CHAPTER SEVENTEEN

RNA-seq Profiling of Small Numbers of *Drosophila* Neurons

Katharine Abruzzi*, Xiao Chen*, Emi Nagoshi†, Abby Zadina*, Michael Rosbash*,1

*Department of Biology, Howard Hughes Medical Institute and National Center for Behavioral Genomics, Brandeis University, Waltham, Massachusetts, USA
†Department of Genetics and Evolution, University of Geneva, Geneva, Switzerland
1Corresponding author: e-mail address: rosbash@brandeis.edu

Contents

Abstract

Drosophila melanogaster has a robust circadian clock, which drives a rhythmic behavior pattern: locomotor activity increases in the morning shortly before lights on (M peak) and in the evening shortly before lights off (E peak). This pattern is controlled by ~75 pairs of circadian neurons in the *Drosophila* brain. One key group of neurons is the M-cells (PDF^+ large and small LN_vs), which control the M peak. A second key group is the E-cells, consisting of four LN_ds and the fifth small LN_v, which control the E peak. Recent studies show that the M-cells have a second role in addition to controlling the M peak; they communicate with the E-cells (as well as DN_1s) to affect their timing, probably as a function of environmental conditions (Guo, Cerullo, Chen, & Rosbash, 2014).

To learn about molecules within the M-cells important for their functional roles, we have adapted methods to manually sort fluorescent protein-expressing neurons of interest from dissociated *Drosophila* brains. We isolated mRNA and miRNA from sorted M-cells and amplified the resulting DNAs to create deep-sequencing libraries. Visual inspection of the libraries illustrates that they are specific to a particular neuronal subgroup; M-cell libraries contain timeless and dopaminergic cell libraries contain ple/TH. Using these data, it is possible to identify cycling transcripts as well as many mRNAs and miRNAs specific to or enriched in particular groups of neurons.

Methods in Enzymology, Volume 551
ISSN 0076-6879
http://dx.doi.org/10.1016/bs.mie.2014.10.025

© 2015 Elsevier Inc.
All rights reserved.

1. INTRODUCTION

Circadian clocks allow organisms to predict and respond to daily fluctuations in their environments. In most organisms, these clocks oscillate with an ~24-h period and are entrained by environmental cues such as light. In *Drosophila*, the clock is driven by a several well-defined transcriptional feedback loops, one of which is focused on a heterodimer of the transcriptional factors CLK and CYC. CLK/CYC drives the transcription of the repressors PER and TIM in the early evening. PER and TIM levels accumulate and repress CLK/CYC-driven transcription in the late night. This negative feedback loop contributes to oscillating gene expression, which has a major impact on the circadian outputs including locomotor activity rhythms.

In *Drosophila*, there are ~75 pairs of neurons in the brain that express high levels of these clock components (CLK, CYC, PER, TIM) and are therefore considered circadian neurons. They have been divided into two main subgroups: dorsal neurons (DNs) and the lateral neurons (LNs). DNs are further subdivided into four groups based primarily on their location in the brain: DN_{1a}, DN_{1p}, DN_2, and DN_3. LNs have been divided into two main groups based on their expression of the neuropeptide pigment-dispersing factor (PDF; Helfrich-Forster, 1995): LPNs, LN_ds, and the fifth small LN_v are PDF^-, whereas the large and four of the five small LN_vs are PDF^+. Further experiments have shown that the PDF^+ s- and l-LN_vs are critical for driving the morning activity period in *Drosophila* and are known as the morning cells (M-cells). The PDF^- LN_ds and the 5th small LN_v are important for driving evening behavior and are known as the evening cells (E-cells; Grima, Chelot, Xia, & Rouyer, 2004; Stoleru, Peng, Agosto, & Rosbash, 2004). Immunostaining studies have begun to suggest the function(s) of these different groups, by revealing different expression patterns. For example, the circadian photoreceptor Cryptochrome (CRY) and different neuropeptides that impact the circadian system are differentially expressed within the circadian network (reviewed in Yoshii, Rieger, & Helfrich-Forster, 2012). In addition, different cell-specific drivers from the GAL4/UAS system have been used to manipulate subsets of these neurons with different UAS proteins to determine changes in circadian behavior. For example, electrical silencing of the M-cells causes a severe deficit in free-running locomotor rhythms (Depetris-Chauvin et al., 2011; Nitabach, Blau, & Holmes, 2002).

Studies in the last decade have provided further evidence that the control of circadian rhythms is not a simple case of attributing a specific task to a single group of neurons. Evidence indicates that interactions between different circadian neurons are necessary to achieve the complex regulation that drives circadian behaviors. M-cells are considered the master pacemakers since they can keep time in constant darkness, but they communicate with both the DN_1s and the E-cells via PDF signaling (Guo et al., 2014; Zhang, Chung, et al., 2010; Zhang, Liu, Bilodeau-Wentworth, Hardin, & Emery, 2010). Moreover, manipulating the E-cells as well as the DN_1s impacts the morning peak (Guo et al., 2014; Zhang, Chung, et al., 2010; Zhang, Liu, et al., 2010). The data indicate that the DN_1s and E-cells are downstream of PDF signaling, but they may also feed back to influence the M-cells.

This complexity indicates that it will be important to characterize subgroups of neurons and eventually single neurons. Although immunostaining has been a valuable tool to start to decipher the unique expression patterns of circadian neurons, a genome-wide view of differential gene expression patterns would greatly expand our vision. To this end, we and others manually sorted subgroups of neurons from dissociated *Drosophila* brains and used microarrays to assay neuron-specific gene expression (Kula-Eversole et al., 2010; Mizrak et al., 2012; Nagoshi et al., 2010). We have now used deep-sequencing technologies to sequence the mRNA and miRNA populations of these M-cells. We also show here a bit of data from a large group of noncircadian neurons (e.g., dopaminergic: ~130 neurons per brain; Mao & Davis, 2009) as well as the smaller group of M neurons (M-cells; l- and s-LN_vs; ~14 neurons/brain). By identifying mRNAs that are enriched and/or undergo cycling in particular groups of neurons, we hope to learn more about the roles of these neurons in contributing to particular aspects of circadian rhythms.

2. RESULTS/METHODS

2.1. Isolating neurons of interest

To isolate neurons of interest, we express a fluorescent protein in a specific subset of neurons and then manually sort these fluorescent neurons from dissociated brains. One of the key steps of this procedure is ensuring that the fluorescent protein (1) is sufficiently bright to make cell sorting possible and (2) has no leaky expression outside of the neurons of interest. Although we

have used UAS-MCD8-GFP (Lee & Luo, 1999; Bloomington Stock Center #56168) and UAS-EGFP (Bloomington stock center #1522) effectively, EGFP is generally brighter. We have used this strategy to successfully sort a variety of neurons. They include large subsets of the brain such as all ELAV-expressing cells (*elav*-GAL4, UAS-EGFP flies) or dopaminergic cells (*TH*-GAL4, UAS-EGFP) as well as smaller subsets of neurons such as M-cells (*pdf*-GAL4, UAS-MCD8GFP) and E-cells (*Dv-pdf*-GAL4, UAS-EGP, *pdf*-RFP).

Neurons are isolated from adult fly brains at different circadian times essentially as described previously by Nagoshi et al. (2010). Young adult flies are entrained for 4 days in 12:12 LD cycles and harvested every 4 h to collect six timepoints throughout the day. Flies are chilled on ice and ~100 brains (for M-cells; fewer brains needed if cells are abundant, e.g., ~10 brains are used for *elav*-GAL4, UAS-EGFP) are dissected in cold dissecting saline (9.9 m*M* HEPES-KOH buffer, 137 m*M* NaCl, 5.4 m*M* KCl, 0.17 m*M* NaH_2PO_4, 0.22 m*M* KH_2PO_4, 3.3 m*M* glucose, 43.8 m*M* sucrose, pH 7.4) containing 20 μ*M* 6,7-dinitroquinoxaline-2,3-dione (DNQX; Sigma), 50 μ*M*D(−)-2-amino-5-phosphonovaleric acid (AP5/APV; Sigma), and 0.1 μ*M* tetrodotoxin (Tocris). Brains are collected in cold SM-active Bis–Tris media (SM^{active} media plus the same drugs included in the saline) and are centrifuged at room temperature for 2 min at 1000 rpm, the supernatant is removed, and the brains are washed in 300 μl cold dissecting saline. Approximately 2 μl/brain of L-cysteine-activated papain (50 units/ml, Worthington; heat activated at 37 °C for 10 min) is added, and samples are incubated at room temperature for 20–30 min with occasional mixing using a 20-μl pipette tip. A fivefold volume of cold SM-active Bis–Tris media is added to quench the digestion. Brains are centrifuged for 2 min at room temperature and resuspended in 7 μl of cold SM-active Bis–Tris media per brain (with a minimum volume of 400 μl to prevent difficulties in trituration).

To break apart the digested brains, we make a collection of flame-rounded 1-ml filter tips with either large, medium, or small orifices left at the tip after flaming. Brains are triturated by pipetting up and down 30 × with a large flame-rounded tip, 20 × with the medium flame-rounded tip, and 10 × with the small flame-rounded tip. To ensure that the samples stay chilled, samples are placed on ice after every 10 iterations of pipetting up and down. The amount to triturate is not exact; it will differ depending on the number of brains. The small tip should be used until the liquid goes through easily without getting stuck. Do not over triturate since it can cause cells to burst.

Once cells are dissociated, they are placed in Sylgard plates (Sylgard 184 Elastomer Base and Curing Agent; Dow Corning) for sorting under the fluorescence scope (Leica M165 FC). Cold SM-active Bis–Tris media are added to Sylgard plates: 2.5 ml into two 6-cm Sylgard plates and 1.5 ml into two 3.5-cm plates. Half of the cell suspension is added to the center of each of the 6-cm plates. The plates are incubated on ice for 20–30 min so that cells can settle. Cells are sorted using micropipettes pulled from capillary tubes (World Precision Glass Capillaries #1B100-4) using a micropipette puller (Sutter Instrument Company). Before use, the micropipette tips are broken by punching them through a Kimwipe until they are ~1 cell wide. Cells are identified under the scope and then a cell aspirator (described in Hempel, Sugino, & Nelson, 2007) is used to move them to a Sylgard plate filled with fresh buffer. Since there is some probability that a nonfluorescent cell is accidently aspirated with a fluorescent cell, the cells undergo three rounds of sorting before being placed in a 0.2-ml tube. Approximately 60–100 cells are sorted for each timepoint and cell type of interest. To isolate microRNAs from specific neurons, at least 100 cells are placed in a 0.2-ml tube with 100 μl TRIzol (Invitrogen) and frozen at −80 °C. To isolate and amplify mRNA from neurons of interest, the isolated neurons were placed in 50 μl of lysis/binding buffer (Invitrogen; Dynabeads mRNA direct kit) and frozen at −80 °C.

2.2. Amplification of mRNA

From ~100 manually sorted neurons, we can isolate approximately 200–500 pg of total RNA or ~5 pg of mRNA. In order to generate libraries for RNA deep sequencing (Illumina HiSeq 2000), this RNA needs to be amplified >2000-fold to generate between 10 and 100 ng of mRNA (recommended starting amount Illumina RNA Tru-Seq v2). To this end, we have modified a linear amplification method traditionally used to amplify mRNA for microarray analysis (Kula-Eversole et al., 2010; Nagoshi et al., 2010; Sugino et al., 2006) and currently utilized in single-cell sequencing approaches (Hashimshony, Wagner, Sher, & Yanai, 2012). In this method, small amounts of RNA are reverse transcribed using dT primers containing a T7 promoter to make a cDNA template that can be used for multiple rounds of linear amplification using *in vitro* transcription. Although this strategy works well, it only amplifies the 3′-most end of the mRNA. To overcome this limitation, we isolated the mRNA population on oligo-dT beads and then made the cDNA template for *in vitro* transcription using both

oligo-dT T7 and random-T7 primers (random deoxyribonucleotides combined with a T7 promoter sequence) in an attempt to circumvent the 3′ bias. We settled on this strategy after trying dT-T7 priming alone, random-T7 priming alone, and the mixture of the two. Indeed, we found that dT-T7 priming typically led to strongly 3′-biased libraries in which the 5′-ends of mRNAs are not represented (Fig. 1; *Act5c*). Random-T7 priming alone led to the underrepresentation of short mRNA molecules and the 3′-ends of longer mRNAs (Fig. 1; *RpS26* and 3′-end of *Act5c*). A mixture of random-T7 and dT-T7 did the best job at balancing the 5′- and 3′-ends, and we therefore moved forward with this approach (Fig. 1, note the presence of 5′-most exon of *Act5c*).

Poly-A RNA is isolated from ~100 neurons using Dynabeads oligo-dT using a scaled-down version of the manufacturer's instructions (Dynabeads

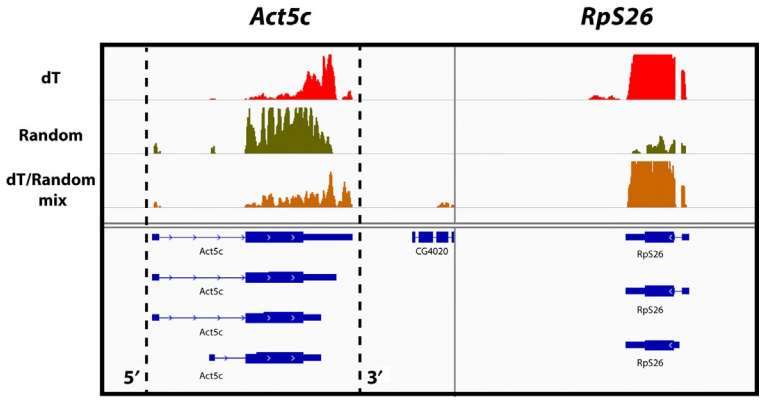

Figure 1 RNA is amplified using a mix of random and dT-T7 primers. Fluorescent cells were sorted from dissociated brains of *elav*-GAL4, UAS-EGFP flies, and mRNA was isolated from ~100 cells. The mRNA was reverse transcribed into cDNA using dT-T7 primers (top), random-T7 primers (middle), or a mix of both dT- and random-T7 primers (bottom). The resulting cDNA was used as a template for *in vitro* transcription and sequencing libraries were made. The resulting data for *Act5c* and *RpS26* are shown here using the Integrative Genomics Viewer (IGV; Robinson et al., 2011; Thorvaldsdottir et al., 2013). Solid blue boxes (dark gray in the print version) indicate exons and the intervening lines represent introns. *Act5c* is on the top strand and transcription is going from left to right; the 3′- and 5′-ends of the *Act5c* mRNA are denoted with dashed lines. *RpS26* is on the bottom strand and transcription is going from right to left. Priming with dT-T7 yields 3′-biased libraries that lack signal in the 5′-most exon of *Act5c*. Priming with random-T7 primer shows good coverage of the 5′-most exon of *Act5c*, but the 3′-end of the mRNA is not represented. Small mRNAs such as RpS26 are underrepresented with random priming. A mixture of dT and random priming leads to a decrease in 3′-bias and the coverage of both 3′-ends of mRNAs and shorter mRNAs.

mRNA direct kit; Invitrogen). The frozen cell suspension is thawed on ice, brought to 100 µl using lysis/binding buffer, and mixed well by pipetting. The lysate is transferred to 20 µl of washed Dynabeads (Dynabeads are prewashed 1 × in lysis/binding buffer and then resuspended in 20 µl of lysis/binding buffer), mixed gently by pipetting up and down, and then rotated at room temperature for 5 min. Samples are placed on the magnet (Invitrogen; DynaMag-2 Magnet) and the supernatant discarded. Beads are washed 2 × with 100 µl of wash buffer A, resuspended in wash buffer B, and transferred to a new tube. Samples are then washed 1 × in ice-cold 10 m*M* Tris–HCl before being resuspended in 10 µl of 10 m*M* Tris–HCl and incubated in a heat block at 70 °C for 2 min. The poly-A mRNA-containing supernatant is moved to a 0.2-ml tube and concentrated to 1 µl using a Speed-Vac (RC1010; Jouan, Winchester, VA). It is important not to let the samples dry out.

The resulting mRNA is reverse transcribed using a mixture of dT and random primers fused to a T7 promoter to generate a double-stranded cDNA to be used as a template for *in vitro* transcription. dT-T7 (GGCCAGTGAATTGTAATACGACTCACTATAGGGAGGCGGT(24)) and random-T7 (GGCCAGTGAATTGTAATACGACTCACTATAGGG AGGCGGN(20)) were added to mRNA at a concentration of 1.25 µ*M* each, and SuperScript III (Invitrogen) was used to generate double-stranded cDNA according to the manufacturer's guidelines. The resulting double-stranded cDNA was ethanol-precipitated overnight and washed 2 × in 75% ethanol. The cDNA is resuspended in 5 µl of RNase-free H_2O.

To validate the cell sorting, 1 µl of the cDNA can be diluted 1:30 and used as a template for q-PCR to determine whether the sample is enriched for genes of interest and/or shows cycling gene expression. For example, when cDNA is made from six timepoints of M-cells (*pdf*-GAL4, UAS-MCD8 GFP), PDF is greatly enriched compared to a similar experiment done with a more heterogeneous group of neurons isolated using the *elav*-GAL4, UAS-GFP line (Fig. 2A). Indeed, q-PCR from the same M-cell samples shows the circadian oscillation of the *tim* mRNA whose expression peaks in the evening (ZT14; Fig. 2B; there are six timepoints, so the data are double-plotted).

The remaining cDNA is used as a template for *in vitro* transcription using T7 polymerase (MEGAscript kit; Ambion) in a 10-µl reaction. Samples are incubated overnight at 37 °C. The resulting mRNA is isolated using RNA MinElute columns (Qiagen) and eluted in 14 µl of RNase-free H_2O. This typically yields between 10 and 50 ng/µl when quantified using

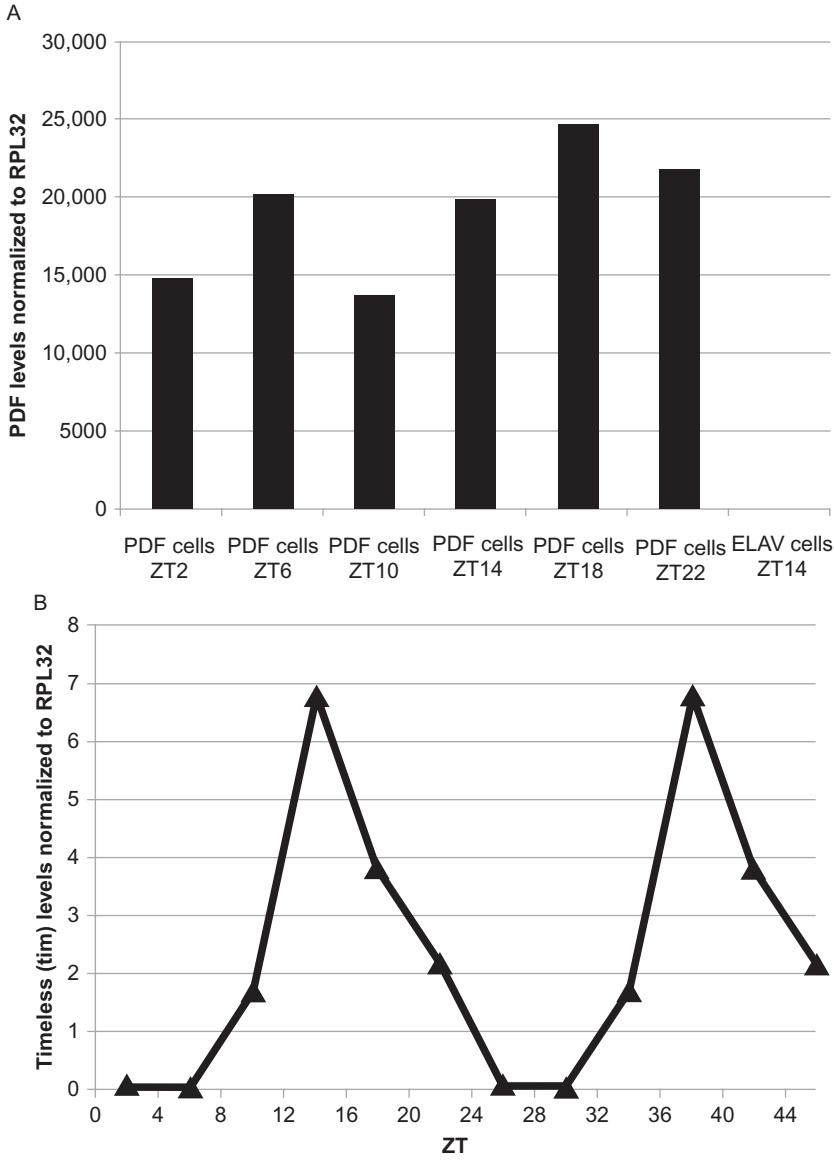

Figure 2 cDNA isolated from M-cells shows both an enrichment for PDF and cycling of tim. Approximately 100 fluorescent cells were isolated from dissociated brains of *pdf*-GAL4, UAS-MCD8-GFP flies or *elav*-GAL4, UAS-EGFP at timepoints throughout the day. mRNA was isolated from these cells and reverse transcribed using a combination of dT- and random-T7 primers. The resulting cDNA was diluted and used as a template for q-PCR using primers for *PDF*, *tim*, and *RPL32* (for normalization purposes). (A) Cells sorted from *pdf*-GAL4, UAS-MCD8-GFP are highly enriched for PDF in comparison to cells isolated from *elav*-GAL4, UAS-EGFP. (B) tim mRNA cycles in cells isolated from *pdf*-GAL4, UAS-MCD8-GFP. Tim levels peak in early evening (ZT14). Data from six timepoints are shown double-plotted.

the NanoDrop (Thermo Scientific). 100 ng of this sample is concentrated to 2 μl using a Speed-Vac and used as input for the RNA Tru-Seq library generation protocol (Illumina). To be more cost-effective, the Tru-Seq library kit was used according to manufacturer's recommendations except that one-third volume was used for every step. The size of the libraries is verified using a High Sensitivity DNA kit on the Bioanalyzer (Agilent), and library concentration is determined by using q-PCR using the PCR primers (Illumina Tru-Seq kit, forward primer: AATGATACGGCGACCACCGA, reverse primer: CAAGCAGAAGACGGCATACGA) and a library of known concentration. A 2 n*M* mixture of six different barcoded libraries is mixed together, and 10 p*M* of this sample is sequenced in a single lane of the Illumina HiSeq 2000.

The resulting data were parsed to separate the barcoded reads, and the resulting sequence files (fastq format) were aligned to the *Drosophila* genome (version dm3) via TopHat (Trapnell, Pachter, & Salzberg, 2009) using the following criteria: −m 1 -F 0 -p 6 -g 1 –microexon-search –no-closure-search –solexa-quals -I 50000. In our initial experiments, the sequencing reads from RNA amplified from sorted neurons mapped very poorly to the *Drosophila* genome, i.e., only approximately 10% of the libraries corresponded to the *Drosophila* genome due to contamination (see Section 3). By cleaning all work surfaces (PCR machines, centrifuges, pipettes, etc.) extremely carefully prior to each experiment (DNA-Off; Takara), we were able to increase the percentage of our libraries that mapped to the transcriptome (non-rRNA) to ∼40–70%. Visualization files were generated by converting the bam output file (TopHat) to a bigwig file using a custom script. These files were viewed using the Integrated Genomics Viewer (Robinson et al., 2011; Thorvaldsdottir, Robinson, & Mesirov, 2013).

To validate the cell sorting and deep-sequencing approach, we visually examined the sequencing results for known neuronal group specific genes. As a simple example, we compared the expression of the circadian gene *timeless* (*tim*) and the dopaminergic cell-enriched enzyme tyrosine hydroxylase (*TH* or *pale*) in sequencing data derived from M-cells or dopaminergic cells (*pdf*-GAL4, UAS-MCD8-GFP, and *TH*-GAL4, UAS-EGFP, respectively). Figure 3 shows the IGV browser showing the number of sequencing reads found for the housekeeping gene *rpl32* as well as *TH* and *tim*. Similar levels of Rpl32 are found in both M-cells and dopaminergic cells. As expected, only dopaminergic cells express *pale* (*TH*). There are two main isoforms of *ple*; dopaminergic cells preferentially express the central nervous system-specific isoform that lacks exons 3 and 4 and not the hypodermal

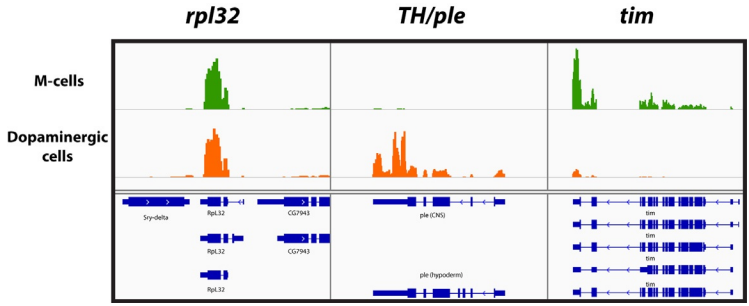

Figure 3 RNA sequencing data from neuronal groups show specificity. Sequencing data from sorted M-cells (*pdf*-GAL4, UAS-MCD8-GFP) or dopaminergic cells (*TH*-GAL4) are visualized on the IGV. Rpl32 shows similar levels in both M-cells and dopaminergic cells. In contrast, ple/TH (tyrosine hydroxylase) is only detected in dopaminergic cells. As expected, dopaminergic neurons show the central nervous system-specific isoform of *ple* that is lacking exons 3 and 4. The circadian mRNA tim is only found in M-cells and not in dopaminergic cells.

form (Fig. 3; Friggi-Grelin, Iche, & Birman, 2003). In addition, only the circadian M-cells express substantial levels of *timeless*.

It is also possible to detect cycling gene expression and differential gene expression in sequencing data from specific neuronal populations. As shown earlier (Fig. 2), cycling *tim* mRNA was observed in the cDNA made directly from M-cell mRNA. Not surprisingly, cycling *tim* mRNA is also easily detected in the sequenced libraries from *in vitro* transcription of this cDNA. Figure 4 shows *tim* gene expression in six timepoints of RNA from M-cells with peak expression at ZT14 as expected. In addition to the core clock genes, other cycling genes have been identified in M-cells. The inward-rectifying potassium channel, Ir, is one example of such a gene. *Ir* mRNA cycles in M-cells with expression peaking at ZT12 as seen with microarrays (Kula-Eversole et al., 2010; Mizrak et al., 2012). There are also genes that are expressed predominantly in either M-cells or E-cells (Fig. 5). CG18343 is an unknown gene detected in mRNA isolated from E-cells but not M-cells. The tetraspanin, *Tsp42Eo*, shows the reverse pattern; it is detected in M-cells but not E-cells.

2.3. Amplification of miRNA

The method for making miRNA libraries from isolated neurons is adapted from Hafner et al. (2012) as well as the Mello Lab Small RNA Cloning Protocol (http://www.umassmed.edu/PageFiles/43096/Mello%20lab%

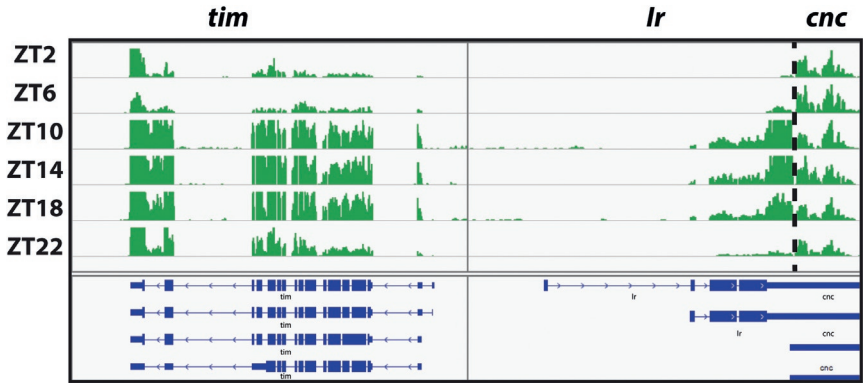

Figure 4 *Timeless (tim)* and *Ir* cycle in M-cells. RNA from M-cells isolated at six different timepoints throughout the day was isolated, amplified, and made into Illumina sequencing libraries. The resulting sequencing data were visualized using the IGV. Data are shown for *tim* (bottom strand; transcription going from right to left), *Ir* (top strand, transcription going from left to right), and *cnc* (bottom strand, transcription going from right to left). Solid blue boxes (dark gray in the print version) indicate exons and the intervening lines represent introns. The dashed line indicates the location where the *Ir* and *cnc* 3′-ends overlap. Tim levels cycle throughout the day in M-cells with the highest expression detected at ZT14. Ir levels also cycle in M-cells with peak phase at ZT14. In contrast, the adjacent gene *cnc* shows equal expression throughout the day.

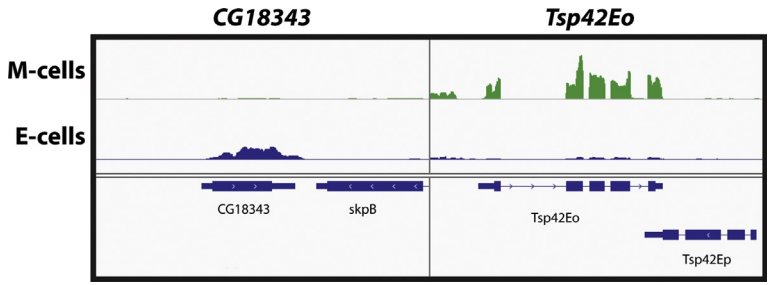

Figure 5 Identification of differentially expressed genes in E-cells and M-cells. RNA-seq libraries from isolated M-cells and E-cells were compared on the IGV. There are many genes that are differentially expressed in subsets of neurons. Two examples are shown here. CG18343 is expressed more strongly in E-cells than in M-cells. In contrast, *Tsp42Eo* is expressed more strongly in M-cells than E-cells.

20small%20RNA%20cloning%20protocol.pdf). In brief, total RNA is extracted from isolated neurons as described above, polyadenylated adaptors are ligated to the 3′-end, and the RNA is size-selected. 5′-Adaptors are then ligated, and the resulting RNAs are reverse transcribed to generate cDNA that can be amplified by PCR. A second size-selection helps to ensure that

cDNA contains primarily miRNAs and not other similar-sized RNA contaminants. Finally, libraries are made and sequenced as described in detail below.

To extract total RNA, cells are thawed and lysed by pipetting up and down several times. 100 μl of TRIzol (Invitrogen) is added, and total RNA is extracted according to the manufacturer's instructions with 40 μl chloroform and precipitated overnight at −20 °C. GlycoBlue (Life Technologies) is used as coprecipitant. Total RNA is precipitated the next day by centrifuging at maximum speed at 4 °C for 30 min, washed with 75% ethanol, and dried for approximately 5 min at room temperature. The amount of total RNA obtained here may be too low to be detected with the NanoDrop (Thermo Scientific), but the quality of the total RNA should be assessed using the RNA 6000 Pico kit in combination with the Bioanalyzer (Agilent).

3′ Preadenylated adaptors are then ligated to the total RNA using T4 RNA ligase 2, truncated (NEB; rAppTGGAATTCTCGGGTGCCAAGG/ddC/; adaptor is specific to Illumina HiSeq 2000 but other adaptors could be designed for other platforms). The adaptors should be at least sixfold in excess of the total RNA to optimize ligation efficiency. 10% DMSO can be added to denature the RNA. A longer incubation time, for example, 6 h, can also increase ligation efficiency. Ligated RNA products are then fractionated on urea gels (Novex precast 15% TBE-UREA gels (Life Technologies) run at 180 V for 40 min in TBE buffer) to size-select miRNAs (Fig. 6A). Since *Drosophila* 2S rRNA (30 nt) is close in size to the miRNA fraction (18–27 nt), it is critical to remove the 2S rRNA region to avoid contamination of the libraries. If the adaptor described above is used, the region corresponding to 40–50 nt is extracted from the gel with the traditional "crush and soak" method or electroelution. The precipitated RNA products are then ligated to 5′ RNA adaptors (5′-GUUCAGAGUU CUACAGUCCGACGAUC-3′) with T4 RNA ligase (Ambion). The efficiency of this step is usually very low, and so overnight incubation at room temperature is advisable.

After 5′ ligation, RNA products are reversely transcribed with SuperScript II with primers that are complementary to the 3′ adaptors. cDNAs are then amplified with Phusion High-Fidelity DNA polymerase (NEB) with universal forward primers and indexed reverse primers for barcoding (Illumina Customer Sequence Letter, Section: Oligonucleotide sequences for Tru-Seq™ Small RNA Sample Prep Kits) for 12–15 cycles. The number of cycles in the PCR program is critical, because overamplification will

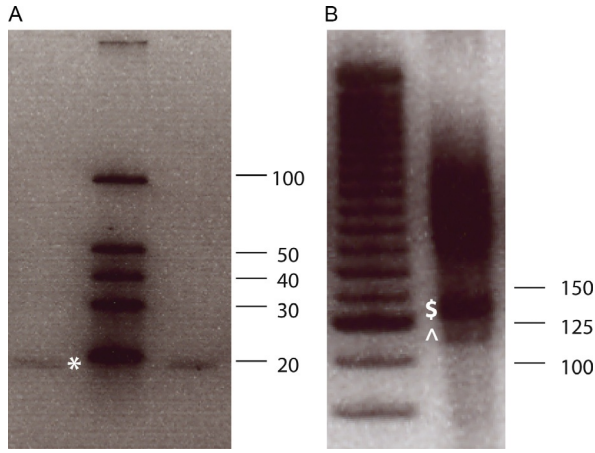

Figure 6 Preparation and amplification of miRNAs from sorted neurons. (A) Adaptor-bound total RNA is fractionated on a 15% TBE-urea gel (Novex) stained with cyber gold (Life Technologies). * indicates excess 3′ preadenylated adaptors. The 3′ ligated products are too little to be observed at this step. A region corresponding to ~40–50 bp is cut from the gel. (B) Amplified products are size-selected on a 4–12% TBE gel (Novex; Life Technologies) stained with cyber gold. ^ indicates self-ligated adaptors. $ indicates amplified miRNA libraries. Some high-molecular weight smear is visible here due to overamplification.

generate large amounts of a high-molecular weight smear, which affects library quality.

The amplified products are then subjected to a second size-selection to remove unwanted products, including self-ligated adaptors, residual amplified 2S rRNA products, and the high-molecular weight smear (Fig. 6B). If the exact adaptors and primers are used as described in this protocol, the DNA from the region containing 130–140 bp products should be extracted as previously described. If no bands or only faint bands are observed in the gel in the 130–140 bp region, a wider region (125–150 bp) can be extracted and amplified a second time.

Quality and quantity of miRNA libraries are assessed with an Agilent bioanalyzer using either the DNA 1000 (Fig. 7A) or a high-sensitivity DNA Chip. miRNA libraries should then be diluted to 2 n*M* with EB buffer and mixed together equally as described for mRNA libraries. Libraries are sequenced on an Illumina HiSeq 2000 and the resulting data analyzed using custom scripts. Data from a miRNA library made from sorted EGFP-labeled M-cells are shown in the IGV (Fig. 7B, top). The miRNA library shows sequencing reads for mir-14 but not for any of the protein-encoding genes

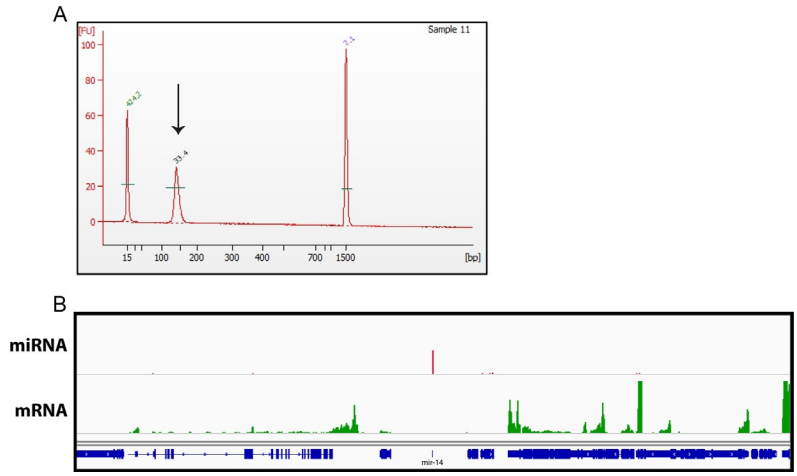

Figure 7 Visualization of miRNA sequencing libraries. (A) Bioanalyzer profile of miRNA sequencing library on a DNA 1000 chip (Agilent). The library (labeled with a black arrow) is ~150 bp in size and shows a clear, defined peak. (B) miRNA libraries viewed on the IGV. Sequencing data from a miRNA and mRNA library made from isolated M-cells are shown in the IGV centered on the miR-14 locus. miR-14 is visible in the miRNA library but there is no signal from the surrounding mRNAs. In contrast, mir-14 signal is not present in the mRNA library.

surrounding it. In contrast, a mRNA library from M-cells is shown below in green (gray in the print version) (Fig. 7B).

3. DISCUSSION

Recent studies suggest that the control and regulation of circadian behavior is due to the coordinated response of several different groups of circadian neurons acting as a network. One critical part of dissecting this network (or any network) is learning more about the role of specific cells and groups of cells. We present here a method for isolating specific groups of neurons from the *Drosophila* brain and using deep sequencing to profile their miRNA and mRNA populations present in these cells. Any neuronal population marked with a fluorescent protein can be manually sorted from dissociated *Drosophila* brains. In this chapter, we show data from both a medium (dopaminergic) and a small (M-cells) population, about 130 and 20 cells/brain, respectively. The two populations show the expected divergent gene expression patterns for two well-characterized genes and many others (not shown), i.e., only M-cells express well the circadian gene *tim* and only

dopaminergic cells express *TH* or *ple* (Fig. 3). It is unclear whether the small amount of *tim* signal from dopaminergic cells reflects cell contamination or bona fide clock gene expression.

In addition to identifying cell-specific gene expression patterns, we have also identified cycling and differential gene expression using this approach. Both *tim* and *Ir* mRNAs cycle throughout the day in M-cells. There are also genes that are expressed in one subgroup of neurons but not another. Two such examples are shown in Fig. 5: CG18343 and *Tsp42Eo*. We also describe the isolation of miRNAs from M-cells. This approach allows for the analysis of neuronal miRNAs, e.g., changes in level with circadian time and/or light and firing, and allows a dissection of their role in circadian and neuronal posttranscriptional gene regulation.

The ability to manually sort neurons from dissociated *Drosophila* brains is not new. Several studies published in the past years have illustrated the usefulness of this approach in combination with microarrays (Kula-Eversole et al., 2010; Mizrak et al., 2012; Nagoshi et al., 2010). The advent of deep sequencing has made it appealing to extend these previous studies and adapt existing methods for this purpose. Since *Drosophila* microarrays contain antisense probes for the 3′-ends of transcripts, amplifying mRNA for this assay only requires material from the 3′-end of the mRNA. In addition, contamination that often occurs when amplifying very small amounts of RNA (picograms) is not an issue because most contaminants, i.e., from human sources, would not hybridize to the *Drosophila* microarray probes. Therefore, adapting existing methods to deep sequencing presents two major hurdles: (1) creating an mRNA library that contains information beyond the 3′-end of the transcript and (2) preventing contamination in both mRNA and miRNA libraries.

All current mRNA amplification methods used in conjunction with deep sequencing start with a reverse transcription step in which the RNA is made into a cDNA template. In some approaches, only the 3′-ends are amplified using a T7-dT oligo (Hashimshony et al., 2012), whereas in other strategies more full-length transcripts can be obtained using strand switching (Deng, Ramskold, Reinius, & Sandberg, 2014; Picelli et al., 2014; Ramskold et al., 2012). We have tried to obtain less 3′-biased libraries by poly-A selecting mRNA from isolated cells and amplifying using a combination of dT- and random-T7 primers. Although these changes have allowed us to obtain sequencing information along the entire length of genes, there is still a 3′-bias present in the libraries. It is not clear whether this bias comes from the amplification method and/or from partial

degradation of the RNA during the cell sorting procedure followed by oligo-dT selection.

Contamination of both mRNA and miRNA libraries can be a substantial problem since libraries generated can contain non-*Drosophila* sequences that dramatically reduce the amount of usable data. (In our early experiments, as much as 90% of the sequencing reads did not map to the *Drosophila* genome). It is recommended to have a separate clean room and/or a PCR workstation to avoid contamination. However, it is possible to drastically reduce contamination by working extremely carefully and cleaning all equipment prior to each experiment using DNA-OFF (Takara).

Deep-sequencing data provide us with much more information than was gleaned from microarray data. The mRNA-seq data can be analyzed to reveal much more detail, for example, isoform-specific expression, differential splicing, and mRNA editing. As we and others have examined many of these processes in larger more heterogeneous *Drosophila* tissues (Brown et al., 2014; Khodor et al., 2011; Rodriguez, Menet, & Rosbash, 2012; St Laurent et al., 2013), it is possible that studies of neuronal populations will reveal novel and perhaps important regulatory events that were masked in whole head or whole body studies.

miRNA-seq will identify miRNAs that are present in specific neurons. Targets can be experimentally identified in many ways, for example, by depleting the miRNA of interest using miRNA sponges (Ebert, Neilson, & Sharp, 2007; Loya, Lu, Van Vactor, & Fulga, 2009) and performing mRNA-seq from the same neuronal population. These approaches could also be used in conjunction with different environmental stimuli (light, feeding, etc.) to determine how the mRNA and miRNA profiles of these neurons respond to these stimuli.

New drivers will certainly be available in the near future to express EGFP in only one or two key neurons, opening the possibility of understanding the circuit at a single-neuron level. These more specific drivers would also facilitate single-neuron sequencing. We note in this context that RNA-seq has been used successfully to profile gene expression in single cells (Hashimshony et al., 2012; Ramskold et al., 2012; Tang et al., 2009). Adapting these methods to *Drosophila* neurons would allow us to accompany behavioral experiments and define each key cell by its gene expression profile, creating a complete and accurate map of where and at what circadian time specific genes are being expressed in the circadian neuronal system.

REFERENCES

Brown, J. B., Boley, N., Eisman, R., May, G. E., Stoiber, M. H., Duff, M. O., et al. (2014). Diversity and dynamics of the Drosophila transcriptome. *Nature*, *512*, 393–399.

Deng, Q., Ramskold, D., Reinius, B., & Sandberg, R. (2014). Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science*, *343*(6167), 193–196.

Depetris-Chauvin, A., Berni, J., Aranovich, E. J., Muraro, N. I., Beckwith, E. J., & Ceriani, M. F. (2011). Adult-specific electrical silencing of pacemaker neurons uncouples molecular clock from circadian outputs. *Current Biology*, *21*(21), 1783–1793.

Ebert, M. S., Neilson, J. R., & Sharp, P. A. (2007). MicroRNA sponges: Competitive inhibitors of small RNAs in mammalian cells. *Nature Methods*, *4*(9), 721–726.

Friggi-Grelin, F., Iche, M., & Birman, S. (2003). Tissue-specific developmental requirements of Drosophila tyrosine hydroxylase isoforms. *Genesis*, *35*(4), 260–269.

Grima, B., Chelot, E., Xia, R., & Rouyer, F. (2004). Morning and evening peaks of activity rely on different clock neurons of the Drosophila brain. *Nature*, *431*(7010), 869–873.

Guo, F., Cerullo, I., Chen, X., & Rosbash, M. (2014). PDF neuron firing phase-shifts key circadian activity neurons in Drosophila. *Elife*, *3*, e02780.

Hafner, M., Renwick, N., Farazi, T. A., Mihailovic, A., Pena, J. T., & Tuschl, T. (2012). Barcoded cDNA library preparation for small RNA profiling by next-generation sequencing. *Methods*, *58*(2), 164–170.

Hashimshony, T., Wagner, F., Sher, N., & Yanai, I. (2012). CEL-Seq: Single-cell RNA-Seq by multiplexed linear amplification. *Cell Reports*, *2*(3), 666–673.

Helfrich-Forster, C. (1995). The period clock gene is expressed in central nervous system neurons which also produce a neuropeptide that reveals the projections of circadian pacemaker cells within the brain of Drosophila melanogaster. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(2), 612–616.

Hempel, C. M., Sugino, K., & Nelson, S. B. (2007). A manual method for the purification of fluorescently labeled neurons from the mammalian brain. *Nature Protocols*, *2*(11), 2924–2929.

Khodor, Y. L., Rodriguez, J., Abruzzi, K. C., Tang, C. H., Marr, M. T., 2nd., & Rosbash, M. (2011). Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. *Genes & Development*, *25*(23), 2502–2512.

Kula-Eversole, E., Nagoshi, E., Shang, Y., Rodriguez, J., Allada, R., & Rosbash, M. (2010). Surprising gene expression patterns within and between PDF-containing circadian neurons in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(30), 13497–13502.

Lee, T., & Luo, L. (1999). Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. *Neuron*, *22*(3), 451–461.

Loya, C. M., Lu, C. S., Van Vactor, D., & Fulga, T. A. (2009). Transgenic microRNA inhibition with spatiotemporal specificity in intact organisms. *Nature Methods*, *6*(12), 897–903.

Mao, Z., & Davis, R. L. (2009). Eight different types of dopaminergic neurons innervate the Drosophila mushroom body neuropil: Anatomical and physiological heterogeneity. *Frontiers in Neural Circuits*, *3*, 5.

Mizrak, D., Ruben, M., Myers, G. N., Rhrissorrakrai, K., Gunsalus, K. C., & Blau, J. (2012). Electrical activity can impose time of day on the circadian transcriptome of pacemaker neurons. *Current Biology*, *22*(20), 1871–1880.

Nagoshi, E., Sugino, K., Kula, E., Okazaki, E., Tachibana, T., Nelson, S., et al. (2010). Dissecting differential gene expression within the circadian neuronal circuit of Drosophila. *Nature Neuroscience*, *13*(1), 60–68.

Nitabach, M. N., Blau, J., & Holmes, T. C. (2002). Electrical silencing of Drosophila pacemaker neurons stops the free-running circadian clock. *Cell*, *109*(4), 485–495.

Picelli, S., Faridani, O. R., Bjorklund, A. K., Winberg, G., Sagasser, S., & Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. *Nature Protocols*, *9*(1), 171–181.

Ramskold, D., Luo, S., Wang, Y. C., Li, R., Deng, Q., Faridani, O. R., et al. (2012). Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nature Biotechnology*, *30*(8), 777–782.

Robinson, J. T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., et al. (2011). Integrative genomics viewer. *Nature Biotechnology*, *29*(1), 24–26.

Rodriguez, J., Menet, J. S., & Rosbash, M. (2012). Nascent-seq indicates widespread cotranscriptional RNA editing in Drosophila. *Molecular Cell*, *47*(1), 27–37.

St Laurent, G., Tackett, M. R., Nechkin, S., Shtokalo, D., Antonets, D., Savva, Y. A., et al. (2013). Genome-wide analysis of A-to-I RNA editing by single-molecule sequencing in Drosophila. *Nature Structural & Molecular Biology*, *20*(11), 1333–1339.

Stoleru, D., Peng, Y., Agosto, J., & Rosbash, M. (2004). Coupled oscillators control morning and evening locomotor behaviour of Drosophila. *Nature*, *431*(7010), 862–868.

Sugino, K., Hempel, C. M., Miller, M. N., Hattox, A. M., Shapiro, P., Wu, C., et al. (2006). Molecular taxonomy of major neuronal classes in the adult mouse forebrain. *Nature Neuroscience*, *9*(1), 99–107.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods*, *6*(5), 377–382.

Thorvaldsdottir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, *14*(2), 178–192.

Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics*, *25*(9), 1105–1111.

Yoshii, T., Rieger, D., & Helfrich-Forster, C. (2012). Two clocks in the brain: An update of the morning and evening oscillator model in Drosophila. *Progress in Brain Research*, *199*, 59–82.

Zhang, L., Chung, B. Y., Lear, B. C., Kilman, V. L., Liu, Y., Mahesh, G., et al. (2010). DN1 (p) circadian neurons coordinate acute light and PDF inputs to produce robust daily behavior in Drosophila. *Current Biology*, *20*(7), 591–599.

Zhang, Y., Liu, Y., Bilodeau-Wentworth, D., Hardin, P. E., & Emery, P. (2010). Light and temperature control the contribution of specific DN1 neurons to Drosophila circadian behavior. *Current Biology*, *20*(7), 600–605.