Extensive Oscillatory Gene Expression During *C. elegans* Larval Development

Inauguraldissertation
zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von

Gerardus Johannes Hendriks

von Leiderdorp, die Niederlande

Basel, 2017

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von

Prof. Dr. Mihaela Zavolan

Dr. Helge Großhans

Prof. Dr. René Ketting

Basel, 8. Dezember 2015

Prof. Dr. J. Schibler

Acknowledgements

First of all I would like to thank my thesis advisor Dr. Helge Grosshans. I am very grateful for your tireless support and advice, especially when the project moved into unfamiliar territory. Thank you for letting me pursue the science and for the inspiring discussions about the project and latest results. Your optimism and support helped me to see the good side in any result, even if they were less than stellar.

I would also like to express my gratitude to Prof. Dr. Mihaela Zavolan and Prof. Dr. René Ketting for acting as my faculty representative and co-examiner respectively. I value our discussion and appreciate all the feedback that you gave on the project. Thank you for investing your valuable time.

I would like to take this opportunity to thank our collaborators from Elanco Animal Health, Dr. Jacques Bouvier and Dr. Lucien Rufener, for our exciting collaboration.

Of course I could not have gotten to this point without the help of Dimos. Thank you for all the work and especially for helping me analyze some of the data by myself. I have really enjoyed learning from you and working with you in these years. I am also grateful to the other people from the Grosshans lab with whom I have worked on this project in the last years; Florian, Yannick, Nicole, Giovanna, Sarah, and Milou. It has been a great working with you all and I am sure we all have much more good (oscillating) science ahead of us! Many thanks also go to Florian for making the days and weeks that we spent behind the microscopes a lot less boring than it perhaps should have been.

I am especially indebted to Monika for teaching me everything worm-related. She really took care of me in the lab and was always ready to give advice and help out. I will always be grateful for this.

Many thanks go out to all Grosshans lab members, past and present, for the great atmosphere in the lab. It has truly been a pleasure to work (and sometimes not-work) with all of you.

Special thanks to all the great facilities at the FMI that have been instrumental to this work. I am especially grateful to Iskra from the *C. elegans* facility for helpful and inspiring discussion on different projects and topics. I am very grateful to the members of the functional genomics facility; Kirsten, Sophie, Birgitte, Stephane

and Tim for help, advice and tips with preparing sequencing libraries. I also want to thank everyone at the media kitchen for the amazing service and taking all this work out of my hands.

Prof. Dr. Nancy Bonini and Dr. Nan Liu as well as Dr. Rute Marques, who I have had the pleasure of working with before I came to Switzerland, are also owed a great deal of gratitude. Thank you for all that you have taught me about top-quality science and for your words of motivation and inspiration.

I would like to especially thank my friends and family who have been a great support in both the good times and the not-so-good times. Thank you for your Belgian beer supply runs and for crashing my place. It was nice to have you all here! I hope to see a little more of you all in the next years!

Finally, I want to take the opportunity to thank my favorite person in the world; Ana. Thank you for bearing with me, and the 1391 km of distance between us. I am pleased (and I admit, quite impressed,) to see that after years of listening to me go on and on about nematodes, you start to become a *C. elegans* specialist. I could not have done this without you!

1. Introduction

1.1 Oscillations

Oscillations are periodic fluctuations away from a state of equilibrium. Oscillatory systems are widely studied in mathematics, physics, chemistry and, more recently, biology. Three main variables characterize an oscillation. The amplitude describes the extent of fluctuation from the baseline or equilibrium. The period describes the duration of one full oscillation. Finally, the phase describes the relative displacement of an oscillation. Harmonic oscillations, which are built up of sine and cosine waves, are the most common type of oscillations. Interactions with oscillating systems are common in Biology. Visual cues, which are encoded in spatial oscillations of photons, are detected by neurons in the retina. The inner ear on the other hand detects oscillations of pressure in air that we refer to as sound. Modifications to the basic characteristics of these oscillations affect the way we perceive them. Changing of the amplitude of sound waves for example increases the volume of the sound we hear. Changes in the period on the other hand affect the pitch of the sound. For the oscillations of photons, a change in period affects the color of the light that we perceive. Mammals and many other organisms are equipped to sense oscillations of various origins. Additionally, many organisms are themselves able to predict rhythmic changes in their environment through the establishment of an internal oscillatory system. Circadian rhythm is an important example of this interaction with an external cue that sets the phase of oscillatory gene expression and behavior. Another example of an oscillatory system in biology is found in the formation of the somites during embryogenesis of vertebrates. In this case, the size and shape of newly formed somite is dictated by an oscillation that occurs specifically during development. Although there are numerous more examples of oscillating systems, we focus on these two widely studied oscillatory systems below.

1.1.1 Circadian rhythm

Circadian rhythms allow organisms to adjust their behavior to the time of the day. The systems that establish this rhythm are the most well studied oscillating systems and together represent the most well-known example of a biological clock. The word circadian stems from the Latin words, '*circa'* and '*dies'*, which mean 'approximately' and 'day', respectively. The term circadian rhythm is commonly used to describe a rhythm that is induced by a timer that has a period of approximately 24 hours. Biological timers are a common feature in biology, and especially circadian timers are plentiful. Circadian timing of behavior was first reported in the 18th century based on the observation that the leaves of the plant *Mimosa pudica* fold and spread at daily intervals. Through these types of behavioral modifications, circadian rhythms offer the organism an ability to cope with the day-night cycle, as well as the changing of the seasons (Stoleru et al., 2007). Although circadian rhythms and the core clock components that establish them may differ between species, there are some core concepts that many of them have in common. First of all, the rhythm is temperature compensated. This means that regardless of the temperature, the period of the rhythm or oscillation will remain unchanged (Pittendrigh, 1954). Secondly, there is an environmental stimulus (also called zeitgeber) that serves to impose and enforce the phase of the rhythm. Light, temperature and food availability are examples of external cues that can function as zeitgebers (Barrett and Takahashi, 1995; Krieger, 1974; Pittendrigh, 1981). Extended signaling of the zeitgeber on the internal circadian clock results in entrainment of the rhythm. The period is adjusted and the phase of the oscillations is determined by the zeitgeber. As a result of this entrainment, the phase of the rhythm is specified and maintained even in the absence of the zeitgeber signal. The most well-known and recognizable example of this is the jet lag that occurs when traveling between time zones. This stems from a mismatch between the entrained circadian rhythm and the signals received from the zeitgeber and is overcome by the

resetting of the clock. Circadian clocks have been found in all kingdoms of life. On the molecular level however, there is a strong divergence between the clocks in different kingdoms. This suggests that the mechanisms that link behavior and gene expression to time may have evolved independently (Young and Kay, 2001; Stanewsky et al., 1998). In humans, the disruption of the circadian rhythm can, among others, affect cognitive function (reviewed in Yaffe et al., 2014), inflammatory disease (reviewed in Lebailly et al., 2015), and even susceptibility to cancer (reviewed in Sahar and Sassone-Corsi, 2009). Early experiments into the mechanism behind circadian regulation took advantage of the eclosion rhythm of newly hatching *Drosophila melanogaster*. These fruit flies hatch from their pupae in the early morning. This specific timing suggested that a mechanism that keeps circadian time might exist. These studies eventually led to the discovery of the first gene involved in circadian rhythm; *period* (*per*) (Reddy et al., 1984). In addition to these mechanistic insights, studying of circadian behavior has led to insights into the characteristics of circadian rhythms in higher eukaryotes. The mechanisms that drive oscillatory gene expression according to a circadian rhythm are extremely well conserved from fly to human (reviewed in Panda et al., 2002). Although there are some reports of circadian rhythms in the nematode *Caenorhabditis elegans*, it has been the topic of debate. Some, low amplitude, not fully entrainable oscillations have been described, but the regulatory mechanisms of this system are not fully understood. For that reason, the focus below will be on the *Drosophila* and human circadian networks. The special case of *C. elegans* circadian rhythm is described in more detail in the last paragraph.

Circadian rhythm in Drosophila melanogaster

After the discovery of Period as the first core clock component in *Drosophila*, the fruit fly became one of the most commonly used model-organisms to study circadian rhythms. The three decades following the discovery of Per saw the discovery of many additional factors that seemed to be involved in the regulation of circadian rhythm in *Drosophila*.

One gene encoding such a factor was *timeless* (*tim*) (Sehgal et al., 1994). It is aptly named, since flies that carry mutations in this gene show strong arrhythmia of circadian behavior. Interestingly, as was the case for *per*, mutations in *tim* resulted in varying effects on circadian behavior and locomotor activity rhythms. Although there are not always one-to-one homologues, the core mechanism that both Per and Tim proteins are a part of is conserved from fly to man (reviewed in Panda et al., 2002). This core mechanism consists mainly of a transcriptional feedback model (illustrated in figure 1.1). The two transcriptional regulators; Per and Tim do not have DNA-binding capacity but, instead, interact with the DNA binding proteins that are encoded by *clock* (*dClk*) and *cycle* (*Cyc*). These DNA-binding proteins are transcriptional activators that reside in the nucleus. They bind to a specific DNA element (CACGTG) that is known as the E-box and is found in the promoters of *per* and *tim* themselves (Lee et al., 1999). The levels of Tim and Per are regulated directly and indirectly by light. Tim is bound by the cryptochrome protein (dCRY) when it is activated by light in the early morning and interacts with the protein Jetlag to initiate degradation of Tim by the proteasome (Lin et al., 2001; Koh et al., 2006). Per on the other hand is stabilized by binding Tim, while non-dimerized Per will be phosphorylated, ubiquitinated and degraded by double-time (DBT) (Kloss et al., 1998). When the levels of Per and Tim in the cytoplasm reach high enough levels, the individual proteins, and Per/Tim complexes translocate to the nucleus. The phosphorylation of Per that is performed by casein kinase 2 (Ck2) promotes the transition of the Per/Tim complex to the nucleus (Lin et al., 2002). In addition to Per and Tim, Dbt also translocates to the nucleus. In the nucleus, the Per/Tim dimers bind to dClk/Cyc dimers, and hyper-phosphorylate the complex (Yu et al., 2006). The hyper-phosphorylated complex does not bind DNA and transcription of its targets, including *per* and *tim*, is reduced. The light sensitivity of Tim results in a rapid drop in levels of Tim in the early day. This is followed by phosphorylation of Per by Dbt and the subsequent degradation of Per. Once the Per/

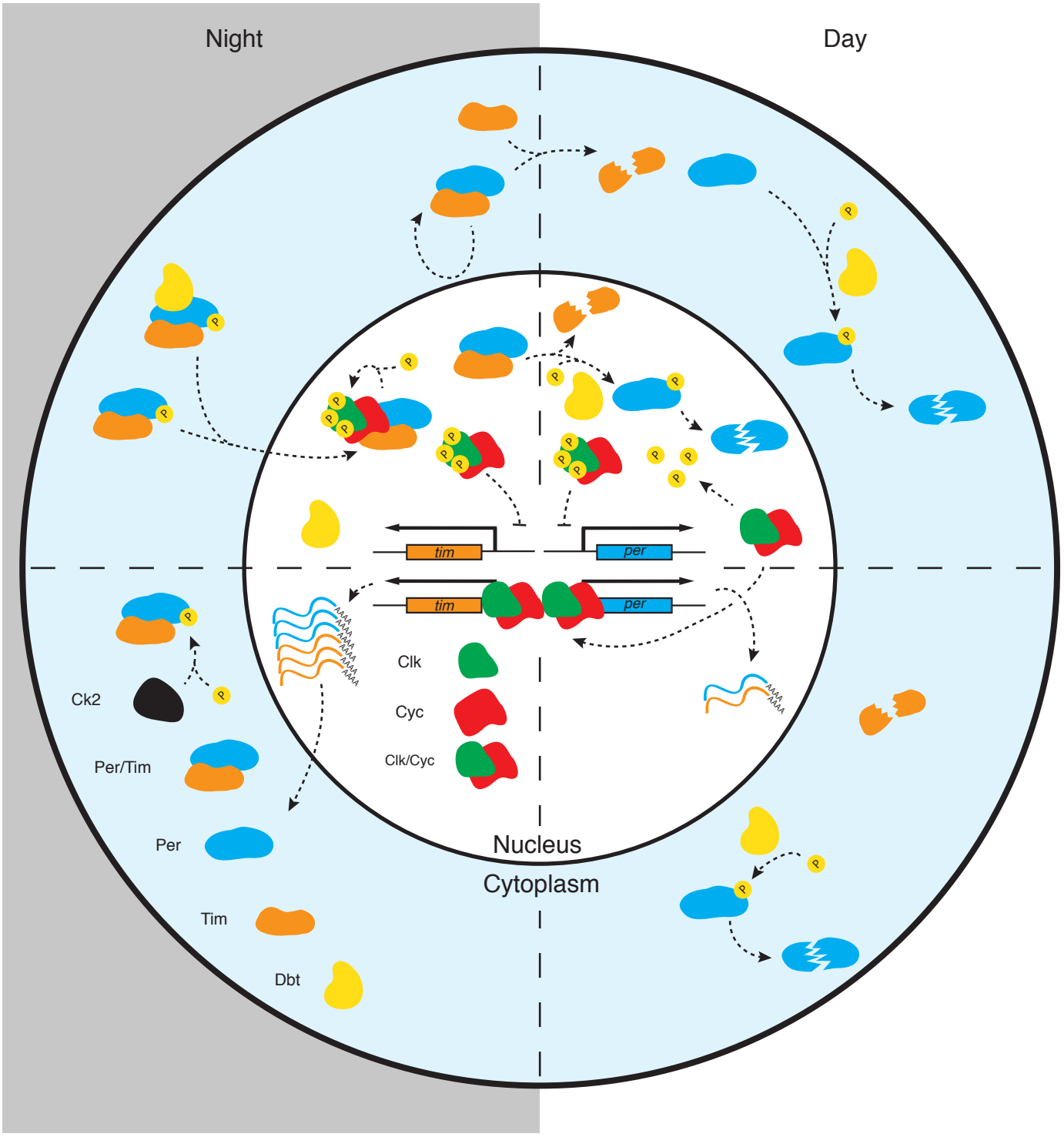


Fig 1.1 THE *DROSOPHILA* CIRCADIAN CLOCK

The circadian clock in *Drosophila* is based on the cycling transcription of *tim* and *per*. Transcription of *tim* and *per* at the end of the day results in rising levels of Per and Tim in the cytoplasm. Ck2 phosphorylates Per. Phosphorylated Per can then, in complex with Tim and Dbt, translocate into the nucleus where it hyperphosphorylates and inactivates Clk/Cyc dimers. In the morning, Tim is degraded in a light-dependent manner and since Tim no longer stabilizes Per, Dbt can now phosphorylate Per, which results in degradation of Per by the proteasome. This process happens both in the nucleus and in the cytoplasm. The hyperphosphorylation of Clk/Cyc is lost and at the end of the day it again activates transcription of *tim* and *per* and thereby restarts the cycle. Figure was inspired by Peschel and Helfrich-Förster, 2011.

Tim dimer levels have dropped sufficiently, the repression is relieved and *dClk*/*Cyc* induce transcription of *per* and *tim* and allow the cycle to start from the beginning. Additional feedback loops can be linked to the core clock. An example of this is the transcriptional activation of genes encoding for Vrille (Vri) and PAR domain protein 1 (Pdp1ε) by the dClk/Cyc dimer (Blau and Young, 1999; Cyran et al., 2003). Vri and Pdp1ε then respectively inhibit and induce transcription of *dClk*. Despite the direct feedback of this system back into the main clock, stable overexpression or inhibition of Pdp1ε cannot stabilize oscillations of *dClk* (Benito et al., 2007). It can however affect circadian behavior, suggesting that it may function downstream of the core clock.

The core clock that we described above is also functional in a range of tissues such as the brain, thorax and abdomen. This has led to the division of the clock into the central- and the peripheral oscillators. As this naming suggests, the central oscillator, in the nervous system of the fly, may regulate the peripheral oscillators, as is the case in mammalian systems (see below). However, rhythms in the periphery can be established and maintained independently of the central clock (Plautz et al., 1997). These peripheral clocks are also referred to as distributed clocks.

As mentioned previously, the oscillating core clock gives rise to rhythmic behavior. On the molecular level, the core clock drives oscillatory expression of large numbers of genes. With the help of genome-wide transcriptome analysis, it has been estimated that more than 2000 of the transcripts that are expressed in the circadian neurons of *Drosophila* show oscillatory expression (Kula-Eversole et al., 2010). The core clock machinery, as described above, is driven to oscillate by a transcriptional regulatory mechanism. To investigate whether this could be the case for all oscillating transcripts, the Rosbash lab investigated pre-mRNA levels and mRNA levels in dissected heads. Since pre-mRNA is considered to have a very short half-life it can be used as a proxy for transcription

(Gaidatzis et al., 2015). Interestingly, many oscillating mature mRNAs do not show pre-mRNA oscillations and vice versa (Rodriguez et al., 2013). This suggests that many of the oscillating mRNAs are oscillating due to posttranscriptional regulation.

Circadian rhythm in mammalian systems

Similarly to *Drosophila*, the mammalian clocks revolve around a simple negative feedback loop. The names in the mammalian systems differ in some cases, but the conservation, especially on the functional level, is extensive. The human proteins BMAL1 and Cryptochrome replace CYC and TIM respectively. In the second feedback loop system, REV-ERBα and RORA act to functionally replace VRI and PDP1ε respectively (reviewed in Panda et al., 2002).

The incorporation of circadian rhythms into large organisms, like mammals, requires a complex network of signal transduction throughout the body. Peripheral oscillations cannot be directly entrained and are regulated through the suprachiasmatic nucleus (SCN). Ablation of this nucleus that is located in the hypothalamus disrupts the circadian behavior of the organism and results in desynchronization of the clock in peripheral tissues (Moore and Eichler, 1972; Yoo et al., 2004). The retina signals to the SCN in response to light stimulation and thereby provides the link between the zeitgeber (light) and the central oscillator. Other zeitgebers, such as food availability, do not require the SCN for entrainment of circadian rhythm (Krieger et al., 1977). The SCN then signals to the peripheral oscillators and thereby establishes and maintains the circadian rhythm in peripheral tissues. Interestingly, the core clock machinery is expressed both in the SCN and peripheral tissues. However, peripheral tissues cannot be entrained in the absence of signal from the SCN. This suggests that there is a critical and currently poorly understood difference between the clock in the SCN and the clock in the periphery.

Circadian rhythm in C. elegans

Circadian rhythms in *C. elegans* have been studied relatively poorly compared to some of the other popular model organisms. Despite this, there have been multiple studies confirming the presence of circadian behavior. Locomoter behavior oscillates with a circadian period in adults (Saigusa et al., 2002), as does the resistance of arrested L1 larvae to environmental stresses (Kippert et al., 2002). Although these studies suggest the presence of a circadian oscillator, the system seems to be highly variable (Simonetta et al., 2009). In addition to these behavioral rhythms, oscillating transcription has been detected for numerous genes in adult worms (van der Linden et al., 2010). Unlike traditional circadian system however, two different zeitgebers (light and temperature) each regulate independent groups of oscillating transcripts. Most of the genes for which oscillatory expression was driven by either stimulus do not show oscillatory expression in free-running conditions.

Some of the core clock components that were described and illustrated in figure 1.1 are well conserved in *C. elegans*. Many of these, however, seem to play a role in development (Jeon et al., 1999; Banerjee et al., 2005). What other functions they may have in the establishment or maintenance of an oscillating system remains unknown.

1.1.2 Somitogenesis

The rhythmic system that we discussed above is the most well studied example of a biological clock that is synchronized by outside stimuli. Another biological clock, that is well studied, but not regulated by external signaling, is the segmentation clock that orchestrates somitogenesis. This system is dependent on an internal timer or clock and does not require any external signaling. It is a key example of a developmental timer.

Somitogenesis occurs during embryogenesis and results in the segmentation of the newly developing body and the formation of so-called somites. These embryonic structures

eventually give rise to the vertebrae and ribs, the skeletal muscle and other tissues. Vertebrate tissue formation and differentiation occurs from the anterior to the posterior of the developing organism. Accordingly, somites are formed from the anterior to the posterior. These cubic structures are formed by groups of cells that bud off on the anterior side of the presomitic mesoderm (PSM). Cells that migrate into the PSM from the tail bud region replenish the posterior PSM. Somite segmentation occurs at defined intervals, with the length and number of repetitions depending on the species. The length of these intervals ranges from 8 hours in humans, to 30 minutes in zebrafish. The biological mechanism of segmentation is well studied and is known to require two regulatory systems that are brought together in a single model; the “clock and wavefront” model. One the one hand, an oscillator is present in the posterior PSM, which drives oscillations that travel to the anterior along the anterior-posterior axis. On the other side, a determination wavefront travels to the posterior from the anterior along the anterior-posterior axis (Cooke and Zeeman, 1976). These two systems establish the timely expression of the master-regulator of somitogenesis; mesoderm posterior 2 (Mesp2) (Saga et al., 1997). Somitogenesis is mainly studied in mouse, chicken and zebrafish. The models and experiments described below focus on segmentation during mouse development.

Determination wavefront

The “wavefront” part of the “clock and wavefront” model consists of two opposing gradients over the anterior-posterior axis of the PSM. From the posterior, FGF8, FGF4 and beta catenin form a gradient towards the anterior PSM (Aulehla et al., 2008; Dubrulle et al., 2001; Naiche et al., 2011). Reversely, from the anterior, retinoic acid (RA) forms a gradient towards the posterior. During development, the front where the FGF proteins, beta-catenin and RA are all present shifts to the posterior. The secreted FGFs are only expressed in the tail bud region. The continuous migration of cells from the tail bud region

to the PSM, combined with the slow degradation of Fgf8 mRNA results in a FGF8 protein gradient (Dubrulle and Pourquié, 2004). The posterior-anterior gradient of WNT3A, and its downstream signaling factor beta catenin are thought to be established in the same way (Aulehla et al., 2003). Both the WNT and FGF signaling pathways regulate the expression of different genes and proteins that are also present in a posterior-anterior gradient throughout the PSM. FGF and WNT signaling block differentiation and maintain the pluripotent state of the posterior PSM (Aulehla et al., 2008; Dubrulle et al., 2001). A gradient of the morphogen RA is present in the anterior to posterior orientation and acts to decrease FGF levels and counter FGF signaling (Sirbu and Duester, 2006). This gradient is established by diffusion of RA from cells in the anterior PSM. The high levels of RA in these cells are regulated by the lack of RA metabolizing enzyme in these cells (Sakai et al., 2001). In mutants, the loss of RA signaling enforces the FGF8 gradient, resulting in the formation of smaller and non-synchronous pairs of somites (Vermot et al., 2005).

Segmentation clock

While the determination wavefront slowly progresses over the anterior-posterior axis of the PSM, a system of spatiotemporal oscillations contributes to the somite boundary formation. These oscillations nucleate in the posterior PSM and travel to the anterior. The first mRNA that was shown to oscillate is the chick homologue of *Drosophila* basic helix-loop-helix transcription factor *hairy* (*chairy-1*) (Palmeirim et al., 1997). The authors showed that one period of the oscillation corresponds to the formation of one pair of somites. Homologues of this group of transcription factors were soon found to show oscillating expression over the PSM in mouse as well as zebrafish. In mice, the genes encoding these transcription factors are: *Hes1, Hes5, Hes7* and *Hey2*. Another transcript that shows oscillations over the PSM anterior-posterior axis is *Lunatic fringe* (*Lfng*), which encodes a protein that affects notch signaling (Forsberg et al., 1998; Panin et al., 1997).

Perhaps not surprisingly, the oscillations of the segmentation clock, like oscillations in circadian rhythm are, at least in part, driven by a simple negative feedback loop. One of the key cycling genes that is widely studied in the mouse is *Hes7*. The promoter of *Hes7* is transcriptionally activated by Notch signaling and inhibited by HES7. NOTCH signaling is also rhythmic due to a negative feedback loop. The finding that NOTCH signaling induces expression of *Lfng* has led to the identification of a *cis*-regulatory element that induce transcription in response to NOTCH signaling and inhibit transcription in response to HES7 (Cole et al., 2002; Morales et al., 2002; Bessho et al., 2003). In these regulatory networks, the timing of induction or inhibition of transcription is essential. The system that was described, in combination with a transcriptional delay, can induce oscillatory rhythms (Lewis, 2003). Mice that constitutively express core segmentation clock components, such as *Lfng*, in the PSM have severe somite development phenotypes. This shows that not just the expression of these factors, but the oscillating expression is important for segmentation (Serth et al., 2003).

1.2 *C. elegans* as a model organism in developmental biology

The field of developmental biology focuses on the regulation of cellular proliferation, differentiation and the fate determination decisions that are made along the way. To study this, we take advantage of the nematode *Caenorhabditis elegans*.

1.2.1 Physiology

The adult roundworm *C. elegans* has only 959 somatic cells. These cells make up all different tissues ranging from the hypodermis (skin) to the nervous system. Additionally, the worm is transparent, which allows for *in vivo* microscopy. Nematodes are surrounded by a cuticle that serves as an exoskeleton and gives the worm its structure and shape. It is made up of many different collagens that are secreted by the cells of the hypodermis that are situated directly beneath the cuticle. In addition to the longitudinal syncytium

formed by the seam cells, the hypodermis consists of 13 different cells, named Hyp 1 – 13. Nine of these cells are also syncytia, the largest of which, Hyp 7, contains 139 nuclei and makes up the bulk of the hypodermis (Shemer and Podbilewicz, 2000). *C. elegans* has a network of 302 neurons as well as musculature, a pharynx and a gut. In terms of volume however, the reproductive system takes up most space in the adult. Since a single worm hermaphrodite lays approximately 200 eggs during its lifetime, the large size of the germline may not come as a surprise. In a population of *C. elegans*, most worms are hermaphrodites. The progeny is therefore generally the product of self-fertilization. Males occur in the population at an incidence of approximately 0.1% and are important for the genetic diversity of the species. For scientists, the presence of males is essential for genetic studies. Worms can survive extreme conditions, but thrive in temperatures between 15°C and 25°C.

1.2.2 Larval Development

C. elegans is a very popular model organism to study development, in part because of its invariant development. This means that all cell proliferation, differentiation and migration events occur identically among different animals. The speed of worm development increases with the environmental temperature. The development times (in hours) that are mentioned below describe development at 25°C (Altun et al.). After the egg is laid a larva hatches from it after approximately 8-9 hours. Larvae in this first stage of larval development (L1) will detect the presence or absence of a food source of bacteria. If no bacteria are present the larvae arrest and survive in this state of diapause for up to 10 days. If a food source is present however, development starts and after approximately 9 hours of development the worms reach larval stage 2 (L2). The transition between larval stages is characterized by a molt. During this molt the worm synthesizes a new cuticle and sheds the old one. The ecdysis from the old cuticle marks the beginning of the next larval stage. The larva develops through an additional 3 stages that each last

approximately 8 hours before it reaches the adult stage. After an additional 8 hours of adulthood the worm starts to lay eggs and the life cycle is completed. In addition to this continuous developmental pathway, worms can enter another arrested state that is referred to as dauer. Any stressful or unfavorable conditions before the L1-L2 molt can divert development from regular L2 to the dauer stage. Instead of the L1-L2 molt, worms undergo the L1-L2d molt before they form the dauer animal. While in dauer diapause, worms can survive extreme and unfavorable conditions for many months.

The correct timing and coordination of cell divisions is crucial for successful development of any multicellular organism. Despite the importance of this process, the timing of cell divisions during development is not well understood in higher eukaryotes, and is little informative in single cell organisms. *C. elegans* has been widely used as a model to study embryonic development. A series of asymmetric divisions in the embryo give rise to the founder cells of the different cell lineages. When the egg eventually hatches, the L1 larva that emerges has 558 cells (Sulston et al., 1983). During larval development this number increases to the final 959 cells in the adult (Sulston and Horvitz, 1977). The relatively small number of cells as well as the fact that the cell divisions are readily visible through the transparent cuticle have allowed for the full tracing of the worm cell lineage, from the fertilized egg cell, to the adult animal. Postembryonic development starts with the hatching of the L1 from the egg. Of the 558 cells that an L1 larva emerges with, only 51 are dividing. Over the 4 larval stages of worm development, these 51 cells divide and ultimately add a total of 401 cells to form the adult worm. Despite this relatively small number of dividing cells, the worm undergoes dramatic changes of the hypodermis, muscle and nervous system. Development of *C. elegans* is inherently well regulated and orchestrated. One example of developmental timing is found in the heterochronic pathway. Heterochronic genes make up a system of developmental timing and cell fate specification of the hypodermis. The word heterochrony stems from the Greek words

‘*heteros’*, meaning other, and ‘*chronos’*, which means time. Extensive screening for genes that, when mutated, cause or relieve developmental lethality and defects of developmental timing has led to the identification of a large number of genes that are involved in the heterochronic pathway. The identification of these mutants has allowed for the mechanistic dissection of this pathway that keeps developmental time and outlines the adoption of cell-fates in the hypodermis. Importantly, this system must also time cell divisions in relation to molting, since the failed synchronization of developmental timing with the molts is lethal (Ruaud and Bessereau, 2006).

The cell-fate decisions that occur during larval development of hermaphrodite nematodes are widely studied. The seam cells, a group of 20 stem cell-like cells in the hypodermis of the newly hatched L1 larvae, are of particular interest. These cells are distributed on two sides over the length of the worm. Although their exact lineages differ slightly between them, they generally divide once per larval stage. After each seam cell division the posterior daughters retain the seam cell fate, while the anterior daughters fuse to the hypodermal syncytium (Hyp 7). Before this fusion takes place however, the anterior daughters undergo endoreduplication. During this process the cells undergo S-phase, but do not divide. Because of this, the cells that fuse to Hyp 7 are tetraploid. In addition to the four asymmetrical cell divisions, 10 of the seam cells also undergo a symmetrical cell division during early L2. This increases the number of seam cells to two pairs of 16 seam cells. The regulation of the seam cell divisions is highly regulated and many different heterochronic mutants show abolished or additional symmetrical cell divisions (Ambros and Horvitz, 1984).

1.2.3 Molting

The cuticle provides the essential rigidity and flexibility to the worm. The body muscles of the worm are anchored to it, allowing the animal to move. In addition to this, it provides

physical protection from the environment, and in the case of parasitic nematodes (such as *H. contortus*, described below) provides mechanisms to evade and cope with the immune response of the host. The cuticle can however only accommodate limited growth of the worm and eventually impedes growth during larval development. Therefore, the cuticle is shed at the end of every larval stage. Over development, the cuticle is therefore synthesized five times and shed four times. This process of synthesis and shedding of the cuticle is altogether referred to as molting.

The process starts with a slow reduction of activity and feeding. The worm eventually stops moving altogether. This state is referred to as lethargus. The subsequent synthesis of the new cuticle, and release and loosening of cuticle from the hypodermis, also known as apolysis, both occur during lethargus. Eventually, the period of lethargus ends and the worm escapes from the cuticle. This process is also referred to as ecdysis.

Cuticle synthesis

The cuticle is built up mostly of cross-linked collagens and a number of accessory proteins and lipids. The *C. elegans* genome encodes 167 cuticular collagens. Only 22 of these collagens give strong phenotypes when mutated, suggesting that the others are either partially redundant or that the phenotypes of single collagen mutants are very minor and have not been detected (reviewed in Page and Johnstone, 2007). The biosynthesis of the collagen matrix that forms the cuticle is a complex process that requires a number of different enzymes (reviewed in Prockop and Kivirikko, 1995). A characteristic of collagen proteins is the repetition of a Glycine-X-Y tripeptide sequence. The X- and Y- positions here are often proline and hydroxyproline respectively. The hydroxylation of the proline on the Y-position takes place in the endoplasmic reticulum (ER) and is catalyzed by the multiprotein complex; collagen prolyl 4-hydroxylase (C-P4H). This hydroxylation of the Y-position proline occurs co-translationally. An essential subunit of the C-P4H complex is

encoded by *dpy-18* (Winter and Page, 2000). After the individual polypeptide has been synthesized, disulphide bonds are formed between the conserved C-terminal cysteine residues by protein disulphide isomerases (PDIs, PDI-2 in *C. elegans*) (Winter et al., 2007b). In the next step of collagen biosynthesis, associated collagen proteins will undergo extensive folding resulting in trimerization and formation of procollagen. This trimerization reaction is dependent on peptidyl prolyl *cis-trans* isomerase (PPIase) (Bächinger, 1987). In *C. elegans* it is unclear exactly which proteins are required, however, FKB-4 and FKB-5 have been shown to be required for procollagen trimerisation in cold-stress conditions (Winter et al., 2007a). After the trimerization, the proteins are exported from the ER to the extracellular space. When the procollagen is in the extracellular space it is further processed by two cleavages that occur on the N-terminal and C-terminal side of the protein trimer. The enzymes that are involved in these cleavage reactions are BLI-4 (Thacker et al., 2006) and DPY-31 (Novelli et al., 2004) respectively. Finally, the collagen trimers will be cross-linked by the NADPH dual oxidase enzyme, BLI-3 (Edens et al., 2001) and its cofactor MLT-7 (Thein et al., 2009).

Apolysis & Ecdysis

To detach the cuticle from the hypodermis, the worm produces different proteases to partially degrade the cuticle and allow for the worm to molt. Specifically, NAS-36 and -37 are two proteases that have been implicated in this process (Suzuki et al., 2004). The loss of the genes encoding these two metallopeptidases results in molting phenotypes. Complementation of *nas-36* and *nas-37* expression specifically in the hypodermis rescued the molting phenotypes. Similarly, the loss of *acn-1*, which is a metallopeptidase that is, surprisingly, lacking an active site, shows clear molting phenotypes, as well as abnormal seam and vulva development (Brooks et al., 2003). In addition, a large number of cysteine – and serine proteases is encoded in the *C. elegans* genome, some of which have been implicated in the molting process. When the new cuticle is synthesized and the

worm has partially detached itself from the cuticle, ecdysis occurs. During ecdysis, the worm performs a number of rapid, rotational movements to physically separate itself from the surrounding old cuticle. Once the cuticle is sufficiently loosened, the worm simply crawls out of it.

Regulation of molting

The formation of the *C. elegans* cuticle and the degradation of the old cuticle are two highly complex processes that occur multiple times during nematode development. A number of regulatory pathways are known to affect the expression of collagens and proteases and thereby regulate worm development. Two important transcription factors that have been implicated in the regulation of molting are the nuclear hormone receptors (NHR), NHR-23 and NHR-25 (Gissendanner and Sluder, 2000; Kostrouchova et al., 1998); Kostrouchova et al., 2001). These two NHRs are the orthologues of *Drosophila* DHR-3 and βFTZ-F1, both of which are involved in the ecdysone response in the fly (reviewed in (Thummel, 2001)). Worms that lack these genes show molting phenotypes as well as abnormal hypodermal development. Additionally, an RNAi screen identified 159 genes, including 7 putative transcriptional regulators, that are involved in molting (Frand et al., 2005).

In addition to the transcription factors and other proteins that are implicated in molting, cholesterol has been implicated as an important factor for molting in *C. elegans*. In the absence of cholesterol the worms arrest and show molting phenotype (Merris et al., 2003; Yochem et al., 1999). Since *C. elegans* requires only very small amounts of cholesterol for proper development (Merris et al., 2003), it has been suggested to act as a precursor for the synthesis of sterol-based steroid hormones (Matyash et al., 2004). Additionally, a mutant of a sterol-modifying enzyme, LET-767, shows increased dependence on cholesterol (Kuervers et al., 2003). Although it has been shown that steroidal ligands can

bind *C. elegans* NHRs, few ligand-NHR interactions have been identified (Motola et al., 2006).

1.2.4 Evolution and conservation of nematode species

The phylum Nematoda contains 25,043 species and represents a very diverse group of species (Zhang, 2013). As a reference, there are approximately 16,000 reported mammalian species (Zhang, 2013). The species that is closest to *C. elegans* on an evolutionary scale is the relatively distant; *C. briggsae*. These two nematodes are estimated to share their closest common ancestor approximately 30 million years ago (Cutter, 2008). Physiologically, *C. briggsae* and *C. elegans* are almost impossible to distinguish. As may be expected, coding regions of essential genes are relatively well conserved. The conservation of non-coding regions, such as promoters and 3'UTRs, however, is poor. Other commonly studied Caenorhabditis species include; *C. remanei*, *C. brenneri* and *C. japonica*, which are all dioecious (i.e. non-hermaphroditic) species.

Other members of the Nematoda phylum that are the subject of study are the parasitic nematodes. These worms infect a wide variety of hosts, ranging from insects to human. Of particular interest to researchers in the veterinary sciences is the ruminant pathogenic parasitic species *Haemonchus contortus*. This gastrointestinal nematode infects, among others, grazing sheep and goats. Its life cycle consists of the free-living larval stages L1-L3 and the parasitic L4 and adult stages. Unless they are ingested, L3 larvae arrest and can remain arrested under harsh conditions in the field. When the L3 is ingested it develops to form the L4 and finally the adult. In these stages, *H. contortus* attaches to the wall of the abomasum and feeds on blood, causing parasitic gastroenteritis. This is also where the worms mate and the female produces and lays eggs. These eggs are shed from the host with the feces and when the L1 larvae hatch in the field the cycle begins anew. Since *H. contortus* is drawing blood from its host, the host organism can show

symptoms that are related to blood loss. The infection results in reduction in the production of milk and wool, a reduction in animal growth and reduced fertility (Parkins and Holmes, 1989). In severe cases the infection may result in the death of the host due to the high number of parasites that is feeding in the abomasum. Because the hosts of these parasites are commonly kept as livestock for the production of meat, milk and wool, these symptoms cause a major financial burden on the industry. Treating infected animals with anthelminthic drugs helps to reduce the parasite burden. Antihelminthic drugs can be toxic to the host animal and resistance to the different drugs is common (discussed in Besier et al., 2016). Drugs targeting molting and cuticle development are of great interest because of the absence of these pathways in the host organism and the therefore relatively low chance of toxicity. Additionally, the important role of cuticle development in physiology may also provide some barrier for the development of resistance. Unfortunately, anti-helminthic drug resistance is rising and control of helminthic infections is proving challenging (Besier et al., 2016).

1.3 The regulation of gene expression

During the development of a multicellular organism, a single omnipotent cell gives rise to all the cells in the organism. While this cell and its daughter cells share the same genomic DNA, they have highly variable characteristics. Differential regulation of gene expression allows for the adoption of different cell fates and proliferative profiles. Steady state mRNA levels are a direct effect of the balance between mRNA transcription and mRNA decay. Posttranscriptional regulation of an mRNA affects the half-life of the transcript and how efficiently it is translated. Below we will discuss the process of transcription, transcriptional regulation and posttranscriptional regulation.

1.3.1 Transcription

In eukaryotes, three multi-subunit enzymes transcribe DNA to produce RNA; DNA-dependent RNA polymerase I, II and III (RNA Pol I, II and III respectively). Pol I is responsible for the transcription of ribosomal RNA (rRNA). Pol III transcribes a wide range of small non-coding RNAs including transfer RNA (tRNA). Pol II is the enzyme that is responsible for the transcription of messenger RNA (mRNA) from protein coding genes. Since mRNAs are the only species of RNA that are translated into protein, these are most widely studied and the following introduction will focus on transcription by RNA Pol II.

The formation of the pre-initiation complex (PIC) is the first step in transcription and is commonly referred to as transcription initiation. The PIC is formed on the DNA in the core promoter of a gene. The core promoter can be defined as “the minimal stretch of contiguous DNA sequence that is sufficient to direct accurate initiation of transcription by the RNA polymerase II machinery” (Butler and Kadonaga, 2002). General transcription factors, including RNA Pol II, are sequentially recruited to this region (reviewed in Sainsbury et al., 2015). The final transcription factor that is recruited phosphorylates pol II on serine 5, marking the initiation of transcription (Komarnitsky et al., 2000). Upon successful initiation, transcription starts and the transcription machinery moves away from the promoter region. The PIC then dissociates and gives way to the elongation complex. Some of the transcription factors that made up the PIC remain associated while other dissociate. Another set of transcription factors, including P-TEFb, can interact with the elongation complex. P-TEFb is composed of multiple proteins, including Cdk9 kinase which phosphorylates serine 2 in the CTD of pol II (Shim et al., 2002) and thereby marks Pol II as actively elongating (Komarnitsky et al., 2000).

1.3.2 Co-transcriptional and posttranscriptional processes.

In addition to transcription itself, there are a number of processes that occur co-transcriptionally and post-transcriptionally. As soon as the 5' end of the mRNA has been synthesized, it is capped to protect it from 5'-to-3' degradation and stimulate translation. The first nucleotide on the 5'-end of a transcript undergoes a modification that forms a protective cap structure. In addition to coding sequences (exons), many genes contain large stretches of non-coding sequence; introns. During pre-mRNA splicing, the non-coding sequences are removed and the resulting coding sequences are fused together. This co-transcriptional process that results in the production of mature mRNA is heavily regulated. Alternative splicing has the potential to exponentially increase the number of possible gene products. Although the genome of the nematode *C. elegans* encodes slightly more protein coding genes than the human genome, the additional complexity of human biology is, in part, made possible through the large number of gene isoforms. Alternative splicing occurs in the nucleus and gives rise to multiple different gene products that can be encoded by a single gene. After the transcribing PolII and associated factors have reached the end of the coding region and the 3' untranslated region (3'UTR), transcription is terminated through mechanisms that are not yet fully understood. When the primary transcript produced by Pol II has been cleaved, a stretch of non-templated adenosines is added to the 3' end. This is performed by a dedicated nucleotide transferase, poly-A polymerase (PAP). After capping, splicing, cleavage and poly-adenylation, the mRNA is considered mature and is ready to be exported from the nucleus and translated.

1.3.3 Regulation of transcription

The entire process of transcription as described above is dependent on the general transcription factors that make up the PIC and elongation complex. The studies that

identified the general transcription factors mostly took advantage of cell-free systems that consist of expressed or purified proteins or protein fractions and ‘naked’ DNA. However, in an *in vivo* situation, transcription initiation and elongation are complicated by the presence of DNA-binding proteins, most importantly, histones. The presence of histones and the compaction of the DNA provide an impediment that can be overcome in a targeted manner to allow gene specific regulation of transcription. We will describe the roles of chromatin regulation and transcription factors below. Although they are described separately, it is important to note that chromatin-based regulation and transcription factors interact with- and influence each other.

Chromatin

Initiation of transcription, as described above, often starts with the binding of TBP to the TATA-box, and the subsequent binding of the entire PIC in the promoter region. However, the PIC is a very large protein complex and DNA is usually wrapped around a complex of histone proteins, making the promoter region relatively inaccessible. One octameric histone complex that is associated with DNA is also known as a nucleosome. Highly compacted DNA stretches, called heterochromatin, are generally transcriptionally silent, while transcriptionally active DNA, or euchromatin, is less heavily compacted. The regulation of DNA accessibility intrinsically allows for gene-specific, or locus-specific regulation. For this reason it has been extensively studied in many biological contexts.

One nucleosome consists of 147 basepairs (bp) of DNA and a complex of eight histone proteins. The histone proteins, or variants, that make up the complex affect the chromatin structure and thereby the transcriptional efficiency of the locus (reviewed in Luger et al., 2012). The chromatin structures in the promoters of transcriptionally regulated genes can be highly dynamic. Chromatin remodeling complexes are able to remodel nucleosome positioning and actively evict them from DNA. An example of a chromatin remodeler is

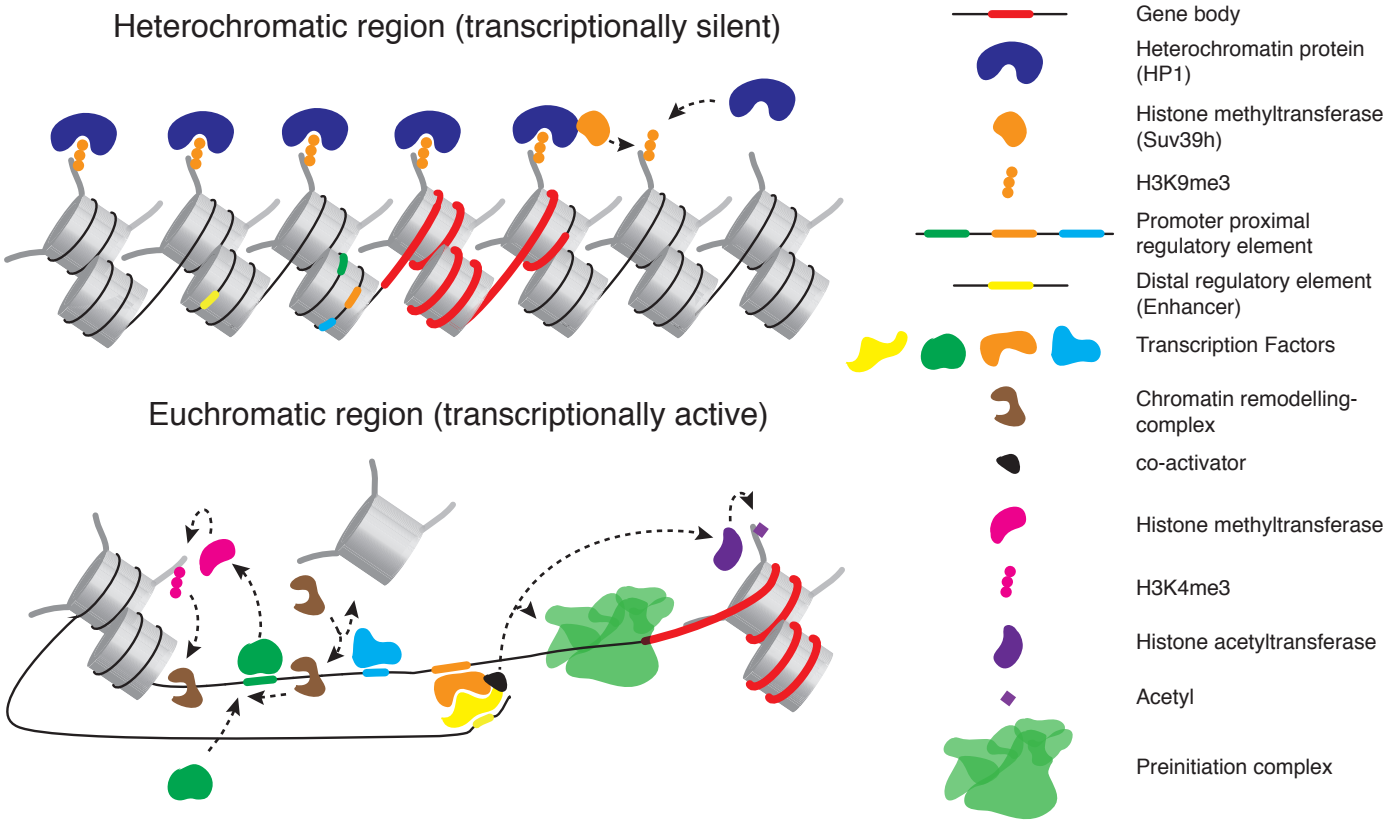


Fig 1.2 TRANSCRIPTIONAL CONTROL

The top panel illustrates the heterochromatic, or transcriptionally silent, state of chromatin. HP1 binds to deposited H3K9me3 mark and recruits a histone methyltransferase that will then deposit the H3K9me3 mark on neighbouring histones, resulting in spreading of heterochromatin.
The bottom panel illustrates a hypothetical transcriptionally active domain. Nucleosomes are positioned further from each other and even evicted by chromatin remodeling complexes. These complexes can be recruited by transcription factors that bind in the region (such as a pioneer transcription factor). Transcription factors can also recruit histone methyltransferases, which can in turn affect chromatin remodeling activity. Gene looping can bring together multiple regulatory elements and create complex binding sites where for example a co-activator can bind. Transcription factors or co-activators can induce the deposition of histone acetyl marks that affect the general chromatin density of the locus.

SWI/SNF, which both remodels and evicts nucleosomes (Lorch et al., 1999; Whitehouse et al., 1999). Additionally, posttranslational modifications (PTMs) of histone proteins correlate with the different transcriptional activity and chromatin states (Barski et al., 2007). The main modifications of histones that occur are the methylation and acetylation of lysines (K) on histones 3 and 4 (H3 & H4). Acetylation of histone tails directly influences nucleosome compaction and leads to a more ‘open’ chromatin state. The function of histone methylation depends on the position and the extent of the methylation. The promoters of transcriptionally active genes, for example, tend to be associated with nucleosomes that have a tri-methylated lysine on position 4 of histone 3 (H3K4me3). On the other hand, the repressive trimethylation of histone 3 on lysine 9 (H3K9me3) induces

the formation and spread of heterochromatin through the recruitment of heterochromatin protein 1 (HP1) (Lachner et al., 2001), and the histone methyltransferase that deposits the H3K9me3 mark. Many of the enzymes responsible for the reading and writing of the most common posttranslational modifications have been identified (reviewed in Chen and Dent, 2014; Musselman et al., 2012). Large scale mapping of PTMs has revealed the correlation of specific marks with specific states of chromatin, the function of many of these modifications however remains poorly understood. In addition to the recruitment of heterochromatin proteins, PTMs have been reported to recruit chromatin-remodeling complexes and enzymes that further modify histone tails (Lachner et al., 2001; Shi et al., 2006; Wysocka et al., 2006).

Transcription Factors

The transcription of DNA in an endogenous system generally requires specific transcription factors that provide access to the DNA. Any protein that affects the transcription of an mRNA, by binding directly to DNA, is referred to as a transcription factor. Factors that do not bind DNA but act to influence transcription are referred to as co-activators or co-repressors. The general transcription factors that are described above are, as the name implies, required for transcription of all mRNA transcripts. However, in addition to these general transcription factors there is a host of transcription factors that is involved in the regulation of transcription of specific mRNAs. Despite the degenerate sequence composition of the recognized target sites, these transcription factors show specificity of binding over the genome. The specificity is likely due to the DNA structure as well as cooperative binding. Transcription factors that bind to their targets may function to repress or increase expression through different mechanisms. One example of highly specific transcription factor binding and transcriptional activation is the binding of the transcription factor AST-1 in *C. elegans*. Binding of AST-1 to a specific motif that

present in a specific a set of genes, increases gene expression. This finally drives the differentiation to the dopaminergic neuronal fate (Flames and Hobert, 2009).

Traditionally, transcription factor binding was thought to correlate directly with transcriptional activity of the locus. Recent studies however have shown that transcription factor binding only altered gene expression of a corresponding gene in approximately 13% of binding events (Vokes et al., 2008). Since the remaining binding events are nevertheless specific, they may have a function other than the immediate transcriptional activation of the locus (reviewed in Spitz and Furlong, 2012). One important function that does not necessarily affect transcription directly (although it may indirectly induce transcription) is that of a pioneer transcription factor. These transcription factors can bind to chromatin that is not accessible for other factors and recruit chromatin remodeling complexes that open the chromatin. This leads to more transcription factor binding sites becoming accessible and increases the regulatory potential of the locus.

In eukaryotes, transcription factors often bind to elements that are present in a cluster of *cis*-regulatory elements. These sequences can be located in the promoter or in distal regulatory loci such as enhancers. The distance between these enhancer elements and their regulated genes may differ greatly. These characteristics are best explained by the proposed model of enhancer mechanism of action; gene looping (Amano et al., 2009). According to this model, a secondary structure within the chromatin fiber is responsible for the co-localization of two regions that are linearly separated by hundreds of kilobases. Another level of complexity is added when more than two of these sites localize together. The combinatorial effects of different enhancer elements as well as different transcription factors may result in more precise regulation of transcription (reviewed in Spitz and Furlong, 2012).

There are numerous classes of transcription factors. One family, of which we have already discussed a few members, is the family of nuclear hormone receptors (NHRs). This family is especially abundant in *C. elegans* (reviewed in Antebi, 2006). These transcription factors can be located in the cytoplasm or nucleus. Cytoplasmic NHRs present a nuclear localization signal (NLS) upon ligand binding, allowing them to be transported to the nucleus. Instead, constitutively nuclear NHRs can be associated with DNA in their inactive state and thereby allow for very rapid initiation of transcription when the ligand binds. In the absence of the ligand, the NHR can be associated with a co-repressor to further suppress spurious transcription from the locus. Although NHRs are well conserved, they are particularly abundant in *C. elegans*. There are an impressive 284 NHRs that have been identified in *C. elegans*. Of these 284 NHRs approximately 15 are conserved in the metazoan subkingdom (Gissendanner et al., 2004). As we have discussed previously, animals lacking specific NHRs show diverse phenotypes, including heterochronic defects.

1.3.4 Posttranscriptional regulation of gene expression

In addition to transcriptional regulation, the modulation of translation as well as the regulation of mRNA decay can directly affect gene expression. These processes are commonly referred to as posttranscriptional regulation of gene expression.

mRNA degradation

mRNA is degraded by three classes of ribonucleases; 5'-to-3' exonucleases, endonuclease and 3'-to-5' exonucleases. Since mature mRNA is capped at the 5' end, it is generally protected from degradation. Decapping of the transcript is required before any 5' exonucleases can degrade it. The poly-A tail and the associated poly-A binding protein protect the 3' end of the transcript. There are numerous quality control mechanisms that act to control transcription, splicing and translation and can initiate degradation by overcoming the mRNA protection at either the 3', or the 5' end. An

example of this is the NMD pathway that functions to recognize transcripts that contain a premature stop codon and induces decapping of the transcript. Another example is the miRNA-induced degradation of mRNA that we discuss below.

Translational repression or stimulation

Another way to regulate gene expression is to inhibit the translation of a particular mRNA transcript. RNA binding proteins can bind and thereby translationally repress them. Since degradation is not induced, this method of regulation can temporarily inhibit the translation of a transcript without affecting mRNA levels. An example of this is the silencing of maternal transcripts of *pal-1* in the *C. elegans* oocyte by the interaction with two RNA binding proteins: GLD-1 and MEX-3 (Mootz et al., 2004).

Degradation and translational inhibition are two ways in which posttranscriptional regulation can affect gene expression. Some posttranscriptional regulatory mechanisms affect both translation efficiency and mRNA stability. The poly-A binding proteins (PABPs), for example, are recruited to the newly formed transcripts and play a key role in the regulation of both mRNA levels (Coller et al., 1998) and translation efficiency (Allen et al., 2001; Imataka et al., 1998). Another example of posttranscriptional regulation on both the mRNA stability and translational efficiency is found in a class of small RNAs; miRNAs.

miRNA mediated posttranscriptional gene silencing

miRNA-mediated post-transcriptional regulation is an example of a well-conserved regulatory system. The small non-coding miRNAs were originally discovered when Ambros and colleagues found that the gene product from the lin-4 region is not protein coding, but rather produces a small RNA. The 3'UTR of lin-14 was found to contain a sequence that is the reverse complement sequence of the lin-4 small RNA (Lee et al., 1993). At the same time, it was shown that lin-4 regulates lin-14 protein levels without affecting mRNA levels (Wightman et al., 1993). For years after the initial discovery of lin-4,

the system was considered unique to *C. elegans* and no evidence of conservation was found. The identification of another miRNA, let-7, prompted the discovery and characterization of a class of small RNAs that post-transcriptionally regulates their mRNA targets by binding in the 3'UTR; miRNAs (Reinhart et al., 2000). The highly conserved let-7 and many other miRNAs have since been found in many different organisms, including vertebrates.

Most miRNAs are transcribed as a primary transcript from dedicated miRNA genes (Lau et al., 2001). This transcript is bound by the RNA binding protein Pasha (or DGCR8 in humans) and its associated endonuclease, Drosha (Denli et al., 2004; Gregory et al., 2004; Lee et al., 2003). The endonuclease cleaves the primary transcript to produce the precursor miRNA (pre-miRNA) stem-loop. Upon export from the nucleus by Exportin-5, the pre-miRNA is bound by Dicer (Grishok et al., 2001; Ketting et al., 2001; Lund et al., 2004). This enzyme, that is also an endonuclease, functions to cleave the pre-miRNA and produces the mature miRNA guide and passenger strands. The miRNA passenger strand (or miRNA*) is low in abundance and relatively unstable when compared to the guide strand (Lim et al., 2003). This is likely caused by the lack of protection from nucleases that Argonaute proteins offer to the guide strand (Vaucheret et al., 2004). The preferential loading of the guide strand into Argonaute depends on the thermodynamic stability of the miRNA:miRNA* complex (Khvorova et al., 2003). While loaded in Argonaute however, miRNAs can be very stable and often have half-lives of 12 hours or more (van Rooij et al., 2007; Gatfield et al., 2009). While the miRNA is loaded in Argonaute, GW182 is recruited to form, together with members of the CCR4-NOT complex, the miRNA induced RNA Silencing Complex (miRISC) (Behm-Ansmant et al., 2006). As the name implies, the miRISC functions to silence gene expression in sequence specifically. It can do so by binding a miRNA target site in the 3'UTR of a transcript and inducing translational

repression and degradation of the transcript (Bazzini et al., 2012; Djuranovic et al., 2012; Guo et al., 2010).

In the following work, we leverage the power of *C. elegans* as a model to study a complex network of oscillating gene expression and we study the role this network plays in nematode development.

2. Extensive Oscillatory Gene Expression during *C. elegans* Larval Development

2.1 Article

In the print version: Reprinted from Molecular Cell, vol. 53, Hendriks GJ, Gaidatzis D, Aeschimann F, Großhans H, Extensive Oscillatory Gene Expression during *C. elegans* Larval Development, 380-392, Copyright 2014, with permission from Elsevier.

http://dx.doi.org/10.1016/j.molcel.2013.12.013

In the electronic version of this thesis, a manuscript of “Extensive Oscillatory Gene Expression during *C. elegans* Larval Development” replaces the article described above. For the printed version of the paper, please see the reference above. Please note that the page numbers in the rest of the online version of this thesis have been kept identical to the print version.

Extensive oscillatory gene expression during *C. elegans* larval development

Gert-Jan Hendriks[1,2,4], Dimos Gaidatzis[1,3,4], Florian Aeschimann[1,2], and Helge Großhans[1,*]

[1] Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, CH-4058 Basel, Switzerland

[2] University of Basel, Petersplatz 1, CH-4003 Basel, Switzerland

[3] Swiss Institute of Bioinformatics, CH-4058 Basel, Switzerland

[4] Equal contribution

[*] Author for correspondence (helge.grosshans@fmi.ch)

Abstract

Oscillations are a key to achieving dynamic behavior and thus occur in biological systems as diverse as the beating heart, defecating worms, and nascent somites. Here we report pervasive, large-amplitude, and phase-locked oscillations of gene expression in developing *C. elegans* larvae, caused by periodic transcription. Nearly one fifth of detectably expressed transcripts oscillate with an 8h period, and hundreds change >10-fold. Oscillations are important for molting but occur in all phases, implying additional functions. Ribosome profiling reveals that periodic mRNA accumulation causes rhythmic translation, potentially facilitating transient protein accumulation as well as coordinated production of stable, complex structures such as the cuticle. Finally, large amplitude oscillations in RNA sampled from whole worms indicate robust synchronization of gene expression programs across cells and tissues, suggesting that these oscillations will be a powerful new model to study coordinated gene expression in an animal.

Introduction

As a ubiquitous feature in biology, oscillations have been considered one of its fundamental dynamic principles that drive processes away from equilibrium (Hasty et al., 2010). Part of their utility stems from the broad range of time-scales over which oscillations can drive dynamic behavior, from the millisecond periods of rhythmic neuronal spiking to the 24-hour periods of circadian clocks. Accordingly, the theoretic understanding and experimental design of oscillators has also been a major field of research in mathematical and synthetic biology (Hogenesch and Ueda, 2011; Tyson et al., 2008).

Biological oscillations drive not only rhythmic behaviors of cells, tissues, and organisms, but also periodic gene expression programs. Circadian clocks in particular direct extensive rhythmic gene expression to help organisms anticipate environmental changes caused for instance by daily cycles of light and temperature (Hogenesch and Ueda, 2011). In mammals, separate, peripheral clocks act in distinct tissues. They are synchronized through a central pacemaker in the suprachiasmatic nucleus of the hypothalamus and drive rhythmic expression of large sets of genes (Mohawk et al., 2012). Interestingly, although individual peripheral clocks can direct oscillations in 3-10% of active genes in a given tissue, there is little overlap in the sets of genes that undergo periodic expression in different tissues (Mohawk et al., 2012). As exemplified by the vertebrate segmentation clock, oscillators can also be utilized to drive periodic developmental events (Kageyama et al., 2012), but generally much less is known about rhythmic gene expression in developmental

than circadian contexts. In *C. elegans*, only ~20 genes are known to be periodically expressed during development (Johnstone and Barry, 1996; Lassandro et al., 1994; Hashmi et al., 2004; Davis et al., 2004; Frand et al., 2005; Hao et al., 2006; Kostrouchova et al., 2001; Gissendanner et al., 2004; Jeon et al., 1999; Monsalve et al., 2011; McMahon et al., 2003) . However, because systematic and quantitative studies have not been performed, the prevalence of periodic gene expression has remained unknown and insights into general principles underlying the rhythmic expression patterns have been lacking.

Here, using genome-wide and temporally highly resolved gene expression studies, we reveal extensive periodic gene expression during *C. elegans* larval development, comprising a fifth of expressed genes. Our characterization reveals robust, transcriptionally driven oscillations across a continuum of phases that result in periodic translation, and thus promote periodic developmental processes such as molting. These results highlight an unanticipated dynamic and complexity of gene expression patterns during *C. elegans* development. Moreover, we propose that a unique combination of features makes these oscillations a powerful model to study coordinated gene expression in an animal.

Results

The expression of thousand of genes oscillates extensively during development

To obtain insight into the dynamics of gene expression during *C. elegans* larval development, we performed a high-resolution and genome-wide time-course analysis. Synchronized L1 stage larvae were placed on food at 25°C, and samples collected hourly over a 16h period that covered development from L3 to the young adult stage. Unexpectedly, cross correlation plots of gene expression profiles obtained by mRNA sequencing revealed a periodical increase in similarity, rather than a progressive decrease during development (Fig. 1A). For instance, expression patterns at t=21h were noticeably more similar to expression patterns at t=27h and 28h than to those at 24h.

To identify trends in gene expression levels that could explain this observation, we performed principle component analysis. Three principle components (PC) explained ~92% of the variation (Fig. 1B). The loadings of the first displayed a monotonic increase over time, thus revealing a set of genes whose expression is altered continuously along the time course. These are mostly related to germline development and function (see below).

The loadings of the second and third PCs displayed an oscillatory pattern with a period of eight hours (Fig. 1B). This suggested that a large number of genes oscillated at a common period but with different phases (Experimental Procedures). To test this possibility, we fitted a general cosine function with a fixed period of 8h and unknown phase and amplitude. Since the first PC explained a large part of the

data, we included it as an additional component when performing the fit. Comparing PC1 to the oscillation amplitude for every gene revealed two distinct classes of dynamically changing genes with largely mutually exclusive membership: genes that increase their expression and those that oscillate (Fig. 1C). This allowed us to categorize genes into 'flat' (black), 'rising' (green) and 'oscillating' (red) based on empirically chosen cut-offs (Supplemental Experimental Procedures). The oscillating class contained the previously reported ~20 genes with periodic expression. However, oscillatory expression is much more pervasive in that this class comprised 2,718 genes, corresponding to 18.9% of 14,378 expressed genes (Table S1).

Following cosine wave fitting, we could examine amplitudes and phases of oscillations more readily and observed large changes (Fig. 1D; note that the amplitude of the cosine function equals half the magnitude of change between trough and peak expression). For instance, the levels of all 2,718 'oscillating' genes changed more than 2.1-fold, and those of >400 genes exceeded a 10-fold change. For comparison, comprehensive studies of circadian clocks revealed typical median expression changes of 2-fold (Hughes et al., 2009; Duffield, 2003). The extent of changes seen here is yet more striking when considering that we examined RNA from whole, undissected worms, not from specific tissues or cells.

A unifying theme among the previously identified periodically expressed genes appears to be a connection to molting as determined by the molecular nature of the encoded proteins and/or mutant phenotypes (Johnstone and Barry, 1996; McMahon et al., 2003; Lassandro et al., 1994; Hashmi et al., 2004; Davis et al., 2004;

Frand et al., 2005; Hao et al., 2006; Kostrouchova et al., 2001; Gissendanner et al., 2004; Jeon et al., 1999; Monsalve et al., 2011) . Molting occurs with an eight hour periodicity at the end of each larval stage and involves the generation of a new cuticle and shedding of the old one (Monsalve and Frand, 2012). As we will discuss below, several classes of genes with roles in molting are also enriched among the oscillatory genes that we identify here. Nonetheless, rather than being restricted to the times of the molt, rhythmic gene expression occurred in all phases (Fig. 1D; a full period corresponds to 8h or 360°, and a phase difference of 45° thus to a peak shift by 1h; the molt occurs roughly between 180° and 270°). This was also readily apparent when plotting the expression changes of all 'oscillating' genes in a heatmap (Fig. 1E). A somewhat larger number of genes were in the 90°, 135° and 315° bins. RT-qPCR confirmed the oscillations and phases of expression for 18 genes (Fig. 1F) in RNA sampled from an independent time course.

Oscillations are independent of life history and synchronization procedure

To assess the robustness of the oscillations and potential effects of life history (Karp et al., 2011), we analyzed mRNA expression patterns in *C. elegans* that had exited from dauer diapause arrest. In contrast to continuous larval development, which proceeds from L1 through the L4 stage, dauer animals arrest development in an alternative L3 stage for instance in response to starvation (Hu, 2007). Synchronous exit from dauer can be induced by providing food to such dauer larvae, which will

subsequently go through an L3-to-L4 molt and continue development through the L4 into the adult stage.

We placed dauer-arrested animals on food and collected hourly samples over 22h. Cross correlation plots again revealed periodical similarity of the expression patterns (Fig. 2A). A heatmap that displayed changes in the expression patterns of all genes that were periodically expressed following continuous development revealed strikingly similar oscillatory expression patterns for post-dauer animals (Fig. 2B). Thus, oscillations are highly robust and not substantially influenced by life history. This was in fact also true for the steadily upregulated genes, whereas more diversity was seen with the bulk of genes that did not follow either of these two patterns (Fig. 2B).

Oscillations are locked in phase over time

To further explore the robustness of the oscillations, we followed them over time by preparing in an extended time course where we sampled worms every two hours between 18 h and 36 h of development at 25°C, and thus covering the entire L3 stage along with the later stages investigated in the first experiment. Beyond confirming that the genes with oscillatory expression during the L4 stage were also rhythmically expressed during the L3 stage (see Fig. 5, below), this permitted us to examine whether phase relationships were maintained across these stages. Strikingly, and consistent with a robust oscillator, we found that phases were tightly

locked between stages. Thus, peaks occurred in the same phase during the earlier and the later stage (Fig. 2C).

Oscillations are not temperature-compensated

Although 8h is a harmonic of the 24h circadian cycle, it is also the period of the molting cycle. Unlike circadian clocks, which are temperature-compensated (Hogenesch and Ueda, 2011), the length of the molting cycle in *C. elegans* is strongly temperature-dependent and approximately doubles when ambient temperature is decreased from 25°C to 15°C. To test whether the oscillator that drove the transcript oscillations that we had identified was compensated, we grew worms at 15°C and collected samples every 3h for 36h between late L3 and adult stage. RT-qPCR-based examination of the transcripts that oscillated with an ~8h period at 25°C (Fig. 1F) revealed that the oscillations themselves were maintained but that the period now increased to ~18h (Fig. 2D,E). This finding not only provides further evidence for the robustness of the oscillations, it also reveals that, unlike circadian clocks, the underlying oscillator is not compensated. This is consistent with, yet does not prove, a mechanism important to keep developmental time.

Oscillations occur in diverse somatic tissues

To gain insight into the developmental function of the oscillations, we sought to determine where periodic gene expression occurred. To this end, we plotted the levels of all genes in the soma versus the levels in the gonad of young adult animals and color-coded them according to the three categories, 'oscillating', 'rising' and 'flat' (Fig. 3A). Whereas genes whose expression neither oscillated nor increased steadily tended to be equally expressed in soma and gonad (Fig. 3A,B), genes with steadily increasing expression exhibited typically higher levels in the gonad (Fig. 3A,B). By contrast, 'oscillating' genes were more highly expressed in the soma than the gonad (Fig. 3A,B).

To confirm that somatic expression contributed significantly to the oscillations, we profiled gene expression of germline-less *glp-4* mutant animals across a developmental time course. Because *glp-4* animals experience a developmental delay, we harvested them at later time points than wild-type animals to obtain animals of comparable stages (Supplemental Experimental Procedures). Consistent with our hypothesis, we could readily observe periodic gene expression in *glp-4* animals (Fig. 3C), despite the fact that these animals grow more asynchronously than wild-type animals, which will obscure oscillations. Collectively, these results suggest that periodic gene expression occurs preferentially in somatic tissues. Moreover, the extensive loss of upregulation among 'rising' genes at the later time points of *gpl-4* animal development demonstrates that these mostly correspond to germline-expressed genes.

The spatial expression patterns of some 2,000 *C. elegans* genes, mostly with human homologues, were recently inferred from an analysis of GFP expression

driven by their promoters (Hunt-Newbury et al., 2007). Using these data, we found that epidermal cells (hypodermis and seam cells) were enriched for expression of oscillating genes whereas body wall muscle cells and neurons appeared to be depleted (Fig. 3D). However, the extents of depletion and enrichment were generally small, not exceeding 1.2-fold, which reveals that oscillatory gene expression can occur in many somatic tissues.

Periodic expression of cuticular collagen genes may facilitate assembly of a complex structure

To identify molecular function and processes subject to periodic gene expression, we performed gene ontology (GO) term enrichment and depletion analysis (Ashburner et al., 2000). Consistent with oscillations as a preferentially somatic event, various terms related to function and development of the germline were strongly depleted (Table S2). Moreover, many of the most significantly depleted terms were linked to translation and the ribosome, e.g. among 123 expressed genes associated with "Structural constituent of ribosome", mostly encoding ribosomal proteins, none was periodically expressed.

By contrast, GO:0003735 "Structural constituent of cuticle", a term associated almost exclusively with cuticular collagen genes, was highly enriched (Fig. 4A). Collagen genes had accounted for half of the ~20 genes previously known to undergo periodic expression (Johnstone and Barry, 1996; McMahon et al., 2003),

but we were nonetheless surprised by the extent of oscillations: 91 of 126 expressed genes, or 72%, were periodically expressed. All but one of the remaining 35 genes increased expression across the time course, and these presumably represent adult-specific collagens.

Closer examination of expression phases further revealed that collagen genes preferentially peak between 180° and 270° (Fig. 4B, C), the time when animals molt as determined by the occurrence of lethargus in a separate time course (data not shown). Nonetheless, different classes of collagens displayed noticeably dissimilar expression patterns: for members of the largest class, the *col* genes, expression peaked indeed mostly during the molt (Fig. 4C). Cuticular collagens are assigned to this gene class solely by sequence homology. By contrast, collagens of the *rol* family were identified genetically through mutations that caused worms to display an abnormal movement (roller) phenotype due to helical distortions of the cuticle. *sqt* (squat) genes can be mutated to cause heterozygous roller and homozygous dumpy (short, fat) phenotypes. We find that expression of *sqt* and *rol* genes occurs in a highly coordinated manner, with peaks occurring all before the molt within a window of 10° (Fig. 4C), and almost identical amplitudes ($\log_2$(A) ranging from 2.67 to 2.72). Taking further into account the specific genetic interactions (Kusch and Edgar, 1986) as well as sequence similarities among these cuticular proteins (Kramer et al., 1990), we propose that these proteins form a cuticular substructure.

Such a model is consistent with findings for the *dpy* (dumpy) collagen genes: Individual DPY proteins contribute to one of two specific cuticular substructures (McMahon et al., 2003), and this functional specification coincides with expression

in two separate clusters (Fig. 4C; (McMahon et al., 2003; Johnstone and Barry, 1996). The specific function of the 'solitary' *dpy-17* is currently unknown and may be unique.

We conclude that although collagen expression occurs preferentially during the molt, individual collagen genes reveal highly specific temporal expression patterns, which permits their incorporation into the cuticle at the right time and place.

Rhythmic expression frequently occurs for genes that encode proteases or steroid metabolism and signaling proteins

In addition to cuticular collagens, two other major themes became apparent following GO-term enrichment analysis of 'oscillating' genes (Fig. 4A). The first was proteolysis and resulted from frequent rhythmic expression among three peptidase families (Table 1; Fig. 4D). These families contain peptidases with known functions in molting, e.g., NAS-37 (Suzuki et al., 2004; Davis et al., 2004) and CPZ-1 (Hashmi et al., 2004), but the diversity of phases (Fig. 4D) suggests roles beyond molting. Periodic protease activity may also be achieved or refined by rhythmic expression of protease inhibitors, as indicated by the enrichment of the GO-term "Serine-type endopeptidase inhibitor activity" (Fig. 4A,B).

The second theme comprised steroid hormone metabolism and signaling. Thus, cytochrome P450s (CYPs), UDP-glucuronosyl/glucosyl transferases (UGTs), and short chain dehydrogenases/reductases enzymes are enriched among

oscillatory genes (Table 1, Fig. 4E). These types of enzymes, and the steroid hormones they metabolize, play key roles in molting in insects (King et al., 2000; Iga and Kataoka, 2012) and, presumably, *C. elegans* (Entchev and Kurzchalia, 2005). Hence, their enrichment might reflect a function in molting hormone metabolism in *C. elegans*.

The Hedgehog (Hh) pathway genes of *C. elegans*, which frequently exhibit oscillatory expression (Fig. 4A,F, Table 1), may also be linked to this theme. This is because they may function in sterol metabolism and signaling rather than canonical Hh signaling (Bürglin and Kuwabara, 2006), and because many of them are important for molting (Zugasti et al., 2005; Hao et al., 2006). Periodic expression was previously noted for four Hh-related genes (Hao et al., 2006), but we find it to be pervasive both among the genes in these families and the putative Hh receptors, which are characterized by the presence of a Patched-like domain (Table 1, Fig. 4F). An attractive hypothesis is that oscillations can act as filters for signaling by helping to match specific receptor and ligand pairs through finely tuned co-expression, allowing them to execute distinct functions in molting and possibly other processes.

mRNA level oscillations cause rhythmic protein production

The examples provided above suggest that mRNA levels oscillations could permit either transient protein accumulation, or production of stable proteins at a given time when they are required for integration into a complex structure. However, both

scenarios are based on the assumption that mRNA level oscillations cause rhythmic protein production. This assumption is not trivial as evidenced by the discovery of extensive posttranscriptional regulation that refines and modifies circadian transcriptional oscillations (Hogenesch and Ueda, 2011). Hence, to rigorously test this notion, we examined whether transcript level oscillations in *C. elegans* larvae resulted in periodic transcript translation. We performed ribosome profiling to assess the translational status of cellular mRNAs (Ingolia et al., 2009) over a separate time course. We sequenced ribosome protected fragments (RPFs) as well as rRNA-depleted total RNA from worms collected every 2 hours during development from L3 to young adult stage. Confirming that RPFs reflect translation, they displayed robust phasing and were depleted from 5'UTRs and, yet more extensively, 3'UTRs (Fig. S1), exactly as expected (Ingolia et al., 2009). Strikingly, the oscillations that occurred at a transcript level were also present at the translational level (Fig. 5). Conversely, the oscillations seen at the translational level were well explained by transcript level oscillations. Hence, we conclude that periodic protein production is indeed the default result of periodic transcript accumulation in *C. elegans* larvae.

Oscillatory gene expression is driven by periodic transcription

Although the circadian clock was long considered to drive oscillations through periodic transcription, pre-mRNA level oscillations appear to be a poor predictor of

mRNA level oscillations (Koike et al., 2012). On the other hand, the promoters of a few periodically expressed *C. elegans* genes have been shown to drive periodic accumulation of reporter proteins (Frand et al., 2005; Davis et al., 2004; Hao et al., 2006). We therefore wished to test quantitatively and comprehensively whether the present oscillator drives rhythmic accumulation of mRNAs through rhythmic transcription. Hence, we sequenced rRNA-depleted total RNA samples to determine pre-mRNA levels as a proxy of transcription. For this, we used the same RNA sample as those that we had subjected to mRNA sequencing (Fig. 1). When analyzing transcripts for which sufficient intronic reads were available, we found that pre-mRNAs and mRNAs oscillated with highly comparable phase and amplitudes (Fig. 6A, B). Consistent with transcriptional oscillations preceding mature transcript level oscillations, pre-mRNA peaks occurred on average some 15 min prior to mature mRNA peaks (Fig. S2). We conclude that transcriptional regulation is a major contributor to the oscillations that we observe. At the same time, 'oscillating' genes are found across all chromosomes and without any apparent clustering of genes according to phase of expression (Fig. 6C), suggesting that there are no specific chromosomal domains that drive, or are particularly permissive to, periodic gene expression.

Discussion

Extensive oscillatory gene expression serves a developmental function in C. elegans *larvae*

We report here extensive oscillatory gene expression during *C. elegans* larval development. Superficially, the oscillations may appear reminiscent of transcriptional bursts (Raj and van Oudenaarden, 2008). However, we can rule out that the two are linked based on the fact that transcriptional bursts occur stochastically. Hence, they cannot be observed in ensembles of cells, let alone the populations of whole animals, comprised of numerous distinct cell types and tissues, that we studied here.

Formally, it is nonetheless possible that periodic gene expression does not serve a particular purpose but represents mere noise, perhaps as a side effect of another periodic process. The fact that oscillations involve high-amplitude expression level changes of a large number of genes, and that the process is highly robust, strongly argue against this possibility. Thousands of genes change >2-fold, and hundreds >10-fold. If these changes reflected 'noise' rather than function, it would be difficult to conceive mechanisms by which animals and their cells could prevent this noise from drowning the signal of 'meaningful' changes in gene expression.

At the same time, oscillations were highly reproducible over three independent time courses examined by different detection technologies; independent of life history; and, although not temperature-compensated, robust under different temperature regimens. Finally, we not only observed reproducibly oscillations for the same genes, but these oscillations were also phase-locked, i.e., the phase difference between pairs of different genes remained the same in different experiments. Kim et al. (2013) also recently observed extensive oscillatory gene expression but grouped oscillatory genes in a small number of distinct expression clusters instead of specifically examining phases and amplitudes. However, when we analyzed their data, we found a similar continuum of phases as in our data, and phase-locking was maintained across stages (Fig. S3). Summarily, the impressive robustness of the process thus provides additional strong evidence for a functional role.

A role of periodic gene expression in molting

The above considerations support that widespread rhythmic gene expression in *C. elegans* serves a biological purpose. Because oscillations are not temperature-compensated, they cannot function analogously to a circadian clock. Instead, they are likely to control a developmental process. The repetitious nature of molting makes it the strongest candidate, and gene expression oscillations and molting occur indeed both with the same periodicity.

Cuticular collagens are a particularly clear example of oscillatory expression of genes involved in molting. Consistent with cuticle generation during the molt, their expression peaks preferentially during, or in close temporal connection with, the molt. Surprisingly, however, this pattern is atypical, i.e., a continuum of gene expression phases is visible when examining the entirety of 2,719 genes. When combined with the fact that this number corresponds to nearly a fifth of expressed genes, it seems indeed likely that oscillations occur for many genes that are not required for the actual molting process. However, we speculate that extensive oscillatory gene expression permits coordination of molting with other cellular, developmental, or behavioral processes, which can be essential for viability (Ruaud and Bessereau, 2006). In fact, given the extensive and robust phase-locking, it seems possible that the succession of periodically expressed genes may define a larval growth or development module in *C. elegans.*

The utility of gene expression oscillations extends beyond driving transient protein accumulation

Atypical as the expression patterns of cuticular collagen genes relative to the overall patterns may be, they provide particularly interesting insights into the utility of periodic gene expression by highlighting a function that goes beyond achieving transient protein accumulation. Cuticular collagens are stable proteins that remain associated with the cuticle once incorporated. At the same time, the cuticle is a

complex structure of several layers that need to be sequentially assembled (Page and Johnstone, 2007). Rhythmic and phase-locked collagen production then permits streamlining of this production through "just in time" delivery of individual components, facilitating faithful and efficient cuticle assembly. We propose that other oscillators may similarly be utilized to this end.

Phase information provides insights into the biology of cuticle synthesis

Johnstone and colleagues previously noticed the co-expression of *dpy* collagens destined for the same cuticular sub-structure (Johnstone and Barry, 1996; McMahon et al., 2003) and reported that collagens generally fell in three expression clusters termed 'early', 'intermediate' and ‘late’ (Johnstone and Barry, 1996; McMahon et al., 2003). However, these conclusions were based on semi-quantitative expression analysis of only ten genes. Our comprehensive analysis reveals a much more sophisticated choreography of collagen gene expression that involves a broad distribution of phases. This finding suggests the possibility that individual collagens may indeed be more functionally distinct than previously appreciated, and that this, and not a need to achieve massive production of a generic collagen, has given rise to the large number of collagen genes present the *C. elegans*.

Another unexpected finding in our data is the observation of collagen expression after the molt. It had previously been assumed that cuticle synthesis was a molt-specific event (Page and Johnstone, 2007), and that the larval growth that

does occur between molts is facilitated by physical stretching of the flexible cuticle (Knight et al., 2002). Our data suggest the possibility that synthesis of specific collagens could additionally contribute to cuticle growth between molts. Consistent with this notion, cuticle regeneration can occur after physical injury (Pujol et al., 2008).

Rhythmic transcription drives mRNA level oscillations

Our finding that pre-mRNA expression patterns faithfully mirror mature mRNA oscillations reveals rhythmic transcription as a key driver of oscillations and implies a need for rhythmically modulated transcription factor activity. The diversity of expression phases potentially suggests the involvement of many different factors, and we observe periodic expression for 92 transcription factors, including several that are required for molting (data not shown).

Our efforts to identify motifs in the promoters of co-expressed genes have so far failed to yield convincing leads as to the identity of involved transcription factors (our unpublished data). However, *lin-42*, the rhythmically expressed *C. elegans* orthologue of the core circadian clock gene and transcriptional regulator *Period* (Jeon et al., 1999; Monsalve et al., 2011) is a particularly interesting candidate, since loss of *lin-42* activity causes arrhythmic molts (Monsalve et al., 2011). Whether LIN-42 indeed drives rhythmic transcription to time molts remains to be established, since no targets are known. Validating this hypothesis will presumably require the

use of single animal-based techniques, because arrhythmia, in population-based studies, will inevitably generate an appearance of decreased oscillations. At any rate, it will be of great interest to identify in future studies the transcription factors at the heart of these oscillations, and understand how they cross-talk to one another. Possibly, this may occur in analogy with periodic transcription during the budding yeast cell cycle (Bähler, 2005), where transcription factors sequentially regulate one another (Simon et al., 2001) and additionally, differ in their activities when present alone or in pairs (Kato et al., 2004). These two features may then suffice to achieve both phase-locking and a continuum of expression phases. At the same time, we would like to emphasize that although our data demonstrate that the oscillator manifests by driving extensive oscillatory transcription, it remains to be established whether the functionality of the core oscillator itself equally depends on transcription.

A unique combination of features defines a new model to study gene expression oscillations and coordinated gene expression in an animal

It has been emphasized that the study of diverse oscillators in various systems has been instrumental to identify unifying themes and idiosyncrasies (Hogenesch and Ueda, 2011; Tyson et al., 2008). We propose that the oscillations that we present here have a number of unusual characteristics that merit investigation. For instance, insights into the mechanisms that achieve robust phase-locking and broad

distribution of phases may be illuminating for our understanding of coordinated gene expression more generally. Moreover, a relatively short period and high amplitudes generate an interesting problem for gene expression kinetics: Rapid induction needs to be balanced with rapid degradation. In other words, are the affected transcripts inherently unstable, which would necessitate yet higher transcription levels upon induction, or is their degradation rate increased once expression declines, and if so, by which means? Finally, it is striking that oscillations are robustly detectable in RNA from whole animals, as this suggests the presence of effective mechanisms, yet to be uncovered, that coordinate oscillations spatially and temporally. The diversity of genetic and other tools available for *C. elegans* will permit extensive and productive investigation of this novel model oscillator to resolve these and other issues.

Experimental Procedures

Worm culture

Animals were grown and synchronized at the L1 or dauer stages according to standard procedures detailed in the Supplemental Experimental Procedures.

RNA sequencing, ribosome profiling, and RT-qPCR

For RNA sequencing, RNA was extracted from extensively washed animals and extracted by freeze-thawing or mortar and pestle in Tri Reagent. DNase-treated,

quality-controled RNA was used for preparation of mRNA sequencing or rRNA-depleted total RNA sequencing using commercial kits and protocols. Ribosome profiling was performed following an adaptation of published protocols (Ingolia et al., 2012; Bazzini et al., 2012). RT-qPCR-based validation of mRNA-seq data was performed on a separately collected time course, with candidate genes chosen based on a wide distribution of their respective phases. All relevant steps are further detailed in the Supplemental Experimental Procedures

Processing of the RNA-seq and ribosome profiling data

All the RNA-seq data (50bp read length) were mapped to the *C. elegans* genome (ce6) using the spliced alignment algorithm SpliceMap included with the R package QuasR (Au et al., 2010) as detailed in the Supplemental Experimental Procedures. For ribosome profiling, bowtie (Langmead et al., 2009) was used to map the ~30 bp ribosome-protected fragments obtained after 3' adapter removal to the *C. elegans* genome (ce6) allowing only uniquely mapping reads. Gene expression quantification was performed by counting reads and analyzing them as detailed in the Supplemental Experimental Procedures.

Principal Component Analysis and cosine curve fitting

After mean-normalization of the log2 gene expression levels, we performed PCA using the function princomp in R with default parameters. A combination of the principle components PC2 and PC3 permitted representation of sinusoidal waves

with an eight hour period and any phase angle (Supplemental Experimental Procedures). Hence, for each gene we fitted a separate cosine curve with a known period of 8 hours and thus a frequency $\omega = 2*\pi/8h$ and unknown variables C and φ. Since ($C*\cos(\omega t + \varphi) = A*\cos(\omega t) - B*\sin(\omega t)$ with $A = C*\cos(\varphi)$ and $B = C*\sin(\varphi)$), we performed the fit using a linear regression including the two components $\cos(\omega t)$ and $-\sin(\omega t)$ as regressors. Because a large proportion of the variance in the data was explained by the non-periodic first principal component PC1, we included it as a separate regressor during the fit. Based on a scatterplot comparing PC1 to the oscillation amplitude, we classified the genes into three categories, oscillating, increasing and flat. A detailed description is provided in the Supplemental Experimental Procedures.

GO enrichment analysis

GO annotations for *C. elegans* were downloaded from http://www.geneontology.org/gene-associations/gene_association.wb.gz and enrichments and depletions calculated as detailed in in the Supplemental Experimental Procedures. To display the normalized phase distributions for all the oscillating genes that belong to enriched GO terms in a heatmap, we binned the phases into 8 equally sized intervals and corrected for gene number variations across phase bins and GO terms (Supplemental Experimental Procedures).

Tissue enrichment analysis

To determine whether periodic gene expression occurred preferentially in the soma

or the germline, we examined mRNA sequencing data obtained for gonads dissected out of wild-type young adult animals and for germline-less *glp4ts* mutant young adult animals, kindly provided by Dr. Rafal Ciosk (C. Scheckel, D.G., and R. Ciosk, unpublished data). *C. elegans* promoter::GFP fusions expression data (Hunt-Newbury et al., 2007) were obtained from http://gfpweb.aecom.yu.edu and analyzed for those cell types that had at least 100 expressed genes as detailed in the Supplemental Experimental Procedures.

Accession numbers

All sequencing and ribosome profiling data generated for this study have been deposited in NCBI's Gene Expression Omnibus (Edgar et al., 2002) and are accessible through GEO Series accession number GSE52910 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= GSE52910).

Acknowledgements

We thank Dirk Schübeler and David Gatfield for comments on an earlier version of the manuscript. We are grateful to Rafal Ciosk for sharing unpublished data, Kirsten Jacobeit, Sophie Dessus-Babus, and Stéphane Thiry for library preparation and sequencing, and to Monika Fasler for technical support. The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under European Research Council grant agreement n° 241985, the Novartis Research Foundation through the FMI, and the Swiss

National Science Foundation [SNF 31003A_127052]. GJH was supported by a Boehringer Ingelheim Fonds PhD Fellowship.

References

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet *25*, 25-29.

Au, K.F., Jiang, H., Lin, L., Xing, Y., and Wong, W.H. (2010). Detection of splice junctions from paired-end RNA-seq data by SpliceMap. Nucleic Acids Res *38*, 4570-78.

Bazzini, A.A., Lee, M.T., and Giraldez, A.J. (2012). Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. Science *336*, 233-37.

Bähler, J. (2005). Cell-cycle control of gene expression in budding and fission yeast. Annu Rev Genet *39*, 69-94.

Davis, M.W., Birnie, A.J., Chan, A.C., Page, A.P., and Jorgensen, E.M. (2004). A conserved metalloprotease mediates ecdysis in Caenorhabditis elegans. Development *131*, 6001-08.

Duffield, G.E. (2003). DNA microarray analyses of circadian timing: the genomic basis of biological time. J Neuroendocrinol *15*, 991-1002.

Edgar, R., Domrachev, M., and Lash, A.E. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository Nucleic Acids Res. 30, 207-210

Entchev, E.V., and Kurzchalia, T.V. (2005). Requirement of sterols in the life cycle of the nematode Caenorhabditis elegans. Semin Cell Dev Biol *16*, 175-182.

Fisher, A.I., and Lee, N.I. (1983). A correlation coefficient for circular data coefficient for circular data. Biometrika *70*, 327-332.

Frand, A.R., Russel, S., and Ruvkun, G. (2005). Functional genomic analysis of C. elegans molting. PLoS Biol *3*, e312.

Gissendanner, C.R., Crossgrove, K., Kraus, K.A., Maina, C.V., and Sluder, A.E. (2004). Expression and function of conserved nuclear receptor genes in Caenorhabditis elegans. Dev Biol *266*, 399-416.

Hao, L., Johnsen, R., Lauter, G., Baillie, D., and Bürglin, T.R. (2006). Comprehensive analysis of gene expression patterns of hedgehog-related genes. BMC Genomics *7*, 280.

Hashmi, S., Zhang, J., Oksov, Y., and Lustigman, S. (2004). The Caenorhabditis elegans cathepsin Z-like cysteine protease, Ce-CPZ-1, has a multifunctional role during the worms' development. J Biol Chem *279*, 6035-045.

Hasty, J., Hoffmann, A., and Golden, S. (2010). Systems biology of cellular rhythms: from cacophony to symphony. Curr Opin Genet Dev *20*, 571-73.

Hogenesch, J.B., and Ueda, H.R. (2011). Understanding systems-level properties: timely stories from the study of clocks. Nat Rev Genet *12*, 407-416.

Hu, P.J. (2007). Dauer. WormBook , 1-19.

Hughes, M.E., DiTacchio, L., Hayes, K.R., Vollmers, C., Pulivarthy, S., Baggs, J.E., Panda, S., and Hogenesch, J.B. (2009). Harmonics of circadian gene transcription in mammals. PLoS Genet *5*, e1000442.

Hunt-Newbury, R., Viveiros, R., Johnsen, R., Mah, A., Anastas, D., Fang, L., Halfnight, E., Lee, D., Lin, J., et al. (2007). High-throughput in vivo analysis of gene expression in Caenorhabditis elegans. PLoS Biol *5*, e237.

Iga, M., and Kataoka, H. (2012). Recent studies on insect hormone metabolic pathways mediated by cytochrome P450 enzymes. Biol Pharm Bull *35*, 838-843.

Ingolia, N.T., Brar, G.A., Rouskin, S., McGeachy, A.M., and Weissman, J.S. (2012). The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. Nat Protoc *7*, 1534-550.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218-223.

Jeon, M., Gardner, H.F., Miller, E.A., Deshler, J., and Rougvie, A.E. (1999). Similarity of the C. elegans developmental timing protein LIN-42 to circadian rhythm proteins. Science *286*, 1141-46.

Johnstone, I.L., and Barry, J.D. (1996). Temporal reiteration of a precise gene expression pattern during nematode development. EMBO J *15*, 3633-39.

Kageyama, R., Niwa, Y., Isomura, A., González, A., and Harima, Y. (2012). Oscillatory gene expression and somitogenesis. Wiley Interdiscip Rev Dev Biol *1*, 629-641.

Karp, X., Hammell, M., Ow, M.C., and Ambros, V. (2011). Effect of life history on microRNA expression during C. elegans development. RNA *17*, 639-651.

Kato, M., Hata, N., Banerjee, N., Futcher, B., and Zhang, M.Q. (2004). Identifying combinatorial regulation of transcription factors and binding motifs. Genome Biol *5*, R56.

Kim, D.H., Grün, D., and van Oudenaarden, A. (2013). Dampening of expression oscillations by synchronous regulation of a microRNA and its target. Nat Genet

King, C.D., Rios, G.R., Green, M.D., and Tephly, T.R. (2000). UDP-glucuronosyltransferases. Curr Drug Metab *1*, 143-161.

Knight, C.G., Patel, M.N., Azevedo, R.B.R., and Leroi, A.M. (2002). A novel mode of ecdysozoan growth in Caenorhabditis elegans. Evol Dev *4*, 16-27.

Koike, N., Yoo, S.-H., Huang, H.-C., Kumar, V., Lee, C., Kim, T.-K., and Takahashi, J.S. (2012). Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. Science *338*, 349-354.

Kostrouchova, M., Krause, M., Kostrouch, Z., and Rall, J.E. (2001). Nuclear hormone receptor CHR3 is a critical regulator of all four larval molts of the nematode Caenorhabditis elegans. Proc Natl Acad Sci U S A *98*, 7360-65.

Kramer, J.M., French, R.P., Park, E.C., and Johnson, J.J. (1990). The Caenorhabditis elegans rol-6 gene, which interacts with the sqt-1 collagen gene to determine organismal morphology, encodes a collagen. Mol Cell Biol *10*, 2081-89.

Kusch, M., and Edgar, R.S. (1986). Genetic studies of unusual loci that affect body shape of the nematode Caenorhabditis elegans and may code for cuticle structural proteins. Genetics *113*, 621-639.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

Lassandro, F., Sebastiano, M., Zei, F., and Bazzicalupo, P. (1994). The role of dityrosine formation in the crosslinking of CUT-2, the product of a second cuticlin gene of Caenorhabditis elegans. Mol Biochem Parasitol *65*, 147-159.

McMahon, L., Muriel, J.M., Roberts, B., Quinn, M., and Johnstone, I.L. (2003). Two sets of interacting collagens form functionally distinct substructures within a Caenorhabditis elegans extracellular matrix. Mol Biol Cell *14*, 1366-378.

Mohawk, J.A., Green, C.B., and Takahashi, J.S. (2012). Central and peripheral circadian clocks in mammals. Annu Rev Neurosci *35*, 445-462.

Monsalve, G.C., and Frand, A.R. (2012). Toward a unified model of developmental timing: A “molting” approach. Worm *1*, 221-230.

Monsalve, G.C., Van Buskirk, C., and Frand, A.R. (2011). LIN-42/PERIOD controls cyclical and developmental progression of C. elegans molts. Curr Biol *21*, 2033-045.

Page, A.P., and Johnstone, I.L. (2007). The cuticle. WormBook , 1-15.

Pujol, N., Cypowyj, S., Ziegler, K., Millet, A., Astrain, A., Goncharov, A., Jin, Y., Chisholm, A.D., and Ewbank, J.J. (2008). Distinct innate immune responses to infection and wounding in the C. elegans epidermis. Curr Biol *18*, 481-89.

Raj, A., and van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. Cell *135*, 216-226.

Ruaud, A.-F., and Bessereau, J.-L. (2006). Activation of nicotinic receptors uncouples a developmental timer from the molting timer in C. elegans. Development *133*, 2211-222.

Simon, I., Barnett, J., Hannett, N., Harbison, C.T., Rinaldi, N.J., Volkert, T.L., Wyrick, J.J., Zeitlinger, J., Gifford, D.K., et al. (2001). Serial regulation of transcriptional regulators in the yeast cell cycle. Cell *106*, 697-708.

Suzuki, M., Sagoh, N., Iwasaki, H., Inoue, H., and Takahashi, K. (2004). Metalloproteases with EGF, CUB, and thrombospondin-1 domains function in molting of Caenorhabditis elegans. Biol Chem *385*, 565-68.

Tyson, J.J., Albert, R., Goldbeter, A., Ruoff, P., and Sible, J. (2008). Biological switches and clocks. J R Soc Interface *5 Suppl 1*, S1-S8.

Zugasti, O., Rajan, J., and Kuwabara, P.E. (2005). The function and expansion of the Patched- and Hedgehog-related homologs in C. elegans. Genome Res *15*, 1402-410.

Table 1: Selected gene classes enriched among periodically expressed genes

Gene class	**Definition[1]**	**Periodic Expression[2]**
Cuticular collagen	GO:0003735	72% (91/126)
Cytochrome P450 (CYP)	IPR001128	57% (27/47)
UDP-glucuronosyl/glucosyl transferase (UGT)	IPR002213	54% (37/69)
Short-chain dehydrogenase/reductase	WB: *dhs* class	40% (12/30)
Hedgehog receptor	GO:0008158	70% (19/27)
Hh-related genes:		
Warthog (*wrt*)	WB: *wrt* class	78% (7/9)
Ground-like (*grl)*	WB: *grl* class	83% (15/18)
Groundhog (*grd*)	WB: *grd* class	100% (13/13)
Quahog (*qua*)	WB: *qua* class	100% (1/1)
Hog only (*hog*)	WB: *hog* class	100% (1/1)
Metallopeptidase M14	GO:0004181	77% (10/13)
Astacin/peptidase M12	IPR001506	75% (24/32)
Peptidase C1A, papain C-terminal (Peptidase_C1A_C)	IPR000668	52% (11/21)

[1] GO: Gene ontology; IPR: Interpro; WB: Wormbase

[2] Percentage of detectably expressed genes. Numbers in brackets: number of genes with oscillatory expression/number of genes detectably expressed

Figure Legends

Figure 1: The expression of thousands of genes oscillates at an eight hour-period.

(A) Cross-correlation plot of gene expression patterns obtained by sequencing of mRNA sampled from *C. elegans* grown for the indicated time from L1 at 25°C.

(B) Proportion of variance of gene expression profiles in (A) explained by 16 principle components (PCs). *Inset*, PC1 through PC3 changes (loadings) over time.

(C) Plotting of oscillation amplitude, derived by cosine wave fitting, over PC1 reveals three separable classes of genes the expression of which is 'flat' (black), rises over time (green), or oscillates (red).

(D) A radar chart plotting oscillation amplitude over the phase of peak expression reveals that periodic gene expression occurs in all phases.

(E) Heat map showing gene expression changes of genes assigned to classes as defined in (C). 'Oscillating' genes were sorted by phase prior to plotting and assigned to indicated 45° bins (equaling 1h of development in this time course where one period corresponds to 8 h).

(F) Genes from each bin were analyzed by RT-qPCR on RNA collected from a separate time course experiment. Expression patterns from the mRNA sequencing and the RT-qPCR experiment are overlaid.

Note that amplitudes in (C), (D) are in $\log_2$ and, by definition, correspond to only half the change from peak to trough.

Figure 2: Periodic gene expression is robust and phase-locked but not temperature-compensated.

(A) Cross-correlation plot of gene expression patterns obtained by sequencing of mRNA sampled from *C. elegans* synchronously released from dauer and grown for the indicated time at 25°C.

(B) Heat maps comparing side-by-side changes in expression for each gene in each of the three categories identified in Fig. 1C for animals grown continuously after L1, or transiently arrested in dauer prior to resuming development for the time indicated.

(C) Gene expression data were collected over an extended time course (Fig. 5), and phases fitted separately for the first and the second half of the time course. A comparison of the two resulting phases in a scatter plot reveals close correlation of phases (circular correlation coefficient $r_C = 0.79$; (Fisher and Lee, 1983). The points in the top left and bottom right corners are a consequence of the circular nature of the data where 0° = 360°.

(D), (E) Expression of the genes shown in Fig. 1F was determined by RT-qPCR for animals grown for the indicated times at 15°C or 25°C, respectively. Principle component analysis demonstrates an increased period of oscillations at (D) 15°C relative to (E) 25°C. Data from Fig. 1F were used for the 25°C timecourse PCA.

Figure 3: Oscillations occurs in various somatic tissues

(A) Scatter plot comparing absolute gene expression in gonads and soma. Each dot corresponds to a gene and is color-coded according to the categories in Fig. 1C ('oscillating': red; 'rising': green; 'flat': black).

(B) Distributions of relative expression levels (comparing soma to gonad) for the categories shown in Fig. 1C. Negative values denote enriched expression in the gonad, positive values denote enriched expression in the soma. 'Oscillating' genes are shifted towards higher expression in the soma relative to the two other groups ($p<<2.2x10^{-16}$; Kolmogorov-Smirnov test).

(C) Heat maps comparing side-by-side changes in expression for each gene in each of the three categories identified in Fig. 1C for wild-type and germline-less *glp-4* mutant animals. Because *glp-4* animals are developmentally delayed, they were sampled at later time points.

(D) Analysis of promoter activity (Hunt-Newbury et al., 2007) of 'oscillating' genes reveals only modest enrichment for specific tissues within the soma.

Figure 4: Specific molecular functions and processes are enriched among periodically expressed genes.

(A), (B) GO-term enrichment of 'oscillating' genes. (B) The heatmap shows the normalized phase distributions, in 45° bins, for all the oscillating genes that belong to enriched GO terms. Two normalization steps (Supplemental Experimental Procedures) correct for the fact that gene numbers vary across phase bins (Fig. 1E) and GO terms (Fig. 4A). (MF) = Molecular Function, (CC) = Cellular Compartmentalization, (BP) = Biological Process.

(C)-(F) 'Oscillating' genes coding for proteins of the indicated families were plotted by phase. The processes and/or gene families selected drive the following GO-term enrichments: (C) "Structural constituent of the cuticle"; (D) “Metallopeptidase acitivity”, "Metallocarboxypeptidase activity", "Metalloendopeptidase activity", and "Proteolysis"; (E) “Electron carrier activity”, “Iron ion binding”, “Heme binding”, and “Oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen" (CYPs) and “Transferase activity, transferring hexosyl groups” (UGTs); and (F) "Hedgehog receptor activity"(F).

Figure 5: Rhythmic mRNA accumulation causes periodic translation.
Ribosome profiling was performed to examine the translational status of expressed mRNAs and compared to mRNA expression analysis for ‘oscillating’ genes. The timecouse was performed at two-hour intervals, but the data was resampled at a one hour grid through spline interpolation for the heatmap to permit better comparison with the other datasets. ‘Translational efficiency’ depicts the residual changes observed in the ribosome profiling data after subtracting mRNA expression data.
See also Figure S1

Figure 6: Oscillations are driven by periodic transcription
(A) Heat maps comparing expression changes of ‘oscillating’ genes at pre-mRNA levels (‘RiboZero intronic’; obtained by sequencing of rRNA-depleted RNA and

counting intronic reads) and mature mRNA levels ('Poly(A) exonic'; obtained by sequencing polyadenylated mRNA and counting exonic reads; Fig. 1E). Genes for which sufficient pre-mRNA reads were detectable are shown. The same RNA samples were utilized for both sequencing reactions.

(B) Following cosine wave fitting of the data shown in Fig. 6A, amplitudes derived for mature mRNA and pre-mRNA were plotted against one another.

(C) Genes were plotted by chromosomal location and phase.

See also Figure S2.

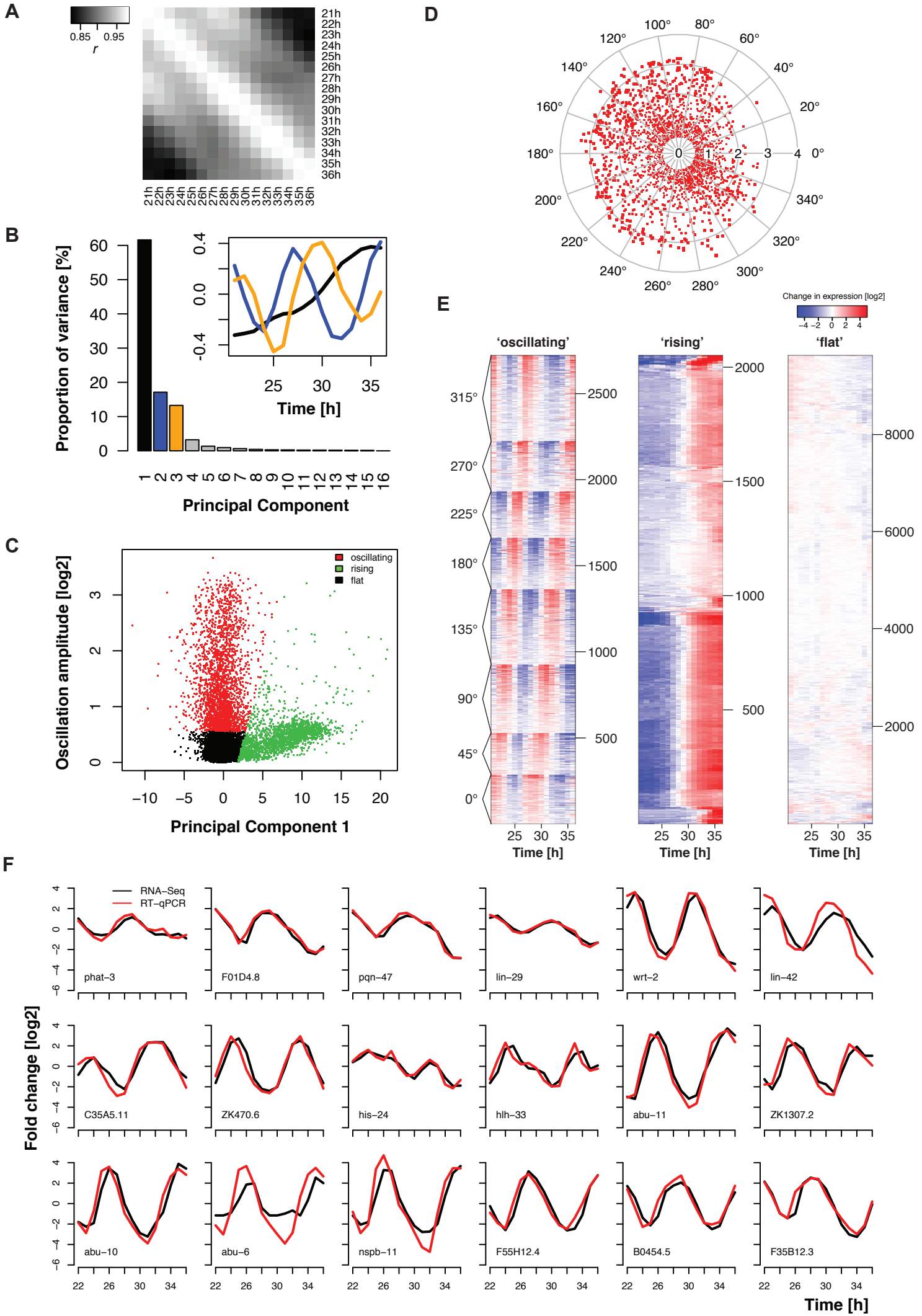

A
0.85
0.95
r
21h
22h
23h
24h
25h
26h
27h
28h
29h
30h
31h
32h
33h
34h
35h
36h
B
Proportion of variance [%]
Principal Component
Time [h]
C
Oscillation amplitude [log2]
Principal Component 1
oscillating
rising
flat
D
E
Change in expression [log2]
'oscillating'
'rising'
'flat'
Time [h]
F
RNA−Seq
RT−qPCR
Fold change [log2]
phat−3
F01D4.8
pqn−47
lin−29
wrt−2
lin−42
C35A5.11
ZK470.6
his−24
hlh−33
abu−11
ZK1307.2
abu−10
abu−6
nspb−11
F55H12.4
B0454.5
F35B12.3
Time [h]

Figure 1

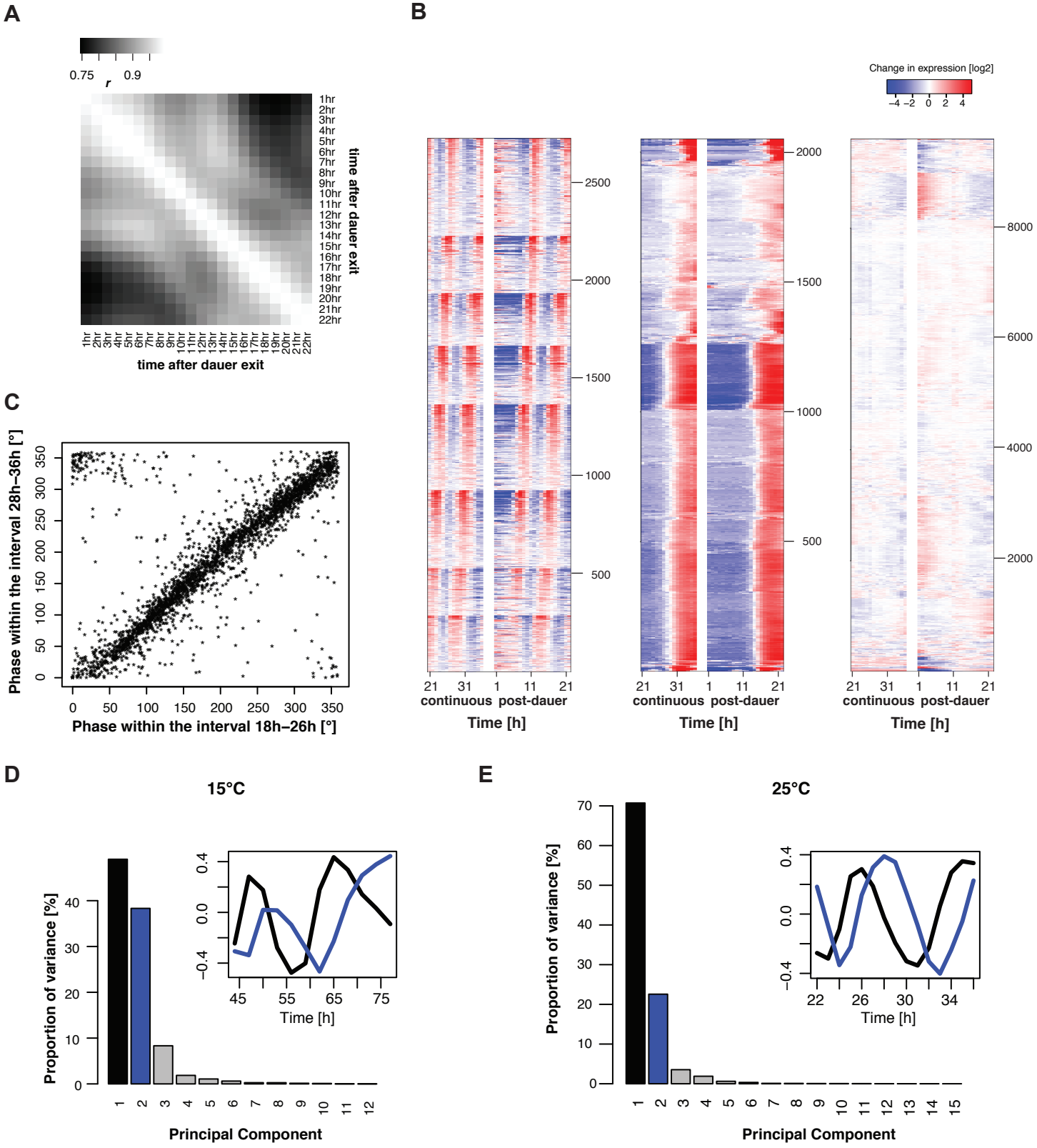

A
0.75
r
0.9
time after dauer exit
time after dauer exit
B
Change in expression [log2]
−4 −2 0 2 4
continuous post-dauer
Time [h]
C
Phase within the interval 28h–36h [°]
Phase within the interval 18h–26h [°]
D
15°C
Proportion of variance [%]
Principal Component
Time [h]
E
25°C
Proportion of variance [%]
Principal Component
Time [h]

Figure 2

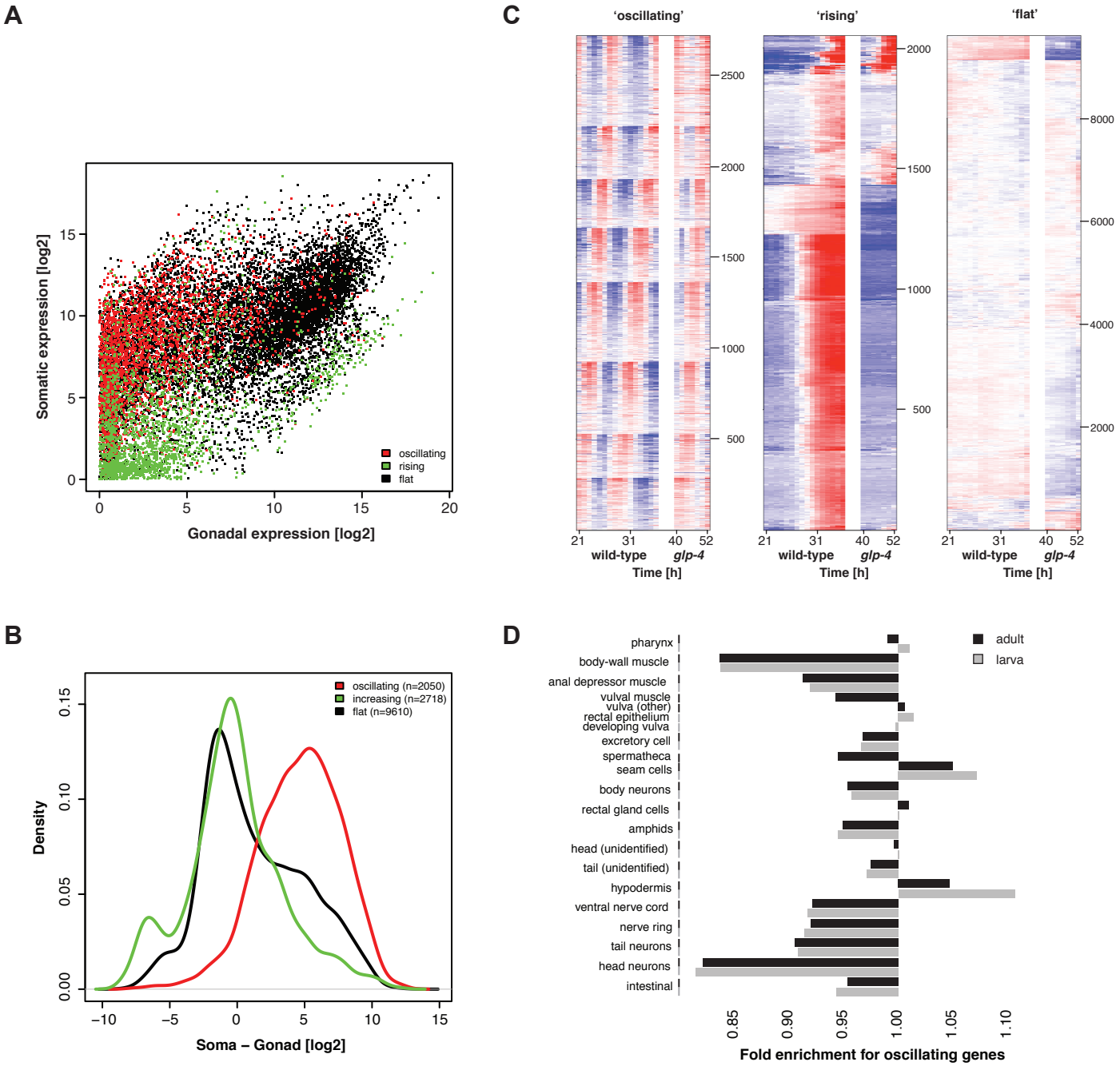

A
Somatic expression [log2]
Gonadal expression [log2]
oscillating
rising
flat
B
Density
Soma – Gonad [log2]
oscillating (n=2050)
increasing (n=2718)
flat (n=9610)
C
'oscillating'
'rising'
'flat'
wild-type
glp-4
Time [h]
D
pharynx
body-wall muscle
anal depressor muscle
vulval muscle
vulva (other)
rectal epithelium
developing vulva
excretory cell
spermatheca
seam cells
body neurons
rectal gland cells
amphids
head (unidentified)
tail (unidentified)
hypodermis
ventral nerve cord
nerve ring
tail neurons
head neurons
intestinal
adult
larva
Fold enrichment for oscillating genes

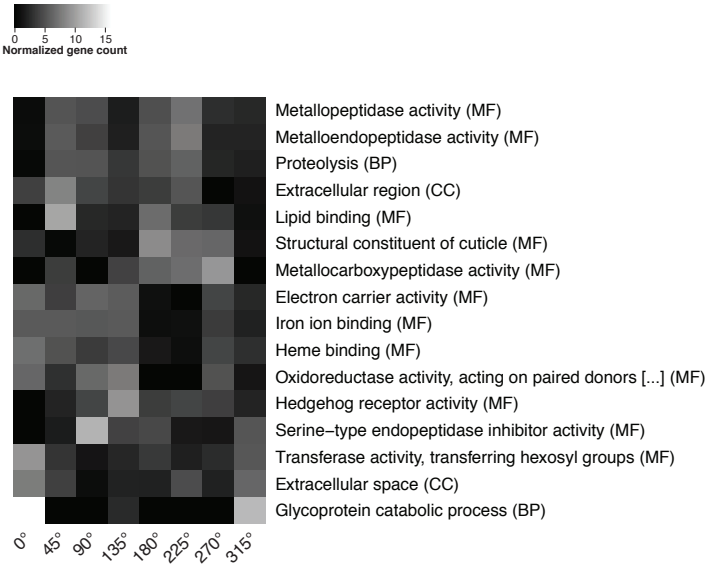
Figure 3

A

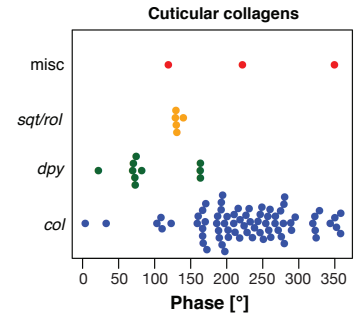
GO term	periodic expression[1]	p-value	fold enrichment
Structural constituent of cuticle (MF)	91/126	4.5e-58	3.82
Proteolysis (BP)	85/244	1.7e-22	1.84
Metallopeptidase activity (MF)	46/94	1.1e-19	2.59
Metalloendopeptidase activity (MF)	42/93	1.8e-16	2.39
Transferase activity, transferring hexosyl groups (MF)	38/82	1.7e-15	2.45
Serine-type endopeptidase inhibitor activity (MF)	26/43	1.2e-14	3.2
Oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen (MF)	29/56	9.7e-14	2.74
Heme binding (MF)	41/109	7.5e-13	1.99
Hedgehog receptor activity (MF)	19/27	8.4e-13	3.72
Extracellular space (CC)	32/72	1.2e-12	2.35
Electron carrier activity (MF)	34/91	8.2e-11	1.98
Iron ion binding (MF)	38/111	1.3e-10	1.81
Extracellular region (CC)	33/88	1.4e-10	1.98
Glycoprotein catabolic process (BP)	10/11	3.2e-09	4.81
Metallocarboxypeptidase activity (MF)	10/13	6.8e-08	4.07
Lipid binding (MF)	18/40	8e-08	2.38

[1] periodically expressed/all expressed genes in category

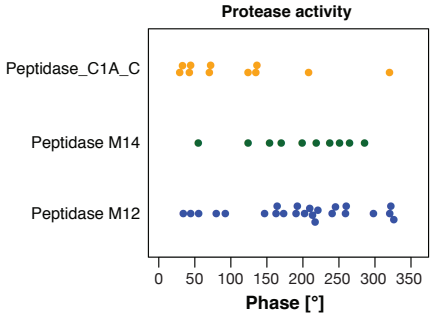
B



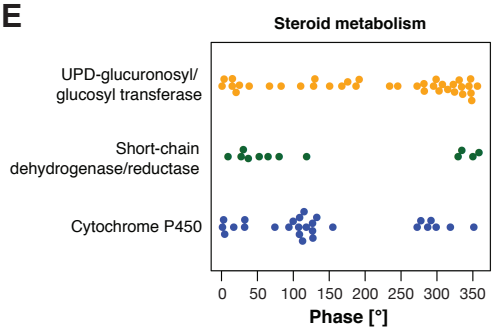
C



D



E



F

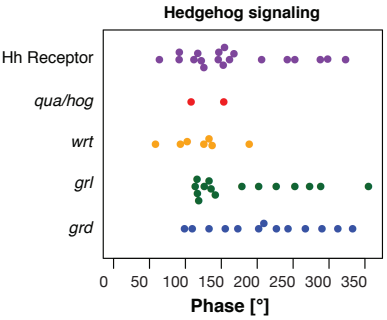


Figure 4

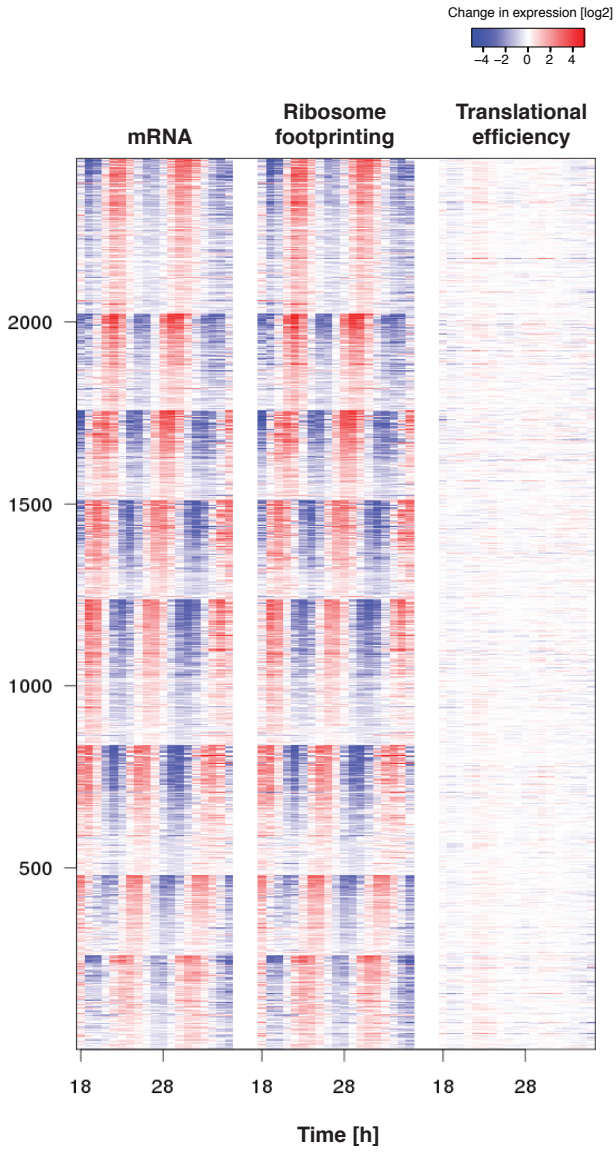
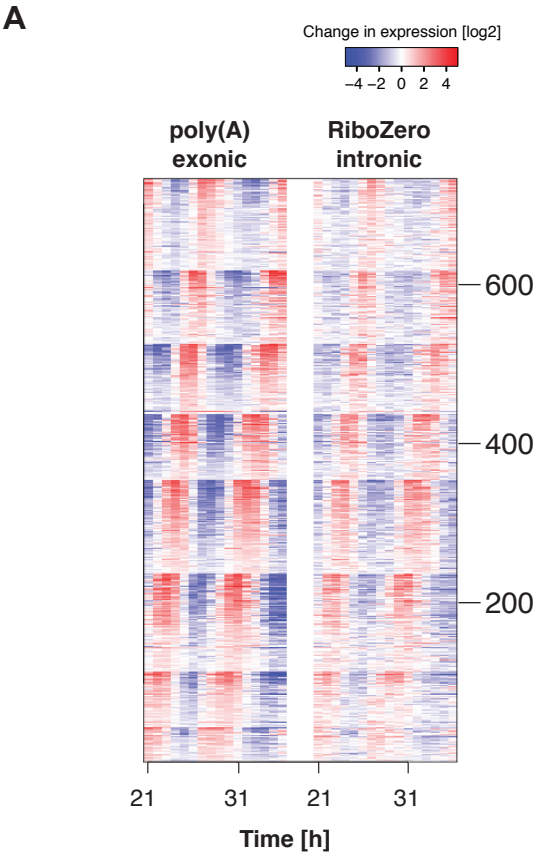
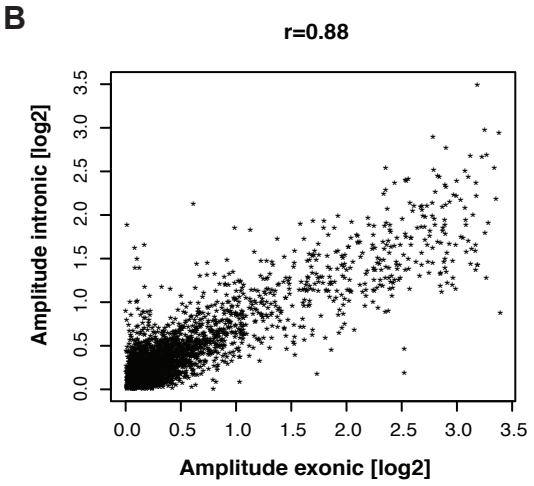

Change in expression [log2]
−4 −2 0 2 4
mRNA
Ribosome footprinting
Translational efficiency
2000
1500
1000
500
18
28
18
28
18
28
Time [h]

Figure 5


A
Change in expression [log2]
−4 −2 0 2 4
poly(A)
exonic
RiboZero
intronic
600
400
200
21
31
21
31
Time [h]


B
r=0.88
Amplitude intronic [log2]
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5
Amplitude exonic [log2]

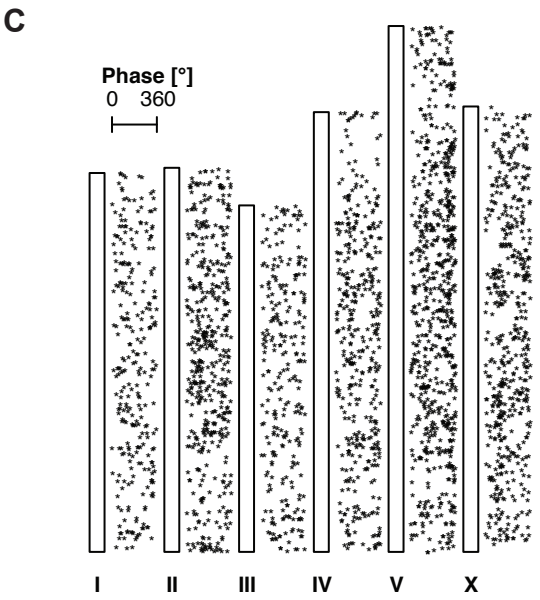

C
Phase [°]
0 360
I
II
III
IV
V
X

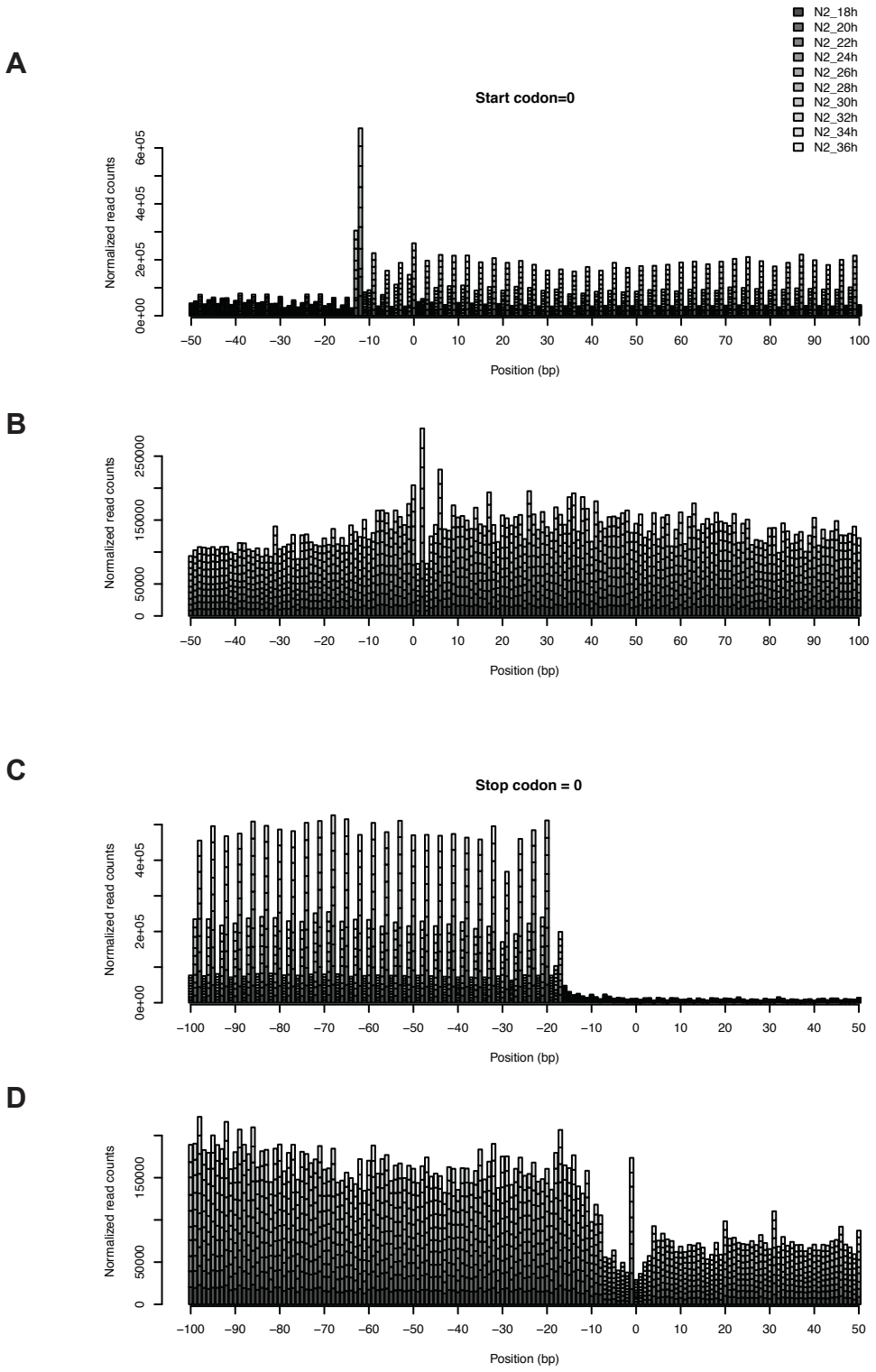
Figure 6

2.2 Supplemental materials and methods

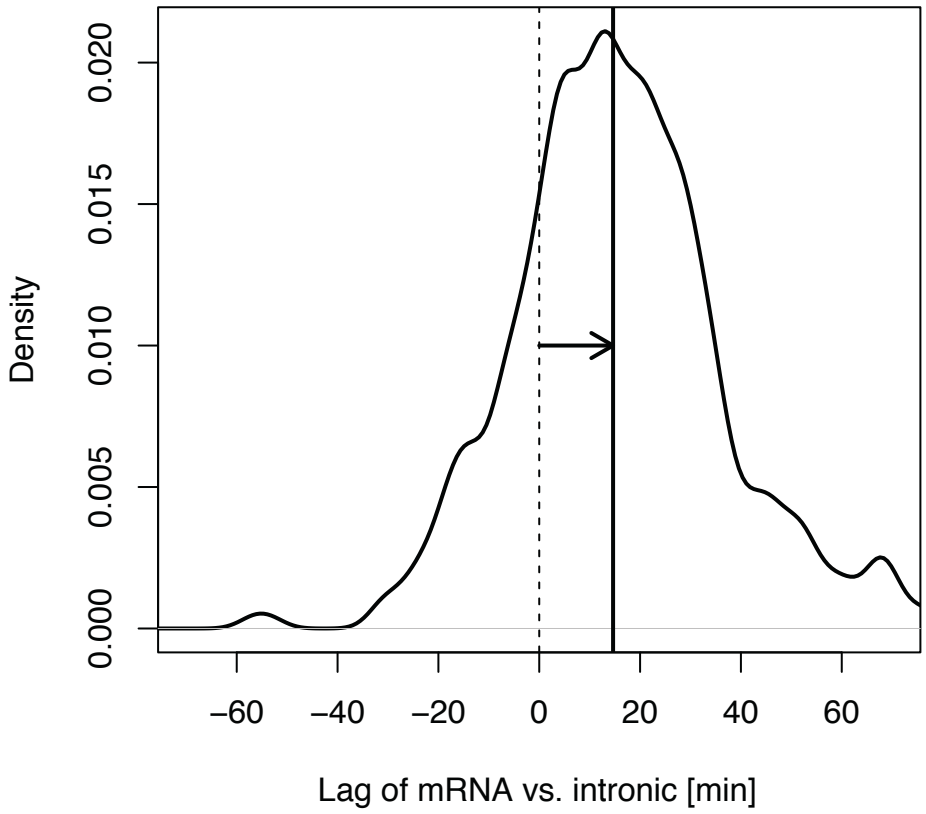
In the print version: Reprinted from Molecular Cell, vol. 53, Hendriks GJ, Gaidatzis D, Aeschimann F, Großhans H, Extensive Oscillatory Gene Expression during *C. elegans* Larval Development, 380-392, Copyright 2014, with permission from Elsevier.

http://dx.doi.org/10.1016/j.molcel.2013.12.013

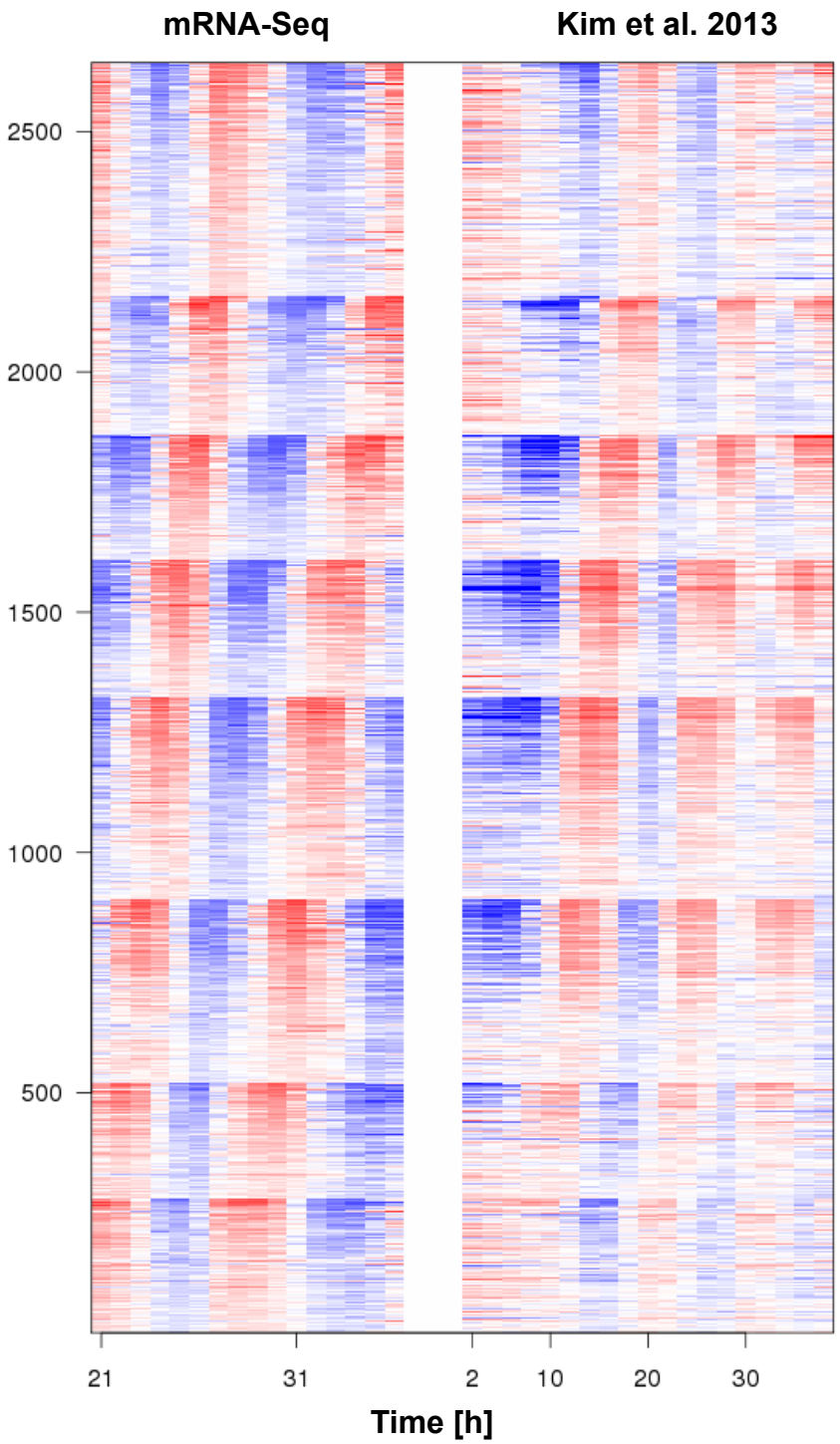
In the electronic version of this thesis, a manuscript of the supplemental materials and methods from “Extensive Oscillatory Gene Expression during *C. elegans* Larval Development” replaces the supplemental materials and methods described above. For the printed version of the paper, please see the reference above. Please note that the page numbers in the rest of the online version of this thesis have been kept identical to the print version.



Supplemental Figure S1. Related to Figure 5. Composite start codon and stop codon profiles for ribosome protected fragments (RPFs) and RNA-Seq reads.
The positions of the 5' ends of the reads are used for the counting. Only genes with an annotated 3'UTR or 5'UTR of at least 50bp are used in the respective panels. To reduce the disproportionate impact of highly expressed transcripts, we normalized the coverage of each transcript by its expression (coverageNorm=coverage/(expression+8)*avgExpression). The pseudocount of 8 was used to reduce the impact of transcripts with very low expression levels. avgExpression denotes the average transcript expression and was used as a global constant to scale back the counts. (A),(C) 3 nucleotide periodicity and depletion of RPFs from (A) 5'UTRs and (C) 3'UTRs supports their origin from mRNAs undergoing translation. (B), (D) RNA-Seq reads from the same transcripts are shown for comparison.



Supplemental Figure S2. Related to Figure 6. Pre-mRNA levels peak before mature mRNA levels.



Supplemental Figure S3. Related to Discussion. Oscillations and phase relationships are robust. Heat maps showing gene expression changes in the time course presented in Fig. 1 (*left panel*), and in Kim et al., 2013 *(doi: 10.1038/ng.2763*) (*right panel*) reveal that phase relationships among genes are maintained between the two experiments and across each time course.

Supplemental Table S1: Gene expression classes. Related to Figure 1

See separate Excel file.

Supplemental Table S2: GO-terms depleted among 'oscillating' genes. Related to Figure 4

GO-term	overlap	genes in set	p-value	fold enrichment
embryo development ending in birth or egg hatching (BP)	177	2628	8.00E-17	0.36
reproduction (BP)	109	1841	2.30E-16	0.31
receptor-mediated endocytosis (BP)	25	651	4.60E-12	0.2
hermaphrodite genitalia development (BP)	30	678	1.30E-10	0.23
translation (BP)	0	136	8.20E-08	0
structural constituent of ribosome (MF)	0	123	3.90E-07	0
nucleic acid binding (MF)	20	429	9.70E-07	0.25
ribosome (CC)	0	114	1.20E-06	0
RNA interference (BP)	1	126	4.70E-06	0.04
intracellular (CC)	16	344	1.10E-05	0.25
cytokinesis (BP)	3	143	4.20E-05	0.11
helicase activity (MF)	0	81	6.10E-05	0
ATP binding (MF)	45	660	6.30E-05	0.36
meiosis (BP)	2	116	0.00011	0.09
negative regulation of vulval development (BP)	6	177	0.00014	0.18
protein binding (MF)	9	214	2.00E-04	0.22
RNA binding (MF)	5	156	0.00024	0.17
germ cell development (BP)	2	107	0.00029	0.1
nematode larval development (BP)	147	1659	0.00053	0.47
growth (BP)	118	1368	0.00058	0.46
ATP-dependent helicase activity (MF)	0	60	0.00076	0
protein kinase activity (MF)	21	335	0.0012	0.33
nucleotide binding (MF)	17	289	0.0012	0.31
protein phosphorylation (BP)	22	343	0.0015	0.34
P granule (CC)	0	53	0.0017	0
meiotic chromosome segregation (BP)	3	105	0.0017	0.15
pronuclear migration (BP)	0	50	0.0025	0
apoptotic process (BP)	27	389	0.0026	0.37

nucleus (CC)	66	797	0.0029	0.44
GTP binding (MF)	7	149	0.0041	0.25
ubiquitin-dependent protein catabolic process (BP)	1	61	0.0059	0.09
neuron projection (CC)	0	42	0.0065	0
intracellular protein transport (BP)	1	59	0.0073	0.09
protein dephosphorylation (BP)	4	102	0.0077	0.21
cell division (BP)	5	115	0.0078	0.23
inductive cell migration (BP)	6	126	0.009	0.25
cell death (BP)	0	39	0.0094	0
positive regulation of growth rate (BP)	124	1329	0.01	0.49
DNA repair (BP)	0	38	0.011	0
protein tyrosine phosphatase activity (MF)	3	84	0.011	0.19
cell fate specification (BP)	0	37	0.012	0
mitochondrion (CC)	6	122	0.012	0.26
unfolded protein binding (MF)	0	35	0.015	0
axon (CC)	3	80	0.016	0.2
transcription, DNA-dependent (BP)	0	34	0.017	0
embryo development (BP)	9	153	0.017	0.31
neuronal cell body (CC)	2	65	0.018	0.16
transferase activity, transferring phosphorus-containing groups (MF)	32	399	0.02	0.42
mitotic spindle organization (BP)	3	72	0.031	0.22
secretion by cell (BP)	3	72	0.031	0.22
gonad development (BP)	8	127	0.043	0.33
cytosol (CC)	1	41	0.046	0.13
cell migration (BP)	6	102	0.05	0.31
striated muscle dense body (CC)	5	90	0.051	0.29
gamete generation (BP)	4	78	0.052	0.27
regulation of cell proliferation (BP)	4	78	0.052	0.27
signal transduction (BP)	8	122	0.058	0.35
response to DNA damage stimulus (BP)	4	74	0.069	0.29
morphogenesis of an epithelium (BP)	30	345	0.071	0.46
cytoplasm (CC)	51	551	0.071	0.49
regulation of meiosis (BP)	4	73	0.074	0.29
small GTPase mediated signal transduction (BP)	5	84	0.077	0.31
GTPase activity (MF)	6	92	0.094	0.34
striated muscle myosin thick filament assembly (BP)	9	124	0.096	0.38

axonal fasciculation (BP)	4	65	0.13	0.33
DNA binding (MF)	35	366	0.17	0.51
structural molecule activity (MF)	10	121	0.18	0.44
protein catabolic process (BP)	10	120	0.19	0.44
nucleoside-triphosphatase activity (MF)	9	108	0.21	0.44
determination of adult lifespan (BP)	73	710	0.22	0.54
G-protein coupled receptor signaling pathway (BP)	12	135	0.23	0.47
plasma membrane (CC)	13	137	0.31	0.5
oogenesis (BP)	16	163	0.33	0.52

Supplemental Table S3: Primer sequences. Related to Figure 1.

	Oscillating targets	
Target	Primer Left	Primer Right
Wrt-2	agccagctcaagtcgcttac	aatgcttggtgctgttgttg
nspb-11	atgttcgctaagtgcttcgccg	atggatagtatgatgggtagtacgctgg
Phat-3	tgctctcttctcggtcttga	gggtgtacacgaacgctgtt
Abu-10	ctattgtcgccctggcactttc	gtcttgagctggagcttgttgg
Abu-6	ttgggtggtggtagtagtggt	actcaacaagttcaggttcaa
C35A5.11	tggccagcgtgtaattctgta	ttgtgatcttctgtgcccgg
ZK1307.2	tcgaatgatccgcgtatccc	cagcaggaaagatacggacca
abu-11	gagcgcggaggaacacattc	tcactcactctgacaagcttga
Lin-42	tcttgttcacgtgaccttc	ggctccgtctggcatagtaa
F01D4.8	gcctccattttgattcatcgtct	ccgttttctgttgctgacga
F35B12.3	tttgagagtcgtcggtgctc	tctggaatggactcttcagaaca
ZK470.6	tcgttcgagccagctacttc	aagaagtgcctgctcgttttc
B0454.5	gccttgacgaaatcttcattcga	gttgcccgtcacattcttcg
F55H12.4	tcgagcatgcattctgaaggt	gcatgttggcaaaggaacca
His-24	aatgtcatccagatcaatgctcatctcc	ttctctggcacacggaaacgtc
hlh-33	gctgcgaagagtggatcaga	gcagcggtggtttataatcact
Lin-29	ccgacgagtacgaagaatgg	gtgattgtgggttgaacacg
Pqn-47	gaccagcgttactgtgtgga	gtaccggtgattcgctttgt

	Reference	
Target	Primer Left	Primer Right
Act-1	gttgcccagaggctatgttc	caagagcggtgatttccttc

Supplemental Experimental Procedures

Preparation of dauers

Dauer animals were obtained through a previously established protocol (Sinha et al., 2012) that we used with minor modifications. Briefly, 30,000 N2 larvae were added to 300 ml of S-medium supplemented with 1x Antibiotic-Antimycotic (Life Technologies; #15240-062) and grown at room-temperature. Five times 0.5 grams of OP50 bacterial pellet were added during the first 6 days of the culture. Dauer larvae were typically harvested from cleared media after 10-12 days. Visual inspection of the cultures showed that >90% of all worms were dauers. The worms were washed and incubated, while rolling, with 1% SDS for 15 minutes, followed by two washes with 0.1 M NaCl and floatation on 35% Ficoll-400 for 15 minutes at 100g. The yellow band that formed during the flotation was transferred to a new 15ml conical tube and washed three times with 0.1 M NaCl. Finally, the worms were precipitated through 15% Ficoll-400 to remove any remaining debris and carcasses and washed three times with M9. The resulting culture was visually inspected and the worms were left in M9 overnight, before they were plated on 2% NGM plates (approximately 1500 worms per plate) with OP50 and left to develop for the desired amount of time.

Preparation of L1 larvae and time courses

Gravid adults (N2 or *glp-4(bn2)*) were bleached and the eggs recovered left to hatch overnight at room-temperature in M9. The hatched L1 larvae were plated on 2% NGM plates (approximately 1500 worms per plate) with OP50 and left to develop for the desired amount of time at the appropriate temperature. Note that N2 samples at 25°C were named 21h through 36h although the actual harvesting occurred between 22h and 37h. This was done to account for a slight developmental delay in this relative to other time courses under the same conditions (data not shown). The time that is mentioned in the sample name more closely resembles the actual developmental time it takes to reach this point in development. For *glp-4*, worms were harvested bi-hourly for RNA isolation between 40 and 52 hours of development because a pilot experiment had revealed a strong developmental delay of the mutant animals (data not shown).

Harvesting of worms for RNA purification

Worms were washed off the plates with M9 and collected in 15ml conical tubes. The culture was washed three additional times with M9 to remove any remaining bacteria. Finally, the worms were pelleted and resuspended in Tri Reagent (Molecular Research Center; TR 118) and frozen in liquid nitrogen.

RNA isolation and sequencing library preparation

Samples were treated with 6 freeze-thaw cycles from liquid nitrogen to a heatblock at 42°C. Subsequently, debris was spun down and supernatant was transferred to a fresh tube. RNA isolation was performed using Tri Reagent and standard protocols. Total RNA was DNAse treated and RNA quality was assessed with an Agilent Bioanalyzer prior to library preparation. For mRNAseq we prepared the libraries with a Truseq stranded mRNA sample preparation kit (Illumina). For "RiboMinus" libraries, a Ribo-Zero Magnetic Kit (Epicentre; MRZH11124) was used to remove ribosomal RNA from total RNA samples and depletion validated through Agilent Bioanalyzer analysis. Subsequent library preparation was performed with a ScriptSeq v2 RNA-Seq library preparation kit (Epicentre). The quality of the resulting libraries was assessed with an Agilent Bioanalyzer and concentrations were measured with a Qubit fluorometer prior to pooling.

Ribosome profiling

We adapted published protocols (Ingolia et al., 2012; Bazzini et al., 2012) to perform ribosome profiling. Synchronized L1 stage larvae were placed on peptone rich plates seeded with NA22 bacteria, incubated at 25°C and samples collected every two hours between 18 – 36 hours thereafter. Depending on the time point, between 100’000 (late time points) and 200’000 (early time points) worms were harvested. After collection, the worm pellet was washed three times with M9 buffer to remove bacteria, then washed once with buffer A (20 mM Tris-HCl (pH 8.5), 140 mM KCl, 1.5 mM $MgCl_2$, 0.5 % Nonidet P40, 1 mM DTT, 0.1 mM Cycloheximide), snap-frozen in liquid

nitrogen and stored at -80 °C. Worm pellets were resuspended in 400 µl (time points 18 h, 20 h and 22 h) or 500 µl (all other time points) of cold lysis buffer (buffer A with 2 % PTE (polyoxyethylene-10-tridecylether) and 1 % DOC (sodiumdeoxycholate monohydrate)) and then crushed with mortar and pestle pre-cooled with liquid nitrogen. The thawed lysates were clarified by centrifugation (10 minutes, 10'000 g, 4 °C) and their absorbance at 260 nm measured. In a total volume of 385 µl, 110 absorbance units of lysate were mixed with lysis buffer and 2 µl of RNase I (100 Units/µl, Ambion) and incubated for 1 hour at 23 °C, 300 rpm. From the remainder of the lysates, total RNA was isolated with Tri Reagent (Molecular Research Center; TR 118) and an rRNA-depleted library for RNA sequencing prepared as described above.

To isolate monosomes, 350 µl of the digested lysates were loaded on linear sucrose gradients and centrifuged for 3 hours at 39'000 rpm, 4 °C, using a SW-40 rotor and an Optima™L-80 XP Ultracentrifuge (Beckman Coulter). Gradients were mixed with a Gradient Master (Biocomp) from 5% (w/v) and 45% (w/v) sucrose solutions containing 20 mM Tris pH 8.5, 140 mM KCl, 1.5 mM MgCl2, 1 mM DTT and 0.1 mM cycloheximide. After centrifugation, the gradients were fractionated using a Tris Pump (Teledyn ISCO), a Gradient Fractionator (BR-184-X, Brandel) and a fraction collector (FC-203B, Gilson). Absorbance profiles at 254 nm were recorded with an Econo UV monitor EM-1 (Biorad) coupled to a LabJack U6 data acquisition device using the DAQFactory-Express software. Gradients were fractionated in 24 fractions of equal volume and the RNA from fractions corresponding to the monosomal peak (i.e. fractions 13 and 14 or fractions 14 and 15) isolated with Tri Reagent (Molecular Research Center; TR 118).

The RNA from the monosomal fraction was separated using a 15% TBE-Urea Gel (Invitrogen) and the region around 28-30 nucleotides excised to isolate Ribosome protected fragments (RPFs). The gel piece was forced through a pierced small tube inside an eppendorf tube by centrifugation and RNA from the gel debris was eluted by overnight incubation in 600 µl cracking buffer (20 mM Tris-HCl (pH 7.9), 1 mM EDTA, 400 mM NH_4Acetate, 0.5 % SDS). RNA was precipitated with isopropanol at -80 °C for at least 4 hours (isopropanol precipitation). RPFs were 3' dephosphorylated with 10 Units of T4 polynucleotide kinase (NEB) in T4 PNK buffer with 40 Units of RNasin for 1 hour at 37 °C. Following isopropanol

precipitation, the RNA samples were ligated to 3' adapters according to the Illumina® TruSeq™ Small RNA Sample Preparation protocol and using the reagents of the kit, then again precipitated with isopropanol. Ligation products were 5' phosphorylated for 30 minutes at 37 °C with 15 Units of T4 polynucleotide kinase (NEB) in T4 PNK buffer, 1 mM ATP and 40 Units of RNasin. Following heat-inactivation of the enzyme for 10 minutes at 70 °C, the RNA was precipitated by isopropanol. Ligation to 5' adapters, reverse transcription, PCR amplification with barcoded primers and gel-purification of the PCR products were performed using the Illumina TruSeq Small RNA Sample Prep kit. Four different barcodes (RPIX 2, 4, 5, 6) were used.

Processing of the RNA-seq data

All the RNA-seq data (50bp read length) were mapped to the *C. elegans* genome (ce6) using the spliced alignment algorithm SpliceMap included with the R package QuasR (www.bioconductor.org/packages/2.12/bioc/html/QuasR.html) (Au et al., 2010). The command used to perform the alignments was "proj <-
qAlign("samples.txt","BSgenome.Celegans.UCSC.ce6",splicedAlignment=TRUE)". Gene expression was quantified by counting the number of reads that started within any of the exons belonging to a particular gene (WormBase, WS190). The command used to create the count table was qCount(proj,exons,orientation="same") in the case of ribosome depleted samples and qCount(proj,exons,orientation="opposite",selectReadPosition="end") in the case of the polyA selected samples. The library preparation protocol for polyA selected RNAs creates reads that correspond to the reverse complement of the original RNA and therefore we had to count the reads on the opposite strand of the actual gene. To compensate for differences in the read depths of the various libraries, we divided each sample by the total number of reads and multiplied by the average library size. To minimize the large differences in expression caused by genes with small number of counts, log2 expression levels were calculated after adding a pseudocount of 8 (y=log2(x+8). Note that we did not normalize by transcript length as we mostly performed differential analyisis during the various timecourse datasets. Moreover, by normalizing to the average library size instead of the arbitrary number of 1 million, we avoided

distortions when adding the pseudocount. Intronic expression for each gene was quantified by subtracting the reads that fall within exons from the reads that cover the whole gene body. Exon coordinates were extended by 10bp on both sides to ensure that exonic reads close to the junctions are not counted as intronic reads.

Processing of the ribosome footprinting data

The 3' adaptor (TGGAATTCTCGGGTGCCAAGG) was removed from the reads using the function preprocessReads from within the R package QuasR (default parameters). Mapping of the short fragments (about 30bp length) to the *C. elegans genome* (ce6) was performed using bowtie (Langmead et al., 2009) allowing only for uniquely mapping reads. The command used to perform the alignments was "proj <- qAlign("samples.txt","BSgenome.Celegans.UCSC.ce6"). Gene expression quantification was performed analogous to the case of the RNA-seq data.

Principal Component Analysis

After mean-normalization of the log2 gene expression levels, we performed PCA using the function princomp in R (default parameters). The loadings corresponding to the second and third principal component (PC) appeared to be sinusoidal waves of the same period of roughly eight hours. We noticed that the phase difference of the two PCs appeared shifted by approximately two hours, i.e. a quarter of the period. Hence, combinations of PC2 and PC3 can represent sinusoidal waves with any phase angle, because the phase difference of PC2 and PC3 is the same as the phase difference between cos(x) and sin(x). To be precise, since $C*\cos(\omega t + \varphi) = A*\cos(\omega t) - B*\sin(\omega t)$ with $A = C*\cos(\varphi)$ and $B = C*\sin(\varphi)$, it follows that a weighted combination of PC2 and PC3 can represent expression patterns with arbitrary phases. PCA revealed this relationship without any prior trigonometric knowledge.

Cosine curve fitting

For each gene we fitted a separate cosine curve $y=C*\cos(\omega t + \varphi)$ with a known period of 8 hours ($\omega = 2*p/8$) and unknown variables C and

φ. Since a cosine curve with an arbitrary amplitude C and angle φ can be represented as a weighted sum of a cosine and a sine function with no phase ($C*\cos(\omega t + \varphi) = A*\cos(\omega t) - B*\sin(\omega t)$ with $A = C*\cos(\varphi)$ and $B = C*\sin(\varphi)$) we performed the fit using a linear regression including the two components $\cos(\omega t)$ and $-\sin(\omega t)$ as regressors. Principal Component Analysis (see previous section) indicated that a large proportion of the variance in the data is explained by the non-periodic first principal component. We therefore also included it as a separate regressor during the fit. Given 16 datapoints, for each regression we obtained three coefficients. A and B, which represent the amplitude and the phase of the oscillatory component, as well as the contribution from the first principal component PC1. Based on the scatterplot comparing PC1 to the oscillation amplitude=$\sqrt{(A^2 + B^2)}$ we classified the genes into three categories, 'oscillating', 'increasing' and 'flat'. The necessary cutoffs were largely dictated by the structure of the plot. The precise locations of the cutoffs were optimized manually in order to reduce false positives after inspecting the resulting expression heatmaps for the three classes. The following three lines of R code were used to perform the classification: increasing <- 2*amplitude-PC1 < -1.7; oscillating <- !increasing & (amplitude > 0.55); flat <- !increasing & !oscillating; Note that the amplitude of a sinusoidal wave corresponds to only half the fold change between trough and peak.

GO enrichment analysis

GO annotations for *C. elegans* were downloaded from http://www.geneontology.org/gene-associations/gene_association.wb.gz (06-Jun-2013) and combined with the expression data using the WormBase gene identifier. Overrepresented GO terms in the set of oscillating genes were determined by calculating the fold enrichment of the number of overlapping genes compared to what to expect by change given the number of genes in a particular GO term and the number of oscillating genes with only expressed genes considered. To minimize the large enrichments that would otherwise be caused by GO terms with small number of genes, we added a pseudocount of 12 before calculating the actual ratio. For example in the case of the GO terms "structural constituent of cuticle", there were 91 genes in the overlap while one would only expect 126*2718/14378=23.8 by chance. Thus the pseudocount-corrected enrichment was

(91+12)/(23.8+12)=2.87. This quantity was calculated for all the GO terms and used to select the most enriched ones based on a cutoff of 1.5 fold. This approach can be seen as a compromise between pure enrichment calculations, which can create artificially high enrichments for terms with very few members, and hypergeometric tests, which can create very low p-values for terms with a large number of members even when the actual enrichments are very low.

To display the normalized phase distributions for all the oscillating genes that belong to enriched GO terms in a heatmap, we binned the phases between 0 and 360 degrees into 8 equally sized intervals. For each pathway, we counted the number of genes that fell into each of the 8 phase bins. This resulted in a table with rows corresponding to pathways and columns corresponding to the phase bins. Then we performed two steps of normalization. In the first we corrected for the fact that not all phases (over all the pathways) contain the same number of genes. We thus divided each column by the total number of genes in that column and multiplied by the average number of genes per column. In the second normalization step, we converted the counts in the table into densities to account for the fact that not all pathways have the same number of genes. Therefore we divided each row in the table by the total number of genes in that row and multiplied by the average number of genes per row.

Tissue enrichment analysis

To determine whether periodic gene expression occurred preferentially in the soma or the germline, we examined mRNA sequencing data obtained for gonads dissected out of wild-type young adult animals and for germline-less *glp-4(bn2)* mutant young adult animals, kindly provided by Dr. Rafal Ciosk (C. Scheckel, D.G., and R. Ciosk, unpublished data). To gain more detailed insight into the spatial expression of the oscillating genes, we downloaded data from the *C. elegans* promoter::GFP fusions database http://gfpweb.aecom.yu.edu (04-Mar-2013). For each strain, a list of cell types is provided in which the construct is expressed. We converted that data into a rectangular matrix where rows correspond to strains, columns correspond to cell types and expression is denoted by either 0 or 1 within the matrix. The number of expressed genes in each cell type varied strongly. We thus only considered cell types with at least 100 expressed genes for further analysis. The

strain identifiers were mapped to Wormbase gene identifiers using mappings from WormMart.

RT-qPCR-based validation of oscillations

Candidate targets for RT-qPCR validation of RNAseq data were chosen based on a wide distribution of their respective phases. Primers were designed to be exon-junction spanning if possible. cDNA synthesis was performed using the ImpromII Reverse Transcription System (Promega; A3800) according to the manufacturers protocol using random primers and 1 µg of total RNA per sample. qPCR was performed on a StepOnePlus realtime PCR system (Applied Biosystems) using SYBR Green Mastermix (Applied Biosystems; 4309155) and the primers listed in Supplementary Table 3. Ct values were corrected for *act-1* expression and normalized to the mean expression for every candidate. PCA on qPCR data was performed as described above.

Supplemental References

Au, K.F., Jiang, H., Lin, L., Xing, Y., and Wong, W.H. (2010). Detection of splice junctions from paired-end RNA-seq data by SpliceMap. Nucleic Acids Res *38*, 4570-78.

Bazzini, A.A., Lee, M.T., and Giraldez, A.J. (2012). Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. Science *336*, 233-37.

Ingolia, N.T., Brar, G.A., Rouskin, S., McGeachy, A.M., and Weissman, J.S. (2012). The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. Nat Protoc *7*, 1534-550.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

Sinha, A., Sommer, R.J., and Dieterich, C. (2012). Divergent gene expression in the conserved dauer stage of the nematodes Pristionchus pacificus and Caenorhabditis elegans. BMC Genomics *13*, 254.

3. Results

3.1 Oscillations occur specifically during larval development

To further characterize and investigate the robustness of oscillatory expression that occurs during development we examined gene expression during all stages of development. We collected a continuous time course from 5 hours until 48 hours after arrested L1s were placed on a medium containing food. We will refer to the time of feeding as developmental time (dt) 0 hours. The overall correlation, similarly to the shorter developmental time course (figure 2.1A), shows a pattern of oscillating correlation (data not shown). The expression patterns of oscillating genes, rising genes and flat genes are all recapitulated in the long time course (figure 3.1A-C). Oscillating genes peak once per larval stage from the L1 to L4 stages and do not show oscillatory expression once they reach adulthood. The maintenance of the oscillating pattern furthermore shows that phases are locked through larval stages.

3.2 A handful of genes shows non-typical oscillations

Oscillatory expression, visualized in the heat maps in figure 3.1A, seems highly homogeneous. To investigate whether there are genes that do not follow the typical pattern of expression that we see in expression heat maps, we examined expression profiles of all oscillating genes. We found that approximately 1.5% of genes oscillate in a non-typical fashion. Some examples of non-typical oscillating genes are shown in figure 3.1D, panels III-VI. Since our data is from a single biological replicate, the classification is rather arbitrary and likely includes many typical oscillators. Despite the low number, these non-typical oscillators may shed a light on the mechanisms that drive oscillatory gene expression. They include genes that peak only three times during development and some genes that peak five times. Interestingly, the five genes that peak five times during

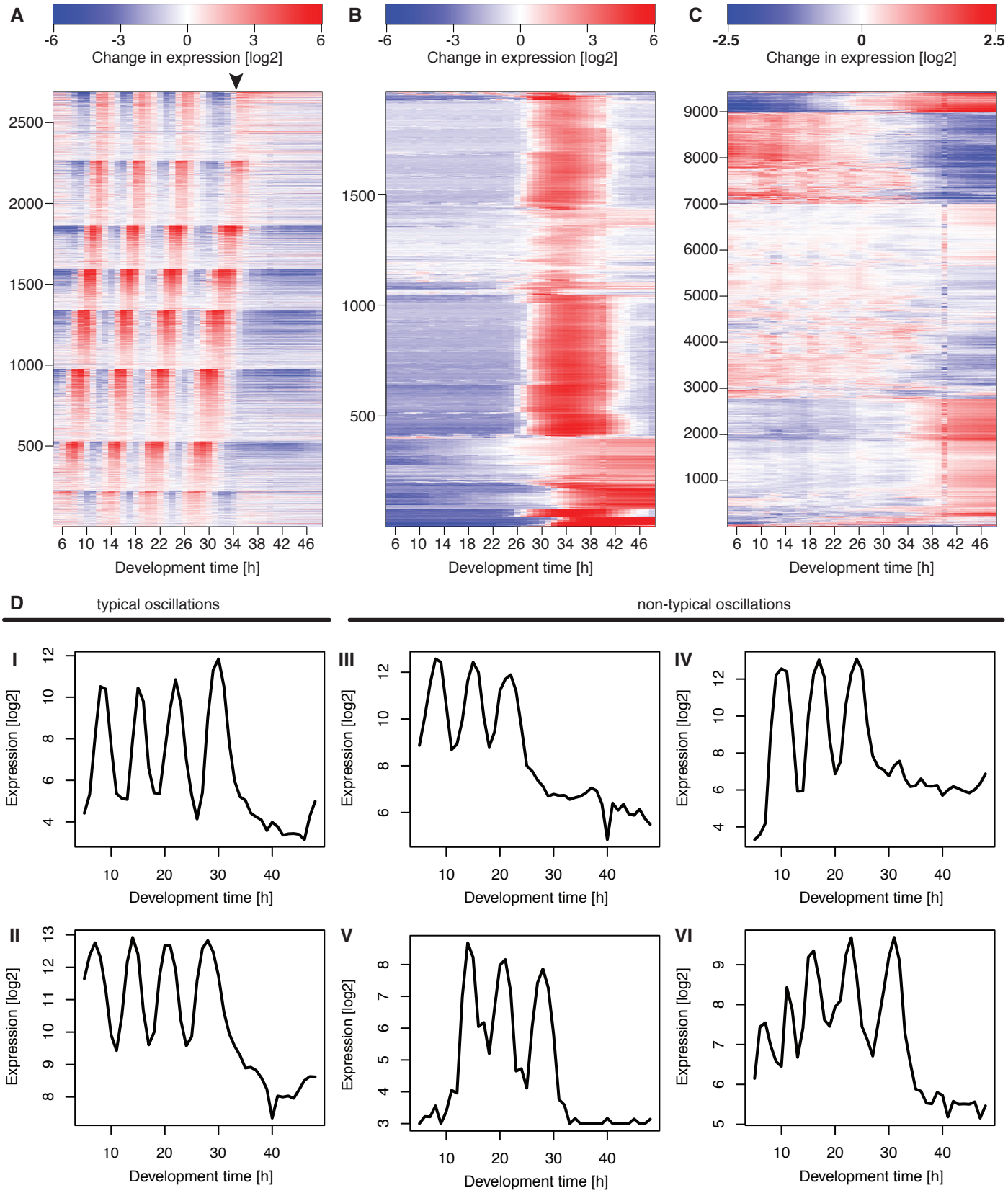


Fig 3.1 GENE EXPRESSION DURING *C. ELEGANS* LARVAL DEVELOPMENT

Full developmental expression profiles during worm development of the three classes of genes that we previously classified (see subchapter 2.2). A) The expression profiles of oscillating genes show that transcripts peak once per larval stage. Additionally, no oscillations are visible after the worms reach the adult stage (arrow-head). B) Gene expression genes previously classified as 'rising'. Expression profiles reveal that these genes are strongly downregulated during adulthood. C) The expression patterns of the set of genes that we previously classified as 'flat'. Note the different scale of the plot. Also, note that the 40 hour timepoint is an outlier and RNAseq will be repeated for this sample. D) Individual expression profiles of typical (panels I-II) and non-typical (Panels III-VI) oscillating genes.

development are all histone genes. There is no evidence to suggest that the genes that peak three times during development are related. Taken together, the presence of non-typical oscillators suggests that although the system is highly robust, specific regulation of a subset of genes can overcome the typical oscillatory expression pattern.

3.3 Oscillations arrest during L2 developmental arrest in *daf-2* mutants

Since adult animals do not show oscillatory gene expression, we were interested to further investigate the link between development and rhythmic gene expression. We investigated gene expression in worms carrying a mutation in the nematode orthologue of insulin receptor *daf-2(e1370)* (Ruaud and Bessereau, 2006). These animals show a penetrant, and synchronous, delay during L2 development when grown at 20°C. We collected worms between DT 10 hours and 56 hours and performed mRNA sequencing. As can be seen in figure 3.2A, expression profiling shows that oscillatory expression temporarily arrests and re-starts shortly before the L2-L3 molt. As with continuous development, a detailed examination of expression profiles of oscillating genes revealed a small group, approximately 0.5% of all oscillating genes, that showed minor abnormalities in their oscillating expression profiles. The most striking examples are shown in figure 3.1D, panels V-VI. However, since only one of these genes reliably maintains oscillating expression during developmental arrest, we conclude that oscillatory expression of thousands of genes during *C. elegans* larval development is tightly linked to larval development.

3.4 Developmental oscillatory gene expression is conserved in *C. briggsae* and *H. contortus*

The link between oscillatory gene expression and animal development that we identified previously prompted us to investigate the conservation of the system in the nematodes

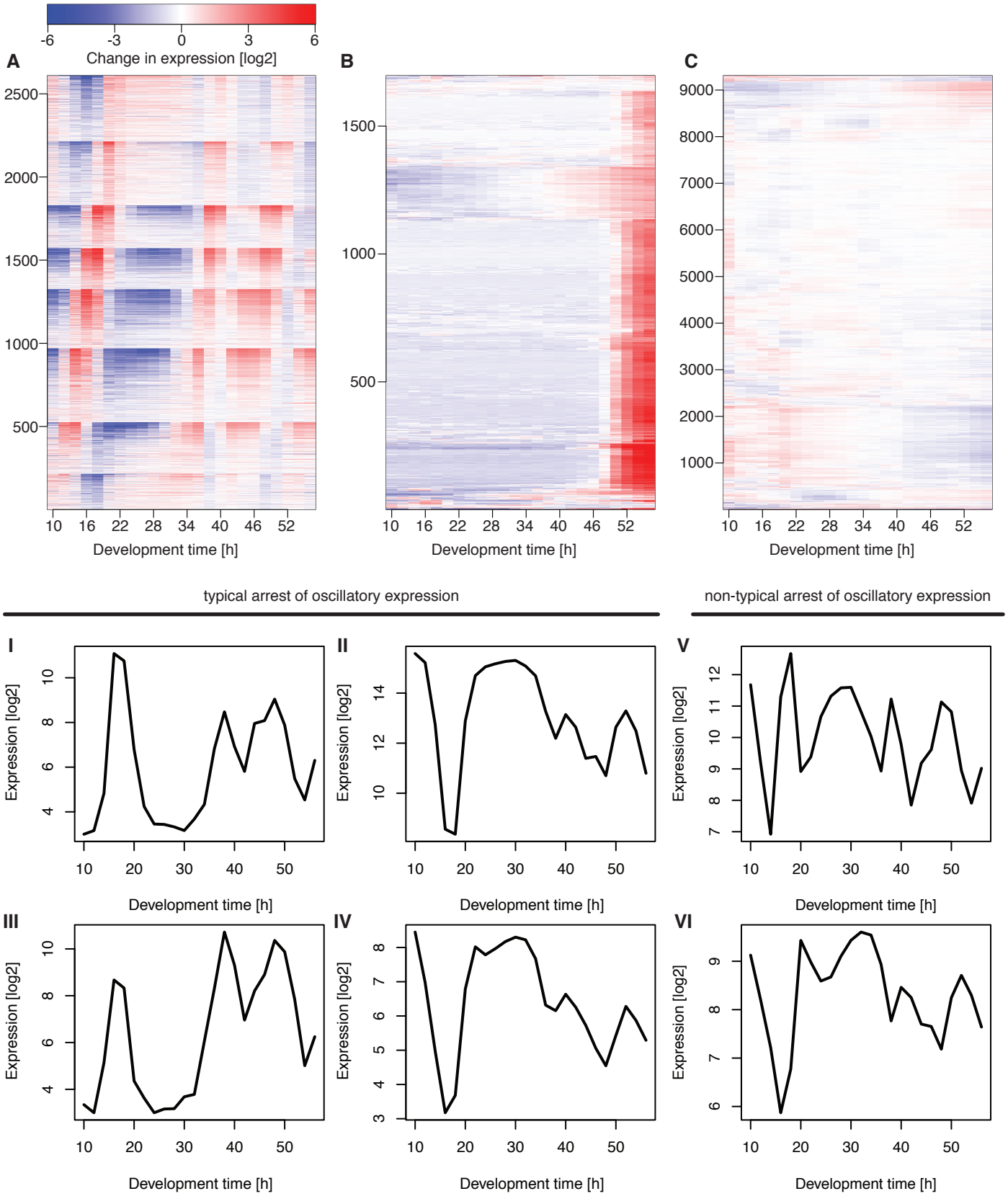


Fig 3.2 GENE EXPRESSION DURING DEVELOPMENT OF *daf-2* (*e1370*) MUTANT WORMS

Gene expression patterns of *daf-2(e1370)* worms grown at 20°C, from late L1 until L4 larval stages. A) Gene expression profiles show that oscillations slow down and freeze during the arrested L2 stage and reset upon reinitiation of development. B) Rising gene expression profiles are not affected in the mutant. C) Gene expression profiles of genes classified as 'flat'. D) Expression patterns of four typical oscillating genes (panels I-IV) and the two most striking atypical genes (panels V-VI).

Caenorhabditis briggsae and *Haemonchus contortus*. As previously described, *C. briggsae* represents the species closest in terms of evolutionary distance to *C. elegans*, while the parasitic *H. contortus* is more evolutionarily distinct. To investigate conservation in this obligate parasite we collaborated with Lucien Rufener, André Wenger, and Jacques Bouvier from Elanco Animal Health (previously Novartis Animal Health).

Time course samples were collected to span more than a single stage. Since *C. briggsae* develops slightly faster than *C. elegans* we collected a time course starting at 18 hours DT until 33 hours. For *H. contortus*, where stage lengths escalate, we collected samples to cover all free-living stages of the parasite's lifecycle. The samples were collected every 12 hours between 0 hours DT and 96 hours with added time points at 30 hours DT and 146 hours DT. Both sets of samples were subjected to poly-A selected mRNA sequencing.

We performed principal component analysis (PCA) to investigate overall trends in the *C. briggsae* and *H. contortus* datasets. The PCA of the *C. briggsae* dataset shows that, as with *C. elegans* (figure 3.3A), one rising, and two oscillating principal components can be detected (figure 3.3B). The PCA of the *H. contortus* data shows what looks like a single oscillation that is represented by principal components 1 and 3 (figure 3.3C). As in *C. elegans* and *C. briggsae* data, a single principal component also shows a rising pattern.

To investigate the possible oscillations of gene expression in more detail, we performed independent sine fitting for both the *C. briggsae* and *H. contortus* data with period-lengths of 7 hours and 60 hours respectively. These estimates were based on the principal component analyses. The method and categorization requirements that we have described previously (see subchapter 2.2) translated well to these new datasets (data not shown). The phases of genes that were classified as "oscillating" based on this fitting of the *C. briggsae* data show a relatively uniform distribution as was also seen for *C. elegans*

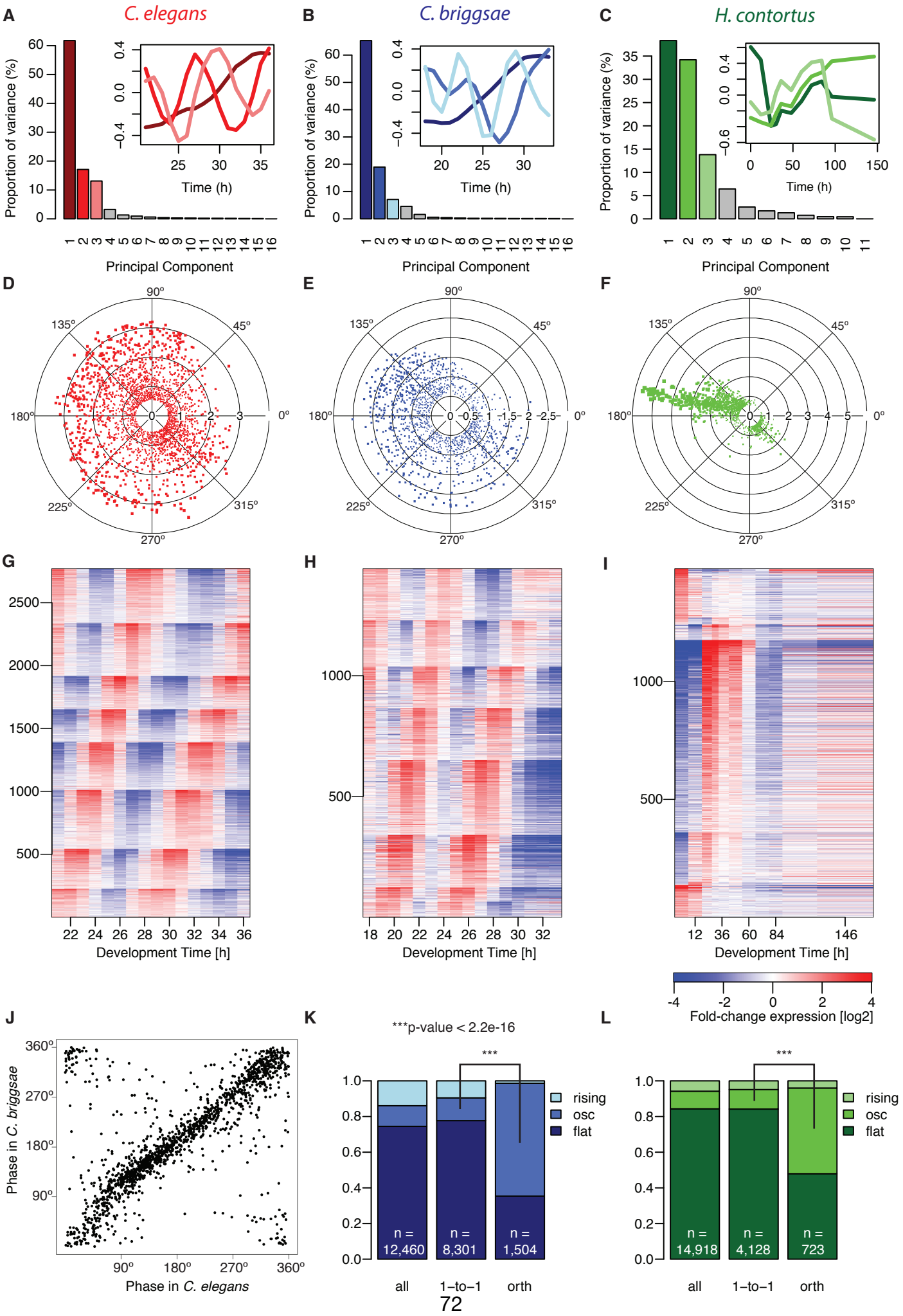

A
C. elegans
B
C. briggsae
C
H. contortus
Proportion of variance (%)
Principal Component
Time (h)
D
E
F
90°
135°
45°
180°
0°
225°
315°
270°
G
H
I
Development Time [h]
Fold-change expression [log2]
J
Phase in C. briggsae
Phase in C. elegans
K
***p-value < 2.2e-16
***
rising
osc
flat
n = 12,460
n = 8,301
n = 1,504
all
1–to–1
orth
L
n = 14,918
n = 4,128
n = 723

Fig 3.3 CONSERVATION OF OSCILLATORY GENE EXPRESSION IN NEMATODES

A) Principal component analysis (PCA) reveals a single rising and two oscillating principal components that are visible in the RNAseq data from *C. elegans*. The period of the oscillations that are visualized by principal components two and three is approximately eight hours. B) PCA performed on data collected from a *C. briggsae* time course also reveals one rising principal component and two oscillating components. The period of the oscillations of principal components two and three is approximately seven hours. C) PCA on data from a *H. contortus* developmental time course reveals a single rising principal component, and two principal components that have an oscillating appearance. The period of oscillatory gene expression that we estimated from these principal components is 60 hours. D) The phases of *C. elegans* oscillating genes are universally distributed. E) Like in *C. elegans*, the phases of *C. briggsae* genes that we classify as oscillating are relatively uniformly distributed. There does seem to be a depletion of genes with phases between 0° and 90°. F) Unlike *C. elegans* and *C. briggsae*, oscillating genes in *H. contortus* do not show a uniform distribution of phases. G) Expression profiles of oscillating genes in *C. elegans*. H) Expression profiles of genes that oscillate in *C. briggsae* confirm the uniform distribution of phases. I) The expression profiles of *H. contortus* genes confirm the discrete phase of oscillating genes. J) Phase-correlation between oscillating genes in *C. elegans* and all their 1-1 orthologues in *C. briggsae*. This also includes the genes that we did not classify as oscillating in *C. briggsae*. The phases correlate well and the circular correlation score is 0.6351. K) Among all *C. briggsae* genes that have a 1-1 orthologue in *C. elegans*, approximately 13% oscillate. Of the *C. briggsae* genes that have an oscillating 1-1 orthologue in *C. elegans*, approximately approximately 63% oscillate. This enrichment is highly significant ($p < 2.2e-16$, binomial test). L) Among all *H. contortus* genes that have a 1-1 orthologue in *C. elegans*, approximately 11% oscillate. Of the *H. contortus* genes that have an oscillating 1-1 orthologue in *C. elegans*, approximately 48% oscillate. This enrichment is highly significant ($p < 2.2e-16$, binomial test).

(figures 3.3D and 3.3E). The *H. contortus* data on the other hand suggests that most oscillating transcripts oscillate in the same phase (figure 3.3F). These findings are corroborated by the individual expression profiles of the genes that were classified as oscillating in either *C. briggsae* or *H. contortus* (figures 3.3G, 3.3H, 3.3I).

Since in *C. briggsae* genes oscillate in different phases we plotted the phases of genes that oscillate in *C. elegans* and compared them to the predicted phases of all 1-to-1 orthologues of these genes in *C. briggsae* (figure 3.3J). This illustrates the remarkable extent of phase-conservation between *C. elegans* and *C. briggsae*. Finally, we compared the percentage of oscillating genes among all genes to the percentage present among orthologues of genes that oscillate in *C. elegans*. For both *C. briggsae* and *H. contortus* we see a strong enrichment of oscillating genes among orthologues of *C. elegans* oscillating genes, suggesting that the system is conserved in both *C. briggsae* and *H. contortus* (figures 3.3K and 3.3L). A little less than 40% of the orthologues of genes that oscillate in *C. elegans* are not categorized as oscillating in *C. briggsae* (figure 3.3K). However, the predicted amplitudes of genes that are not categorized as oscillating still

correlate well with the phases in *C. elegans* (figure 3.3J). This argues that we underestimate the percentage of oscillating genes in *C. briggsae*, and the enrichment of oscillating genes is even greater.

In conclusion, we see extensive conservation of oscillatory gene expression in both *C. briggsae* and *H. contortus*. The phases of oscillations in *C. elegans* are very well conserved in *C. briggsae*. In *H. contortus* we see changes in period and a singular phase of oscillatory gene expression and importantly, we only detect a single wave of expression rather than a sustained oscillation. Nevertheless, we show highly significant enrichment of orthologues of genes that oscillate in *C. elegans* among the genes showing oscillating expression patterns in *H. contortus*. The strong conservation of the system argues that it is important in nematode physiology.

3.5 Oscillatory gene expression is not limited to coding genes

As we have reported, oscillatory expression of coding genes is widespread during *C. elegans* larval development. To examine whether non-coding genes would also show oscillating gene expression profiles, we examined small RNA expression profiles during development. The small RNA data additionally allowed us to investigate the hypothesis that was put forth by Kim *et al* (2013). that suggests that oscillations of miRNAs would functionally dampen oscillations of mRNAs (discussed below). Our results show that oscillatory expression also affects miRNA expression level. The miRNAs that are plotted in figure 3.4A are both the guide (green sidebar) and passenger strands (red sidebar) of any miRNAs that show oscillatory expression of either strand. Interestingly, many of the oscillating miRNA guide and passenger strands belong to the miR-35 family of miRNAs that includes 7 miRNAs. All of the members of this family show oscillatory expression of both the guide and passenger strands, with varying amplitudes and robustness. The characteristic of the oscillations of the miRNA guide and passgenger strands vary widely

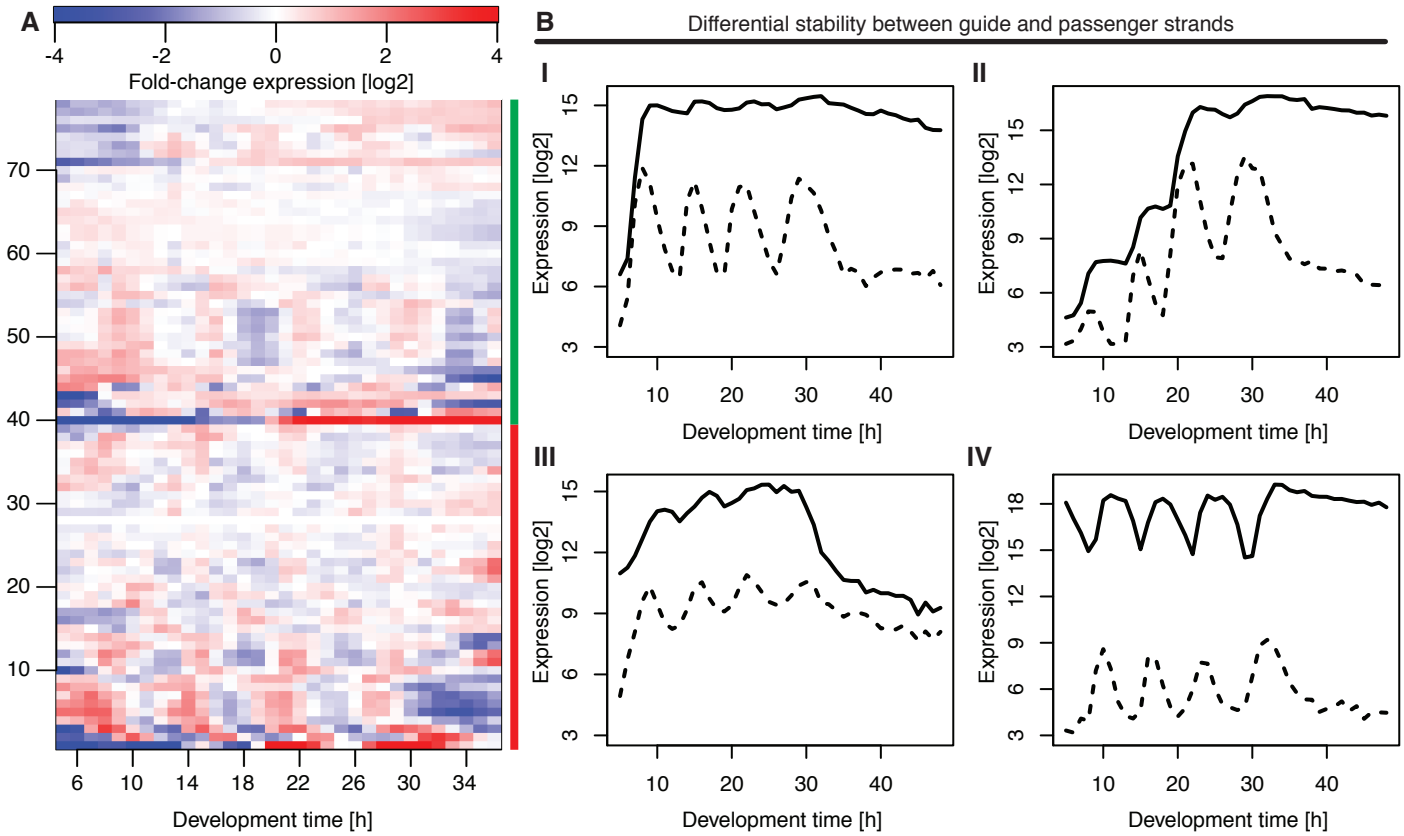


Fig 3.4 OSCILLATIONS OF MIRNAS DURING DEVELOPMENT

A) Oscillatory expression affects expression levels of miRNA guide (green sidebar) and passenger (red sidebar). miRNA guide and passenger strands were identified based on mean expression level in the time course. B) Expression profiles of miRNAs that show differential stability (panels I-II). The two miRNA and miRNA*s that are shown here, *lin-4* and *let-7* respectively, show differential stability between miRNA strands. Two other miRNAs, *miR-788* and *miR-235*, show regulated differential stability between guide and passenger strands (panels III-IV).

(figure 3.4B). Since miRNA guide and passenger strands are synthesized together, the difference between the two strands is likely mostly due to the difference in stability. The wide range of miRNA oscillation charactersitics shown in figure 3.4B illustrates the effect of miRNA degradation on the temporally resolved expression levels of miRNAs. Taken together, this data clearly shows that oscillating expression is not limited to coding genes and also affects miRNA expression.

4. Discussion

4.1 Thousands of genes oscillate during *C. elegans* development

Developmental expression profiles have been widely used to investigate development in model organisms. Previous transcriptome-wide expression profiling studies on both miRNA and mRNA level have been performed at a resolution of approximately 1-2 timepoints per larval stage (Kato et al., 2009; Spencer et al., 2011). We have shown that gene expression is highly dynamic even within a single larval stage. Approximately 19% of the transcriptome shows oscillatory gene expression levels and changes more than 2.1-fold. Oscillations occur only during larval development, and expression peaks once per larval stage. Our results support and are supported by independent studies that have reported evidence of oscillating gene expression in *C. elegans* larval development (Francesconi and Lehner, 2014; Grün et al., 2014; Kim et al., 2013; Snoek et al., 2014; Turek and Bringmann, 2014). Although the other studies detect pervasive and reproducible oscillatory expression, the exact number of oscillating genes differs slightly depending on the data quality and classification method (Kim et al., 2013; Turek and Bringmann, 2014). A detailed analysis indicates that thousands of genes that we classified as “flat”, actually show oscillating expression with very low amplitudes. In conclusion, oscillating gene expression is robust and highly pervasive during *C. elegans* larval development.

4.2 Larval oscillatory expression is conserved in nematodes

The conservation of oscillatory gene expression in *C. briggsae* and *C. elegans* furthermore suggests a role of importance in animal development. These findings are supported by qPCR data that demonstrates that a handful of *C. briggsae* transcripts show oscillatory expression (Grün et al., 2014). The strength of phase conservation that we detect, despite

the evolutionary distance, suggests that the phases of the oscillations may serve an important role (figure 3.3J). For cuticle synthesis for example, a series of sequential reactions process the collagens to produce the final secreted complex. The sequential expression of these enzymes and proteins may optimize complex formation and minimize energy expenditure. Genes that peak at the same time may therefore be part of the same protein complexes. Accordingly, mutations in a series of genes that are involved in cuticle formation and oscillate with highly similar phases, all result in similar rolling phenotypes (figure 2.4). These genes, despite being encoded at different loci, oscillate with highly similar phases and amplitudes. In addition to conservation in *C. briggsae*, we also see conservation of the system, although not of phases of oscillatory expression, in *H. contortus*. The period length of 60 hours, measured using principal component analysis, is higher than expected based on previous timing experiments (Julia Tietz, Lucien Rufener & Jacques Bouvier, personal communication). Surprisingly, the phases of genes that show oscillatory expression in *H. contortus* are not universally distributed. The long development time per stage may, combined with our relatively low temporal resolution, result in the detection of oscillatory expression in a single phase. When we investigate the orthologues of oscillating genes in *C. elegans* however we see clear enrichment among the genes that show these long oscillations with a single phase in H. contortus. The strong conservation in the larval development of *C. briggsae* and *H. contortus* suggests that oscillatory expression during larval development is important in nematode development.

4.3 Molecular mechanism

Oscillatory expression can be established through oscillating transcription, oscillating degradation, or a combination of the two. The finding that intron-containing transcripts (pre-mRNAs) show oscillatory expression suggests that oscillating transcription drives

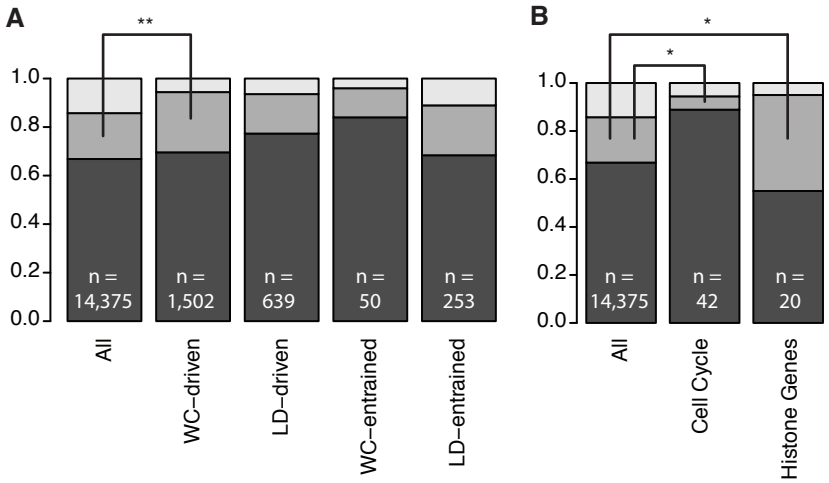


Fig 4.1 ENRICHMENT OF OSCILLATING GENES AMONG CIRCADIAN RHYTHM AND CELL CYCLE GENES

A) Little or no enrichment of oscillating genes can be found in any of the groups of transcripts that was reported by van der Linden and colleagues (van der Linden et al., 2010) to be regulated in a circadian manner. A mild, but significant enrichment was seen among the genes that are driven by warm-cold (WC) cycles (p = 1.6e-8, binomial test). No significant enrichments were seen in genes that were driven by light-dark (LD) cycles.

B) Significant depletion of oscillation genes among genes that are associated with cell divisions (p = 0.02, binomial test). Among histone genes, a significant enrichment of oscillating genes was seen (p = 0.04, binomial test).

rhythmic gene expression. To investigate this further, we created reporter constructs where the promoters of oscillating genes were used to drive expression of a fluorescent reporter (in collaboration with Yannick Hauser). The GFP reporter we used is destabilized by the addition of a PEST sequence and localized to the nucleus because it is fused to histone protein H2B. We performed microscopy and could show that GFP protein levels oscillate during larval development (data not shown). We furthermore showed that *gfp* RNA levels oscillate and recapitulate the rhythmic expression of the endogenous gene that is under control of the same oscillating promoter (YP Hauser, GJ Hendriks, H Grosshans, unpublished). This suggests that, at least in the case of our reporters, the promoter contains all the regulatory elements that are required to initiate oscillating transcription. Taken together, these results support our previous finding that oscillatory expression is driven on the transcriptional level.

4.4 Involvement of miRNAs

Kim *et al*. showed that the miRNA *lin-4* dampens oscillatory expression of *lin-14* (Kim et al., 2013). They suggest that dampening of oscillatory expression may be a general function of miRNAs during *C. elegans* development. To investigate if miRNAs are likely to perform this role in development, we investigated miRNA expression profiles. We found

that although there is a subset of miRNAs that is expressed in oscillatory fashion, approximately two-thirds of them are miRNA passenger strands (figure 3.4A). Of the miRNAs that do show oscillatory expression, the fold-changes of the miRNA are relatively minor compared to most oscillations on the mRNA level (figure 3.1A). This indicates that it is unlikely that the oscillations of a single miRNA dampen oscillations of any mRNA fully. The shared function of different oscillating miRNAs, like the miR-35 family or unrelated oscillating miRNAs, may however dampen relatively strong oscillations on the mRNA level.

miRNAs can assert their function on gene expression through either mRNA decay, or translational repression. We have previously shown that translational efficiency of oscillating mRNAs is constant over development (figure 2.5). This argues against translational inhibition of oscillating mRNA as a way of repressing oscillations on the protein level. Additionally, we found no transcripts that show oscillations on the pre-mRNA level (based on intronic reads), but do not oscillate on the mature mRNA (data not shown). This suggests that there are no miRNAs that induce degradation of an oscillating mRNA and thereby repress the oscillatory expression. We conclude that, contrary to what was described by Kim *et al*., oscillating miRNAs, in general, do not function to dampen oscillatory expression of coding transcripts.

4.5 Relation to other rhythmic phenomena

Rhythmic gene expression is a common feature of many oscillating systems. We investigate the relationship between *C. elegans* larval oscillatory expression and two of these widely-studied systems; circadian rhythm and the cell cycle.

4.5.1 Larval oscillatory expression is not related to circadian rhythm in *C. elegans*

A circadian rhythm has been described in many different species, including *C. elegans*. As described previously (see chapter 1.1.1 Circadian rhythm in *C. elegans*), the *C. elegans* circadian rhythm is less well characterized and seems less extensive than in most other organisms. One of the orthologues of the key clock-component period, *lin-42,* shows strong, and previously described, oscillating expression (Jeon et al., 1999) (Monsalve et al., 2011). This raises the question whether oscillatory gene expression during larval development is linked to circadian rhythm. Van der Linden and colleagues reported on a large number of genes that are periodically expressed (van der Linden et al., 2010). They distinguished between genes that always oscillate, only oscillate in cycling dark-light or warm-cold conditions, and genes that can be entrained to either rhythm. They only classified very few genes as always oscillating, these were therefore not included in our analysis. In the other categories, we could only detect a small but significant enrichment of oscillating genes among the set of genes that are driven by warm-cold cycles, but cannot be entrained (figure 4.1A). Furthermore, Rodriguez *et al*. showed that oscillatory expression of transcripts that are under circadian control, in Drosophila, is regulated at the post-transcriptional level (Rodriguez et al., 2013). We on the other hand have shown that oscillations occur at the pre-mRNA level and are therefore transcriptionally regulated. Additionally, we have shown previously that as opposed to circadian clocks, larval oscillatory gene expression is not temperature-compensated (figures 2.2D-E). Taken together, this indicates that the system is not related to circadian rhythm in *C. elegans*. We can however not rule out that in *C. elegans*, a small number of orthologues of circadian rhythm factors are involved in the maintenance of oscillatory expression during larval development (see below).

4.5.2 Larval oscillatory expression is not related to cell division cycles

Oscillatory gene expression has been reported in synchronously cycling *Saccharomyces cerevisiae* (Spellman et al., 1998). The number of cells that divide during nematode larval development is, however, relatively limited. Additionally, overall cell divisions throughout the worm are not synchronized and rhythmic (Sulston and Horvitz, 1977). There is however one specific tissue where cell divisions do mostly follow this rhythm; the hypodermis. This tissue is of special interest since many of the oscillating genes are expressed here (YP. Hauser, G. Brancati, GJ. Hendriks, H. Grosshans, unpublished). The hypodermis grows during larval development by the fusion of seam cells to the hypodermal syncytium. These cell fusions occur once per larval stage. The seam cells that fuse to the hypodermis are the product of asynchronous cell divisions, where one of the daughter cells remains in a stem cell-like state, while the other undergoes differentiation and fusion. In addition to these four asynchronous cell divisions, the seam cells also undergo one synchronous cell division in the early L2 stage. If cell division would induce periodic transcription, this additional, synchronous, cell division would drive five expression peaks during larval development and shorten the period length during the L1 and L2 stages. With few exceptions (see below), oscillating genes do not show five expression peaks during larval development. Accordingly, we do not see enrichment, but rather a depletion of oscillating genes among genes that are involved in cell cycle regulation (figure 4.1B). One group of genes that is involved in cell cycle regulation, and does show significant enrichment of oscillating genes (figure 4.1B), encodes histone genes. During S phase of mitosis, the coordinated DNA replication requires large numbers of histone proteins to be available (Marzluff et al., 2008). The RNA transcripts of these histone-coding genes, which are usually not poly-adenylated, accumulate during S phase, and are rapidly degraded after the conclusion of replication. Since we perform poly-A selection, we only detect low levels of histone transcripts in our data. Out of the

eight oscillating genes, five show a clear, atypical pattern with five peaks of expression (for example figure 1.3 panel VI). Detailed examination of all oscillating gene expression profiles revealed that these are the only genes that show this fifth peak. This indicates that repetitive divisions of the seam cells could cause oscillations of this subset of genes. The thousands of genes that oscillate in a typical fashion, and peak four times during development are however not periodically transcribed in response to the cell cycle.

4.6 Rhythmic development as driver or function of oscillatory expression

The large number of genes that oscillate, and the high fold-changes of gene expression that they undergo, exact a heavy energetic toll on the organism. It is unlikely that such an energetically expensive network of co-regulated expression would not play an important role in development. The findings that oscillatory expression occurs specifically during, and at all stages of larval development, support this hypothesis. This is reinforced further by the strong correlation between development and oscillating gene expression, that we see during a transient developmental arrest (figure 3.2A). Finally, the strong conservation of the network of oscillating genes to *C. briggsae* and *H. contortus*, argues that it plays an important role in development. Although these findings do not provide direct evidence that oscillations are essential for development, they imply that oscillatory gene expression and development are closely linked. The driver as well as function of the system is therefore likely to be developmental. Since we know that many oscillating genes are expressed in the hypodermis (YP Hauser, G. Brancati, GJ Hendriks, Grosshans H, unpublished), we focus on repetitive developmental events that occur in the hypodermis.

Two types of repetitive development have been previously linked to the hypodermis; hypodermal cell lineage specification and the molting cycle (Sulston and Horvitz, 1977). Below we will elaborate on the evidence that implicates these two processes with

oscillatory expression. Since the data that we have obtained is mostly correlative, it is impossible to distinguish a potential cause from a potential effect of oscillatory gene expression. We will therefore discuss both rhythmic processes as possible drivers and outputs of the oscillating expression during larval development.

4.6.1 Molting as a possible function or driver of oscillatory gene expression

The rhythmic nature of the molting process makes it a good candidate output of oscillating expression. Many of the genes that we found to be oscillating are collagens and are involved in the synthesis of the cuticle. In addition, many proteases that may be required for cuticle synthesis also oscillate. Moreover, many of these genes peak specifically during the molt, where they likely exert their function in the construction or destruction of the cuticle. Periodic expression could function to minimize energy expenditure and orchestrate the sequential steps of cuticle production by availability of enzymes and structural components of the cuticle. All things considered, oscillatory gene expression likely functions to regulate molting cycles.

Although a function in the regulation of molting is likely, it is worth noting that in addition to the oscillating collagens and proteases, there are many oscillating genes that have no known role in cuticle generation and show expression peaks outside of the molt. This suggests that oscillatory expression may have a role unrelated to molting.

The molting cycle induces a number of changes in worm behavior. One example of this is the occurrence of lethargus. During this period in the molting cycle the worms do not feed. Due to its small size and volume, changes in nutrient supply can have rapid effects on physiology throughout the animal. This change in nutrition supply might be an interesting possible driver of rhythmicity since some of the oscillating transcription factors are nuclear hormone receptors (see below). The ligands that these NHRs require to initiate transcription of their targets could be directly or indirectly provided through feeding.

NHRs are often pre-assembled on DNA and can therefore rapidly induce transcription for many targets. Repetitive molting, specifically through repetitive feeding, has the potential to cause repetitive gene expression patterns.

4.6.2 Cellular differentiation and fusion as a possible function or driver of oscillatory gene expression

During development, seam cells undergo four asymmetrical divisions and one symmetrical division in the early L2 stage. Since they divide a total of five times, cell divisions are unlikely to cause the rhythmic gene expression profiles that we observe. Cellular differentiation and fusion, however, occur only once per stage, during the asymmetrical divisions. Both differentiation and cellular fusion coincide with dramatic changes in gene expression and availability of transcription factors. Interestingly, a recent study that has examined gene expression profiles in different mammalian, models of differentiation, reported that there are sequential waves of transcriptional regulation required for development (Arner et al., 2015). These sequential waves of different genes being upregulated would correspond to different phases of *C. elegans* oscillating genes. The repetitive nature of these cellular differentiation events would finally result in an oscillating pattern. During cellular fusion, 20 (from L2-L4) seam cells fuse to the hypodermis. Before the fusion these cells undergo endoreduplication, resulting in 40 diploid genomes. When these cells fuse, transcription factor availability shifts instantly and gene expression follows. Co-regulation of transcription over these 40 genomes could cause synchronous changes in gene expression that can be detected in whole-worm expression profiling. The extremely rapid and well-synchronized expression could be induced by activation of ligand dependent NHRs that are pre-assembled on the genome. Induction by other, perhaps slower, mechanisms may result in the universal distribution of the phases of oscillating genes.

In summary, differentiation and fusion of the seam cells could be either a driver or a result of the gene expression profiles that we observe. Since we see many oscillating genes being expressed in the hypodermis we hypothesize that they are linked to either seam cell differentiation in the hypodermis, the molting cycle or perhaps connect these two modular developmental pathways. Interestingly, both of these systems have been previously described as possible modular systems, where simple reiteration of the system allows for further growth (Monsalve and Frand, 2012; Sternberg, 1991). We propose that the oscillating gene expression is part of the module of development that regulates seam cell divisions and molting and may function to synchronize the two modular systems as well as connect modular development to linear (non-modular) development.

4.7 Open questions and outlook

4.7.1 The role of transcript stability in oscillatory expression

Stability of oscillating transcripts

We have shown that transcriptional oscillations drive oscillating expression levels of thousands of genes. In addition to rapid and oscillating transcription, we see equally rapid degradation of oscillating transcripts. Since the promoter of an oscillating gene can drive expression of *gfp* in an oscillating pattern, that closely resembles that of the endogenous gene (YP Hauser, GJ Hendriks & H Grosshans, unpublished), the degradation of oscillating RNA is likely aspecific. Additionally, oscillating, but non-specific, mRNA decay would not support the universal distribution of phases of oscillating genes of *C. elegans* and *C. briggsae*. Specific degradation of transcripts could explain this, however, as mentioned previously, transcript stability does not differ between endogenous transcripts and reporters that are driven from the same promoter. Additionally, we did not detect any oscillations on the transcriptional level of genes that do not show oscillations on the mature mRNA level (data not shown). Taken together, this indicates that oscillating

destabilization does not cause oscillatory expression of any genes. Instead, oscillating transcription overcomes the extreme instability of oscillating transcripts. Whether only oscillating transcripts are degraded, or all transcripts in the cells where oscillations occur are rapidly turned over, remains unclear. The high transcript instability as well as the dynamics and possible regulation of degradation during and after larval development remains a point of interest.

Stability of miRNAs

While many, if not all, miRNA passenger strands show oscillatory expression, there are numerous guide strands that do not show this pattern. This is in line with the concept that differential Argonaute loading of miRNA strands results in differential stability of the strands (Vaucheret et al., 2004). Since we do not see any oscillating miRNA guide strands where the passenger strand does not oscillate, we propose that the passenger strands of these miRNAs are inherently unstable and can function as a measure of transcription. This is supported by qPCRs measuring the highly similar pri-miR-788 and miR-788* levels during development (data not shown). Two examples of differential stability between the guide and passenger strands are *let-7* and *lin-4*. Expression profiles of these two essential and well-studied miRNAs (figures 3.4B I & 3.4B II) show clear differences between the strands. Oscillatory expression of *let-7* miRNA passenger strands suggest that transcription of *let-7* oscillates while steady-state levels of the *let-7* guide strand rise in a stepwise manner. *lin-4* guide levels however rise rapidly and stay relatively stable until transcription of *lin-4* stops in the adult stage. Both *lin-4* and *let-7* are examples of miRNAs that show a constitutive difference between guide and passenger stability. Two other miRNAs of particular interest are miR-788 and miR-235 (figure 3.4B III and 3.4B IV). Transcription of these two miRNAs oscillates, as can be seen by the oscillations in the miRNA passenger strand levels. The mature miRNA levels however show two distinct expression patterns (figure 3.4). miR-788 is high during L2 and L3 and is rapidly degraded

during the L4 stage. This suggests that the miRNA is temporarily stabilized, and then rapidly degraded. miR-235 on the other hand oscillates with a phase that is distinct from its passenger strand. When the worms reach the adult stage however, the guide strand is stabilized in the absence of transcription. This miRNA seems to be highly unstable during development, but is stabilized in adult animals. Taken together this data shows that differential stabilization between miRNA guide and passenger strands is widespread. It furthermore shows that this stabilization of the guide strand can be regulated and change over development in *C. elegans*. miRNA oscillations provide an excellent model to study the mechanisms that govern miRNA stability modification and identify the key factors involved.

4.7.2 Transcription factors that drive oscillations

The molecular mechanisms driving oscillatory expression during nematode larval development remain unclear. As we have shown previously, the oscillations are driven on a transcriptional level (figure 2.6A). A number of transcriptional regulators have been implicated in seam cell development and timing of molting. Strikingly, worms that carry a mutation in *lin-42*, which is the *C. elegans* orthologue of human *period*, have severe defects in the timing of molting as well as seam cell divisions (Monsalve et al., 2011). The homology to *period*, cycling expression pattern of *lin-42*, and phenotypes of *lin-42* mutant animals have led to the hypothesis that LIN-42 acts as the main timer of seam cell divisions and the molting cycle (Monsalve et al., 2011). More recent data however has shown that oscillatory expression of certain miRNAs is maintained in the absence of *lin-42* (McCulloch and Rougvie, 2014). Additionally, available chromatin immunoprecipitation (ChIP) DNA sequencing data does not suggest that LIN-42 targets are enriched for oscillating genes (Perales et al., 2014) (figure 4.2A). Taken together, this indicates that although *lin-42* may play an important role, it is unlikely to be the main timer of *C. elegans* larval oscillatory expression.

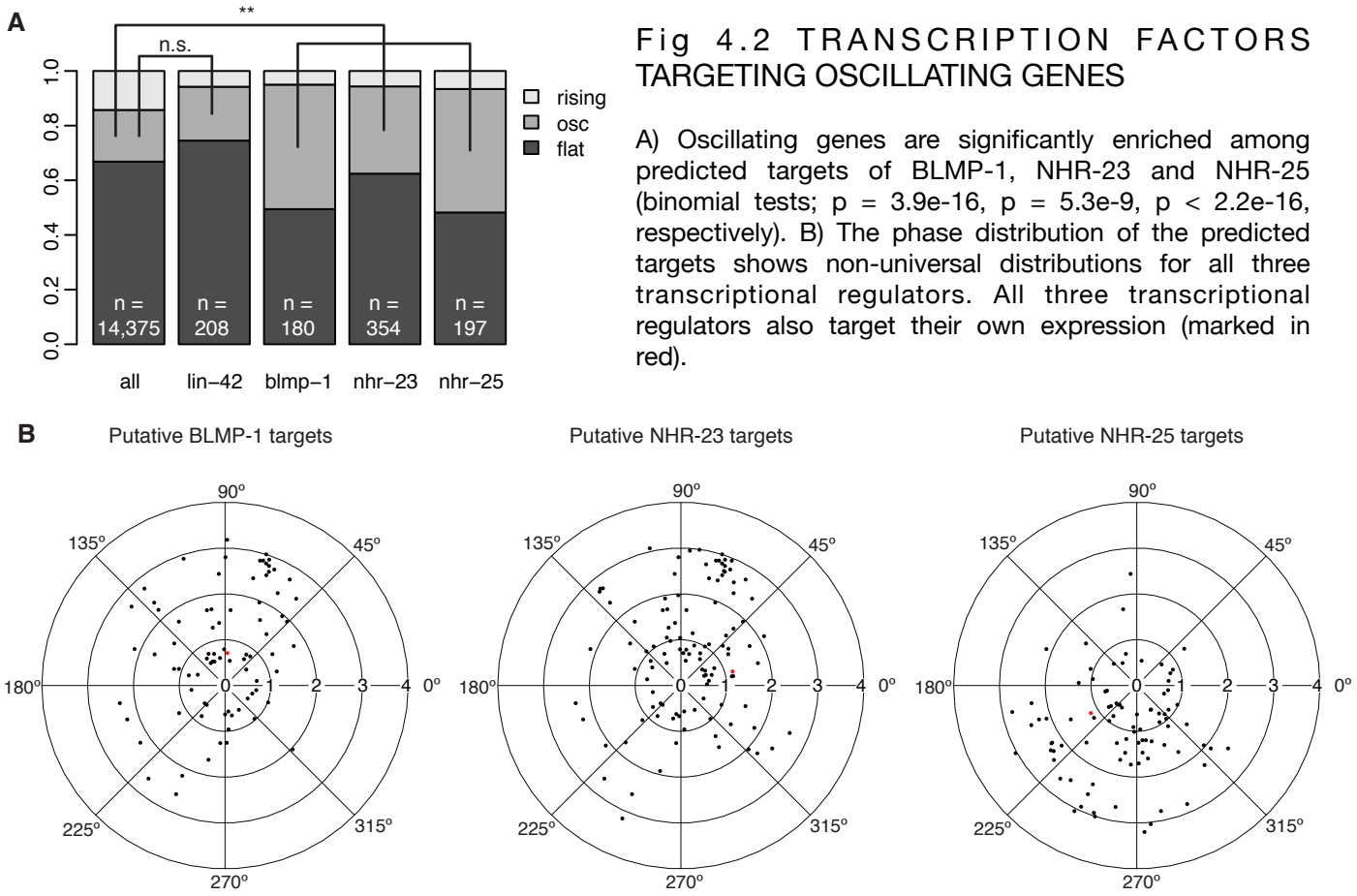


Fig 4.2 TRANSCRIPTION FACTORS TARGETING OSCILLATING GENES

A) Oscillating genes are significantly enriched among predicted targets of BLMP-1, NHR-23 and NHR-25 (binomial tests; p = 3.9e-16, p = 5.3e-9, p < 2.2e-16, respectively). B) The phase distribution of the predicted targets shows non-universal distributions for all three transcriptional regulators. All three transcriptional regulators also target their own expression (marked in red).

nhr-23 on the other hand encodes for orthologues of the accessory mammalian clock proteins REV-ERB and RORA. Interestingly, knockdown of *nhr-23* results in molting defects (Monsalve et al., 2011). RNAi of another NHR encoding gene, *nhr-25*, gives rise to molting phenotypes as well as hypodermal differentiation defects (Kostrouchova et al., 1998). All three of these transcriptional regulators have been previously suggested to be involved in the cyclic or modular development of *C. elegans* during larval development (Monsalve and Frand, 2012). NHR-23 and NHR-25, as well as another transcriptional regulator, BLMP-1, show significant enrichments of oscillating genes among their targets, as predicted by ChIP (figure 4.2A) (Celniker et al., 2009). In addition, the phases of these predicted targets seem to be relatively restricted (figure 4.2B). Interestingly, preliminary data from promoter dissection studies in the lab identified a putative *blmp-1* binding site that is essential for oscillatory expression of a reporter (YP Hauser, GJ Hendriks & H Grosshans, data not shown). *blmp-1* is the *C. elegans* orthologue of the highly conserved

mammalian B lymphocyte-induced maturation protein 1 (blimp1). This transcriptional repressor was originally discovered to transcriptionally repress human beta-interferon in mammalian cell culture (Keller and Maniatis, 1991). In *C. elegans*, a mutation in *blmp-1(tm548)* derepresses *unc-5* in the distal tip cells, which results in distal tip migration defects. A recent study showed that BLMP-1 in *C. elegans* may also function as a transcriptional activator (Yang et al., 2015). Our preliminary data shows that oscillations are strongly reduced in this background. However, asynchrony of the population of worms could also cause extensive reduction of oscillatory expression. The asynchrony that we see in the strain does not appear to be more extreme than in other mutants where oscillatory expression is maintained, but this needs to be quantified in more detail. We therefore hypothesize that BLMP-1 is a major regulator of oscillatory expression in *C. elegans* development. Quantitative measures of synchrony will be performed in the future to further examine this hypothesis.

4.7.3 Tissue specificity of oscillating genes

We know that oscillatory expression is abundant in the hypodermis. Our reporters of oscillating genes are however not exclusively expressed in the hypodermis. In many cases, the hypodermis is only one of the tissues where we see expression. It may be that expression in some additional tissues, where no oscillation occurs, merely dampens the amplitude that we detect. However, since in most cases, the addition of two cosine waves produces another cosine wave with modified amplitude and phase, we cannot conclude that phases in the hypodermis are universally distributed. It is a possibility that oscillations in the hypodermis all occur in the same phase, and oscillations in another tissue occur in another phase. Co-expression and expression levels in these two tissues would determine the eventual phase and amplitude that we detect in whole-worm experiments. According to this model, it would even be possible that all genes expressed in the tissue oscillate. Genes that are specifically expressed in the hypodermis, such as collagens, for

which we see strong enrichment, and genes required for hypodermal fate specification, would show the highest amplitudes and similar phases. Genes that are also expressed in the other tissue, or even ubiquitously, would have lower amplitudes and be less likely to be classified as oscillating. This could explain the thousands of genes that oscillate at very low amplitude and were classified as ‘flat’. Whether oscillations occur in one or multiple tissues, and if the phases are distributed within a tissue, remains unknown and a question we hope to answer in the near future.

5. References

Allen, N.P., Huang, L., Burlingame, A., and Rexach, M. (2001). Proteomic analysis of nucleoporin interacting proteins. J. Biol. Chem. *276*, 29268–29274.

Altun, Z.F., Herndon, L.A., Wolkow, C.A., Crocker, C., Lints, R., and Hall, D.H. WormAtlas. *2002-2015*.

Amano, T., Sagai, T., Tanabe, H., Mizushina, Y., Nakazawa, H., and Shiroishi, T. (2009). Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. Dev. Cell *16*, 47–57.

Ambros, V., and Horvitz, H.R. (1984). Heterochronic mutants of the nematode Caenorhabditis elegans. Science *226*, 409–416.

Antebi, A. (2006). Nuclear hormone receptors in C. elegans. WormBook 1–13.

Arner, E., Daub, C.O., Vitting-Seerup, K., Andersson, R., Lilje, B., Drablos, F., Lennartsson, A., Rönnerblad, M., Hrydziuszko, O., Vitezic, M., et al. (2015). Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. Science.

Aulehla, A., Wehrle, C., Brand-Saberi, B., Kemler, R., Gossler, A., Kanzler, B., and Herrmann, B.G. (2003). Wnt3a plays a major role in the segmentation clock controlling somitogenesis. Dev. Cell *4*, 395–406.

Aulehla, A., Wiegraebe, W., Baubet, V., Wahl, M.B., Deng, C., Taketo, M., Lewandoski, M., and Pourquié, O. (2008). A beta-catenin gradient links the clock and wavefront systems in mouse embryo segmentation. Nat. Cell Biol. *10*, 186–193.

Banerjee, D., Kwok, A., Lin, S.-Y., and Slack, F.J. (2005). Developmental timing in C. elegans is regulated by kin-20 and tim-1, homologs of core circadian clock genes. Dev. Cell *8*, 287–295.

Barrett, R.K., and Takahashi, J.S. (1995). Temperature compensation and temperature entrainment of the chick pineal cell circadian clock. J. Neurosci. *15*, 5681–5692.

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823–837.

Bazzini, A.A., Lee, M.T., and Giraldez, A.J. (2012). Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. Science *336*, 233–237.

Bächinger, H.P. (1987). The influence of peptidyl-prolyl cis-trans isomerase on the in vitro folding of type III collagen. J. Biol. Chem. *262*, 17144–17148.

Behm-Ansmant, I., Rehwinkel, J., Doerks, T., Stark, A., Bork, P., and Izaurralde, E. (2006). mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. Genes & Development *20*, 1885–1898.

Benito, J., Zheng, H., and Hardin, P.E. (2007). PDP1epsilon functions downstream of the circadian oscillator to mediate behavioral rhythms. J. Neurosci. *27*, 2539–2547.

Besier, R.B., Kahn, L.P., Sargison, N.D., and Van Wyk, J.A. (2016). Diagnosis, Treatment and Management of Haemonchus contortus in Small Ruminants (Elsevier Ltd).

Bessho, Y., Hirata, H., Masamizu, Y., and Kageyama, R. (2003). Periodic repression by the bHLH factor Hes7 is an essential mechanism for the somite segmentation clock. Genes & Development *17*, 1451–1456.

Blau, J., and Young, M.W. (1999). Cycling vrille expression is required for a functional Drosophila clock. Cell *99*, 661–671.

Brooks, D.R., Appleford, P.J., Murray, L., and Isaac, R.E. (2003). An essential role in molting and morphogenesis of Caenorhabditis elegans for ACN-1, a novel member of the angiotensin-converting enzyme family that lacks a metallopeptidase active site. J. Biol. Chem. *278*, 52340–52346.

Butler, J.E.F., and Kadonaga, J.T. (2002). The RNA polymerase II core promoter: a key component in the regulation of gene expression. Genes & Development *16*, 2583–2592.

Celniker, S.E., Dillon, L.A.L., Gerstein, M.B., Gunsalus, K.C., Henikoff, S., Karpen, G.H., Kellis, M., Lai, E.C., Lieb, J.D., MacAlpine, D.M., et al. (2009). Unlocking the secrets of the genome. Nature *459*, 927–930.

Chen, T., and Dent, S.Y.R. (2014). Chromatin modifiers and remodellers: regulators of cellular differentiation. Nature Publishing Group *15*, 93–106.

Cole, S.E., Levorse, J.M., Tilghman, S.M., and Vogt, T.F. (2002). Clock regulatory elements control cyclic expression of Lunatic fringe during somitogenesis. Dev. Cell *3*, 75–84.

Coller, J.M., Gray, N.K., and Wickens, M.P. (1998). mRNA stabilization by poly(A) binding protein is independent of poly(A) and requires translation. Genes & Development *12*, 3226–3235.

Cooke, J., and Zeeman, E.C. (1976). A clock and wavefront model for control of the number of repeated structures during animal morphogenesis. J. Theor. Biol. *58*, 455–476.

Cutter, A.D. (2008). Divergence times in Caenorhabditis and Drosophila inferred from direct estimates of the neutral mutation rate. Mol. Biol. Evol. *25*, 778–786.

Cyran, S.A., Buchsbaum, A.M., Reddy, K.L., Lin, M.-C., Glossop, N.R.J., Hardin, P.E., Young, M.W., Storti, R.V., and Blau, J. (2003). vrille, Pdp1, and dClock form a second feedback loop in the Drosophila circadian clock. Cell *112*, 329–341.

Denli, A.M., Tops, B.B.J., Plasterk, R.H.A., Ketting, R.F., and Hannon, G.J. (2004). Processing of primary microRNAs by the Microprocessor complex. Nature *432*, 231–235.

Djuranovic, S., Nahvi, A., and Green, R. (2012). miRNA-mediated gene silencing by translational repression followed by mRNA deadenylation and decay. Science *336*, 237–240.

Dubrulle, J., McGrew, M.J., and Pourquié, O. (2001). FGF signaling controls somite boundary position and regulates segmentation clock control of spatiotemporal Hox gene activation. Cell *106*, 219–232.

Dubrulle, J., and Pourquié, O. (2004). fgf8 mRNA decay establishes a gradient that couples axial elongation to patterning in the vertebrate embryo. Nature *427*, 419–422.

Edens, W.A., Sharling, L., Cheng, G., Shapira, R., Kinkade, J.M., Lee, T., Edens, H.A., Tang, X., Sullards, C., Flaherty, D.B., et al. (2001). Tyrosine cross-linking of extracellular matrix is catalyzed by Duox, a multidomain oxidase/peroxidase with homology to the phagocyte oxidase subunit gp91phox. The Journal of Cell Biology *154*, 879–891.

Flames, N., and Hobert, O. (2009). Gene regulatory logic of dopamine neuron differentiation. Nature *458*, 885–889.

Forsberg, H., Crozet, F., and Brown, N.A. (1998). Waves of mouse Lunatic fringe expression, in four-hour cycles at two-hour intervals, precede somite boundary formation. Current Biology *8*, 1027–1030.

Francesconi, M., and Lehner, B. (2014). The effects of genetic variation on gene expression dynamics during development. Nature *505*, 208–211.

Frand, A.R., Russel, S., and Ruvkun, G. (2005). Functional genomic analysis of C. elegans molting. Plos Biol *3*, e312.

Gaidatzis, D., Burger, L., Florescu, M., and Stadler, M.B. (2015). Analysis of intronic and exonic reads in RNA-seq data characterizes transcriptional and post-transcriptional regulation. Nat. Biotechnol. *33*, 722–729.

Gatfield, D., Le Martelot, G., Vejnar, C.E., Gerlach, D., Schaad, O., Fleury-Olela, F., Ruskeepää, A.-L., Oresic, M., Esau, C.C., Zdobnov, E.M., et al. (2009). Integration of microRNA miR-122 in hepatic circadian gene expression. Genes & Development *23*, 1313–1326.

Gissendanner, C.R., and Sluder, A.E. (2000). nhr-25, the Caenorhabditis elegans ortholog of ftz-f1, is required for epidermal and somatic gonad development. Dev. Biol. *221*, 259–272.

Gissendanner, C.R., Crossgrove, K., Kraus, K.A., Maina, C.V., and Sluder, A.E. (2004). Expression and function of conserved nuclear receptor genes in Caenorhabditis elegans. Dev. Biol. *266*, 399–416.

Gregory, R.I., Yan, K.-P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., and Shiekhattar, R. (2004). The Microprocessor complex mediates the genesis of microRNAs. Nature *432*, 235–240.

Grishok, A., Pasquinelli, A.E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D.L., Fire, A., Ruvkun, G., and Mello, C.C. (2001). Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. Cell *106*, 23–34.

Grün, D., Kirchner, M., Thierfelder, N., Stoeckius, M., Selbach, M., and Rajewsky, N. (2014). Conservation of mRNA and Protein Expression during Development of C. elegans. Cell Reports.

Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature *466*, 835–840.

Imataka, H., Gradi, A., and Sonenberg, N. (1998). A newly identified N-terminal amino acid sequence of human eIF4G binds poly(A)-binding protein and functions in poly(A)-dependent translation. The EMBO Journal *17*, 7480–7489.

Jeon, M., Gardner, H.F., Miller, E.A., Deshler, J., and Rougvie, A.E. (1999). Similarity of the C. elegans developmental timing protein LIN-42 to circadian rhythm proteins. Science *286*, 1141–1146.

Kato, M., de Lencastre, A., Pincus, Z., and Slack, F.J. (2009). Dynamic expression of small non-coding RNAs, including novel microRNAs and piRNAs/21U-RNAs, during Caenorhabditis elegans development. Genome Biol *10*, R54.

Keller, A.D., and Maniatis, T. (1991). Identification and characterization of a novel repressor of beta-interferon gene expression. Genes & Development *5*, 868–879.

Ketting, R.F., Fischer, S.E., Bernstein, E., Sijen, T., Hannon, G.J., and Plasterk, R.H. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. Genes & Development *15*, 2654–2659.

Khvorova, A., Reynolds, A., and Jayasena, S.D. (2003). Functional siRNAs and miRNAs exhibit strand bias. Cell *115*, 209–216.

Kim, D.H., Grün, D., and van Oudenaarden, A. (2013). Dampening of expression oscillations by synchronous regulation of a microRNA and its target. Nature Publishing Group *45*, 1337–1344.

Kippert, F., Saunders, D.S., and Blaxter, M.L. (2002). Caenorhabditis elegans has a circadian clock. Current Biology *12*, R47–R49.

Kloss, B., Price, J.L., Saez, L., Blau, J., Rothenfluh, A., Wesley, C.S., and Young, M.W. (1998). The Drosophila clock gene double-time encodes a protein closely related to human casein kinase Iepsilon. Cell *94*, 97–107.

Koh, K., Zheng, X., and Sehgal, A. (2006). JETLAG resets the Drosophila circadian clock by promoting light-induced degradation of TIMELESS. Science *312*, 1809–1812.

Komarnitsky, P., Cho, E.J., and Buratowski, S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes & Development *14*, 2452–2460.

Kostrouchova, M.M., Krause, M.M., Kostrouch, Z.Z., and Rall, J.E.J. (2001). Nuclear hormone receptor CHR3 is a critical regulator of all four larval molts of the nematode Caenorhabditis elegans. Proc Natl Acad Sci U S A *98*, 7360–7365.

Kostrouchova, M., Krause, M., Kostrouch, Z., and Rall, J.E. (1998). CHR3: a Caenorhabditis elegans orphan nuclear hormone receptor required for proper epidermal development and molting. Development *125*, 1617–1626.

Krieger, D.T. (1974). Food and water restriction shifts corticosterone, temperature, activity and brain amine periodicity. Endocrinology *95*, 1195–1201.

Krieger, D.T., Hauser, H., and Krey, L.C. (1977). Suprachiasmatic nuclear lesions do not abolish food-shifted circadian adrenal and temperature rhythmicity. Science *197*, 398–399.

Kuervers, L.M., Jones, C.L., O'Neil, N.J., and Baillie, D.L. (2003). The sterol modifying enzyme LET-767 is essential for growth, reproduction and development in Caenorhabditis elegans. Mol. Genet. Genomics *270*, 121–131.

Kula-Eversole, E., Nagoshi, E., Shang, Y., Rodriguez, J., Allada, R., and Rosbash, M. (2010). Surprising gene expression patterns within and between PDF-containing circadian neurons in Drosophila. Proc Natl Acad Sci U S A *107*, 13497–13502.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116–120.

Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. (2001). An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science *294*, 858–862.

Lebailly, B., Boitard, C., and Rogner, U.C. (2015). Circadian rhythm-related genes: implication in autoimmunity and type 1 diabetes. Diabetes Obes Metab *17 Suppl 1*, 134–138.

Lee, C., Bae, K., and Edery, I. (1999). PER and TIM inhibit the DNA binding activity of a Drosophila CLOCK-CYC/dBMAL1 heterodimer without disrupting formation of the heterodimer: a basis for circadian transcription. Mol. Cell. Biol. *19*, 5316–5325.

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell *75*, 843–854.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Rådmark, O., Kim, S., et al. (2003). The nuclear RNase III Drosha initiates microRNA processing. Nature *425*, 415–419.

Lewis, J. (2003). Autoinhibition with transcriptional delay: a simple mechanism for the zebrafish somitogenesis oscillator. Current Biology *13*, 1398–1408.

Lim, L.P., Lau, N.C., Weinstein, E.G., Abdelhakim, A., Yekta, S., Rhoades, M.W., Burge, C.B., and Bartel, D.P. (2003). The microRNAs of Caenorhabditis elegans. Genes & Development *17*, 991–1008.

Lin, F.J., Song, W., Meyer-Bernstein, E., Naidoo, N., and Sehgal, A. (2001). Photic signaling by cryptochrome in the Drosophila circadian system. Mol. Cell. Biol. *21*, 7287–7294.

Lin, J.-M., Kilman, V.L., Keegan, K., Paddock, B., Emery-Le, M., Rosbash, M., and Allada, R. (2002). A role for casein kinase 2alpha in the Drosophila circadian clock. Nature *420*, 816–820.

Lorch, Y., Zhang, M., and Kornberg, R.D. (1999). Histone octamer transfer by a chromatin-remodeling complex. Cell *96*, 389–392.

Luger, K., Dechassa, M.L., and Tremethick, D.J. (2012). New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? Nat. Rev. Mol. Cell Biol. *13*, 436–447.

Lund, E., Güttinger, S., Calado, A., Dahlberg, J.E., and Kutay, U. (2004). Nuclear export of microRNA precursors. Science *303*, 95–98.

Marzluff, W.F., Wagner, E.J., and Duronio, R.J. (2008). Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. Nature Publishing Group *9*, 843–854.

Matyash, V., Entchev, E.V., Mende, F., Wilsch-Bräuninger, M., Thiele, C., Schmidt, A.W., Knölker, H.-J., Ward, S., and Kurzchalia, T.V. (2004). Sterol-derived hormone(s) controls entry into diapause in Caenorhabditis elegans by consecutive activation of DAF-12 and DAF-16. Plos Biol *2*, e280.

McCulloch, K.A., and Rougvie, A.E. (2014). Caenorhabditis elegans period homolog lin-42 regulates the timing of heterochronic miRNA expression. Proc Natl Acad Sci U S A *111*, 15450–15455.

Merris, M., Wadsworth, W.G., Khamrai, U., Bittman, R., Chitwood, D.J., and Lenard, J. (2003). Sterol effects and sites of sterol accumulation in

Caenorhabditis elegans: developmental requirement for 4alpha-methyl sterols. J. Lipid Res. *44*, 172–181.

Monsalve, G., and Frand, A. (2012). Toward a unified model of developmental timing: A “molting” approach. Worm *1*, 221–230.

Monsalve, G.C., Van Buskirk, C., and Frand, A.R. (2011). LIN-42/PERIOD Controls Cyclical and Developmental Progression of C. elegans Molts. Current Biology *21*, 2033–2045.

Moore, R.Y., and Eichler, V.B. (1972). Loss of a circadian adrenal corticosterone rhythm following suprachiasmatic lesions in the rat. Brain Res. *42*, 201–206.

Mootz, D., Ho, D.M., and Hunter, C.P. (2004). The STAR/Maxi-KH domain protein GLD-1 mediates a developmental switch in the translational control of C. elegans PAL-1. Development *131*, 3263–3272.

Morales, A.V., Yasuda, Y., and Ish-Horowicz, D. (2002). Periodic Lunatic fringe expression is controlled during segmentation by a cyclic transcriptional enhancer responsive to notch signaling. Dev. Cell *3*, 63–74.

Motola, D.L., Cummins, C.L., Rottiers, V., Sharma, K.K., Li, T., Li, Y., Suino-Powell, K., Xu, H.E., Auchus, R.J., Antebi, A., et al. (2006). Identification of ligands for DAF-12 that govern dauer formation and reproduction in C. elegans. Cell *124*, 1209–1223.

Musselman, C.A., Lalonde, M.-E., Côté, J., and Kutateladze, T.G. (2012). Perceiving the epigenetic landscape through histone readers. Nat Struct Mol Biol *19*, 1218–1227.

Naiche, L.A., Holder, N., and Lewandoski, M. (2011). FGF4 and FGF8 comprise the wavefront activity that controls somitogenesis. Proc Natl Acad Sci U S A *108*, 4018–4023.

Novelli, J., Ahmed, S., and Hodgkin, J. (2004). Gene interactions in Caenorhabditis elegans define DPY-31 as a candidate procollagen C-proteinase and SQT-3/ROL-4 as its predicted major target. Genetics *168*, 1259–1273.

Page, A.P., and Johnstone, I.L. (2007). The cuticle. WormBook 1–15.

Palmeirim, I., Henrique, D., Ish-Horowicz, D., and Pourquié, O. (1997). Avian hairy gene expression identifies a molecular clock linked to vertebrate segmentation and somitogenesis. Cell *91*, 639–648.

Panda, S., Hogenesch, J.B., and Kay, S.A. (2002). Circadian rhythms from flies to human. Nature *417*, 329–335.

Panin, V.M., Papayannopoulos, V., Wilson, R., and Irvine, K.D. (1997). Fringe modulates Notch-ligand interactions. Nature *387*, 908–912.

Parkins, J.J., and Holmes, P.H. (1989). Effects of gastrointestinal helminth parasites on ruminant nutrition. Nutr Res Rev *2*, 227–246.

Perales, R., King, D.M., Aguirre-Chen, C., and Hammell, C.M. (2014). LIN-42, the Caenorhabditis elegans PERIOD homolog, Negatively Regulates MicroRNA Transcription. PLoS Genet.

Peschel, N., and Helfrich-Förster, C. (2011). Setting the clock--by nature: circadian rhythm in the fruitfly Drosophila melanogaster. FEBS Lett. *585*, 1435–1442.

Pittendrigh, C.S. (1954). ON TEMPERATURE INDEPENDENCE IN THE CLOCK SYSTEM CONTROLLING EMERGENCE TIME IN DROSOPHILA. Proc Natl Acad Sci U S A *40*, 1018–1029.

Pittendrigh, C.S. (1981). Circadian Systems: Entrainment. Biological Rhythms *II*, 95–124.

Plautz, J.D., Kaneko, M., Hall, J.C., and Kay, S.A. (1997). Independent photoreceptive circadian clocks throughout Drosophila. Science *278*, 1632–1635.

Prockop, D.J., and Kivirikko, K.I. (1995). Collagens: molecular biology, diseases, and potentials for therapy. Annu. Rev. Biochem. *64*, 403–434.

Reddy, P., Zehring, W.A., Wheeler, D.A., Pirrotta, V., Hadfield, C., Hall, J.C., and Rosbash, M. (1984). Molecular analysis of the period locus in Drosophila melanogaster and identification of a transcript involved in biological rhythms. Cell *38*, 701–710.

Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R., and Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature *403*, 901–906.

Rodriguez, J., Tang, C.-H.A., Khodor, Y.L., Vodala, S., Menet, J.S., and Rosbash, M. (2013). Nascent-Seq analysis of Drosophila cycling gene expression. Proc Natl Acad Sci U S A *110*, E275–E284.

Ruaud, A.-F., and Bessereau, J.-L. (2006). Activation of nicotinic receptors uncouples a developmental timer from the molting timer in C. elegans. Development *133*, 2211–2222.

Saga, Y., Hata, N., Koseki, H., and Taketo, M.M. (1997). Mesp2: a novel mouse gene expressed in the presegmented mesoderm and essential for segmentation initiation. Genes & Development *11*, 1827–1839.

Sahar, S., and Sassone-Corsi, P. (2009). Metabolism and cancer: the circadian clock connection. Nat. Rev. Cancer *9*, 886–896.

Saigusa, T., Ishizaki, S., Watabiki, S., Ishii, N., Tanakadate, A., Tamai, Y., and Hasegawa, K. (2002). Circadian behavioural rhythm in Caenorhabditis elegans. Current Biology *12*, R46–R47.

Sainsbury, S., Bernecky, C., and Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. Nat. Rev. Mol. Cell Biol. *16*, 129–143.

Sakai, Y., Meno, C., Fujii, H., Nishino, J., Shiratori, H., Saijoh, Y., Rossant, J., and Hamada, H. (2001). The retinoic acid-inactivating enzyme CYP26 is essential for establishing an uneven distribution of retinoic acid along the anterio-posterior axis within the mouse embryo. Genes & Development *15*, 213–225.

Sehgal, A., Price, J.L., Man, B., and Young, M.W. (1994). Loss of circadian behavioral rhythms and per RNA oscillations in the Drosophila mutant timeless. Science *263*, 1603–1606.

Serth, K., Schuster-Gossler, K., Cordes, R., and Gossler, A. (2003). Transcriptional oscillation of lunatic fringe is essential for somitogenesis. Genes & Development *17*, 912–925.

Shemer, G., and Podbilewicz, B. (2000). Fusomorphogenesis: cell fusion in organ formation. Dev. Dyn. *218*, 30–51.

Shi, X., Hong, T., Walter, K.L., Ewalt, M., Michishita, E., Hung, T., Carney, D., Peña, P., Lan, F., Kaadige, M.R., et al. (2006). ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. Nature *442*, 96–99.

Shim, E.Y., Walker, A.K., Shi, Y., and Blackwell, T.K. (2002). CDK-9/cyclin T (P-TEFb) is required in two postinitiation pathways for transcription in the C. elegans embryo. Genes & Development *16*, 2135–2146.

Simonetta, S.H., Migliori, M.L., Romanowski, A., and Golombek, D.A. (2009). Timing of locomotor activity circadian rhythms in Caenorhabditis elegans. PLoS ONE *4*, e7571.

Sirbu, I.O., and Duester, G. (2006). Retinoic-acid signalling in node ectoderm and posterior neural plate directs left-right patterning of somitic mesoderm. Nat. Cell Biol. *8*, 271–277.

Snoek, L.B., Sterken, M.G., Volkers, R.J.M., Klatter, M., Bosman, K.J., Bevers, R.P.J., Riksen, J.A.G., Smant, G., Cossins, A.R., and Kammenga, J.E. (2014). A rapid and massive gene expression shift marking adolescent transition in C. elegans. Sci Rep *4*, 3912.

Spellman, P.T., Sherlock, G., Zhang, M.Q., Iyer, V.R., Anders, K., Eisen, M.B., Brown, P.O., Botstein, D., and Futcher, B. (1998). Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. Molecular Biology of the Cell *9*, 3273–3297.

Spencer, W.C., Zeller, G., Watson, J.D., Henz, S.R., Watkins, K.L., McWhirter, R.D., Petersen, S., Sreedharan, V.T., Widmer, C., Jo, J., et al. (2011). A spatial and temporal map of C. elegans gene expression. Genome Research *21*, 325–341.

Spitz, F., and Furlong, E.E.M. (2012). Transcription factors: from enhancer binding to developmental control. Nature Publishing Group *13*, 613–626.

Stanewsky, R., Kaneko, M., Emery, P., Beretta, B., Wager-Smith, K., Kay, S.A., Rosbash, M., and Hall, J.C. (1998). The cryb mutation identifies cryptochrome as a circadian photoreceptor in Drosophila. Cell *95*, 681–692.

Sternberg, P.W. (1991). Control of cell lineage and cell fate during nematode development. Curr Top Dev Biol. 177–225.

Stoleru, D., Nawathean, P., Fernández, M. de L.P., Menet, J.S., Ceriani, M.F., and Rosbash, M. (2007). The Drosophila circadian network is a seasonal timer. Cell *129*, 207–219.

Sulston, J.E., and Horvitz, H.R. (1977). Post-embryonic cell lineages of the nematode, Caenorhabditis elegans. Dev. Biol. *56*, 110–156.

Sulston, J.E., Schierenberg, E., White, J.G., and Thomson, J.N. (1983). The embryonic cell lineage of the nematode Caenorhabditis elegans. Dev. Biol. *100*, 64–119.

Suzuki, M., Sagoh, N., Iwasaki, H., Inoue, H., and Takahashi, K. (2004). Metalloproteases with EGF, CUB, and thrombospondin-1 domains function in molting of Caenorhabditis elegans. Biol. Chem. *385*, 565–568.

Thacker, C., Sheps, J.A., and Rose, A.M. (2006). Caenorhabditis elegans dpy-5 is a cuticle procollagen processed by a proprotein convertase. Cell. Mol. Life Sci. *63*, 1193–1204.

Thein, M.C., Winter, A.D., Stepek, G., McCormack, G., Stapleton, G., Johnstone, I.L., and Page, A.P. (2009). Combined extracellular matrix cross-linking activity of the peroxidase MLT-7 and the dual oxidase BLI-3 is critical for post-embryonic viability in Caenorhabditis elegans. J. Biol. Chem. *284*, 17549–17563.

Thummel, C.S. (2001). Molecular mechanisms of developmental timing in C. elegans and Drosophila. Dev. Cell *1*, 453–465.

Turek, M., and Bringmann, H. (2014). Gene Expression Changes of Caenorhabditis elegans Larvae during Molting and Sleep-Like Lethargus. PLoS ONE *9*, e113269.

van der Linden, A.M., Beverly, M., Kadener, S., Rodriguez, J., Wasserman, S., Rosbash, M., and Sengupta, P. (2010). Genome-Wide Analysis of Light- and Temperature-Entrained Circadian Transcripts in Caenorhabditis elegans. Plos Biol *8*, e1000503.

van Rooij, E., Sutherland, L.B., Qi, X., Richardson, J.A., Hill, J., and Olson, E.N. (2007). Control of stress-dependent cardiac growth and gene expression by a microRNA. Science *316*, 575–579.

Vaucheret, H., Vazquez, F., Crété, P., and Bartel, D.P. (2004). The action of ARGONAUTE1 in the miRNA pathway and its regulation by the miRNA

pathway are crucial for plant development. Genes & Development *18*, 1187–1197.

Vermot, J., Gallego Llamas, J., Fraulob, V., Niederreither, K., Chambon, P., and Dollé, P. (2005). Retinoic acid controls the bilateral symmetry of somite formation in the mouse embryo. Science *308*, 563–566.

Vokes, S.A., Ji, H., Wong, W.H., and McMahon, A.P. (2008). A genome-scale analysis of the cis-regulatory circuitry underlying sonic hedgehog-mediated patterning of the mammalian limb. Genes & Development *22*, 2651–2663.

Whitehouse, I., Flaus, A., Cairns, B.R., White, M.F., Workman, J.L., and Owen-Hughes, T. (1999). Nucleosome mobilization catalysed by the yeast SWI/SNF complex. Nature *400*, 784–787.

Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell *75*, 855–862.

Winter, A.D., and Page, A.P. (2000). Prolyl 4-hydroxylase is an essential procollagen-modifying enzyme required for exoskeleton formation and the maintenance of body shape in the nematode Caenorhabditis elegans. Mol. Cell. Biol. *20*, 4084–4093.

Winter, A.D., Eschenlauer, S.C.P., McCormack, G., and Page, A.P. (2007a). Loss of secretory pathway FK506-binding proteins results in cold-sensitive lethality and associate extracellular matrix defects in the nematode Caenorhabditis elegans. J. Biol. Chem. *282*, 12813–12821.

Winter, A.D., Keskiaho, K., Kukkola, L., McCormack, G., Felix, M.-A., Myllyharju, J., and Page, A.P. (2007b). Differences in collagen prolyl 4-hydroxylase assembly between two Caenorhabditis nematode species despite high amino acid sequence identity of the enzyme subunits. Matrix Biol. *26*, 382–395.

Wysocka, J., Swigut, T., Xiao, H., Milne, T.A., Kwon, S.Y., Landry, J., Kauer, M., Tackett, A.J., Chait, B.T., Badenhorst, P., et al. (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. Nature *442*, 86–90.

Yaffe, K., Falvey, C.M., and Hoang, T. (2014). Connections between sleep and cognition in older adults. Lancet Neurol *13*, 1017–1028.

Yang, J., Fong, H.T., Xie, Z., Tan, J.W.H., and Inoue, T. (2015). Direct and positive regulation of Caenorhabditis elegans bed-3 by PRDM1/BLIMP1 ortholog BLMP-1. Biochim. Biophys. Acta *1849*, 1229–1236.

Yochem, J., Tuck, S., Greenwald, I., and Han, M. (1999). A gp330/megalin-related protein is required in the major epidermis of Caenorhabditis elegans for completion of molting. Development *126*, 597–606.

Yoo, S.-H., Yamazaki, S., Lowrey, P.L., Shimomura, K., Ko, C.H., Buhr, E.D., Siepka, S.M., Hong, H.-K., Oh, W.J., Yoo, O.J., et al. (2004).

PERIOD2::LUCIFERASE real-time reporting of circadian dynamics reveals persistent circadian oscillations in mouse peripheral tissues. Proc Natl Acad Sci U S A *101*, 5339–5346.

Young, M.W., and Kay, S.A. (2001). Time zones: a comparative genetics of circadian clocks. Nature Reviews Genetics *2*, 702–715.

Yu, W., Zheng, H., Houl, J.H., Dauwalder, B., and Hardin, P.E. (2006). PER-dependent rhythms in CLK phosphorylation and E-box binding regulate circadian transcription. Genes & Development *20*, 723–733.

Zhang, Z.Q. (2013). Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness (Addenda 2013). (Zootaxa).

Curriculum Vitae

In the electronic version of this thesis the CV has been omitted.