# Upstream open reading frames dynamically modulate CLOCK protein translation to regulate circadian rhythm and sleep

Yuanqiang Sun[1,#], Ke Shui[2,3,#], Qinyu Li[3], Wanting Jin[1], Chenlu Liu[1], Jian Lu[1,*], Luoying Zhang[3,4,*]

[1] China State Key Laboratory of Protein and Plant Gene Research, Center for Bioinformatics, School of Life Sciences, Peking University, Beijing 100871, China

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

[3] Key Laboratory of Molecular Biophysics of Ministry of Education, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei, 430074, China

[4] Hubei Province Key Laboratory of Oral and Maxillofacial Development and Regeneration, Wuhan, Hubei 430022, China

#These authors contributed equally to this work.

*Corresponding author.

Email: luj@pku.edu.cn (Jian Lu); zhangluoying@hust.edu.cn (Luoying Zhang)

**Keywords**: *Drosophila*, *Clock*, uORF, circadian rhythm, sleep

**Summary**

The circadian rhythm is an evolutionarily conserved mechanism with translational regulation increasingly recognized as pivotal in its modulation. In this study, we show evidence that the upstream open reading frames (uORFs) of the core circadian clock gene, *Clock* (*Clk*), rhythmically and substantially attenuate CLK protein translation in *Drosophila*, with pronounced suppression occurring during daylight hours. Eliminating *Clk* uORFs results in elevated CLK protein levels during the day and a compressed circadian cycle, along with a broad shift in clock gene expression rhythms. Interestingly, *Clk* uORF deletion also augments morning sleep by reducing dopaminergic activity. Beyond daily circadian adjustments, *Clk* uORFs play a role in modulating sleep patterns in response to the varying day lengths of different seasons, inhibiting translation in a day-length contingent manner. Furthermore, the *Clk* uORFs act as a master regulator to shape the rhythmic expression of a vast array of genes and influencing multifaceted physiological outcomes. Collectively, our research sheds light on the intricate ways uORFs dynamically adjust downstream coding sequences to acclimate to environmental shifts.

## Introduction

Circadian clocks drive ~24 h (or circadian) rhythms in a myriad of biological processes, with sleep-wake rhythm being one of the most prominent manifestations. Molecularly, these rhythms are underpinned by a series of transcriptional and post-transcriptional feedback loops that are evolutionarily conserved (*1*). In the model organism, *Drosophila*, two transcription factors, CLOCK (CLK) and CYCLE (CYC), constitute the positive limb of the loops. CLK and CYC heterodimerize and activate the transcription of *period* (*per*) and *timeless* (*tim*) via E-box enhancer elements (*2*). As PER and TIM proteins accumulate in the cytoplasm, they bind to each other and translocate into the nucleus. They then suppress the transcriptional activities of CLK and CYC, thus repressing their own transcription and forming a pivotal negative feedback loop. During this process, post-translational modifications (PTMs) influence the stability of PER and TIM, leading ultimately to their degradation. This enables CLK/CYC to embark on another transcriptional cycle (*1*). While the majority of the studies regarding clock regulation focus on transcriptional and post-translational mechanisms, an emerging area of interest lies in the realm of translational regulation. A few studies have demonstrated several proteins such as TWENTY-FOUR and ATAXIN2 play critical roles in regulating the translation of PER, indicating a critical role for translational regulation in the circadian clockwork (*3*-*6*). Overall, the mechanisms underlying translational control of the clock remain under-characterized.

Protein synthesis is tightly regulated in eukaryotic cells, with translational initiation serving as the rate-limiting step of mRNA translation (*7*). Upstream open reading frames (uORFs) are short ORFs with start codons located in the 5' untranslated region (5'UTR) of eukaryotic mRNAs. uORFs are widely distributed in eukaryotic genomes, and they suppress translation initiation of downstream coding

sequences (CDSs) by sequestering ribosomes or competing for binding to translation initiation complexes (*8-13*). The notion that uORFs orchestrate global circadian translatomic regulation is supported in both fly and mammalian studies (*14-16*). Notwithstanding, empirical validation of uORFs in circadian regulation remains sparse. In *Neurospora*, uORFs of the core clock gene *frequency* participate in the temperature-sensing process of the circadian clock (*17*). By generating transgenic plants expressing core clock gene *TIMING OF CAB EXPRESSION 1* with WT or mutant uORFs, Wu and colleagues identified a role for these uORFs in contributing to the robust operation of the circadian clock (*18*). Two other groups reported uORFs in the mouse core clock gene *Period2* to be involved in temperature entrainment of the circadian clock (*19*) and regulating sleep (*20*). While the regulatory potential of uORFs in circadian rhythm is established, a detailed molecular delineation remains a coveted objective.

This study embarks on an in-depth exploration of uORFs in *Clk* and *cyc* of *Drosophila melanogaster*. While the uORFs of both genes suppress translation *in vitro*, only the uORFs of *Clk* appear to substantially inhibit protein translation *in vivo*. Knockout (KO) of *Clk* uORFs amplified daytime CLK protein levels and shortened the circadian period of locomotor rhythm in the mutants. Alongside these molecular deviations, *Clk*-uORF-KO mutants manifest increased morning sleep, likely stemming from diminished dopaminergic activity. These mutants also exhibit defects in modulating sleep duration in response to photoperiodic shifts. Molecular investigations underscore the indispensability of *Clk* uORFs for the photoperiod-driven regulation of CLK protein, highlighting their central role in both circadian and circannual orchestrations. Beyond circadian rhythm and sleep phenotypes, these mutants, intriguingly, also showed diminished fecundity and

reduced starvation resilience, signifying broader physiological ramifications of *Clk* uORF ablations.

## Results

### uORFs suppress CLK and CYC translation in cultured cells

In *Drosophila melanogaster*, the 5'UTR of *Clk* mRNA contains five canonical uORFs with AUG start codons (uAUG), among which the second (uORF2) and the third uORF (uORF3) are in-frame to each other and overlapping. None of these uORFs overlaps with downstream CDS (Figure 1A). The 5'UTR of *cyc* mRNA contains four canonical uORFs overlapping with each other but not with downstream CDS (Figure S1A). Since the start codon is the most important and definitive feature of an uORF (*11, 21*), we conducted multiple sequence alignments (MSA) of the uATGs of these uORFs across 22 insect species. Notably, among the five uORFs of *Clk*, the start codon of uORF2 (uAUG2) was the most conserved, as it was identical across all the species (Figure 1A). In comparison, none of the start codons of *cyc* uORFs was universally conserved across these species, and the uAUGs of *cyc* were generally less conserved than those of *Clk* (Figure S1A).

To investigate whether these *Clk* and *cyc* uORFs show evidence of translation, we analyzed our previously published transcriptome and translatome data from multiple tissues of *Drosophila melanogaster* (*22*). We detected prominent mRNA expression and ribosome occupancy for uORF2 and uORF3 of *Clk* in the head and body of male and female flies (Figure 1B). As for *cyc*, prominent mRNA expression and ribosome occupancy was observed for uORF2, uORF3 and uORF4 (Figure S1B). These findings implicate substantial translational activity at these uORFs and thus they are likely to be functional.

To further investigate the regulatory effects of these uORFs on translation, we knocked out *Clk* uORF2 and uORF3 simultaneously by mutating their start codon ATGs to ATTs, which showed prominent ribosome occupancies and higher conservation, and established a homozygous knock-out line (*Clk*-uORF-KO) (Figure 1C). Similarly, we generated a *cyc* uORF KO line by mutating the ATGs of uORF2-4 to ATTs (Figure 1C). We cloned wild-type (WT) and mutant 5'UTRs of *Clk* and *cyc* mRNA into a dual-luciferase reporter construct, and measured relative luminescence intensity to assess the CDS translational activity under the controls of different 5'UTRs (Figure 1D). The relative luminescence level of mutant 5'UTRs are significantly higher than that of the WT, indicating that these uORFs are indeed functional and repress the translation of downstream CDSs (Figure 1D).

**_Clk_ uORFs suppress translation of CLK protein during the day time**

To investigate the function of *Clk* and *cyc* uORFs *in vivo*, we carried out ribosome fractionation to compare *Clk* and *cyc* mRNA abundance in the polysome vs. monosome fractions (P-to-M ratio) (Figure 2A). A larger P-to-M ratio means more mRNAs are enriched in the polysome fractions and bound by more ribosomes, thus indicative of higher translation efficiency (*23-25*). We found that the P-to-M ratios of *Clk* mRNA in *Clk*-uORF-KO flies are significantly higher at Zeitgeber Time (ZT) 0, 4 and 8 (ZT0 is defined as the time of lights on) (Figure 2B). On the other hand, the P-to-M ratios of *cyc* mRNA in *cyc* uORF KO flies are only modestly elevated at ZT16 (Figure 2C). These results indicate that *Clk* uORFs substantially impact translation, while *cyc* uORFs only exert mild influences. Therefore, we focused on *Clk* uORFs for the remainder of this study.

Next, we examined the consequences of knocking out *Clk* uORFs on CLK protein

by immunoblotting. CLK protein level is significantly elevated at ZT0 and 4, with a trend of increase at ZT8 (Figure 2D and 2E). This well agrees with the changes in P-to-M ratios (Figure 2B). These series of results indicate that *Clk* uORFs inhibit the translation of CLK protein during the daytime, and the effects are particularly prominent in the morning.

**_Clk_ uORFs modulate the pace of the circadian clock**

To explore the role of *Clk* uORFs in circadian regulation, we adopted a previously developed mathematical model to simulate the *Drosophila* molecular clock (*26*). Given that the system relies on nonlinear ordinary differential equations to model the regulatory weights among mRNAs or proteins molecules in the circadian regulatory network, it is applicable to simulate the enhanced *Clk* translation caused by the disruption of uORFs. To achieve this, we increased the original value of regulatory weight from *Clk* mRNA to CLK protein (0.026, baseline) to three larger values (0.027, 0.028 and 0.030, respectively). This modification allowed us to model the enhanced *Clk* translation resulting from knockout of uORFs (Figure 3A). Based on the simulation results, both *Clk* mRNA and protein abundance increase as the regulatory weight of *Clk* mRNA to protein increases (Figure 3B). Consequently, the mRNA and protein levels of *per* and *tim* also increase (Figure 3B). To quantitatively simulate the length of circadian period as the regulatory weight of *Clk* mRNA to protein increases, we adopted four different regulatory weights (0.026 as the baseline and three other increased values: 0.027, 0.028 and 0.030) in the model and calculated the time intervals between the peaks of *tim* mRNA as the proxy of circadian period length. The simulation results show a shortening of circadian period as the regulatory weight of *Clk* mRNA to protein increases, regardless of whether measured by the median or

mean (Figure 3C; Table S1).

To validate these simulations, we assessed the mRNA level of *per* and *tim*, observing significant elevation in *Clk*-uORF-KO flies with the peak level occurring at an earlier time (Figure 3D). Consistent with this phase advance of the molecular oscillations, the period of locomotor rhythm is shortened by over one hour in *Clk*-uORF-KO flies under constant darkness condition (Figure 3E and 3F; Table S2). These experimental findings match the simulations quite well, strongly suggesting that *Clk* uORFs suppress CLK translation to slow down the pace of the molecular clock by reducing the expression of CLK target genes such as *per* and *tim*.

**_Clk_ uORFs shape the temporal transcriptomic landscape**

Besides being a master transcriptional regulator in the molecular clock, CLK also drives cyclic expression of a substantial portion of the transcriptome (*27*). Therefore, we investigated the influences of *Clk* uORFs on global rhythms of gene expression. We collected heads from WT and *Clk*-uORF-KO flies every 4 hours throughout the day with three biological replicates and conducted transcriptome profiling by RNA sequencing (RNA-seq). JTK-CYCLE was employed to identify genes with abundance that oscillates at the mRNA level (*28*). In total, 3,503 genes are rhythmically expressed in WT females while 3,868 genes are rhythmically expressed in *Clk*-uORF-KO females with 2,085 overlapping genes (Figure 4A and 4B). As for males, 3,317 genes are rhythmically expressed in WT males while 3,328 genes are rhythmically expressed in *Clk*-uORF-KO males with 1,582 overlapping genes (Figure S2A and S2B). The oscillatory expression of 40-52% of the genes in WT are abolished by knocking out *Clk* uORFs, while roughly 1,700 genes acquire cyclic expression patterns in *Clk*-uORF-KO flies. We next quantified the amplitude and

phase of the mRNA rhythms. Knocking out *Clk* uORFs appears to increase the amplitude of the oscillations with only a slight advance of the phase in males (Figure 4C, 4D, and S2C-E). In females, on the other hand, knocking out *Clk* uORFs leads to desynchronized oscillations with peak expression time distributed throughout the day (Figure 4E).

To probe the physiological functions of these alterations in the transcriptome caused by knocking out *Clk* uORFs, we conducted a Gene Ontology (GO)-based enrichment analysis. We found that genes which lost rhythms in *Clk*-uORF-KO flies are mainly involved in translation, ribosome processes, transport and metabolism (Figure 4F and S4F). On the other hand, genes that gain rhythms when *Clk* uORFs are knocked out belong to pathways such as proton transport, development and mitochondrial processes (Figure 4G and S4G). Taken together, these results indicate that *Clk* uORFs exert pervasive influences on global rhythms of gene expression.

**_Clk_ uORFs promote morning wakefulness by tuning up the dopaminergic tone**

CLK has been shown to promote sleep (*29*). Therefore, we tested whether *Clk* uORFs also play a role in sleep regulation. We found that knocking out *Clk* uORFs significantly lengthens sleep duration, with more prominent alteration in the females (Figure 5A and 5B) compared to males (Figure S3A and S3B). More detailed analysis demonstrates that sleep duration is increased in the first 8 hours of the light phase in *Clk*-uORF-KO flies, indicating that *Clk* uORFs function to promote wakefulness in the morning (Figure 5A-C). This increase in sleep duration is caused by lengthening average sleep bout duration, indicating that knocking out *Clk* uORFs facilitates sleep maintenance and thus results in more consolidated sleep (Figure 5D). The waking activity is also increased in *Clk*-uORF-KO flies, which means lengthening of sleep

duration is not a result of impaired motor activity (Figure 5D).

Since it has been reported that CLK down-regulates the expression of the gene encoding tyrosine hydroxylase (*ple*) which leads to decreased dopamine signaling and reduced activity level, we tested whether dopamine signaling mediates the influences of *Clk* uORFs on sleep as well (*30*). We first measured *ple* mRNA level and found it decreased in *Clk*-uORF-KO flies (Figure 5E). Next, we treated flies with dopamine precursor L-DOPA which should facilitate dopamine synthesis (*31*). This treatment rescues the increased sleep duration in *Clk*-uORF-KO flies (Figure 5F and 5G). In short, *Clk* uORF promotes morning wakefulness by enhancing the dopaminergic tone.

**_Clk_ uORFs limit starvation-induced wakefulness and are beneficial for survival**

Starvation induces sleep loss in a CLK-dependent manner, and this enhanced wakefulness is believed to be beneficial for survival (*32*). We tested whether *Clk* uORFs are involved in mediating the impact of starvation on sleep/wakefulness. We observed an exaggerated sleep loss in *Clk*-uORF-KO flies, which means *Clk* uORFs function to restrict starvation-induced sleep loss (Figure S4A-C). Consistent with the excessive wakefulness, *Clk*-uORF-KO flies exhibit significantly accelerated mortality rates compared to WT flies under starvation conditions ($P = 2.4\times10^{-9}$ and $8.8\times10^{-13}$ for males and females, respectively, log-rank test) (Figure S4D). These results identify a role for *Clk* uORFs in adjusting the appropriate amount of sleep loss under starvation condition which may be beneficial for survival.

**_Clk_ uORFs modulate sleep/wakefulness in adaptation to seasonal photoperiodic changes by suppressing CLK protein translation in a photoperiod-dependent manner**

Our results thus far demonstrate how *Clk* uORFs dynamically regulate translation throughout the day to influence circadian rhythm and sleep. Since the circadian clock is believed to be important for adaptations to seasonal changes in the environment, we suspected that *Clk* uORFs also regulate translation in a dynamic manner on the circannual time scale (*33*). To address this issue, we subjected *Clk*-uORF-KO and WT flies to conditions that mimic seasonal changes. Because the most important seasonal signal for the circadian clock is the shortening/lengthening of photoperiod, we monitored locomotor rhythm and sleep under photoperiods of different lengths, ranging from 4 h light/20 h dark (4L20D) to 20L4D (*33*). *Clk* uORFs do not appear to alter locomotor activity rhythm under different photoperiods, indicating that *Clk* uORFs are not required for the adaptation of locomotor activity rhythm to seasonal photoperiod changes (Figure S5).

On the other hand, sleep duration is increased more prominently in *Clk*-uORF-KO flies as the photoperiod shortens (Figure 6A-E). Interestingly, we found that sleep duration reduces substantially in WT when the photoperiod becomes shorter (Figure 6E). In *Clk*-uORF-KO flies, however, the magnitude of this sleep reduction is smaller. When comparing the sleep duration between 4L20D and 8L16D, there is significant decrease in WT flies but not in *Clk*-uORF-KO. These findings uncover a role for *Clk* uORFs in adapting sleep/wakefulness to seasonal changes of photoperiod.

Given these behavioral phenotypes, we hypothesized that *Clk* uORFs inhibit translation in a photoperiod-dependent manner. To test this, we carried out ribosome fractionation under 4L20D and 12L12D. We observed a consistent trend of decreased P-to-M ratios throughout the day under 12L12D vs. 4L20D in WT but not in *Clk*-uORF-KO flies (Figure 7A and 7B). We further calculated the area under curve

(AUC) for the data in Figure 7A and 7B, which can reflect gross daily CLK protein translation. Notably, long photoperiod substantially reduces total daily CLK translation in WT (~30%) while the magnitude of this reduction is much smaller in *Clk*-uORF-KO flies (~8%) (Figure 7C and 7D). A photoperiod-dependent alteration is also observed at the protein level. In WT flies, CLK protein level is significantly reduced from ZT4-12 under 4L20D compared to 12L12D (Figure 7E and 7F). On the other hand, CLK protein level is only decreased at ZT0 in *Clk*-uORF-KO flies under 4L20D relative to 12L12D (Figure 7E and 7F). AUC values demonstrate that short photoperiod significantly reduces total daily CLK level in WT but not *Clk*-uORF-KO flies (Figure 7G and 7H). All in all, these series of results indicate that *Clk* uORFs suppress translation in a photoperiod-dependent manner to promote wakefulness in adaptation to seasonal changes.

**Knocking out *Clk* uORFs impairs fecundity**

Recognizing the roles of *Clk* uORFs in facilitating the adaptation to environmental changes, we further investigated whether they enhance evolutionary fitness by assessing fecundity. Notably, mated *Clk*-uORF-KO females laid considerably fewer eggs compared to their WT counterparts during an eight-day monitoring period (Figure S6A). Accordingly, the cumulative egg count over eight days for mutants was significantly lower than that of WT ($P < 0.01$). Similar trends were observed when comparing virgin *Clk*-uORF-KO females to WT assessed over a 2-day interval or a total of 10 days (Figure S6B). We suspected this observed decline in egg-laying should lead to a corresponding decrease in progeny. Indeed, female *Clk*-uORF-KO flies produced significantly fewer offspring than WT counterparts at 25°C ($P = 1.5\times10^{-5}$ Wilcoxon rank sum test) (Figure S6C). Even at 29°C, when both mutants

and WT flies exhibit reduced offspring production compared to 25°C, the offspring counts for *Clk*-uORF-KO females were significantly fewer than that for WT females ($P$ = 0.00029, Wilcoxon rank sum test) (Figure S6C).

These findings suggest that *Clk* uORFs exert profound influences on physiology and behavior, extending beyond circadian rhythm and sleep regulation to impact fitness.

## Discussion

The evolutionary conservation of the circadian mechanisms across species, from *Drosophila* to mammals, underscores their fundamental significance. Our understanding of circadian rhythms has grown substantially over the years, providing critical insights into a range of physiological processes. Earlier investigations, including our own, have noted a very low correlation between the temporal variation pattern of the transcriptome and that of the proteome, indicating that regulations at the translational and post-translational level play dominant roles in determining the daily variations of protein abundance (*34*). Within this complex regulatory network, our research has highlighted a pivotal role for translational regulation, specifically mediated by uORFs, in fine-tuning these rhythms. Beyond circadian rhythm and sleep, *Clk*-uORF-KO flies show diminished fecundity and reduced starvation resilience, suggesting that *Clk* uORFs might act as a master regulator, shaping the rhythmic expression of a vast array of genes and influencing multifaceted physiological outcomes.

In this study, we observed over one hour shortening of circadian period when *Clk* uORFs are knocked out, which means these uORFs play a crucial role in the core circadian timing mechanism. Consistent with this idea, we found that *Clk* uORFs

mediate an instructive signal to the molecular clock by suppressing CLK protein translation during the daytime. Besides setting the pace of the clock, *Clk* uORFs also modulate cyclic gene expression globally. The oscillation of approximately half of the rhythmic genes requires *Clk* uORFs, as these oscillations are dampened in *Clk*-uORF-KO flies. This reveals the delicacy of the circadian system: an increase of CLK translation and thus CLK protein level caused by knocking out *Clk* uORFs can exert detrimental effects on the rhythmicity of the transcriptome. It is surprising that genes cycle with a more robust amplitude in *Clk*-uORF-KO flies relative to WT. Among these genes, about half do not display oscillatory expression pattern in WT. In other words, *Clk* uORFs suppress the mRNA rhythms of these genes. We reason that knocking out *Clk* uORFs increases CLK protein level in a time-dependent manner, which results in enhanced transcription of CLK target genes at a certain time of the day and thus these genes may acquire a "*de novo*" rhythmic expression pattern. Considering that *Clk* uORFs suppress CLK translation more under longer photoperiod and less under shorter photoperiod, we speculate this photoperiod-dependent variation in the translational activities of *Clk* uORFs can shape the rhythmic transcription landscape in adaptation to seasonal photoperiod changes.

*Clk* uORFs promote morning wakefulness, consistent with the time-dependent modulation of these uORFs on CLK protein translation and abundance. This regulation is not only time of the day-dependent, but also time of the year-dependent. WT flies display increased wakefulness and decreased sleep in adaptation to a shortening of photoperiod (occurring in fall and winter seasons), while *Clk* uORFs participate in modulating this process by suppressing CLK translation in a photoperiod-dependent manner. Intriguingly, human individuals with seasonal affective disorder (SAD, commonly known as winter depression) report a winter

increase in sleep, ranging from 30 min to 2 h longer in duration compared to controls, while the underlying mechanism is yet unknown (*35*). The long sleep phenotype caused by knocking out *Clk* uORFs is most striking under short photoperiod, reminiscent of the winter hypersomnolence in SAD. Notably, while we observed stronger sleep phenotype in female flies, SAD is also predominant in women(*36*). From a mechanistic perspective, our results strongly suggest that *Clk* uORFs promote wakefulness by increasing tyrosine hydroxylase expression, thus up-regulating the dopaminergic tone. Similarly, the mammalian CLOCK protein has been shown to repress the transcription of *Tyrosine hydroxylase*, while *Clock* mutation leads to decreased sleep and depression-like behaviors in mice due to elevated dopamine signaling (*37, 38*). Taken together, our findings here may provide some insights for understanding the pathological mechanism of SAD.

It seems contradictory that CLK translation is down-regulated under 12L12D vs. 4L20D, while CLK protein level is higher under 12L12D relative to 4L20D. We reason that CLK protein abundance is determined by coordinated actions of transcription, translation and post-translational modifications. Longer photoperiod may promote transcription and/or inhibit degradation of CLK protein, while *Clk* uORFs counteract these influences by suppressing CLK synthesis. In other words, *Clk* uORFs may serve as a buffer against the impacts of transcription and/or PTMs, adjusting CLK protein abundance to maintain an optimal level throughout the year. Of course, further investigations are needed to elucidate how *Clk* uORFs mediate photoperiodic signals.

uORFs have been shown to dynamically modulate translation of CDSs in stress responses, and here we also observed a role for *Clk* uORFs in responses to starvation stress (Figure S4)(*11*). Although under food *ad libitum* condition *Clk* uORFs promote

wakefulness, under starvation they function to reduce wakefulness, implicating the complexity of uORF regulation and their multifaceted roles in varying physiological states. Despite these complicated and sometimes contradictory actions, in the end the presence of *Clk* uORFs confers adaptive advantages and is beneficial for the survival (KO flies die earlier when starved) and fecundity of the organism.

To summarize, in this study we identified functional uORFs in *Clk* that repress translation of CLK protein in both circadian and circannual manner. These regulations slow down the pace of the clock, modulate global rhythms of gene expression and promote morning wakefulness in adaption to seasonal photoperiod changes. These findings, situated within the broader evolutionary, clinical, and interdisciplinary contexts, underscore the pivotal role of uORFs in shaping biological processes and potential therapeutic avenues.

**Acknowledgements**

This work was supported by grants from the Natural Science Foundation of China (31930021 and 32022035 to L.Z., 32070597 to J. L, and 32200956 to K.S.), the Ministry of Science and Technology of China (STI 2030-Major Projects 2021ZD0203200-02 to L.Z, and 2022YFE0132000 to J.L.), and Natural Science Foundation of Beijing (grant 5212006 to J.L.). We thank the National Center for Protein Sciences at Peking University for technical assistance. Some of the analyses were performed on the High-Performance Computing Platform of the Center for Life Sciences. We thank Dr. Joanna Chiu for CLK antibody used in this study.

**Declaration of interests**

The authors declare no competing interests.

## Data and code availability

- The raw fastq files for the RNA-seq libraries were deposited to NCBI Sequence Read Archive (SRA) with BioProject accession PRJNA1026336 and are publicly available as of the date of publication.
- Original Western blot images have been deposited to Mendeley and are publicly available on the date of publication.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

## References


1. S. Li, L. Zhang, Circadian Control of Global Transcription. *Biomed Res Int* **2015**, 187809 (2015).
2. P. E. Hardin, Molecular genetic analysis of circadian timekeeping in Drosophila. *Adv Genet* **74**, 141-173 (2011).
3. J. Lee, E. Yoo, H. Lee, K. Park, J. H. Hur, C. Lim, LSM12 and ME31B/DDX6 Define Distinct Modes of Posttranscriptional Regulation by ATAXIN-2 Protein Complex in Drosophila Circadian Pacemaker Neurons. *Mol Cell* **66**, 129-140 e127 (2017).
4. C. Lim, R. Allada, ATAXIN-2 activates PERIOD translation to sustain circadian rhythms in Drosophila. *Science* **340**, 875-879 (2013).
5. C. Lim, J. Lee, C. Choi, V. L. Kilman, J. Kim, S. M. Park, S. K. Jang, R. Allada, J. Choe, The novel gene twenty-four defines a critical translational step in the Drosophila clock. *Nature* **470**, 399-403 (2011).
6. Y. Zhang, J. Ling, C. Yuan, R. Dubruille, P. Emery, A role for Drosophila ATX2 in activation of PER translation and circadian behavior. *Science* **340**, 879-882 (2013).
7. R. J. Jackson, C. U. Hellen, T. V. Pestova, The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* **11**, 113-127 (2010).
8. H. Zhang, Y. Wang, X. Wu, X. Tang, C. Wu, J. Lu, Determinants of genome-wide distribution and evolution of uORFs in eukaryotes. *Nature Communications* **12**, 1076 (2021).
9. A. G. Hinnebusch, I. P. Ivanov, N. Sonenberg, Translational control by 5'-untranslated regions of eukaryotic mRNAs. *Science* **352**, 1413-1416 (2016).

10. D. R. Morris, A. P. Geballe, Upstream open reading frames as regulators of mRNA translation. *Mol Cell Biol* **20**, 8635-8642 (2000).

11. H. Zhang, Y. Wang, J. Lu, Function and Evolution of Upstream ORFs in Eukaryotes. *Trends Biochem Sci* **44**, 782-794 (2019).

12. C. Barbosa, I. Peixeiro, L. Romão, Gene expression regulation by upstream open reading frames and human disease. *PLoS Genet* **9**, e1003529 (2013).

13. J. Ruiz-Orera, M. M. Alba, Translation of Small Open Reading Frames: Roles in Regulation and Evolutionary Innovation. *Trends Genet* **35**, 186-198 (2019).

14. Y. Huang, J. A. Ainsley, L. G. Reijmers, F. R. Jackson, Translational profiling of clock cells reveals circadianly synchronized protein synthesis. *PLoS Biol* **11**, e1001703 (2013).

15. P. Janich, A. B. Arpat, V. Castelo-Szekely, M. Lopes, D. Gatfield, Ribosome profiling reveals the rhythmic liver translatome and circadian clock regulation by upstream open reading frames. *Genome Res* **25**, 1848-1859 (2015).

16. C. Jang, N. F. Lahens, J. B. Hogenesch, A. Sehgal, Ribosome profiling reveals an important role for translational control in circadian gene expression. *Genome Res* **25**, 1836-1847 (2015).

17. A. C. Diernfellner, T. Schafmeier, M. W. Merrow, M. Brunner, Molecular mechanism of temperature sensing by the circadian clock of Neurospora crassa. *Genes Dev* **19**, 1968-1973 (2005).

18. H. W. Wu, E. Fajiculay, J. F. Wu, C. S. Yan, C. P. Hsu, S. H. Wu, Noise reduction by upstream open reading frames. *Nat Plants* **8**, 474-480 (2022).

19. T. Miyake, Y. Inoue, X. Shao, T. Seta, Y. Aoki, K. T. Nguyen Pham, Y. Shichino, J. Sasaki, T. Sasaki, M. Ikawa, Y. Yamaguchi, H. Okamura, S. Iwasaki, M. Doi, 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 (2023).

20. A. Millius, R. G. Yamada, H. Fujishima, K. Maeda, D. M. Standley, K. Sumiyama, D. Perrin, H. R. Ueda, Circadian ribosome profiling reveals a role for the Period2 upstream open reading frame in sleep. *Proc Natl Acad Sci U S A* **120**, e2214636120 (2023).

21. H. Zhang, Y. Wang, X. Wu, X. Tang, C. Wu, J. Lu, Determinants of genome-wide distribution and evolution of uORFs in eukaryotes. *Nat Commun* **12**, 1076 (2021).

22. H. Zhang, S. Dou, F. He, J. Luo, L. Wei, J. Lu, Genome-wide maps of ribosomal occupancy provide insights into adaptive evolution and regulatory roles of uORFs during Drosophila development. *PLoS Biol* **16**, e2003903 (2018).

23. K. Abdelmohsen, A. C. Panda, M. J. Kang, R. Guo, J. Kim, I. Grammatikakis, J. H. Yoon, D. B. Dudekula, J. H. Noh, X. Yang, J. L. Martindale, M. Gorospe, 7SL RNA represses p53 translation by competing with HuR. *Nucleic Acids Res* **42**, 10099-10111 (2014).

24. A. C. Panda, K. Abdelmohsen, J. L. Martindale, C. Di Germanio, X. Yang, I. Grammatikakis, J. H. Noh, Y. Zhang, E. Lehrmann, D. B. Dudekula, S. De, K. G. Becker, E. J. White, G. M. Wilson, R. de Cabo, M. Gorospe, Novel RNA-binding activity of MYF5 enhances Ccnd1/Cyclin D1 mRNA translation during myogenesis. *Nucleic Acids Res* **44**, 2393-2408 (2016).

25. A. C. Panda, J. L. Martindale, M. Gorospe, Polysome Fractionation to Analyze mRNA Distribution Profiles. *Bio Protoc* **7**, (2017).

26. H. M. Fathallah-Shaykh, J. L. Bona, S. Kadener, Mathematical model of the Drosophila circadian clock: loop regulation and transcriptional integration. *Biophys J* **97**, 2399-2408 (2009).

27. K. C. Abruzzi, J. Rodriguez, J. S. Menet, J. Desrochers, A. Zadina, W. Luo, S. Tkachev, M. Rosbash, Drosophila CLOCK target gene characterization: implications for circadian tissue-specific gene expression. *Genes Dev* **25**, 2374-2386 (2011).

28. M. E. Hughes, J. B. Hogenesch, K. Kornacker, JTK_CYCLE: an efficient nonparametric algorithm for detecting rhythmic components in genome-scale data sets. *J Biol Rhythms* **25**, 372-380 (2010).

29. S. Liu, A. Lamaze, Q. Liu, M. Tabuchi, Y. Yang, M. Fowler, R. Bharadwaj, J. Zhang, J. Bedont, S. Blackshaw, T. E. Lloyd, C. Montell, A. Sehgal, K. Koh, M. N. Wu, WIDE AWAKE mediates the circadian timing of sleep onset. *Neuron* **82**, 151-166 (2014).

30. S. Kumar, D. Chen, A. Sehgal, Dopamine acts through Cryptochrome to promote acute arousal in Drosophila. *Genes Dev* **26**, 1224-1234 (2012).

31. R. G. Pendleton, A. Rasheed, P. Paluru, J. Joyner, N. Jerome, R. D. Meyers, R. Hillman, A developmental role for catecholamines in Drosophila behavior. *Pharmacol Biochem Behav* **81**, 849-853 (2005).

32. A. C. Keene, E. R. Duboue, D. M. McDonald, M. Dus, G. S. Suh, S. Waddell, J. Blau, Clock and cycle limit starvation-induced sleep loss in Drosophila. *Curr Biol* **20**, 1209-1215 (2010).

33. S. Wood, A. Loudon, Clocks for all seasons: unwinding the roles and mechanisms of circadian and interval timers in the hypothalamus and pituitary. *J Endocrinol* **222**, R39-59 (2014).

34. C. Wang, K. Shui, S. Ma, S. Lin, Y. Zhang, B. Wen, W. Deng, H. Xu, H. Hu, A. Guo, Y. Xue, L. Zhang, Integrated omics in Drosophila uncover a circadian kinome. *Nat Commun* **11**, 2710 (2020).

35. D. L. Wescott, A. M. Soehner, K. A. Roecklein, Sleep in seasonal affective disorder. *Curr Opin Psychol* **34**, 7-11 (2020).

36. T. L. Jepson, M. E. Ernst, M. W. Kelly, Current Perspectives on the Management of Seasonal Affective Disorder. *Journal of the American Pharmaceutical Association (1996)* **39**, 822-829 (1999).

37. K. Roybal, D. Theobold, A. Graham, J. A. DiNieri, S. J. Russo, V. Krishnan, S. Chakravarty, J. Peevey, N. Oehrlein, S. Birnbaum, M. H. Vitaterna, P. Orsulak, J. S. Takahashi, E. J. Nestler, W. A. Carlezon, Jr., C. A. McClung, Mania-like behavior induced by disruption of CLOCK. *Proc Natl Acad Sci U S A* **104**, 6406-6411 (2007).

38. M. M. Sidor, S. M. Spencer, K. Dzirasa, P. K. Parekh, K. M. Tye, M. R. Warden, R. N. Arey, J. F. Enwright, J. P. Jacobsen, S. Kumar, E. M. Remillard, M. G. Caron, K. Deisseroth, C. A. McClung, Daytime spikes in dopaminergic activity drive rapid mood-cycling in mice. *Mol Psychiatry* **20**, 1479-1480 (2015).

39. X. Ren, Z. Yang, J. Xu, J. Sun, D. Mao, Y. Hu, S. J. Yang, H. H. Qiao, X. Wang, Q. Hu, P. Deng, L. P. Liu, J. Y. Ji, J. B. Li, J. Q. Ni, Enhanced specificity and efficiency of the CRISPR/Cas9 system with optimized sgRNA parameters in Drosophila. *Cell Rep* **9**,

1151-1162 (2014).

40. J. Q. Ni, R. Zhou, B. Czech, L. P. Liu, L. Holderbaum, D. Yang-Zhou, H. S. Shim, R. Tao, D. Handler, P. Karpowicz, R. Binari, M. Booker, J. Brennecke, L. A. Perkins, G. J. Hannon, N. Perrimon, A genome-scale shRNA resource for transgenic RNAi in Drosophila. *Nat Methods* **8**, 405-407 (2011).

41. S. Sisobhan, C. Rosensweig, B. C. Lear, R. Allada, SleepMat: a new behavioral analysis software program for sleep and circadian rhythms. *Sleep* **45**, (2022).

42. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).

43. S. Anders, P. T. Pyl, W. Huber, HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169 (2015).

44. C. Gene Ontology, The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res* **49**, D325-D334 (2021).

## STAR Methods

## RESOURCE AVAILABILITY

### Lead contact

Further information and requests for resources and reagents should be directed to the lead contact, Luoying Zhang (zhangluoying@hust.edu.cn).

### Materials availability

All fly lines generated in this study are available from the Lead Contact.

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

All flies were reared on standard cornmeal-yeast-sucrose medium and kept in light/dark (LD) cycles at 25 °C except for specific experimental design denoted in the text. Food was accessible *ad libitum*. *{nos-Cas9}attP40* and *{nos-Cas9}attP2* fly stocks used for embryo injection are from the TsingHua Fly Center. The tool fly stocks (*y sc v* and *Dr, e/TM3, Sb*) used for mutant screening are from the TsingHua Fly Center. *Clk*-uORF-KO and *cyc*-uORF-KO were generated and backcrossed to the Canton-S strain for 9 generations. *Drosophila* S2 cells used in this study were purchased from Life Technologies Corp (http://www.lifetechnologies.com).

## METHOD DETAILS

### Generation of uORF-KO strain

We searched for possible sgRNA target sites near the start codon of the uORFs of *Clk* and *cyc* genes using the Benchling website (https://www.benchling.com/crispr/) to design optimal single guide RNA (sgRNA) sequences with high specificity and low off-target effects. Two pairs of sgRNAs were used for each gene. We then synthesized single-stranded complementary DNAs (ssDNAs) and annealed them to obtain

double-stranded DNA (dsDNA), which served as the template for sgRNA expression. The dsDNA was then ligated into the BbsI-digested pU6B vector. The template sequences of the sgRNA uATG-KO are listed in Key resources table.

To achieve single base editing (ATG to ATT,) by CRISPR/Cas9-mediated Homology-Directed Repair (HDR), a plasmid DNA donor with about 2 kilobase flanking the targeted site were constructed following a previous protocol (*39*). The donors and pU6B-sgRNA plasmid were co-injected into the embryos of transgenic Cas9 flies collected within one hour of laying at the TsingHua Fly Center as described in Ni *et al*. (*40*). The injected embryos were kept at 25°C and 60% humidity until adulthood (G0). The G0 adult flies that hatched from the injected embryos were individually crossed with another strain (*y sc v*) to increase the number of offspring. Then, the F1 progeny were crossed with flies carrying an appropriate balancer (*Dr, e/TM3, Sb*). After F2 spawning, the F1 individuals were screened for mutations of interest by genotyping. The primers used for genotyping are listed in Key resources table. The F2 progeny whose parents showed positive genotyping results were then screened for the $y^-$ allele to separate the chromosome carrying *nos-Cas9*. The screened F2 males were crossed with flies containing the same balancer as mentioned above. After F2 genotyping, the progeny (F3) of positive F2 individuals carrying the same mutation were crossed individually to generate homozygous mutants in the F4 generation. The original homozygous mutants were backcrossed with *Canton-S* flies for 9 generations.

**uORF mutation sequencing validation**

Mutant fly was homogenized with DNA sequencing extract buffer [10 mM Tris-HCl at pH 8.0, 1 mM EDTA at pH 8.0, 25 mM NaCl and 0.02 % proteinase K (CW2584,

CWBIO)] and incubated at 37 °C for 30 min. PCR was performed with Taq Plus MasterMix (CW2849, CWbiotech). The PCR reaction was performed as follows: 94 °C for 2 min followed by 94 °C for 10 s, 57 °C for 15 s and 72 °C for 1 min 20 s for 35 cycles. The primers used were listed in Key resources table. PCR products were sequenced using commercial Sanger sequencing (AuGCT Co.).

**Cell culture and transfection**

S2 cells were cultured in Schneider's *Drosophila* Medium (Sigma) containing 10% heat-inactivated fetal bovine serum (FBS), 100 U/ml penicillin and 100 μg/ml streptomycin (Thermo Fisher) at 25°C for 24 h to reach 2-4 × $10^6$ cells/ml before further treatments. Plasmid transfection was conducted with Lipofectamine 3000 (L3000001, Thermo Fisher) according to the supplier's protocol.

**Dual-luciferase assay**

The wild-type (WT) 5'UTR of *Clk* and *cyc* was cloned from cDNA by PCR, and uATG mutations were introduced into 5'UTR by the amplification primers. The PCR primers used for this purpose are listed in Key resources table. The WT and mutated 5'UTR sequences were ligated into a linearized reporter plasmid (psiCHECK-2 vector, Promega). The entire sequence of all the plasmids were validated by Sanger sequencing. The *Renilla* luciferase activity associated with WT or uORF-mutated 5'UTRs was measured according to the manual of the Dual-Luciferase Reporter Assay System (Promega) 32 h after transfection and were normalized to the activity of firefly luciferase.

**Fly head collection**

Flies were collected within 3-7 days of eclosion and entrained in LD for 3 days. After that, they were transferred into constant darkness (DD). Flies were collected during LD or on the first day of DD and frozen at -80 °C. Frozen flies were vortexed for 10 s to separate the head from the body.

**Ribosome fraction analysis**

The heads obtained from WT and mutant flies were collected and homogenized in a Dounce homogenizer with lysis buffer [50 mM Tris pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 1% Triton X-100, 2 mM dithiothreitol (DTT), 20 U/ml SuperaseIn (Ambion), 0.5 tablets of proteinase inhibitor (Roche), 100 μg/ml emetine (Sigma Aldrich), and 50 μM guanosine 5′-[β,γ-imido]triphosphate trisodium salt hydrate (GMP-PNP) (Sigma Aldrich)] at 4°C. The lysates were clarified by centrifugation at 4°C and 20,000×g for 10 min, and the supernatants were transferred to new 1.5 ml tubes. 10-45% sucrose gradients were prepared in buffer (250 mM NaCl, 50 mM Tris pH 7.5, 15 mM $MgCl_2$, 0.5 mM DTT, 12 U/ml RNaseOUT, 0.5 tablets of protease inhibitor, and 20 μg/ml emetine) using a Gradient Master (Biocomp Instruments) in ULTRA-CLEAR Thinwall Tubes (Beckman Coulter). A sample volume of up to 500 μl was applied to the top of each gradient. After ultracentrifugation with a Hitachi P40ST rotor at 35,000 × rpm for 3 h at 4°C, the monosome and polysome fractions were collected, flash-frozen in liquid nitrogen, and stored at -80°C until further use.

The RNA in the monosome and polysome fractions was extracted separately using TRIzol reagent (Life Technologies, Inc.) and chloroform (Beijing Chemical Works) following the manufacturer's instructions and were reverse transcribed into cDNA using the PrimeScript™ II 1st Strand cDNA Synthesis Kit (Takara). RT-qPCR analysis of *Clk* and *cyc* cDNA was performed using PowerUp™ SYBR™ Green

Master Mix (Thermo Fisher) following the manufacturer's instructions. The primer sequences employed for RT-qPCR are listed in Key resources table. For each sample, the ratio of *Clk* and *cyc* mRNA abundance in the polysome fraction to that in the monosome fraction was calculated as the P-to-M ratio. Six biological replicates were performed for each sample.

**Western blot**

Proteins were extracted from fly heads and cultured cells using SDS lysis buffer [50 mM Tris-HCl at pH 7.6, 150 mM NaCl, 1% Triton X-100, 0.5% SDS, 1 mM EDTA at pH8.0, 2 mM DTT, 1 × proteinase inhibitor and phosphatase inhibitor (4693116001 and 04906845001, Roche)]. After homogenization, protein lysates were centrifuged at 12,000 × *g* for 15 min at 4 °C and incubated at 95 °C in loading buffer for 5 min. Equal amounts of protein were loaded into each well on 6% SDS-PAGE gels and then transferred to nitrocellulose membranes for 2 h at 90 V. Membranes were incubated with primary antibody at 4 °C overnight, followed by secondary antibody at room temperature for 1 h. The primary antibodies used were as follows: guinea pig anti-CLK (1:1000, gift from Dr. Joanna Chiu) and rabbit anti-ACTIN 57B (1:5000, AC026, ABclonal, CN). Donkey secondary antibodies (1:10000 dilution) were conjugated either with IRDye 680 or IRDye 800 (926-68072 and 926-32213, LI-COR Biosciences, US). Bands were visualized and quantified with an Odyssey Infrared Imaging System (LI-COR Biosciences).

**Model simulation for the outcome of enhanced *Clk* translation**

We adopted a previously published mathematical model of the *Drosophila* circadian clock (*26*). This system relies on a nonlinear, autonomous, first-order system of

ordinary differential equations to simulate the regulatory weights among mRNAs or proteins molecules in the molecular clock network. Briefly, assuming that genes or proteins $j \in \{1,\ldots, n\}$ regulate the generation of gene/protein *i*, the subsequent type of differential equations as a general model were employed:

$$\frac{dx_i(t)}{dx} = \rho_i g\left(\sum_{j=1}^{n} \lambda_{ji}\, x_j(t) - \delta_i x_i(t)\right) x_i(t)\big(s_i - x_i(t)\big), 1 \le j \le n, \qquad (1)$$

where the state vector $x_i$ represents the concentration of molecule *i* at its site of action. The parameters $\lambda_{ji}$ represents regulatory weights that indicate the influence of molecule *j* on the production rate of molecule *i*. Positive and negative values of $\lambda_{ji}$ are indicative of activating or repressing effects, respectively. The larger absolute value of $\lambda_{ji}$, the stronger effect. It is worth noting that $\lambda_{Clk\,\rightarrow\mathrm{CLK}}$ in the model is applicable to the regulation of *Clk* mRNA to CLK protein (i.e., translation efficiency) in our study. To mimic the enhanced CLK translation caused by uORFs removal, we increased the original $\lambda_{Clk\,\rightarrow\mathrm{CLK}}$ from 0.026 (baseline) to three larger values: 0.027, 0.028 and 0.030, respectively.

An odd sigmoid function modulates the cumulative regulatory influences, g: $\mathbb{R}\rightarrow\mathbb{R}$, of the form

$$g(\mu) = \frac{\mu}{\sqrt{1+\mu^2}} = tanh\left(ln\,(\mu + \sqrt{1+\mu^2})\right),$$

together with a parameter $\rho_i > 0$ that indicates the maximum rate of *i* production. The model incorporates logistic terms $x_i(s_i - x_i)$, where constants $s_i \ge 0$ indicate the saturation level of molecule *i*. The real parameter $\delta_i$ is the decay rate of *i*.

Simulations were performed on MATLAB (The MathWorks, Natick, MA). The system of ordinary differential equations is solved numerically by the subroutine "ode45". The actual equations and parameters are the same as those provided in the original model (*26*). During each simulation cycle, the mRNA and protein levels of each molecule in the circadian network was recoded. We calculated the time intervals between the peaks of *tim* mRNA, serving as the proxy of circadian period length under the different regulatory weights.

## Total RNA extraction and RT-qPCR

RNA was extracted from frozen fly heads maintained at -80 °C. 50 fly heads per sample were homogenized in Total RNA Isolation (TRIzol) Reagent (15596026, Invitrogen), by using a homogenizer with plastic beads. After mixing with trichloromethane, homogenates were centrifuged at 12,000 × *g* and suspension was precipitated with 75% ethanol. After air dry, total RNA was reverse-transcribed into cDNA and genomic DNA was removed with One-step cDNA Synthesis SuperMix (AT311, Transgen). Quantitative real-time PCR was performed with One-Step RT-PCR SuperMix (AT141, Transgen). The PCR reaction was performed as follows: 45 °C for 5 min; 94 °C for 2 min; 94 °C for 5 s, 60 °C for 15 s, 72 °C for 20 s for 40 cycles (Applied Biosystems). The ΔΔCT method was used for quantification. *Beta-Actin* was used as internal control. The primers used are listed in Key resources table.

## Locomotor activity monitoring and analysis

Locomotor activity levels of adult flies were monitored by *Drosophila* Activity Monitoring System (TriKinetics) for 7 days of LD followed by 7 days of DD. For DD rhythmicity, chi-squared periodogram analyses were performed by Clocklab (Actimetrics). Rhythmic flies were defined as those in which the chi-squared power was ≥ 10 above the significance line. Period calculations considered only rhythmic flies. Dead flies were defined by 0 activity on DD7 and removed from the analysis.

## Sleep measurement and analysis

*Drosophila* Activity Monitor system (Trikinetics) was used to analyze fly sleep.

Virgin female and male flies of 3~5 days old were used for the experiments unless otherwise specified. Flies were entrained under LD at 25°C for 3 days, and then their activities in the next 3 days under LD condition were analyzed.

For sleep analysis during starvation, 3~5 days old virgin females were used. After 3 days of entrainment in monitoring tubes with standard media (2% agar and 5% sucrose) baseline sleep was measured for 2 days. Standard food was then replaced at ZT0 by starvation media (2% agar) for 24 h. Flies were then re-fed with standard media for an additional 24 h recovery period.

Sleep is defined as 5 min consecutive inactivity. Sleep was analyzed with SleepMat following previously published protocol (*41*).

**RNA-seq and circadian analysis**

We collected heads from *Clk*-uORF-KO mutants and WT flies raised under LD cycles at 25°C at ZT 0, 4, 8, 12, 16 and 20. Library construction and sequencing with PE150 were conducted by Annoroad on the Illumina Nova6000 platform. Three biological replicates were sequenced for each sample. The clean data were mapped to the reference genome of *D. melanogaster* (FlyBase, r6.04) using STAR (*42*). Reads mapped to the exons of each gene were tabulated with htseq-count (*43*). Rhythmically expressed genes were identified by JTK_CYCLE under the default threshold of $P < 0.05$.

**GO-based enrichment**

For GO-based enrichment analysis, GO annotation files were downloaded on 18 November 2020 from the Gene Ontology Resource (http://geneontology.org/)(*44*) and contained 12,616 fly genes with at least one GO term. For each GO term *t*, we defined the following:

*N* = number of genes annotated by at least one GO term.

*n* = number of genes annotated by GO term *t*.

*M* = number of genes annotated by at least one GO term.

*m* = number of genes annotated by GO term *t*.

The enrichment ratio (E-ratio) of *t* was then computed, and the *P* was calculated with the hypergeometric distribution as follows:

$$\mathrm{E-ratio} = \frac{m}{M} / \frac{n}{N} \#(1)$$

$$P = \sum_{m'=m}^{n} \frac{\binom{M}{m'}\binom{N-M}{n-m'}}{\binom{N}{n}} (\mathrm{E-ratio} \geq 1), \mathrm{or} \#(2)$$

$$P = \sum_{m'=0}^{m} \frac{\binom{M}{m'}\binom{N-M}{n-m'}}{\binom{N}{n}} (\mathrm{E-ratio} < 1) \#(3)$$

**Drug treatment**

For pharmacological experiments, 5 mg/ml L-DOPA (BBI Life Sciences) was mixed in the fly food and the flies were fed with L-Dopa food during sleep monitoring.

**Egg numbers quantification**

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. Subsequently, ten female parents from each group were transferred to a fresh dish containing a grape juice-based medium overlaid with yeast. This procedure was replicated in five separate dishes for each strain.

After a one day of egg laying, the number of eggs deposited was meticulously documented. Following this assessment, the female parents were relocated to a new dish. This egg quantification process was conducted over a total duration of eight days. For the virgin female, egg counts were performed every two days, extending over a span of ten days.

**Quantification of offspring number per female fly**

Newly hatched virgins were picked out and allowed to mature for two days in individual 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 and the offspring number produced in 10 days was measured 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 starvation medium (1.5% agar), with 10 flies per vial and 10 vials for both males and females from each strain. We observed every 6 h or 12 h to count the number of deaths under starvation conditions until all flies starved to death. The survival curves were plotted by ggsurvplot package in R.

**Quantification and statistical analysis**

Quantifications in all data graphs represent the mean of at least four biological replicates, and error bars represent standard error of the mean (SEM). Two-tailed Mann-Whitney *U* test, Wilcoxon sign rank test, log-rank test and two-tailed Student's *t* test were used to calculate significance. All statistical analysis was carried out using *Statskingdom*. In each statistical analysis, number of biological replicates and significant *P* values were noted in corresponding figure legends.

## Figures

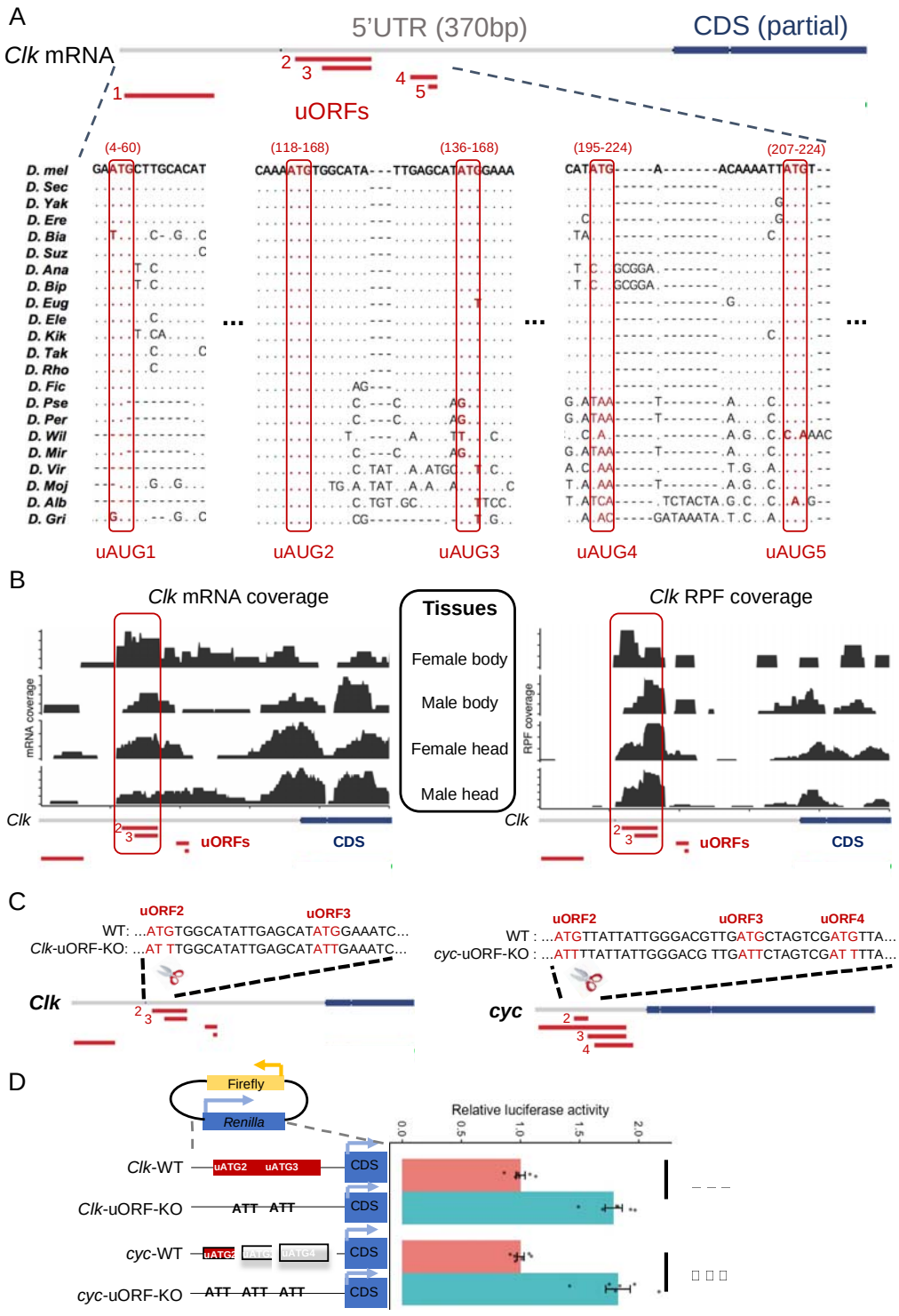



**Figure 1. uORFs suppress CLK and CYC protein translation.**

(A) Multiple sequence alignment (MSA) of *Clk* uORFs among 22 *Drosophila* species.

The start codons (uAUGs) of uORFs are highlighted by red boxes. The position schemes of uORFs and partial CDSs are denoted above the MSA, with red and blue color, respectively. The start and ending positions of each uORF (separated by a "-") are given in the parenthesis above each uAUG.

(B) The mRNA reads coverage (left) and ribosome-protected footprints (RPF) coverage (right) of uORFs of *Clk* mRNA from head and body. The position schemes of uORFs and partial CDSs were denoted at the bottom.

(C) Schematics showing the genotypes of WT and two uORF knock-out strains (*Clk*-uORF-KO and *cyc*-uORF-KO) generated with CRISPR-Cas9 technology. The position schemes of uORFs and partial CDSs were denoted below the DNA sequences. The uORFs and CDS were denoted by red and blue boxes, respectively.

(D) Dual-luciferase assay for WT 5'UTR and uORF-KO 5'UTR of *Clk* and *cyc* gene respectively. The reporter constructs are illustrated on the top left. The relative activity of *Renilla* luciferase was normalized to that of firefly luciferase.

Data are presented as the mean ± SEM. Asterisks indicate statistical significance ($n$ = 6. Two-tailed Student's $t$ test. ***, $P < 0.001$).

See also Figure S1

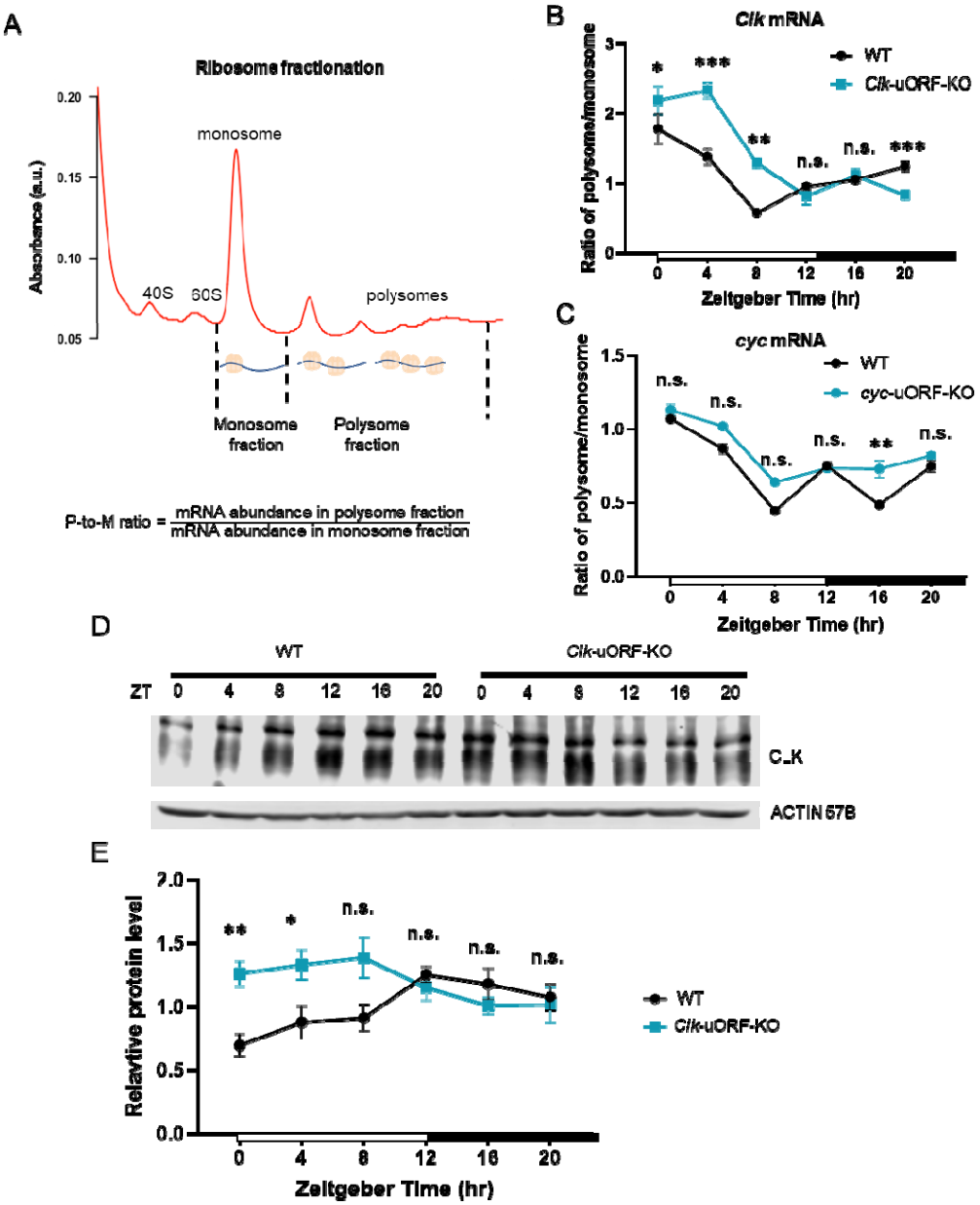



**Figure 2. *Clk* uORFs suppress CLK protein translation *in vivo*.**

(A) Schematic showing the monosome and polysome fraction separated in a sucrose density gradient. Relative RNA abundance in the monosome and polysome fractions (P-to-M ratio) was quantified by real-time qPCR after reverse transcription. The P-to-M ratio was calculated by comparing mRNA abundance in the polysome fraction to that in monosome fraction.

(B-C) P-to-M ratios of *Clk* mRNA (B) and *cyc* mRNA (C) from whole-head extracts

in *Clk*-uORF-KO and WT flies. ($n$ = 6. Wilcoxon sign rank test for unpaired comparisons.)

(D) Representative Western blots for CLK protein from whole-head total protein extracts collected at the indicated time points. ACTIN 57B was used as a loading control.

(E) The relative CLK abundance normalized to ACTIN 57B in (D). The average intensity of WT is set to 1. ($n$ = 4. Two-tailed Mann-Whitney $U$ test for unpaired comparisons.)

Data are presented as the mean ± SEM. Asterisks indicate statistical significance (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; n.s., $P > 0.05$). ZT, Zeitgeber Time.

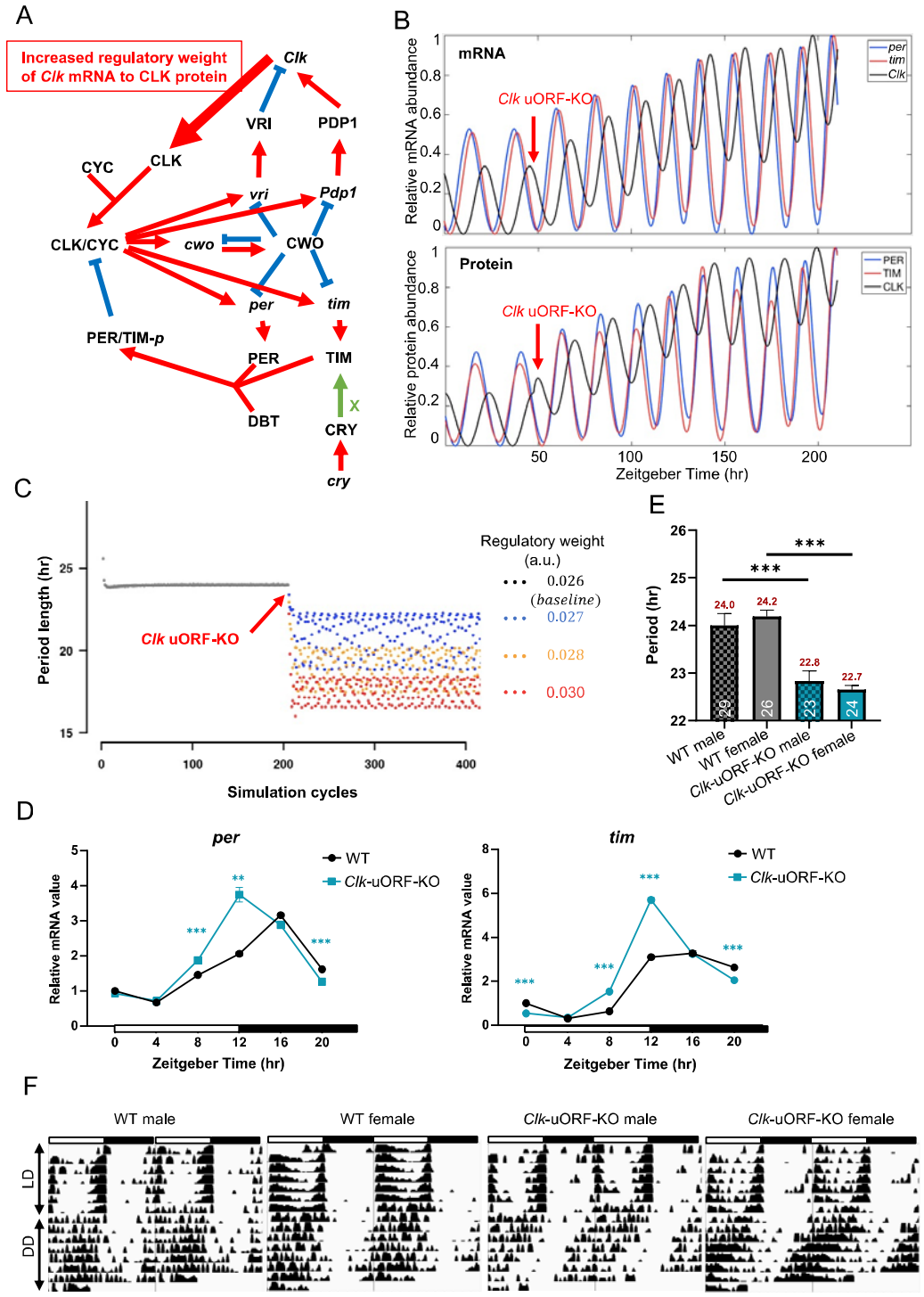



**Figure 3. C*lk* uORFs regulate core clock gene expression and circadian period.** (A) The molecular network of *Drosophila* circadian clock, adapted from a previous study (*26*). The mRNA molecules are denoted by italicized lowercase letters, and protein molecules are denoted by uppercase letters. Red arrows and blue lines indicate

up-regulation and down-regulation, respectively. The bold red arrow from *Clk* to CLK represents the increased regulatory weight from *Clk* mRNA to CLK protein, representing the elevation of CLK translation resulting from the knockout of *Clk* uORFs. The green arrow ending in an 'X' denotes that the CRY protein stimulates the degradation of TIM.

(B) The relative mRNA and protein abundance of *Clk*, *per*, and *tim* from model simulations (A). The time point of increase of *Clk* mRNA to protein regulatory weight is indicated by the red arrow.

(C) The distribution of circadian period (hr) under the indicated regulatory weight of *Clk* mRNA to CLK protein in about 400 simulation cycles. 0.026 is the original setting of regulatory weight in the model (*26*) and serves as the baseline to mimic the presence of WT uORFs. The time point of increase of *Clk* mRNA to protein regulatory weight is indicated by the red arrow.

(D) The mRNA levels of *per* and *tim*, measured by qRT-PCR from whole-head total RNA extracts, collected at the indicated time points ($n$ = 4 for each time point). The average intensity of WT at ZT0 is set to 1.

(E) Period of locomotor rhythms of *Clk*-uORF-KO flies and WT flies under constant darkness conditions (DD). The period length (hr) and number of tested flies ($n$) for each strain are shown at the top and bottom of each column.

(F) Representative double-plotted actograms of the locomotor activity of the indicated flies, monitored under LD followed by DD. White and black bars represent light and dark phases, respectively. Data are presented as mean ± SEM. Asterisks indicate statistical significance (two-tailed Student's $t$ test. **, $P < 0.01$; ***, $P < 0.001$).

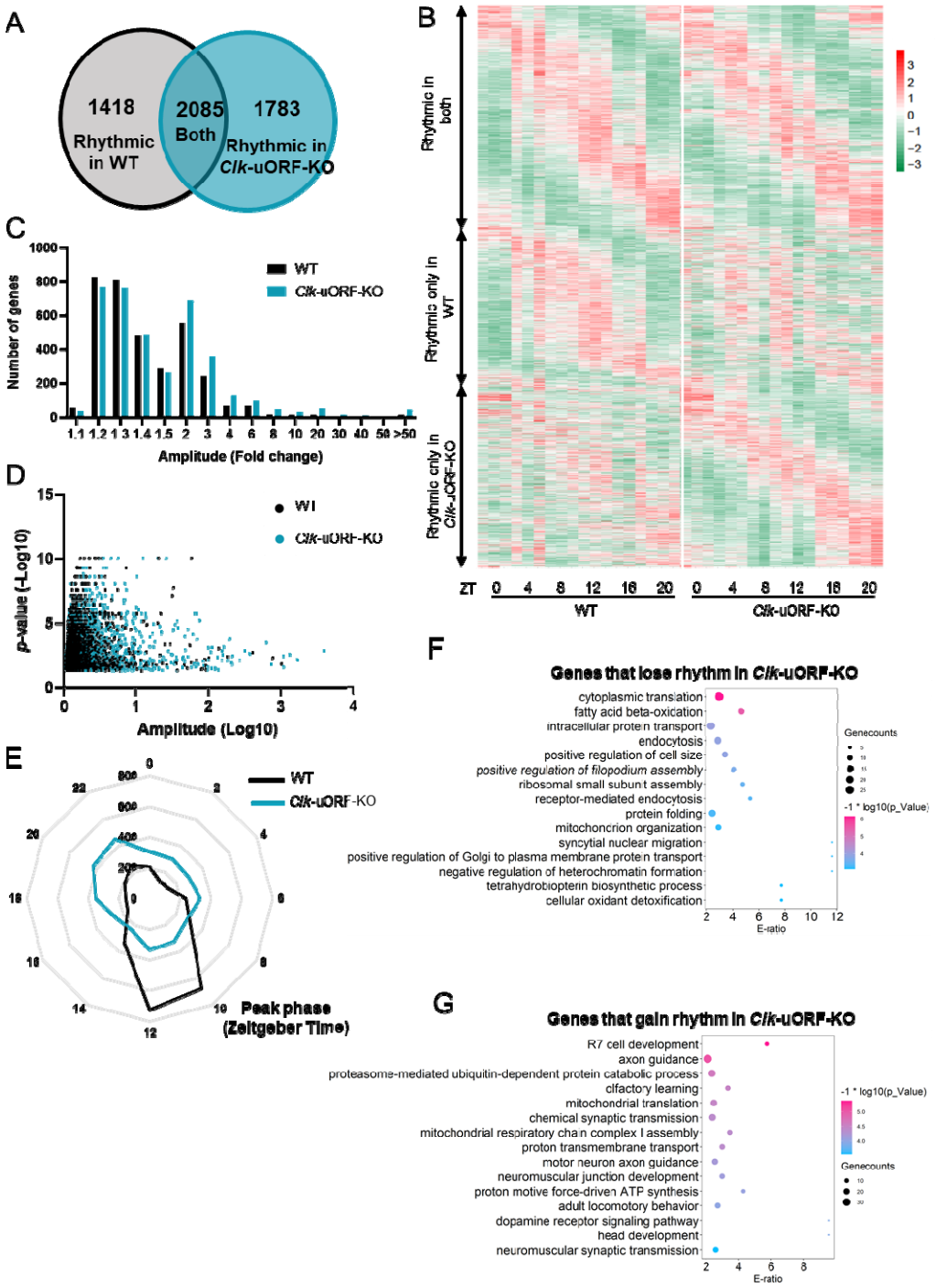



**Figure 4. *Clk* uORFs regulates the temporal expression pattern of the transcriptome.**

(A) Number of rhythmically expressed genes identified from RNA-seq samples of

WT and *Clk*-uORF-KO fly heads.

(B) Hierarchical clustering of rhythmically expressed genes ordered by the phase of oscillation. Expression values of genes are color-coded based on their intensities. Colors indicate low (green) and high (red) Z-scores.

(C) Amplitude distribution of rhythmically expressed genes.

(D) Volcano plot showing the statistical significance of rhythmically expressed genes and their corresponding amplitude.

(E) Phase distribution of rhythmically expressed genes. The *y*-axis of the radar chart indicates the number of rhythmic genes.

(F and G) GO-based enrichment analysis of genes rhythmically expressed only in WT (F) or *Clk*-uORF-KO flies (G).

ZT denotes Zeitgeber Time.

See also Figure S2.

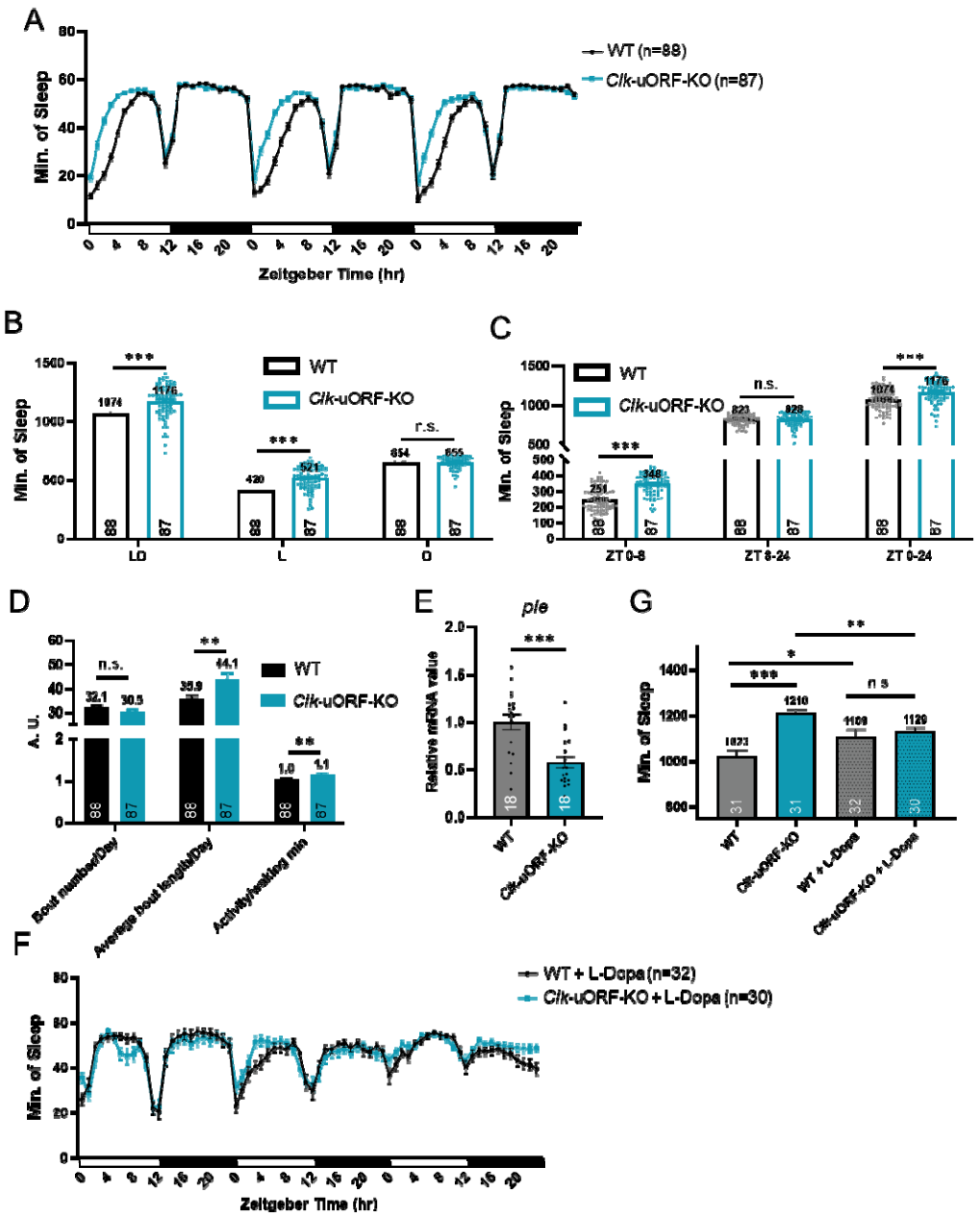



**Figure 5. *Clk* uORFs promote morning wakefulness by up-regulating dopaminergic tone.**

(A) Sleep profile of female *Clk*-uORF-KO and WT flies under LD condition.

(B) Sleep duration during daytime (L), nighttime (D) and the entire day (LD) of female *Clk*-uORF-KO and WT flies.

(C) Sleep duration during ZT0-8, ZT8-24 and ZT0-24 of female *Clk*-uORF-KO and

WT flies under LD condition.

(D) Daily sleep bout number, average bout length and average waking activity of female *Clk*-uORF-KO and WT flies under LD condition.

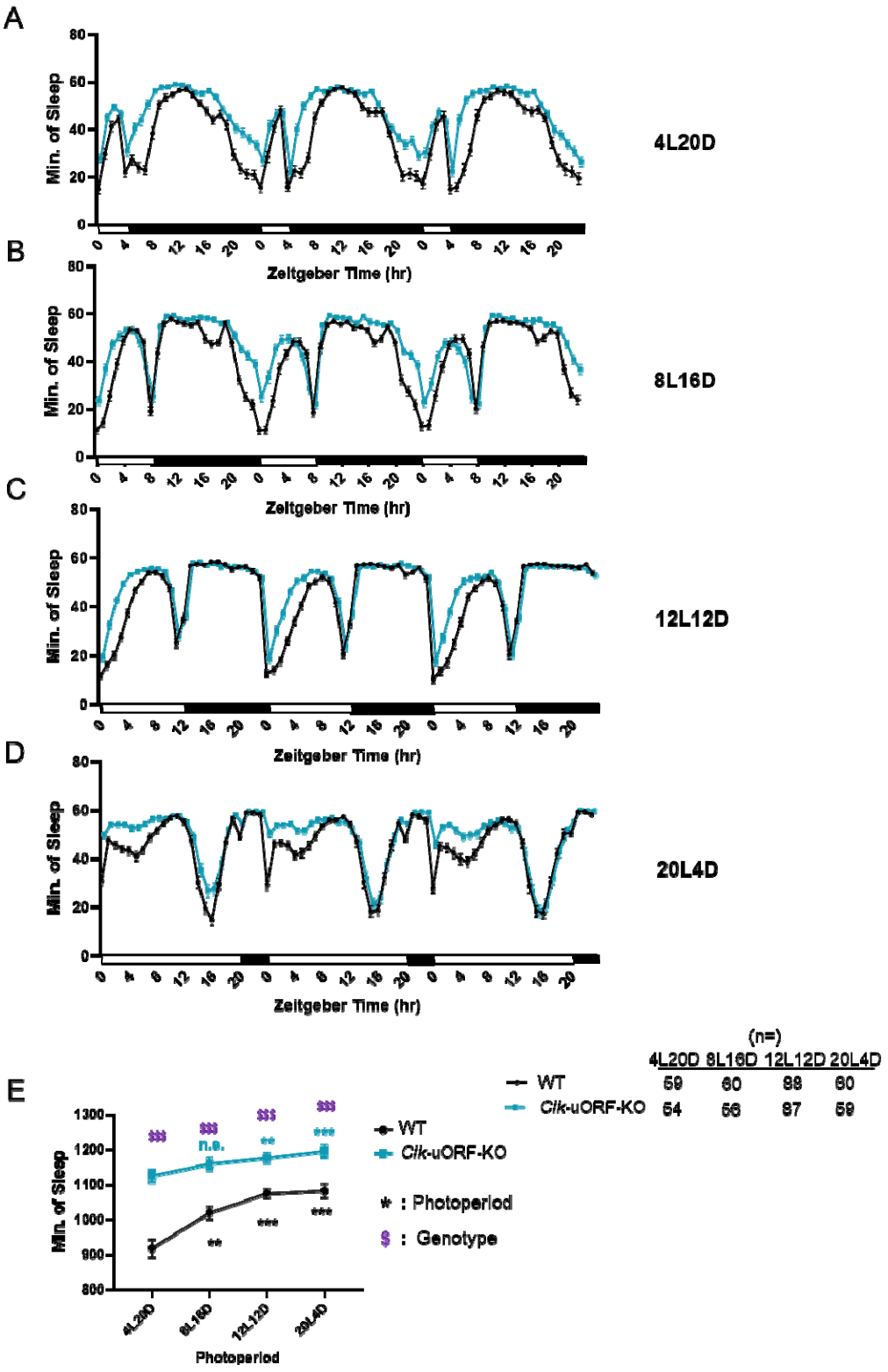
(E) The relative mRNA abundance of *ple* determined by RT-qPCR using whole-head total RNA extracts collected from female flies under LD condition. The average value of WT is set to 1.

(F) Sleep profile of female *Clk*-uORF-KO and WT flies treated with L-Dopa under LD condition.

(G) Daily sleep duration of female *Clk*-uORF-KO *vs*. WT flies treated with L-Dopa.

The numbers of flies tested are shown in the brackets and histograms. Data are presented as the mean ± SEM. Asterisks indicate statistical significance (two-tailed Student's *t* test. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; n.s., $P > 0.05$). ZT, Zeitgeber Time.

See also Figure S3 and S4.




885

886 **Figure 6. *Clk* uORFs promote wakefulness in adaptation to seasonal photoperiod**

887 **changes.**

(A-D) Sleep profile of female *Clk*-uORF-KO and WT flies under 4 h L : 20 h D (4L20D) (A), 8h L : 16 h D (8L16D) (B), 12 h L : 12 h D (12L12D) (C) and 20 h L : 4 h D (20L4D) (D) condition.

(E) Daily sleep duration of female *Clk*-uORF-KO and WT flies under different photoperiod.

The numbers of flies tested (n) are shown. Data are presented as the mean ± SEM. Asterisks and dollar signs indicate statistical significance (two-tailed Student's *t* test. **, $P < 0.01$; ***, $P < 0.001$; $$$, $P < 0.001$; n.s., $P > 0.05$). ZT, Zeitgeber Time.

See also Figure S5.

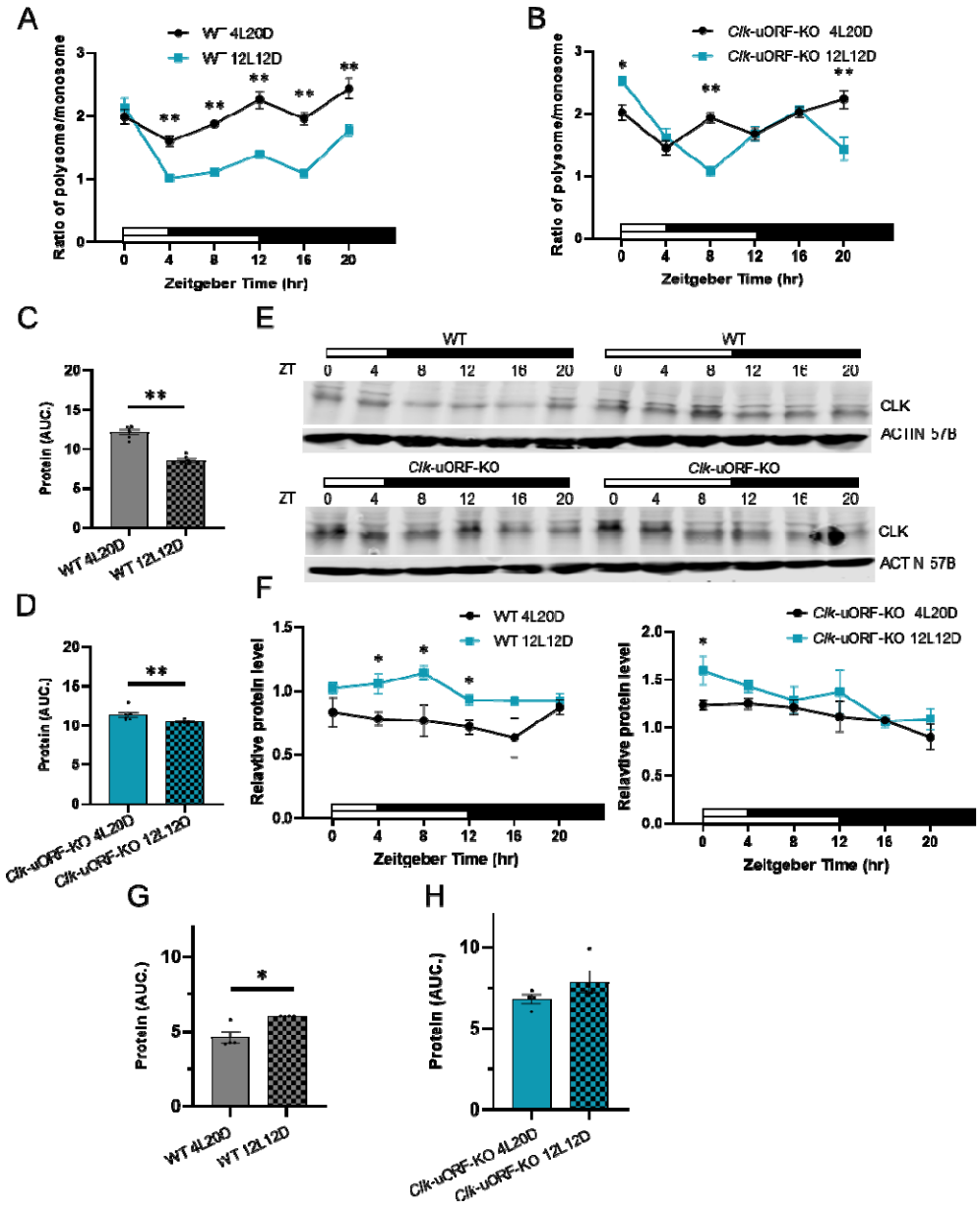



**Figure 7. *Clk* uORFs suppress CLK protein translation in a photoperiod-dependent manner.**

(A-B) P-to-M ratios of *Clk* mRNA from whole-head extracts collected from female flies at indicated time point under 4L20D and 12L12D condition. ($n$ = 6)

(C-D) AUC analysis of daily P-to-M ratio in (A) and (B) of female WT (C) and *Clk*-uORF-KO (D) flies under indicated photoperiod. ($n$ = 6.)

(E) Representative Western blots for CLK protein from whole-head total protein

extracts collected from femail flies at indicated time points under 4L20D (left) and 12L12D (right). ACTIN 57B is used as a loading control.

(F) The relative CLK abundance normalized to ACTIN 57B in (E). The average value of WT under 12L12D is set to 1. ($n$ = 4.)

(G-H) AUC analysis of daily CLK protein level of female WT (G) and *Clk*-uORF-KO (H) flies under indicated photoperiod in (F). ($n$ = 4.)

Data are presented as the mean ± SEM. Signs indicate statistical significance (two-tailed Mann-Whitney $U$ test for unpaired comparisons. **, $P < 0.01$; ***, $P < 0.001$). ZT, Zeitgeber Time.

See also Figure S6.

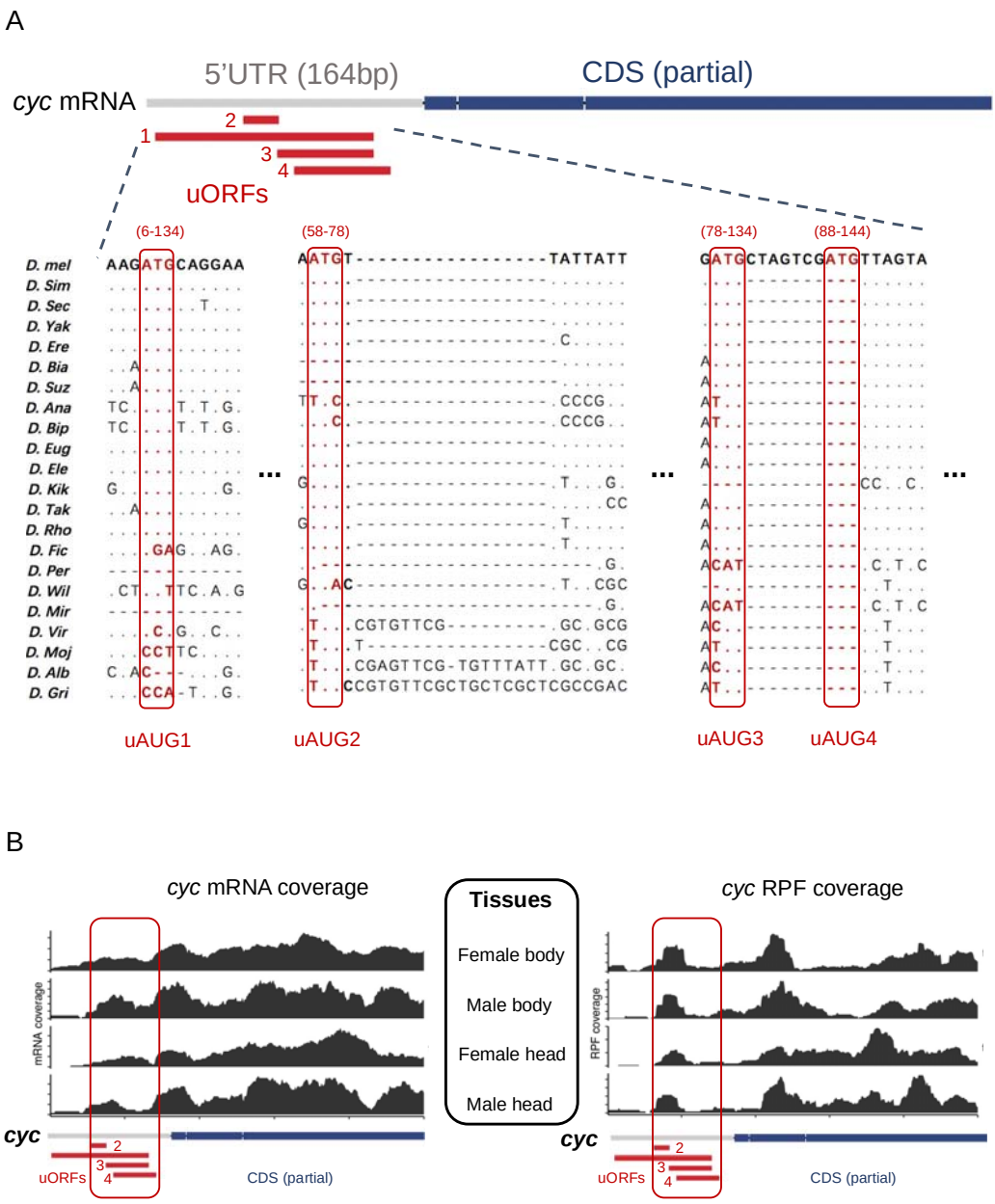



**Figure S1. uORFs suppress CLK and CYC protein expression**

(A) MSA of *cyc* uORFs among 22 *Drosophila* species. The start codons (uAUGs) of uORFs are highlighted by red boxes. The position schemes of uORFs and partial CDSs are denoted above the MSA, with red and blue color, respectively. The start and ending positions of each uORF (separated by a "-") are given in the parenthesis above each uAUG.

(B) The mRNA reads coverage (left) and ribosome-protected footprints (RPF)

coverage (right) of uORFs of *cyc* mRNA from head and body. The position schemes of uORFs and partial CDSs are denoted at the bottom.

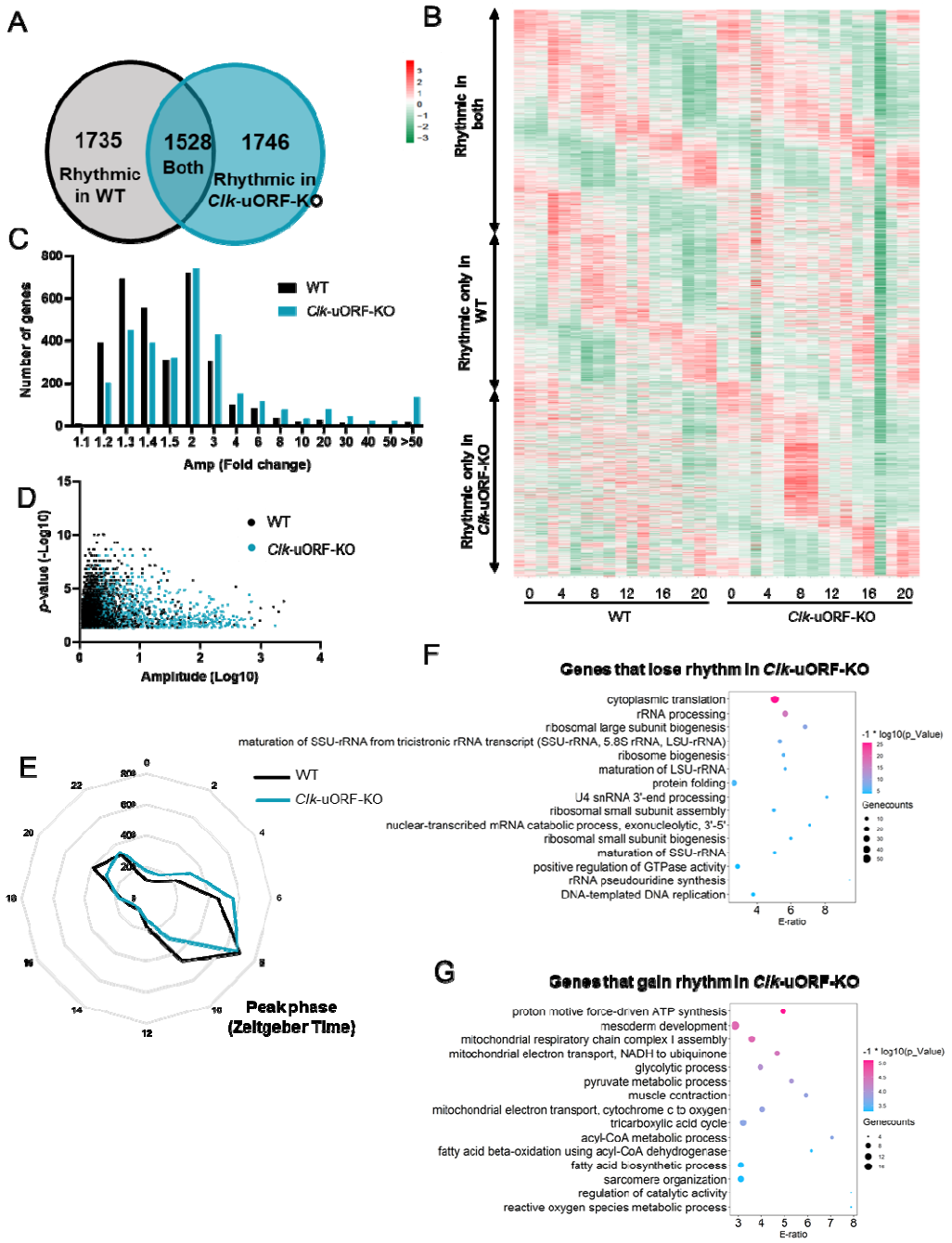



**Figure S2. *Clk* uORFs regulate the temporal expression pattern of the transcriptome in male flies.**

(A) The number of rhythmically expressed genes identified from RNA-seq samples of male WT and *Clk*-uORF-KO flies.

(B) A hierarchical clustering of the rhythmically expressed genes in male flies ordered by the phase of the oscillation. Values of genes are color-coded based on the intensities. Colors indicate low (green) and high (red) Z-scores.

(C) The amplitude distribution of rhythmically expressed genes in male flies.

(D) Volcano plot shows the statistical significance of rhythmically expressed genes in male flies and their corresponding amplitude.

(E) The phase distribution of rhythmically expressed genes in male flies. The y axis in the radar chart indicates the number of rhythmic genes.

(F and G) GO-based enrichment analysis of genes rhythmically expressed on in male WT (F) or *Clk*-uORF-KO (G) flies.

ZT, Zeitgeber Time.

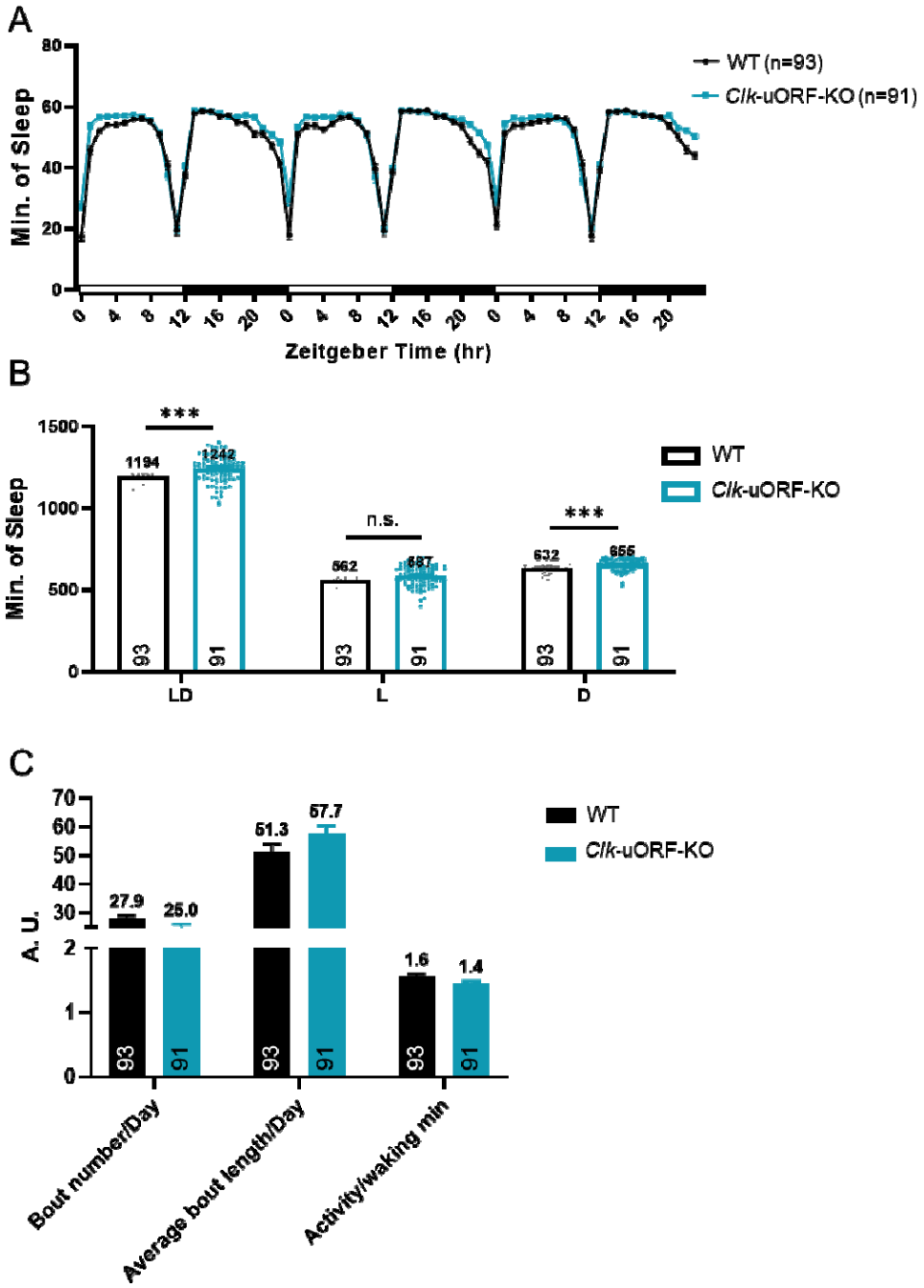


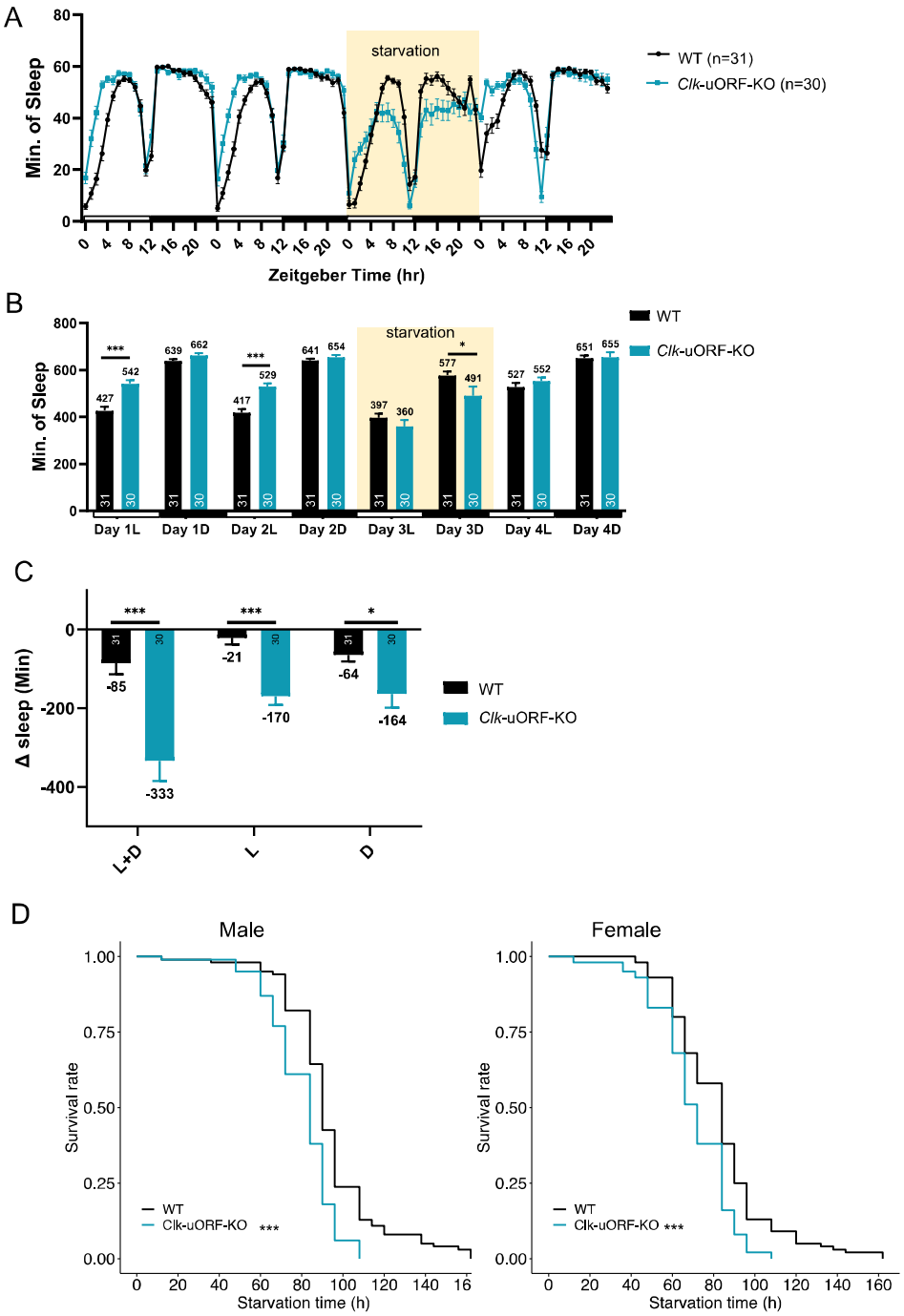

**Figure S3. Knocking out *Clk* uORF does not substantially alter sleep in males.**

(A) Sleep profile of male *Clk*-uORF-KO and WT flies under LD condition.

(B) Sleep duration during daytime (L), nighttime (D) and the entire day (LD) of male *Clk*-uORF-KO and WT flies under.

(C) Daily sleep bout number, average bout length and waking activity of male *Clk*-uORF-KO and WT flies under LD condition.

Data are presented as the mean ± SEM. The numbers of flies tested are shown in the brackets and histograms. Asterisks indicate statistical significance (two-tailed Student's *t* test. *, $P < 0.05$; ***, $P < 0.001$; n.s., $P > 0.05$).




957

958 **Figure S4. *Clk* uORFs suppresses starvation-induced wakefulness**

959 (A) Sleep profile of female uORF-KO and WT flies under food ad libitum and

960 starvation (the 3rd day) condition. Yellow shadow indicates period of starvation.

961 (B) Daily sleep duration of female *Clk*-uORF-KO and WT flies under baseline and

962 starvation condition.

(C) Histogram shows starvation induced sleep change (Δ sleep) compared with normal condition during daytime (L), nighttime (D) and the entire day (LD).

(D) Survival curves of male and female flies under starvation condition. (n=200, log-rank test).

Data are presented as the mean ± SEM. The numbers of flies tested are shown in the histograms. Asterisks indicate statistical significance [two-tailed Student's *t* test for (B) and (C). *, $P < 0.05$; ***, $P < 0.001$].

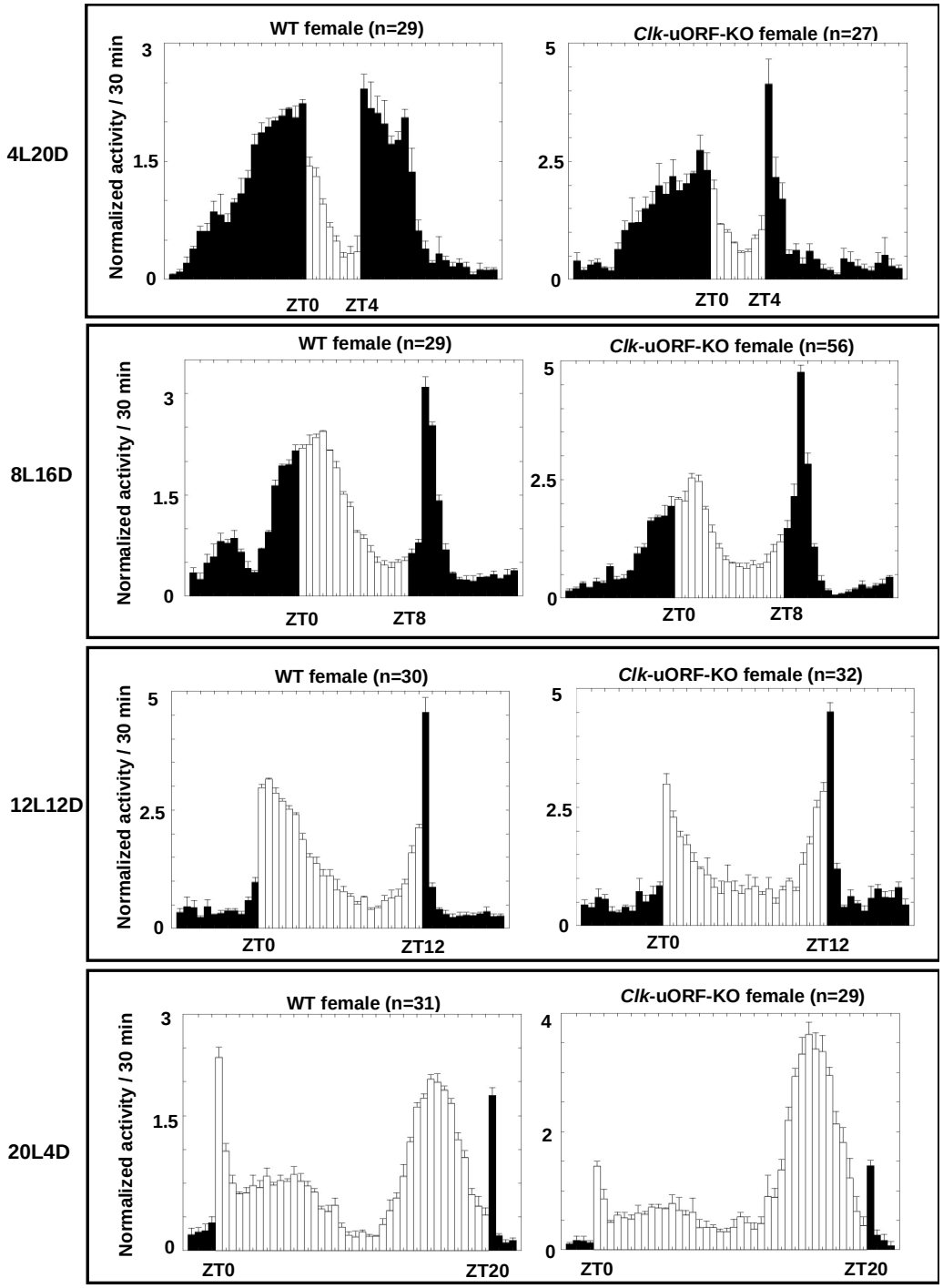



**Figure S5. Knocking out *Clk* uORFs does not alter the adaptation of locomotor rhythm to seasonal photoperiod changes.**

Histograms show the locomotor activity profiles of WT and *Clk*-uORF-KO female flies under 4 h L : 20 h D (4L20D), 8h L : 16 h D (8L16D), 12 h L : 12 h D

(12L12D) and 20 h L : 4 h D (20L4D) condition. ZT, Zeitgeber Time.

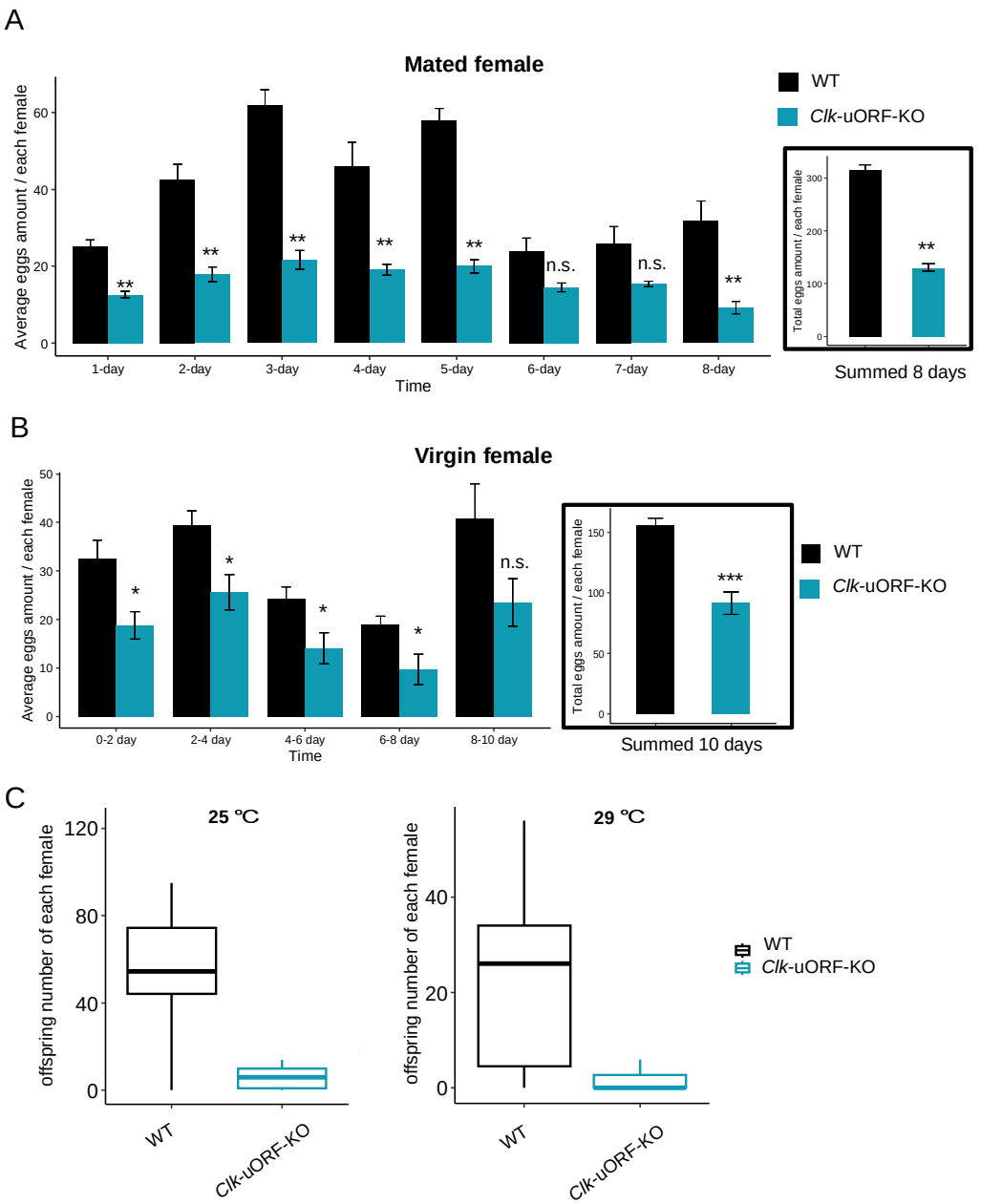



**Figure S6. Knocking out *Clk* uORFs impairs fecundity**

(A) The average egg number (left) and summed number (right) laid by mated female of *Clk*-uORF-KO and WT over 8 consecutive days. Each gray dot represents the average from ten females.

(B) The average egg number (left) and summed number (right) laid by virgins of *Clk*-uORF-KO and WT over 10 consecutive days.

(C) The offspring number per female parent of *Clk*-uORF-KO and WT over 10 days at 25°C and 29°C.

Data are presented as the mean ± SEM. The numbers of tested flies are shown in the histograms. Asterisks indicate statistical significance (n=20, Wilcoxon rank sum test; **, $P < 0.01$; ***, $P < 0.001$; n.s., $P > 0.05$).