes according to whether a gene had a conserved and dominantly translated uORF or not. Boxplots showing interspecific CDS TE variability $|log_2(\beta_c)|$ of different gene classes. *P* values were calculated using Wilcoxon rank-sum tests between the neighboring groups. ***, *P* < 0.001. (D) Genes expressed in female heads were classified into three classes according to the length of translated uORFs. Boxplots showing interspecific CDS TE variability $|log_2(\beta_c)|$ of different gene classes. *P* values were calculated using Wilcoxon rank-sum tests between the neighboring groups. ***, *P* < 0.001.


A
CV of TE$_{CDS}$ during *Dm* development
Gene with translated uORF in *Dm*
uORF
CDS
Dm
Ds
embryo 0-2 h
embryo 2-6 h
embryo 6-12 h
embryo 12-24 h
larva
pupa
Adult
Orthologous gene without translated uORF in *Ds*
CV of TE$_{CDS}$ during *Ds* development
CV of TE$_{CDS}$ during development
***
With translated uORF in *Dm*
Without translated uORF in *Ds*
B
CV of TE$_{CDS}$ during *Dm* development
Gene without translated uORF in *Dm*
Orthologous gene with translated uORF in *Ds*
CV of TE$_{CDS}$ during *Ds* development
With translated uORF in *Ds*
Without translated uORF in *Dm*
C
Genes with translated uORF
VS.
Genes without translated uORF
TE$_{CDS}$ CV during development
TE$_{CDS}$ CV during development
Dm
Ds
With translated uORFs
Without translated uORFs

1325
1326

1327 **Figure 4. uORFs could reduce CDS translational fluctuation during *Drosophila* development.** (A)
1328 The CV of TE_{CDS} across 10 *Dm* (*D. melanogaster*) samples and 10 *Ds* (*D. simulans*) samples. The
1329 selected gene with uORFs translated (TE > 0.1) in at least one *Dm* sample but its homologous gene
1330 without translated uORF in *Ds* samples. Each pair of dots linked by a gray line represents a pair of
1331 homologous genes in *Dm* and *Ds*. ***, $P < 0.001$, Wilcoxon signed-rank test. (B) The CV of TE_{CDS}
1332 across 10 *Dm* samples and 10 *Ds* samples. The selected gene with uORFs translated (TE > 0.1) in at
1333 least one *Ds* sample but its homologous gene without translated uORF in *Dm* samples. Each pair of dots
1334 linked by a gray line represents a pair of homologous genes in *Dm* and *Ds*. ***, $P < 0.001$, Wilcoxon
1335 signed-rank test. (C) Within each *Drosophila* species, the CV of TE_{CDS} of genes with translated uORFs
1336 compared to genes without the translated uORFs. The *P* values are obtained by the Wilcoxon rank-sum
1337 test. ***, $P < 0.001$.
1338

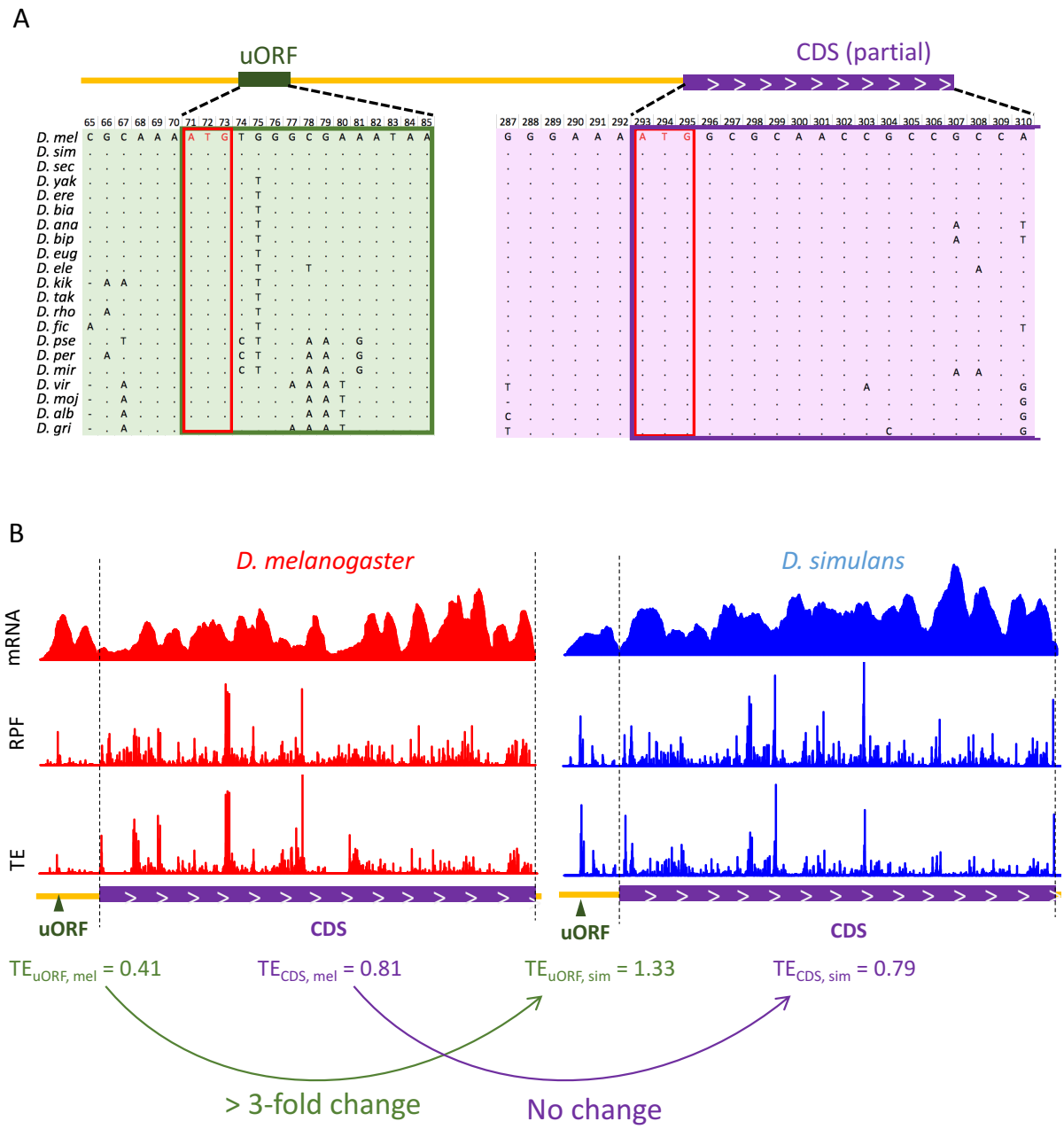
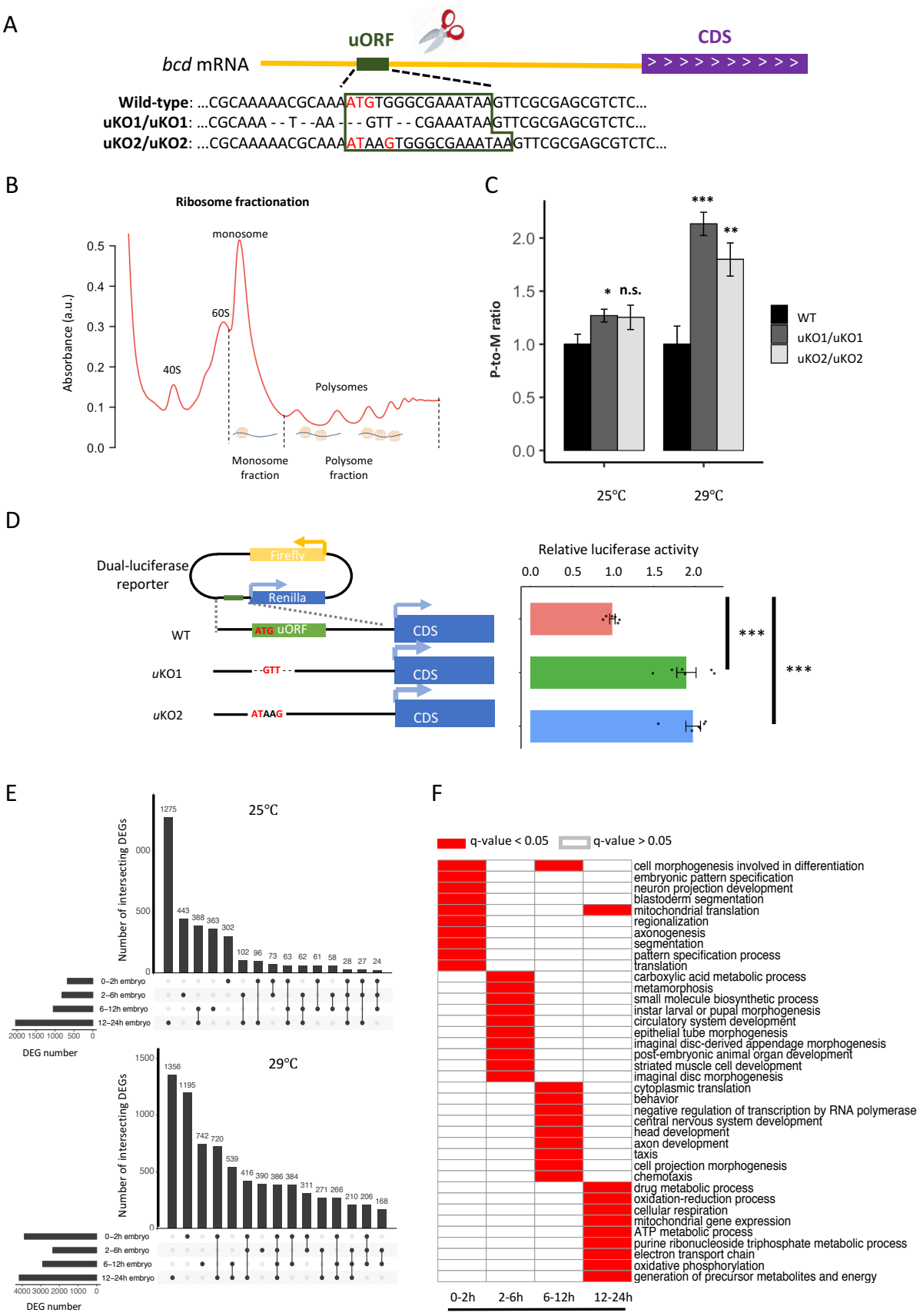
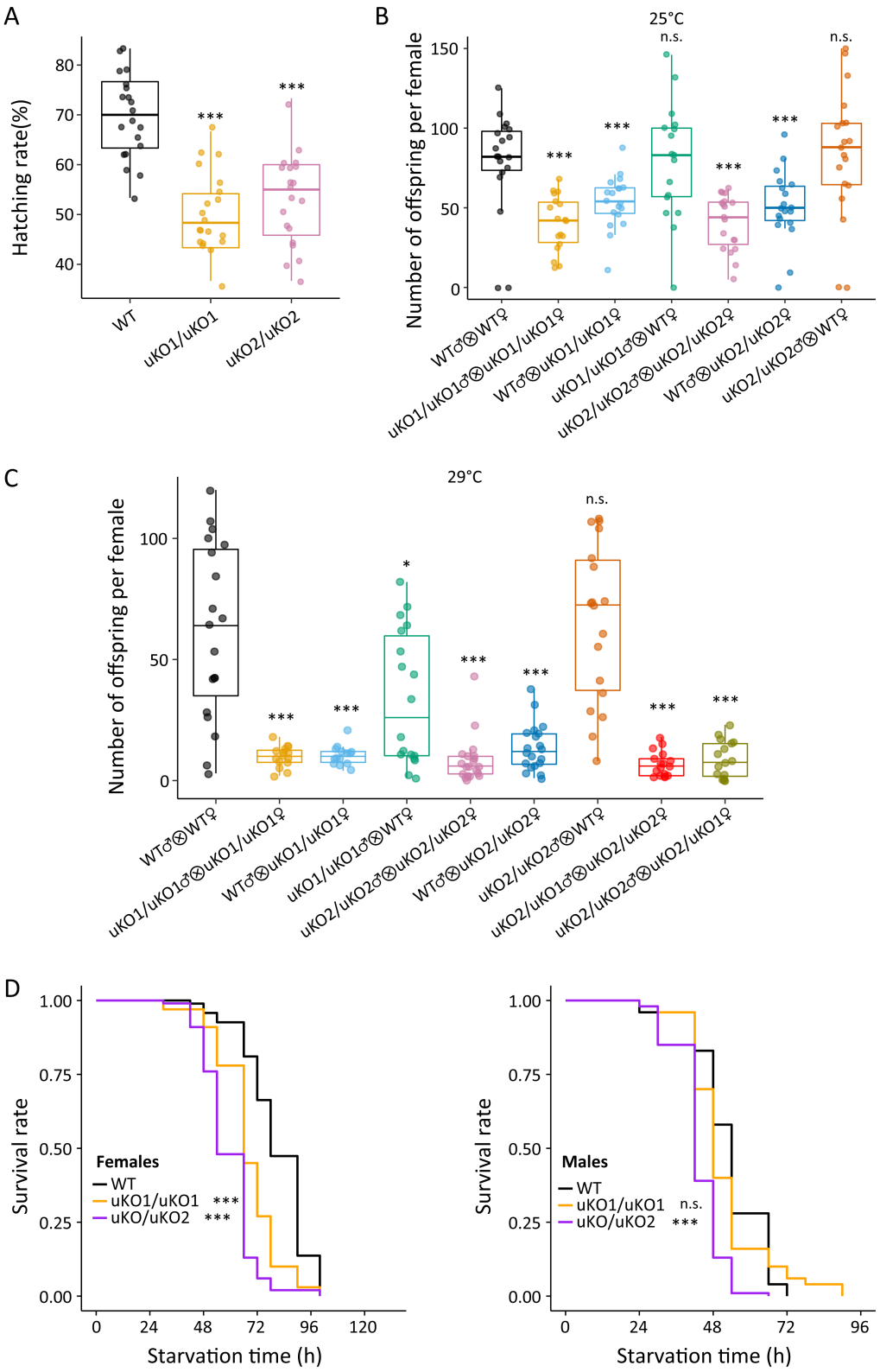


Figure 5. **The strong buffering effect of the *bcd* uORF on CDS translation between the two *Drosophila* species.** (A) Multiple sequence alignment of the *bcd* uORF and partial CDS in *D. melanogaster* and 20 other *Drosophila* species. The uORF and CDS are boxed in green and purple, respectively. The start codons of the uORF and CDS are boxed in red. (B) The coverage of mRNA-Seq (top), Ribo-Seq (middle), and TEs (bottom) of the *bcd* uORF and CDS in 0-2 h embryos of *D. melanogaster* (red) and *D. simulans* (blue). The uORF and CDS are denoted at the lower panel with dark green triangles and purple boxes, respectively. The 2 dashed lines mark the CDS region. The uORF TE, CDS TE and their interspecific changes were labeled at the bottom.


A
uORF
CDS
bcd mRNA
Wild-type: ...CGCAAAAACGCAAAATGTGGGCGAAATAAGTTCGCGAGCGTCTC...
uKO1/uKO1: ...CGCAAA - - T - -AA - - - GTT - - CGAAATAAGTTCGCGAGCGTCTC...
uKO2/uKO2: ...CGCAAAAACGCAAAATAAGTGGGCGAAATAAGTTCGCGAGCGTCTC...
B
Ribosome fractionation
Absorbance (a.u.)
monosome
60S
40S
Polysomes
Monosome fraction
Polysome fraction
C
P-to-M ratio
n.s.
WT
uKO1/uKO1
uKO2/uKO2
25°C
29°C
D
Dual-luciferase reporter
Firefly
Renilla
WT
uKO1
uKO2
ATG uORF
GTT
ATAAG
CDS
Relative luciferase activity
E
25°C
29°C
Number of intersecting DEGs
0–2h embryo
2–6h embryo
6–12h embryo
12–24h embryo
DEG number
F
q-value < 0.05
q-value > 0.05
cell morphogenesis involved in differentiation
embryonic pattern specification
neuron projection development
blastoderm segmentation
mitochondrial translation
regionalization
axonogenesis
segmentation
pattern specification process
translation
carboxylic acid metabolic process
metamorphosis
small molecule biosynthetic process
instar larval or pupal morphogenesis
circulatory system development
epithelial tube morphogenesis
imaginal disc-derived appendage morphogenesis
post-embryonic animal organ development
striated muscle cell development
imaginal disc morphogenesis
cytoplasmic translation
behavior
negative regulation of transcription by RNA polymerase
central nervous system development
head development
axon development
taxis
cell projection morphogenesis
chemotaxis
drug metabolic process
oxidation-reduction process
cellular respiration
mitochondrial gene expression
ATP metabolic process
purine ribonucleoside triphosphate metabolic process
electron transport chain
oxidative phosphorylation
generation of precursor metabolites and energy
0-2h
2-6h
6-12h
12-24h
Embryo

1351 **Figure 6. Knocking out the *bcd* uORF increases CDS translation and perturbs the transcriptome**
1352 **during *D. melanogaster* embryogenesis.** (A) Genotypes of WT and two uORF knock-out strains
1353 (uKO1 and uKO2) generated by CRISPR-Cas9 technology. The uORF is boxed in dark green, and the
1354 red ATG represents the start codon of the uORF in the *D. melanogaster* genome. (B) Two ribosome
1355 fractions (monosome and polysome) of 0-2 h embryos were separated in a sucrose density gradient.
1356 Relative RNA abundance in the monosome and polysome fractions was quantified by real-time
1357 quantitative PCR. (C) P-to-M ratio of *bcd* mRNA (*bcd* mRNA abundance in polysome fraction/*bcd*
1358 mRNA abundance in monosome fraction) at 25°C (left) and 29°C (right). The P-to-M ratios of mutants
1359 were normalized to WT controls at 25°C and at 29°C, respectively. Error bars represent the S.E. of six
1360 biological replicates. Asterisks indicate statistical significance (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$;
1361 n.s., $P > 0.05$). (D) Dual-luciferase assay for *bcd* WT uORF and mutated uORF. The reporter structures
1362 of the WT and uORF mutants are illustrated on the left. The uORF mutant sequence was the same as
1363 that in the fly mutant created with CRISPR-Cas9 technology. The relative activity of *Renilla* luciferase
1364 was normalized to that of firefly luciferase. Error bars represent the S.E. of six biological replicates.
1365 Asterisks indicate statistical significance (***, $P < 0.001$). (E) The number of DEGs in each stage and
1366 their intersection with each other at 25°C (top) and 29°C (bottom). (F) Gene ontology analysis of DEGs
1367 at 29°C in each stage. The biological process (BP) terms with q-values < 0.05 in each stage are indicated
1368 in red and others are indicated in white.
1369

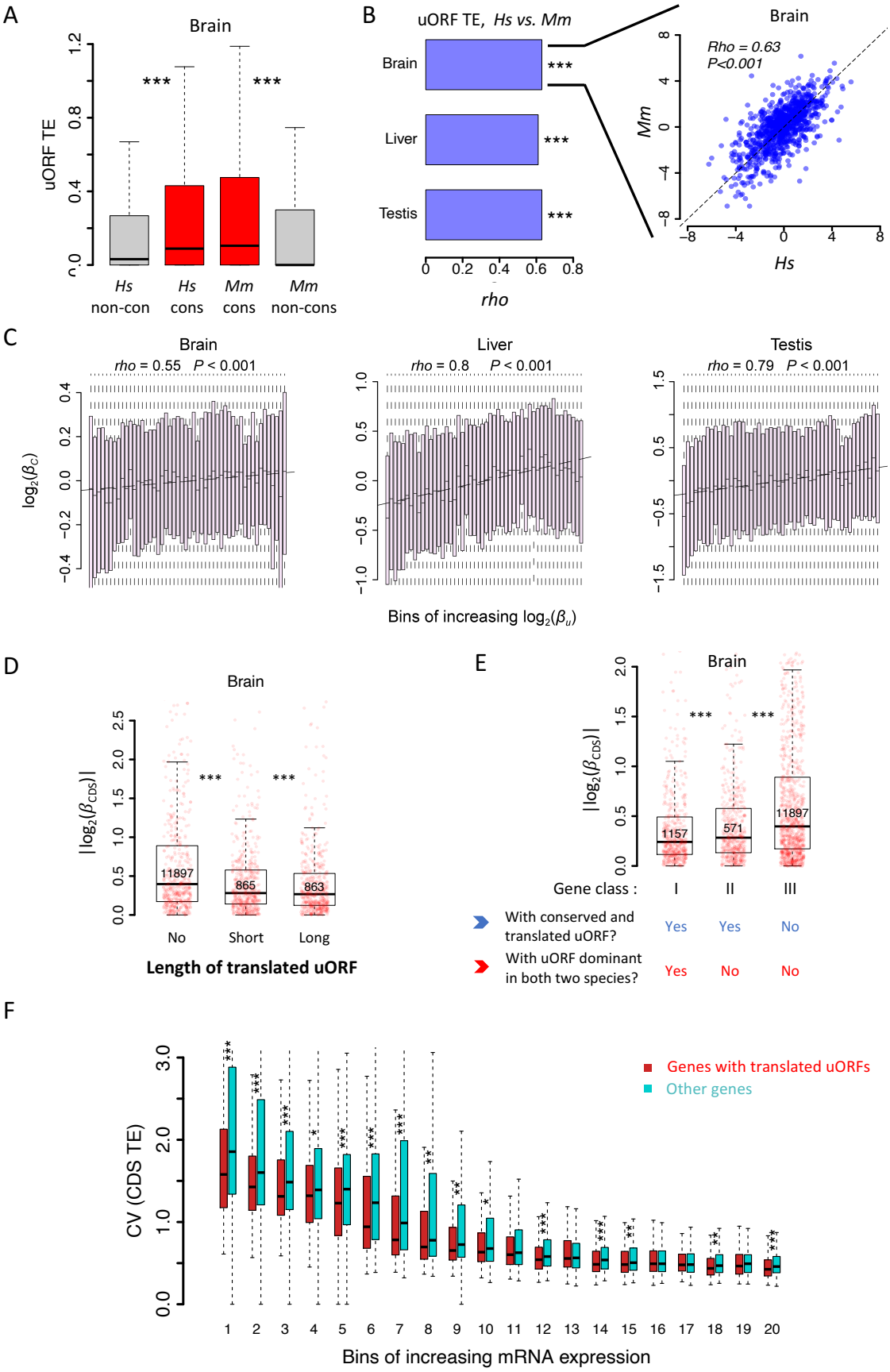
1370



1371

1372 **Figure 7. Knockout of the *bcd* uORF reduces offspring number and starvation resistance**. (A)
1373 Comparison of the hatching rates (%) of mutant and WT offspring (n = 20, Wilcoxon rank-sum test;
1374 ***, $P < 0.001$). (B) The offspring number per maternal parent in different crosses over 10 days at 25°C.
1375 Asterisks indicate significant differences between various crosses and crosses of WT females with WT

1376 males (n = 20, Wilcoxon rank-sum test; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; n.s., $P > 0.05$). The
1377 different crosses were denoted as the x-axis labels. (C) The offspring number per maternal parent in
1378 different crosses over 10 days at 29°C. (D) Survival curves of WT and mutant adult flies of females
1379 (left) and males (right) under starvation conditions. The black line represents the WT, the red line
1380 represents the uKO1/uKO1 mutant, and the blue line represents the uKO2/uKO2 mutant. Asterisks
1381 indicate significant differences compared to the WT. (n = 200, log-rank test; ***, $P < 0.001$; n.s., $P >$
1382 0.05).
1383


A
Brain
uORF TE
***
***
Hs non-con
Hs cons
Mm cons
Mm non-cons
B
uORF TE, Hs vs. Mm
Brain
Liver
Testis
rho
Brain
Rho = 0.63
P<0.001
Mm
Hs
C
Brain
rho = 0.55 P < 0.001
Liver
rho = 0.8 P < 0.001
Testis
rho = 0.79 P < 0.001
Bins of increasing log2(βu)
D
Brain
11897
865
863
No
Short
Long
Length of translated uORF
E
Brain
1157
571
11897
Gene class : I II III
With conserved and translated uORF? Yes Yes No
With uORF dominant in both two species? Yes No No
F
Genes with translated uORFs
Other genes
CV (CDS TE)
Bins of increasing mRNA expression

1386 **Figure 8. uORFs function as translational buffers in primates.** (A) Boxplots showing the TEs of
1387 conserved and species-specific uORFs between *Hs* (*H. sapiens*) and *Mm* (*M. mulatta*). Data for the
1388 brain is shown as an example. Wilcoxon rank-sum tests. ***, $P < 0.001$. (B) Spearman's correlation
1389 coefficient (*rho*) of uORFs' TE between humans and macaques. The *rho* values in the brain, liver, and
1390 testis were shown as bar plots. ***, $P < 0.001$. Data for the brain is shown as an example in the right
1391 panel. (C) Correlation between interspecific uORF TE changes ($\log_2\beta_u$) and corresponding CDS TE
1392 changes ($\log_2\beta_C$) in three tissues. The x-axis was divided into 50 equal bins with increasing β_u. (D)
1393 Genes expressed in brains were classified into three classes according to the total length of translated
1394 uORFs. Boxplots showing interspecific CDS TE variability $|log_2(\beta_c)|$ of different gene classes. *P* values
1395 were calculated using Wilcoxon rank-sum tests between the neighboring groups. ***, $P < 0.001$. (E)
1396 Genes expressed in brains (mRNA RPKM > 0.1 in both species) were classified into three classes
1397 according to whether a gene had a conserved and dominantly translated uORF (TE > 0.1) in both species
1398 or not. Boxplots showing interspecific CDS TE variability $|log_2(\beta_c)|$ of different gene classes. *P* values
1399 were calculated using Wilcoxon-rank sum tests between the neighboring groups. ***, $P < 0.001$. (F)
1400 Boxplot showing the coefficients of variation (CVs) of CDS TE among the 69 lymphoblastoid cell lines
1401 (LCLs). Expressed genes (mean mRNA RPKM > 0.1) were divided into 20 bins with increased mRNA
1402 expression levels. In each bin, the genes were divided into two fractions according to whether the gene
1403 had a translated uORF or not. Wilcoxon rank-sum tests. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.
1404

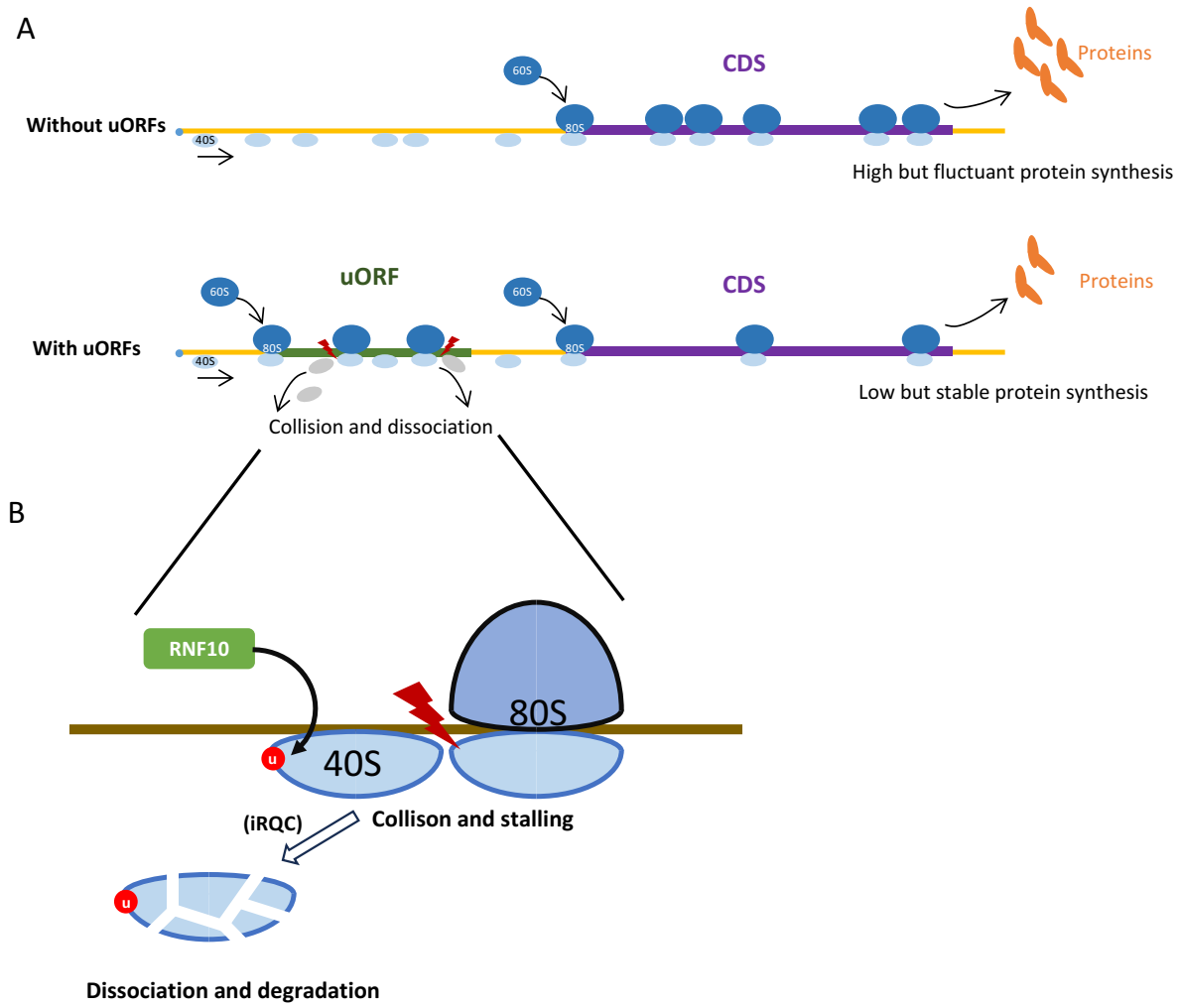


Figure 9. Model illustrating the uORF-mediated translational buffering. (A) uORFs can reduce and stabilize the downstream CDS translation through "ribosomes collision and dissociation" mechanism in an mRNA containing uORFs. In contrast, the CDS translation is higher and more fluctuant in an mRNA without uORFs. (B) The collided and stalled 40S subunit dissociates from mRNA, is ubiquitinated by the E3 ubiquitin ligase RNF10 and degraded by the initiation ribosome quality control (iRQC) pathway.

1413 **Supplementary Materials and Methods**

1414

1415 **The detailed operations per action in the ICIER simulation**

1416 The translation process of an mRNA molecule is modeled as a series of discrete actions. During each
1417 action, certain events may occur with certain probabilities for each entity (40S or 80S ribosome). The
1418 operations per action are described as follows:

1419 1. If the first 10 triplets were vacant of the mRNA, a 40S ribosome would be added there with a
1420 probability of R_{in}.

1421 2. If a 40S ribosome was located at the start codon of the uORF, it would transform into an 80S ribosome
1422 with a probability of I_{uORF}. If a 40S ribosome was located at the start codon of the CDS, it would
1423 transform into an 80S ribosome with a probability of I_{CDS}.

1424 3.If an 80S ribosome was located at the stop codon of the uORF or CDS, it would dissociate from the
1425 mRNA. The dissociation of an 80S ribosome from the uORF or CDS stop codon was counted as a
1426 completed translation event, and the values of N_{EC} and N_{Eu} were incremented by one, respectively.

1427 4. If a 40S ribosome arrived at the 3' end of the mRNA, it would dissociate from the mRNA.

1428 5. For each 40S ribosome on the mRNA, if its next position was vacant (in the downstream-dissociation
1429 model) or not occupied by another 40S ribosome (in the upstream-dissociation model and double-
1430 dissociation model), it would move to the next position with a probability of ν_S.

1431 6. For each 80S ribosome in the uORF, if its next position was vacant (in the upstream-dissociation
1432 model) or not occupied by another 80S ribosome (in the downstream-dissociation model and double-
1433 dissociation model), it would move to the next position with a probability of ν_{Eu}. Similarly, for each
1434 80S ribosome in the CDS, if its next position was vacant (in the upstream-dissociation model) or not
1435 occupied by an 80S ribosome (in the downstream-dissociation model and double-dissociation model),
1436 it would move to the next position with a probability of ν_{EC}.

1437 7. It was checked whether any 80S ribosome collided with a 40S ribosome. If such a collision occurred,
1438 the 40S ribosome involved would dissociate from the mRNA with a probability of K_{up} (upstream
1439 dissociation or double-dissociation model) or K_{down} (downstream dissociation or double-dissociation
1440 model).

1441

1442 These above operations are executed repeatedly until the action count reaches the limit given by the
1443 user (1,000,000 in this study).

1444

Supplementary Figures and Tables

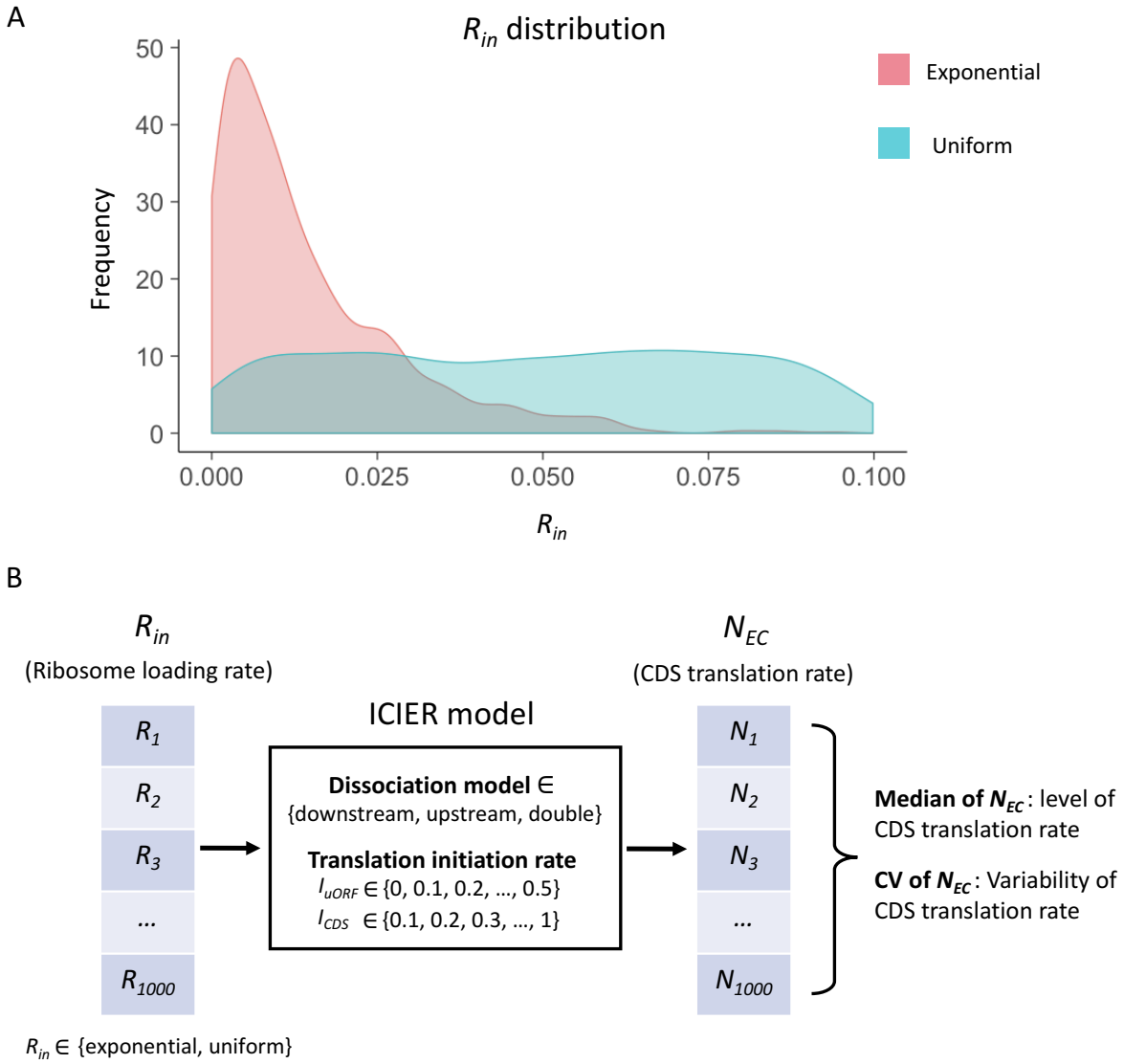
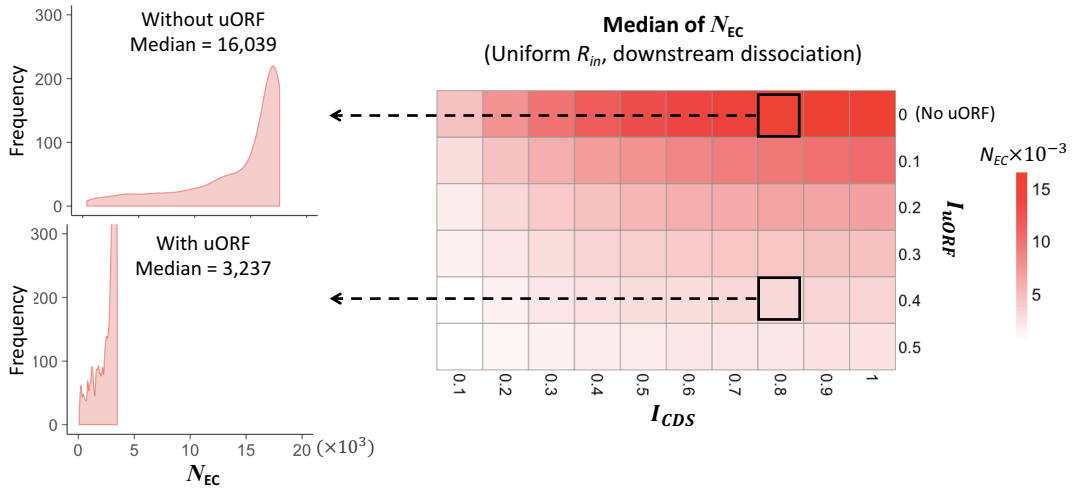
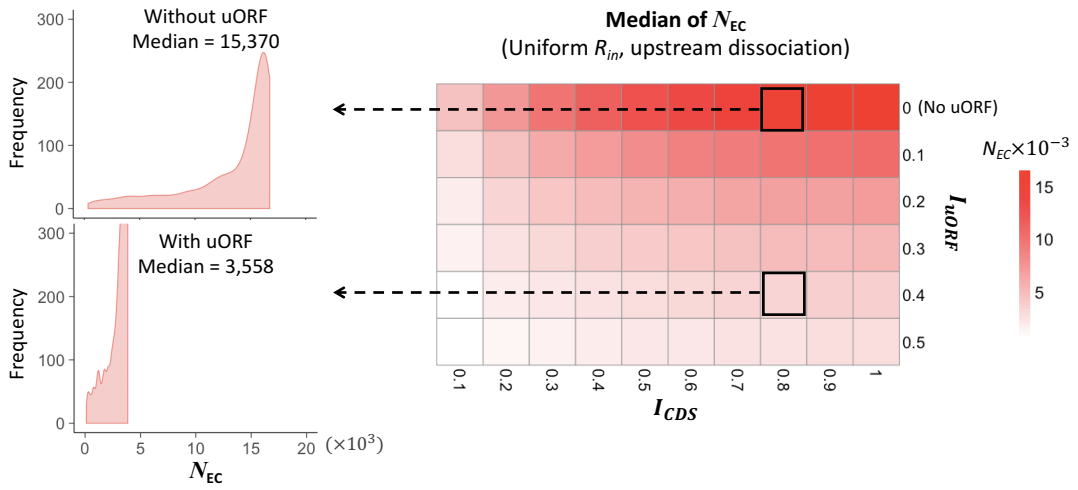


Figure S1. The input distribution and model simulation flow. (A) Two distributions of the R_{in} input were used in our simulation: an exponential distribution and a uniform distribution. A total of 1,000 R_{in} values (ranging from 0 to 0.1) were randomly generated following either distribution. (B) Under a fixed combination of parameters (R_{in} distribution, dissociation model, I_{uORF} and I_{CDS}), the model simulation produces 1000 N_{EC} values from 1000 R_{in} input. The median of these 1000 N_{EC} represented the median of translation rate of CDS under the 1000 varying R_{in} inputs and the other fixed parameters. The the coefficient of variation (CV) of these 1000 N_{EC} reflected the variability of translation rate of CDS under the 1000 varying R_{in} inputs and the other fixed parameters.

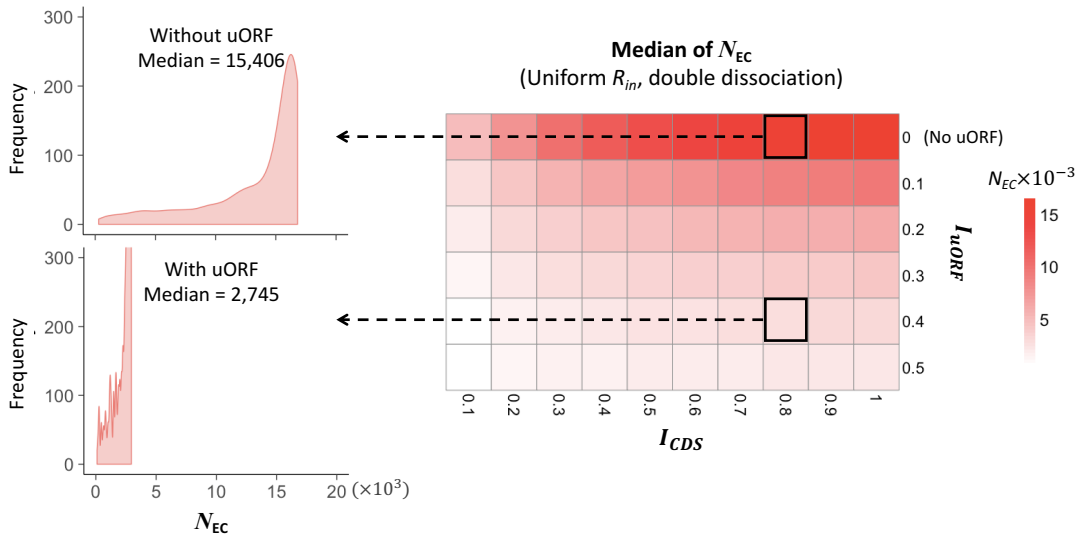
A


Without uORF
Median = 16,039
With uORF
Median = 3,237
Frequency
N_{EC}
(×10³)
Median of N_{EC}
(Uniform R_{in}, downstream dissociation)
0 (No uORF)
I_{uORF}
I_{CDS}
N_{EC}×10⁻³

B


Without uORF
Median = 15,370
With uORF
Median = 3,558
Frequency
N_{EC}
(×10³)
Median of N_{EC}
(Uniform R_{in}, upstream dissociation)
0 (No uORF)
I_{uORF}
I_{CDS}
N_{EC}×10⁻³

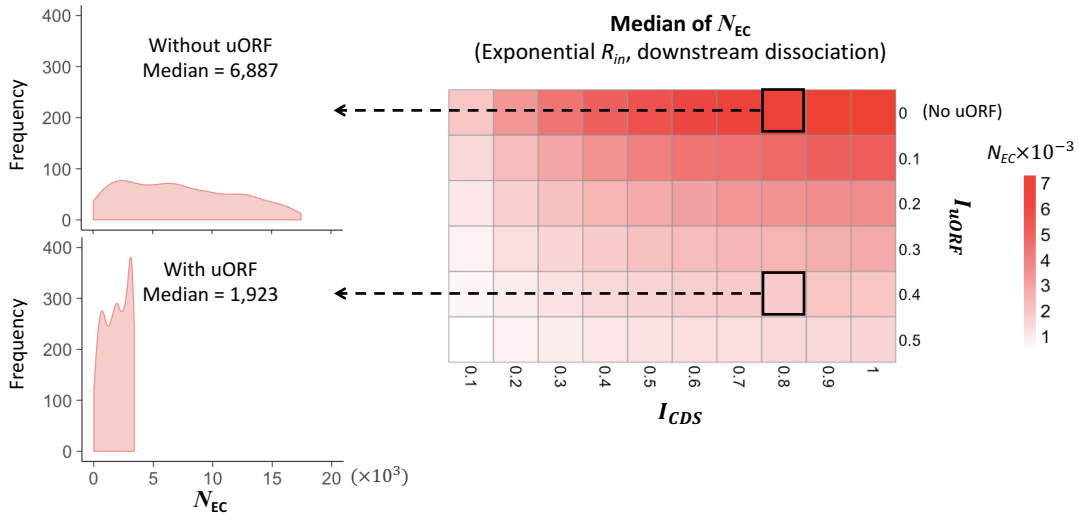
C


Without uORF
Median = 15,406
With uORF
Median = 2,745
Frequency
N_{EC}
(×10³)
Median of N_{EC}
(Uniform R_{in}, double dissociation)
0 (No uORF)
I_{uORF}
I_{CDS}
N_{EC}×10⁻³

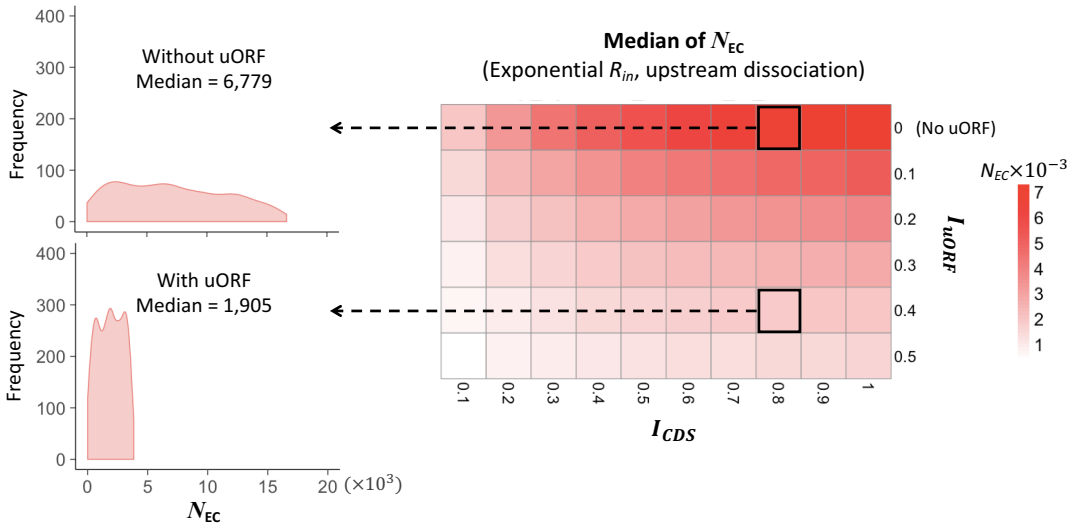
1458

Figure S2. Heatmaps showing the median CDS translation rate (N_{EC}) under different I_{CDS} (x-axis) and I_{uORF} (y-axis) combinations with a uniform distribution of R_{in} input and the downstream dissociation model (A), a uniform distribution of R_{in} input and the upstream dissociation model (B), a uniform distribution of R_{in} input and the double dissociation model (C). For each heatmap, the left panels elicited by the dotted lines from specific squares of right heatmap were two examples showing the distribution of N_{EC} under $I_{CDS} = 0.8$ & $I_{uORF} = 0$ (top panel, without uORF) and $I_{CDS} = 0.8$ & $I_{uORF} = 0.4$ (bottom panel, with uORF).

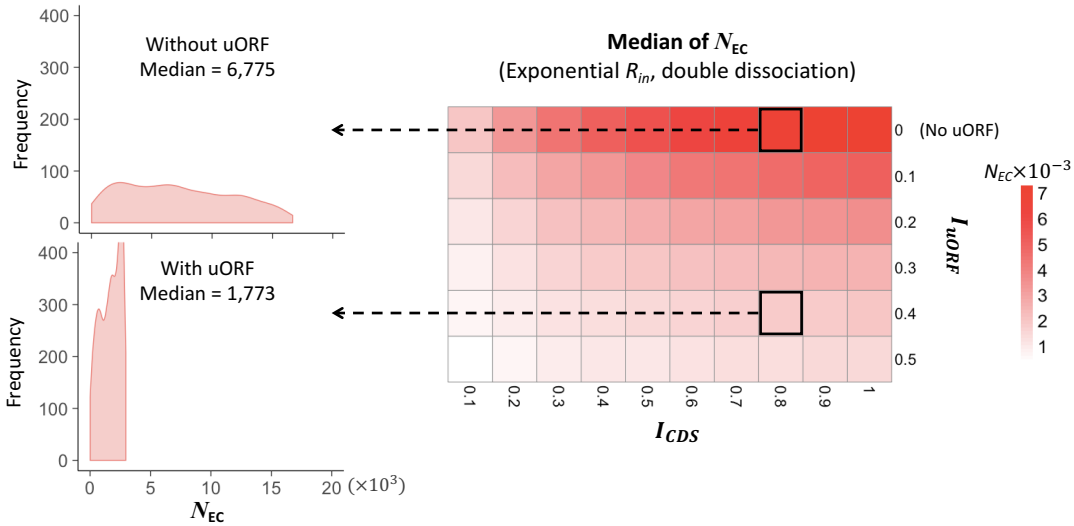
A


Without uORF
Median = 6,887
With uORF
Median = 1,923
Frequency
N_{EC}
(×10³)
Median of N_{EC}
(Exponential R_{in}, downstream dissociation)
0 (No uORF)
I_{uORF}
I_{CDS}
N_{EC}×10⁻³

B


Without uORF
Median = 6,779
With uORF
Median = 1,905
Frequency
N_{EC}
(×10³)
Median of N_{EC}
(Exponential R_{in}, upstream dissociation)
0 (No uORF)
I_{uORF}
I_{CDS}
N_{EC}×10⁻³

C


Without uORF
Median = 6,775
With uORF
Median = 1,773
Frequency
N_{EC}
(×10³)
Median of N_{EC}
(Exponential R_{in}, double dissociation)
0 (No uORF)
I_{uORF}
I_{CDS}
N_{EC}×10⁻³

1467

Figure S3. Heatmaps showing the median CDS translation rate (N_{EC}) under different I_{CDS} (x-axis) and I_{uORF} (y-axis) combinations with an exponential distribution of R_{in} input and the downstream dissociation model (A), an exponential distribution of R_{in} input and the upstream dissociation model (B), an exponential distribution of R_{in} input and the double dissociation model (C). For each heatmap, the left panels elicited by the dotted lines from specific squares of right heatmap were two examples showing the distribution of N_{EC} under I_{CDS} = 0.8 & I_{uORF} = 0 (top panel, without uORF) and I_{CDS} = 0.8 & I_{uORF} = 0.4 (bottom panel, with uORF).

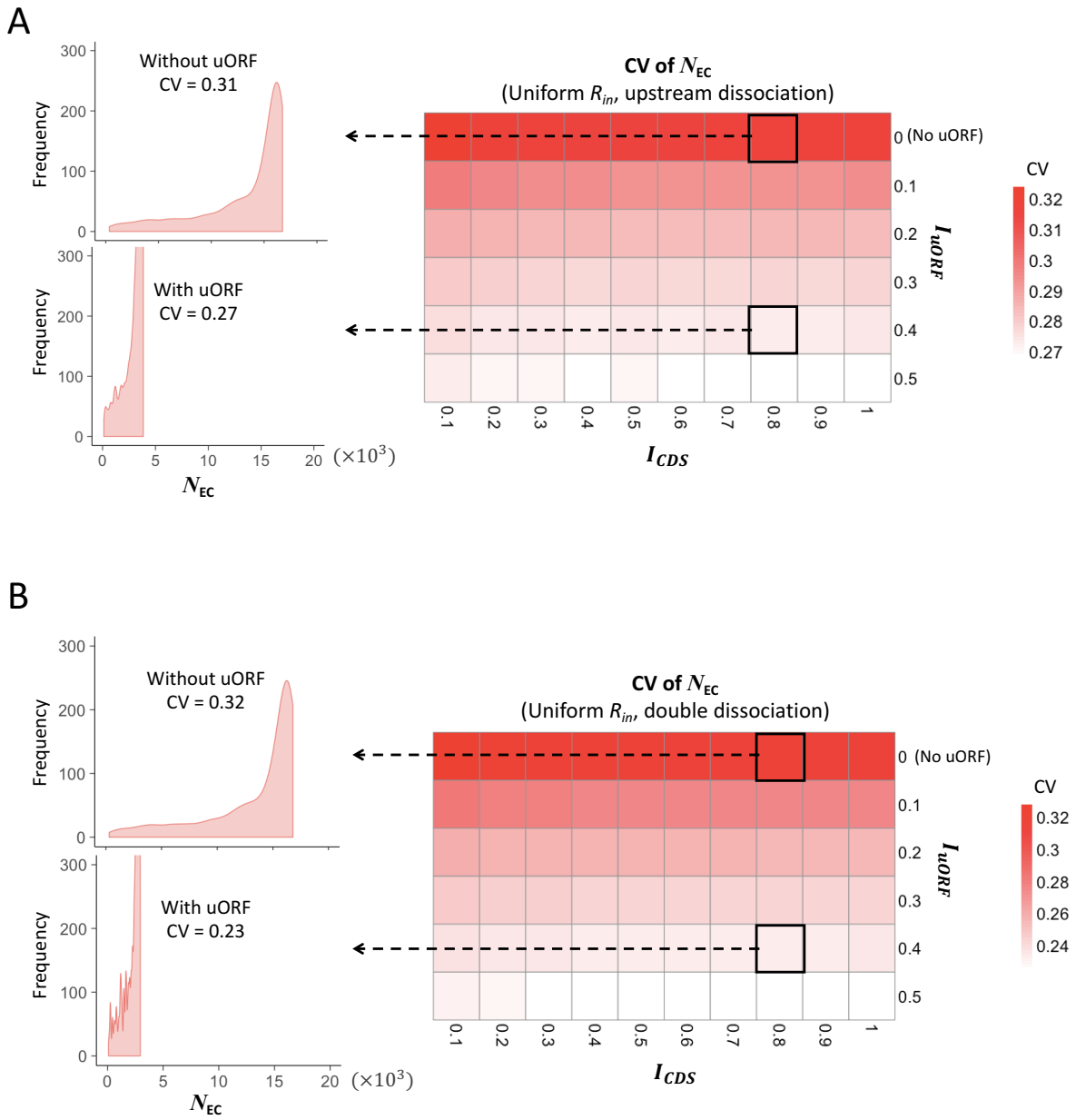
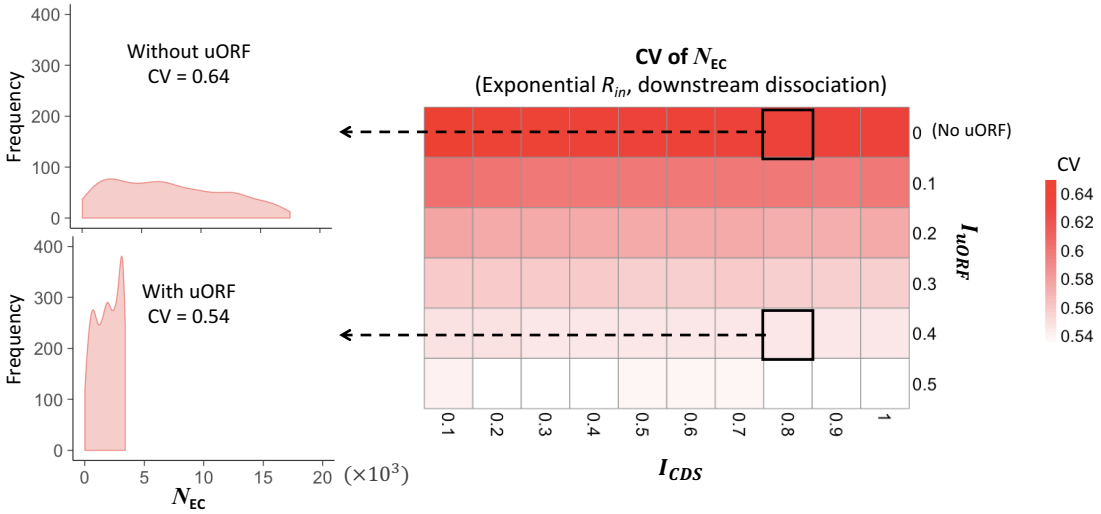
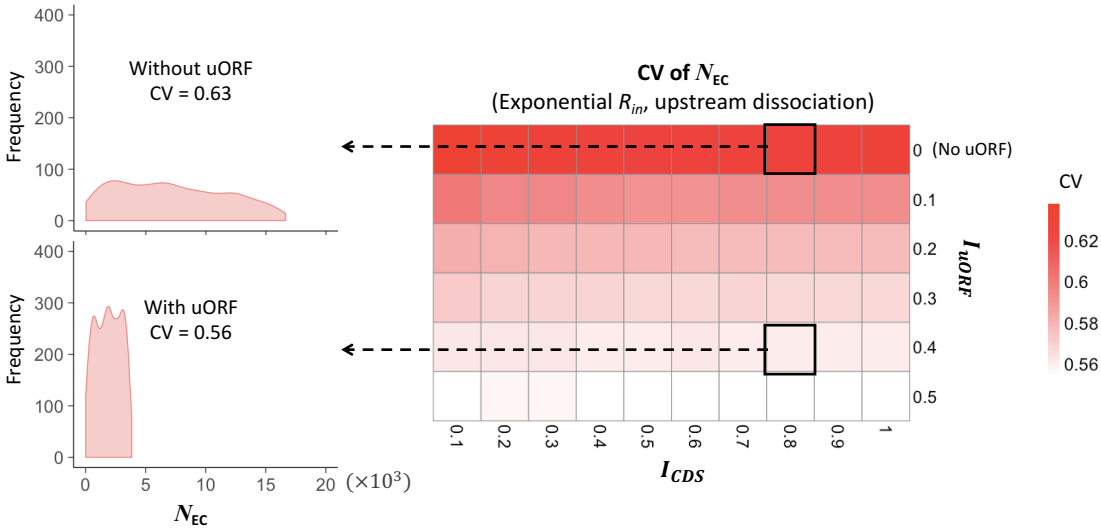


Figure S4. Heatmaps showing the CVs of CDS translation rate (N_{EC}) under different I_{CDS} (x-axis) and I_{uORF} (y-axis) combinations with a uniform distribution of R_{in} input and the upstream dissociation model (A), a uniform distribution of R_{in} input and the double dissociation model (B). For each heatmap, the left panels elicited by the dotted lines from specific squares of right heatmap were two examples showing the distribution of N_{EC} under I_{CDS} = 0.8 & I_{uORF} = 0 (top panel, without uORF) and I_{CDS} = 0.8 & I_{uORF} = 0.4 (bottom panel, with uORF).

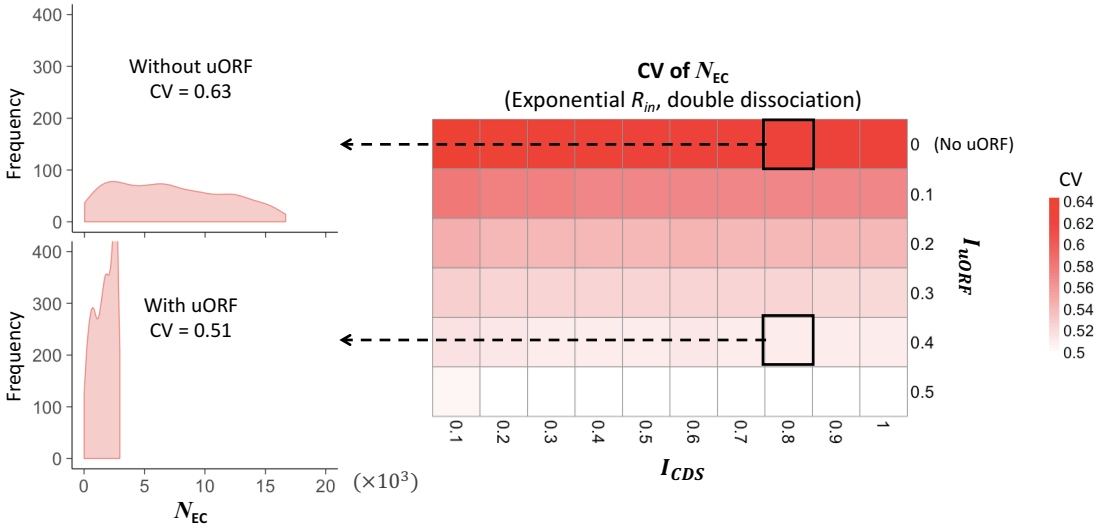
A


Without uORF
CV = 0.64
With uORF
CV = 0.54
Frequency
N_{EC}
(×10³)
CV of N_{EC}
(Exponential R_{in}, downstream dissociation)
0 (No uORF)
I_{uORF}
I_{CDS}
CV

B


Without uORF
CV = 0.63
With uORF
CV = 0.56
Frequency
N_{EC}
(×10³)
CV of N_{EC}
(Exponential R_{in}, upstream dissociation)
0 (No uORF)
I_{uORF}
I_{CDS}
CV

C


Without uORF
CV = 0.63
With uORF
CV = 0.51
Frequency
N_{EC}
(×10³)
CV of N_{EC}
(Exponential R_{in}, double dissociation)
0 (No uORF)
I_{uORF}
I_{CDS}
CV

1485

Figure S5. Heatmaps showing the CVs of CDS translation rate (N_{EC}) under different I_{CDS} (x-axis) and I_{uORF} (y-axis) combinations with an exponential distribution of R_{in} input and the downstream dissociation model (A), an exponential distribution of R_{in} input and the upstream dissociation model (B), an exponential distribution of R_{in} input and the double dissociation model (C). For each heatmap, the left panels elicited by the dotted lines from specific squares of right heatmap were two examples showing the distribution of N_{EC} under I_{CDS} = 0.8 & I_{uORF} = 0 (top panel, without uORF) and I_{CDS} = 0.8 & I_{uORF} = 0.4 (bottom panel, with uORF).

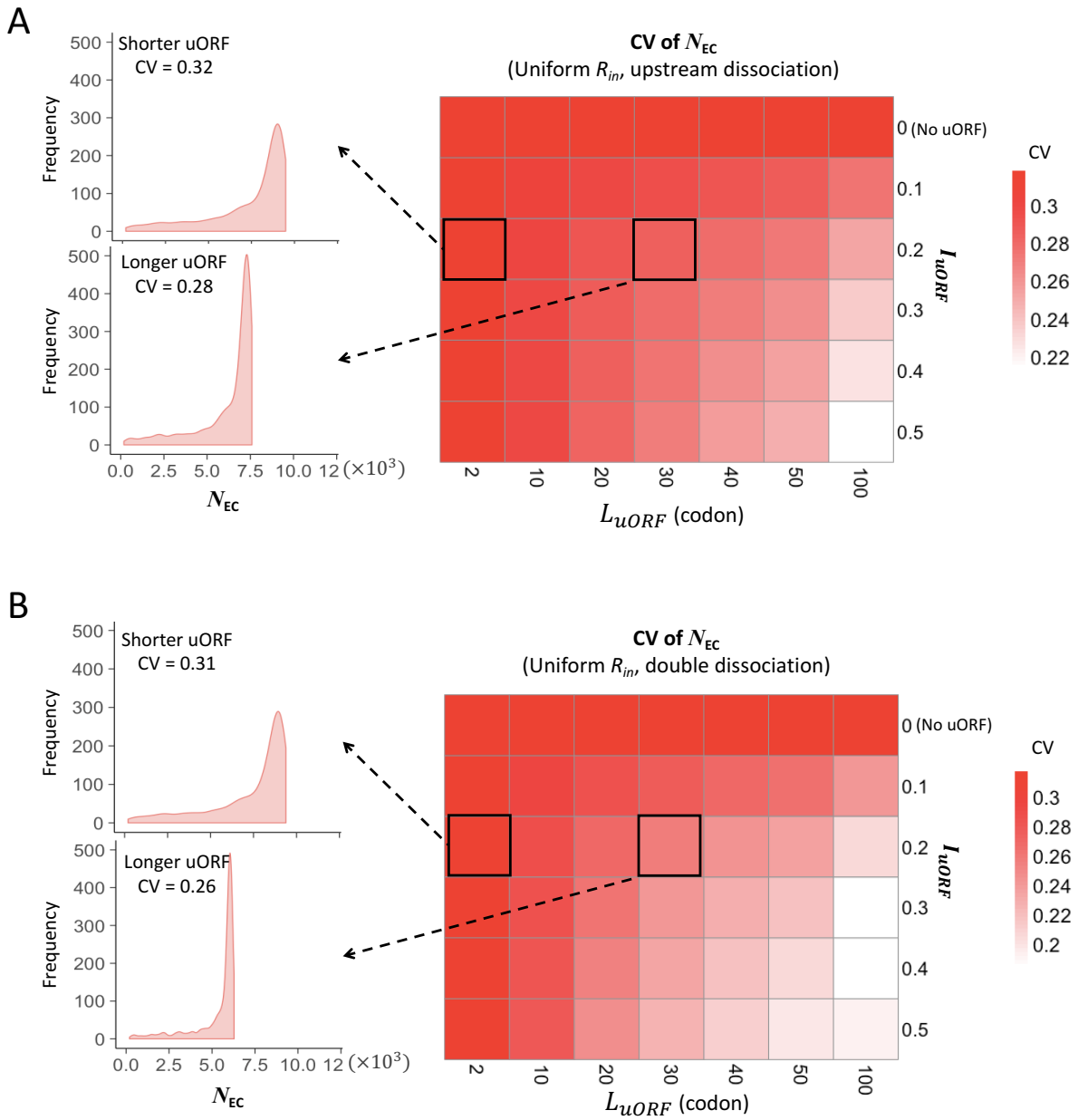
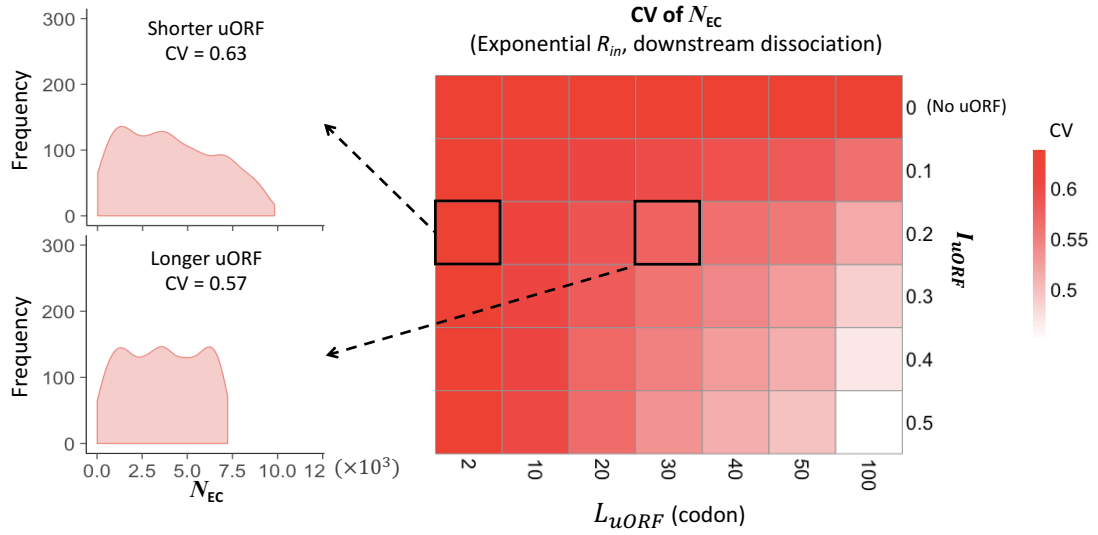
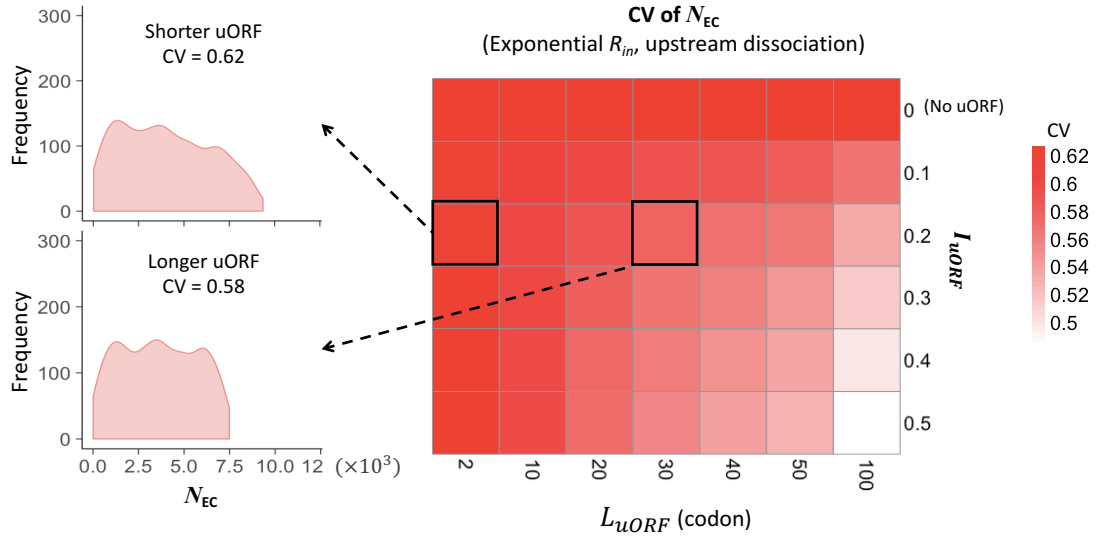


Figure S6. Heatmaps showing the CVs of CDS translation rate (N_{EC}) under different L_{uORF} (x-axis) and I_{uORF} (y-axis) combinations with a uniform distribution of R_{in} input and the upstream dissociation model (A), a uniform distribution of R_{in} input and the double dissociation model (B). For each heatmap, the left panels elicited by the dotted lines from specific squares of right heatmap were two examples showing the distribution of N_{EC} under $L_{uORF} = 2$ & $I_{uORF} = 0.2$ (top panel, shorter uORF) and $L_{uORF} = 30$ & $I_{uORF} = 0.2$ (bottom panel, longer uORF).

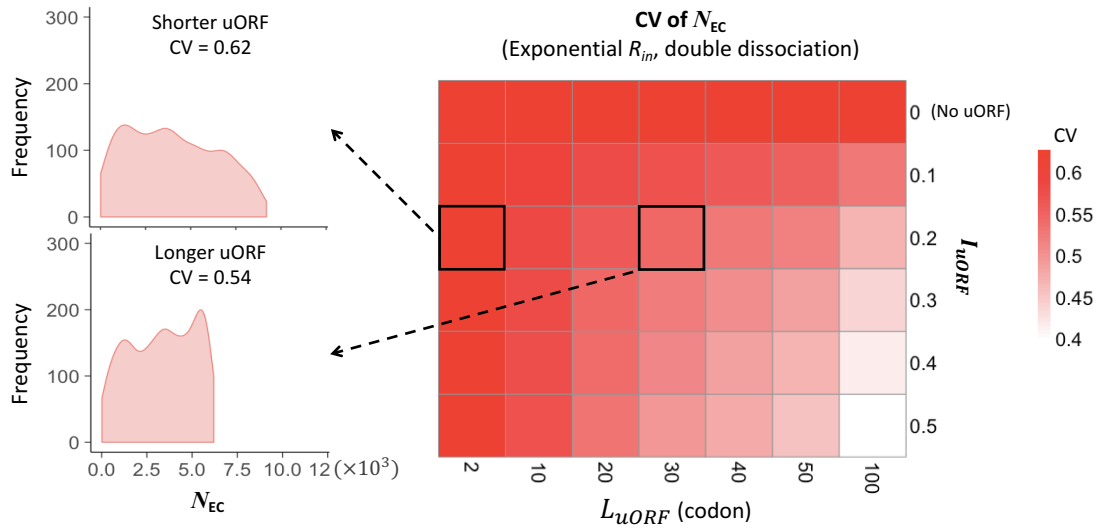
A


Shorter uORF
CV = 0.63
Longer uORF
CV = 0.57
Frequency
300
200
100
0
0.0 2.5 5.0 7.5 10.0 12 (×10³)
N_{EC}
CV of N_{EC}
(Exponential R_{in}, downstream dissociation)
0 (No uORF)
0.1
0.2
0.3
0.4
0.5
I_{uORF}
2 10 20 30 40 50 100
L_{uORF} (codon)
CV
0.6
0.55
0.5

B


Shorter uORF
CV = 0.62
Longer uORF
CV = 0.58
Frequency
300
200
100
0
0.0 2.5 5.0 7.5 10.0 12 (×10³)
N_{EC}
CV of N_{EC}
(Exponential R_{in}, upstream dissociation)
0 (No uORF)
0.1
0.2
0.3
0.4
0.5
I_{uORF}
2 10 20 30 40 50 100
L_{uORF} (codon)
CV
0.62
0.6
0.58
0.56
0.54
0.52
0.5

C


Shorter uORF
CV = 0.62
Longer uORF
CV = 0.54
Frequency
300
200
100
0
0.0 2.5 5.0 7.5 10.0 12 (×10³)
N_{EC}
CV of N_{EC}
(Exponential R_{in}, double dissociation)
0 (No uORF)
0.1
0.2
0.3
0.4
0.5
I_{uORF}
2 10 20 30 40 50 100
L_{uORF} (codon)
CV
0.6
0.55
0.5
0.45
0.4

1502

1503 **Figure S7.** Heatmaps showing the CVs of CDS translation rate (N_{EC}) under different L_{uORF} (x-axis) and
1504 I_{uORF} (y-axis) combinations with an exponential distribution of R_{in} input and the downstream
1505 dissociation model (A), an exponential distribution of R_{in} input and the upstream dissociation model
1506 (B), an exponential distribution of R_{in} input and the double dissociation model (C). For each heatmap,
1507 the left panels elicited by the dotted lines from specific squares of right heatmap were two examples
1508 showing the distribution of N_{EC} under $L_{uORF} = 2$ & $I_{uORF} = 0.2$ (top panel, shorter uORF) and $L_{uORF} = 30$
1509 & $I_{uORF} = 0.2$ (bottom panel, longer uORF).
1510

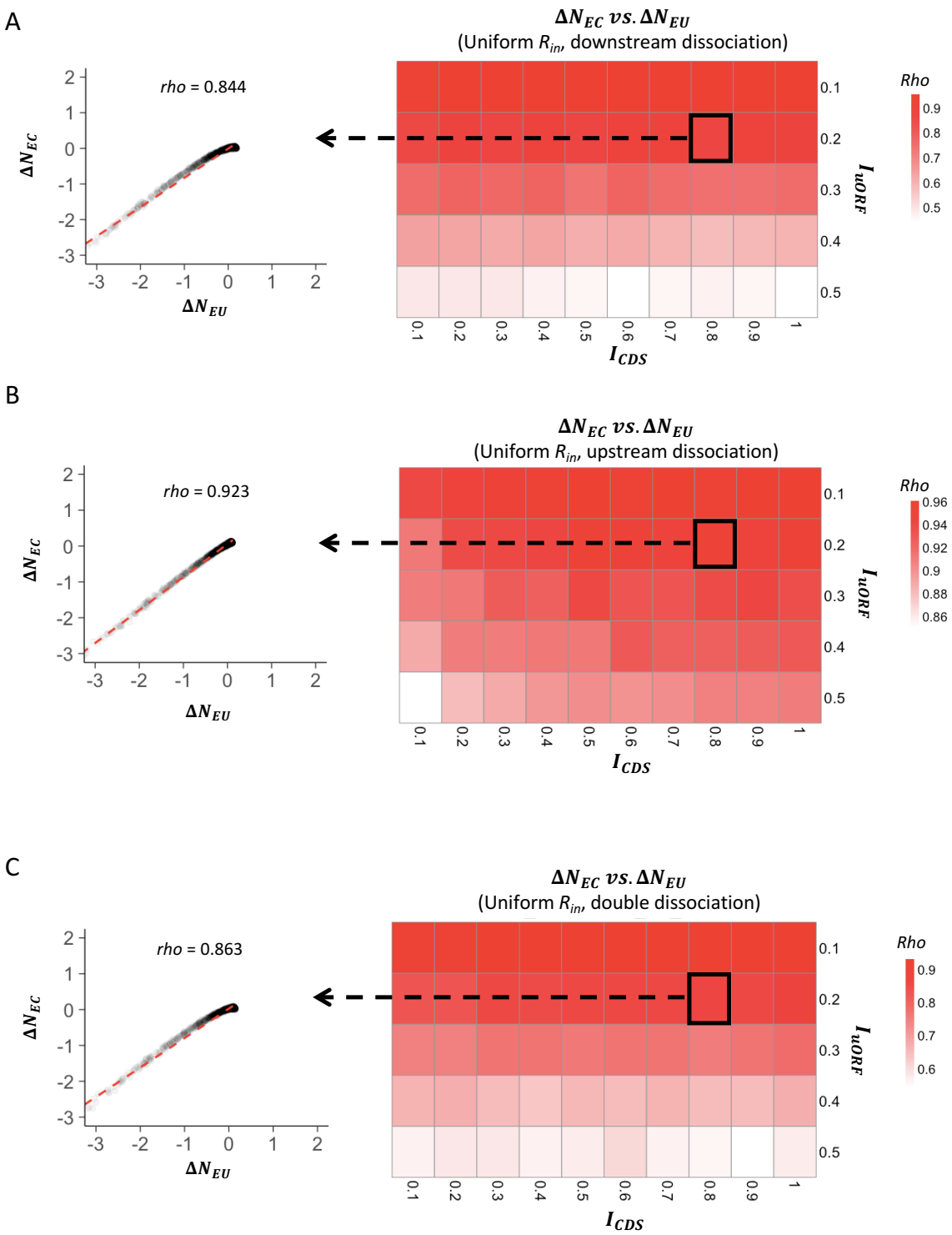


Figure S8. Heatmaps showing the Spearman's correlations between changes of uORF translation rate (ΔN_{EU}) and downstream CDS translation rate (ΔN_{EC}) under different I_{CDS} (x-axis) and I_{uORF} (y-axis) combinations with a uniform distribution of R_{in} input and the downstream dissociation model (A), a uniform distribution of R_{in} input and the upstream dissociation model (B), a uniform distribution of R_{in} input and the double dissociation model (C). For each heatmap, the left panel elicited by the dotted line from the specific square of right heatmap was an example showing the Spearman's correlations between changes of uORF translation rate (ΔN_{EU}) and downstream CDS translation rate (ΔN_{EC}) under I_{CDS} = 0.8 & I_{uORF} = 0.2. All P values < 0.001.

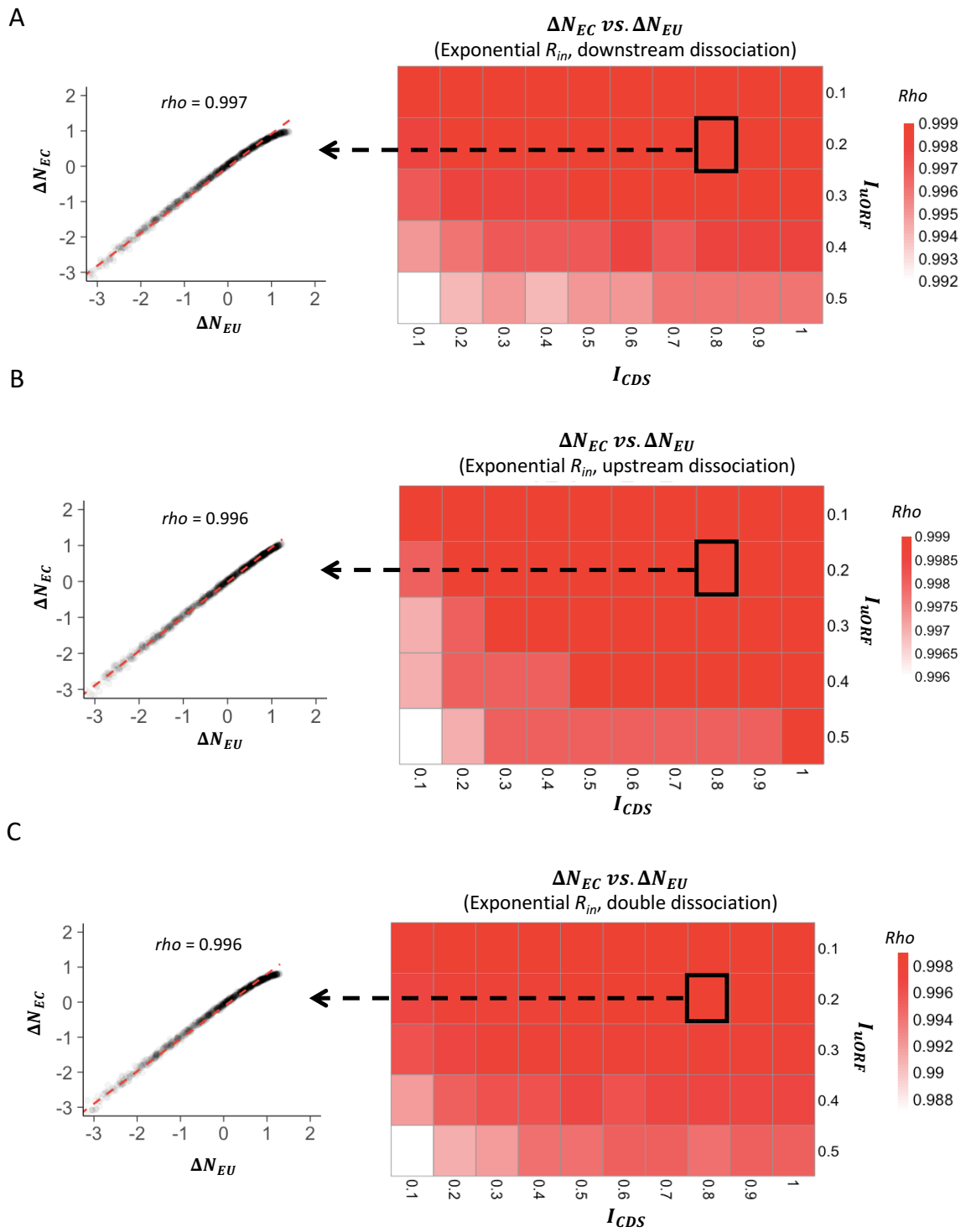


Figure S9. Heatmaps showing the Spearman's correlations between changes of uORF translation rate (ΔN_{EU}) and downstream CDS translation rate (ΔN_{EC}) under different I_{CDS} (x-axis) and I_{uORF} (y-axis) combinations with an exponential distribution of R_{in} input and the downstream dissociation model (A), an exponential distribution of R_{in} input and the upstream dissociation model (B), an exponential distribution of R_{in} input and the double dissociation model (C). For each heatmap, the left panel elicited by the dotted line from the specific square of right heatmap was an example showing the Spearman's correlations between changes of uORF translation rate (ΔN_{EU}) and downstream CDS translation rate (ΔN_{EC}) under $I_{CDS} = 0.8$ & $I_{uORF} = 0.2$. All P values < 0.001.

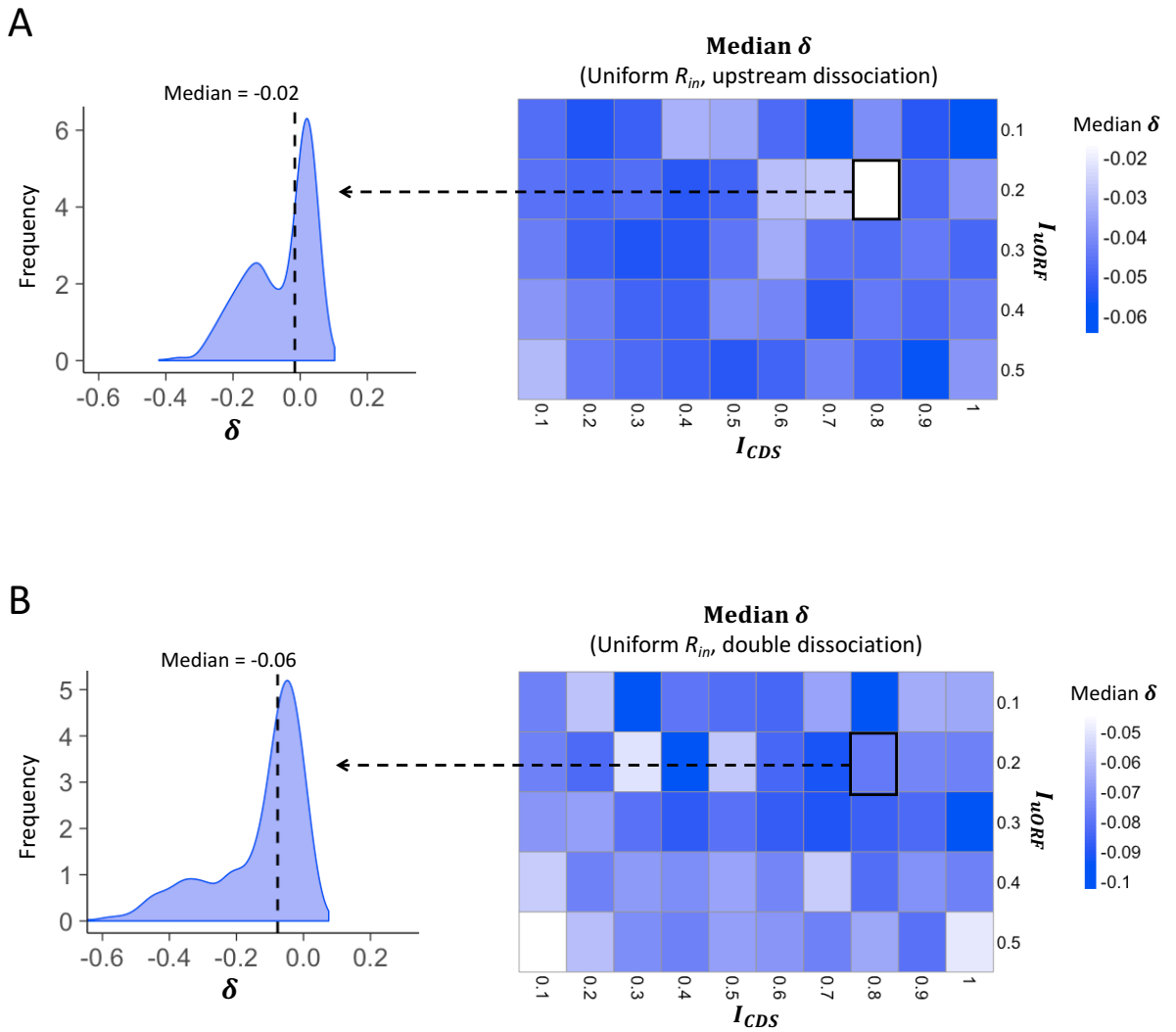


Figure S10. Heatmaps showing the median δ under different I_{CDS} (x-axis) and I_{uORF} (y-axis) combinations with a uniform distribution of R_{in} input and the upstream dissociation model (A), a uniform distribution of R_{in} input, and the double dissociation model (B). For each heatmap, the left panel elicited by the dotted line from the specific square of right heatmap was an example showing the distribution of δ under I_{CDS} = 0.8 & I_{uORF} = 0.2. The vertical dashed line indicated the median value of δ.

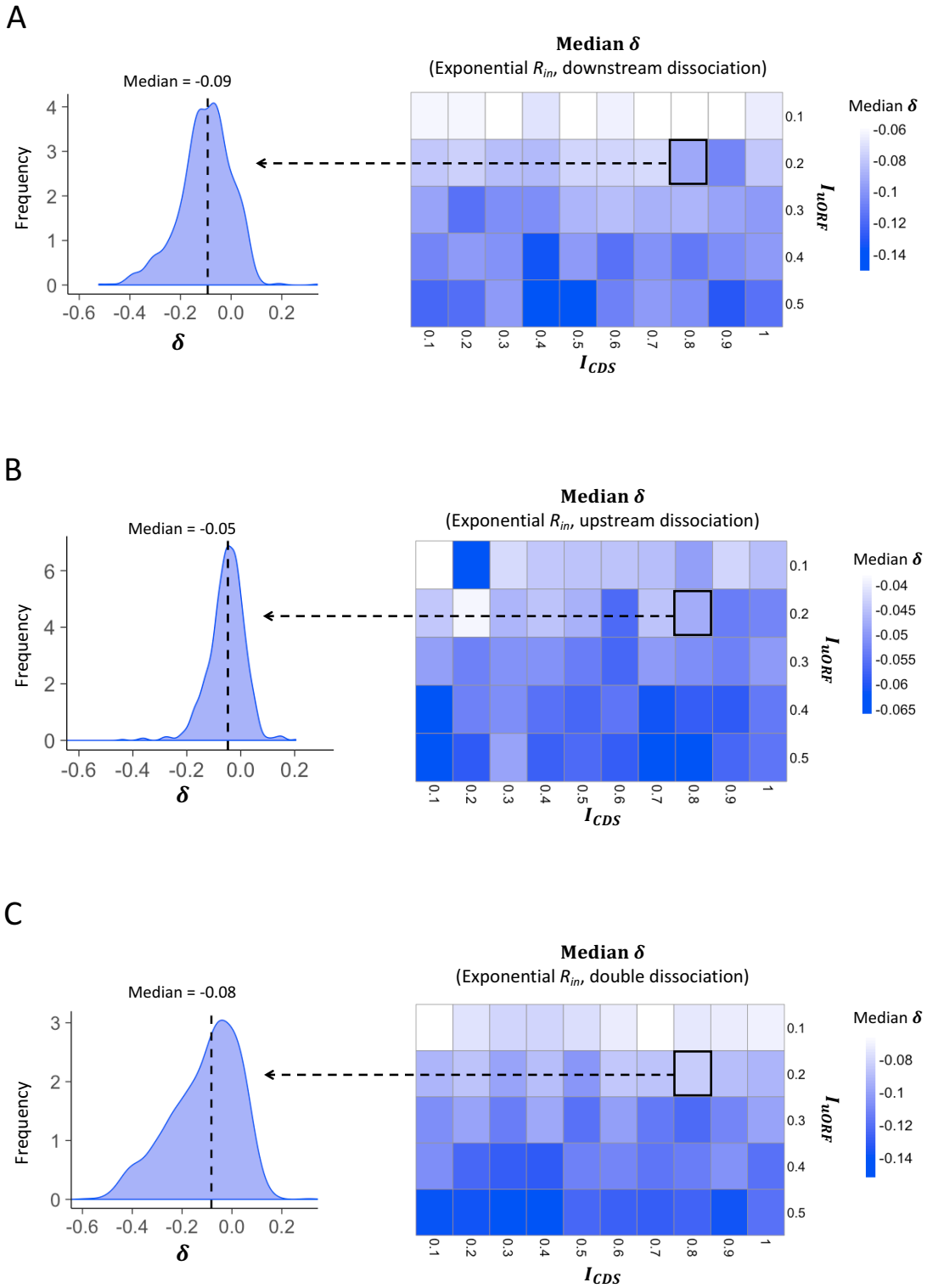


Figure S11. Heatmaps showing the median δ under different I_{CDS} (x-axis) and I_{uORF} (y-axis) combinations with an exponential distribution of R_{in} input and the downstream dissociation model (A), an exponential distribution of R_{in} input and the upstream dissociation model (B), an exponential distribution of R_{in} input and the double dissociation model (C). For each heatmap, the left panel elicited by the dotted line from the specific square of right heatmap was an example showing the distribution of δ under $I_{CDS} = 0.8$ & $I_{uORF} = 0.2$. The vertical dashed line indicated the median value of δ.

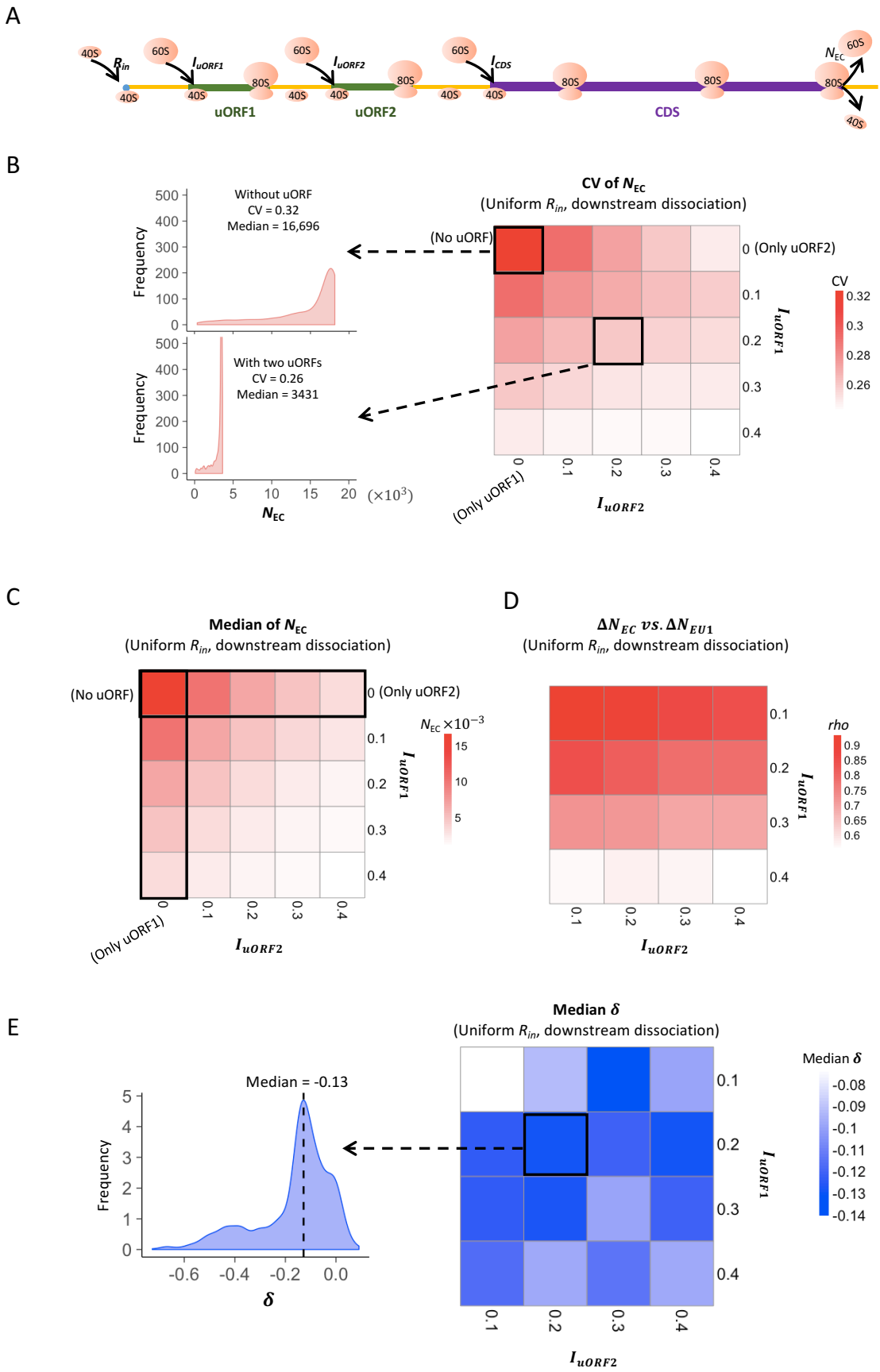


Figure S12. Two-uORF model simulation measurements. (A) The two-uORF model schema. The factors and parameters were the same as those illustrated in Fig. 1A except where specifically indicated.

(B) Heatmap showing the CVs of CDS translation rate (N_{EC}) under different I_{uORF2} (x-axis) and I_{uORF1} (y-axis) combinations with a uniform distribution of R_{in} input and the downstream dissociation model. The left panels elicited by the dotted lines from specific squares of right heatmap were two examples showing the distribution of N_{EC} under $I_{uORF1} = 0$ & $I_{uORF2} = 0$ (top panel, without uORF) and $I_{uORF1} = 0.2$ & $I_{uORF2} = 0.2$ (bottom panel, with two uORFs). (C) Heatmap showing the median CDS translation rate (N_{EC}) under different I_{uORF2} (x-axis) and I_{uORF1} (y-axis) combinations with a uniform distribution of R_{in} input and the downstream dissociation model. (D) Spearman's correlations between changes of uORF1 translation rate (ΔN_{EU1}) and downstream CDS translation rate (ΔN_{EC}) under different I_{uORF2} (x-axis) and I_{uORF1} (y-axis) combinations with a uniform distribution of R_{in} input and the downstream dissociation model. (E) Heatmap showing the median δ $[log_2(\Delta N_{EC}/\Delta N_{EU1})]$ under different I_{uORF2} (x-axis) and I_{uORF1} (y-axis) combinations with a uniform distribution of R_{in} input and the downstream dissociation model. The left panel elicited by the dotted line from the specific square of right heatmap was an example showing the distribution of δ under $I_{uORF1} = 0.2$ & $I_{uORF2} = 0.2$. The vertical dashed line indicated the median value of δ.

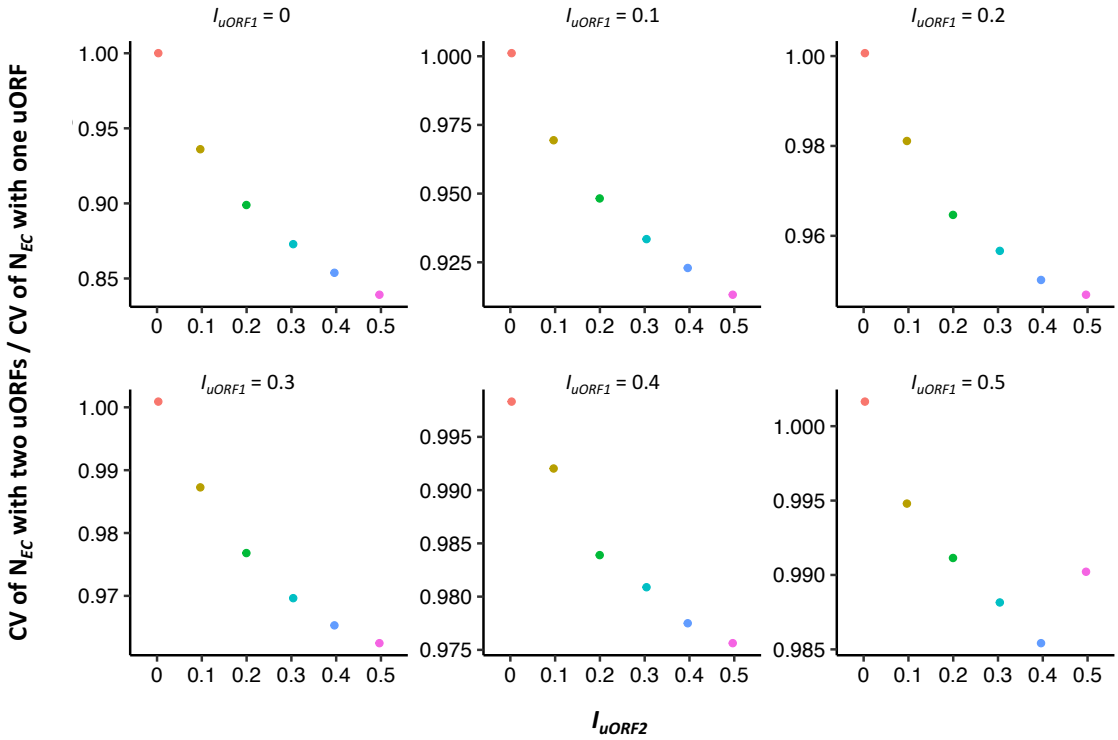


Figure S13. Comparison of the buffering effects between single uORF and two uORFs. In each panel, the I_{uORF} of single-uORF model equals to $I_{uORF1}$of two-uORF model, with both values ranging from 0 to 0.5. The x-axis in each panel denotes the values of different I_{uORF2} in the two-uORF model, ranging from 0 to 0.5. The y-axis in each panel represents the ratio of the CV of N_{EC} with two uORFs to that with a single uORF.

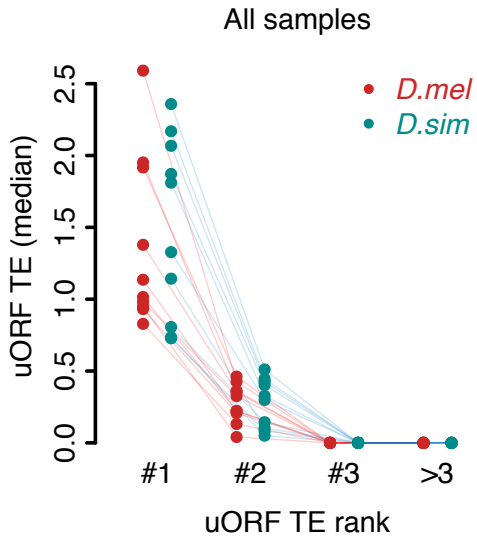
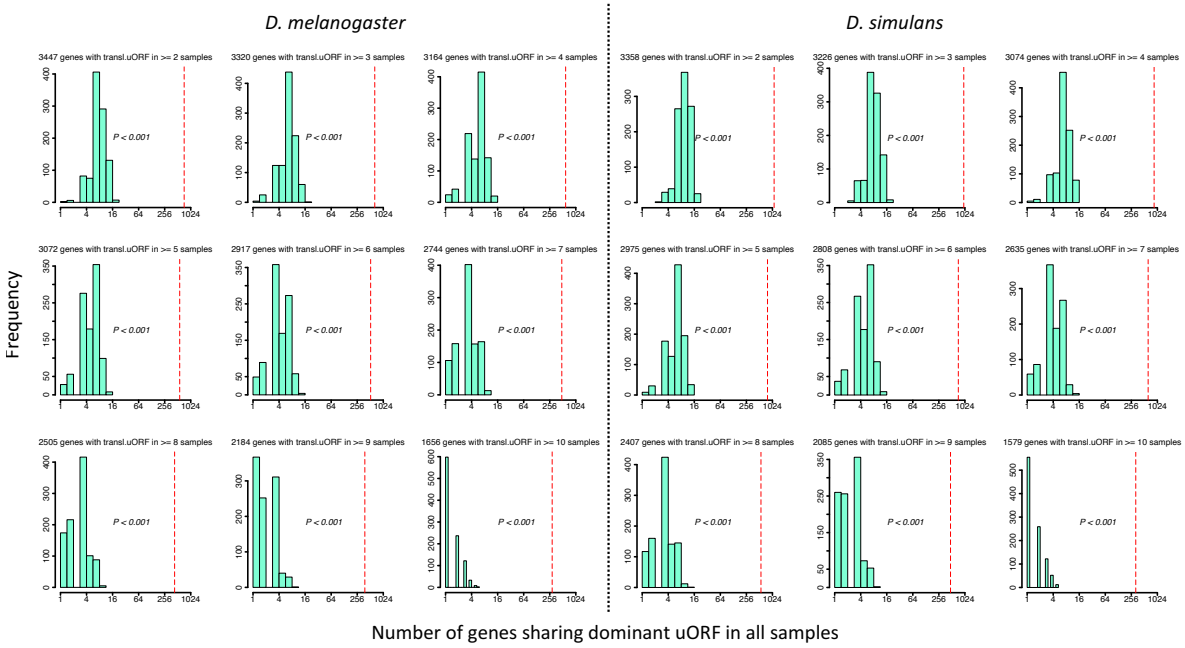


Figure S14. **The dominant uORF showed highest TE than other uORFs with a same gene.** uORFs were ranked by decreasing TEs within each gene. The uORF with the highest TE within each gene was defined as the dominant uORF (#1). "#2" represents the second highest uORF TE and the same goes for "#3" and ">3". Each dot represents the median TE of a sample.

1580



1581
1582 **Figure S15**. **Observed and expected numbers of genes sharing dominant uORFs in all samples.**
1583 Among the genes of *D. melanogaster* (where only the longest transcripts were considered), 7,259
1584 (52.2%) genes had no uORF ("no-uORF" genes), 2,687 (19.3%) genes had one uORF ("one-uORF"
1585 genes), and 3,961 (28.5%) genes had multiple (≥ 2) uORFs ("multiple-uORF" genes). The numbers in
1586 the header of each panel represent genes with translated uORFs (TE > 0.1) in ≥ N samples (2 ≤ N ≤ 10).
1587 The TEs of all uORFs were shuffled 1,000 times. *P* values were obtained from randomization tests.
1588

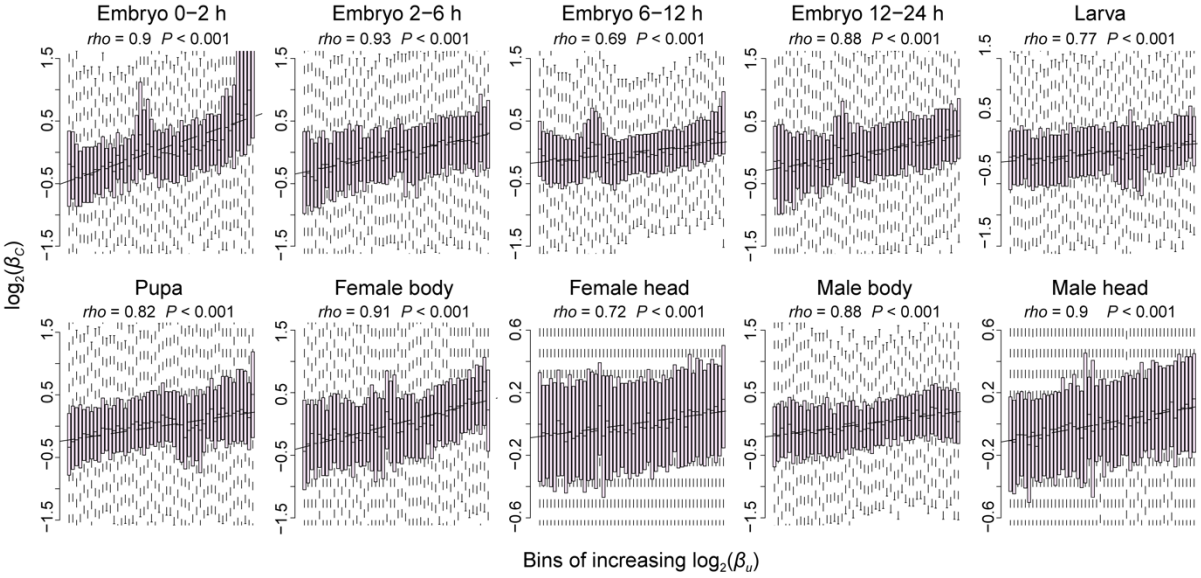


Figure S16.The positive correlation of interspecific TE changes between uORFs and CDSs. Correlations between interspecific uORF TE changes ($\log_2\beta_u$) and CDS TE changes ($\log_2\beta_C$) in 10 samples. The x-axis was divided into 50 equal bins with increasing β_u. Spearman's correlation coefficients (*rho*) are shown at the top left. ***, $P < 0.001$ in the correlation test.

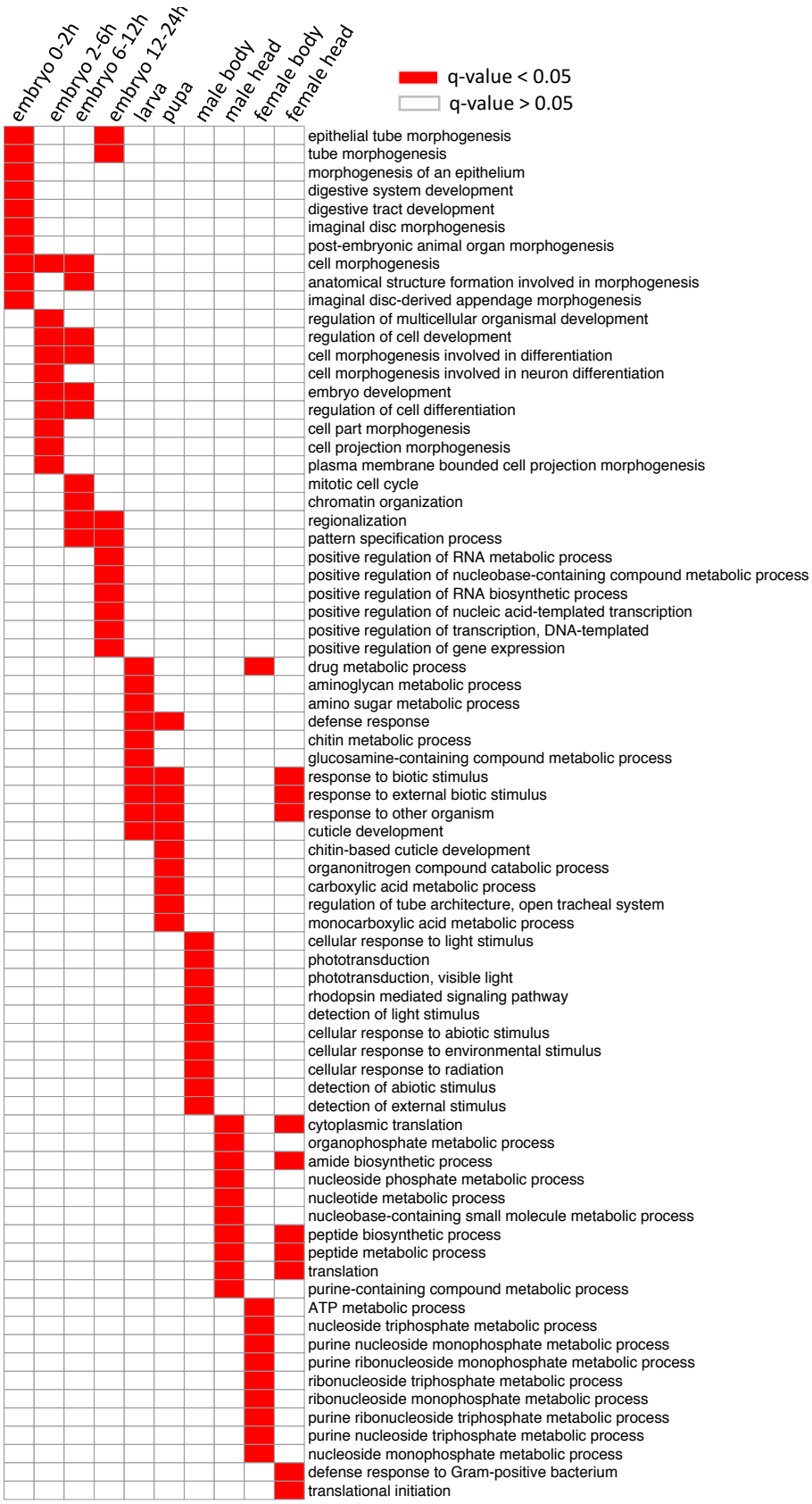


Figure S17. **Gene ontology analysis of the genes with $log_2(\beta_c) \neq 0$ in each stage and tissue.** The biological process (BP) terms with q-values < 0.05 in each sample type are indicated with red, and others are indicated with white.

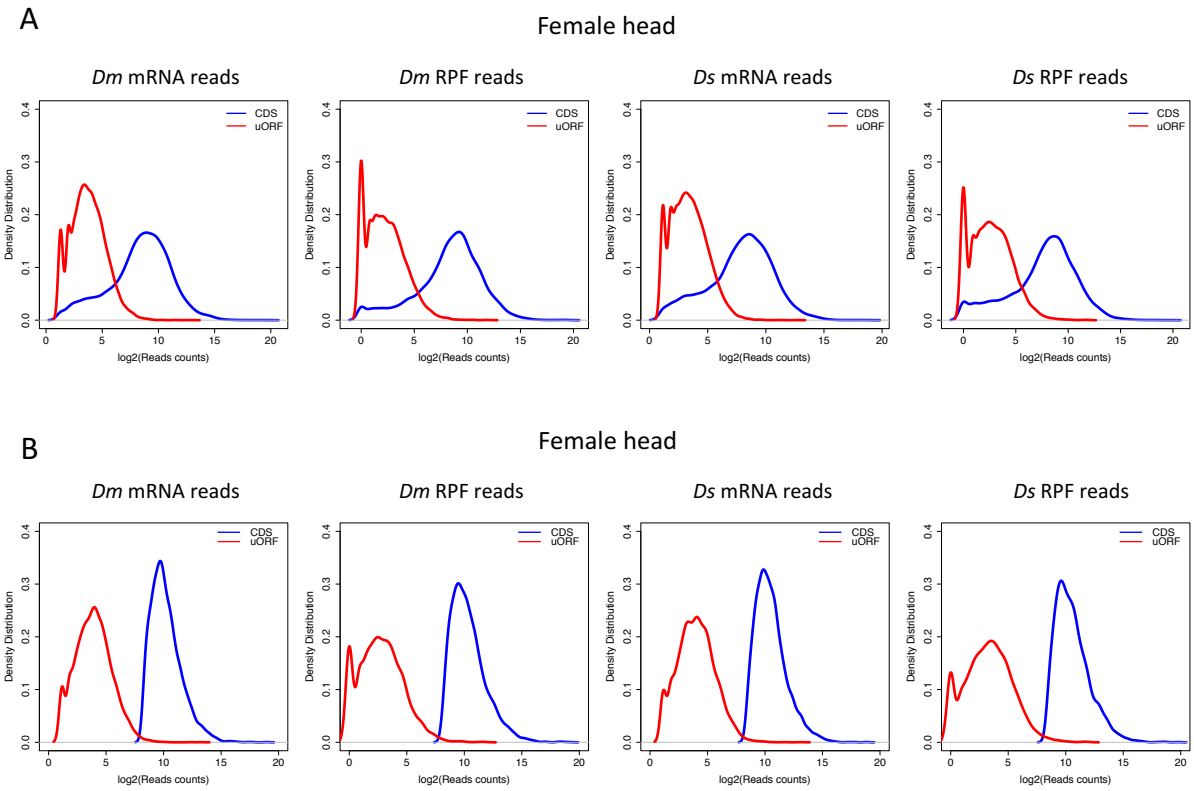


Figure S18. Reads count distribution of uORFs and CDSs. Distribution of mRNA reads counts and RPF counts mapped to uORFs and CDSs for all expressed uORFs (A) or only highly expressed genes (B) in female head sample. The distribution patterns were similar across other samples, and the data are not shown here. *Dm*, *D. melanogaster*; *Ds*, *D. simulans*. The reads counts were $\log_2$ transformed.

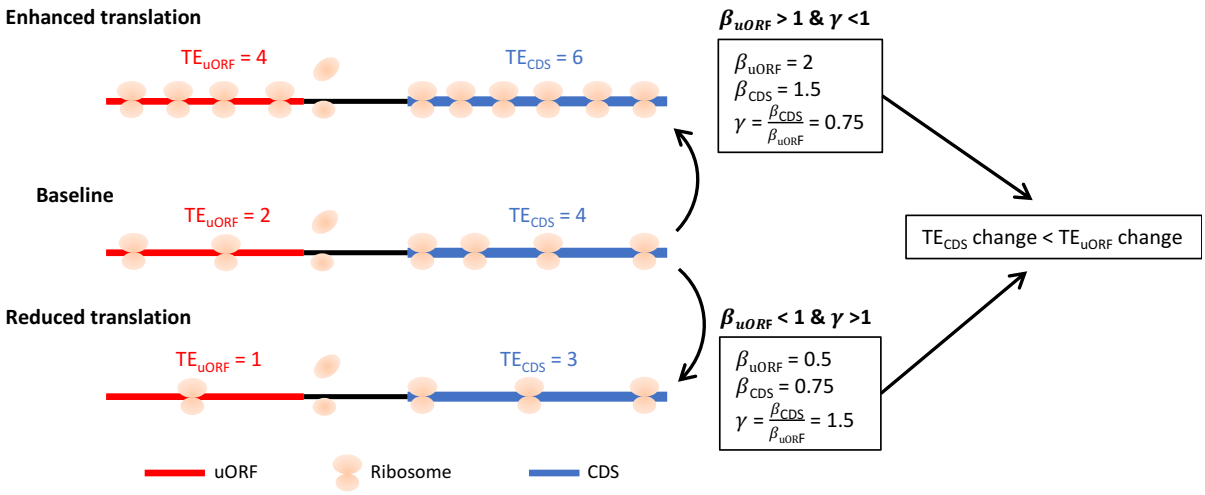


Figure S19. **The scheme illustrating the calculation of β_u, β_c and γ.** When the cellular environment causes mRNA translation to be enhanced from baseline level, the TE_{uORF} changed from 2 to 4 (β_u = 2), while the TE_{uORF} changed with a smaller degree due to uORF's buffering, from 4 to 6 (β_C = 1.5). This resulted in γ <1. Conversely, When the cellular environment causes mRNA translation to be reduced from the baseline level, the TE_{uORF} changed from 2 to 1 (β_u = 0.5), while the TE_{uORF} changed with a smaller degree due to uORF's buffering, from 4 to 3 (β_C = 0.76). This resulted in γ >1. Overall, both β_u >1 &γ <1, and β_u <1 &γ >1, indicated the existence of uORFs' buffering.

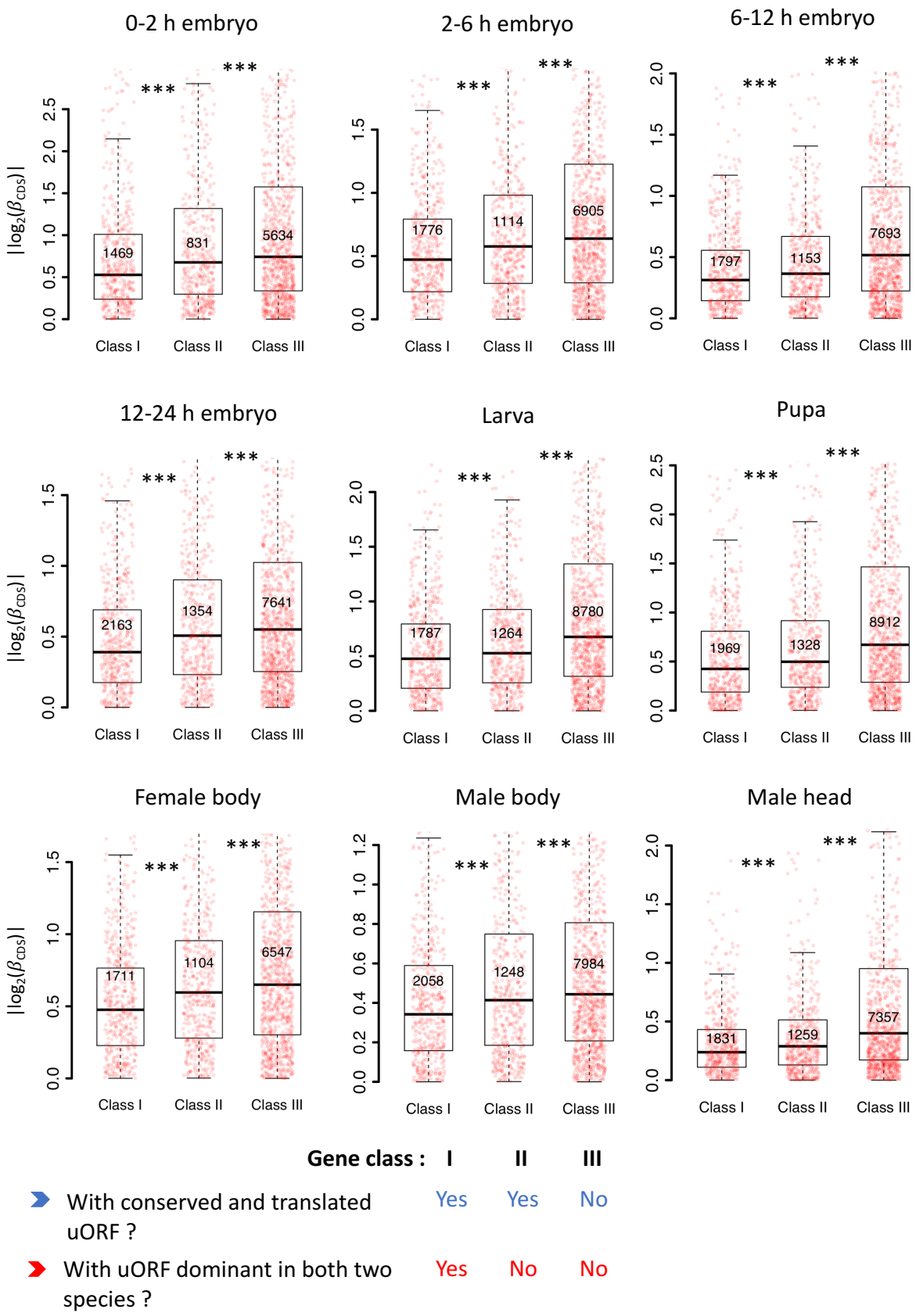


Figure S20. Conserved and dominantly translated uORFs showed the stronger buffering effect. Genes expressed in different stages/tissues (mRNA RPKM > 0.1 in both species) were classified into three classes according to whether a gene had a conserved and dominantly translated uORF or not.

1622 Boxplots showing interspecific CDS TE variability $|log_2(\beta_c)|$ of different gene classes. *P* values were
1623 calculated using Wilcoxon rank-sum tests between the neighboring groups. ***, $P < 0.001$.
1624

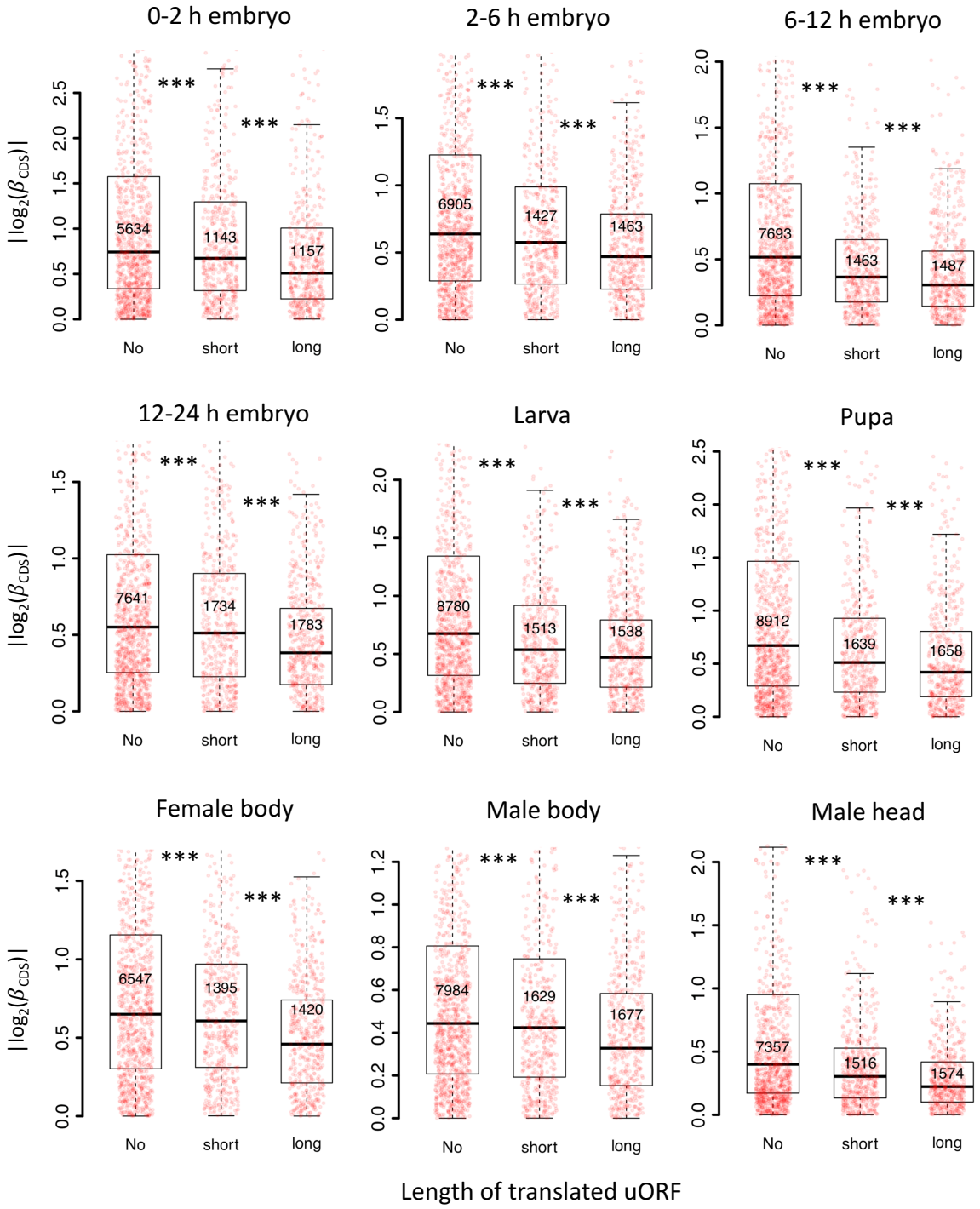


Figure S21. Longer uORFs showed the stronger buffering effect. Genes expressed in different stages/tissues (mRNA RPKM > 0.1 in both species) were classified into three classes according to the length of translated uORFs. Boxplots showing interspecific CDS TE variability $|log_2(\beta_c)|$ of different gene classes. *P* values were calculated using Wilcoxon rank-sum tests between the neighboring groups. ***, $P < 0.001$.

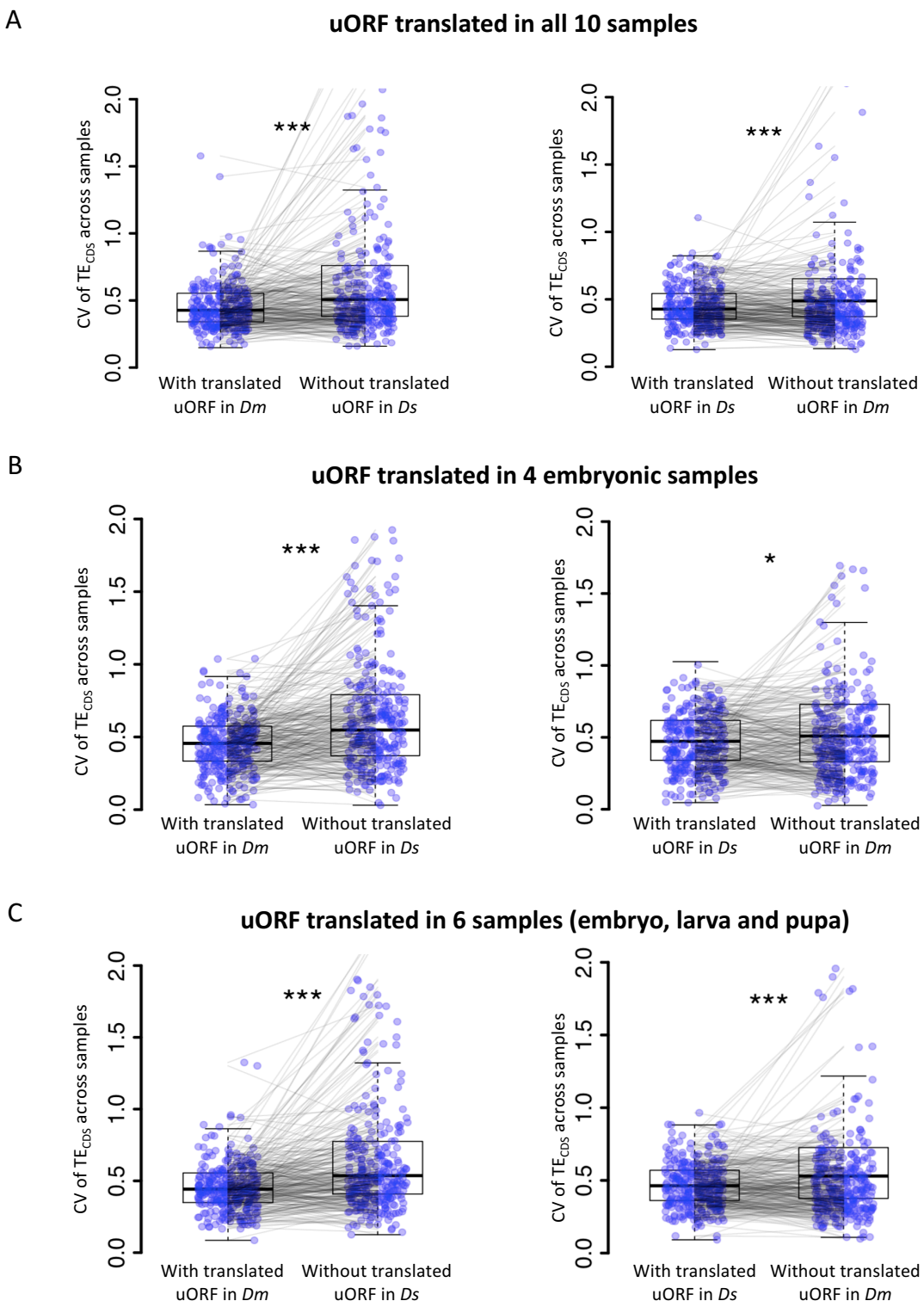
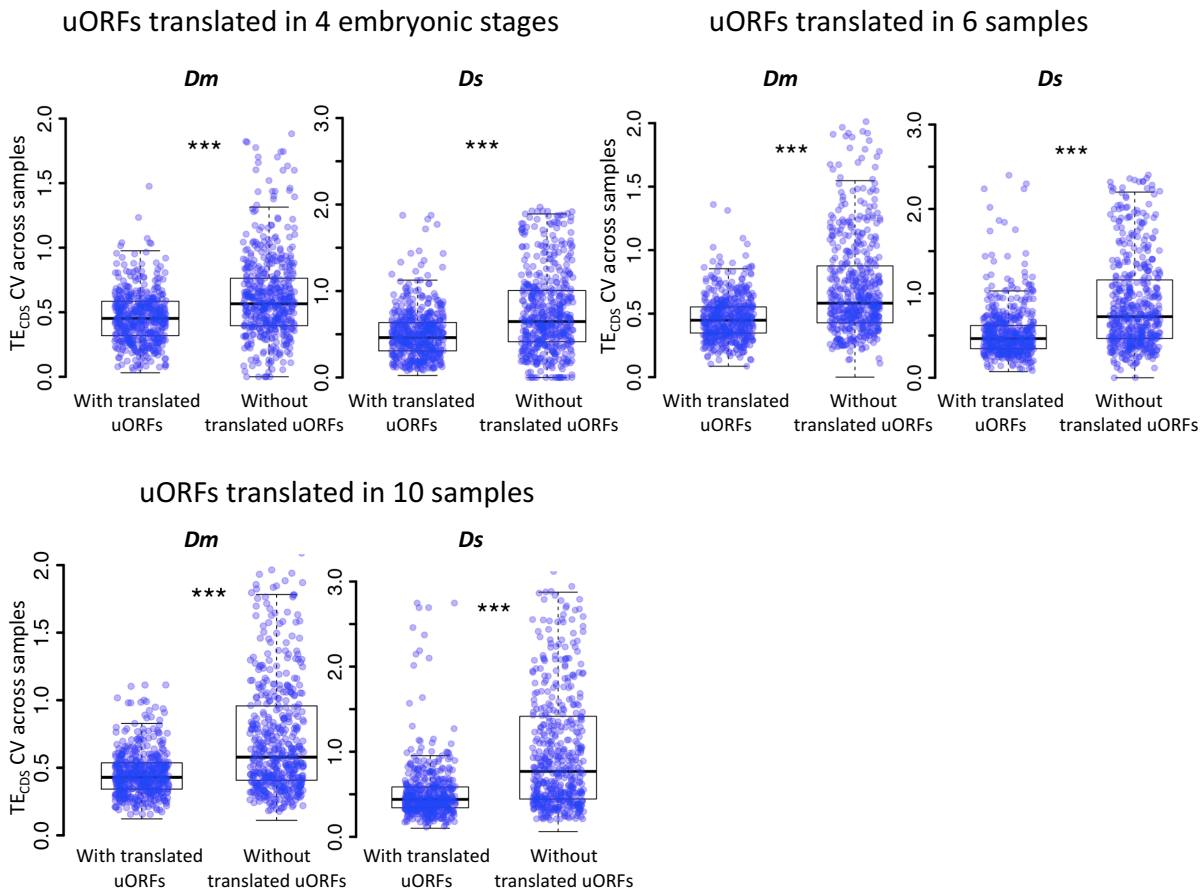


Figure S22. **uORFs reduce CDS translational fluctuation during *Drosophila* development under different cutoffs on defining "translated uORFs".** The CV of TE_{CDS} across *Dm* (*D. melanogaster*) samples and *Ds* (*D. simulans*) samples was shown as boxplots. The selected genes harbor the translated uORFs in one of *Drosophila* species but its homologous gene without translated uORFs in another *Drosophila* species. Each pair of dots linked by a gray line represents a pair of homologous genes in *Dm* and *Ds*. (A) Translated uORF was defined as uORF translated (TE > 0.1) in all 10 samples. (B) Translated uORF was defined as uORF translated (TE > 0.1) in 4 embryonic stages. The CV of TE_{CDS} across these 4 stages was calculated. (C) Translated uORF was defined as uORF translated (TE > 0.1)

1641 in the 6 developmental stages including 4 embryonic stages, larva, and pupa. The CV of TE_{CDS} across
1642 these 6 stages was calculated. Wilcoxon signed-rank sum test. *, $P < 0.05$; ***, $P < 0.001$.

1643



1644
1645 **Figure S23**. **uORFs could reduce CDS translational variation during *Drosophila* development**
1646 **under different cutoffs on defining "translated uORFs".** Within each species, the CV of TE_{CDS}
1647 across different samples was measured for genes with translated uORFs and genes without translated
1648 uORFs, respectively. Each dot in the boxplot represents one gene. Translated uORF was defined as
1649 uORF translated (TE > 0.1) in at least 4 embryonic stages, six developmental stages (left) or translated
1650 or in all 10 samples. Wilcoxon rank-sum tests. ***, $P < 0.001$. *Dm*, *D. melanogaster*; *Ds*, *D. simulans*.
1651

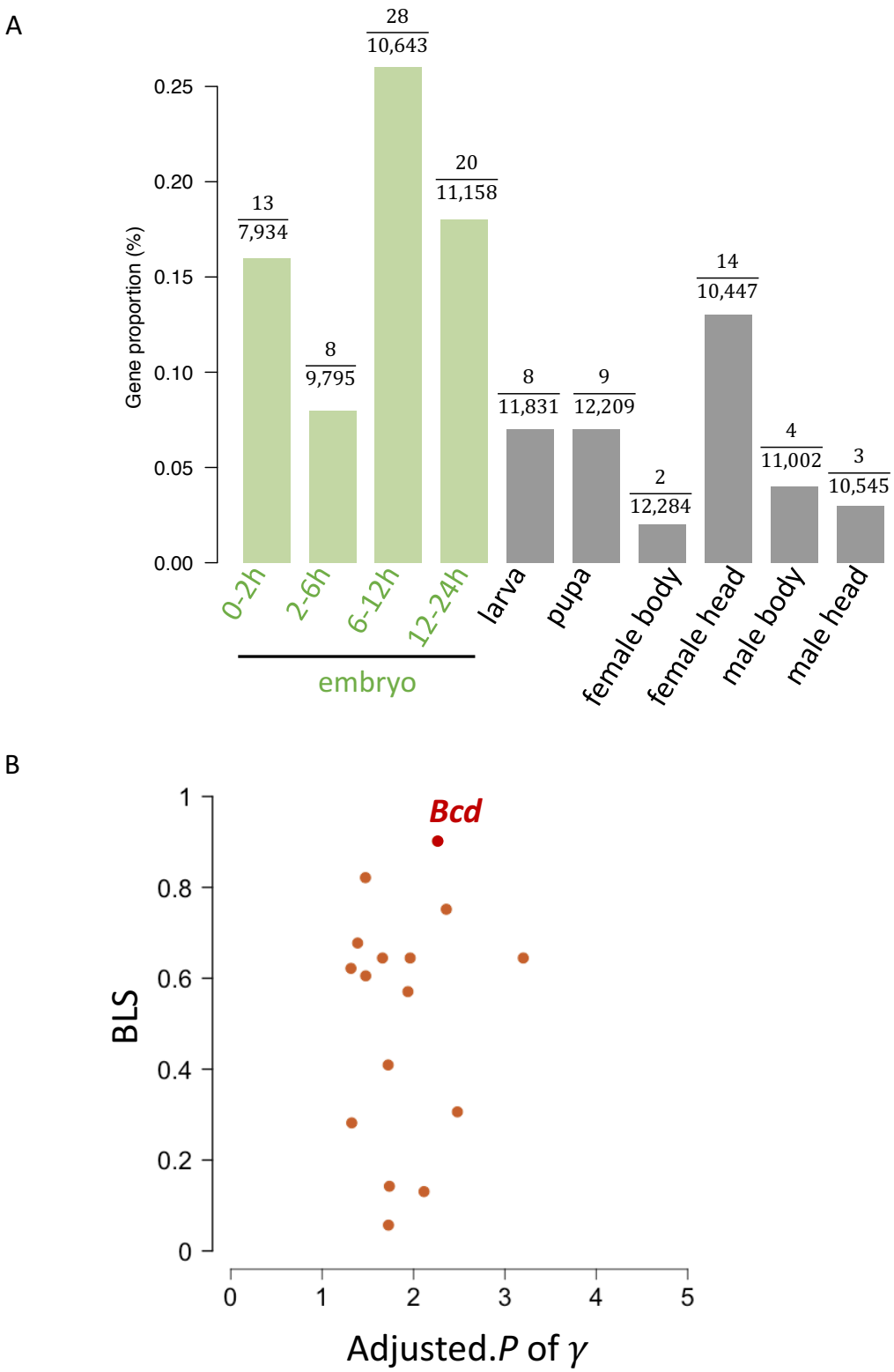


Figure S24. (A) The proportion (%) of genes with buffering uORFs in all expressed genes across developmental stages/tissues. (B) The distribution of adjusted. *P* of γ (x-axis) and BLS (y-axis) of dominant uORFs in embryonic stages. The *bcd* was highlighted by red color.

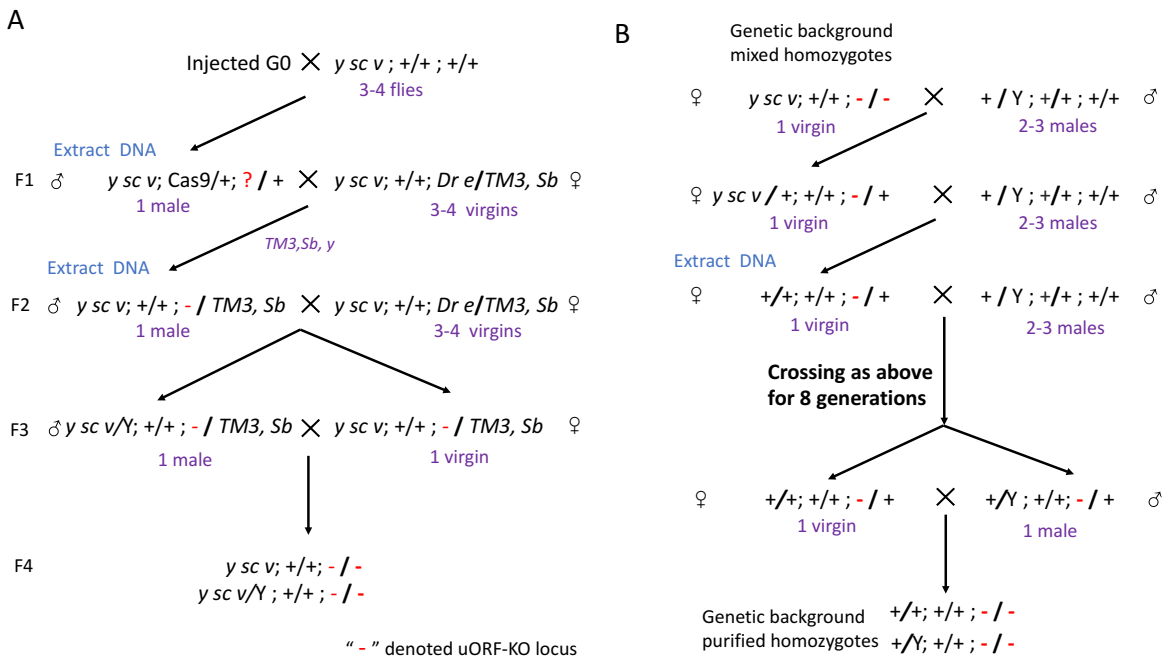


Figure S25. Overview of homozygous uORF-KO mutant screening. (A) Procedures for the primary screening of CRISPR/Cas9-induced mutants. Red "-" denotes the *bcd*-uORF-KO allele. (B) Purification of the genetic background of homozygous mutants by back-crossing with WT lines for nine generations.

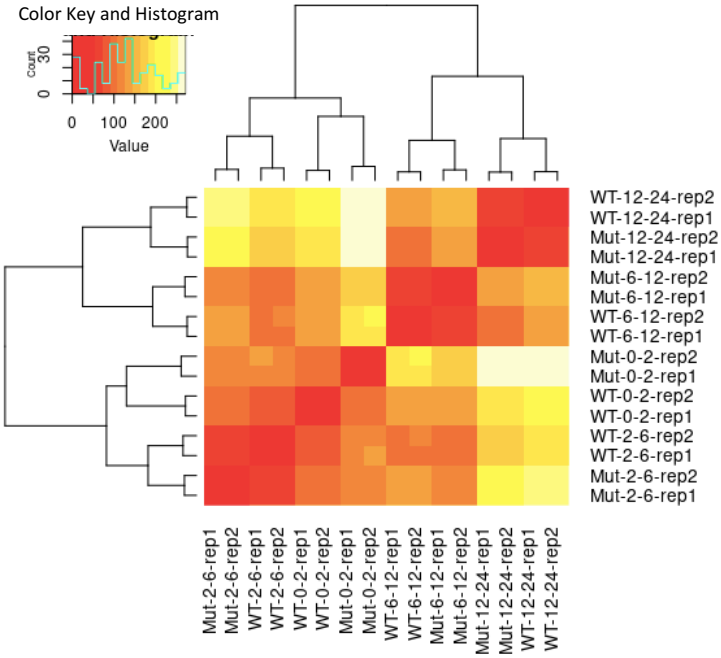


Figure S26. Heatmap of sample-to-sample distances of RNA-Seq libraries. Mut, mutant; WT, wild-type; rep1/rep2, biological replicate 1/biological replicate 2.

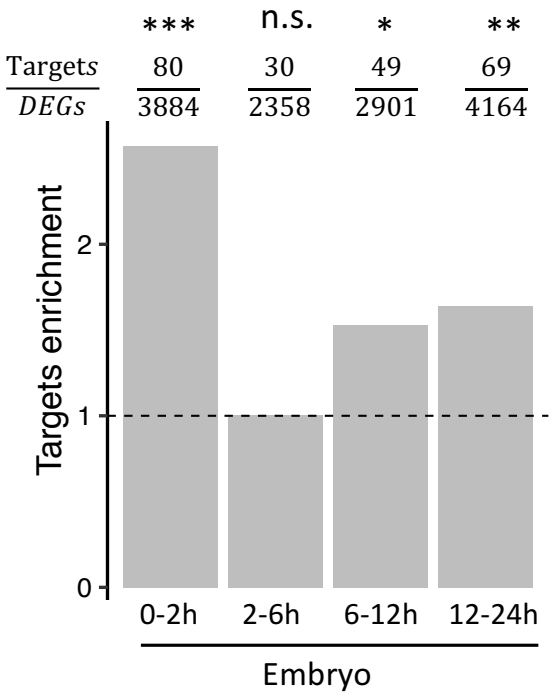
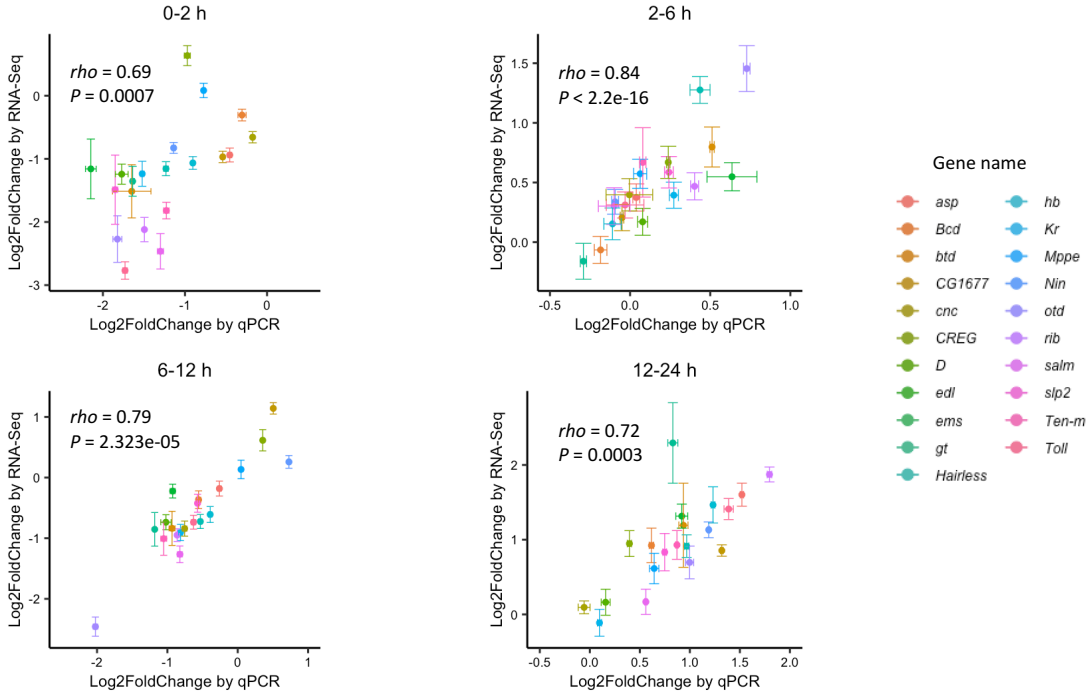


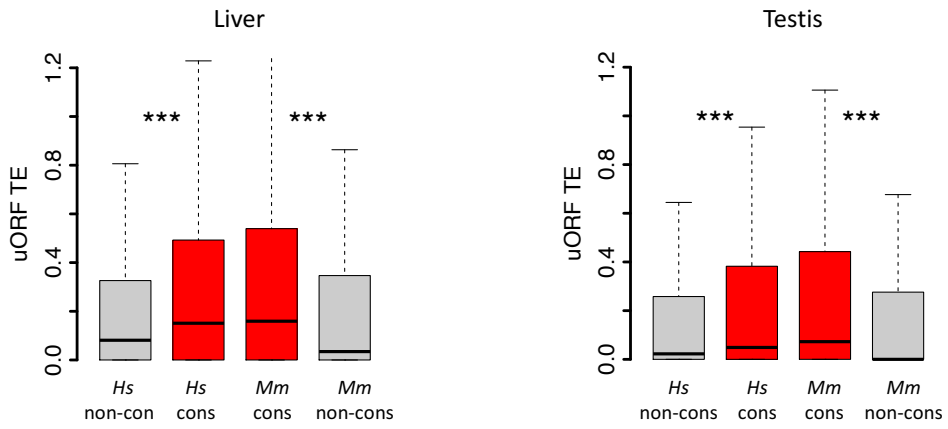
Figure S27. (A) The enrichment of Bcd targets in each embryo stage at 29°C. Enrichment was calculated as the ratio of the Bcd target proportion in DEGs to the Bcd target proportion in non-DEGs. χ^2 tests were performed to compare the difference in Bcd target distribution in DEGs and non-DEGs. The number of DEGs and the number of Bcd targets in DEGs for each stage are provided above the bar. Asterisks indicate statistical significance (χ^2 tests; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; n.s., $P > 0.05$). (B) The expression [$\log_2$(fold change)] levels of *bcd* and 3 of its targets, *Kr*, *btd*, and *otd,* in mutants compared to those in the WT in 4 embryonic stages. Asterisks are present at the top of bars only for |fold-changes| > 1.5 and adjusted *P* values < 0.05 (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$). (C) Expression of *bcd* and its targets (*Kr*, *btd*, *otd*) measured by RNA-Seq in uKO2/uKO2 mutant (red line) and WT (blue line) of *D. melanogaster*. The expression of each gene in each stage is shown as the RPKM value.

1685



1686
1687 **Figure S28.** The correlation of expression changes (log_2FoldChange) between RNA-Seq (y-axis) and
1688 RT-qPCR (x-axis) of 20 *bcd* targets in four embryo stages in uKO2/uKO2 mutant compared to WT.
1689

1690



1691
1692
1693 **Figure S29.** Boxplots showing the TEs of conserved and species-specific uORFs in the liver and testis
1694 of *H. sapiens* and *M. mulatta*. Wilcoxon rank-sum tests. ***, $P < 0.001$.
1695

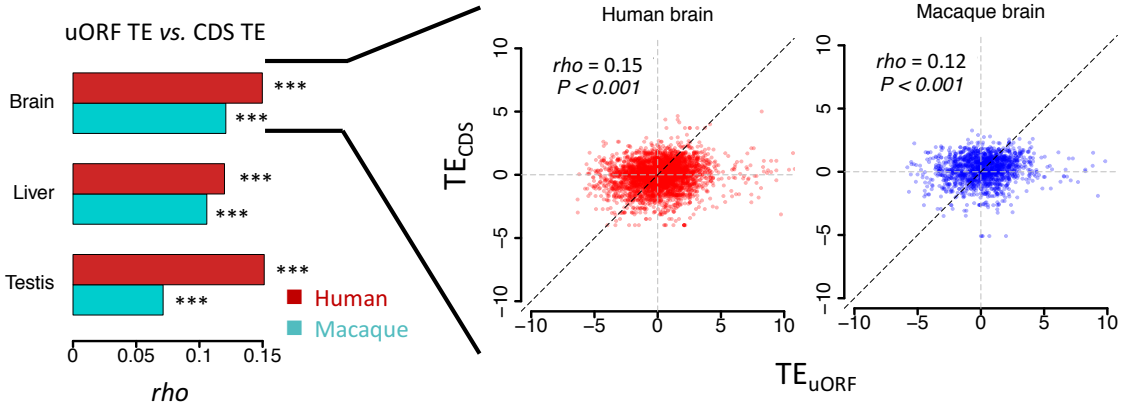


Figure S30. Correlations of uORF TE and corresponding CDS TE in 3 tissues of *H. sapiens* and *M. mulatta*. The bars represent Spearman's correlation coefficient (*rho*). In all samples, we obtained *P* values < 0.001. Data from brains are shown as examples in the right panel.

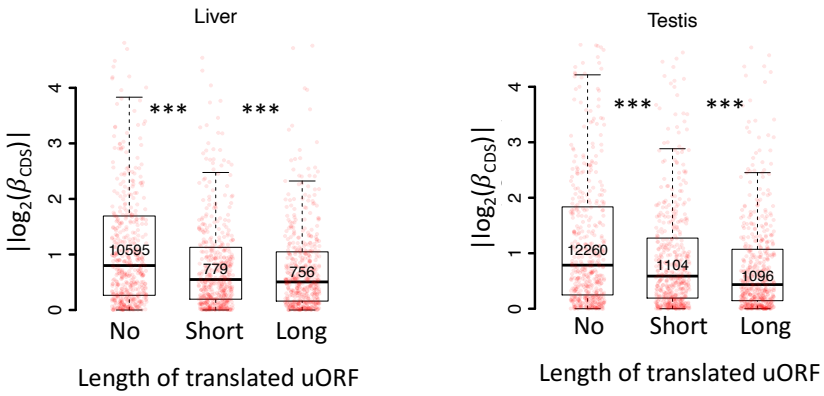


Figure S31. Boxplots showing interspecific CDS TE variability $|log_2(\beta_{CDS})|$ for three gene groups according to the length of their translated uORFs (No, short, long) in the liver (left) and testis (testis). Wilcoxon rank-sum tests. ***, $P < 0.001$.

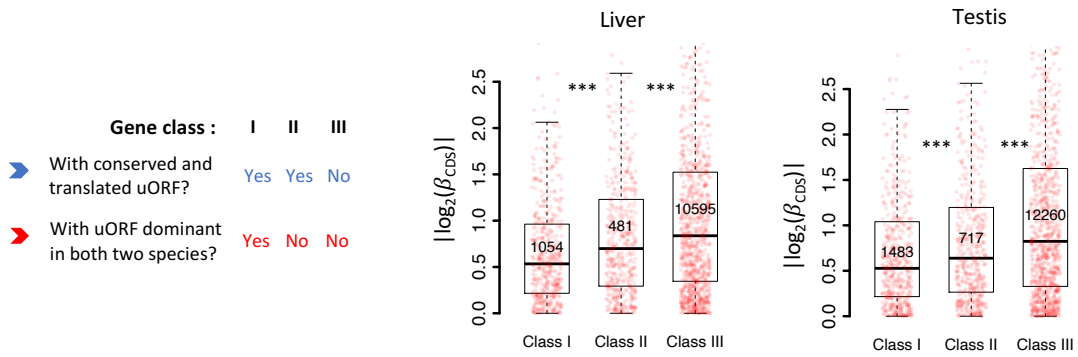


Figure S32. Boxplots showing interspecific CDS TE variability $|log_2(\beta_{CDS})|$ for different gene classes in the liver and testis. Genes expressed in liver or testis (mRNA RPKM > 0.1 in both species) were classified into three classes according to whether a gene had a conserved and dominant translated uORF (TE > 0.1) in both species or not. Wilcoxon rank-sum tests. ***, $P < 0.001$.

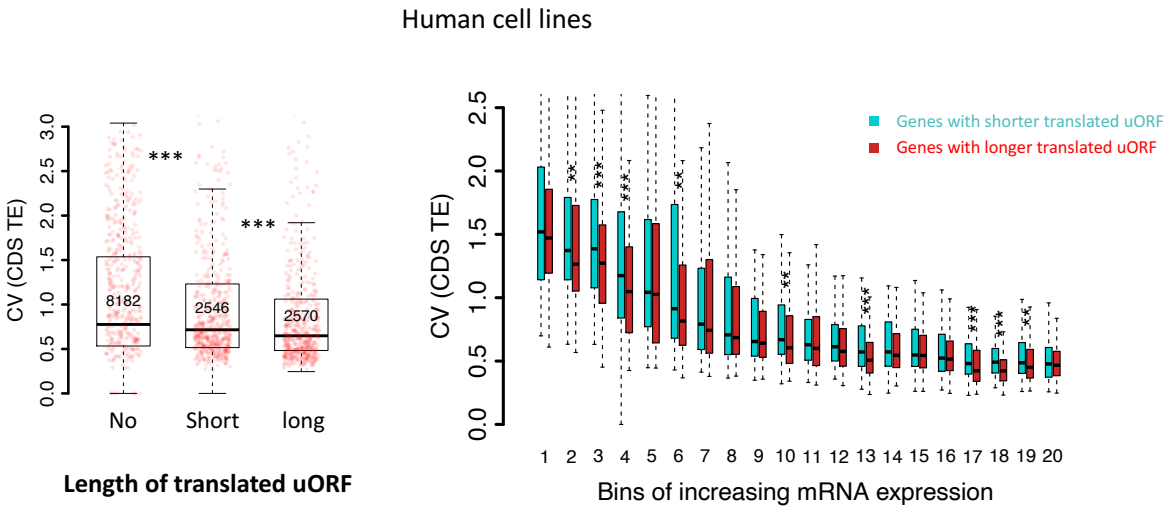


Figure S33. Left panel: Boxplot showing the coefficients of variation (CVs) of CDS TEs among 69 human lymphoblastoid cell lines (LCLs) for three gene groups. Genes were classified into 3 groups according to the total length of translated uORFs. Right panel: boxplot showing the coefficients of variation (CVs) of CDS TEs among the human cell lines. Genes with translated uORFs were divided into 20 bins with increasing mRNA expression levels. In each bin, genes were divided into two fractions according to the total length of translated uORFs.

Table S1. Parameters used in our simulation

Symbol	Description	Value
Len_a	Length of 5'-leader before uORF	50 triplets
Len_u	Length of uORF	Default: 30 triplets Variable: 2, 10, 20, 30, 40, 50, 100 triplets
Len_b	Distance between uORF and CDS	50 triplets
Len_m	Length of CDS	500 triplets
Len_c	Length of 3'UTR	50 triplets
v_s	Probability of movement of a 40S ribosome to the next position in a single action	0.3*
v_{Eu}	Probability of movement of an 80S ribosome to the next position in uORF in a single action	0.3**
v_{EC}	Probability of movement of an 80S ribosome to the next position in CDS in a single action	0.5***
R_{in}	Probability of loading a new 40S ribosome at the 5'-terminus of the mRNA in a single action	1000 values generated from uniform distribution or exponential distribution
K_{up}	Dissociation probability of upstream 40S ribosome colliding with downstream 80S ribosome	0 for downstream dissociation; 1 for upstream or double dissociation
K_{down}	Dissociation probability of downstream 40S ribosome colliding with upstream 80S ribosome	0 for upstream dissociation; 1 for downstream or double dissociation
I_{uORF}	Probability of translation initiation at the uORF start codon in a single action	Single-uORF model: 0, 0.1, 0.2, 0.3, 0.4, 0.5 Double-uORF model: 0, 0.1, 0.2, 0.3, 0.4
I_{CDS}	Probability of translation initiation at the CDS start codon in a single action	Single-uORF model: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0 Double-uORF model: 0.9

*Adopted from Andreev *et al.*'s original ICIER model [1], corresponding to a movement rate of 5 triplets/s.
**Adopted from Andreev *et al.*'s original ICIER model [1], corresponding to a movement rate of 5 codons/s.
*** Considering that uORFs usually encode blocking peptides [2-6] or contain stalling codons [7-9], we set v_{Eu} slightly lower than v_{EC}.

1740
1741 **Table S2.** Mapping statistics of Ribo-Seq and matched mRNA-Seq libraries
1742

Sample	Species*	Library type	Total reads (M)	After quality control (M) **	Unique mapping rate (%)	Multiple mapping rate (%)
0-2 h embryos	*D. melanogaster*	mRNA-Seq	61.01	52.74	93.14	2.84
2-6 h embryos	*D. melanogaster*	mRNA-Seq	59.44	52.07	93.26	3.24
6-12 h embryos	*D. melanogaster*	mRNA-Seq	53.56	45.11	92.33	3.09
12-24 h embryos	*D. melanogaster*	mRNA-Seq	61.18	50.64	91.89	4.31
Third-instar larvae	*D. melanogaster*	mRNA-Seq	58.28	45.46	88.53	6.09
P7-8 pupae	*D. melanogaster*	mRNA-Seq	52.99	41.16	92.65	4.46
Female adult heads	*D. melanogaster*	mRNA-Seq	21.48	16.87	93.00	3.27
Male adult heads	*D. melanogaster*	mRNA-Seq	16.32	14.18	91.17	4.24
Female adult bodies rep1	*D. melanogaster*	mRNA-Seq	16.84	11.20	88.83	5.76
Female adult bodies rep2	*D. melanogaster*	mRNA-Seq	24.94	19.37	88.73	6.00
Male adult bodies rep1	*D. melanogaster*	mRNA-Seq	19.79	16.01	91.62	6.39
Male adult bodies rep2	*D. melanogaster*	mRNA-Seq	11.24	7.64	89.86	6.69
0-2 h embryos	*D. melanogaster*	Ribo-Seq	64.81	51.96	90.36	6.31
2-6 h embryos	*D. melanogaster*	Ribo-Seq	53.14	41.19	69.47	16.92
6-12 h embryos	*D. melanogaster*	Ribo-Seq	66.24	46.76	86.02	8.94
12-24 h embryos	*D. melanogaster*	Ribo-Seq	63.7	47.45	88.96	6.30
Third-instar larvae	*D. melanogaster*	Ribo-Seq	39.53	22.13	58.83	6.82
P7-8 pupae	*D. melanogaster*	Ribo-Seq	41.53	20.33	80.92	5.51
Female adult heads	*D. melanogaster*	Ribo-Seq	63.94	39.90	94.39	2.92
Male adult heads	*D. melanogaster*	Ribo-Seq	58.22	23.87	93.36	3.24
Female adult bodies rep1	*D. melanogaster*	Ribo-Seq	44.11	19.85	84.20	6.36
Female adult bodies rep2	*D. melanogaster*	Ribo-Seq	41.95	18.73	70.74	11.33
Male adult bodies rep1	*D. melanogaster*	Ribo-Seq	42.56	18.98	79.39	8.65
Male adult bodies rep2	*D. melanogaster*	Ribo-Seq	39.52	28.66	81.19	6.93
0-2 h embryos	*D. simulans*	mRNA-Seq	19.92	18.33	90.85	5.82
2-6 h embryos	*D. simulans*	mRNA-Seq	13.88	7.47	88.32	7.46
6-12 h embryos	*D. simulans*	mRNA-Seq	21.13	18.64	89.26	7.06
12-24 h embryos	*D. simulans*	mRNA-Seq	16.17	13.96	87.77	8.47
Third-instar larvae	*D. simulans*	mRNA-Seq	17.56	12.69	89.88	6.57
P7-8 pupae	*D. simulans*	mRNA-Seq	23.87	19.04	89.92	5.62
Female adult heads	*D. simulans*	mRNA-Seq	17.13	15.11	89.70	6.39
Male adult heads	*D. simulans*	mRNA-Seq	17.72	15.96	90.73	6.15
Female adult bodies	*D. simulans*	mRNA-Seq	21.64	15.37	85.24	9.20
Male adult bodies	*D. simulans*	mRNA-Seq	21.54	17.34	87.45	8.84
0-2 h embryos	*D. simulans*	Ribo-Seq	63.38	49.48	80.09	10.81
2-6 h embryos	*D. simulans*	Ribo-Seq	53.46	38.09	77.27	13.72
6-12 h embryos	*D. simulans*	Ribo-Seq	57.76	43.84	81.58	10.81
12-24 h embryos	*D. simulans*	Ribo-Seq	50.55	35.05	83.55	8.21
Third-instar larvae	*D. simulans*	Ribo-Seq	67.69	35.40	86.44	7.10
P7-8 pupae	*D. simulans*	Ribo-Seq	61.74	40.45	87.45	4.76
Female adult heads	*D. simulans*	Ribo-Seq	62.52	26.18	86.51	4.58
Male adult heads	*D. simulans*	Ribo-Seq	65.07	28.99	89.62	3.96
Female adult bodies	*D. simulans*	Ribo-Seq	59.47	12.91	83.01	5.72
Male adult bodies	*D. simulans*	Ribo-Seq	53.55	18.35	87.44	5.61

1743 * The mRNA-seq and Ribo-Seq datasets for *D. melanogaster* were previously generated from Zhang *et al.* [10]. The
1744 datasets for *D. simulans* were generated in this study.
1745
1746 ** After removing the reads mapped to rRNA, miscRNA (snoRNA, snRNA, rRNA, tRNA), the yeast genome,
1747 and the Wolbachia genome.

1748
1749 **Table S3.** Genes with uORFs showing strong evidence of translational buffering.
1750 (see TableS3_strong_buffer_uORFs.csv)
1751

1752 **Table S4**. Mapping statistics of mRNA-Seq libraries for *bcd* uKO2/uKO2 mutant and WT flies
1753

Sample	Condition	Stain	Stage	Total read-pairs (M)	Unique mapping reads (M)	Unique mapping rate (%)	Multiple mapping rate (%)
uKO2-0-2 h-embryo-29-rep1	29℃	Mutant	0-2 h	19.39	16.73	86.31	9.13
uKO2-0-2 h-embryo-29-rep2	29℃	Mutant	0-2 h	18.72	12.72	67.94	26.64
uKO2-2-6 h-embryo-29-rep1	29℃	Mutant	2-6 h	19.60	16.76	85.51	11.60
uKO2-2-6 h-embryo-29-rep2	29℃	Mutant	2-6 h	18.01	15.40	85.52	11.53
uKO2-6-12 h-embryo-29-rep1	29℃	Mutant	6-12 h	18.55	16.26	87.64	7.52
uKO2-6-12 h-embryo-29-rep2	29℃	Mutant	6-12 h	19.76	17.31	87.58	7.79
uKO2-12-24 h-embryo-29-rep1	29℃	Mutant	12-24 h	17.48	15.23	87.14	8.36
uKO2-12-24 h-embryo-29-rep2	29℃	Mutant	12-24 h	18.78	15.98	85.09	10.05
w1118-0-2 h-embryo-29-rep1	29℃	WT	0-2 h	19.10	17.24	90.25	6.87
w1118-0-2 h-embryo-29-rep2	29℃	WT	0-2 h	19.02	17.12	90.01	7.38
w1118-2-6 h-embryo-29-rep1	29℃	WT	2-6 h	18.54	15.60	84.11	12.40
w1118-2-6 h-embryo-29-rep2	29℃	WT	2-6 h	19.24	16.34	84.93	11.84
w1118-6-12 h-embryo-29-rep1	29℃	WT	6-12 h	18.61	15.87	85.28	10.50
w1118-6-12 h-embryo-29-rep2	29℃	WT	6-12 h	19.48	16.83	86.39	9.45
w1118-12-24 h-embryo-29-rep1	29℃	WT	12-24 h	18.93	16.44	86.85	6.96
w1118-12-24 h-embryo-29-rep2	29℃	WT	12-24 h	19.70	16.95	86.03	7.69
uKO2-0-2 h-embryo-25-rep1	25℃	Mutant	0-2 h	16.22	14.94	92.12	4.27
uKO2-0-2 h-embryo-25-rep2	25℃	Mutant	0-2 h	16.23	14.91	91.91	4.15
uKO2-2-6 h-embryo-25-rep1	25℃	Mutant	2-6 h	16.80	14.50	86.31	7.53
uKO2-2-6 h-embryo-25-rep2	25℃	Mutant	2-6 h	19.34	16.58	85.71	7.61
uKO2-6-12 h-embryo-25-rep1	25℃	Mutant	6-12 h	16.17	14.20	87.80	8.49
uKO2-6-12 h-embryo-25-rep2	25℃	Mutant	6-12 h	16.34	14.36	87.90	8.11
uKO2-12-24 h-embryo-25-rep1	25℃	Mutant	12-24 h	16.85	14.35	85.20	11.33
uKO2-12-24 h-embryo-25-rep2	25℃	Mutant	12-24 h	16.88	14.25	84.44	12.32
w1118-0-2 h-embryo-25-rep1	25℃	WT	0-2 h	17.29	15.74	91.06	5.27
w1118-0-2 h-embryo-25-rep2	25℃	WT	0-2 h	19.53	17.83	91.26	5.06
w1118-2-6 h-embryo-25-rep1	25℃	WT	2-6 h	19.23	16.85	87.62	7.76
w1118-2-6 h-embryo-25-rep2	25℃	WT	2-6 h	17.22	14.86	86.27	9.67
w1118-6-12 h-embryo-25-rep1	25℃	WT	6-12 h	19.04	16.02	84.15	11.76
w1118-6-12 h-embryo-25-rep2	25℃	WT	6-12 h	17.59	15.00	85.31	10.71
w1118-12-24 h-embryo-25-rep1	25℃	WT	12-24 h	16.25	14.03	86.35	10.29
w1118-12-24 h-embryo-25-rep2	25℃	WT	12-24 h	16.23	14.08	86.73	9.37

1754

1755 **Table S5.** Numbers of genes showing different magnitudes of TE changes between uORFs and CDS at the
1756 interspecific level, *H. sapiens* and *M. mulatta*.
1757

Tissues	# of expressed uORFs	$\beta_u \neq 1$ (%)	# of expressed CDSs	$\beta_c \neq 1$ (%)	uORF-CDS pairs with $\beta_u > 1$			uORF-CDS pairs with $\beta_u < 1$		
					Total	$\gamma > 1$	$\gamma < 1$	Total	$\gamma > 1$	$\gamma < 1$
Brain	7,380	80 (1.08)	15,086	507 (3.36)	51	0	27	29	14	0
Liver	4,429	28 (0.63)	13,246	149 (1.24)	10	0	4	18	10	0
Testis	7,384	53 (0.72)	15,134	272 (1.80)	33	0	9	20	5	0

1758
1759 Only uORFs and CDSs with an mRNA RPKM > 0.1 in both *H. sapiens and M. mulatta* were considered in each
1760 sample pair in the analysis. $\beta_u = TE_{uORF,macaque}/TE_{uORF,human}$, is the fold change in TE_{uORF} in *M. mulatta*
1761 relative to *H. sapiens* for each sample. $\beta_c = TE_{CDS,macaque}/TE_{CDS,human}$, is the fold change in TE_{CDS} in *M. mulatta*
1762 relative to *H. sapiens* for each sample. $\gamma = \beta_c/\beta_u$. For each CDS-uORF pair, β_u>1 and $\gamma < 1$ or β_u<1 and $\gamma > 1$
1763 means that the magnitude of TE change is lower for a CDS than a uORF. The statistical significance of β_c, β_u,
1764 and γ were all determined at an FDR < 0.05.
1765

1766 **Table S6.** The primer sequences used in this study

Primer	Sequence (5'->3')	Application
Bcd-uORF-sgRNA-F	ATCGCAAAAACGCAAAATGT	sgRNA for Bcd uORF knock-out
Bcd-uORF-sgRNA-R	ACATTTTGCGTTTTTGCGAT	
Bcd-genotyping-F	GCTTTGCCGTACTGTTCGAT	Primers used for Bcd genotyping
Bcd-genotyping-R	AACTGAAGCTGCGGATGTTG	
Bcd-qPCR-F	GATGTATCTGGGTGGCTGCT	RT–qPCR primers used for Bcd mRNA P-to-M ratio quantification
Bcd-qPCR-R	CCGAAATGTGGGACGATAAC	
Rp49-qPCR-F	CACTTCATCCGCCACCAGTC	RT–qPCR primers used for rp49 as inter reference
Rp49-qPCR-R	CGCTTGTTCGATCCGTAACC	
Bcd-5utr-PC1-wF1	CATCCCTAAATAACGGCACT	Used for WT and uATG-mutated *bcd* 5'UTR amplification
Bcd-5utr-PC1-wR	CGCGAACTTATTTCGCCCACATTTTGCGTTTTTG	
Bcd-5utr-PC1-mR1	CGCGAACTTATTTCGCCCACTTATTTTGCGTTTTTG	
Bcd-5utr-PC1-mR2	CGCGAACTTATTTCGAACTTATTTGCGATCCACTGCT	
Bcd-5utr-PC2-wF1	CGAAATAAGTTCGCGAGCGT	
Bcd-5utr-PC2-wR1	GTACACCTTGGAAGCCGGCGGTTGCGCCATTTTCC	
btd_qP_F	CACCAATTACCGCCCGTATG	RT-qPCR Primers for Bcd target genes
btd_qP_R	CGGCCAATACCTTCTTGCTC	
kr_qP_F	CCAGTGGCAGCTCTCCTAAT	
kr_qP_R	GCGAGATGGATCCTTGGAGA	
otd_qP_F	CTACAACATGGGCCATTCGG	
otd_qP_R	GCGACATGTAATCCAGGCTG	
Toll_qP_F	TTACTTCAGTGCCTCGCTCA	
Toll_qP_R	CCAGAGCCGATAACATTGCC	
hb_qP_F	GGAGAAGGAGCACGATCAGA	
hb_qP_R	TGCCGTGCGAATTGTAGATG	
Hairless_qP_F	GCATACATTCCCTCGCCTTC	

Hairless_qP_R	CGCTCCTGTCTGATAGTGGT
salm_qP_F	GCCCATTCAAGTGCAGGATC
salm_qP_R	GAAGTTTCTCATCGGCGGAC
D_qP_F	CAAGAGGGTCACATCAAGCG
D_qP_R	CAGAGTTGTGCATCTTGGGG
Ten-m_qP_F	CCACCGCATCCTAAAGCATG
Ten-m_qP_R	CTCTTTCTCCTCCTCGGTCC
rib_qP_F	TTCGCCCTACACAGATCGTT
rib_qP_R	CGCACATCATATCCGCTCTG
CG1677_qP_F	TTGGCCGGGGATCATATGAA
CG1677_qP_R	GGTGGAATACGACGAGTTGC
Nin_qP_F	TAATGACCGCGATCACTGGA
Nin_qP_R	GCACCAGTTCAAGTTCCGTT
gt_qP_F	CCCTCGCCGTATCAAACAAG
gt_qP_R	GGGCGGCATACAGAAGATTG
asp_qP_F	AGCTACCATTTTCCCGGACA
asp_qP_R	CATCTGGAGTGATTCGACGC
CREG_qP_F	CTGATTTTGGCCCTCGTGAG
CREG_qP_R	GGCGATCTTGGCGTGATTTA
slp2_qP_F	GCAAGCCGGTGAAGGATAAG
slp2_qP_R	ACTCTGACGAATGGCCATCA
edl_qP_F	TCAAATGCTGGACAAGTGCC
edl_qP_R	GCGCTCATCAGTCGATTGTT
Mppe_qP_F	TGGACTTTAGCAACCCGGAT
Mppe_qP_R	CGCTGATGATGGCTCTCAAC
cnc_qP_F	GGAAACAGCAGCAACTTGGA
cnc_qP_R	CCGGCACTGAAATGGGTATG
ems_qP_F	ATCTATCCGGTCCCCAGAGT
ems_qP_R	CCTGGGATCAAGGTGAGTGT

SOM Reference

1. Andreev, D.E., Arnold, M., Kiniry, S.J., Loughran, G., Michel, A.M., Rachinskii, D., and Baranov, P.V. (2018). TASEP modelling provides a parsimonious explanation for the ability of a single uORF to derepress translation during the integrated stress response. Elife *7*. 10.7554/eLife.32563.
2. Ivanov, I.P., Shin, B.S., Loughran, G., Tzani, I., Young-Baird, S.K., Cao, C., Atkins, J.F., and Dever, T.E. (2018). Polyamine Control of Translation Elongation Regulates Start Site Selection on Antizyme Inhibitor mRNA via Ribosome Queuing. Mol Cell *70*, 254-264 e256. 10.1016/j.molcel.2018.03.015.
3. Luo, Z., and Sachs, M.S. (1996). Role of an upstream open reading frame in mediating arginine-specific translational control in Neurospora crassa. J Bacteriol *178*, 2172-2177. 10.1128/jb.178.8.2172-2177.1996.
4. Lovett, P.S., and Rogers, E.J. (1996). Ribosome regulation by the nascent peptide. Microbiol Rev *60*, 366-385. 10.1128/mr.60.2.366-385.1996.
5. Vilela, C., and McCarthy, J.E. (2003). Regulation of fungal gene expression via short open reading frames in the mRNA 5'untranslated region. Mol Microbiol *49*, 859-867. 10.1046/j.1365-2958.2003.03622.x.
6. Raney, A., Law, G.L., Mize, G.J., and Morris, D.R. (2002). Regulated translation termination at the upstream open reading frame in s-adenosylmethionine decarboxylase mRNA. J Biol Chem *277*, 5988-5994. 10.1074/jbc.M108375200.
7. Lin, Y., May, G.E., Kready, H., Nazzaro, L., Mao, M., Spealman, P., Creeger, Y., and McManus, C.J. (2019). Impacts of uORF codon identity and position on translation regulation. Nucleic Acids Res *47*, 9358-9367. 10.1093/nar/gkz681.
8. Meijer, H.A., and Thomas, A.A. (2003). Ribosomes stalling on uORF1 in the Xenopus Cx41 5' UTR inhibit downstream translation initiation. Nucleic Acids Res *31*, 3174-3184. 10.1093/nar/gkg429.
9. Bottorff, T.A., Park, H., Geballe, A.P., and Subramaniam, A.R. (2022). Translational buffering by ribosome stalling in upstream open reading frames. PLoS Genet *18*, e1010460. 10.1371/journal.pgen.1010460.
10. Zhang, H., Dou, S., He, F., Luo, J., Wei, L., and Lu, J. (2018). Genome-wide maps of ribosomal occupancy provide insights into adaptive evolution and regulatory roles of uORFs during Drosophila development. PLoS biology *16*, e2003903.