Streamlined, cell-specific single-molecule RNA-FISH of core clock mRNAs in clock neurons in whole mount *Drosophila* brains

Ye Yuan[1*], Marc-Antonio Padilla[2], Dunham Clark[3] & Swathi Yadlapalli[1,2,3*]

[1]Cellular and Molecular Biology Program, University of Michigan, Ann Arbor, MI 48109, USA

[2]Neuroscience Graduate Program, University of Michigan, Ann Arbor, MI 48109, USA

[3]Cell and Developmental Biology Department, University of Michigan, Ann Arbor, MI 48109, USA

*co-corresponding authors: yeyu@umich.edu, swathi@umich.edu

KEYWORDS: Circadian rhythms, single molecule Fluorescence In Situ Hybridization (smFISH), machine learning-based analysis pipeline, spatiotemporal organization of clock mRNAs

Abstract

Circadian clocks are ~24-hour timekeepers that control rhythms in almost all aspects of our behavior and physiology. While it is well known that subcellular localization of core clock proteins plays a critical role in circadian regulation, very little is known about the spatiotemporal organization of core clock mRNAs and its role in generating ~24-hr circadian rhythms. Here we describe a streamlined single molecule Fluorescence *In Situ* Hybridization (smFISH) protocol and a fully automated analysis pipeline to precisely quantify the number and subcellular location of mRNAs of *Clock*, a core circadian transcription factor, in individual clock neurons in whole mount *Drosophila* adult brains. Specifically, we designed ~48 fluorescent oligonucleotide probes that can bind to an individual *Clock* mRNA molecule, which can then be detected as a diffraction-limited spot. Further, we developed a machine learning-based approach for 3-D cell segmentation, based on a pretrained encoder-decoder neural network, to automatically identify the cytoplasm and nuclei of clock neurons and combined it with a spot counting algorithm to detect *Clock* mRNA spots in individual clock neurons. Our results demonstrate that the number of *Clock* mRNA molecules cycle in large ventral lateral clock neurons (lLNvs) with a peak at ZT4 (4 hours after lights are turned on, ZT-Zeitgeber time) with ~80 molecules/neuron and a trough at ZT16 with ~30 molecules/neuron. Our streamlined smFISH protocol and deep learning-based analysis pipeline can be employed to quantify the number and subcellular location of any mRNA at the single cell level in *Drosophila* brains. Further, this method can open mechanistic and functional studies into how spatiotemporal organization of clock mRNAs affect circadian rhythms.

Introduction

Almost all organisms have evolved internal clocks to tell time. Circadian clocks are cell-autonomous timekeepers that generate ~24-hour oscillations in the expression of many genes and control rhythms in sleep-wake cycles, metabolism, immunity[1-4]. While past studies highlight the critical role of subcellular localization of core clock proteins in controlling circadian rhythms[5-9], what remains poorly understood is how core clock mRNAs are spatiotemporally organized over the circadian cycle and the role of such organization in generating ~24-hour circadian rhythms at the single-cell level. A major drawback of techniques such as qPCR and RNA-sequencing currently used to study circadian gene expression in clock neurons is lack of spatial information. Single-molecule RNA Fluorescence *In Situ* Hybridization (smFISH) has recently emerged as a powerful technique which enables quantitative measurement of the number as well as the subcellular location of mRNA molecules at the single-cell level. In the field of chronobiology specifically, smFISH method has been recently adapted in one study to understand role of RNA localization on circadian rhythms in *Neurospora Crassa*. In this study, the authors showed that core clock transcripts are clustered in the perinuclear cytoplasmic region in a time-of-day dependent manner.

Drosophila melanogaster is a powerful model system[10] for understanding spatiotemporal regulation of circadian rhythms, as it has a highly conserved clock similar to humans along with powerful genetic and molecular tools. Circadian clocks in all eukaryotes are based on highly conserved transcriptional-translational delayed negative feedback loops. The *Drosophila* circadian clock is made up of four key components[10]–Clock (CLK), Cycle (CYC), Period (PER), and Timeless (TIM). CLK and CYC are the positive transcription factors and PER and TIM act as core clock repressors. During the activation phase, CLK and CYC drive transcription of

thousands of by binding to the E-boxes of clock target genes, including *per* and *tim*. PER and TIM, after a time delay, enter the nucleus and inhibit CLK/CYC activity thereby silencing their own expression as well as the expression of other clock-regulated genes. PER and TIM are then degraded leading to the end of the repression phase and the start of the activation phase of a new cycle. In addition to this core feedback loop, there exists a second interlocked feedback loop that controls rhythmic expression of *Clk* mRNA[11,12]. CLK/CYC proteins bind to the E-boxes of the transcription factors, Vrille (VRI) and PAR domain protein 1ε (PDP1ε), creating rhythms in their mRNA with peak levels during early night and trough levels during early day. Interestingly, VRI and PDP1ε protein levels peak at different times during the night, with VRI peaking during the early evening and PDP1ε later in the night. VRI and PDP1ε proteins in turn bind to the VRI/PDP1ε-binding boxes in the *Clk* enhancer to repress or activate *Clk* transcription, respectively, thus resulting in rhythmic *Clk* mRNA expression. These molecular feedback loops operate in ~150 clock neurons in *Drosophila* brains, creating rhythms in sleep-wake cycles and other physiological processes.

Gene expression, as characterized by the central dogma, refers to the process in which genetic information encoded in DNA is used for synthesis of proteins. In addition to transcriptional regulation, post-transcriptional regulation plays a prominent role in regulating gene expression. In fact, as demonstrated in multiple contexts including stress granules and nuclear speckles, localization of the mRNAs to specific subcellular structures is a vital mechanism that controls post-transcriptional regulation. To visualize mRNA localization, RNA fluorescence in situ hybridization (RNA-FISH) methods have been widely adopted in the past in many model systems. In general, RNA-FISH technique uses a number of fluorescent oligonucleotide probes that can specifically bind to particular sequences of an RNA molecule,

and direct detection or amplification-based methods can be used to visualize the RNAs. While amplification-based methods offer more flexibility with regard to target length and yield better signal-to-noise ratios, direct detection methods generate individual diffraction-limited spots for each single RNA molecule with well-defined intensity and shape profiles.

Here, we describe a streamlined single molecule FISH (smFISH) protocol and fully automated reproducible analysis pipeline to quantify the number and localization of mRNAs of *Clock*, a key clock transcription factor, in individual clock neurons in whole-mount *Drosophila* adult brains (Fig. 1A). We developed a simple, robust smFISH protocol that takes 3 days from. brain dissections to slide preparation. Next, we generated an accurate three-dimensional segmentation of clock neurons using machine learning-based approaches by adopting a pretrained decoder-encoder convolutional neural network model. Finally, we combined our segmentation analysis and a spot-fitting algorithm, "AIRLOCALIZE", into a workflow management system to acquire cell-specific smFISH data in an automated and reproducible manner. To systematically evaluate the performance of our analysis pipeline, we characterized *Clock* mRNA distribution in Large ventral lateral clock neurons (lLNvs), one of the main *Drosophila* clock neuron groups. Our results show that *Clock* mRNA cycles in the light-dark conditions in lLNvs, consistent with past biochemical studies, with peak levels at ZT4 (ZT-Zeitgeber time, ZT0-time of lights on, ZT12-time of lights off) with ∼80 molecules per neuron and trough levels at ZT16 with ∼30 molecules per neuron. Consistent with past studies, we show that *Clock* mRNA cycling is abolished in per^{01} null mutants. Further, we show that clock-neuron specific knockdown of PDP1ε led to consistently low levels of *Clock* mRNA levels throughout the circadian cycle, and clock-neuron specific knockdown of VRI led to consistently high levels of *Clock* mRNA levels. Together, these studies describe a powerful technique to directly

visualize and quantify the number and subcellular location of core clock mRNAs in individual neurons, which can open new investigations into the role of spatiotemporal localization of mRNAs in controlling circadian rhythms. The methods described here can be easily adopted to directly visualize gene expression at the single-cell level in *Drosophila* brains.

Results

To directly visualize the subcellular localization and quantify the number of core clock mRNAs in individual clock neurons, we adapted single molecule RNA-Fluorescence *In Situ* Hybridization (smFISH) method for use in *Drosophila* clock neurons and developed a machine-learning based analysis pipeline to analyze the data. Below, we describe the smFISH method and analysis pipeline in detail and then our present our findings.

Procedure overview

Our smFISH procedure consists of five parts: smFISH sample preparation, image acquisition, data analysis, quality control check, and final plot generation. For sample preparation, *Drosophila* brains are dissected, fixed, and subjected to a three-day experiment protocol with little hands-on time (~1 hour/day) required for sample preparation. For imaging, we used a glycerol-based mountant for tissue clearing and acquired images using a high-resolution Airyscan confocal microscope, which enabled us to acquire high signal-to-noise ratio (SNR) images in which well-defined diffraction-limited spots, corresponding to individual mRNA molecules, can be detected in smFISH probe channels. To automate the process of identification of the cytoplasm and nucleus of lLNvs, a group of clock neurons, we developed a custom convolutional neural network (CNN) segmentation model to generate three-dimensional lLNv

clock neuron masks. Performance of our CNN segmentation model was validated by manual inspection of CNN-generated masks on a testing dataset. To quantify the intensity and visualize the subcellular location of individual mRNA spots, we adapted a published spot detection algorithm AIRLOCALIZE. Finally, we developed an integrated pipeline in which all data analysis workflows are automated by the workflow manager Snakemake. After data analysis step, we generated quality control plots to verify the consistency of our results, identify noisy image outliers, and optimize analysis parameters. The analysis can be rerun with optimized parameters in a fully automated and reproducible manner. Finally, we report the number of identified mRNA spots per cell, which is a quantitative measurement of absolute mRNA molecule count. In the next subsections, we describe individual steps of the protocol in detail.

1. Fluorescent oligonucleotide probes for smFISH experiments

To enable detection of specific mRNA transcripts, we order commercial fluorescent oligonucleotide probes that can specifically bind to complementary sequences within individual mRNA molecules from Stellaris. First, we design ~48 20nt short oligonucleotide probes that span across the full-length of the target mRNA species covering all exons of the mRNA using the online Stellaris RNA-FISH Probe Designer. For genes with alternative splicing isoforms, shared exons are used to design smFISH probes to ensure detection of all isoforms. The designer provides masking of repetitive sequences to improve probe specificity, and we generally start with the most stringent masking level. In our experience, thirty probes are generally adequate, but the minimum number of probes that can still yield high signal-to-noise ratio images is likely dependent on microscope resolution. It is possible that enough probes cannot be designed given the length of the target mRNA. In such cases, masking level can be adjusted to increase the

number of smFISH probes. Each of the 20nt probe is conjugated at the 5' end to a fluorophore and collaborative binding of ~48 probes to an individual mRNA molecule enables detection of a single mRNA spot by fluorescence microscopy. Upon receipt of the probe pool, we prepare aliquots of 20μM probe stock in 10mM Tris-Cl, pH 8.5 and store them in -20C.

2. Sample preparation

In our smFISH experiments, we typically use transgenic flies in which cell membranes are fluorescently labeled (*e.g.*, CD4-tdTomato, CD8-GFP), which enables identification of specific cell types in *Drosophila* brains. Specifically, we drive the expression of CD4-tdTomato or CD8-GFP exclusively in clock neurons by crossing them with pan-clock neuron driver flies, *Clk856-GAL4*. On day 1, *Drosophila* brains are dissected, fixed in 4% Paraformaldehyde, and permeabilized in 70% ethanol at 4C overnight. Fixation step is a crucial step in smFISH experiments, as both under-fixation as well as over-fixation of tissues can result in samples in which cellular morphology is not well preserved. Permeabilization with 70% ethanol is another necessary step to allow the smFISH probes to pass through the cell membrane and enter the cell. On day 2, brains are washed with PBS and then incubated in hybridization solution with probes overnight at 37C. On day 3, brains are washed with PBS to remove the hybridization solution and mounted on a slide using a glycerol-based mountant for tissue clearing (see Methods section for a detailed protocol). This smFISH protocol is robust, simple to implement, taking 3 days from dissection to slide preparation, and produces highly reproducible results across experimental repeats.

3. Microscope setup for image acquisition

The goal is to obtain clock neuron images where smFISH spots and the neuron cell membrane marker are clearly visible and can be distinguished from the background. In general, image acquisition parameters should be optimized taking into account both the clock neuron marker and smFISH probe channels, and the same settings need to be reused across all experimental conditions and repeats to ensure data consistency. We typically consider tuning the following key parameters: Firstly, filter settings must be compatible with the fluorophore used. Secondly, appropriate laser power and detector gain should be determined. While larger laser power and detector gain will yield higher pixel intensities, high laser power will lead to significant bleaching during acquisition while high detector gain will lead to increased random noise. Finally, to further suppress random noise, pixel averaging might be adopted. Satisfactory settings should yield distinct diffraction-limited spots in the FISH probe channels and clear cell boundaries in the neuron marker channel across all Z-slices, and the quality control results showing detected spots and segmentation masks should be consistent with expectation from manual inspection.

4. Segmentation of clock neurons

We adopted a pretrained encoder-decoder convolutional neural network (CNN) for segmentation of clock neurons and detailed documentation is publicly available online. In brief, backbone of the network with pretrained parameters from the ImageNet dataset is obtained from public packages. Clock neurons pose an intriguing problem because i) three-dimensional information should be incorporated to provide reasonable segmentation across different Z-slices, ii) cell bodies of a neuron cluster such as lLNvs may be overlapping each other due to limited resolution, iii) cell marker intensity varies significantly within each cell and among different experiment

conditions. To solve the varying intensity problem, for each training Z-stack, we transformed our training data to generate a series of Z-stacks with varying intensities. Further, we adopted a simple histogram stretching algorithm to maintain reasonable contrast of different marker signal data (Fig. S1C). To solve the three-dimensional problem, at each Z-slice, we transformed the data to provide current Z-slice and projection of neighboring Z-slices signals (Fig. S1D). Performance of the segmentation model was verified by visual inspection of testing datasets and the pipeline also provides visualization of the segmentation outputs (Fig. S1D).

5. smFISH spot detection

We adopted the AIRLOCALIZE package for spot detection, detailed documentation of which is publicly available online. In short, diffraction-limited spots can be modeled as three-dimensional Gaussian functions with lateral symmetry (i.e., a point spread function, PSF). PSF parameters, which are dependent on the microscope setup, must be estimated with a graphical interface provided by the package, and can be specified in nanometers to the pipeline (Fig. S1A). Therefore, smFISH data can be considered as a Gaussian mixture. AIRLOCALIZE tries to fit Gaussian functions centered around local maxima after applying a global threshold to reject noise. The global threshold parameter affects spot detection significantly, must be determined by experiment, and reflects potentially varying noise level in different samples. We find the threshold to be mostly constant for samples in one experiment batch, and the pipeline provides adequate spot visualization for optimizing threshold settings.

6. Data preprocessing and integration of the workflows

All workflows including spot detection and segmentation require suitable image formats, varying software environment, and configuration of key parameters to run properly. In order to ensure reproducibility and aid optimization of analysis parameters, we developed an automated analysis pipeline using the Snakemake workflow management system (Fig. 3A). First, acquired raw data is converted to lossless, open-source OME-TIFF format by the Bio-Formats package [cite]. Next, smFISH and cell marker channels are split and used for spot detection and cell segmentation respectively [mention packages in methods?]. Finally, quality control plots and visualization of all intermediate outputs are generated, and results are combined to produce the final output. The pipeline with detailed documentation is publicly available online.

7. Quality control check

As is the case with most complex quantitative experiments, unbiased consistency verification and identification of noisy data is vital to reject noises originated from varying conditions during dissection, incubation, mounting, etc. Towards this end, we developed a series of quality control plots. Firstly, intensity histogram of detected dots for all and individual images are provided (Fig. 3B). As each mRNA molecule is bound to a constant number of fluorophores, such intensities should be similar among different images and conditions. Secondly, to understand whether working distance is adequate, dot intensity distribution along Z-stack position is plotted for each image with a linear fit (Fig. 3C). If spot intensity varies significantly along Z-stack position, working distance should be considered insufficient. Finally, to understand performance of the segmentation model, physical volumes of the segmentation masks are computed and summarized as equivalent diameter values assuming spherical shapes.

Visualization and quantification of Clock mRNAs in lLNv neurons

To validate our smFISH pipeline, we performed Clk smFISH specific to large ventral lateral neurons (lLNvs) across the circadian cycle. We observed ~80 Clk mRNAs per cell during peak at ZT4 and ~80 spots per cell at trough ZT16 (Fig. 2B). Moreover, Clk transcript rhythm is disrupted in per^{01} mutant background (Fig. 2C). The molecular clock consists of multiple interlocked transcriptional-translational feedback loops to realize a ~24hr self-sustained oscillator. Apart from the core Period/Timeless negative feedback loop, auxiliary loops are in place for regulation of Clk transcription (Fig. 4A). To understand how the two factors vrille (vri) and PAR-domain protein 1 (Pdp1) affect Clk mRNA at a subcellular level, we performed smFISH with vrille or Pdp1 RNAi knockdown mutants (Fig. 4B). Our results demonstrate that vri is necessary for repression of Clk transcription at trough time points while knockdown of Pdp1 leads to overall lower Clk transcription.

Quality control is an essential component of the pipeline to understand variations among different samples. Individual and merged intensity histogram plots (Fig. 3B) showed that smFISH signals among different samples are highly consistent (Fig. 3D). Further, working distance evaluation plots (Fig. 3C) revealed that most (>80%) images show adequate working distance for individual samples (Fig. 3D). Finally, equivalent diameters of segmentation masks demonstrate consistent volumes expected for lLNv neurons (Fig. 3D). With these three quality controls we were able to confirm consistency of the input data, which is essential for generation of the final molecule count plot (Fig. 2B). Intermediate results of the pipeline including segmentation masks and identified smFISH spots are readily available for visual inspection (Supplementary Video 1-3).

Discussion

It has been well appreciated that mRNAs and protein complexes are localized to specific subcellular compartments at specific times to enable cellular functions. Using live imaging and high-resolution microscopy, we recently showed that PER/CLK proteins are organized into discrete, dynamic nuclear condensates, and are not in fact diffusely distributed in the nucleus as was expected from prior immunostaining experiments, specifically during the repression phase in *Drosophila* clock neurons[7].

smFISH requires hybridization of tens of ~20nt oligonucleotides to the same RNA molecule of interest. Therefore, a relatively long transcript is required for adequate SNR. In our experience, at a minimum 1.2-1.5kb is required. This requirement limits smFISH from visualizing short transcripts as well as individual exons or introns of interest. Towards this end, amplification-based methods, such as hybridization chain reaction (HCR) may be adopted. However, while spatial information is typically retained in amplification-based methods, quantitative capability requires optimization as signal amplification kinetics must be determined empirically.

As is the case with other in situ hybridization methods, smFISH requires extensive sample treatment including fixation, permeabilization and hybridization. Therefore, temporal resolution of smFISH is limited and artefacts may arise from fixation and hybridization such as loss of transient interactions and sample shrinkage. In this regard, live imaging of RNA molecules such as MS2 stem-loop tagging may be preferable. However, addition of stem-loops to the RNA molecule of interest may affect its function and/or transcription/degradation kinetics.

Materials and Methods

Solutions used for smFISH

Prepare all solutions with RNase-free tubes and reagents and filtered pipette tips. Aliquot reagent stocks to reduce chance of contamination.

Fixation solution: 4% formaldehyde in 1X phosphate-buffered saline (PBS) and is made from 16% EM-grade formaldehyde (Electron Microscopy Sciences 15710) and 10X PBS (Invitrogen AM9625) in RNase-free water.

Permeabilization solution: 70% ethanol diluted from 200 proof ethanol (Thermo Scientific T038181000) in RNase-free water.

Hybridization solution: 2X saline-sodium citrate (SSC, Invitrogen AM9763), 10% dextran sulfate (Thermo Scientific AC433240050), 1mg/mL E.coli tRNA (Roche 10109541001), 2mM ribonucleoside vanadyl complex (RVC, New England Biolabs S1402S), 0.5% bovine serum albumin (BSA, Roche 10711454001), 10% formamide (Sigma, F9037), and 5μM oligonucleotide probes conjugated to fluorophores (Stellaris, LGC Biosearch Technologies) in RNase-free water.

Wash solution: 2X SSC and 10% formamide in RNase-free water.

We prepare 20% (weight/volume) dextran sulfate stock from powder. Fixation and permeabilization solutions are prepared fresh on day of experiment and RVC and formamide aliquots are used one-time once thawed.

smFISH protocol

1. Prepare fixative solution. For each sample, put 500uL of fixative to a well of a clean glass dish and leave on ice.

2. Dissect at least 10 brains in 1X PBS for each sample. Put dissected brains directly into the fixative on ice in the glass dish. Dissections should be completed within 30 minutes.
3. Place the glass dish on a nutator and incubate for 20 minutes at room temperature.
4. Remove the fixative and perform three quick rinses with 1X PBS.
5. Remove solution and wash twice with 1X PBS, 5 minutes per wash at room temperature with nutation.
6. Remove 1X PBS completely and add 500uL of permeabilization solution to the glass dish. With a P1000 tip, carefully transfer the brains into a protein low-bind tube. Avoid damaging the brains during transfer.
7. Fill the tubes with extra permeabilization solution to a final volume of 1mL. Incubate the sample overnight at 4°C with nutation, protected from light.
8. Prepare hybridization solution, 100μL per sample and leave on ice protected from light. Prepare wash solution, 1mL per sample and leave at room temperature.
9. Aspirate permeabilization solution and add 1mL of wash solution. Incubate 5 minutes at room temperature with nutation.
10. Aspirate the wash solution completely and add 100μL of hybridization solution containing probe(s). Gently pipette to make sure all brains are collected in the solution.
11. Incubate overnight in a 37°C incubator, protected from light.
12. Add 1mL of wash solution without removing the hybridization solution. Gently pipette to mix without touching the brains. Incubate at 37°C for 30 minutes.
13. Aspirate off the solution and wash one more time with 1mL of wash solution at 37°C for 30 minutes. Supplement the wash solution with 1μg/mL DAPI for nucleus staining.

14. Remove the wash solution and mount the samples in Prolong glass mountant. Make sure the brains are mounted in the correct orientation. Seal the samples immediately after mounting and leave at room temperature overnight protected from light.
15. The samples are ready for imaging the next day.

Statistical tests

All statistical significance tests presented are unpaired, two-tailed Student's *t*-tests assuming unequal variances.

Data Availability Statement

The analysis pipeline is publicly available online at (github).

Author Contributions

Y.Y. and S.Y. developed the smFISH protocol. Y.Y. developed the automated analysis pipeline. Y.Y., M.P., D.C., and S.Y. performed all the experiments. Y.Y., M.P., and D.C., analyzed the data from the experiments. Y.Y. and S.Y. wrote the paper with input from all the authors.

Declaration of competing interest

The authors declare that they have no known competing financial interests.

Acknowledgments

The authors acknowledge help from Dr. Yukiko Yamashita and Dr. Jaclyn Fingerhut for the development of the smFISH protocol in adult *Drosophila* brains. The authors also acknowledge

Yangbo Xiao for help in preparing the figures. This work was supported by funds from the NIH (grant no. R35GM133737 to S.Y.), Alfred P. Sloan Fellowship (to S. Y.) and McKnight Scholar Award (to S. Y.).

References

1 Hardin, P. E., Hall, J. C. & Rosbash, M. Feedback of the Drosophila period gene product on circadian cycling of its messenger RNA levels. *Nature* **343**, 536-540, doi:10.1038/343536a0 (1990).

2 Aronson, B. D., Johnson, K. A., Loros, J. J. & Dunlap, J. C. Negative Feedback Defining a Circadian Clock - Autoregulation of the Clock Gene-Frequency. *Science* **263**, 1578-1584, doi:DOI 10.1126/science.8128244 (1994).

3 Sehgal, A., Price, J. L., Man, B. & Young, M. W. Loss of circadian behavioral rhythms and per RNA oscillations in the Drosophila mutant timeless. *Science* **263**, 1603-1606, doi:10.1126/science.8128246 (1994).

4 Patke, A., Young, M. W. & Axelrod, S. Molecular mechanisms and physiological importance of circadian rhythms. *Nat Rev Mol Cell Bio* **21**, 67-84, doi:10.1038/s41580-019-0179-2 (2020).

5 Curtin, K. D., Huang, Z. J. & Rosbash, M. Temporally regulated nuclear entry of the Drosophila period protein contributes to the circadian clock. *Neuron* **14**, 365-372, doi:10.1016/0896-6273(95)90292-9 (1995).

6 Shafer, O. T., Rosbash, M. & Truman, J. W. Sequential nuclear accumulation of the clock proteins period and timeless in the pacemaker neurons of Drosophila melanogaster. *J Neurosci.* **22**, 5946-5954, doi:20026628 (2002).

7 Xiao, Y., Yuan, Y., Jimenez, M., Soni, N. & Yadlapalli, S. Clock proteins regulate spatiotemporal organization of clock genes to control circadian rhythms. *Proc Natl Acad Sci U S A* **118**, doi:10.1073/pnas.2019756118 (2021).

8 Cohen, S. E. *et al.* Dynamic localization of the cyanobacterial circadian clock proteins. *Curr Biol* **24**, 1836-1844, doi:10.1016/j.cub.2014.07.036 (2014).

9 Zhao, H. *et al.* PARP1- and CTCF-Mediated Interactions between Active and Repressed Chromatin at the Lamina Promote Oscillating Transcription. *Mol Cell* **59**, 984-997, doi:10.1016/j.molcel.2015.07.019 (2015).

10 Dubowy, C. & Sehgal, A. Circadian Rhythms and Sleep in Drosophila melanogaster. *Genetics* **205**, 1373-1397, doi:10.1534/genetics.115.185157 (2017).

11 Cyran, S. A. *et al.* vrille, Pdp1, and dClock form a second feedback loop in the Drosophila circadian clock. *Cell* **112**, 329-341, doi:Doi 10.1016/S0092-8674(03)00074-6 (2003).

12 Glossop, N. R. J. *et al.* VRILLE feeds back to control circadian transcription of Clock in the Drosophila circadian oscillator. *Neuron* **37**, 249-261, doi:Doi 10.1016/S0896-6273(03)00002-3 (2003).

Figure 1

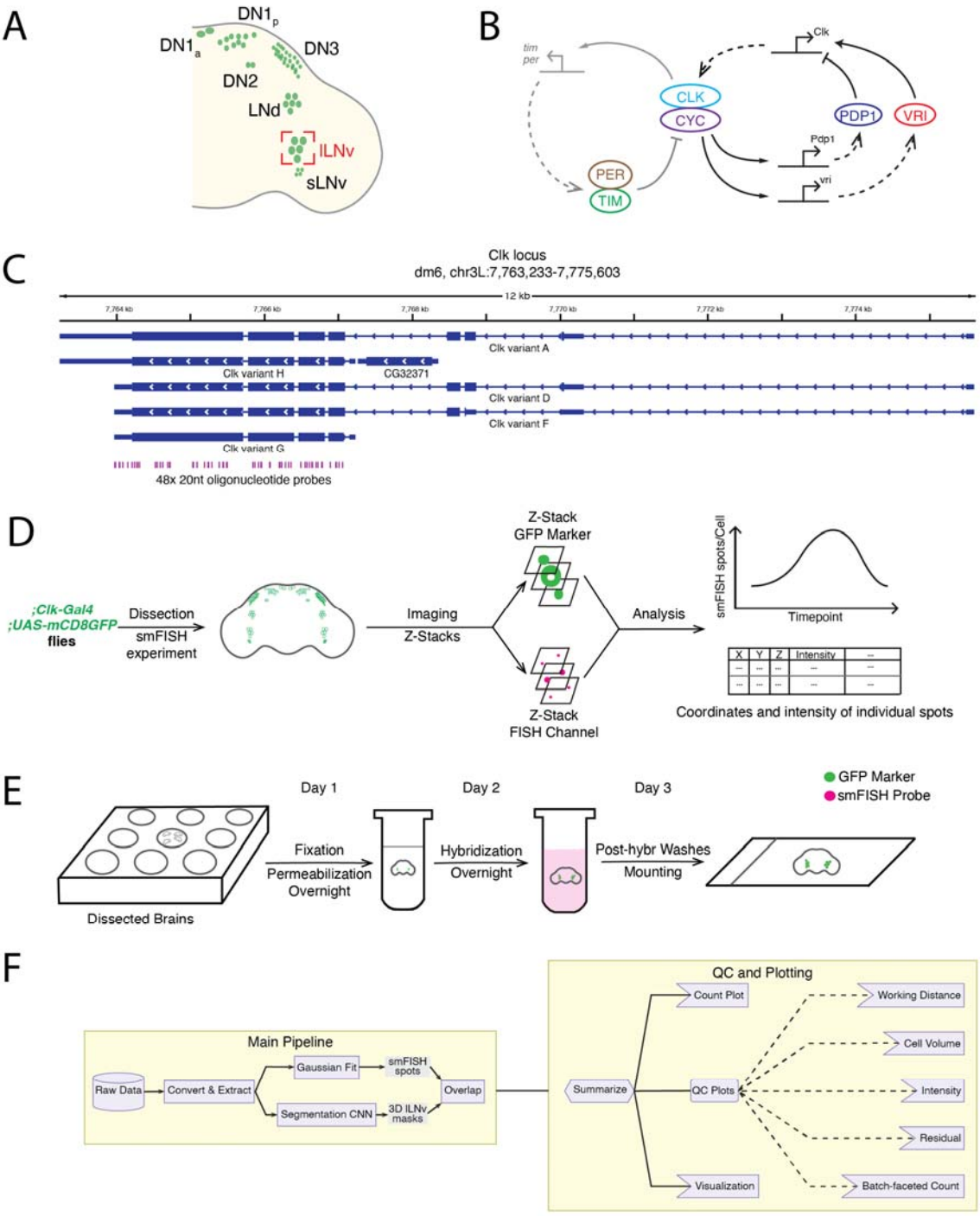

A
DN1p
DN1a
DN3
DN2
LNd
lLNv
sLNv
B
tim
per
CLK
CYC
PER
TIM
Clk
PDP1
VRI
Pdp1
vri
C
Clk locus
dm6, chr3L:7,763,233-7,775,603
12 kb
Clk variant A
Clk variant H
CG32371
Clk variant D
Clk variant F
Clk variant G
48x 20nt oligonucleotide probes
D
;Clk-Gal4
;UAS-mCD8GFP
flies
Dissection
smFISH
experiment
Imaging
Z-Stacks
Z-Stack
GFP Marker
Z-Stack
FISH Channel
Analysis
smFISH spots/Cell
Timepoint
X
Y
Z
Intensity
Coordinates and intensity of individual spots
E
Day 1
Day 2
Day 3
GFP Marker
smFISH Probe
Dissected Brains
Fixation
Permeabilization
Overnight
Hybridization
Overnight
Post-hybr Washes
Mounting
F
Main Pipeline
Raw Data
Convert & Extract
Gaussian Fit
smFISH spots
Segmentation CNN
3D lLNv masks
Overlap
QC and Plotting
Summarize
Count Plot
QC Plots
Visualization
Working Distance
Cell Volume
Intensity
Residual
Batch-faceted Count

Figure 1. Overview of the smFISH method and analysis pipeline.

A) Schematic of different clock neuron groups in a *Drosophila melanogaster* hemi-brain. We focused on large ventrolateral neurons (lLNvs) as circled in red. **B)** Schematic of the molecular clock in *Drosophila*. Clock (Clk) and Cycle (cyc) are at the center of interlocked feedback loops that are intricately balanced to generate the ~24 molecular rhythm. **C)** Genomic locations of the oligonucleotide probes in the *Clk* locus. Probes are put in shared exon sequences to visualize mRNAs of all annotated *Clk* transcript variants. **D)** Overview of the full workflow including experiment and analysis. **E)** illustrates the smFISH procedure and **F)** outlines the analysis pipeline. Two tasks are performed in parallel to enable lLNv-specific analysis: detection of *Clk* smFISH spots and segmentation of lLNvs.

Figure 2

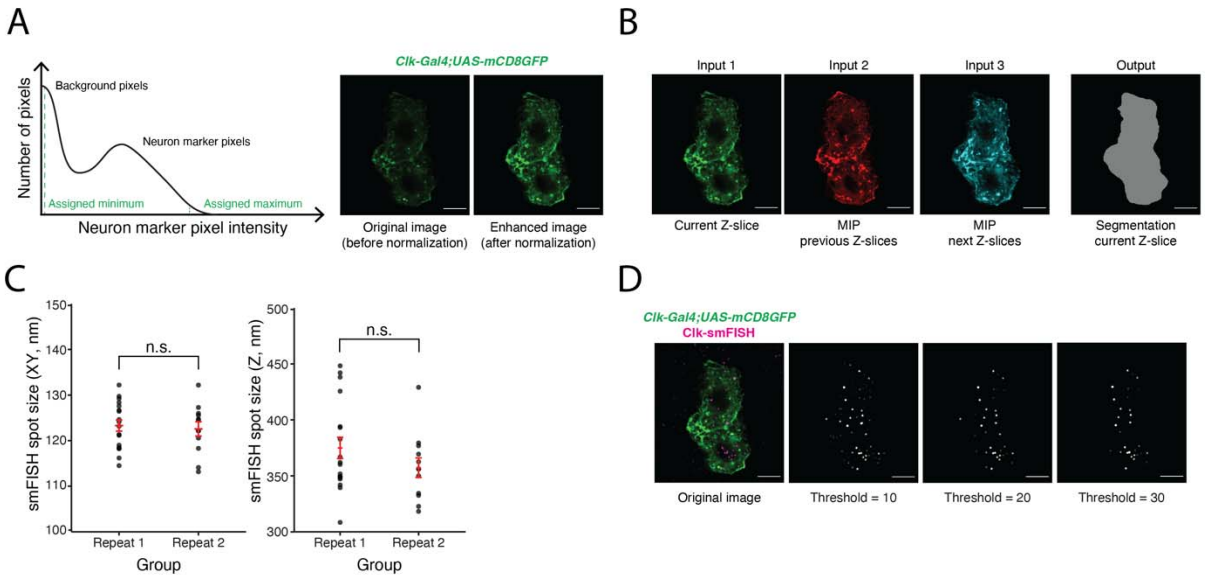


Figure 2. Development of the machine-learning based analysis pipeline.

lLNv cell bodies are relatively large (~10μm) structures spanning multiple Z-slices. To enable robust segmentation, we adopted a publicly available U-net convolutional neural network (CNN) pixel segmentation model. A shows normalization of the neuron marker image. Because of genotype and experimental variations, pixel intensities of the neuron marker signal may vary significantly. Considering a single Z-slice in which there are two clearly visible lLNvs, the original marker signal may be dim and requires normalization. To overcome this issue, we compute the distribution of marker pixel intensities. Typically, a bimodal distribution is observed. Most pixels will show near-zero values (background pixels) and another mode shows significantly higher values (true mCD8GFP signal). We then take the 2% dimmest and brightest intensities and assign that to be the minimum and maximum values. B demonstrates incorporation of three-dimensional information into the CNN model. Publicly available CNN

models are typically trained on two-dimensional image datasets. To get a simple extension into 3D, for each Z-slice, we input the following images to the CNN model: 1) current Z-slice, 2) a maximum intensity projection (MIP) of neighboring Z-slices within 0.4μm in one direction, 3) a MIP of neighboring 0.4μm slices in another Z-direction. With manual inspection using an isolated validation dataset, we found that this method generates robust segmentation results with a small training dataset consisting of ~10 annotated Z-stacks. Another workflow of the analysis is smFISH spot detection. Towards this end, we adopted the AIRLOCALIZE script which performs Gaussian fitting based on empirically measured smFISH spot size using local maxima. C shows measured smFISH spot dimensions using our confocal setup with Cy5 probes. We found the spot sizes in both XY and Z direction to be highly reproducible, with XY size ~120nm and Z size ~400nm. Those values are expected given our Airyscan setup. To perform Gaussian fitting, AIRLOCALIZE adopts a global filter to reject background noise. In D, we show the effect of different threshold settings on the detected smFISH spots. In this example image, the number of detected spots visible in this Z-slice changes with threshold setting. At the minimum threshold tested, ~70 spots are detected. At the maximum threshold tested, ~30 spots are detected. We find it important to manually sample and inspect a portion of images to ensure proper threshold settings and keep it consistent across different experiment groups and biological replicates. Scale bars, 5μm.

Figure 3

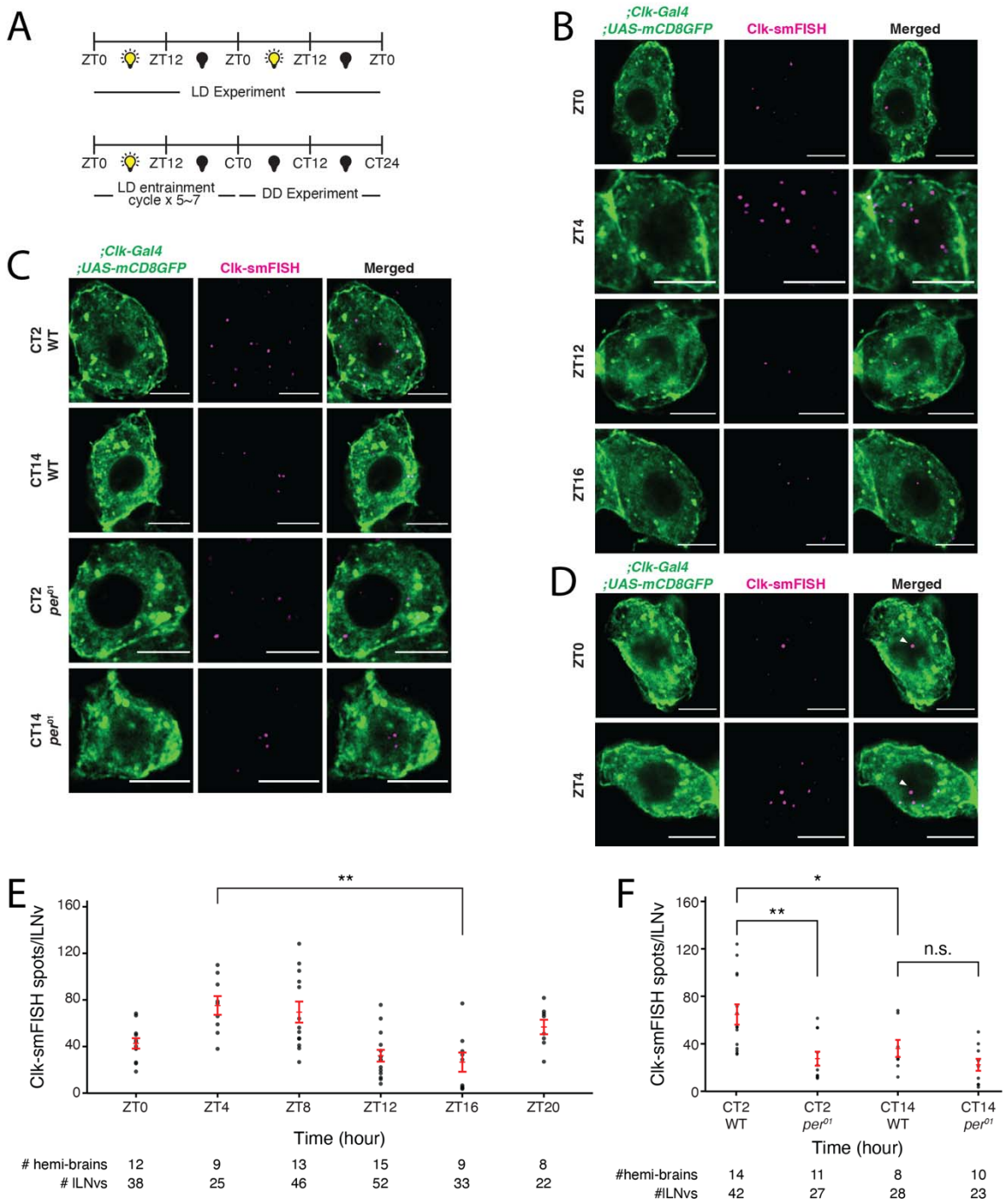
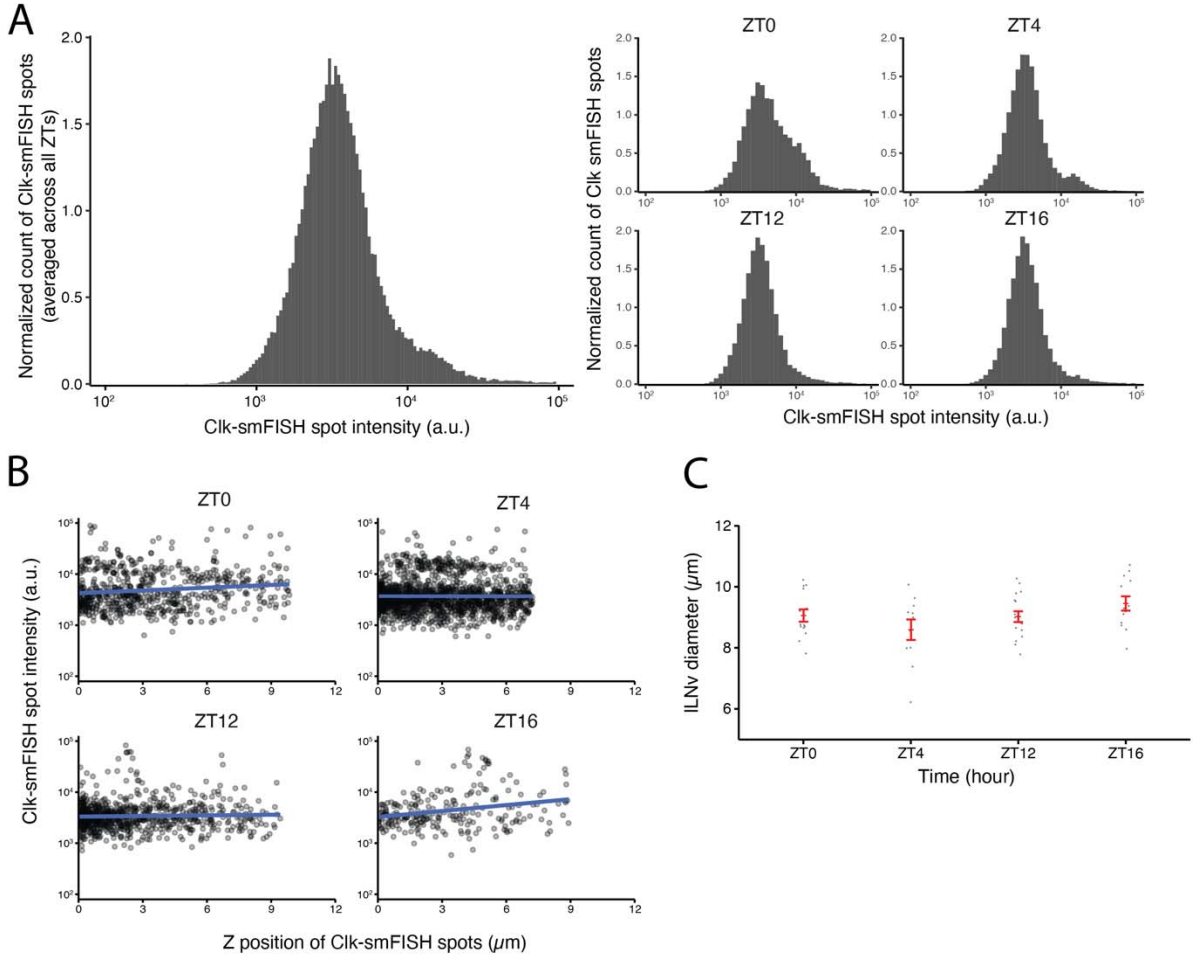

A
ZT0 ZT12 ZT0 ZT12 ZT0
LD Experiment
ZT0 ZT12 CT0 CT12 CT24
LD entrainment cycle x 5~7
DD Experiment
B
;Clk-Gal4 ;UAS-mCD8GFP
Clk-smFISH
Merged
ZT0
ZT4
ZT12
ZT16
C
CT2 WT
CT14 WT
CT2 per01
CT14 per01
D
ZT0
ZT4
E
Clk-smFISH spots/ILNv
Time (hour)
ZT0 ZT4 ZT8 ZT12 ZT16 ZT20
hemi-brains 12 9 13 15 9 8
ILNvs 38 25 46 52 33 22
F
CT2 WT
CT2 per01
CT14 WT
CT14 per01
n.s.
#hemi-brains 14 11 8 10
#ILNvs 42 27 28 23

Figure 3. Clk mRNA cycles in lLNvs in light-dark condition and is abolished in clock neurons in period null mutant per^{01} flies.

A) Entrainment schedule for either light-dark (LD) or constant darkness (DD) experiments. We first performed Clk-smFISH using wildtype flies (Clk>mCD8GFP) under LD condition. **B)** Representative images of Clk mRNA at peak (ZT4) and trough (ZT16) timepoints. **C)** To further confirm that the observed smFISH signal is specific, we performed Clk smFISH experiments using per^{01} (period null mutant) with wildtype control in DD condition. **D)** Single large nuclear spots are observed in individual lLNv nuclei reflecting active transcription site. **E)** and **F)** show quantification of the experiments. Scale bars, 5μm.

Figure 4



D

Time (hour)	ZT0	ZT4	ZT8	ZT12	ZT16	ZT20
log10(spot intensity) (a.u.)	3.80±0.02	3.72±0.03	3.66±0.01	3.61±0.02	3.66±0.03	3.73±0.03
lLNv diameter (um)	9.1±0.2	8.6±0.3	8.8±0.1	9.0±0.2	9.5±0.2	10.0±0.3

Figure 4. Quality control plots for *Clk* smFISH spots in LD condition.

To provide more information on data quality and inform outlier removal, the analysis pipeline generates multiple quality control plots. Firstly, individual mRNAs should be bound by a consistent number of oligonucleotide probes. Therefore, histogram plots of smFISH spot intensity should show predominantly a single peak with consistent distribution among different samples **(A)**. Secondly, as the whole-mount imaging usually scans ~10μm for lLNv imaging, we

looked at how smFISH spot intensity varies with each spot's Z position. **B)** For each image, our pipeline generates an intensity ~ Z-position plot and representative plots for peak and trough timepoints are shown. Z-position is the relative distance of each smFISH spot to the Z-stack acquisition start Z coordinate. **C)** Also, given consistent cell marker signal among different samples, our segmentation model should report similar diameters of lLNvs. **D)** summary table of the quality control metrics.

Figure 5

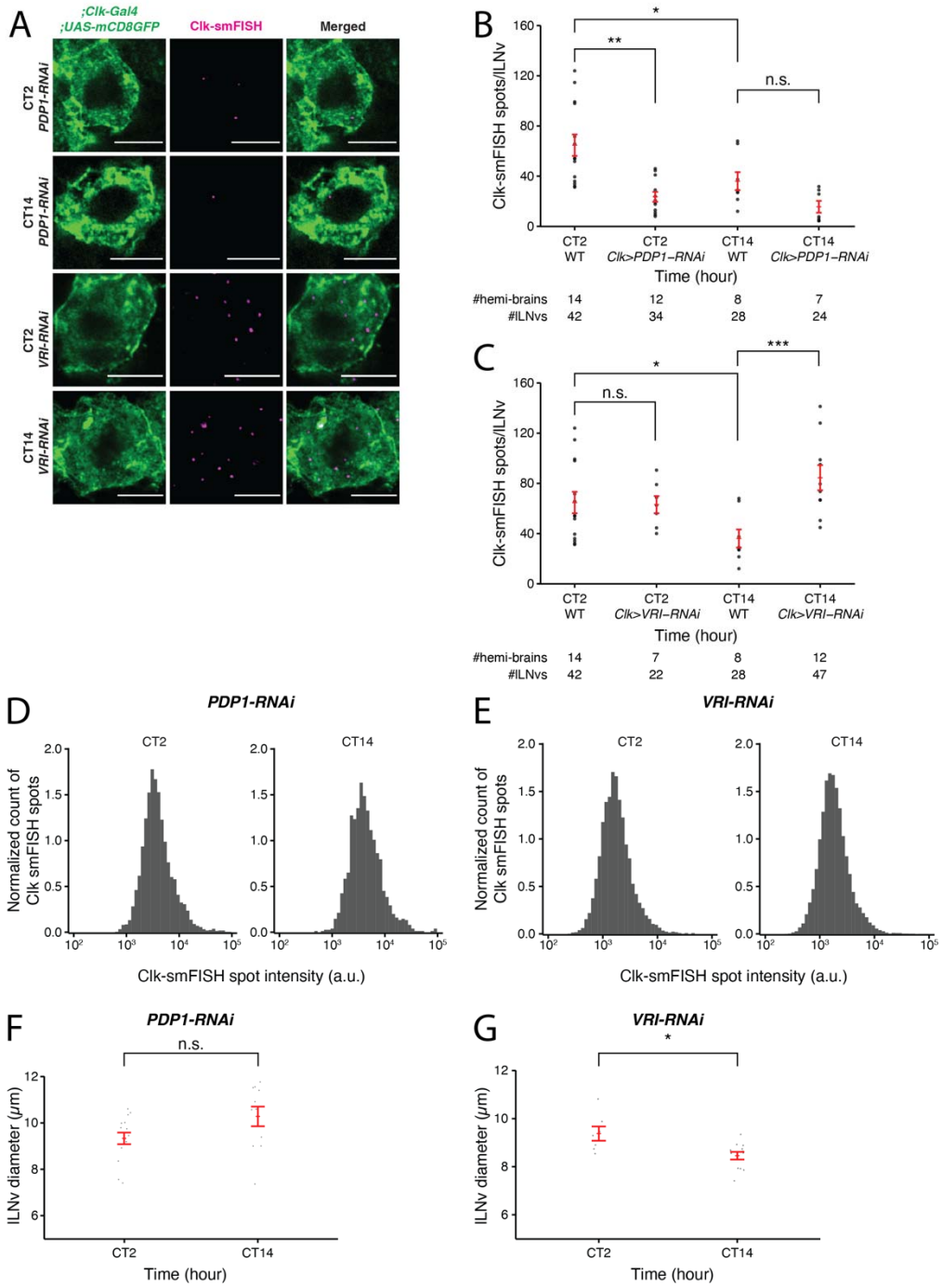
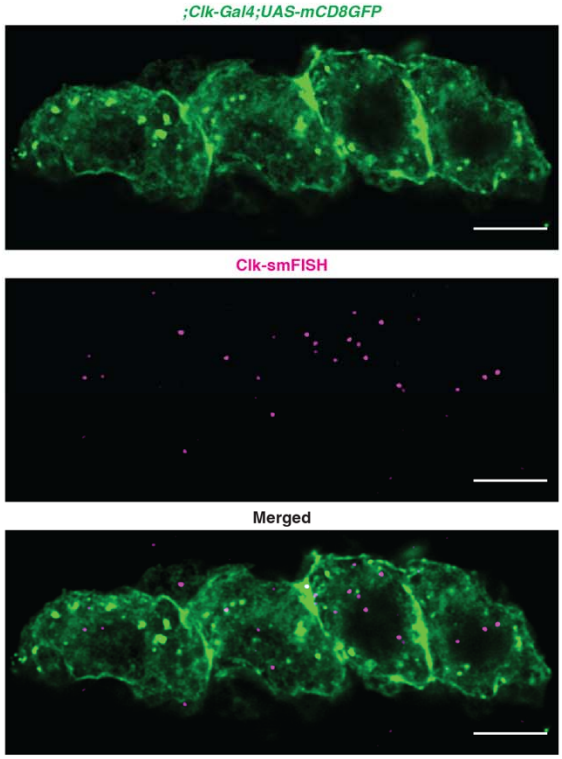

A
;Clk-Gal4
;UAS-mCD8GFP
Clk-smFISH
Merged
CT2 PDP1-RNAi
CT14 PDP1-RNAi
CT2 VRI-RNAi
CT14 VRI-RNAi
B
Clk-smFISH spots/ILNv
CT2 WT
CT2 Clk>PDP1-RNAi
CT14 WT
CT14 Clk>PDP1-RNAi
Time (hour)
#hemi-brains 14 12 8 7
#ILNvs 42 34 28 24
C
Clk-smFISH spots/ILNv
CT2 WT
CT2 Clk>VRI-RNAi
CT14 WT
CT14 Clk>VRI-RNAi
Time (hour)
#hemi-brains 14 7 8 12
#ILNvs 42 22 28 47
D
PDP1-RNAi
CT2
CT14
Normalized count of Clk smFISH spots
Clk-smFISH spot intensity (a.u.)
E
VRI-RNAi
CT2
CT14
Normalized count of Clk smFISH spots
Clk-smFISH spot intensity (a.u.)
F
PDP1-RNAi
n.s.
ILNv diameter (μm)
CT2
CT14
Time (hour)
G
VRI-RNAi
*
ILNv diameter (μm)
CT2
CT14
Time (hour)

Figure 5. Clk mRNA rhythm is disrupted in Pdp1 or vri RNAi mutants.

To further validate our workflow, we performed Clk-smFISH with Pdp1 or vri RNA interference mutants. **A)** Representative images of both mutants at wildtype peak (CT2) and trough (CT14) timepoints. **B,C)** We observed that knockdown of Pdp1 in clock neurons leads to a constant low level of Clk mRNA similar to wildtype CT14 while knockdown of vri leads to constant high Clk mRNA level similar to wildtype CT2. **D-G)** We observed overall consistent quality control metrics for all the experiment groups. Scale bars, 5μm.

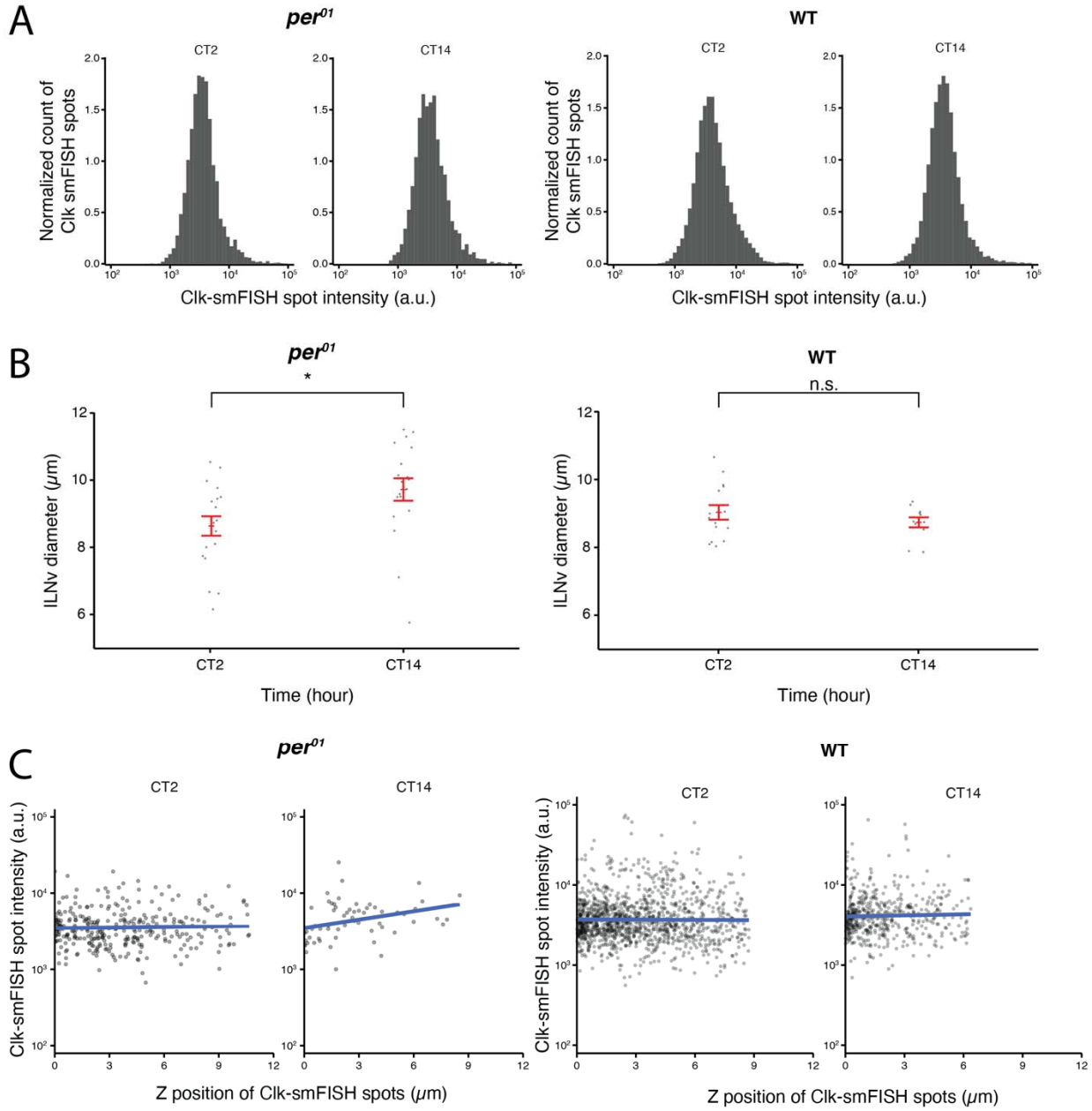
Supplementary Figure 1



Supplementary Figure 1. Zoomed-out representative images of Clk-smFISH shows overall clean signal.

Here we show a representative Z-slice of all four lLNvs observed in a hemi-brain from the wildtype ZT4 sample. Whether all cells can be observed depends on orientation of the brain as well as other experiment conditions. We notice that the *Clk* smFISH signal mostly shows good specificity. This image is a zoomed-out view of the ZT4 representative image shown in Fig. 3B. Scale bars, 5 μm.

Supplementary Figure 2



Supplementary Figure 2. Quality control plots for *Clk* smFISH spots in *per*^{01} null mutant. **A)** The *Clk* smFISH spot intensity plots at timepoints over the circadian cycle. **B)** We note that clock neuron marker signal is visibly affected in the period null mutant similar to the Pdp1 and

vri RNAi samples, which is also reflected in the lLNv diameter plot. However, we note again that the variation remains small (~10%) and possibly has biological origin. **C)** Sample distributions of *Clk* smFISH spot intensity relative to Z-position of the spots.