The *Caenorhabditis elegans* ALA neuron: its transcriptome and role in inducing sleep

Thesis by

Elly Suk Hen Chow

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2014
(Defended May 29, 2014)

To Nathan,

and my family; past, present, and future

Acknowledgments

First and foremost I would like to express my sincere gratitude to my advisor, Paul Sternberg, for having accepted me as a student in his laboratory. It is a privilege to work with Paul. This thesis is a sum of Paul's influence, insights, and inputs. I am grateful for the latitude of freedom he afforded me to explore, to learn, and the luxury to learn from my mistakes, while providing me with encouragement and timely guidance. His kindness to others, scientific insights, endless passion in science and life, and fearlessness in concurring challenges inspire me throughout the journey and beyond. Thank you, Paul, for trusting in me, and for being such an awesome mentor.

I would also like to thank the members of my thesis committee, David Prober, Angela Stathopoulos, and John Allman for kindly reviewing this work. I appreciate their guidance, advice, energy, and time over the years. David is also a collaborator on the zebrafish project; I thank him for the opportunity, inspirational discussions, his critiques, support, and amazing work for making it possible.

I am fortunate to have the opportunity to work with many talented and dedicated scientists. I thank Erich Schwarz for his bioinformatics wizardry, advices, critical comments, and encouragements along the way. Working with Erich allowed me to know him beyond C++, Python, and worm genomes. I appreciate his friendship, humor, and wonderfully fruitful collaboration that I hope for many years to come. I am thankful for Cheryl Van Buskirk, whose pioneering work sets the foundation for this project. Cheryl was my first lab mentor and I thank her for the patience and guidance. I will always remember the excitement we shared when we generated the first transgenic line, as well as the time we shared over scientific discussions and frozen yoghurt. Daniel Lee has been

a great support since we started the zebrafish project. His kindness, enthusiasm in science, and hard work inspire me. And I am thankful for his critical reviews on the manuscript.

Gratitude is owed to the many scientists who aspired and guided me to get to this point. Shuk Han Cheng has been an inspiration since my undergraduate and master research. I want to thank her for nurturing my burgeoning interest in science by first inviting me to work in her lab and showing me what "*hard work*" really means. I am grateful for her unyielding support and care throughout the years. I would like to express my gratitude to Eric Davidson, who showed me how to dissect a question systematically and to answer it logically. I thank Andy Cameron for teaching me the very first concept of *cis*-regulation and genomics. I want to thank Marianne Bronner for teaching me developmental biology and microdissection skills, which I applied on this project to isolate single neurons.

Many thanks to the people I worked with in the Sternberg lab. I could not have hoped for a more supportive, stimulating, and motivating environment than the one I found. I am especially thankful to the members in K217, where the blazing noise of the worm plate pouring machine is broken by intermittent laughter and long discussions on science, life, and food. I thank Mihoko Kato for her friendship, encouragement, and listening ear. I thank Hillel Schwartz for his scientific insights and honest comments, not to mention the scrumptious chocolates from Seattle. I appreciate both of them for spending time reviewing and commenting on my manuscripts. I thank my congenial bench mates over the years: Paul Minor for his humor and for generously sharing his meticulously seeded worm plates when I ran out of plates, which seemed to happen quite

often; Ravi Nath for reminding me why I love science by showing me his enthusiasm in science and through countless causal conversations over espresso and chai.

I am thankful for the friendship and help of my fellow graduate students: Steven Kuntz for introducing me to Mussa; Margaret Ho and Jonathan Liu for reading and critically reviewing my manuscripts; Pei-Yin Shih, James Lee, Wen Chen, Kai Yuet, Alli Akagi, Srimoyee Ghosh, Daniel Leighton, Katie Brugman, and Andrea Choe. Their presence brings joy and laughter to the lab.

I also want to thank past and present members of the Sternberg lab: "Amir Sapir, Yen-Ping Hsueh, Meenakshi Doma, Alon Zaslaver, Rioji Shinya, and Carmie Robinson", for their kindness and for generously exchanging notes on experiments, No work would have been done without the superb technical support from Gladys Medina, Barbara Perry, Shahla Gharib, and John DeModena. My special thanks go to Christopher Cronin, who patiently taught me to use MATLAB and worked with me side by side to tailor analysis programs for my data set; and to Mary Alvarez and Sarah Torress for administrative support. I am also thankful for members at WormBase for the enormous archives of worm knowledge and resources, and for their technical support that was always only an email away.

My students Victoria Gu, Katelyn McKown, and Christy Shen spent the last two summers and their time outside of the classroom working in the lab with me. I thank them for their hard work and pertinent questions that helped me better understand my project and improve my communication skills. Han Wang, our new postdoc, is here to investigate the site of action of the neuropeptide receptors identified in this work. I wish him much luck and success.

Over the years at Caltech, I have met and befriended with many wonderful people, too numerous to list but no less significant. I thank Willem den Besten for his humor and many good conversations; Hesham Azizgolshani for sharing insights in science and in life, and for helping me stay focused on the goal; Markus Hauschild for guiding me to see outside of the box. Also, William Chiu, Icy Ma, and Eve Helguero were there to help me get started at Caltech, while Evelyn Novello, Thomas Ng, John Young, Na Hu, Max Ezin, Jennifer Mok, Felicia Hunt, Portia Harris, and Weizhe Hong were there to help me through the course. Thank you for making this journey special.

Life would not have been balanced and complete without the tremendous patience, encouragement, and love from my friends outside of campus. Priscilla Li has been a cheerleader for many of my endeavors ever since we met in middle school. I thank her for helping to tear apart complex issues and for being the person who seems to always have the magical power to guide me in the right direction without saying a word. I am indebted to the Selicha family, Aggie, Jane, Joe, and Alex, for providing an immediate safety net to fall back on when I am thousands of miles away from my family.

The core propellants originate from my family. Nathan has been amazing and I could not be more grateful to have welcomed him into my life. He is by far the most aspiring mentor, critical judge, and harshest teacher I have ever met, but he is also the bundle of joy that I hold on to at any moment of life. I am eternally appreciative for the support, patience, and unconditional love streaming from my mother, Markus, Ashley, Aaron, Ada, Catherine, Jean-Jacques, and Sacha, even though they may not fully understand exactly what I do and what my science really means. I thank Cezar for encouraging me to accept challenges and to pursue dreams. My father, who planted the

seed of biology and health in my head with the articles he read on Reader's Digest, would have been more thrilled than anyone to witness where I have come to in my own journey of scientific adventure.

Preface

The Corner….

> *"When I was small, not much bigger than a pollywog," said Frog, "my father said to me, 'Son, this is a cold, gray day but spring is just around the corner,'*
>
> *I wanted spring to come. I went out to find that corner. I walked down a path in the woods until I came to a corner. I went around the corner to see if spring was on the other side."*
>
> *"And was it?" asked Toad.*
>
> *"No," said Frog. "There was only a pine tree, three pebbles and some dry grass."*
>
> *I walked in the meadow. Soon I came to another corner. I went around the corner to see if spring was there."*
>
> *"Did you find it?" asked Toad.*
>
> *"No," said Frog. "There was only an old worm asleep on a tree stump."*
>
> *……. Extracted from Frog and Toad All Year by Arnold Lobel*

Nothing is more beautiful than the excitement of conceiving the first idea of a project and the pure curiosity of wanting to know the what, how, and why. Time after time I thought I have come to the right corner to find the answer for one simple and direct question, and yet I found myself opening another can of worms and getting lost in more dangling unanswered questions. This journey of scientific adventure has truly been a test of perseverance and an intellectual challenge. I am grateful for the nurturing and supportive environment that Paul afforded for all my trials and errors. I may not have found all the answers but I have learned to appreciate the art of learning and the few inspirations that came along, which will stay with me beyond the chapter of academic education.

Abstract

A long-standing yet to be accomplished task in understanding behavior is to dissect the function of each gene involved in the development and function of a neuron. The *C. elegans* ALA neuron was chosen in this study for its known function in sleep, an ancient but less understood animal behavior. Single-cell transcriptome profiling identified 8,133 protein-coding genes in the ALA neuron, of which 57 are neuropeptide-coding genes. The most enriched genes are also neuropeptides. In combination with gain-of-function and loss-of-function assays, here I showed that the ALA-enriched FMRFamide neuropeptides, FLP-7, FLP-13, and FLP-24, are sufficient and necessary for inducing *C. elegans* sleep. These neuropeptides act as neuromodulators through GPCRs, NPR-7, and NPR-22. Further investigation in zebrafish indicates that FMRFamide neuropeptides are sleep-promoting molecules in animals. To correlate the behavioral outputs with genomic context, I constructed a gene regulatory network of the relevant genes controlling *C. elegans* sleep behavior through EGFR signaling in the ALA neuron. First, I identified an ALA cell-specific motif to conduct a genome-wide search for possible ALA-expressed genes. I then filtered out non ALA-expressed genes by comparing the motif-search genes with ALA transcriptomes from single-cell profiling. In corroborating with ChIP-seq data from modENCODE, I sorted out direct interaction of ALA-expressed transcription factors and differentiation genes in the EGFR sleep regulation pathway. This approach provides a network reference for the molecular regulation of *C. elegans* sleep behavior, and serves as an entry point for the understanding of functional genomics in animal behaviors.

Contents

Acknowledgments…….. …….3

Preface ……..8

Abstract ……9

Contents ……..……10

Chapter 1. An introduction of *Caenorhabditis elegans*……...11

Chapter 2. Towards the translation of genetic codes to behavioral outputs:

the ALA neuron and *Caenorhabditis elegans* sleep…… ..25

Chapter 3. Sleep – what do we know about it?.....37

Chapter 4. Sleep in *Caerhabditis elegans* and *Danio rerio*…….45

Chapter 5: FMRFamide neuropeptides promote sleep in

Caerhabditis elegans and *Danio rerio*……..53

Chapter 6: Deciphering the gene regulatory network of

a *Caenorhabditis elegans* sleep-inducing neuron……123

Chapter 7: Findings, Implications, and Future Plans…….173

Appendix: Supplementary table of ALA transcriptome…….181

Chapter 1

An introduction of *Caenorhabditis elegans*

Abstract

Much of today's understanding on behavior is at the physiological state and circuitry level; relatively little is known about the underlying molecular mechanisms. A long-standing, yet-to-be accomplished task is to dissect the function of each gene involved in the development and function of a neuron. Behaviors of animals are reflections of integrated specific characteristics and functionalities of molecular context stored in individual cells. In order to understand the function of a neuron, it is essential to identify the cell-specific genomic content as well as to decipher their regulatory interactions. The nematode *Caenorhabditis elegans* is small and simple but capable of executing sophisticated and dynamic behaviors as their vertebrate cousins do. It has a completely sequenced and well-annotated genome. The nervous system is relatively simple and composed of only 302 neurons, precisely organized in an invariant manner. Moreover, the characteristic morphology of each neuron is cataloged and the synaptic connections are carefully delineated. Herein I introduce the *C. elegans*, its anatomy, nervous system, and genome. Also, I discuss the potential of using *C. elegans* as a model organism for functional genomic analysis of animal behavior at the single-neuron resolution.

An overview of *C. elegans*

Caenorhabditis elegans is a free-living, soil-dwelling nematode found in most parts of the world. As the name tells, *C. elegans* has an unsegmented, rod-like body shape that is tapered at the ends (Figure 1; Altun and Hall, 2009). The body turgidity is maintained by an internal hydrostatic pressure enclosed in a layer of muscle bands that

run along the body. The outer surface is covered with a collagenous cuticle secreted by the underlying hypodermis, which attaches to the body-wall muscle. It moves in an elegant sinusoidal fashion, propelled by a sequence of opposing contractions of the dorsal and ventral muscles. *C. elegans* feeds on microbes, mostly bacteria, as it moves. Feeding is indicated by the contraction of the terminal bulb (Figure 2) and is observable under differential interference contrast (DIC) microscopy.

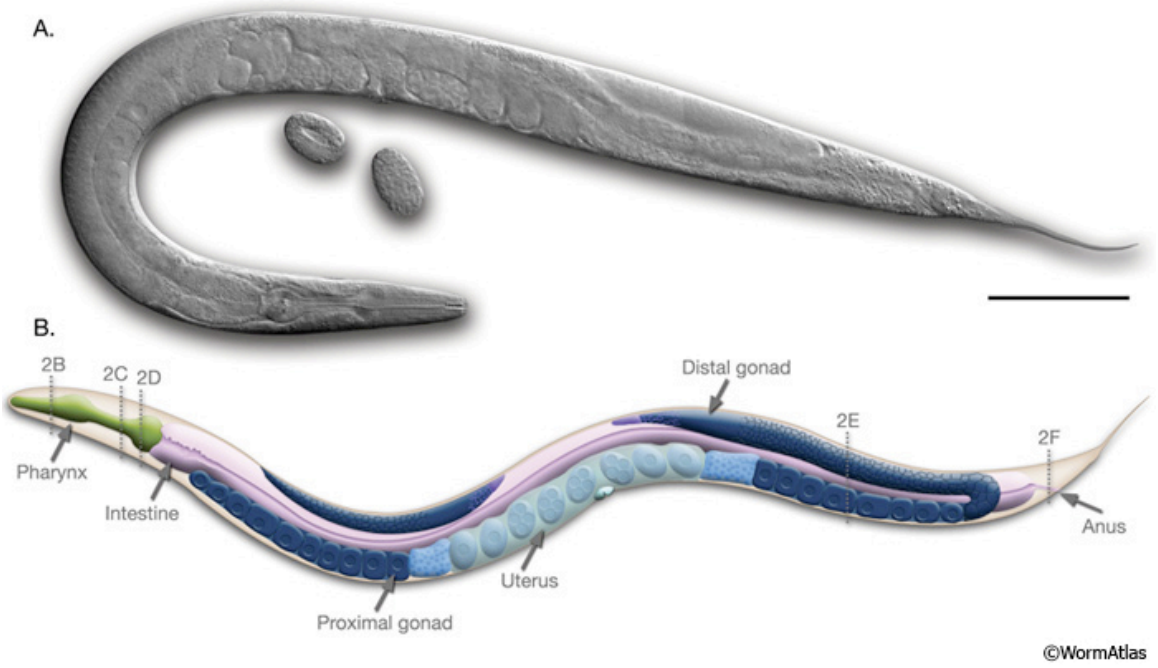


Figure 1. Anatomy of an adult hermaphrodite. (A) DIC image of an adult hermaphrodite, left lateral side. Scale bar 0.1mm. (B) Schematic drawing of anatomical structures, left lateral side. Adopted and unmodified from WormAtlas.

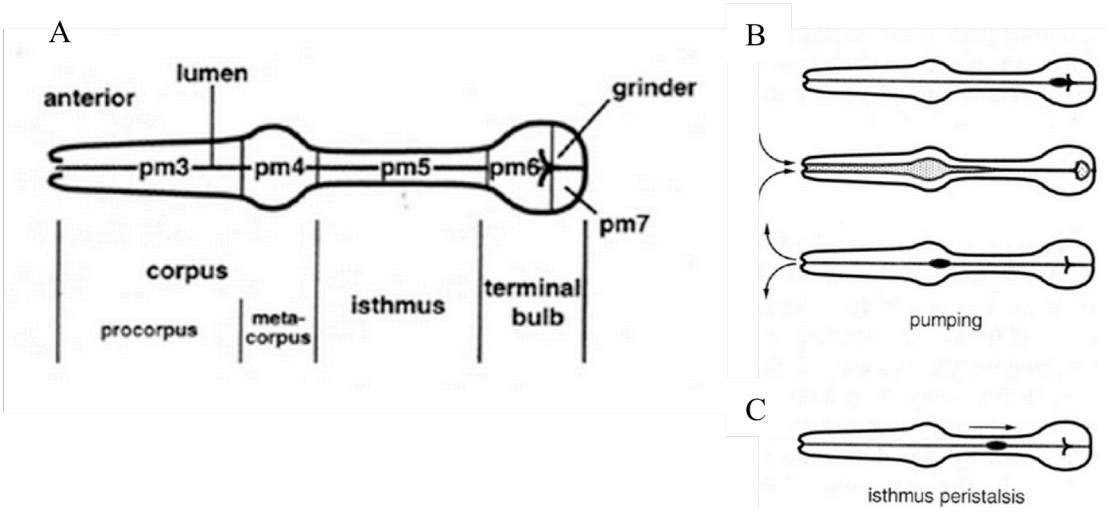


Figure 2. (A) The pharynx is divided into three parts: the corpus, isthmus, and terminal bulb. Lateral view, anterior to the left. (B) Feeding consists of two motions, pumping and isthmus peristalsis (Albertson and Thomson, 1976; Avery and Horvitz, 1989). A pump is a near-simultaneous contraction of the muscle of the corpus, anterior isthmus, and terminal bulb, followed by a near-simultaneous relaxation. Adopted and unmodified from WormAtlas.

Such a little worm, so much to offer

Sydney Brenner first introduced *C. elegans* as a model for molecular and developmental biology research (Brenner, 1974), and ever since then *C. elegans* has been extensively used in a broad spectrum of scientific researches from systems biology, cancer research, aging, and stem cells to neuroscience. Indeed, *C. elegans* presents several traits that make it a powerful and attractive model organism for studying genetics, development, and behavior. It has a small body size (approximately 1.5-mm-long adult, Riddle et al., 1997) and is easy to culture in a laboratory setup, making it an economic and convenient experimental model. The inner structures are visible through the transparent body and can easily be examined in living preparations under differential

interference contrast (DIC) microscopy. In addition, it grows rapidly and reaches adulthood within 3 days after hatching. Each individual produces 300-350 progenies. Over the years, a wide range of behavioral and morphological mutants have been generated and characterized. The complete, high-quality reference genome sequence offers much benefit to the worm community with markers and tools (Ambros, 2006) for genetic and genomic analysis. Last but not least, the connectome, a mapped network of neuronal synaptic connections in the nervous system (Albertson and Thompson, 1976; White et al., 1986; Jarrell et al., 2012), offers itself as an unmatched system for behavioral program and neuromodulation analysis.

The *C. elegans* nervous system

C. elegans hermaphrodites have a simple nervous system with 302 neurons of 118 types (Sulston and Horvitz, 1977), taking up about one third of the cells in the whole organism. These neurons are organized in two sub-systems: the somatic nervous system and the pharyngeal nervous system (Ward et al., 1975; Sulston and Horvitz, 1977; Sulston et al., 1983; White et al., 1986). The hermaphrodite somatic system has 282 neurons, while there are an additional 79 neurons in the males that are primarily for controlling mating. The pharyngeal nervous system consists of 20 neurons. These two sub-systems operate independently and communicate through a pair of interneurons (Altun and Hall, 2011). In general, the somatic nervous system is organized into ganglia in the head and tail, with the primary ganglia in the head (Figure 3A). Individual neurons in the primary ganglia sends out processes and forms a bundle of synaptic connections

that wraps around the isthmus of the pharynx anterior to the terminal bulb. This is called the "nerve ring", or sometimes referred to as the worm "mini brain" (Figure 3B).

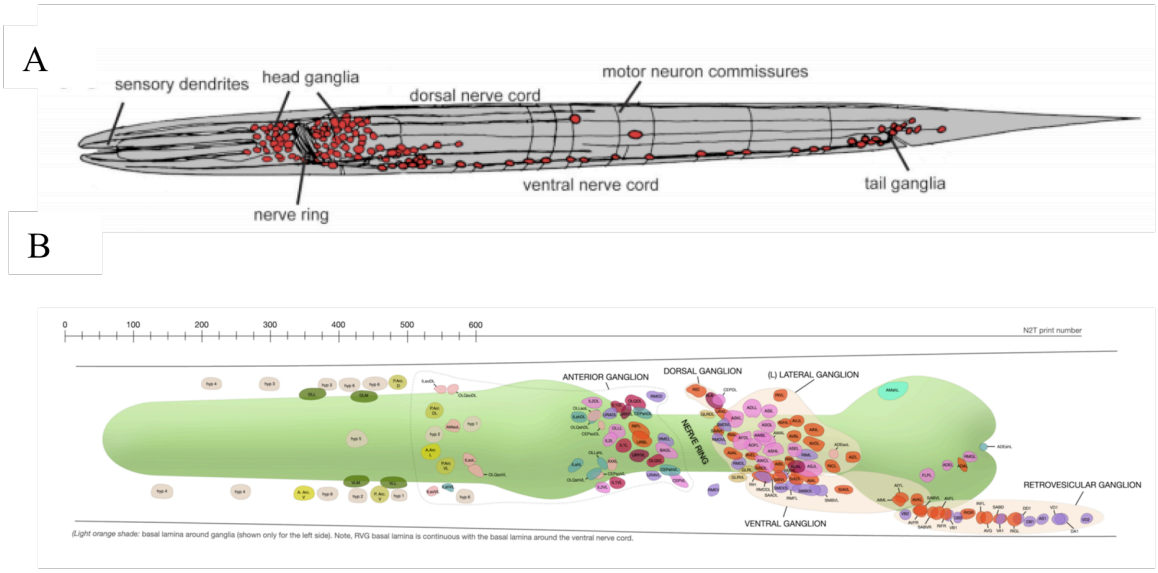


Figure 3. (A) An overview of the *C. elegans* nervous system. Adopted and unmodified from Hobert, 2010. (B) *C. elegans* head neurons. Schematic drawing of all head neurons in the left and right sides of the worm. Pharynx: shaded green. Adopted and unmodified from Wormatlas.

Development of the *C. elegans* somatic nervous system continues after embryogenesis. At the time of birth (hatching), the hermaphrodite larva has 202 somatic neurons. Additional 80 neurons are born at late-L1 and L2 stages (Altun and Hall, 2011). As a general rule, most *C. elegans* neurons are bilaterally symmetric pairs that arise from bilaterally symmetric cell lineages, but there are exceptions to this rule. For instance, the head peptidergic interneuron, ALA, is unpaired (White et al., 1986). The mature *C. elegans* nervous system has an estimation of 6400 chemical synapses, 1500 neuromuscular junctions, and 900 gap junctions (Altun and Hall, 2011).

Besides of executing basic behaviors such as locomotion, food seeking, feeding, and defecation (de Bono and Maricq, 2005), *C. elegans* is capable of computing dynamic environmental cues and giving discriminated responses accordingly. It moves toward favorable temperature, odorant, or food source and withdraws from negative cues such as noxious smell and harsh mechanical stimulation (Riddle and Meyer, 1997). The *C. elegans* nervous system is plastic. It exhibits associative learning and remembers averse stimuli or relevant environmental features that predict food availability (Ardiel and Rankin, 2010). This allows the worm to sense chemical, temperature, and oxygen levels and move to a more favorable environment. *C. elegans* also shows non-associative learning where the worm decreases response, both the amplitude and frequency, to repeated administration of a mechanical disturbance (Rankin, 1990) or odorant (Colbert and Bargmann, 1995). In addition, it can detect the density of nearby nematodes or the presence of a mate based on pheromone, changes in oxygen level, and the presence of short-range diffusible signals (Riddle and Golden, 1982, Cheung et al., 2004; Gray et al., 2004; Jeong et al., 2005; Barr and Garcia, 2006). Furthermore, *C. elegans* feeding, locomotion, and olfactory behavior can be modulated by the presence of food and food quality (Zhang et al., 2005).

The synaptic connections and morphological characteristics of each neuron in the *C. elegans* nervous system are cataloged and composed into connectivity maps (White et al., 1986; Jarrell et al., 2012). These maps, or the connectome as a whole, reveal invariant connections of neurons and provide valuable information for identifying neuronal circuitry that governs a specific behavior. Similar to other organisms, this fixed anatomy is subject to modulations through neurotransmitters and neuropeptides such as dopamine,

serotonin, and acetylcholine (Sawin et al., 2000; Hardaker et al., 2001; Gally et al., 2004). These neuromodulators modify neuronal dynamics, excitability, and synaptic functions, thereby changing the composition and activity of functional circuits (Bargmann, 2012).

The *C. elegans* genome

The *C. elegans* genome is not only well-annotated, but is also the first animal genome to have been completely sequenced. It is compact (100 Mb) and amazingly information-rich. Currently, there are 20,252 protein-coding genes (Schwarz et al., 2012) predicted in the genome, just a little more than the approximately 19,800 protein-coding genes reported when the genome was released (*C. elegans* Sequencing Consortium, 1998). More than 40% of the predicted protein-coding genes are conserved with other organisms (*C. elegans* Sequencing Consortium, 1998). The expression pattern and function of nearly all annotated genes are easily accessible online at WormBase (www.wormbase.org).

One of the most attractive features of *C. elegans* is the ease of generating mutants for gene function analysis. Random mutants are traditionally generated by methods such as chemical mutagens (Jansen et al. 1997; GengyoAndo and Mitani 2000; Edgley et al. 2002), mobilized endogenous transposons (Rushforth et al. 1993), and RNA interference (RNAi, Fire et al., 1998). Targeted mutants are now available by engineered nucleases designed to provide double-strand breaks at specific target sites to induce deletion mutation, examples of which are techniques like TALENs and CRISPRCas (Miller et al., 2011; Chiu et al., 2013).

Another attraction of the *C. elegans* genome is the ample resources for functional genomic analysis. The availability of genome sequences from other *Caenorhabditis* sibling species makes it possible for multispecies comparison within the family (Figure 4), an advantage that is rarely found in other animal models. There are at least 12,000 genes conserved between *C. elegans* and *C. briggsae* (Schwarz, 2005), two sibling species in the Elegans group. Within the Elegans group, *C. brenneri* subdivides an evolutionary branch between *C. elegans* and the siblings *C. briggsae* and *C. remanei.* Comparison between the *C. brenneri* genome and the other sibling genome can filter out nonfunctional DNA sequences that have failed to diverge in the sibling species (Kuntz et a., 2008). Comparison with a more remote species, *C. sp.* 3 PS1010, can further define highly conserved sequences in the *Caenorhabditis* genus. Such an approach has successfully identified the functional *Hox cis*-regulatory elements (Kuntz, et al., 2008). It is anticipated that similar conserved and functional *cis*-regulatory elements of genes of interest can be found using this approach.

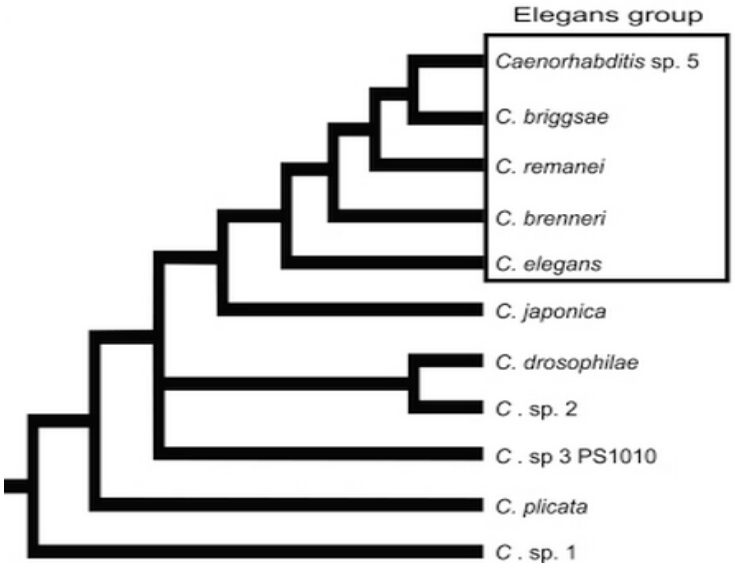


Figure 4. Phylogeny of nematodes within the *Caenorhabditis* genus from Kiontke et al. (2007). Adopted and unmodified from Kuntz et al., 2008.

The best-kept secret of *C. elegans* is the tools for genomic analysis at single-cell resolution. Much to the credit of Martin Chalfie, green fluorescent protein (*gfp*) has been widely used to illuminate protein-coding gene expression since it was first introduced by Chalfie et al., (1994). Building on this success and others, we now have the technology and tools to identify and quantify transcriptome in a group of specific cells (Spencer et al., 2011) or in single cell (Schwarz et al., 2012). Measurement of transcripts in a collection of single cell type is now made possible by fluorescence-activated cell sorting (FACS) and RNA-seq of cell-type specific gfp labeled cells (Spencer et al., 2011). This approach has successfully identified novel transcripts in a collection of given cells. However, the trace of transcripts from gfp unlabeled cells isolated from FACS reduces the accuracy of detected gene expression in a tissue type (Spencer et al., 2011). A more direct and reliable approach is to dissect single cells labeled with gfp whose expression is driven under a cell-specific promoter and perform RNA-seq to profile the cell-specific transcriptomes (Schwarz et al., 2012). Over the years, modENCODE (www.modencode.org) has generated 343 data sets on direct interaction of *C. elegans* genes through chromatin immuno-precipitation followed by sequencing (ChIP-seq), including 209 data sets of transcription factor binding sites (Gerstein et al., 2000). Corroborating data sets from cell-specific transcriptome profiling and ChIP-seq, an elaborated network elucidating gene interaction at the single cell resolution, is to be expected.

C. elegans as a model for single-cell genomics and behavioral analysis

The wealth of resources for biological research that *C. elegans* provides is enormous and

ever-growing. The marriage of an annotated genome and a delineated connectome obvisouly positioned the *C. elegans* for functional genomic studies of behavior at the single-neuron resolution. What remains to be explored is to identify the genes expressed in individual neurons and to investigate their functions. As a first step toward this goal, I chose *C. elegans* as a model for this study. Here I uncover the genomic content of a single neuron, and identify the functional molecules and their regulatory network that governs an ancient but less understood metazoan behavior: sleep.

Reference

1. Albertson, D.G., Thomson, J.N. (1976). The pharynx of *Caenorhabditis elegans.* Philos. Trans. R. Soc. Lond. B Biol. Sci. *275,* 299–325.
2. Altun, Z.F. and Hall, D.H. 2009. Introduction. In *WormAtlas*.
3. Ambros, V. (2006). Worm Methods. WormBook ed. The C. elegans Research Community:doi/10.1895/wormbook, http://www.wormbook.org.
4. Ardiel, E.L., and Rankin, C.H. (2010). An elegant mind: learning and memory in *Caenorhabditis elegans*. Learn Mem. *17(4),* 191-201.
5. Avery L., Horvitz, H .R. (1989). Pharyngeal pumping continues after laser killing of the pharyngeal nervous system of *C. elegans.* Neuron *3,* 473–485.
6. Bargmann, C.I. (2012). Beyond the connectome: How neuromodulators shape neural circuits. Bioessays *34,* 458-465.
7. Barr, M.M. and Garcia, L.R. (2006). Male mating behavior. In *WormBook* (ed. The *C.elegans* Research Community), WormBook, doi/10.1895/wormbook.1.7.1.
8. Brenner, S. (1974). The genetics of *Caenorhabditis elegans*. Genetics *77,* 71-94.
9. Brenner, S. (1988). In The Nematode Caenorhabditis elegans (ed. W. Wood). Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
10. *Caenorhabditis elegans* Sequencing Consortium. (1998). Genome sequence of the nematode *C. elegans:* a platform for investigating biology. Science *282(5396),* 2012-2018.
11. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W., and Prasher, D.C. (1994). Green fluorescent protein as a marker for gene expression. Science *263(5148)*, 802-805.
12. Cheung, B.H.H., Arellano-Carbajal, F., Rybicki, I. and De Bono, M. (2004). Soluble guanylate cyclases act in neurons exposed to the body fluid to promote *C. elegans* aggregation behavior. Curr. Biol. *14,* 1105-1111.
13. Chiu, H., Schwartz, H. T., Antoshechkin, I., and Sternberg, P. W. (2013). Transgene-free genome editing in*Caenorhabditis elegans* using CRISPR‐Cas. Genetics *195(3):*1167-1171.
14. Colbert, H.A., and Bargmann, C.I. (1995). Odorant-specific adaptation pathways generate olfactory plasticity in *C. elegans*. Neuron *14,* 803 –812.
15. Edgley, M., D'Souza, A., Moulder, G., McKay, S., Shen B., Gilchrist, E., Moerman, D., and Barstead, R. (2002). Improved detection of small deletions in complex pools of DNA. Nucleic Acids Res. *30:* e52
16. Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature *391*, 806–811.
17. Gally, C., Eimer, S., Richmond, J.E. and Bessereau, J.L. (2004). A transmembrane protein required for acetylcholine receptor clustering in*Caenorhabditis elegans*. Nature *431,* 578-582.
18. Gengyo-Ando, K., and S. Mitani. (2000). Characterization of mutations induced by ethyl methanesulfonate, UV, and trimethylpsoralen in the nematode Caenorhabditis elegans. Biochem. Biophys. Res. Commun. *269:* 64–69.
19. Gerstein, M.B., Lu, Z.J., Van Nostrand, E.L., and Cheng, C. et al., (2000). Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE

project. Science. 330(6012), 1775-1787.
20. Gray J.M., Karow, D.S., Lu, H., Chang, A.J., Chang, J.S., Ellis, R.E., Marletta, M A. and Bargmann, C I. (2004). Oxygen sensation and social feeding mediated by a C. elegans guanylate cyclase homologue. Nature *430,* 317-322.
21. Hardaker, L.A., Singer, E., Kerr, R., Zhou, G. and Schafer, W.R. (2001). Serotonin modulates locomotory behavior and coordinates egg-laying and movement in *Caenorhabditis elegans*. J. Neurobiol. *49,* 303-313.
22. Hobert, O. (2010). Neurogenesis in the nematode *Caenorhabditis elegans*. WormBook ed. The C. elegans Research Community:doi/10.1895/wormbook, http://www.wormbook.org.
23. Jansen, G., Hazendonk E., Thijssen, K.L., and Plasterk, R.H. (1997). Reverse genetics by chemical mutagenesis in Caenorhabditis elegans. Nat. Genet. *17:* 119–121.
24. Jarrell, T.A., Wang, Y., Bloniarz, A.E., Brittin, C.A., Xu, M., Thomson, J.N., Albertson, D.G., Hall, D.H,, Emmons, S.W. (2012). The connectome of a decision-making neural network. Science *337(6093)*, 437-444.
25. Jeong P.Y., Jung M., Yim, Y.H., Kim, H., Park, M., Hong, E., Lee, W., Kim, Y.H., Kim, K. and Paik, Y.K. (2005). Chemical structure and biological activity of the *Caenorhabditis elegans* dauer-inducing pheromone. Nature *433,* 541-545.
26. Kiontke, K., Barriere, A., Kolotuev, I., Podbilewicz, B., Sommer, R., Fitch, D.H. and Felix, M.A. (2007). Trends, stasis, and drift in the evolution of nematode vulva development. Curr. Biol. *17,* 1925-1937.
27. Kuntz, S.G., Schwarz, E.M., DeModena, J.A., De Buysscher, T., Trout, D, Shizuya, H., Sternberg, P.W., and Wold, B.J. (2008). Multigenome DNA sequence conservation identifies Hox cis-regulatory elements. Genome Res *18(12),* 1955-1968.
28. Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J., Dulay, G.P., Hua, K.L., Ankoudinova, I., Cost, G.J., Urnov, F.D., Zhang, H.S., Holmes, M.C., Zhang, L., Gregory, P.D., Rebar, E.J. (2011). A TALE nuclease architecture for efficient genome editing. Nat. Biotechno. *29(2),* 143-148.
29. Rankin, C.H., Chiba, C., and Beck, C. (1990). Caenorhabditis elegans: A new model system for the study of learning and memory. Behav Brain Res *37,* 89 –92.
30. Riddle, D.L. and Golden, J.W. 1982. A pheromone influences larval development in the nematode *C. elegans*. Science *218,* 578-580.
31. Riddle, D.L., Blumenthal, T., Meyer B.J., and Priess, J.R. (1997). *C. elegans* II, 2nd edition. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.
32. Rushforth, A. M., Saari, B., and Anderson, P. (1993). Site-selected insertion of the transposon Tc1 into a Caenorhabditis elegans myosin light chain gene. Mol. Cell. Biol. *13:* 902–910.
33. Sawin, E.R., Ranganathan, R. and Horvitz, H.R. (2000). *C. elegans* locomotory rate is modulated by the environment through a dopaminergic pathway and by experience through a serotonergic pathway. Neuron *26,* 619-631.
34. Schwarz, E.M. (2005). Genomic classification of protein-coding genes. *WormBook*, ed. The *C. elegans* Research Community, WormBook, doi/10.1895/wormbook.1.7.1, http://www.wormbook.org.

35. Schwarz, E.M., Kato,M., Sternberg, P.W. (2012). Functional transcriptomics of a migrating cell in *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. U.S.A. *109(40),* 16246-16251.
36. Spencer WC, Zeller G, Watson JD, Henz SR, Watkins KL, McWhirter RD, Petersen S, Sreedharan VT, Widmer C, Jo J, Reinke V, Petrella L, Strome S, Von Stetina SE, Katz M, Shaham S, Rätsch G, Miller DM III. 2011. A spatial and temporal map of *C. elegans* gene expression. Genome Res. *21,* 325-341.
37. Sulston, J.E., Horvitz, H.R. (1977). Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. Dev. Biol. *56(1),* 110-156.
38. 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.
39. Ward, S., Thomson, J., White, J. and Brenner, S. 1975. Electron microscopical reconstruction of the anterior sensory anatomy of the nematode*C. elegans*. J. Comp. Neurol. *160,* 313-337.
40. White, J.G., Southgate, E., Thomson, J.N. and Brenner, S. (1986). The structure of the nervous system of the nematode *C. elegans*. Philos. Trans. R. Soc. Lond. Series B. Biol. Sci. *314,* 1-340.
41. You, Y.J., Kim, J., Raizen, D.M., and Avery L. (2008). Insulin, cGMP, and TGF-beta signals regulate food intake and quiescence in *C. elegans*: a model for satiety. Cell Metab. *7(3),* 249-257.
42. Zhang, Y., Lu, H. and Bargmann, C.I. (2005). Pathogenic bacteria induce aversive olfactory learning in *Caenorhabditis elegans.* Nature *438,* 179-184.

Chapter 2

Towards the translation of genetic codes to behavioral outputs:

the ALA neuron and *C. elegans* sleep

Abstract

Small as it appears, the *C. elegans* nervous system is complex in function and diverse in cell types. Individual neurons are highly wired in functional circuits and modules. Choosing the right neuron is crucial for studying single-neuron functional genomics, and there are a few prerequisites: easy identification by morphology, existence of tools such as cell-specific promoter::*gfp* markers to label the neuron, minimal synaptic connections, and known and robust behavioral readout. The head interneuron, ALA, fulfills all criteria and distinguishes itself by the availability of promoter::*gfp* markers for cell recognition. It has one pair of axons connecting to neurons in the tail, making it easy for isolation without contamination of nearby connected neurons. The ALA neuron is required for maintaining locomotor quiescence during a *C. elegans* sleep-like state and is capable of inducing sleep behavior at any time of the worm's life post-hatching. The characterized transcriptional inputs for ALA development are an added bonus for genomic analysis. This chapter introduces the ALA neuron, the function of ALA in mediating *C. elegans* sleep behavior, and the hypothesis of a molecular mechanistic regulation in sleep elicited by the ALA neuron.

The ALA neuron

The ALA neuron is clearly visible under differentiation interference microscopy. The neuron resides dorsal to the pharyngeal isthmus anterior to the posterior pharyngeal terminal bulb (Figure 2.1). Electron microscopy analysis revealed that ALA is a petidergic interneuron with just a few synaptic connections and gap junctions (White et al., 2006). The neuron sends out two processes branching from the anterior portion and

running along the lateral cord to the tail (Figure 2.2). Unlike the majority of neurons in the *C. elegans* nervous system, the ALA neuron is unpaired (White et al., 1986).

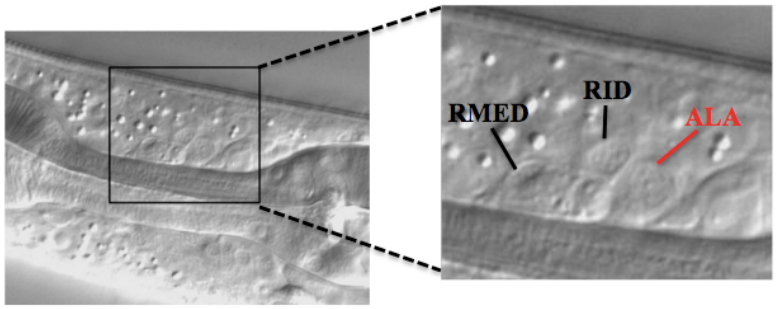


Figure 2.1. Shown here are three neuronal nuclei in the central plane doral to the isthmus of the pharynx: the anterior-most cell is RMED, the middle cell is RID, and the posterior cell is ALA, which is also the largest of the three. Adopted and modified from Yochem, 2006.

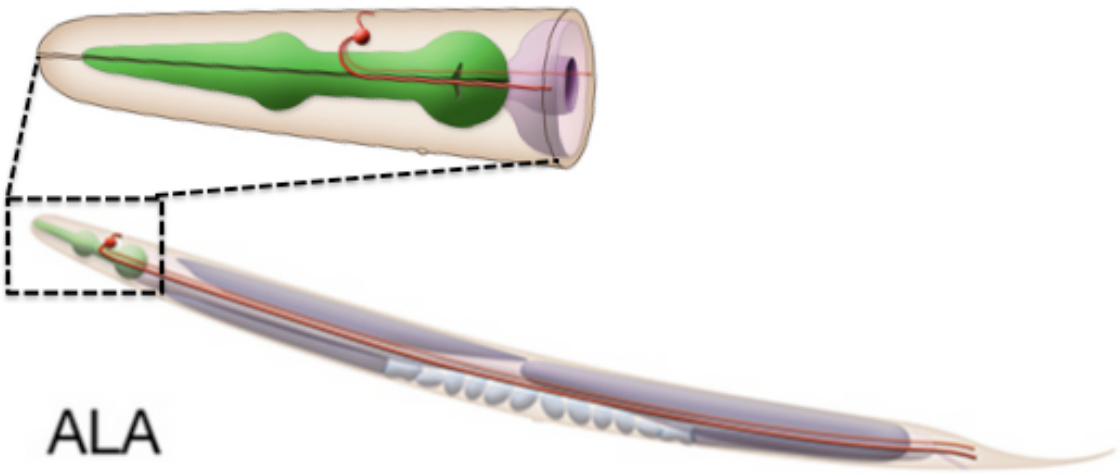


Figure 2.2. ALA has two processes that branch from the anterior portion of the cell body, project into the left or right side of the nerve ring, and then migrate as far as the tail on the lateral cords, adjacent to excretory canals. ALA also sends a process along dorsal cord. Adopted from WormAtlas.

Functions of ALA in *C. elegans*

ALA is one of the less understood neurons in the *C. elegans* nervous system. Categorized as an interneuron, ALA is expected to process and integrate information inputs from other classes of neurons, either motor or sensory, and relays the message to other neurons in the circuit to elicit a unified action response (Altun and Hall, 2011). At present, only a handful of genes are known to express in ALA, including a neuropeptide-coding gene, *flp-7* (Li and Kim, 2008), three transcription factor-coding genes, *ceh-10, ceh-14,* and *ceh-17* (Van Buskirk and Sternberg, 2007), and a few membrane receptor-coding genes, *sra-10* (Troemel et a., 1995), *des-2* and the ortholog of epidermal growth factor receptor-coding gene, *let-23* (Van Buskirk and Sternberg, 2007). The lack of genomic information sets the barrier to understand the function of ALA.

The one known and robust function of ALA is its ability to induce sleep behavior in *C. elegans.* Van Buskirk and Sternberg (2007) overexpressed the EGFR ligand with a heat shock inducible promoter, and observed locomotor and feeding quiescence in normally active young adult worms (Van Buskirk and Sternberg, 2007). This behavior mimics a sleep-like state in *C. elegans*, known as lethargus, when the animal ceases voluntary movements such as locomotion and feeding (Van Buskirk and Sternberg, 2007; Raizen et al., 2008). Further characterization identified expression of the sole EGFR ortholog in *C. elegans, let-23,* in ALA as the receptor of overexpressed EGF ligand for sleep induction (Van Buskirk and Sternberg, 2007). Furthermore, EGFR signaling is one of the components to maintain locomotor quiescence in lethargus (Van Buskirk and Sternberg, 2007). Similar to the necessity of a functional LET-23, the presence of ALA is required to maintain locomotion quiescence in *C. elegans* lethargus (Van Buskirk and

Sternberg, 2007). Interestingly, the axons of ALA are dispensable for mediating sleep induction, indicating that the ALA neuron discharges neuropeptides and neurotransmitters to induce sleep (Van Buskirk and Sternberg, 2007).

Neuropeptides in *C. elegans*

Neuropeptides are small signaling molecules that modulate synaptic connectivity throughout the animal kingdom (Kow and Pfaff, 1988; Li and Kim, 2008). These are short sequences of amino acids derived from large precursor genes as a result of post-translational processing, and sometimes modification such as amidation, to become fully active (Figure 2.3, Li and Kim, 2008). A neuropeptide precursor gene may give rise to multiple identical or different mature neuropeptides, which are then packaged inside dense core vesicles as they are transported to the nerve terminal (Strand, 1999). Increase of calcium level throughout the nerve terminal triggers release of neuropeptides from dense core vesicles (Strand, 1999; Salio et al., 2006). Neuropeptides act through G-protein coupled receptors (GPCRs), and one neuropeptide may bind to multiple receptors, making it difficult to discern the function of specific neuropeptide (Bargmann, 1998).

The existence of neuropeptides in *C. elegans* was inferred by the observation of dense core vesicles in the nervous system in electron microscope images (White et al., 1986). In general, neuropeptides act as neuromodulators and fast neurotransmitters to regulate animal behaviors, and this is no exception in *C. elegans* (Bargmann, 1998). In fact, neuropeptides are extensively utilized in neuromodulators for *C. elegans* locomotion, reproduction, social behavior, foraging, mechano- and chemosensation, learning, and memory (Li and Kim, 2008).

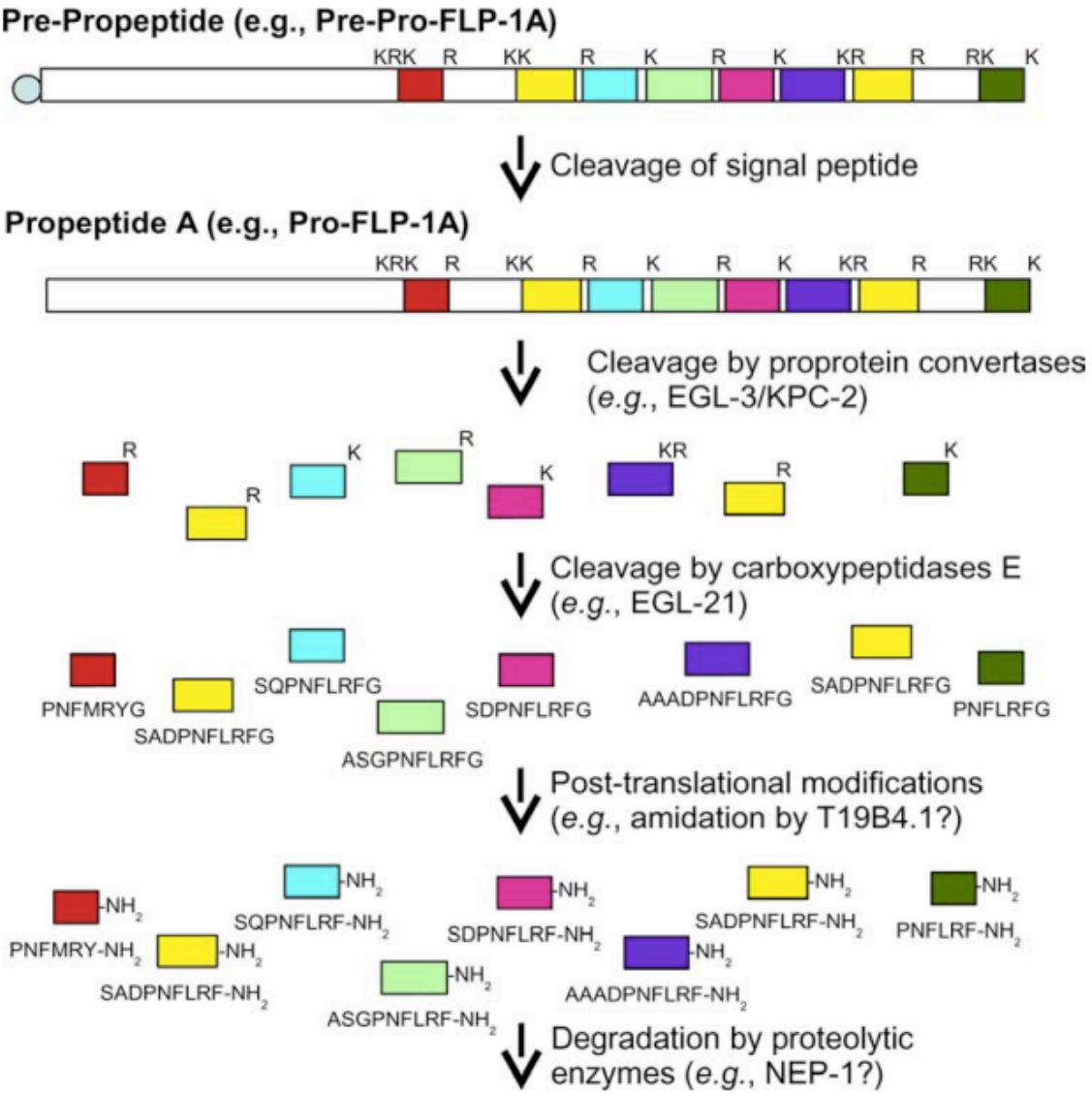


Figure 2.3. Processing of a neuropeptide gene product: flp-1 as an example. After translation of the flp-1A transcript, pre-pro-FLP-1A is cleaved by signal peptidase in the endoplasmic reticulum to release the signal peptide. The propeptide pro-FLP-1A is further cleaved C-terminal to mono-, di-, or tribasic residues (indicated by K and R) by proprotein convertases, such as EGL-3/KPC-2. The basic amino acids are removed by carboxypeptidases E, such as EGL-21, to yield the basic neuropeptides. The FLP-1A peptides are further modified by the addition of an amide group donated from the C-terminal glycine, a reaction which may be catalyzed by T19B4.1, to yield the active peptides. After release, peptides are removed from the synaptic cleft by proteolytic degradation, which may be mediated by NEP-1. With the exception of PNFMRFYamide, all flp-1A encoded peptides have been biochemically isolated (Li, 2005). Adopted and unmodified from Li and Kim, 2008.

The function of neuropeptide in promoting wakefulness was observed in mammals (Sakurai, 2007) and in zebrafish (Prober et al., 2006). Neuropeptide Y (NPY) possesses duo functions in promoting animal sleep and wakefulness, depending on the site and vehicle of NPY introduction (Antonijevic et al., 2000; Dyzma et al., 2010). The *Drosophila* NPY ortholog, sNPF neuropeptide, plays a role in regulating sleep homeostat (Shang et al., 2013). Despite all this, a conserved sleep promoting neuropeptide system has not been reported.

In *C. elegans,* it is proposed that the syntaxin regulator *unc-13* and the calcium dependent activator protein for serotonin (CAPS) *unc-31* mediate movement of dense core vesicle to the cell membrane (Richmond et al., 1999; Sieburth et al., 2007; Renden et al., 2001; Grishanin et al., 2002). Mutants of *unc-13* and *unc-31* do not induce sleep, presumably due to their failure to package and transport sleep-promoting neuropeptides and neurotransmitters in ALA (Van Buskirk and Sternberg, 2007).

The only known neuropeptide in *C. elegans* ALA is a FMRFamide-like neuropeptide encoded by, *flp-7,* which does not induce sleep (Van Buskirk and Sternberg, 2007). It is estimated that at least hundreds if not thousands of genes are expressed in a neuron to define the specific functionality of a neuron (Hobert, 2012). There are 8,011 protein-coding genes expressed in the somatic male linker cell (Schwarz et al., 2012), ALA is anticipated to express similar number of protein-coding genes if not more. Moreover, the related nematode, *Ascaris suum*, has at least 7 orthologous *flp* genes expressed in ALA (Jarecki et al., 2010). Given that there are 122 neuropeptide-coding genes in the *C. elegans* genome (Hobert, 2012), it is possible that other unknown neuropeptides and functional genes expressed in ALA have the ability to induce sleep.

Development of ALA

Transgenic reporter analysis and imaging showed that ALA expresses three transcription factors, including *ceh-10* (Wu et al, 2011) and *ceh-17* (Pujol et al, 2000) in the Paired-like homeodomain class, and *ceh-14* (Cassata et al, 2000) in the LIM homeodomain class families. Van Buskirk and Sternberg (2010) showed that these transcription factors are essential for the generation and differentiation of ALA in a combinatorial and temporal fashion (Van Buskirk and Sternberg 2010). Moreover, absence or reduction of these regulatory genes abolishes expression of component genes in the EGFR pathway, as well as other known ALA-expressed differentiation genes. As a result, these mutants lost their ability to respond to the ALA-induced sleep effect (Van Buskirk and Sternberg 2010). But how might such genetic information translate into functional readout, i.e., what are the local genomic context and gene regulatory logic employed to elicit the sleep-inducing ability of the ALA neuron?

Differentiation genes, the genes that give rise to the functional characteristics of a cell, are often regulated by cell-specific motifs (Hobert, 2008). These motifs contain regulatory binding sites of transcription factors that read them. Genes regulated by a common motif are often co-expressed, and may share the same function (Hobert, 2008). The *C. elegans* nervous system appears to fit into this model. Over the years, several motifs have been identified in *C. elegans,* including a cell type-specific motif (Wenick and Hobert, 2004), a neuronal subtype-specific motif (Kratsios et al, 2012; Zhang et al, 2014), and some neuron function-specific motifs (Doitsidou et al, 2013). Therefore, identifying a cell-specific motif in ALA is essential for sorting out genes that function in the same pathway.

Connecting the dots towards the sleep state

The sleep-inducing property and the morphological features differentiate ALA from the rest of the nervous system for single-neuron functional analysis. ALA induces *C. elegans* sleep through the EGFR signaling pathway (Van Buskirk and Sternberg, 2007) and I hypothesize that neuropeptides, as well as other differentiation genes, are involved in the sleep-inducing pathway. I propose to conduct an unbiased transcriptome profiling to catalog the protein-coding genes expressed in ALA. Deciphering their regulatory interactions may shed light on understanding how genetic codes are translated into behavioral outputs, and provide an entry point to decode the genetic players in the connectome that governs animal behaviors.

Reference

1. Altun, Z.F. and Hall, D.H. 2009. Introduction. In *WormAtlas*.
2. Antonijevic, I.A., Murck, H., Bohlhalter, S., Frieboes, R.M., Holsboer, F., and Steiger, A. (2000). Neuropeptide Y promotes sleep and inhibits ACTH and cortisol release in young men. Neuropharmacology *39(8),* 1474-1481.
3. Doitsidou M, Flames N, Topalidou I, Abe N, Felton T, Remesal L, Popovitchenko T, Mann R, Chalfie M, Hobert O. 2013. A combinatorial regulatory signature controls terminal differentiation of the dopaminergic nervous system in *C. elegans*. *Genes Dev* **27(12):** 1391-1405.
4. Dupuy D, Bertin N, Hidalgo CA, Venkatesan K, Tu D, Lee D, Rosenberg J, Svrzikapa N, Blanc A, Carnec A, Carvunis AR, Pulak R, Shingles J, Reece-Hoyes J, Hunt-Newbury R, Viveiros R, Mohler WA, Tasan M, Roth FP, Le Peuch C, Hope IA, Johnsen R, Moerman DG, Barabási AL, Baillie D, Vidal M. 2007. Genome-scale analysis of in vivo spatiotemporal promoter activity in Caenorhabditis elegans. *Nat Biotechnol* **25(6):** 663-668.
5. Dyzma, M., Boudjeltia, K.Z., Farut, B., and Kerkhofs M. (2010). Neuropeptide Y and sleep. Sleep Med. Rev. *14(3),* 161-165.
6. Grishanin, R.N., Klenchin, V.A., Loyet, K.M., Kowalchyk, J.A., Ann, K., and Martin, T.F. (2002). Membrane association domains in Ca2+-dependent activator protein for secretion mediate plasma membrane and dense-core vesicle binding required for Ca2+-dependent exocytosis. J. Biol. Chem. *277,* 22025–22034.
7. Hobert O. 2008. Regulatory logic of neuronal diversity: Terminal selector genes and selector motifs. *Proc Natl Acad Sci U.S.A.* **105:** 20067-20071.
8. Hobert, O. (2012). Neurogenesis in the nematode *Caenorhabditis elegans.* WormBook ed. The C. elegans Research Community:doi/10.1895/wormbook, http://www.wormbook.org.
9. Jarecki, J.L., Andersen, K., Konop, K.J., Knickelbine, J.J., Vestling, M.M., and Stretton, A.O. (2000). Mapping neuropeptide expression by mass spectrometry in single dissected identified neurons from the dorsal ganglion of the nematode *Ascaris suum*. ACS Chem. Neurosci. *1,* 505-519.
10. Kow, L.M., and D. W. (1988). Pfaff, D.W. Neuromodulatory actions of peptides. *Ann. Rev. Pharmacol. Toxico. 28,* 163
11. Kratsios P, Stolfi A, Levine M, Hobert O. 2011. Coordinated regulation of cholinergic motor neuron traits through a conserved terminal selector gene. *Nat Neurosci* **15(2):** 205-214.
12. Li, C. (2005). The ever-expanding neuropeptide gene families in the nematode Caenorhabditis elegans. Parasitology *131* Suppl. S109–S127.
13. Li, C., and Kim, K. (2008). Neuropeptides. WormBook ed. The C. elegans Research Community:doi/10.1895/wormbook, http://www.wormbook.org.
14. Prober, D.A., Rihel, J., Onah, A.A., Sung, R.J., and Schier, A.F. (2006). Hypocretin/orexin overexpression induces an insomnia-like phenotype in zebrafish. J. Neurosci. *26(51),* 13400-13410.

15. Pujol N, Torregrossa P, Ewbank JJ, Brunet JF. 2000. The homeodomain protein CePhox2/CEH-17 controls antero-posterior axonal growth in *C. elegans. Development* **127:** 3361-3371.
16. Raizen, D.M., Zimmerman, J.E., Maycock, M.H., Ta, U.D., You, Y.J. Sundaram, M.V., and Pack, A.I. (2008). Lethargus is a *Caenorhabditis elegans* sleep-like state. Nature *451(7178),* 569-572.
17. Renden, R., Berwin, B., Davis, W., Kyoungsook, A., Chin, C.-T., Kreber, R., Ganetzky, B, Martin, T.F.J., and Broadie, K. (2001). Drosophila CAPS is an essential gene that regulates dense-cre vesicle release and synaptic vesicle fusion. Neuron *31,* 421–437.
18. Richmond, J.E., Davis, W.S, and Jorgensen, E.M. (1999). UNC-13 I srequired for synaptic vesicle fusion in *C. elegans.* Nat. Neurosci. *2,* 959–964.
19. Sakurai, T. (2007). The neural circuit of orexin (hypocretin): maintaining sleep and wakefulness. Nat. Rev. Neurosci. *8(3),*171-181.
20. Salio, C., Lossi, L., Ferrini, F., and Merighi, A. (2006). Neuropeptides as synaptic transmitters. Cell Tissue Res. *326,* 583–598.
21. Schwarz EM, Kato M, Sternberg PW. 2012. Functional transcriptomics of a migrating cell in *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A.* **109(40):** 16246-16251.
22. Shang, Y., Donelson, N.C., Vecsey, C.G, Guo, F., Rosbash, M., and Griffith, L.C. (2013). Short neuropeptide F is a sleep-promoting inhibitory modulator. Neuron *80(1),* 171-183.
23. Sieburth, D., Madison, J.M., and Kaplan, J.M. (2007). PKC-1 regulates secretion of neuropeptides. Nat. Neurosci.*10,* 49–57.
24. Strand, F.L. (1999). Neuropeptides. MIT Press, Cambridge, MA.
25. Troemel, E.R., Chou, J.H., Dwyer, N.D., Colbert, H.A., and Bargmann, C.I. (1995). Divergent seven transmembrane receptors are candidate chemosensory receptors in *C. elegans.* Cell *83(2),* 207-218.
26. Van Buskirk C, Sternberg PW. 2007. Epidermal growth factor signaling induces behavioral quiescence in *Caenorhabditis elegans. Nat Neurosci* **10(10):** 1300-1307.
27. Van Buskirk C, Sternberg PW. 2010. Paired and LIM class homeodomain proteins coordinate differentiation of the *C. elegans* ALA neuron. *Development* **137:** 2065-2074.
28. Wenick AS, Hobert O. 2004. Genomic cis-regulatory architecture and trans-acting regulators of a single interneuron-specific gene battery in *C. elegans. Dev Cell* **6(6):** 757-770.
29. White, J.G., Southgate, E., Thomson, J.N. and Brenner, S. (1986). The structure of the nervous system of the nematode *C. elegans*. Philos. Trans. R. Soc. Lond. Series B. Biol. Sci. *314,* 1-340.
30. Wu, Y., Ghitani, A., Christensen, R., Santellam A., Du, Z., Rondeau, G., Bao, Z., Colón-Ramos, D., and Shroff, H. (2011). Inverted selective plane illumination microscopy (iSPIM) enables coupled cell identity lineaging and neurodevelopmental imaging in Caenorhabditis elegans. Proc Natl Acad Sci U S A. *108(43),* 17708-17713.

31. Yochem, J. (2005). Nomarski images for learning the anatomy, with tips for mosaic analysis. WormBook ed. The C. elegans Research Community:doi/10.1895/wormbook, http://www.wormbook.org.

Chapter 3

Sleep -

What do we know about it?

Abstract

General understanding of sleep is that it is a prolonged period of motor inactivity and slow response to external environmental changes such as chemical and mechanical stimulation. Because of this seemingly unproductive and vulnerable state, animals are more prone to the danger of being caught by predators, and yet sleep is a behavior observed across the animal kingdom, spanning from worms to humans (Allada and Siegal, 2008; Zimmerman et al, 2008). Such phenomenon highlights the importance of sleep, in that the benefits it offers outweighs the risk of death, and indicates that the need of sleep is an internal drive and is perhaps indispensable for survival. Despite a growing community in sleep research, there remain two fundamental questions regarding sleep that go unanswered: i) What is the function of sleep? ii) What is the molecular regulation of sleep? This chapter introduces current understandings of sleep, defines the definition of sleep, summarizes known molecular players regulating sleep, and discusses approaches to facilitate sleep research.

Definition of sleep

Sleep was initially defined on the basis of electroencephalograms (EEGs), recordings that reflect cortical electrical activity alterations (Sehgal and Mignot, 2011). During the sleep period, three states of behavior are found in EEG: wake, rapid eye movement (REM) sleep, and non-REM (NREM) sleep. In humans, a full night's sleep is composed of 90 minute cycles of REM and NREM sleep. NREM sleep is further divided into stages 1-3, where the deepest sleep occurs in stage 3 NREM sleep (Sehgal and Mignot, 2011). While the measurement of cortical electrical activity is informative for

mammalian and avian models, it precludes sleep research in other models such as fish, reptile, and bees, all of which do not present a well-defined cortex but exhibit a sleep-like state (Campbell and Tobler, 1984). Moreover, this approach is not practical for high-throughput screening and is laborious for day-to-day experiments (Sehgal and Mignot, 2011). A more efficient and economic approach is to use simple animal models and behavioral assays to measure a sleep-like state originally proposed by Campbell and Tobler (1984). This shifted the paradigm of sleep definition. The current definition of sleep behavior consists of four criteria: i) a prolonged but reversible period of voluntary movement inactivity, ii) increased arousal threshold for response to sensory stimulants, iii) regulation by homeostasis, and iv) a circadian clock control (Sehgal and Mignot, 2011).

Function of sleep

Knowledge and recognition of the need of sleep come from complaints of sleep disorders, which are extremely common (Mahowald and Schenck, 2005). There are more than 100 identified sleep disorders, and most of them fall into four categories: hypersomnia (i.e. excessive day-time sleep without obvious explanation, e.g. narcolepsy, obstructive sleep apnoea), insomnia (i.e. trouble falling and staying asleep, e.g. restless leg syndrome), circadian rhythm disorders (e.g. delayed sleep syndrome), and parasomnias (i.e. complex behaviors arising from the sleep period, e.g. sleep walking, sleep terrors).

Despite the prevalence of sleep disorders in the past decades, little is known about the purpose of sleep. Several lines of studies suggest that sleep promotes synaptic

plasticity and supports cognitive function (Diekelmann and Born, 2010; Poe et al., 2010). Others observed association of regulation of sleep and wakefulness with regulation of cerebral energy stores (Benington, et la., 1995). Insights from studies in sleep-deprived mice and flies suggest that sleep has a role in curbing stress, and that the need for sleep is influenced by cellular stress (Naidoo et al., 2007). More recently, using real-time two-photon imaging in mice, Lulu et al., (2013) showed that a critical function of sleep is to clear metabolite from the brain and to maintain metabolic homeostasis.

Molecules regulating sleep

Research in model organisms revealed that sleep is genetically regulated and is evolutionarily conserved among animals (Sehgal and Mignot, 2011). Much discoveries in sleep regulation come from the identification of neuropeptides and neurotransmitters, as well as characterization of intracellular signaling molecules that are essential for regulating sleep/wakefulness. Other molecules include ion channels and channel-regulating proteins, circadian clock genes, metabolic factors, and immune genes (Sehgal and Mignot, 2011). This study will focus on the roles of neuropeptides and intracellular signaling pathway on sleep regulation.

Neuropeptide/receptor systems and neurotransmitters regulating sleep

In mammals and other vertebrates, numerous neuropeptide/receptor systems have been reported to modulate sleep and wakefulness, the best known being the hypocretins/orexins in mammals and fish (Raffa, 1988; Prober et al., 2006; Sakurai, 2007). Hypocretins are wake-promoting neuropeptides that underlie the sleep disorder

narcolepsy (Taheri et al., 2002; Sehgal and Mignot, 2011). Orthologs of hypocretins in flies and worms have yet to be reported. However, the neuropeptide pigment-dispersing factor (PDF) is a wake-promoting peptide and functions in an analogous fashion in fles (*Drosophila melanogaster)* by inhibiting sleep-promoting neurotransmitters such as gamma-aminobutyric acid (GABA; Parisky et al., 2008). Consistent with this view, the wake promoting property of *C. elegans* PDF-1/PDFR system is shown by its ability in governing exit of locomotor quiescence in lethargus (Choi et al., 2013). Other neurotransmitters implicated in sleep regulation are the wake-promoting histamine, dopamine, acetylcholine, and norepinephrine, as well as sleep-promoting serotonin and adenosine (Sehgal and Mignot, 2011).

Intracellular signaling molecules regulating sleep

As the downstream target of neuropeptides and neurotransmitters, intracellular signaling pathways also take part in regulating sleep and wakefulness. The mammalian CREB pathway promotes wakefulness (Graves et al., 2003; Hendricks et al., 2001), whereas *Drosophila* protein kinase A (PKA)/CREB pathway plays duo roles in a site-dependent manner. Pan-neuronal expression of PKA promotes wakefulness, but PKA expression in specific subsets of neurons promotes sleep (Joiner et al., 2006). The cyclic guanosine monophosphate (cGMP) kinase regulates sleep in mammals (Langmesser et al., 2009) and promotes sleep in *C. elegans* and *Drosophila* (Raizen et al., 2008). Lastly, the epidermal growth factor receptor (EGFR) signaling pathway promotes sleep in worms, flies, and mammals. Intracerebroventricular injection of EGF increases NREM sleep in rabbits (Kushikata et al., 1998), and has been shown to inhibit locomotor activity in mice

(Kramer et al., 2001). Overexpression of the EGFR ligand in *Drosophila* increases sleep, and this effect is dependent on a functional EGFR and, at least in part, mediated by the extracellular signal-regulated kinases/mitogen-activated protein kinase (ERK/MAPK) pathway (Foltenyi et al., 2007). *C. elegans* has one EGFR, LET-23, and reduction of this receptor as a result of loss-of-function mutation increases locomotor activity in the sleep-like state of worms (Van Buskirk and Sternberg, 2007). On the contrary, overexpression of the sole EGFR ligand, LIN-3C, activates EGFR signaling through diacylglycerol and phospholipase C-γ to induce a sleep-like state where worms cease locomotor and feeding activities (Van Buskirk and Sternberg, 2007).

What remains to be explored?

Research on sleep had made enormous progress in the past decades due to the utilization of simple animal model organisms (Sehgal and Mignot, 2011). Whether the function of sleep is for building synaptic plasticity or for metabolite clearance, the genetic basis of sleep regulation remains elusive. With the paradigm of sleep definition shifted from measuring EEGs to directly measuring rest/active behavior, research using model organisms will continue to yield insights in sleep regulation. Novel sleep components can be found by high-throughput screening, such as explorative whole-genome sequencing, and the well-established genetic tools that provide experimental validation.

References

1. Allada, R., and Siegel, J.M. (2008). Unearthing the phylogenetic roots of sleep. Curr. Biol. *18,* R670-R679.
2. Benington, J.H., and H.C. Heller. (1995). Restoration of brain energy metabolism as the function of sleep. Prog. Neurobiol. *45,* 347-360.
3. Campbell, S.S., and Tobler, I. (1984). Animal sleep: a review of sleep duration across phyogeny. Neurosci. Biobehav. Rev. *8,* 269-300.
4. Choi, S., Chatzigeorgiou, M., Taylor, K.P., Schafer, W.R., and Kaplan, J.M. (2013). Analysis of NPR-1 reveals a circuit mechanism for behavioral quiescence in *C. elegans*. Neuron *78(5),* 869-880.
5. Diekelmann, S., and Born, J. (2010). The memory function of sleep. Nat. Rev. Neurosci. *11,* 114-126.
6. Graves, L.A., Hellman, K., Veasey, S., Blendy, J.A., Pack, A.I., and Abel, T. (2003). Genetic evidence for a role of CREB in sustained cortical arousal. K. Neurophysiol. *90*, 1152-1159.
7. Hendricks, J.C., Williams, J.A., Panckeri, K., Kirk, D., Tello, M., Yin, J.C., and Sehgal, A. (2001). A non-circadian role for camp signaling and CREB activity in Drosophila rst homeostasis. Nat. Neurosci. *4,* 1108-1115.
8. Joiner, W.J., Crocker, A., White, B.H., Sehgal, A. (2006). Sleep in Drosophila is regulated by adult mushroom bodies. Nature *441(7094),* 757-760.
9. Kramer, A., Yang, F.C., Snodgrass, O., Li, X., Scammell, T.E., Davis, F.C., Weitz, C.J. (2001). Regulation of daily locomotor activity and sleep by hypothalamic EGF receptor signaling. Science *294(5551),* 2511-2515.
10. Kushikata, T., Fang, J., Chen, Z., Wang, Y., and Krueger, J.M. (1998). Epidermal growth factor enhances spontaneous sleep in rabbits. Am. J. Physiol. *275(2 Pt 2)*, R509-514.
11. Langmesser, S., Franken, P., Feil, S., Emmenegger, Y., Albrecht, U., and Feil, R. (2009). cGMP-dependent protein kinase type I is implicated in the regulation of the timing and quality of sleep and wakefulness. PLoS One *4(1),* e4238.
12. Mahowald, M.W., and Schenck., C.H. (2005). Insights from studying human sleep disorders. Nature *437,* 1279-1285.
13. Naidoo, N., Casiano, V., Cater, J., Zimmerman, J., and Pack., A.I. (2007). A role for the molecular chaperone protein BiP/GRP78 in Drosophila sleep homeostasis. Sleep *30,* 557-565.
14. Parisky, K.K., Agosto, J., Pulver, S.R., Shang, Y., Kuklin, E., Hodge, J.J., Kang, K., Liu, X., Garrity, P.A., Rosbash, M., and Griffith, L.C. (2008). PDF cells are a GABA-responsive wake-promoting component of the *Drosophila* sleep circuit. Neuron *60,* 672-682.
15. Poe, G.R., Walsh, C.M., and Bjorness, T.E. (2010). Cognitive neuroscience of sleep. Prog. Brain Res. *185*, 1-19.
16. Prober, D.A., Rihel, J., Onah, A.A., Sung, R.J., and Schier, A.F. (2006). Hypocretin/orexin overexpression induces an insomnia-like phenotype in zebrafish. J. Neurosci. *26(51),* 13400-13410.
17. Raffa, R.B. (1988). The action of FMRFamide (Phe-Met-Arg-Phe-NH2) and related peptides on mammals. Peptides *9(4),* 915-922.

18. Raizen, D.M., Zimmerman, J.E., Maycock, M.H., Ta, U.D., You, Y.J. Sundaram, M.V., and Pack, A.I. (2008). Lethargus is a *Caenorhabditis elegans* sleep-like state. Nature *451(7178),* 569-572.
19. Sakurai, T. (2007). The neural circuit of orexin (hypocretin): maintaining sleep and wakefulness. Nat. Rev. Neurosci. *8(3),*171-181.
20. Sehgal, A., and Mignot, E. (2011). Genetics of sleep and sleep disorders. Cell *146,* 194-207.
21. Taheri, S., Zeitzer, J.M., and Mignot, E. (2002). The role of hypocretins (orexins) in sleep regulation and narcolepsy. Annu. Rev. Neurosci. 25, 283–313.
22. Van Buskirk, C., and Sternberg, P.W. (2007). Epidermal growth factor signaling induces behavioral quiescence in *Caenorhabditis elegans.* Nat. Neurosci. *10(10),* 1300-1307.
23. Xie, L., Kang, H., Xu, Q., Chen, M.J., Liao, Y., Thiyagarajan, M., O'Donnell, J., Christensen, D.J., Nicholson, C., Iliff, J.J., Takano, T., Deane, R., and Nedergaard, M. (2013). Sleep drives metabolite clearance from the adult brain. Science *342(6156),* 373-377.
24. Zimmerman, J.E., Naidoo, N., Raizen D.M., and Pack, A.I. (2008). Conservation of sleep: Insights from non-mammalian model systems. Trends Neurosci. *31*, 371-376.

Chapter 4

Sleep in
Caenorhabditis elegans
and
Danio rerio

Abstract

A sleep-like state was recently characterized in two non-mammalian model organisms: the roundworm *Caenorhabditis elegans* and the zebrafish *Danio rerio* (Zhdanova et al., 2001; Prober et al., 2006; Yokogawa et al., 2007; Van Buskirk and Sternberg, 2007; Raizen et al., 2008). Both of these model organisms employ regulatory molecules similar to the mammalian models for sleep and wake promotion (see also Chapter 3). In addition, they are genetically traceable and are excellent for high-throughput behavioral screens. More importantly, the well-established genetic tools enable easy manipulation on genes for functional analysis. Herein I introduce sleep in *C. elegans* and zebrafish, as well as the behavioral essays for analyzing sleep in each organism, and discuss the potential advantages of corroborating insights gained from these two models for elucidating genes involved in sleep behavior.

Worm sleep

A sleep-like state in *C. elegans* was first described in lethargus, a quiescence behavioral state at the end of each larval stage (Raizen et al., 2008). Lethargus in worms possesses specific characteristics that fulfill the criteria for sleep, i.e. reversibility of quiescence, elevated arousal threshold to sensory stimuli, and homeostasis. Although lethargus is associated with the molting cycles rather than a 24 hr rhythm as in mammals, its timing is synchronized with the expression of *lin-42*, which governs the larval molts. The ortholog of *lin*-42 is *period* (Jeon et al., 1999), a circadian clock gene that regulates the timing of sleep in *Drosophila* and mammals. Therefore, lethargus is regulated by a circadian clock mechanism (Zimmerman et al., 2008; Sehgal and Mignot, 2011).

C. elegans sleep is not restricted to the larval stage (Figure 4.1). In fact, sleep-like states are observed in adults and can be induced by intracellular signaling pathways such as anachronistic expression of the EGFR signaling (Figure 4.2; Van Buskirk and Sternberg, 2007), and by satiety, through the TGFβ and cGMP pathways, in the presence of high-nutrient food or full feeding after a long period of starvation (You et al., 2008; Gallagher et al., 2013). More evidence on EGFR signaling or satiety induced sleep-like state comes from neuronal modulation analysis. Using optogenetic tools, Cho and Sternberg (2014) showed that a sleep-like state of *C. elegans* in response to both EGFR signaling activation and satiety share the same neuronal circuit modulation as in lethargus.

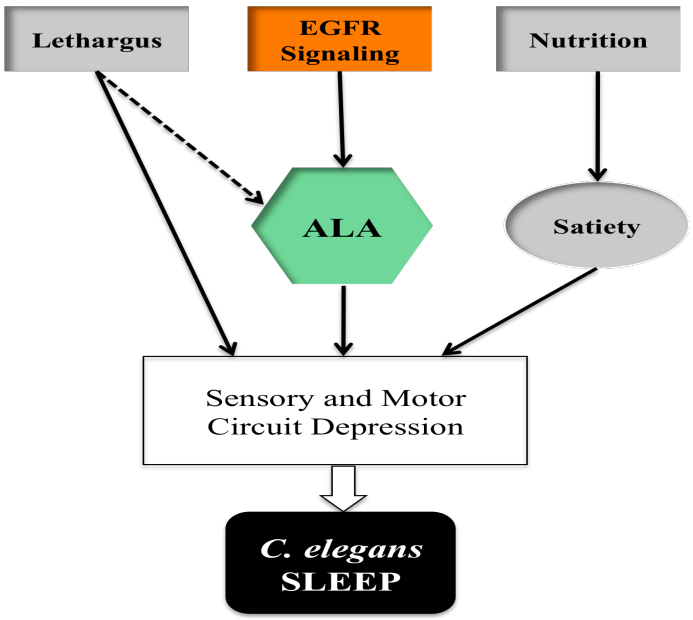


Figure 4.1. Model of sleep regulation in *C. elegans*. EGFR signaling promotes sleep through the ALA neuron, which is also a locomotor quiescence component in lethargus. High nutrition food induces satiety. All three pathways dampen sensory and motor circuit in a sleep-like state.

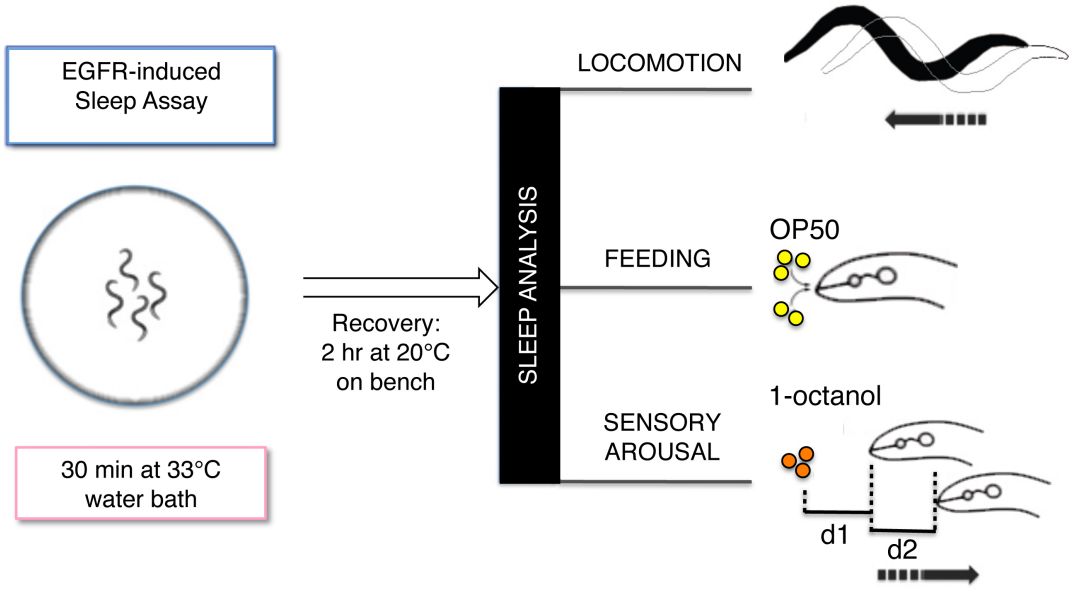


Figure 4.2. Schematic diagram of sleep assay to assess locomotion, feeding and sensory arousal behaviors. Worms were placed in a water bath for heat shock treatment (30 min at 33°C) and recovered at their growing temperature for 2 hr. Worms were examined for the sinusoidal movement as an indication for locomotion, pharyngeal pumping as an indicator for feeding, and rapid reversal movement when it withdraws from noxious smell like 1-octanol as an indication for sensory arousal.

Zebrafish sleep

Zebrafish is a vertebrate that has a sleep-like state, which is regulated by circadian rhythm and homeostasis, and has reduced sensory responsiveness (Chiu and Prober, 2013). Zebrafish rest-activity rhythm is synchronized with the day-night cycle. Similar to their diurnal cousins like human, zebrafish are active in the daytime and quiescent at night (Zimmerman et al., 2008, and references therein). Also, arousal state in zebrafish larvae can be characterized by changes in frequency and intensity of voluntary locomotor activity and changes in responsiveness to sensory or emotional stimuli (Pfaff et al., 2008). Zebrafish larval sleep or arousal can be measured by quantifying the locomotor behavior

using high-speed infrared video capture in conjunction with computational analysis for locomotor behaviors (Figure 4.3; Chiu and Prober, 2013).

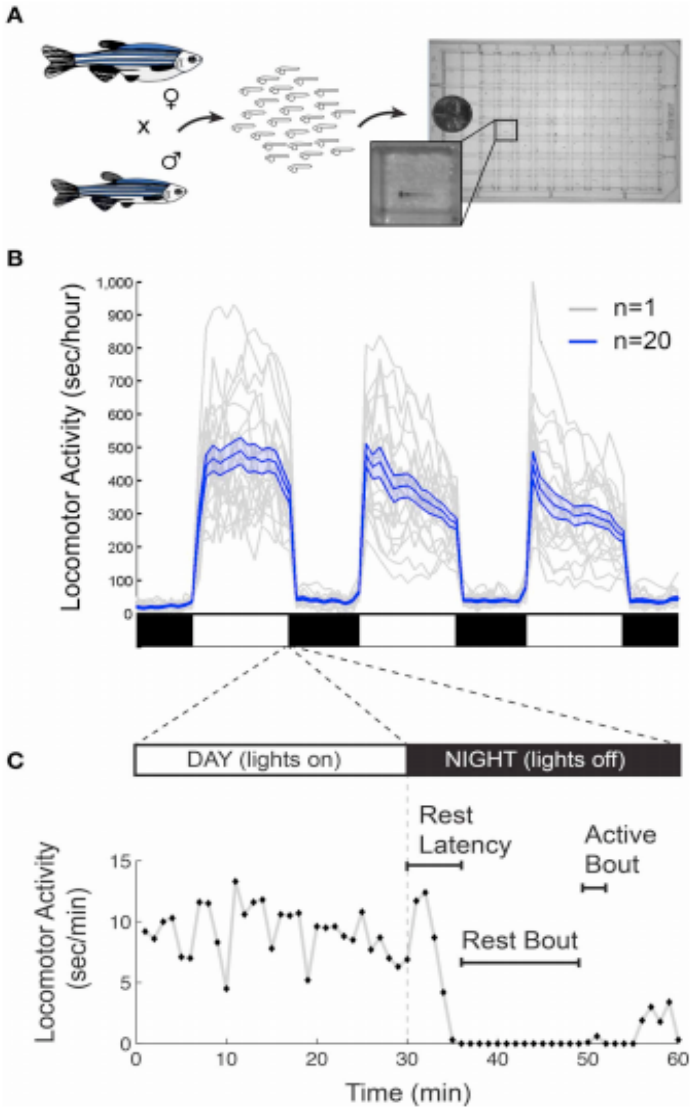


Figure 4.3. Monitoring larval zebrafish sleep and wake behavior. (A) Zebrafish larval locomotor activity assay. Individual zebrafish larva are placed in each well of a 96-well-plate on the 5th day of development. The plate is placed in a temperature-controlled chamber that is illuminated by white lights during the day and continuously illuminated by infrared lights. The larvae are monitored by an infrared camera and the locomotor activity of each larva is recorded by a computer. (B) Representative locomotor activity data for each of 20 individual wild-type larvae (gray traces) and their mean locomotor activity (blue trace, ± standard error bar of the mean) is shown. Black and white bars indicate day and night, respectively. Larvae are more active during the day than at night,

although there is considerable variability among individuals. (C) An example of typical larval zebrafish behavior at the end of the day is shown. A rest bout is defined as a period of at least 1 min of inactivity, which is associated with an increase in arousal threshold (Prober et al., 2006). Rest latency indicates the time between lights-off at night and initiation of the first rest bout. Adopted and unmodified from Chiu and Prober, 2013.

Zebrafish also show physiological and pharmacological characteristics of mammalian sleep (Chiu and Prober, 2013, and references therein). The hypothalamus-expressed neuropeptide hypocretin is the best-known and characterized sleep and arousal regulator in zebrafish. Overexpression of hypocretin using a heat shock promoter induces an insomnia-like behavior in zebrafish (Prober et. al., 2006), where the animals have increased wakefulness, longer latency to sleep after lights off, decreased frequency and length of sleep bouts at night, and are hyperaroused. Today, a sleep promoting neuropeptide has yet to be found in zebrafish.

What happens when a worm meets a fish?

The route to better understand the need of sleep is to find out its functions and regulations, which highlights the importance of uncovering essential genetic molecules involved in the process. Despite intensive research in sleep, little is known about the fundamental molecules and their specific roles in regulating sleep and arousal. Identification of novel genes has been challenging due to the complex genome and a long life cycle of mammalian models.

To understand the role of a component in an assemble, there is no better way than taking it out of the system and observing the consequence of missing the component. In the context of understanding gene function, geneticists generate mutants to dissect the gene functions. *C. elegans* has a short life cycle of about 3 days, and a well-annotated

genome, making it a good model for genetic analysis (WormBook, www.wormbook.org). One of the most attractive features of *C. elegans* is the ease of generating mutants (see Chapter 1). More importantly, the wealth of existing mutant collections provides an instant resource for identification of gene function. Together with the ease of amenability and genetic similarities between worm and vertebrate, *C. elegans* can serve as a quick screening tool for identifying new genes and unraveling their function in animal physiology, such as sleep behavior. Functions of these candidate genes can then be validated in a vertebrate system like zebrafish.

References

1. Chiu, C.N., and Prober, D.A. (2013). Regulation of zebrafish sleep and arousal states: current and prospective approaches. Frontiers in Neural Circuits *7:58.*
2. Cho, J.Y., and Sternberg P.W. (2014). Multilevel modulation of a sensory motor circuit during *C. elegans* sleep and arousal. Cell *156,* 249-260.
3. Gallagher, T., Kim, J., Oldenbroek, M., Kerr, R, and You, Y.J. (2013). ASI regulates satiety quiescence in *C. elegans*. J. Neurosci. *33(23),* 9716-9724.
4. 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(5442),* 1141-1146.
5. Pfaff, D., Ribeiro, A., Matthews, J., and Kow, L.-M. (2008). Concepts and mechanism of generalized central nervous system arousal. Ann. N.Y. Acad. Sci *1129*, 11-25.
6. Prober, D.A., Rihel, J., Onah, A.A., Sung, R.J., and Schier, A.F. (2006). Hypocretin/orexin overexpression induces an insomnia-like phenotype in zebrafish. J. Neurosci. *26(51),* 13400-13410.
7. Raizen, D.M., Zimmerman, J.E., Maycock, M.H., Ta, U.D., You, Y.J. Sundaram, M.V., and Pack, A.I. (2008). Lethargus is a *Caenorhabditis elegans* sleep-like state. Nature *451(7178),* 569-572.
8. Sehgal, A., and Mignot, E. (2011). Genetics of sleep and sleep disorders. Cell *146,* 194-207.
9. TALE nuclease architecture for efficient genome editing. Nat. Biotechnol. *29:* 143–148.
10. Van Buskirk, C., and Sternberg, P.W. (2007). Epidermal growth factor signaling induces behavioral quiescence in *Caenorhabditis elegans.* Nat. Neurosci. *10(10),* 1300-1307.
11. White, J.G., Southgate, E., Thomson, J.N., Brenner, F.R.S. (1996). The structure of the nervous system of the nematode *Caenorhabditis elegans.* Philos Trans R Soc. Lond. B Biol. Sci. *314,* 1-340.
12. Yokogawa, T., Marin, W., Faraco, J., Pezeron, G., Appelbaum, L., Zhang, J., Rosa, F., Mourrain, P., and Mignot, E. (2007). Charactrization of sleep in zebrafish and insomnia in hypocretin receptor mutants. PLoS Biol. *5(10),* e277.
13. You, Y.J., Kim, J., Raizen, D.M., and Avery L. (2008). Insulin, cGMP, and TGF-beta signals regulate food intake and quiescence in *C. elegans*: a model for satiety. Cell Metab. *7(3),* 249-257.
14. Zhdanova, I.V., Wang, S.Y., Leclair, O.U., and Danilova, N.P. (2001). Melantonin promotes sleep-like state in zebrafish. Brain Res. *903,* 263-268.
15. Zimmerman, J.E., Naidoo, N., Raizen D.M., and Pack, A.I. (2008). Conservation of sleep: Insights from non-mammalian model systems. Trends Neurosci. *31*, 371-376.

Chapter 5

FMRFamide neuropeptides promote sleep in *Caenorhabditis elegans* and *Danio rerio*

FMRFamide neuropeptides promote sleep in *Caenorhabditis elegans* and zebrafish (*Danio rerio*)

Elly Suk Chow[1,2], Daniel A. Lee[1], Erich M. Schwarz[3], David A. Prober[1], Paul W. Sternberg[1,2*]

1. Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, U.S.A.
2. Howard Hughes Medical Institute
3. Department of Molecular Biology and Genetics, Biotechnology 351, Cornell University, Ithaca, NY 14853, U.S.A.
* Corresponding author: pws@caltech.edu

SUMMARY

Sleep is an evolutionarily conserved physiological state whose underlying mechanism remains elusive. The nematode *Caenorhabditis elegans* exhibits a sleep-like state characterized by sensory depression and locomotor quiescence in response to intracellular signaling activation, satiety, and its developmental molting cycle. Temperature elevation can activate the epidermal growth factor receptor (EGFR) signaling in the single neuron ALA to induce sleep. However, the downstream molecular pathways are unknown. Using single-cell RNA-seq analysis of ALA, we show that some of the most highly expressed and enriched genes encode neuropeptides, primarily FMRFamide-like neuropeptides. These peptides are necessary for EGFR-induced sleep, and are sufficient to induce sleep in a manner that depends on G-protein coupled receptors NPR-7 and NPR-22. A vertebrate equivalent of these FMRFamide neuropeptides, zebrafish RFRP-1, is sufficient to induce sleep in *C. elegans* via NPR-7 and NPR-22. Furthermore, we show that overexpression of RFRP-1 in zebrafish also increases sleep. These observations

indicate that FMRFamide neuropeptides are novel, evolutionarily conserved regulators of sleep.

HIGHLIGHTS

- Single neuron RNA-seq identified sleep-inducing neuropeptides in *C. elegans*
- The ALA neuron releases FMRFamide neuropeptides to induce *C. elegans* sleep
- FMRFamide neuropeptides promote sleep in worm and zebrafish
- FMRFamide neuropeptides mediates EGFR-induced *C. elegans* sleep via GPCRs

INTRODUCTION

Sleep is a fundamental behavior required by most animals, and is universally characterized by a reversible state during which voluntary movements are inhibited, responsiveness to sensory stimulation is reduced as a result of increased arousal threshold, and sleep is driven by homeostasis and controlled by a circadian clock (Campbell and Tobler, 1984; Sehgal and Mignot, 2011). Sleep behavior, although ancient as reproduction and feeding, is poorly understood at both the functional and molecular levels. Uncovering the essential molecules regulating sleep/wakefulness will expand our knowledge regarding the need for sleep.

Recent research exploiting genetic information and the amenability of animal models had not only uncovered new molecular elements regulating sleep/wakefulness, but also revealed that sleep is genetically controlled and the underlying molecular mechanism is

evolutionarily conserved from worms to mammals (Zimmerman et al., 2008; Sehgal and Mignot, 2011). The roundworm *Caenorhabditis elegans,* along with fish and flies, exhibits a sleep-like state (Campbell and Tobler, 1984; Zimmerman et al., 2008; Sehgal and Mignot, 2011) and has emerged as newer models for sleep studies. Here, we use the relatively simple nervous system and easy accessibility of *C. elegans* and zebrafish (*Danio rerio)* as models to address the fundamental molecular underpinnings regulating sleep.

Sleep behavior in *C. elegans* was first found during lethargus, a developmentally programmed form of quiescence in which animals exhibit a prolonged period of voluntary movement inactivity such as locomotion and feeding quiescence, and increased arousal threshold to sensory and chemical stimulation (Raizen et a., 2008). Further genetic studies and circuit modulation analysis showed that worm sleep is not restricted to lethargus or solely tied to developmental stage (Figure 1A). Similar to their vertebrate cousins, intracellular signaling pathways play pivotal roles in worm sleep regulation. *C. elegans* sleep can be induced in larvae and young adults by anachronistic expression of the epidermal growth factor receptor (EGFR) signaling (Van Buskirk and Sternberg 2007) and by satiety through the TGFβ and cGMP pathways (You et al., 2008; Gallagher et al., 2013). In addition, EGFR-induced sleep and satiety quiescence share common neuronal modulation with lethargus (Cho and Sternberg, 2014). Observations from these distinct but related studies indicate that *C. elegans* is an excellent model for screening sleep-promoting molecules. Indeed, the sleep-promoting role of EGFR signaling is conserved in worms and flies (Van Buskirk and Sternberg 2007; Foltenyi et al, 2007),

and EGFR signaling has been shown to inhibit locomotor activity in mammals (Kramer et la., 2001); however, little is known about the downstream molecules of EGFR signaling that promote sleep.

The sole EGFR ortholog in *C. elegans*, LET-23, is expressed in the petidergic ALA neuron (White et al, 1986), which is a quiescence-promoting component in lethargus and is indispensable for mediating EGFR-induced sleep outside of lethargus (Van Buskirk and Sternberg 2007). The ALA neuron was implicated as a functional equivalent to the mammalian subparaventricular zone of the hypothalamus and the activation of EGFR signaling triggers ALA to secrete sleep-promoting neuropeptides (Van Buskirk and Sternberg 2007). Anachronistic activation of the EGFR signaling can be achieved by overexpressing the ligand, LIN-3C, under a heat-shock promoter or by temperature elevation itself (Hill et al., personal communication). Hill et al. recently demonstrated that heat shock activates EGFR signaling in the ALA neuron to induce a robust and reversible sleep behavior, likely through discharged neuropeptides, in young adult *C. elegans* within minutes of treatment. At present, only a few neuropeptides are known to express in ALA, and less is known about their role in promoting sleep.

Here we use RNA-sequencing to unravel neuropeptides that are highly transcribed in the ALA neuron. We report that FMRFamide neuropeptides are enriched in the ALA neuron and are necessary to mediate EGFR-induced sleep through two G-protein coupled receptors (GPCRs). In addition, these neuropeptides are sufficient to induce sleep without

a mature ALA neuron. Moreover, we present evidence that FMRFamide neuropeptides are evolutionarily conserved sleep-promoters in *C. elegans* and zebrafish.

RESULTS

FMRFamide neuropeptide-coding genes are highly transcribed in the ALA neuron

To test the hypothesis that ALA synthesizes neuropeptides to induce sleep, we characterized the transcriptome of ALA neurons from larvae two hours prior to the L4 molt (Figure S1), and detected the expression of 8,133 protein-coding genes (Table S1). Among the detected genes, 23 were FMRFamide-like neuropeptide-coding (FLP) genes, of which 6 FLP neuropeptide-coding genes were transcribed at least 17-fold more in ALA neurons than in whole larvae (Figure 1B; Table S2). The most ALA-enriched *flp* genes are *flp-24, flp-13,* and *flp-7* (of which only *flp-7* was previously known to express in ALA; Kim and Li, 2004). We then verified expression of *flp-13* and *flp-24* in ALA with GFP reporter constructs (Figure 1C). Previous analysis showed that each of these genes encodes a prepropeptide containing one or more mature neuropeptides (www.uniprot.org, Figure S2).

ALA-synthesized FMRFamide neuropeptides mediate EGFR-induced sleep

To examine if these ALA-enriched neuropeptides are required for EGFR-induced sleep, we tested the response of their null mutants to activated LET-23/EGFR signaling. The transcription factor CEH-14 is required for the expression of *let-23* and downstream signaling components in the ALA neuron (Van Buskirk and Sternberg 2010). *ceh-14* null mutants are resistant to EGFR-induced sleep, presumably due to ALA's failure to

activate downstream signaling components such as neuropeptides (Van Buskirk and Sternberg, 2007). Using a feeding quiescence assay modified from published sleep assays (Figure 2A, Van Buskirk and Sternberg, 2007; Raizen et al., 2008; Hill et al., personal communication), we found that the null mutants of all three FLP peptide-encoding genes (Figure S3A) were resistant to sleep by varying degrees (Figure 2B): *flp-13* mutants exhibited the highest resistance, followed by *flp-24* and *flp-7* mutants. By contrast, we did not observe resistance in mutants that lacked other ALA-enriched neuropeptides, such as *flp-10* and *flp-19* (Figure S3C and Table S3), suggesting that ALA induces sleep through the combined effects of at least three FMRFamide neuropeptides.

ALA-enriched FMRFamide neuropeptides promote sleep

To test if ALA-synthesized neuropeptides are sufficient to induce sleep, we conditionally expressed the top three highly expressed *flp* genes under a heat-shock inducible promoter and assessed sleep behavior after the stress-induced EGF effect has subsided, using three criteria: locomotor quiescence, feeding quiescence, and decreased sensory arousal (Figure 2A). Wild type young adult animals are active and rarely exhibit locomotor or feeding quiescence (Raizen et al., 2008). In response to elevated temperature, wild type animals exhibited transient sleep behavior and eventually resumed to the normal active state after one hour of recovery at 20°C (Hill et al, personal communication). By contrast, overexpression of *flp-13* (hs:FLP-13) and *flp-24* (hs:FLP-24) induced sleep in typically active young adults after the response of temperature elevation per se wore off (Figure 3). FLP-13-overexpressing animals had prolonged periods of inactivity (Figure 3A)

interrupted by brief bouts of movement in which they jerked backward (Movie S1), while FLP-24-overexpressing animals rarely moved at all (Movie S2). Feeding was reduced or completely abolished, as indicated by the pharyngeal pumping rate (Figures 3A and 3B). In contrast to wild-type animals (Troemel et al., 1995), sleeping animals had delayed response to the sensory stimulant. We found that hs:FLP-13 or hs:FLP-24 animals exhibited delayed withdrawal behavior when presented with 1-octanol (Figure 3C)., a behavior similar to sleeping animals (Raizen et al., 2008; Cho and Sternberg, 2014). In agreement with the mild resistance to EGFR-induced sleep in *flp-7* mutants (Figure 3C), we observed a mild increase in response latency in FLP-7-overexpressing (hs:FLP-7) animals (Figure 3A). In all cases, we observed fully recovered active animals three hours post heat shock (data not shown), indicating reversibility of FMRFamide neuropeptide-induced sleep. Together our data suggest that ALA-synthesized FMRFamide neuropeptides are sufficient to promote sleep in active *C. elegans*.

FLP-13 and FLP-24 neuropeptides are the functional outputs of EGFR signaling in ALA

To further characterize if these neuropeptides function downstream of EGFR signaling in the ALA neuron, we overexpressed FLP-13 and FLP-24 in *ceh-14* null mutants (Figure 4) that lack EGFR signaling due to the loss of LET-23/EGFR (Van Buskirk and Sternberg, 2010). As a control of EGFR signaling activation, we overexpressed LIN-3C in wild type and *ceh-14* null mutants (Figure 4A). We observed nearly complete resistance to sleep induction in *ceh-14* mutants when overexpressing LIN-3C alone. However, overexpressing FLP-13 or FLP-24 together with LIN-3C blocked the non-sleep behavior in *ceh-14* mutant (Figure 4A). Our findings suggest that FLP-13 or FLP-24 can induce

sleep without requiring the presence of CEH-14 or LET-23/EGFR in the ALA. Consistent with this hypothesis, we found that the transcriptomes of *flp-13* and *flp-24* are dramatically reduced in the ALA of *ceh-14* null mutant (Chow et al., unpublished data), suggesting transcriptional dependency on *ceh-14* by these ALA-synthesized neuropeptides (Figure 4B), and that these neuropeptides are the functional outputs of CEH-14 regulation and EGFR signaling activation in promoting sleep.

Neuropeptide receptors, NPR-7 and NPR-22, mediate FMRFamide neuropeptides-induced sleep

Neuropeptides exert their effects predominantly through G-protein-coupled receptors (GPCRs, McVeigh et al., 2006). The neuropeptide products of multiple FLP genes can activate the GPCR NPR-22 *in vitro* (Mertens et al., 2006), including ALA-synthesized *flp* genes (*flp-7*, *flp-9*, *flp-13*, and *flp-22*; Tables S2 and S4), suggesting that NPR-22 might be a receptor for ALA neuropeptides. We therefore investigated the responses of animals containing null mutations in *npr-22* (Figure S3B) and paralogous receptors (Hinuma et al., 2000) to EGFR-induced sleep. Mutants defective in *npr-3, npr-7,* or *npr-22* exhibited increased activity in sleep-induced animals (Figure 5A), and all peptides tested required NPR-7 and NPR-22 to varying extents (Figure 5B). Loss of NPR-7 restored feeding and locomotion completely for hs:FLP-13 and partially for hs:FLP-24. Conversely, loss of NPR-22 restored feeding completely for hs:FLP-24 and partially for hs:FLP-13, and restored locomotion partially for both peptides. Moreover, loss of NPR-7 or NPR-22 reduced the effects of hs:FLP-13 or hs:FLP-24 on arousal (Figure 5C). We therefore infer that the GPCRs NPR-7 and NPR-22 are likely to induce sleep by acting as receptors for

the ALA-synthesized neuropeptides FLP-13 and FLP-24, with different sensitivities to the two neuropeptides, thereby eliciting multiple behavioral outputs.

Sleep-promoting effect of FMRFamide neuropeptide/receptor system is evolutionarily conserved

To identify the vertebrate equivalent of NPR-7 and NPR-22, we compared their sequences with vertebrate GPCRs and found them to have high similarity with human NPFFR1 and zebrafish NPFFR1-1 and NPFFR1-3 (Table S5), the cognate receptors of vertebrate FMRFamide neuropeptides (Liu et al., 2001). We then aligned FLP-13 and FLP-24 peptides with the bioactive peptide sequence of NPFFR1 ligands, the RFRP neuropeptides (RFRP-1, RFRP-2, RFRP-3) generated by the zebrafish prepropeptide gene *npvf* (Hinuma et al., 2000; Liu et al., 2001). We observed similar carboxy-termini and hydrophobic backbones in both FLP and RFRP mature peptides (Figures S4 and S5). However, the sequences are short enough to be the product of either divergent or convergent evolution. Therefore, to test the functional relevance of these sequence similarities, we replaced the bioactive peptide of FLP-24 with the zebrafish RFRP bioactive peptides in the FLP-24 backbone (Figure 6A), and overexpressed these chimeric peptides in *C. elegans*. We observed that overexpression of zebrafish RFRP-1 peptide induced sleep in *C. elegans* and that this phenotype was completely blocked in *npr-7* mutants, whereas mutation of *npr-22* only partially blocked the locomotion phenotype (Figure 6B; Figure S6A). By contrast, overexpression of RFRP-3 peptide caused a mild, statistically insignificant effect on locomotion (Figure S6B). We conclude

that vertebrate NPFFR1 is functionally equivalent to nematode NPR-7, and that vertebrate RFRP-1 peptide is a functional equivalent of *C. elegans* FLP-13 and FLP-24.

RFRP neuropeptides promote sleep in zebrafish

To determine whether RFRP neuropeptides can promote sleep in a diurnal vertebrate, we generated transgenic zebrafish in which the RFRP neuropeptide precursor gene expression was regulated by a heat-shock-inducible promoter (hs:RFRP; Figure S7). We continuously monitored the locomotor activity of hs:RFRP larvae and their wild-type siblings over several days using an automated video-tracking system (Prober et al, 2006; Figure 7A; Figure S7). Prior to heat-shock, there was no statistically significant difference in locomotor activity or sleep (Figures 7A and 7B). Following heat-shock, however, we observed a dramatic reduction of locomotor activity (Figure 7A) and a commensurate increase in sleep that was restricted to the day period (Figures 7A and 7B) in hs:RFRP larvae. The increase in sleep resulted from an increase in both the number and length of sleep bouts (Figures 7C and 7D) and a decrease in the length of wake bouts (Figure 7E). We also detected decreased sleep latency, or time to first sleep bout, in the day after heat-shock (Figure 7F). We conclude that FMRFamide neuropeptides are evolutionarily conserved sleep-promoting molecules in vertebrate and nematodes.

DISCUSSION

We have shown that multiple FMRFamide neuropeptides are highly transcribed in the ALA neuron prior to developmentally programmed sleep during lethargus. We also showed that ectopic expression of these neuropeptides could elicit prolonged sleep in

typically active young adults. Moreover, we found that the presence of these neuropeptides in the ALA is necessary to mediate EGFR-induced sleep, and that when overexpressed this activity is independent of a fully developed ALA neuron or the presence of EGFR due to the lack of *ceh-14*. Genetic analysis indicates that *ceh-14*, together with its downstream-regulated transcription factor *ceh-17,* is essential for proper ALA development (Van Buskirk and Sternberg, 2010). Indeed, sleep induction is unperturbed in *ceh-17* mutants that exhibit truncated ALA axons (Van Buskirk and Sternberg, 2010). By contrast, EGFR-induced sleep is dependent on neuropeptides and neurotransmitters release. Mutations of a synaptic vesicle release regulator, *unc-13* and a vesicle docking protein, *unc-18,* severely suppress sleep induction (Van Buskirk and Sternberg, 2007). Consistent with the view, we detected expression of *unc-13* and *unc-18* in the ALA neuron (Table S1), indicating their requirement in secreting neuropeptides from ALA. Our results show that FMRFamide neuropeptides are transcriptionally regulated by CEH-14 in ALA and are likely discharged upon activation of EGFR signaling.

Neuropeptide ligand/receptor systems provide ubiquitous neuromodulation to transmit information flow through their long-range effects on neuronal excitability and presynaptic efficacy at a given time and circumstance (Bargmann, 2012). Here we demonstrate that the ALA-synthesized FMRFamide neuropeptides promote sleep through GPCRs, NPR-7, and NPR-22. In addition, we systematically dissected specific roles of particular ligand/receptor combination. We found that although multiple receptors are involved in mediating FMRFamide neuropeptides to induce sleep, there are varying

degrees of flexibility and specificity (Figure 5). We show that NPR-7 is absolutely required to mediate suppression of all activities elicited by both FLP-13 and FLP-24 neuropeptide overexpression, and lack of NPR-7 dramatically decreased EGFR-induced sleep (Figure 5A). We conclude that NPR-7 is a sleep-promoting modulator for the ALA-discharged neuropeptides. On the contrary, the role of NPR-22 in sleep promoting is less critical. We postulate that NPR-22 may mediate both sleep and wakefulness elicited by neuropeptides, and such behavior is likely subject to a given environmental condition. This observation of FLP/NPR-22 combination in modulating sleep/wakefulness mimics that of the vertebrate NPY neuropeptide/receptor system. In fact, NPR-22 has high sequence homology with NPFFR receptors (Table S5) and tachykinin-like receptor 1 and neuropeptide Y receptor 2 (Mertens et al., 2006). NPY in regulating sleep/wakefulness is thought to be site-dependent and species varied. The short neuropeptide sNPF, a NPY-like peptide in *Drosophila*, potentially promotes sleep (Shang et al., 2013). Depending on dosage and the site of injection into the rat brain, NPY can either promote sleep or wakefulness (Dyzma et la, 2010; Toth et al, 2007). In humans, introduction of NPY through repetitive intravenous injection promotes sleep in young men (Antonijevic et al., 2000).

Current understanding of neuropeptide/ligand receptor systems in *C. elegans* and their vertebrate orthologs is still nascent. *C. elegans* has a compact genome and often has a gene of a family encoded by three or more orthologous genes in vertebrates as a result of genome duplication in chordate evolution (Sidow, 1996), adding complexity to the identification of nematode and vertebrate orthologous neuropeptide/ligand receptor

systems. Given that neuropeptides and receptors often co-evolve through evolution (Bargmann, 1998), we infer that neuropeptides in one organism may act through an equivalent receptor in another organism to elicit similar behavioral outputs. Here, we provide evidence that FMRFamide neuropeptide from a vertebrate model, zebrafish, can promote sleep in *C. elegans,* and that this activity is mediated by the same GPCRs that are required by the sleep-promoting *C. elegans* FMRFamide neuropeptides. Based on the absolute necessity of NPR-7 to mediate sleep induction by zebrafish RFRP-1 in *C. elegans,* we infer that NPR-7 is the equivalent of vertebrate NPFFR1. Furthermore, we show that RFRP neuropeptides, a member of the vertebrate FMRFamide neuropeptide family, can promote sleep in zebrafish in the daytime when the animals are mostly active. To our knowledge, this work is the first report of vertebrate FMRFamide in promoting sleep.

Other neuropeptides impinge on sleep and related behaviors. The neuropeptide-like-protein NLP-22 in *C. elegans* was previously shown to regulate developmentally programmed sleep-like behavior (Nelson et al., 2013). Some vertebrate FMRFamide peptides have been demonstrated to modulate feeding and reproduction (Dockray, 2004); the *C. elegans* FMRFamide peptides, FLP-18 and FLP-21, in combination with GPCRs NPR-5 and NPR-4, and with NPR-1, modulate foraging and locomotion quiescence during lethargus, respectively (Cohen et al., 2009; Choi et al., 2013). Here we demonstrate that the ALA-synthesized FMRFamide peptides can induce sleep at all stages via two additional GPCRs, demonstrating the versatility of FMRFamide peptides in modulating multiple behaviors. We found that these neuropeptides and their receptors

are downstream mediators of EGFR-induced sleep in *C. elegans*, and are functional equivalents of the vertebrate RFRP/NPFFR1 system. Given that EGFR signaling and FMRFamide neuropeptides have similar roles in promoting sleep, it is possible that EGFR signaling will act via neuropeptides in other organisms as well as *C. elegans*. The combination of single-cell profiling and functional analysis in *C. elegans* allowed us to identify candidate vertebrate sleep regulators; validation of RFRP-1 in zebrafish as a sleep-promoting FMRFamide neuropeptide identifies a novel sleep regulator likely conserved among animals.

EXPERIMENTAL PROCEDURESS

Strains

Strains were grown, unless indicated otherwise, on nematode growth medium (NGM) 2% agar Petri plate seeded with *E. coli* strain OP50 and maintained at 20°C under standard conditions (Brenner, 1974).

Wild-type worm strain was N2 (Bristol). Information about mutant strains and transgenic strains in this study is available in the Supplemental Experimental Procedures.

Single ALA neuron dissection and transcriptome profiling

Individual wild-type larvae of strain TB513 at the mid-L4 larval stage were hand picked and glued on an agar pad for microdissection as previously described (Schwarz et al., 2012). GFP-tagged ALA neurons were individually collected with an unpolished patch-clamp tube that served as a pipette, transferred to a prelubricated microcentrifuge tube

(Figure S1), and snap-frozen with liquid nitrogen. Frozen tubes containing individual ALA neurons were kept at -70°C until their RNA was amplified.

RT-PCR, RNA-seq, and computational analysis of individual neurons were done essentially as in Schwarz et al., 2012. To obtain RNA-seq data by Illumina sequencing, aliquots of RT-PCR from individual cell were collected into two pools (4 cells and 5 cells). All RNA-seq reads were single-end, and originally 50 nt in length. Raw reads were quality-filtered as in Schwarz et al., 2012. They were then truncated *in silico* from 50 nt to 38 nt, the read length for previously published control data from mixed whole larvae Schwarz et al., 2012. This truncation allowed the ALA reads to be mapped and quantitated using exactly the same pipeline that had been used for larval data, and thus allowed more exact comparisons between ALA and larvae. After quality filtering and truncation but before mapping, RNA-seq data from the two pools of wild-type ALA comprised 1,164,892,280 nt in 30,655,060 reads and 1,520,526,262 nt in 40,013,849 reads. Of these, 25.2% could be mapped to WS190 protein-coding gene models (i.e., 17,798,207 out of 70,668,909 reads). We used existing whole wild-type larval RNA-seq data (Schwarz et al., 2012) as controls for housekeeping versus ALA-enriched genes. Expression values for genes were computed as in Schwarz et al., 2012. They were defined by pooling reads from both wild-type mid-L4 ALA neuron sets into a single set of expression values, doing likewise for both wild-type larval RNA-seq sets from Schwarz et al., 2012, and computing ALA/larval ratios of gene activity. We detected expression of 7,698 and 4,068 genes in the two ALA pools separately, and 8,133 genes collectively.

Behavior assays

Heat-shock assays

EGFR signaling-induced feeding suppression assays:

Ten young adult hermaphrodites were hand-selected and transferred using a platinum pick to a Petri plate containing NGM coated with a thin lawn of OP50 bacteria. The plates were sealed with Parafilm and placed in a 35°C water bath for 30 min. Heat-shocked animals were permitted to recover at 20°C for 10 min before being examined for feeding behavior, as described below; this was done at 10 min intervals for 60 min.

Heat-shock neuropeptide gain-of-function quiescence assays:

Young adult transgenic or control hermaphrodites were hand-selected and transferred using a platinum pick to Petri plates containing NGM coated with a thin lawn of OP50 bacteria. Approximately 20 animals were transferred to each plate. Heat-shock treatment was performed as described in Van Buskirk and Sternberg, 2007. Briefly, the plates were sealed with Parafilm and placed in a 33°C water bath for 30 min, allowed to recover at 20°C for 2 hours, and then scored for suppression of locomotion and feeding, and for response latency to sensory stimulation. These assays were performed as described below.

Feeding behavior

Each animal was examined for feeding, indicated by the movement of the grinder in the posterior pharyngeal bulb (Raizen et al., 1995) under high magnification on a stereomicroscope for 5 sec. Animals with no pharyngeal movement observed were scored

as quiescent for feeding behavior. For animals with pumping, activity was recorded for 1 min with a Unibrain camera using Unibrain software, and quantified by number of pharyngeal contractions per min.

Locomotion

For locomotion analysis, animals 2 hrs post heat-shock were placed on a Leica steromicroscope base with a Unibrain camera using Unibrain software, illuminated with continuous white light and imaged for 4 min to track forward and backward movements. The centroid velocity plot was calculated over 2-sec intervals as described (Van Buskirk and Sternberg, 2007).

Response latency

For chemical response assay, animals that recovered from heat shock were presented with 30% 1-octanol, as described previously (Raizen et al., 2008; Troemel et al., 1995). 1-octanol (Sigma-Aldrich) was diluted to a final concentration of 30% with Ethanol (volume : volume). Briefly, an eyebrow hair dipped in 30% 1-octanol was presented within the length of a pharynx of the nose of the worm (d1, Figure 1b). The time required for the worm to move backward for the length of a pharynx (d2, Figure 1b) was recorded. The time duration between the time when 1-octanol was presented to the animal and the time when the animal moved backward for the length of a pharynx was documented as response latency. All comparisons between treatments and genotypes were made on the same day.

Statistics for *C. elegans* experiments

Two-tail *p*-values were calculated using InStat software (GraphPad). Means were compared using an unpaired *t*-test with Welch's correction in the case of unequal variances.

Generation of transgenic zebrafish

Full-length zebrafish RFRP precursor gene *npvf* (Ensembl ID ENSDARG00000036227) was amplified from a larval zebrafish cDNA library by PCR (forward primer 5'-ATGTCCTACTTCGCTCTTCTTTCT-3', reverse primer 5'-AGCGTCTAGACGTTTTTGGGGCTTTGTTAG-3'; PFU Ultra II Fusion HS DNA Polymerase, Agilent Technologies) and subcloned downstream of the zebrafish *hsp70c* promoter (Halloran et al., 2000) in a vector containing flanking ISceI meganuclease sites. Transgenic zebrafish were generated by co-injecting plasmids with ISceI enzyme (R0694, New England Biolabs Inc.) into the cytoplasm of embryos of the TLAB wild-type strain at the one-cell stage. All experiments were performed using standard protocols (Westerfield, 1994) in accordance with the California Institute of Technology Institutional Animal Care and Use Committee guidelines.

Zebrafish behavioral analysis

Larval zebrafish were raised on a 14 hour:10 hour light:dark cycle at 28.5°C with lights on at 9 am and off at 11 pm. Individual larvae were placed into each well of a 96-well plate (7701-1651; Whatman) containing 650 μL of E3 embryo medium (5 mM NaCl,

0.17 mM KCl, 0.33 mM $CaCl_2$, 0.33 mM $MgSO_4$, pH 7.4; 7). Plates were sealed with an optical adhesive film (4311971, Applied Biosystems) to prevent evaporation. The sealing process introduced air bubbles into some wells, which were discarded from analysis. Locomotor activity was monitored using an automated videotracking system (Viewpoint Life Sciences, Montreal, Quebec, Canada) with a Dinion one-third inch Monochrome camera (Dragonfly 2, Point Grey) fitted with a variable-focus megapixel lens (M5018-MP; Computar) and infrared filter. The movement of each larva was recorded using the quantization mode, with data from two cameras collected simultaneously by one computer. The 96-well plate and camera were housed inside a custom-modified Zebrabox (Viewpoint Life Sciences, Montreal, Quebec, Canada) that was continuously illuminated with infrared lights and illuminated with white lights from 9 am to 11 pm. The 96-well plate was housed in a chamber filled with recirculating water to maintain a constant temperature of 28.5°C. Larvae were heat shocked by transferring the 96-well plate to a 37°C water bath for 1 hour. The parameters used for detection were: detection threshold: 15; burst: 25; freeze: 3; bin size: 60 seconds. Videotracker data was processed using custom PERL and Matlab (version R2013a; The Mathworks, Inc) scripts. Any 1 minute period with less than 0.1 second of movement was defined as 1 minute of sleep[16]. A sleep bout was defined as a continuous string of sleep minutes. Sleep latency was defined as the length of time from a change in lighting conditions to the start of the first sleep bout. Average activity was defined as the average amount of detected activity in seconds/minute, including all rest bouts. Wells were excluded from behavioral analysis if zero or more than one larva was observed in the well (setup error), large bubbles were observed in the sealed well, or if a larva did not have an inflated swim bladder. Data and

figures are described as mean ± standard error of the mean (SEM). Two-tailed Student's t-test was used to assess the statistical significance where * = $p<0.05$, ** = $p<0.01$, ***=$p<0.001$, unless otherwise stated.

Data Availability

RNA-seq reads for the two pools of wild-type mid-L4 ALA neurons are available in the NCBI Sequence Read Archive (SRA), under accession number SRP038903.

AUTHOR CONTRIBUTIONS

E.S.C. and P.W.S. conceived the project and E.S.C. performed *C. elegans* research. D.A.L. and D.A.P. designed and performed *D. rerio* research, and E.M.S. performed the ALA transcriptome analysis.

ACKNOWLEDGMENTS

We thank the *Caenorhabditis* Genetics Center and S. Mitani for strains, WormBase, Gladys Medina for technical support, Millard and Muriel Jacobs Genetics and Genomics Laboratory for sequencing, and members of our laboratories for discussions. We thank Cheryl Van Buskirk for sharing unpublished observations, and Hillel Schwartz and Mihoko Kato for comments and discussions. This work was supported by the Howard Hughes Medical Institute, with which P.W.S. is an investigator, the Rita Allen Foundation to D.A.P., and the National Institutes of Health (GM084389 to P.W.S.; NS070911 and DA031367 to D.A.P.; 1F32NS084769 (NINDS NRSA) to D.A.L.).

Strains carrying the *tm2427* and *tm1583* alleles are available through the National BioResource Project (NBRP), subject to a materials transfer agreement.

REFERENCES

1. Antonijevic, I.A., Murck, H., Bohlhalter, S., Frieboes, R.M., Holsboer, F., and Steiger, A. (2000). Neuropeptide Y promotes sleep and inhibits ACTH and cortisol release in young men. Neuropharmacology *39(8),* 1474-1481.

2. Bargmann, C. I. (2012). Beyond the connectome: how neuromodulators shape neural circuits. Bioessays *34(6),* 458-465.

3. Bargmann, C.I.. (1998). Neurobiology of the *Caenorhabditis elegans* genome. Science *282(5396),* 2028-2033.

4. Campbell, S.S., and Tobler, I. (1984). Animal sleep: a review of sleep duration across phyogeny. Neurosci. Biobehav. Rev. *8,* 269-300.

5. Cho, J.Y., and Sternberg P.W. (2014). Multilevel modulation of a sensory motor circuit during *C. elegans* sleep and arousal. Cell *156,* 249-260.

6. Choi, S., Chatzigeorgiou, M., Taylor, K.P., Schafer, W.R., and Kaplan, J.M. (2013). Analysis of NPR-1 reveals a circuit mechanism for behavioral quiescence in *C. elegans*. Neuron *78(5),* 869-880.

7. Cohen, M., Reale, V., Olofsson, B., Knights, A., Evans, P., and de Bono, M. (2009). Coordinated regulation of foraging and metabolism in *C. elegans* by RFamide neuropeptide signaling. Cell Metab. *9(4),* 375-385.

8. Dockray, G. J. (2004). The expanding family of –RFamide peptides and their effects on feeding behavior. Exp. Physiol. *89(3),* 229-235.

9. Dyzma, M., Boudjeltia, K.Z., Farut, B., and Kerkhofs M. (2010). Neuropeptide Y and sleep. Sleep Med. Rev. *14(3),* 161-165.

10. Foltenyi, K., Greenspan, R.J., and Newport, J.W. (2007). Activation of EGFR and ERK by rhomboid signaling regulates the consolidation and maintenance of sleep in *Drosophila.* Nat. Neurosci. *10(9),* 1160-1167.

11. Gallagher, T., Kim, J., Oldenbroek, M., Kerr, R, and You, Y.J. (2013). ASI regulates satiety quiescence in *C. elegans*. J. Neurosci. *33(23),* 9716-9724.

12. Hinuma, S., Shintani, Y., Fukusumi, S., Iijima, N., Matsumoto, Y., Hosoya, M., Fujii, R, Watanabe, T., Kikuchi, K, Terao, Y., Yano, T., Yamamoto, T.,

Kawamata, Y., Habata, Y., Asada, M., Kitada, C., Kurokawa, T., Onda, H., Nishimura, O., Tanaka, M., Ibata, Y., and Fujino, M. (2000). New neuropeptides containing carboxy-terminal RFamide and their receptor in mammals. Nat. Cell Biol. *2(10),* 703-708.

13. Kim K., and Li, C. (2004). Expression and regulation of an FMRFamide-related neuropeptide gene family in *Caenorhabditis elegans*. J Comp Neurol. *475(4),* 540-550.

14. Kramer, A., Yang, F.C., Snodgrass, O., Li, X., Scammell, T.E., Davis, F.C., Weitz, C.J. (2001). Regulation of daily locomotor activity and sleep by hypothalamic EGF receptor signaling. Science *294(5551),* 2511-2515.

15. Liu, Q., Guan, X.M., Martin, W.J., McDonald, T.P., Clements, M.K., Jiang, Q., Zeng, Z., Jacobson, M., Williams, D.L.Jr., Yu, H., Bomford, D., Figueroa, D., Mallee, K., Wang, R., Evans, J., Gould, R., and Austin, C.P. (2001). Identification and characterization of novel mammalian neuropeptide FF-like peptides that attenuate morphine-induced antinociception. J Biol Chem. *276(40),* 36961-36969.

16. McVeigh, P., Geary, T.G., Marks, N.J., and Maule, A.G. (2006). The FLP-side of nematodes. Trends Parasitol. *22(8),* 385-396.

17. Mertens, I., Clinckspoor, I., Janssen, T., Nachman R., and Schoofs, L. (2006). FMRFamide related peptide ligands activate the *Caenorhabditis elegans* orphan GPCR Y59H11AL.1. Peptides *27(6),* 1291-1296.

18. Nelson, M.D., Trojanowski, N.F., George-Raizen, J.B., Smith, C.J., Yu, C.C., Fang-Yen, C., and Raizen, D.M. (2013). The neuropeptide NLP-22 regulates a sleep-like state in *Caenorhabditis elegans*. Nat Commun *4,* 2846.

19. Prober, D.A., Rihel, J., Onah, A.A., Sung, R.J., and Schier, A.F. (2006). Hypocretin/orexin overexpression induces an insomnia-like phenotype in zebrafish. J. Neurosci. *26(51),* 13400-13410.

20. Raizen, D.M., Zimmerman, J.E., Maycock, M.H., Ta, U.D., You, Y.J. Sundaram, M.V., and Pack, A.I. (2008). Lethargus is a *Caenorhabditis elegans* sleep-like state. Nature *451(7178),* 569-572.

21. Sehgal, A., and Mignot, E. (2011). Genetics of sleep and sleep disorders. Cell *146,* 194-207.

22. Sidow, A. (1996). Gen(om)e duplications in the evolution of early vertebrates. Curr. Opin. Genet. Dev. *6(6),* 715-722.

23. Shang, Y., Donelson, N.C., Vecsey, C.G, Guo, F., Rosbash, M., and Griffith, L.C. (2013). Short neuropeptide F is a sleep-promoting inhibitory modulator. Neuron *80(1),* 171-183.

24. Toth, A., Hajnik, T., Zaborszky, L., and Detari L. (2007). Effect of basal forebrain neuropeptide Y administration on sleep and spontaneous behavior in freely moving rats. Brain Res. Bull *72(4-6),* 293-301.

25. Troemel, E.R., Chou, J.H., Dwyer, N.D., Colbert, H.A., and Bargmann, C.I. (1995). Divergent seven transmembrane receptors are candidate chemosensory receptors in *C. elegans.* Cell *83(2),* 207-218.

26. Van Buskirk, C., and Sternberg P.W. (2010). Paired and LIM class homeodomain proteins coordinate differentiation of the *C. elegans* ALA neuron. Development *137(12),* 2056-2074.

27. Van Buskirk, C., and Sternberg, P.W. (2007). Epidermal growth factor signaling induces behavioral quiescence in *Caenorhabditis elegans.* Nat. Neurosci. *10(10),* 1300-1307.

28. White, J.G., Southgate, E., Thomson, J.N., Brenner, F.R.S. (1996). The structure of the nervous system of the nematode *Caenorhabditis elegans.* Philos Trans R Soc. Lond. B Biol. Sci. *314,* 1-340.

29. You, Y.J., Kim, J., Raizen, D.M., and Avery L. (2008). Insulin, cGMP, and TGF-beta signals regulate food intake and quiescence in *C. elegans*: a model for satiety. Cell Metab. *7(3),* 249-257.

30. Zimmerman, J.E., Naidoo, N., Raizen D.M., and Pack, A.I. (2008). Conservation of sleep: Insights from non-mammalian model systems. Trends Neurosci. *31*, 371-376.

AUTHOR INFORMATION

ALA neuron transcriptome data are available from NCBI Sequence Read Archive (SRA), under accession number SRP038903. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests.

Correspondence should be addressed to P.W.S. (pws@caltech.edu). Requests for

materials should be addressed to P.W.S. (pws@caltech.edu) for *C. elegans*, and to D.A.P. (dprober@caltech.edu) for *D. rerio*.

Figure 1. ALA expresses multiple FMRFamide neuropeptide-coding genes.

(A) Model of sleep regulation in *C. elegans*. EGFR signaling promotes sleep through the ALA neuron, which is also a locomotor quiescence component in lethargus. Nutrition state induces satiety. All three pathways dampen sensory and motor circuit in a sleep-like state.

(B) Expression data of 8,133 protein-coding genes (grey) collected from two pools of microdissected ALA neurons (4 cells and 5 cells; see also Figure S1) compared with mixed-stage whole larvae. ALA neuron versus whole larval expression ratio showed that eight neuropeptide-coding genes have >17x higher expression in ALA than in whole larvae.

(C) Architecture of GFP expression constructs. GFP is driven by a conserved *cis*-regulatory element (CR; orange blocks) in the vicinity of neuropeptide coding sequence (grey blocks), either intergenic or intronic, via a basal promoter (*pes-10*).

(D-G) Detection of *flp-13* or *flp-24* in ALA (white arrow) was indicated by GFP in an L1 larva. The ALA neuron is located dorsal to the pharynx between the anterior and posterior pharyngeal bulbs. The pharynx (p) was illuminated by a pharynx specific expression marker, *myo-2*::dsRed. Anterior is left. Dorsal is up. Scale bar represents 20 μm.

Figure 2. ALA-synthesized FMRFamide neuropeptides are necessary for EGFR-induced sleep.

(A) Schematic diagram of sleep assay using locomotion, feeding and sensory arousal behavioral assays.

(B) FMRFamide neuropeptide null mutants (see also Figures S2 and S3) had significantly lower fraction quiescent (i.e. percentage of animals not feeding) than wild-type animals in response to EGFR-induced sleep. Feeding quiescence peaked at 20 minutes post heat shock in wild type animals while mutants of the ALA transcriptome regulator, *ceh-14*, and ALA-synthesized *flp-7, flp-13,* and *flp-24* were significantly more active. Error bars, mean s.e.m. ***$p<0.0001$, **$p<0.001$, *$p<0.05$, using two-tailed Student's *t*-test with unequal variance. Unless otherwise stated, n=100 for each group.

Figure 3. FMRFamide neuropeptides promote sleep in active young adult worms.

(A) Heat shock-induced overexpression of ALA-enriched neuropeptides increased feeding and locomotion quiescence compared to heat-shocked controls (no hs:peptide). Overexpression of a non-sleep inducing transcription factor (hs:HLH-13) and a non-functional peptide (hs:FLP-24PS) were used as controls for the heat shock induction assay. Animals overexpressing the EGFR ligand LIN-3C were used as a control for quiescence behaviors.

(B) Animals with overexpressed neuropeptide have less feeding behavior, quantified as pharyngeal pumping per min (ppm).

(C) Response latency to 30% 1-octanol is increased in animals overexpressing LIN-3C/EGF and neuropeptides (n=10). Error bars, mean s.e.m. ***$p<0.0001$, **$p<0.001$, *$p<0.05$, using two-tailed Student's *t*-test with unequal variance. Unless otherwise stated, n=100 for each group.

Figure 4. ALA-synthesized FMRFamide neuropeptides induce sleep without requiring EGFR signaling or CEH-14.

(A) All strains have hs:LIN-3C and thus were assayed after the response to heat shock per se wore off. Wild-type (WT) animals were completely inactive, while *ceh-14* null mutants were resistant to hs:LIN-3C/EGF-induced sleep and thus were active. Overexpression of FLP-13 and FLP-24 restored feeding and locomotion quiescence in *ceh-14* null mutants in response to heat shock overexpression of LIN-3C/EGF-induced sleep. Error bars, mean s.e.m. ***$p<0.0001$, using two-tailed Student's *t*-test with unequal variance. n=100.

(B) Model of FMRFamide neuropeptides functioning downstream of EGFR and CEH-14 in ALA to promote sleep. ALA: oval; Target cells: rectangle; G-protein coupled receptor: orange; FLP-13 peptide: red circle; FLP-24 peptide: blue circle.

Figure 5. GPCRs mediate FMRFamide neuropeptide-induced sleep.

(A) Null mutants of *npr-3, npr-7* and *npr-22* exhibited resistance to sleep induction, demonstrated by reduced feeding quiescence after heat shock. Mutants of *npr-32* exhibited similar response to sleep-induction as wild-type (WT) animals.

(B) Feeding and locomotion quiescence induced by FLP-13 overexpression in wild-type animals were completely blocked in *npr-7* mutants while the effects of FLP-24 were partially blocked. Mutation of *npr-22* completely blocked feeding quiescence and partially blocked locomotion quiescence. Behavioral quiescence is compared to the effect in wild-type animals with corresponding peptide or no hs:peptide (first sample). See also Figure S4.

(C) Mutation of *npr-7* restored sensory arousal in animals with heat shock overexpression of FLP-13 to wild-type levels as shown in animals with no hs:peptide (p = 0.26) and partially restored it in hs:FLP-24 animals (n=10). Mutation of *npr-22* partially blocks the effect of hs:FLP-13 and hs:FLP-24 (n=10). Error bars, mean s.e.m. ***p<0.0001, **p<0.001, *p<0.05, using two-tailed Student's t-test with unequal variance. n.s. indicates no significant difference. Unless otherwise stated, n=100 for each group.

Figure 6. Zebrafish FMRFamide neuropeptides induce *C. elegans* sleep through NPR-7 and NPR-22

(A) Schematic diagram of zebrafish FMRFamide neuropeptide (RFRP peptides) overexpression construct. Synthetic zebrafish RFRP neuropeptides were introduced into the FLP-24 prepropeptide backbone by replacing the single copy of FLP-24 neuropeptide and keeping the FLP-24 signal peptide and cleavage sites intact. The expression construct is attached to a heat shock promoter for conditional activation of the neuropeptide. See also Figure S5.

(B) Zebrafish RFRP-1 neuropeptide overexpression induces locomotion and feeding quiescence in *C. elegans* through neuropeptide receptors *npr-7* and *npr-22*. Behavioral quiescence is compared to the effect in wild-type animals with no hs:peptide (first sample). Wild type worms exhibit locomotion and feeding quiescence in response to overexpression of zebrafish RFRP-1. Mutation of *npr-7* blocks both behavioral quiescence and restored activities to the wild type level (locomotion p = 0.4; feeding p =

0.29). Mutation of *npr-22* partially blocks feeding quiescence in hs:RFRP-1 animals and has no effect on locomotion quiescence. See also Figures S6.

(C) Overexpression of zebrafish RFRP-1 increases response latency to 1-octanol in *C. elegans* and the effect is dependent on neuropeptide receptors *npr-7* and *npr-22*. Animal response agility to sensory stimulation using 1-octanol is completely restored to wild type level by mutation of *npr-7* ($p = 0.35$, n=10) and partially by mutation of *npr-22* ($p<0.05$, n=10). Error bars, mean s.e.m. ***$p<0.0001$, **$p<0.001$, *$p<0.05$, using two-tailed Student's *t*-test with unequal variance. n.s. indicates no significant difference. Unless otherwise stated, n=100 for each group.

Figure 7. Zebrafish RFRP induces sleep and alters sleep architecture.

(A) Zebrafish sleep/wake behavior in wild-type and hs:RFRP siblings. Heat shock (HS, yellow bar) denotes when larvae were moved from the videotracker to a 37°C water bath for 1 hour. RFRP overexpression inhibited locomotor activity and increased sleep (n=130).

(B) RFRP overexpression (red) increased sleep compared to wild-type (WT, n=118) sibling controls (blue) immediately after HS and during the following day (D7), but not during the intervening night (N6). **c-f,** Average number of sleep bouts, sleep bout length, wake bout length and time to first sleep bout after light transition were quantified for each day and night period pre- and post-HS. See also Figure S7. Error bar, mean s.e.m., averaged over 3 independent experiments. ***$p<0.0001$, using two-tailed Student's *t*-test. n.s. indicates no significant difference.

Figure 1.

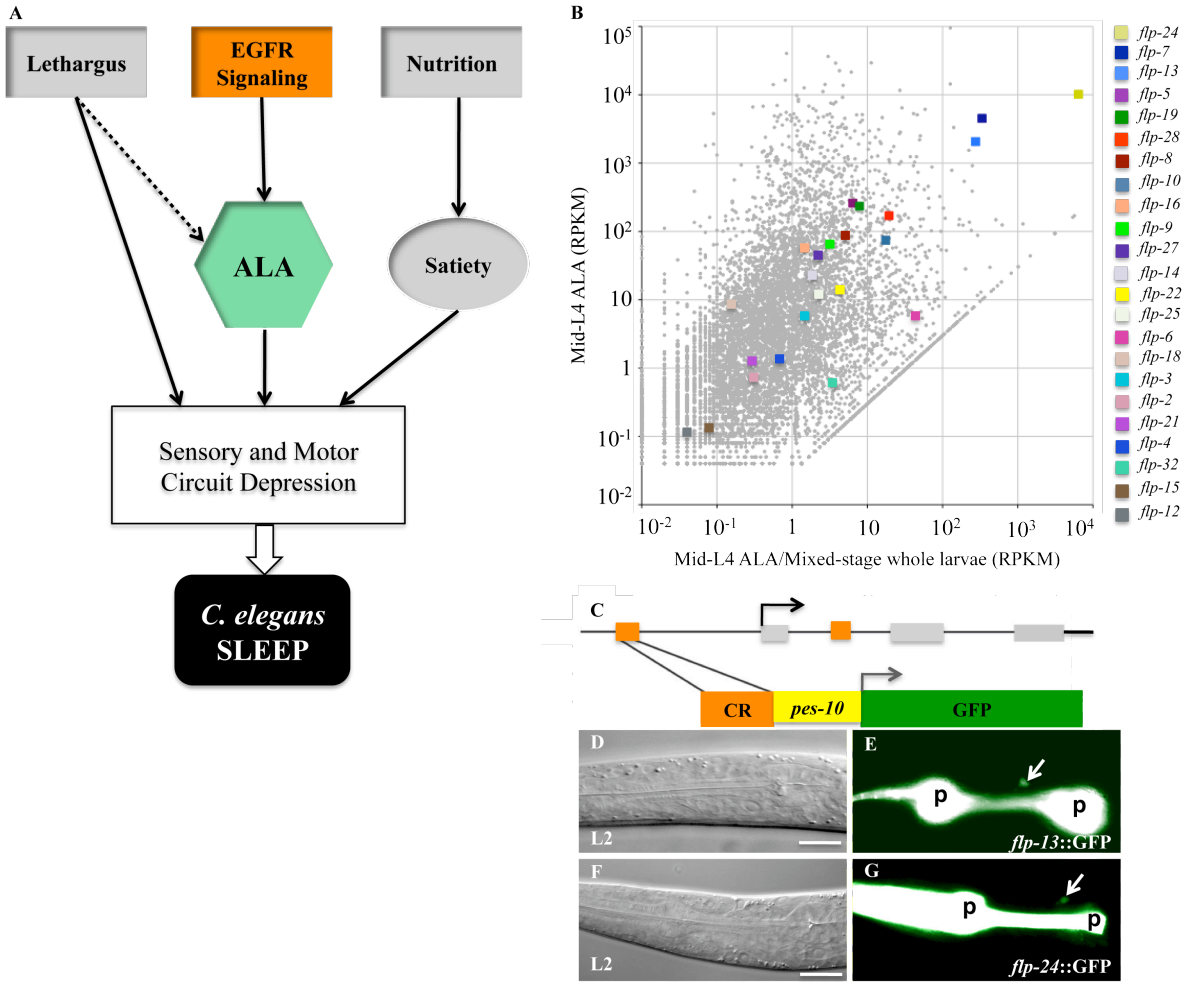
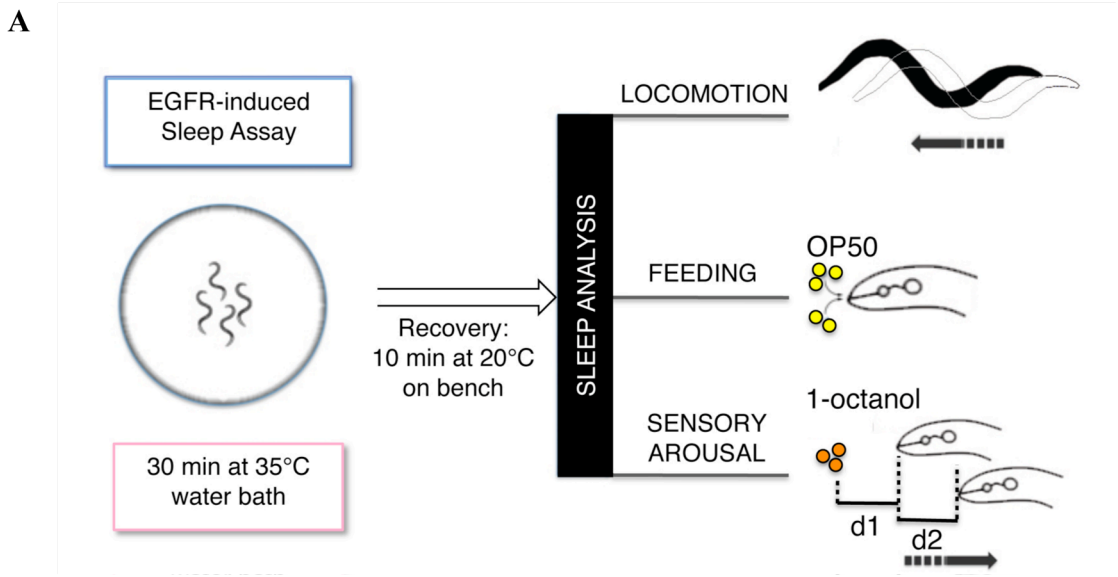

A
Lethargus
EGFR Signaling
Nutrition
ALA
Satiety
Sensory and Motor Circuit Depression
C. elegans SLEEP
B
Mid-L4 ALA (RPKM)
Mid-L4 ALA/Mixed-stage whole larvae (RPKM)
flp-24
flp-7
flp-13
flp-5
flp-19
flp-28
flp-8
flp-10
flp-16
flp-9
flp-27
flp-14
flp-22
flp-25
flp-6
flp-18
flp-3
flp-2
flp-21
flp-4
flp-32
flp-15
flp-12
C
CR
pes-10
GFP
D
L2
E
p
p
flp-13::GFP
F
L2
G
p
p
flp-24::GFP

Figure 2.


A
EGFR-induced
Sleep Assay
30 min at 35°C
water bath
Recovery:
10 min at 20°C
on bench
SLEEP ANALYSIS
LOCOMOTION
FEEDING
OP50
SENSORY
AROUSAL
1-octanol
d1
d2

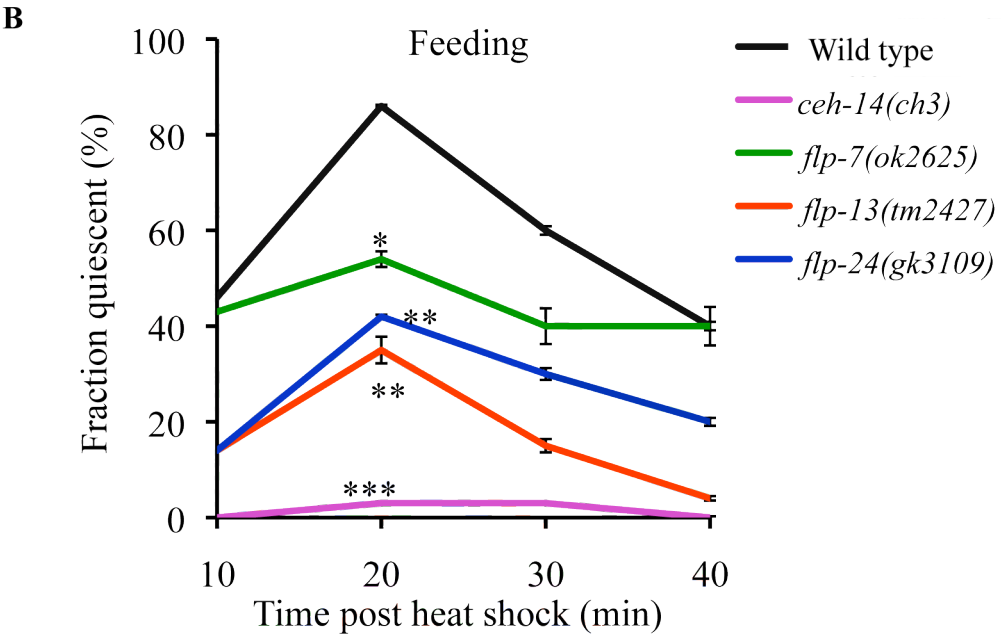

B
Feeding
Fraction quiescent (%)
100
80
60
40
20
0
10
20
30
40
Time post heat shock (min)
*
**
**
***
Wild type
ceh-14(ch3)
flp-7(ok2625)
flp-13(tm2427)
flp-24(gk3109)

Figure 3.

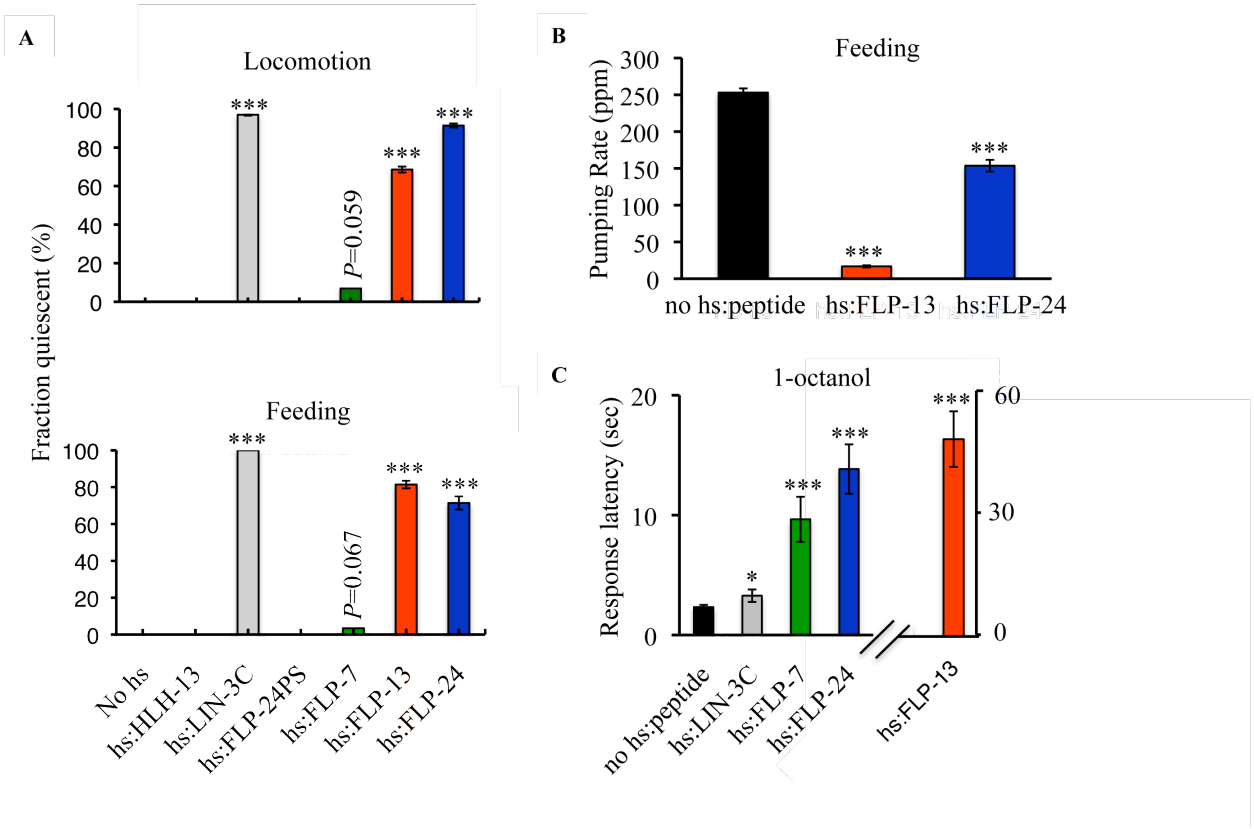

A
Locomotion
Fraction quiescent (%)
100
80
60
40
20
0
***
***
***
P=0.059
Feeding
***
***
***
P=0.067
No hs
hs:HLH-13
hs:LIN-3C
hs:FLP-24PS
hs:FLP-7
hs:FLP-13
hs:FLP-24
B
Feeding
Pumping Rate (ppm)
300
250
200
150
100
50
0
***
***
no hs:peptide
hs:FLP-13
hs:FLP-24
C
1-octanol
Response latency (sec)
20
10
0
60
30
0
*
***
***
***
no hs:peptide
hs:LIN-3C
hs:FLP-7
hs:FLP-24
hs:FLP-13

Figure 4.

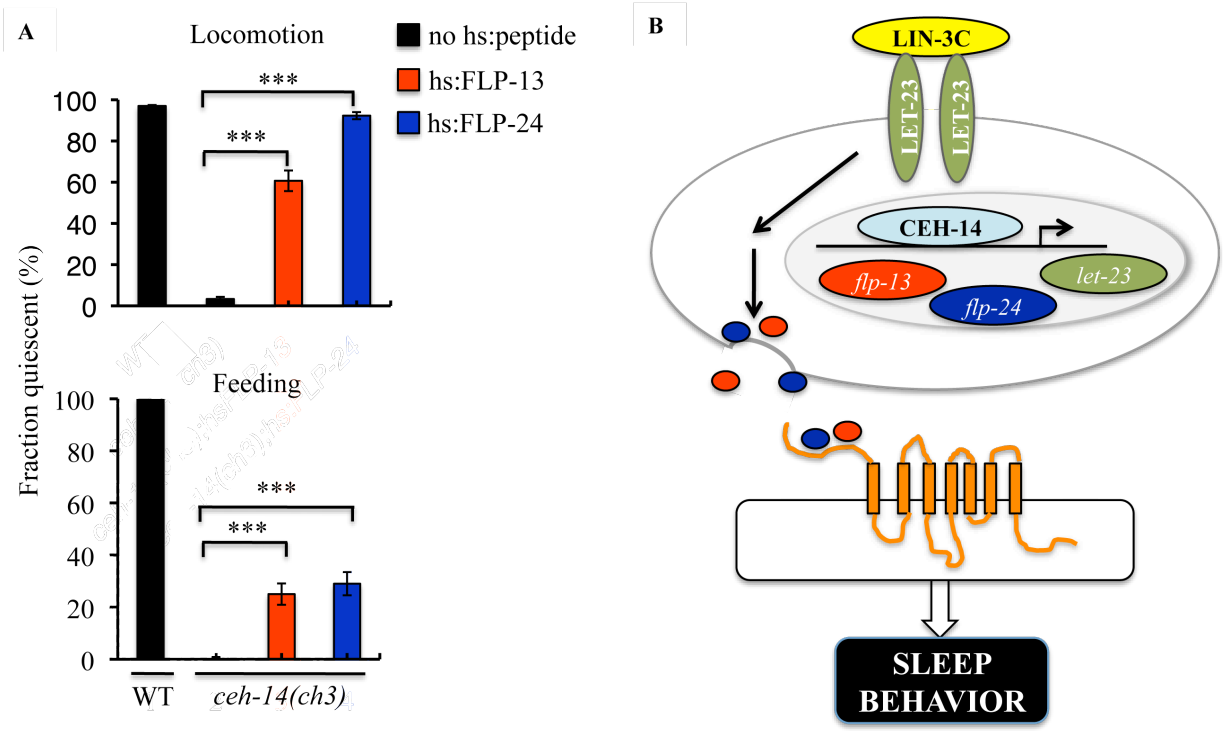

A
Locomotion
no hs:peptide
hs:FLP-13
hs:FLP-24
Fraction quiescent (%)
100
80
60
40
20
0
***
***
Feeding
***
***
WT
ceh-14(ch3)
B
LIN-3C
LET-23
LET-23
CEH-14
flp-13
flp-24
let-23
SLEEP
BEHAVIOR

Figure 5.

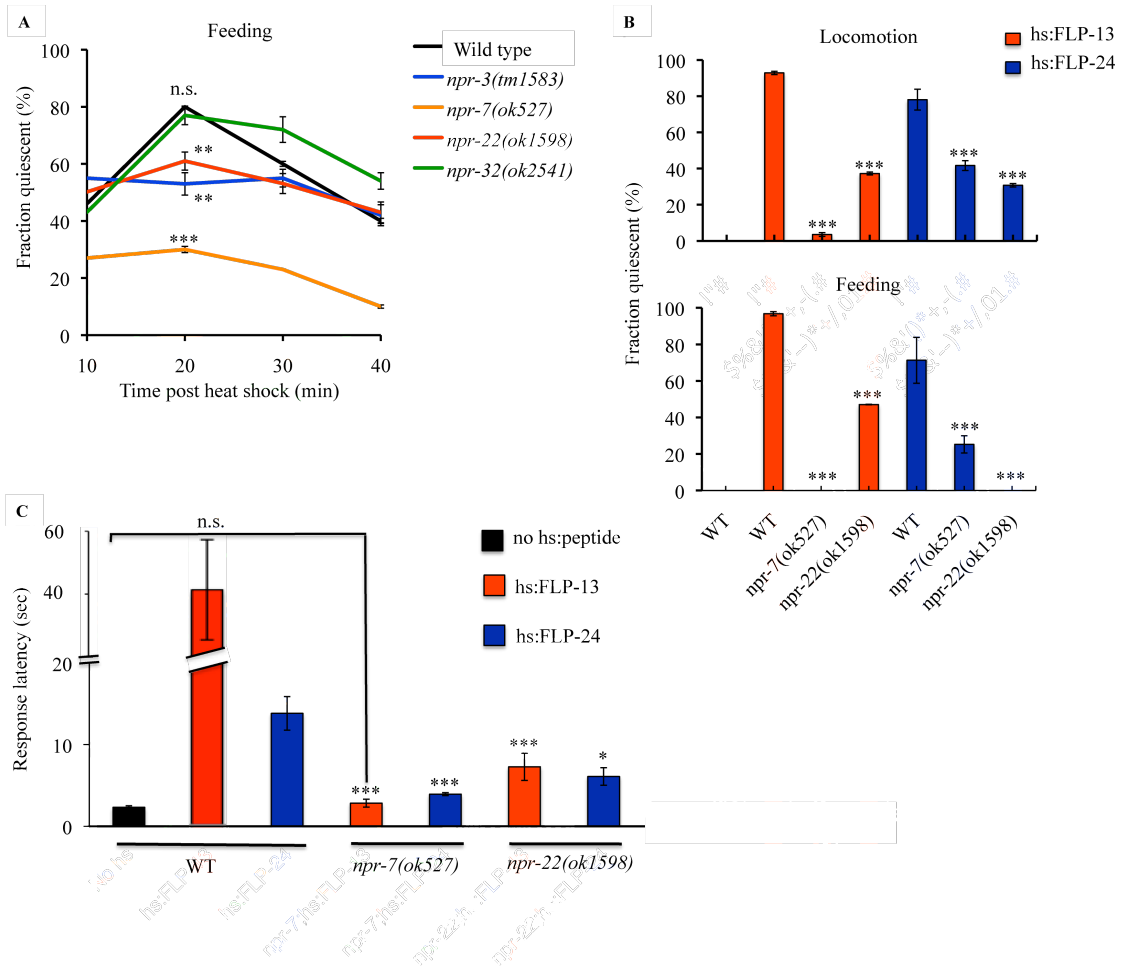

A
Feeding
Wild type
npr-3(tm1583)
npr-7(ok527)
npr-22(ok1598)
npr-32(ok2541)
Fraction quiescent (%)
Time post heat shock (min)
n.s.
B
Locomotion
hs:FLP-13
hs:FLP-24
Fraction quiescent (%)
Feeding
WT
WT
npr-7(ok527)
npr-22(ok1598)
WT
npr-7(ok527)
npr-22(ok1598)
C
n.s.
no hs:peptide
hs:FLP-13
hs:FLP-24
Response latency (sec)
WT
npr-7(ok527)
npr-22(ok1598)

Figure 6.

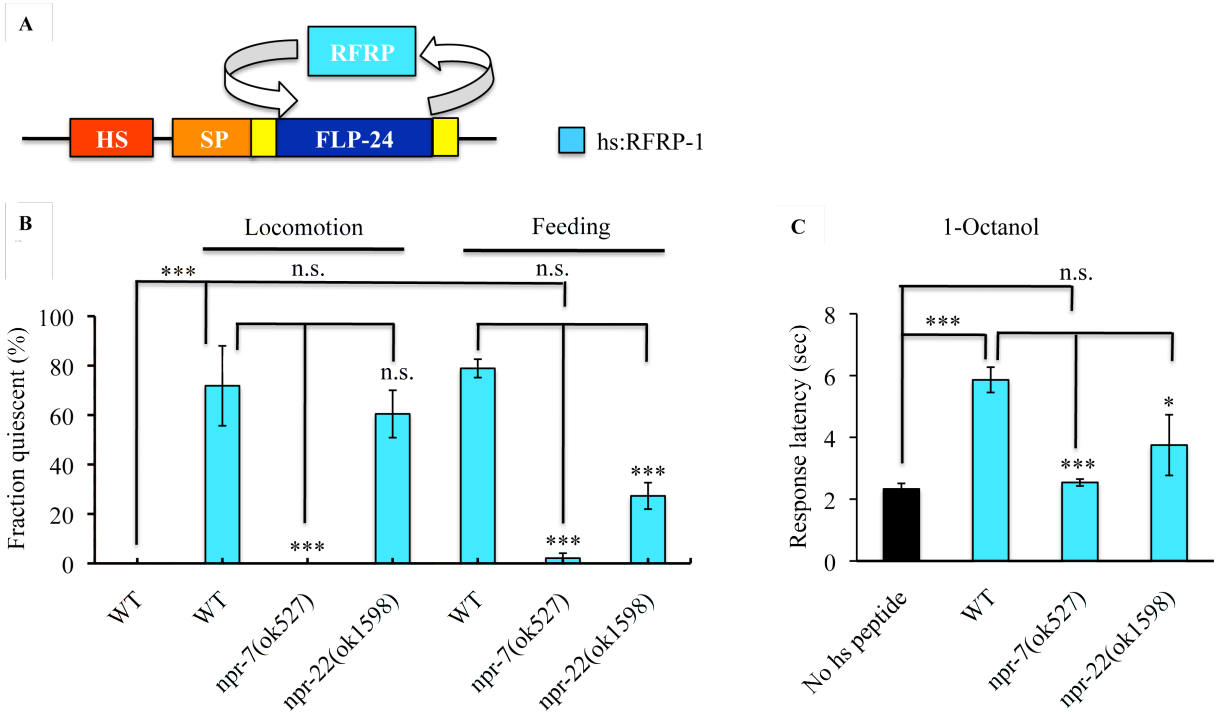
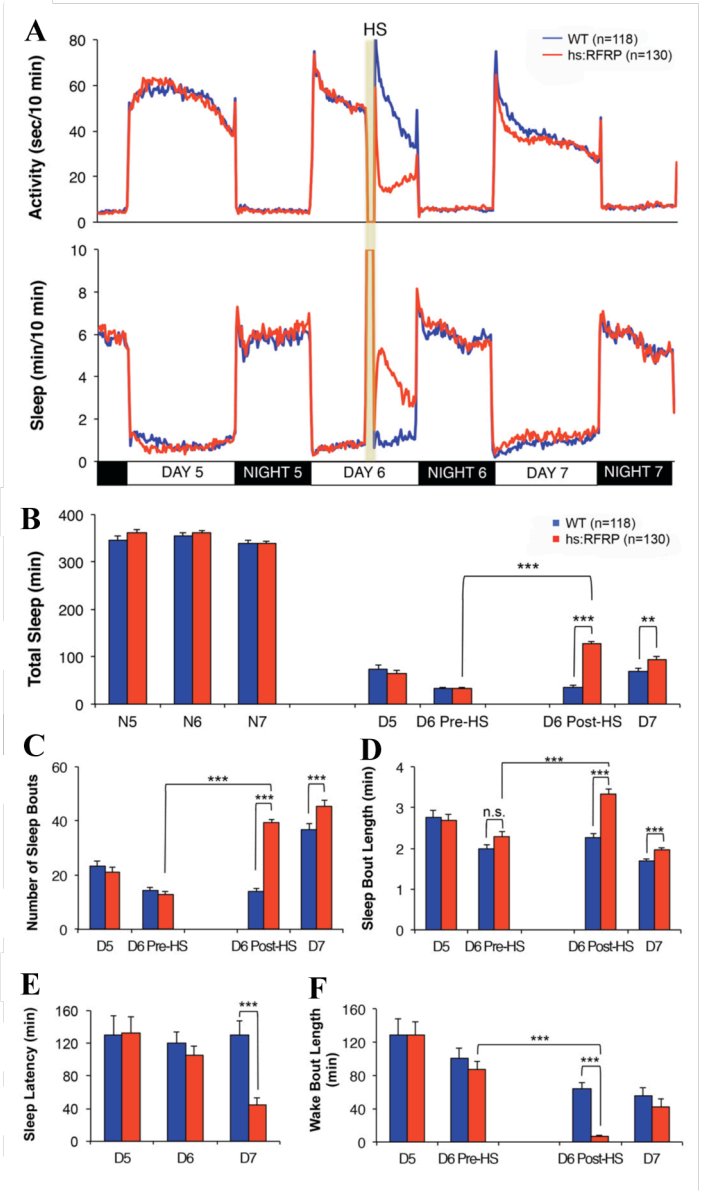

A
RFRP
HS
SP
FLP-24
hs:RFRP-1
B
Locomotion
Feeding
Fraction quiescent (%)
100
80
60
40
20
0
***
n.s.
n.s.
n.s.
***
***
***
WT
WT
npr-7(ok527)
npr-22(ok1598)
WT
npr-7(ok527)
npr-22(ok1598)
C
1-Octanol
Response latency (sec)
8
6
4
2
0
n.s.
***
***
*
No hs peptide
WT
npr-7(ok527)
npr-22(ok1598)

Figure 7.

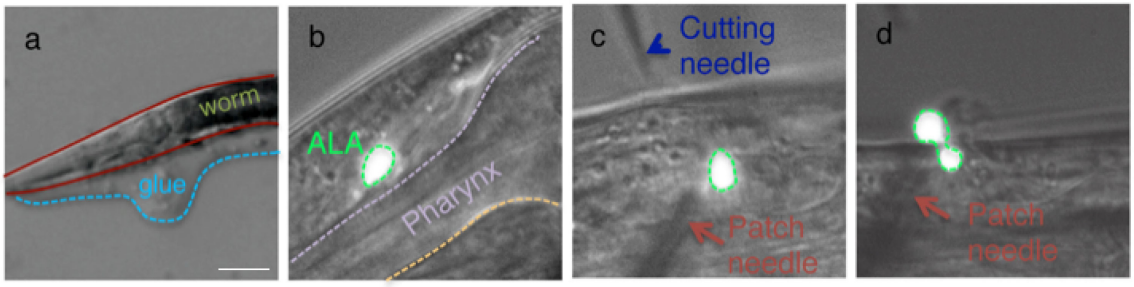

A
HS
WT (n=118)
hs:RFRP (n=130)
Activity (sec/10 min)
Sleep (min/10 min)
DAY 5
NIGHT 5
DAY 6
NIGHT 6
DAY 7
NIGHT 7
B
Total Sleep (min)
N5
N6
N7
D5
D6 Pre-HS
D6 Post-HS
D7
C
Number of Sleep Bouts
D
Sleep Bout Length (min)
n.s.
E
Sleep Latency (min)
D5
D6
D7
F
Wake Bout Length (min)

Supplemental Information

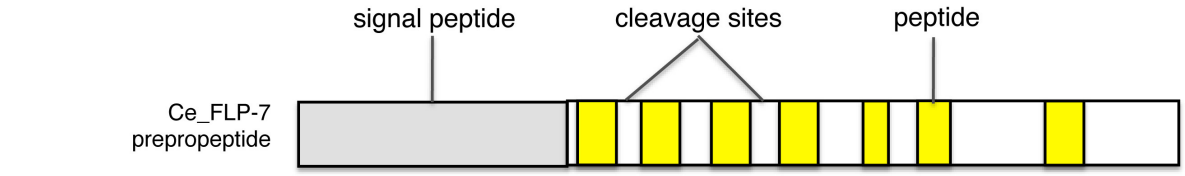
FMRFamide neuropeptides promote sleep in *Caenorhabditis elegans* and zebrafish (*Danio rerio*)

Elly Suk Chow, Daniel A. Lee, Erich M. Schwarz, David A. Prober, Paul W. Sternberg

Supplemental Figure 1



Supplemental Figure 2A



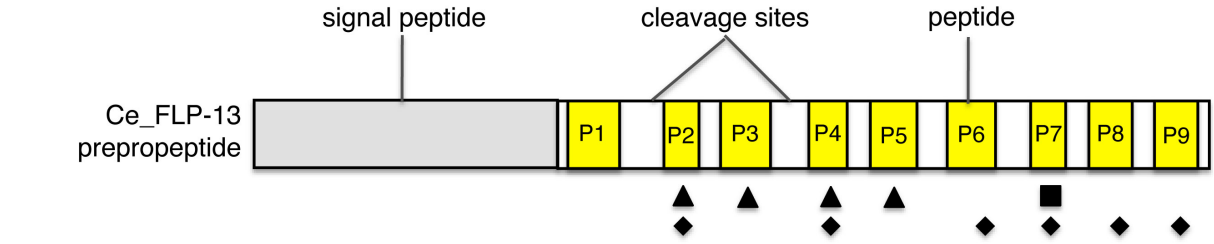

```
C. elegans   -MLGSRFLLLALGLLVLVLAEESAEQQ-V--QEPTELEKSG-EQLSEEDLIDEQKRTPMQ   55  F49E10.3
C. brenneri  MMTSLVFA-MMFAVA-AVQAFDSSEIRMLDEQ----YDKSPYFPFLENEKR-SDRPTRAM   53  CBN24837
C. remanei   -MLGSRFLLLALGLFVLVWAEKSTEQQ-V--QEQTDLDKSG-EQLSEEDLIDEQKRNPMQ   55  CRE00688
C. briggsae  MLGSPRFLLLALGLLVLVLAESSVEQQV---QDQTDLDKSG-EQLSEEDIIEEQKRSPME   56  CBG14771
C. japonica  -MLGPRFFLLALGLLALIRADESTNEQQIQEQEQNELDKSEEQQLTDGDLIDEQKRSPMQ   59  CJA00701
                : .  *  : :.:   : * .* : :    *     :**    : : :   .:: .

C. elegans   RSSMVRFGRSPMQRSSMVRFGKRSPMQRSSMVRFGKRSPMQRSSMVRFGK----------  105  F49E10.3
C. brenneri  DSPLIRFGKRAADGAPLIRFGRAP--EASPFIRFGKRAA-DGAPLIRFGRAP--EASPFI  108  CBN24837
C. remanei   RSSMVRFGRSPMQRSSMVRFGKRSPMQRSSMVCFGKRSPMQRSSMVRFGK----------  105  CRE00688
C. briggsae  RSSMVRFGRSPMERSSMVRFGKRSPMERSAMVRFGKRSPMDRSAMVRFGKRLPSDRSSMV  116  CBG14771
C. japonica  RSSMVRFGRSSNQ---MSQFEKRSPMQRSSMVRFGKRSPMQRAAMVRFGK----------  106  CJA00701
              * ::****:   :   : :* :    : * :: ****:  : : ::***:

C. elegans   ----RSPMERSAMVRFG-RSPMDRSKMVRFGRSSIDRASMV--RLGKRTPMQRSSMVRFG  158  F49E10.3
C. brenneri  RFGKRAS-PSAPLIRFGRS-P-SSAPLIRFGRSAGAP--LI--RFGR---ASSAPLIRFG  158  CBN24837
C. remanei   ----RSPMERSAMVRFG-RSPMDRSKMVRFGRSSIDRASMV--RLGKRTPMQRSSMVRFG  158  CRE00688
C. briggsae  RLGKRSPMDRSAMVRFGKRSPMDRSAMVRFGRSSIDRASMV--RLGKRTPIQRSSMVRFG  174  CBG14771
C. japonica  ----RSPMERSAMVRFG-RSAMDRSSMVRFGRSSIDRASMVSSRLGKRSPMQRSSMVRFG  161  CJA00701
                 *:    : ::***     . : ::*****:     ::  *:*:    . : ::***

C. elegans   KRSMEFEMQSNEKN--IEDSE  177  F49E10.3
C. brenneri  KK------------------  160  CBN24837
C. remanei   KRSAPSDINEIQDNEQIRDNE  179  CRE00688
C. briggsae  KRSADETENTNE----------  186  CBG14771
C. japonica  KRSTQQSSSEN----------  172  CJA00701
             *:
```

Supplemental Figure 2B

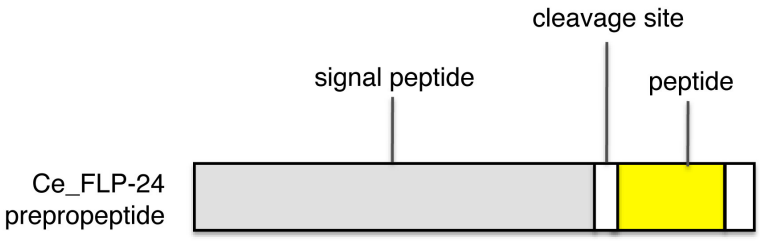



```
C. elegans  ---MMTSLLT-ISMFVVAIQAFDSSEIRMLDEQYDTK---NPFFQFLENSKRSDRPTRAM   53  O44185
C. brenneri ---MMTSLLVIPMMFVVAIQAFDSSEIRMLDEQYETN---HPYFPFLEQ-KRSDRPTRAM   53  A8X1A3
C. remanei  ---MMTPFLVISLMFVVSIQAFDSSEIRMLDEQYEPK---NPYFPFLENEKRSDRPTRAM   54  E3M7H9
C. briggsae ---MMTSLVFAMMFAVAAVQAFDSSEIRMLDEQYD-K---SPYFPFLENEKRSDRPTRAM   53  G0P6W9
C. japonica ---MMTSLAVLLVVFVATSEAFDSSEIRMLDEQYDTK---NPFLAYLENQKRSDRPTRAL   54  H2W239
A. suum     MAINSISCLITMLAITVAVSTLESSEIRMLESEFLDGKRGNSYRSFFNPNEY-TVVERRD   59  D9I8P2
                       ..: .:::********:.::        :  :::  :       *

C. elegans  ---DSPLIRFGKRAADGAPLIRFGRAPEASPFIRFGKRAADGAPLIRFGRAPEASPFIRF  110  O44185
C. brenneri ---DSPLIRFGKRAADGAPLIRFGRAPEASPFIRFGKRAADGAPLIRFGRAPEASPFIRF  110  A8X1A3
C. remanei  ---DSPLIRFGKRAADGAPLIRFGRAPEASPFIRFGKRAADGAPLIRFGRAPEASPFIRF  111  E3M7H9
C. briggsae ---DSPLIRFGKRAADGAPLIRFGRAPEASPFIRFGKRAADGAPLIRFGRAPEASPFIRF  110  G0P6W9
C. japonica DSLDSPLIRFGKRAADGAPLIRFGRAPEASPFIRFGKRA-SAVPLIRFGRAPEASPFIRF  113  H2W239
A. suum     SKLMDPLIRFGKRTNIMGEN-RLNR---------------------NLRAEGLSSPLIRF   97  D9I8P2
               .********:   .   *: *                    .:     :**:***

C. elegans  GKRASPSAPLI-RFGRSPS-----AVPLIRFGRSAAAPLIRFGRASSAPLIRFGRK  160  O44185
C. brenneri GKRAAPSAPLI-RFGRSPS-----AAPLIRFGRSAAAPLIRFGRASSAPLIRFGRK  160  A8X1A3
C. remanei  GKRASPSAPLI-RFGRSPS-----AVPLIRFGRSAAAPLIRFGRASSAPLIRFGRK  161  E3M7H9
C. briggsae GKRASPSAPLI-RFGRSPS-----SAPLIRFGRSAGAPLIRFGRASSAPLIRFGKK  160  G0P6W9
C. japonica GKRASPSAPLI-RFGRSPA-----AVPLIRFGRSAAAPLIRFGRK----------  152  H2W239
A. suum     GKRTPPEEDLLGRFTRDPQQRIVTDETVLRFGRSSAPAAA---------------  137  D9I8P2
            ***: *.  *: ** *.*         ::******:.
```

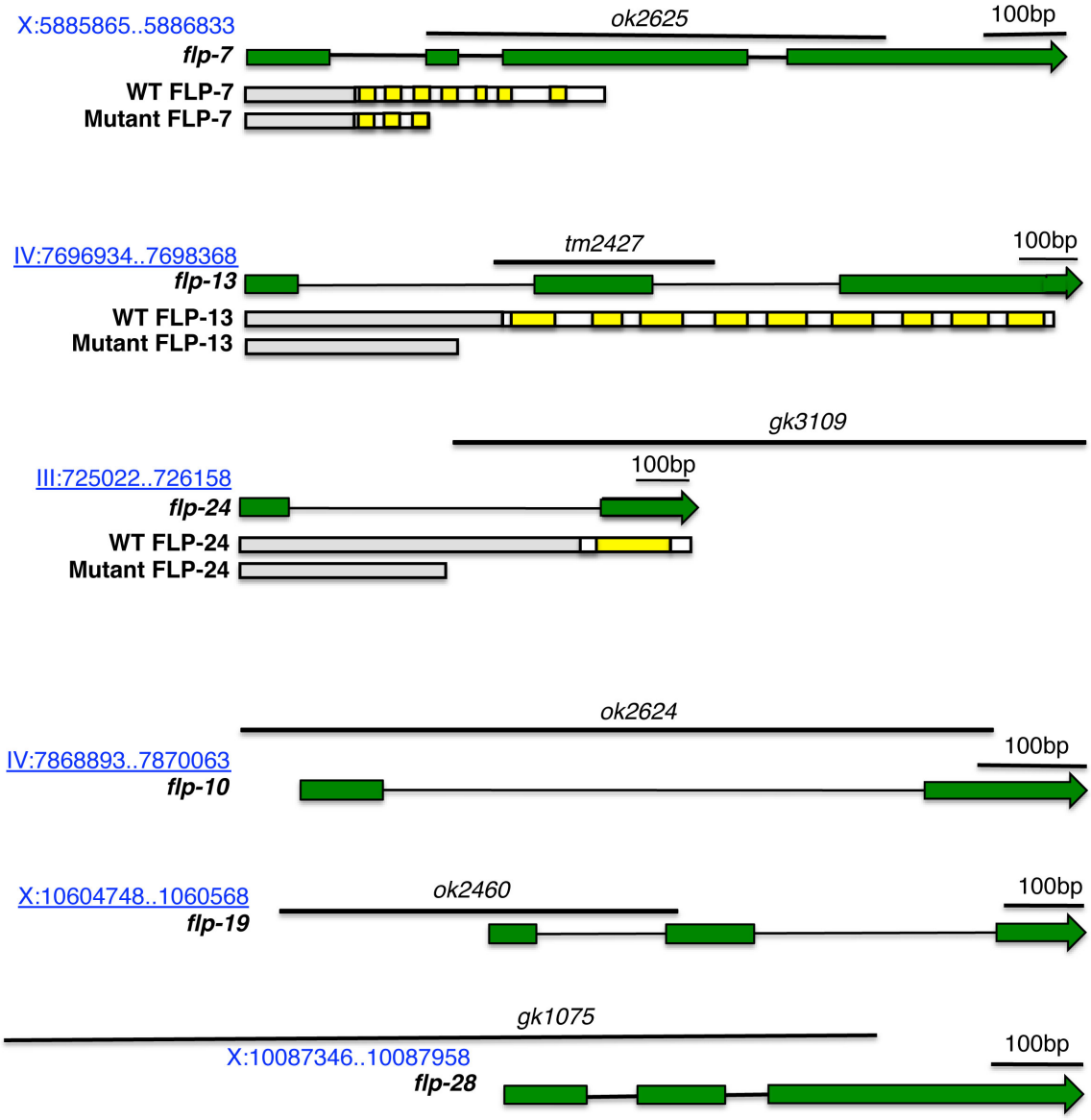
Supplemental Figure 2C



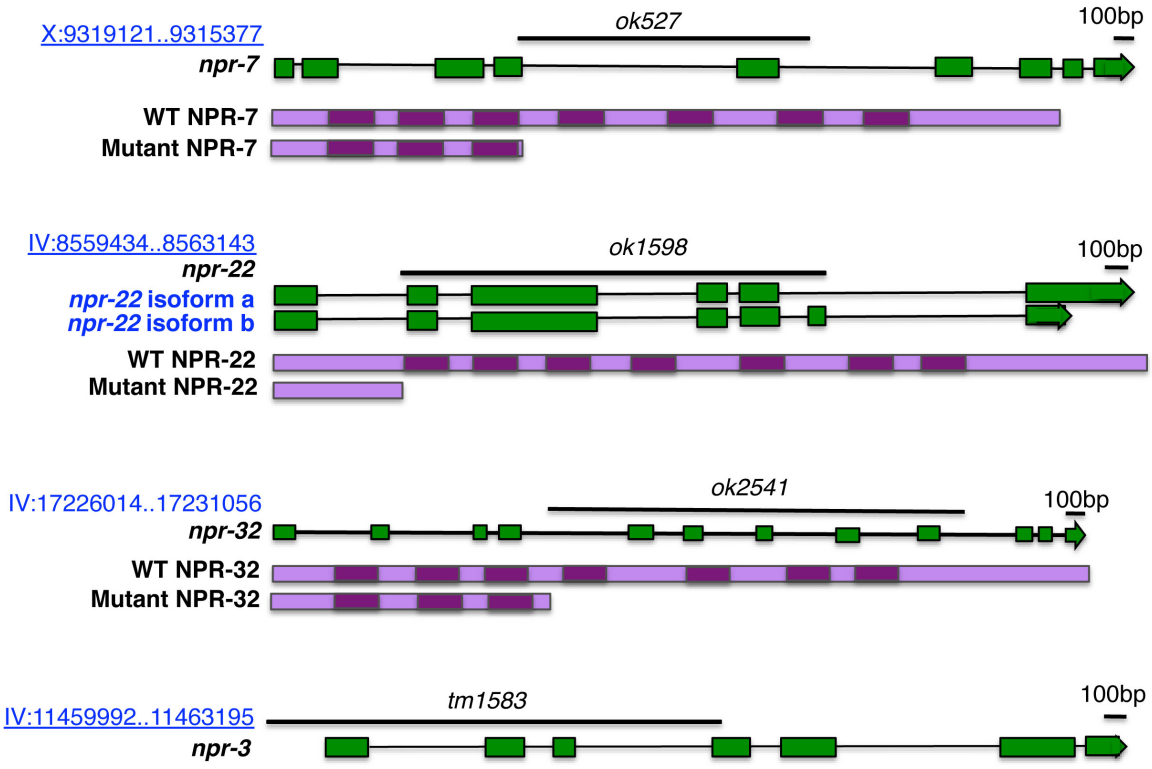

```
C. elegans     MLSSRTSSIILILAILVAIMAVAQCRNIQ-YDVE------EMTPEAAFRYAQWG-EIPHK   52  O17058
C. brenneri    -MISSRSSIILILAILVAIMAVAQCRNIQ-YDID------EIAPESAFRYAQWG-EIPHK   51  G0MYY2
C. remanei     -MLSSRASIVLIFAILVAIMAVAQCRNIQ-YDID------EITPEAAFRYAQWG-EIPHK   51  E3MLA0
C. briggsae    ---MSRTSIILVLAIFVAIAAIAQCRNIQ-YDVD------EISPEAAFRYAQWG-EIPHK   49  A8XLL0
P. pacificus   -----MKSILIILLSFILLLSLSSARNIFPYGPV------GQMGEINAEEREWMIPSMAK   49  H3ENH6
A. suum        --MFSLKAIVMIALVVICTFCISESRRFHDDDFSRQFLFRGIDEPLKNYMRLREARILSK   58  Q5ENY8
                      :*:::   .:   .:::.*.:                                 *

C. elegans     RVPSAGDMMVRFGKRSI------   69  O17058
C. brenneri    RVPSAGDMMVRFGKRAV------   68  G0MYY2
C. remanei     RVPSAGDMMVRFGKRSV------   68  E3MLA0
C. briggsae    RVPSAGDMMVRFGKRSV------   66  A8XLL0
P. pacificus   RVPSAGDMMVRFG----------   62  H3ENH6
A. suum        RVPSAADMMIRFGKRSFIEQDME   81  Q5ENY8
               *****.***:***
```

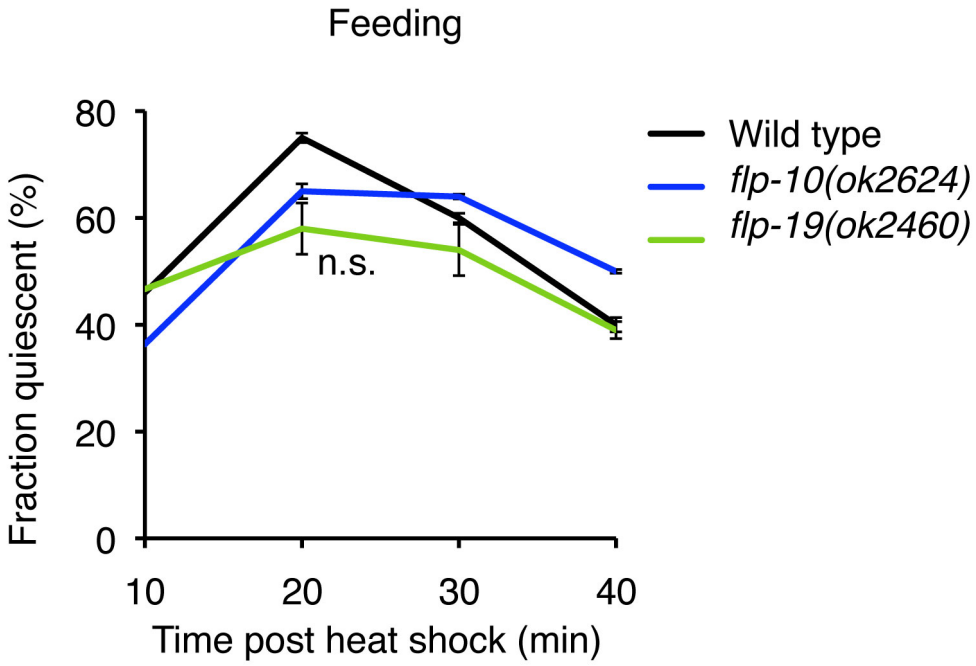
Supplemental Figure 3A



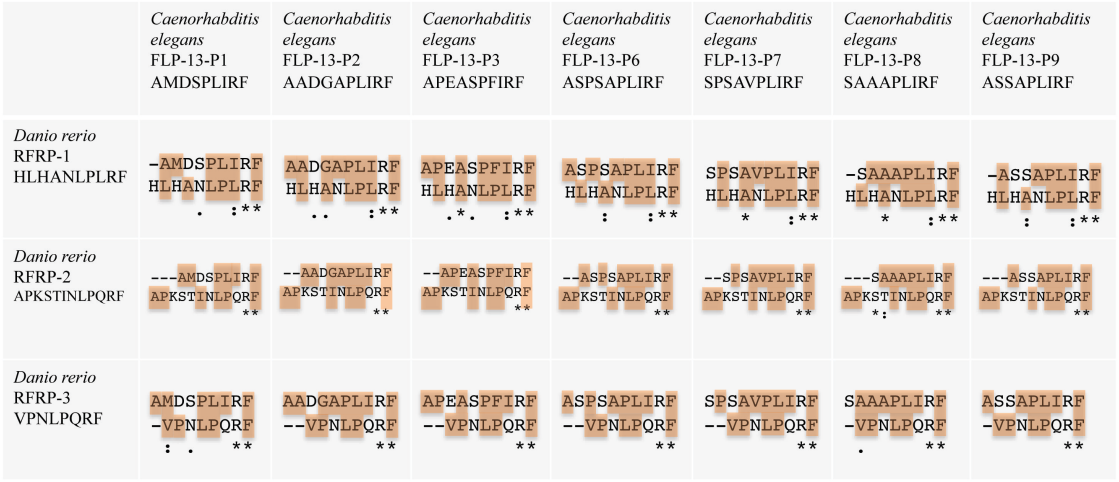
Supplemental Figure 3B



Supplemental Figure 3C



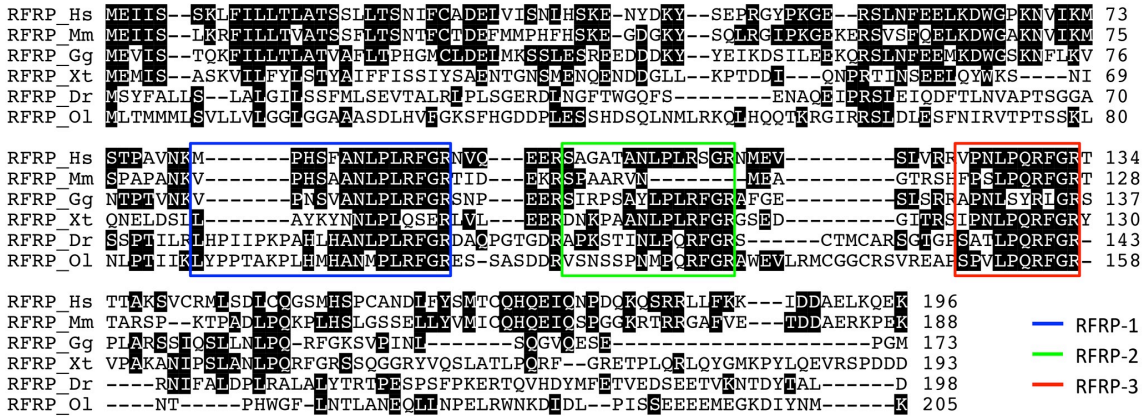
Supplemental Figure 4A



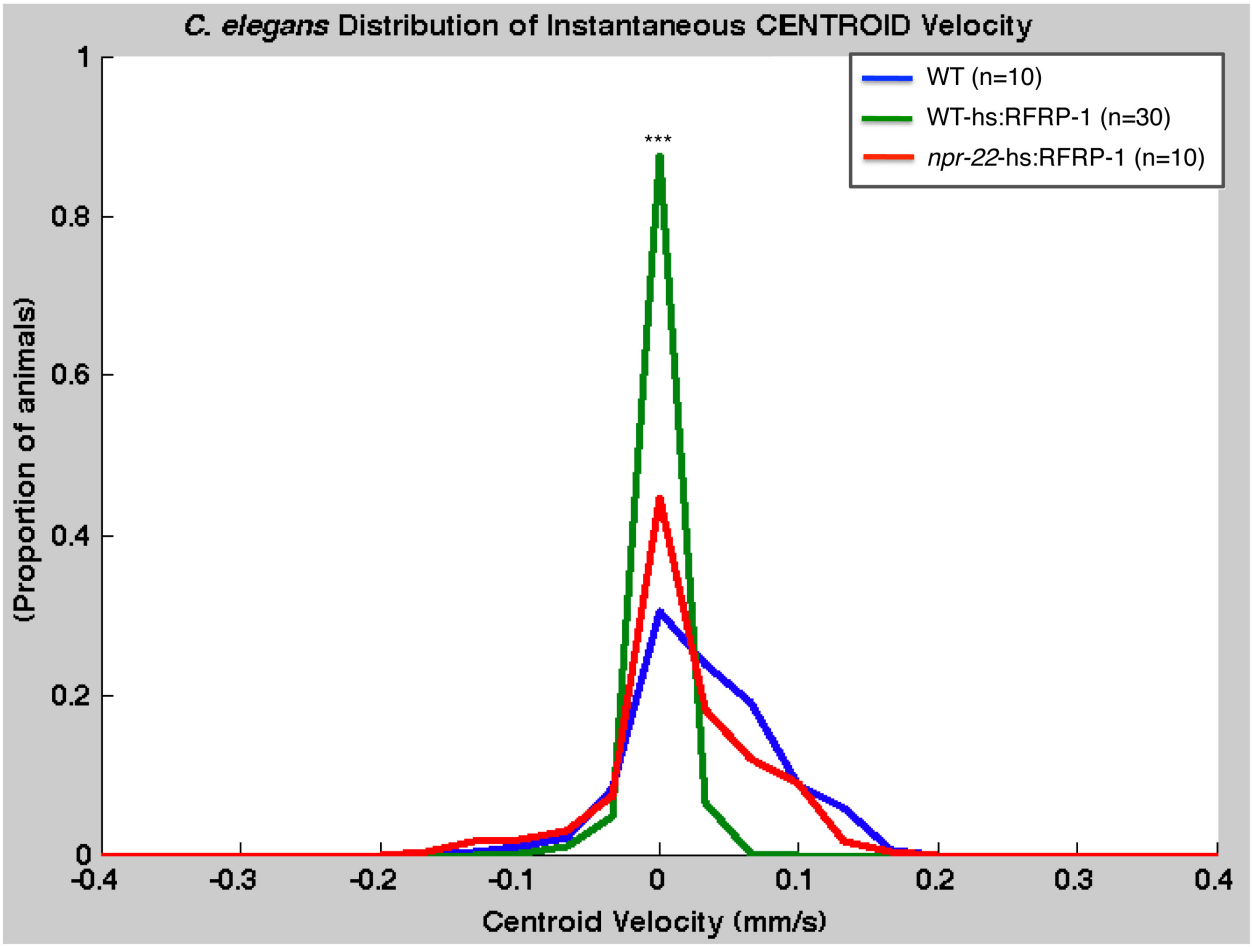
Supplemental Figure 4B

	Caenorhabditis elegans FLP-24 VPSAGDMMVRF
Danio rerio RFRP-1 HLHANLPLRF	VPSAGDMMVRF -HLHANLPLRF .:: :**
Danio rerio RFRP-2 APKSTINLPQRF	VPSAGDMMVRF APKSTINLPQRF *: :: **
Danio rerio RFRP-3 VPNLPQRF	VPSAGDMMVRF -VPNLPQRF . **

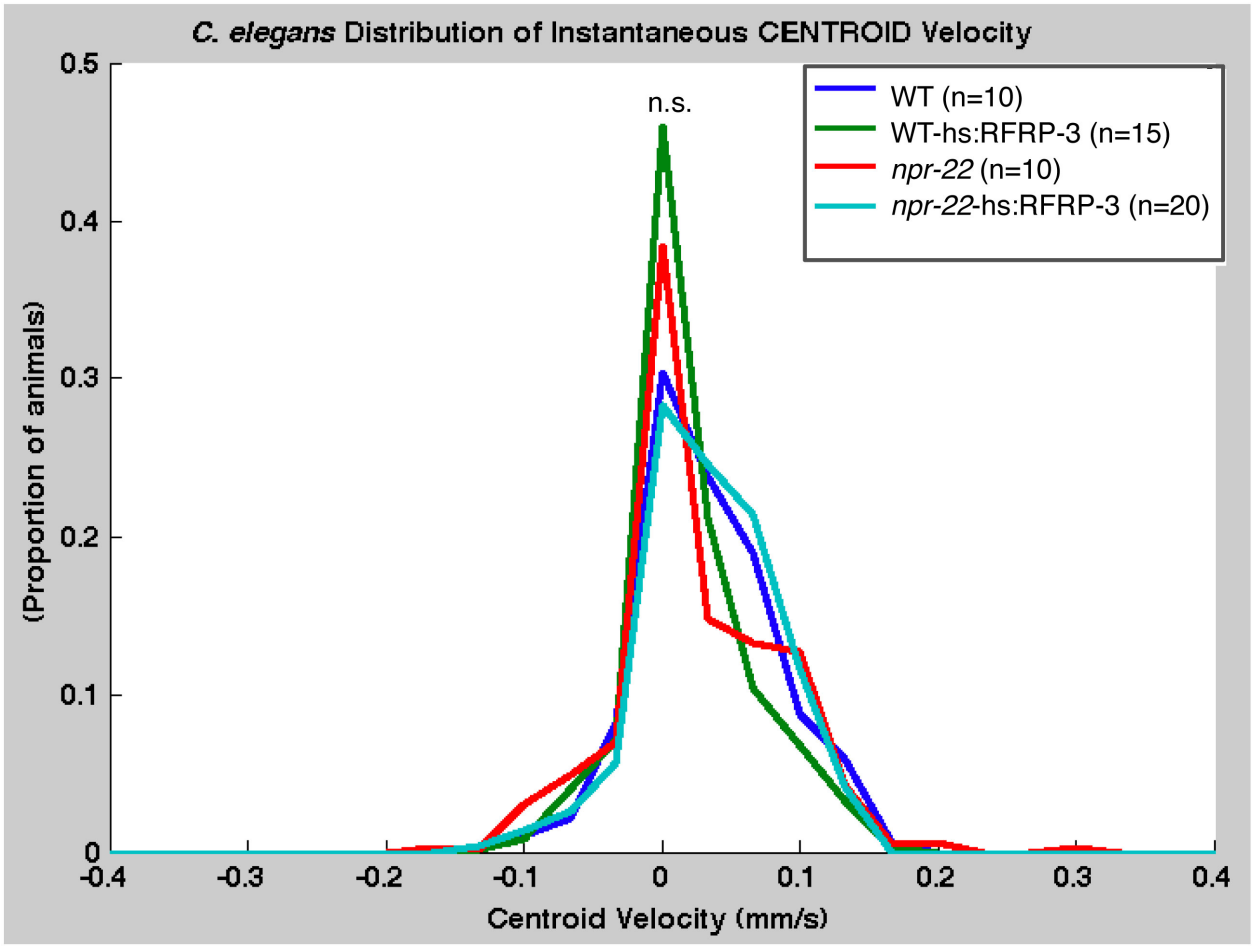
Supplemental Figure 5



Supplemental Figure 6A



Supplemental Figure 6B



Supplemental Figure 7

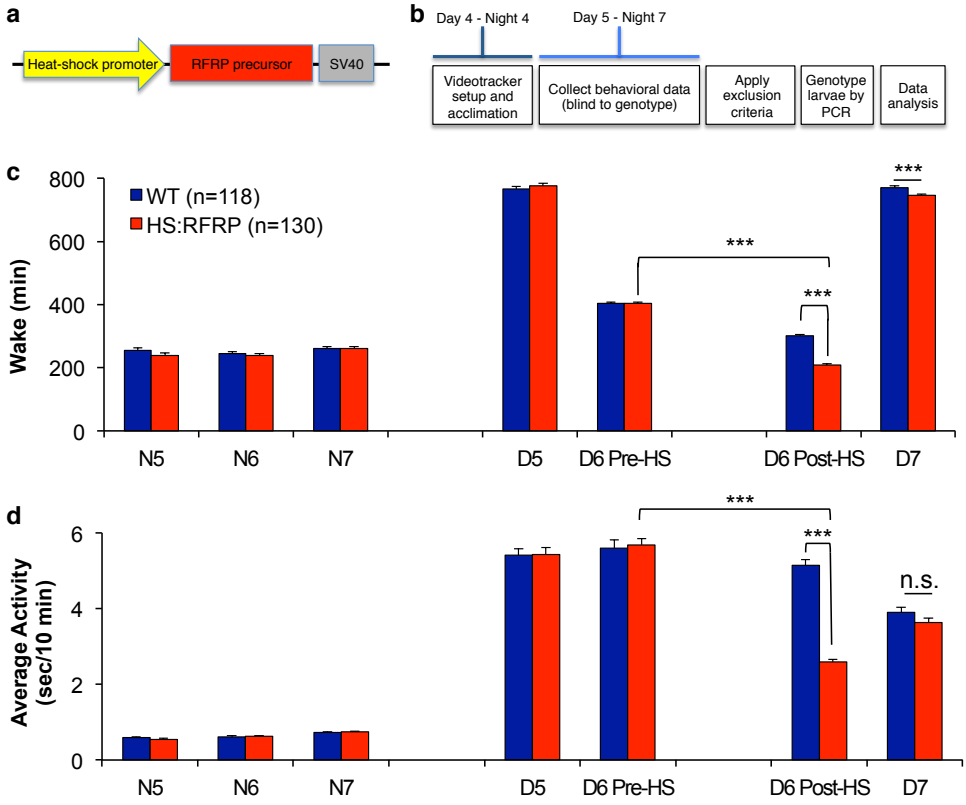


Figure S1. (related to Figure 1) Microdissection of ALA neuron from mid-L4 larva.

Mid-L4 larvae were adhered to a freshly made agar pad with dental glue along the ventral bodyline (Schwarz et al., 2012). (B) The ALA neuron was identified as dorsal to the pharynx and labeled with a GFP reporter driving *ceh-14* expression in the ALA neuron (green circle), the only dorsal head neuron expressing *ceh-14*::GFP. (C) A fine glass cutting needle (blue arrowhead) was used to cut open the dorsal worm body close to the vulva to release body pressure (not shown), and a small puncture was made in the dorsal head just big enough to release the ALA neuron. (D) A glass patch needle (red arrow) was used to collect the released ALA neuron. Anterior is left. Dorsal is up. Scale bar represents 20 μm.

Figure S2. *C. elegans flp-7, flp-13,* and *flp-24* are conserved nematode neuropeptide-coding precursor genes.

(A) *flp-7* encodes a propeptide that generates seven mature FMRFamide neuropeptides. Neuropeptide annotation was based on previous report (Li and Kim, 1999) and EnsemblMetazoa (http://metazoa.ensembl.org/Caenorhabditis_elegans/). Sequences were obtained from EnsemblMetazoa (http://metazoa.ensembl.org/Caenorhabditis_elegans/) and sequence alignment was conducted by www.uniprot.org. Signal peptide and transmembrane helix: grey box; cleavage site: white box; FMRFamide peptide: yellow box. Shown are amino acid sequences of FLP-7 in nematodes. *Caenorhabditis elegans* (F49E10.3), *Caenorhabditis brenneri* (CBN24837), *Caenorhabditis remanei* (CRE00688). *Caenorhabditis briggsae* (CBG14771), *Caenorhabditis japonica* (CJA00701).

(B) *flp-13* encodes a propeptide that generates nine mature neuropeptides (P1-P9): P2 and P4, P3 and P5 are repeated copies and have inhibitory effects (triangles) on *Ascaris suum* muscle strips (Marks et al., 1997), while P7 has an excitatory effect (square) on *A. suum* muscle strips (Marks et al., 2001). P2, P4, P6, P7, P8, and P9 can activate calcium response via NPR-22 *in vitro* (Mertens et al., 2006; diamonds). Neuropeptide annotation and sequence alignment were conducted by www.uniprot.org. Signal peptide and transmembrane helix: grey box; cleavage site: white box; FMRFamide peptide: yellow box. Shown are amino acid sequences of FLP-13 in nematodes (www.uniprot.org). *Caenorhabditis elegans* (O44185), *Caenorhabditis brenneri* (G0P6W9), *Caenorhabditis remanei* (E3M7H9), *Caenorhabditis briggsae* (A8X1A3), *Caenorhabditis japonica* (H2W239), *Ascaris suum (*D9I8P2).

(C) *flp-24* encodes a propeptide that generates one mature neuropeptide. Neuropeptide annotation and sequence alignment were conducted by www.uniprot.org. Signal peptide and transmembrane helix: grey box; cleavage site: white box; FMRFamide peptide: yellow box. Shown are amino acid sequences of FLP-24 in nematodes (www.uniprot.org). *Caenorhabditis elegans* (O44185), *Caenorhabditis brenneri* (G0P6W9), *Caenorhabditis remanei* (E3M7H9), *Caenorhabditis briggsae* (A8X1A3), *Pristionchus pacificus* (H3ENH6), *Ascaris suum (*Q5ENY8).

Figure S3 (related to Figure 2) Gene models with deletions in mutant alleles.

(A) Shown are gene models of *flp-7, flp-10, flp-13, flp-19, flp-24,* and *flp-28* with deletion mutations, along with protein structure and domains annotated by

EnsemblMetazoa (http://metazoa.ensembl.org/Caenorhabditis_elegans/). Horizontal black bars indicate genomic deletions (www.wormbase.org) and green blocks represent exons of coding genes. Grey boxes indicate signal peptide and yellow boxes indicate mature peptides.

(B) Shown are gene models of *npr-3, npr-7, npr-22,* and *npr-32* with deletion mutations, along with protein structure and domains annotated by EnsemblMetazoa (http://metazoa.ensembl.org/Caenorhabditis_elegans/). Horizontal black bars indicate genomic deletions (www.wormbase.org) and green blocks represent exons of coding genes. Lavender boxes indicate protein backbone. Dark purple boxes indicate transmembrane domains.

(C) EGFR-induced sleep does not require all ALA-enriched neuropeptides. Time course of feeding quiescence in wild-type compared with mutants of neuropeptides synthesized in ALA 20 minutes after heat shock. Animals lacking *flp-10* and *flp-19* are not resistant to sleep induction. N > 50.

Figure S4. (related to Figure 6) FLP-13 peptides share similar structure with zebrafish RFRP peptides. (A) Shown are alignments of FLP-13 peptides (Q44185) with zebrafish RFRP-1, RFRP-2, and RFRP-3 peptides (Q9HCQ7) by www.uniprot.org. Hydrophobic amino acids are shaded in orange. Asterisks (*) mark identical amino acids. Dots (. and :) mark similar amino acids.

(B) FLP-24 peptide shares similar structure with zebrafish RFRP peptides. Shown are alignments of an FLP-24 peptide (O17058) with zebrafish RFRP-1, RFRP-2, and RFRP-3 peptides (Q9HCQ7) by www.uniprot.org. Hydrophobic amino acids are shaded in orange. Asterisks (*) mark identical amino acids. Dots (. and :) mark similar amino acids.

Figure S5. (related to Figure 6) Conservation of RFRP among vertebrates.

(A) Alignment of RFRP peptide encoding amino acid sequence among vertebrates is shown. The RFRP propeptide is cleaved into three mature RFRP peptides: RFRP-1, RFRP-2, and RFRP-3. An alignment of the human (Hs; ENSG00000105954), mouse (Mm; ENSMUSG00000029831), chicken (Gg; ENSGALG00000011022), frog (Xt; ENSXETG00000031334), zebrafish (Dr; ENSDARG00000036227), and medaka (Ol; GENSCAN00000071979) orthologs is shown. Amino acids identical to the consensus are shaded in black. The designation of each mature peptide sequence is based on the human peptide (Ubuka et al, 2008).

Figure S6 (related to Figure 6) (A) Overexpression of zebrafish RFRP-1 peptide suppresses locomotion in *C. elegans*. Forward locomotion (positive centroid velocity) and reverse locomotion (negative centroid velocity) were recorded in animals beginning 2 hours after heat shock. Overexpression of zebrafish RFRP-1 peptide in *C. elegans* significantly reduced locomotor activity in wild-type animals, and the effect was blocked in *npr-22* mutants. Basal centroid velocity was normalized by heat-shocked wild-type animals without hs:RFRP-1. WT: wild type. ***$p < 0.0001$.

(B) Overexpression of zebrafish RFRP-3 peptide does not suppress locomotion in *C. elegans*. Forward locomotion (positive centroid velocity) and reverse locomotion (negative centroid velocity) were recorded in animals that had recovered from heat shock. Overexpression of zebrafish RFRP-3 *C. elegans* had no effect on locomotion in wild-type animals or *npr-22* mutants. Basal centroid velocity was normalized by heat-shocked wild-type and *npr-22* mutant animals without hs:RFRP-3. WT: wild type. $p > 0.05$.

Figure S7. Zebrafish RFRP peptide precursor gene overexpression reduces waking behavior. (A) Schematic diagram of the heat shock-inducible transgene. (B) Schematic of zebrafish behavioral experiment. (C) Overexpression of RFRP peptide precursor gene reduces total time spent awake during the day, but not at night, compared to wild-type siblings. (D) RFRP overexpression reduces average locomotor activity during the day but not at night. Data is represented as mean ± SEM for 118 WT and 130 hs:RFRP larvae averaged over 3 independent experiments. ***$p<0.0001$. n.s. indicates no significant difference.

Supplemental Experimental Procedures

Strains

Wild-type worm strain was N2 (Bristol). Mutation strains obtained from the *Caenorhabditis* Genetics Center (CGC) include RB1990 *flp-7(ok2625) X*, RB1989 *flp-10(ok2624) IV,* RB1902 *flp-19(2460) X,* RB761 *npr-7(ok527) X,* and RB1405 *npr-22(ok1598) IV* were provided by the *C. elegans* Gene Knockout Project at OMRF (http://www.mutantfactory.ouhsc.edu/). VC1971 *flp-24(gk3109) III* and VC2502 *flp-28(gk1075)* X were provided by the *C. elegans* Reverse Genetics Core Facility at the

University of British Columbia, of *C. elegans* Gene KO Consortium (http://www.celeganskoconsortium.omrf.org). Mutation strains FX02427 *flp-13(tm2427) IV* and FX01583 *npr-3(tm1583) IV* were obtained from the National Bioresource Project (http://www.shigen.nig.ac.jp/c.elegans/mutants/).

Mutant strains:
TB528 *ceh-14(ch3) X*
RB1990 *flp-7(ok2625) X*
RB1989 *flp-10(ok2624) IV*
FX02427 *flp-13(tm2427) IV*
RB1902 *flp-19(ok2460) X*
VC1971 *flp-24(gk3109) III*
VC2502 *flp-28(gk1075)* X
FX01583 *npr-3(tm1583) IV*
RB761 *npr-7(ok527) X*
RB1405 *npr-22(ok1598) IV*
RB1938 *npr-32(ok2541)* IV

Mutant Alleles

Mutant allele information is available at WormBase (www.wormbase.org). Protein structure and domain annotation are available at Uniprot (www.uniprot.org) and EnsemblMetazoa (http://metazoa.ensembl.org/Caenorhabditis_elegans/).

flp-7(ok2625) is a 548-bp deletion that erases exons 2 and 3 completely and exon 4 partially, resulting in mutation of the gene that abolished the last four peptides synthesized by *flp-7*.

flp-10(ok2624) is a 682-bp deletion that erases the entire first exon and part of the second exon, resulting in a null mutation that completely abolishes all peptides synthesized from *flp-10*.

flp-13(tm2427) is a 382-bp deletion starting in the first intron and extending to the second intron that removes the second exon and causes a frameshift at amino acid residue 31. This mutation deletes the C-terminal of the propeptide, and abolishes all peptide cleavage sites and synthesis of all FLP-13 peptides.

flp-19(ok2460) is a 505-bp deletion that removes the first exon and part of the second exon, resulting in a null mutation that completely abolishes all peptides synthesized from *flp-19*.

flp-24(gk3109) is a 1180-bp deletion extending from within the first intron to the 3' intergenic sequence that removes the second exon and 3'-UTR of *flp-24*. This deletion truncates the protein at amino acid residue 29, removing a majority of the propeptide, all post-translational processing sites and the FLP-24 peptide.

flp-28(gk1075) is a 957-bp deletion that erases the first exon and part of the second exon, resulting in null mutation of the gene and completely abolishing all peptides synthesized by *flp-28*.

*npr-3(tm1583)*consists of a 1817-bp deletion and a 4-bp insertion at the deletion site. The deletion mutation removes exons 1 through 3 and part of exon 4, resulting in a null mutation.

npr-7(ok527) is a 1211-bp deletion that removes part of exon 3 and the entire exon 4. This mutation removes the last four transmembrane domains, causing truncation of the protein.

npr-22(ok1598) is a 2528-bp deletion that removes exons 2 to 5 of isoform a and exons 2 to 6 of isoform b, causing protein truncation before the first transmembrane domain and resulting in null mutations of both isoforms.

npr-32(ok2541) is a complex substitution mutation with about 1500-bp deletion that erases exons 5 through 8, removing the last four transmembrane domains and causing truncation of the protein.

Transgenic lines

Generation of heat shock transgenic lines

Heat-shock transgenic strains:

PS5009 *pha-1(e2123ts); him-5(e1490); syEx723[hsp16-41::lin-3C cDNA(10ng/uL) + myo2:GFP(10ng/uL) + pha-1(+)(pbx-1)(90ng/uL) + bluescript(90ng/uL)]*
PS6562 *syEx1285[hsp16-41::flp-13(10ng/ul), PmyosdsRed(10ng/ul), KS+(90ng/ul)]*
PS6563 *syEx1286[hsp16-41::flp-24(10ng/ul), PmyosdsRed(10ng/ul), KS+(90ng/ul)]*
PS 6571 *syEx1294[hsp16-41::flp-7(10ng/ul), PmyosdsRed(10ng/ul), KS+(90ng/ul)]*
PS6572 *syEx1295[Pflp-13::GFP(10ng/ul), PmyosdsRed(10ng/ul), KS+(90ng/ul)]*
PS6573 *syEx1296[Pflp-24::GFP(10ng/ul), PmyosdsRed(10ng/ul), KS+(90ng/ul)]*
PS6577 *syEx1300[hsp16-41::Dr*RFRP-1*(10ng/ul), PmyosdsRed(10ng/ul), KS+(90ng/ul)]*
PS6578 *syEx1301[hsp16-41::Dr*RFRP-3*(10ng/ul), PmyosdsRed(10ng/ul), KS+(90ng/ul)]*

Conditional expression of cDNAs was achieved by generating a fusion of the coding sequence of a gene under study to the *hsp-16.41* promoter (Stringham et al., 1992). A

synthetic DNA fragment consisting of the *hsp-16.41* promoter, DNA coding sequence, and its endogenous 3'-UTR was generated using fusion PCR (Hobert, 2002).

Mixed-stage populations of wild-type animals were harvested for RNA extraction and subsequently reverse-transcribed into cDNA (Schwarz et al., 2012) for amplification of *flp*-7 cDNA, flp-*13* cDNA, *flp-24* cDNA, and *hlh-13* cDNA. Their corresponding 3'-UTR regions were amplified from wild-type mixed stage animal genomic lysates. The *hsp-16.41* (Stringham et al., 1992) promoter region was amplified from plasmid ppD49.83 (Addgene) and the PCR product was verified by DNA sequencing. The cDNAs and 3'-UTR genomic DNAs amplified were sequenced and blasted against the predictions from www.wormbase.org. The final fusion PCR product was sequenced and blasted against the predicted DNA sequence.

hs:FLP-24PS was constructed by using a scrambled FLP-24 peptide sequence as a control for the neuropeptide function. hs:FLP-24 overexpression DNA was used as a backbone of the construct. The FLP-24 peptide (VPSAGDMMVRF) was replaced by FLP-24PS (SMFMGVAVPRGKRS), keeping the endogenous FLP-24 signal peptide, propeptide, cleavage sites, and 3'-UTR intact.

Primer used for generation of heat-shock transgenic lines.

DNA fragment	Forward primer 5' to 3'	Reverse primer 5' to 3'
P200-201 *Phsp-16.41*	ATGACCATGATTACGCCAAG	GCTAGCCAAGGGTCCTCCT
P206-P207 *hlh-13*	AGGAGGACCCTTGGCTAGCATGACAGCTTCATCTTCTGGGTGT	TAATCAGTATGTTTATTGAAATGAAAGATAGAAAATCATGAGTTGTATTCGTG

P249-P250 *flp-7*	AGGAGGACCCTTGGCTAGCATGCTTGGATCCCGCTTC	AACAGGCGTCGGTTCTTTATTT
P245-P247 *flp-13*	AGGAGGACCCTTGGCTAGCATGATGACGTCACTGCTCACT	TTATTTTCTGCCAAAACGAATG
P209-P212 *flp-24*	AGGAGGACCCTTGGCTAGCATGTTGTCGTCGCGCACATCGTCCATCAT	TCAGATGCTTCTTTTTCCAAATC
P279-P280 *Danio rerio* RFRP-1	CCAGCTCACCTGCATGCAAACCTCCCTCTTCGCTTTGGAAAAAGAAGCATCTGATAA	AGCGAAGAGGGAGGTTTGCATGCAGGTGAGCTGGACGTTTGTGTGGAATCTCTCC
P283-P282 *Danio rerio* RFRP-3	GTGCTGCACCAGCCTCAGCGGTTTGGAAAAAGAAGCATCTGATAATATACCATCTACC	GATGCTTCTTTTTCCAAACCGCTGAGGCTGGTGCAGCACACGTTTGTGTGGAATCTCTCC
P295-P296 hsFLP-24PS	AGCATGTTCATGGGGGTAGCTGTTCCACGTGGAAAAAGAAGCATCTGATAATATACCATCTACC	GCTTCTTTTTCCACGTGGAACAGCTACCCCCATGAACATGCTACGTTTGTGTGGAATCTCTC

Generation of reporter expression transgenic lines

Transgenic strains:
TB513 *dpy-20(e2017); chIs513[ceh-14::GFP, dpy-20(+)]*
PS6572 *syEx1295[Pflp-13::GFP(10ng/ul), PmyosdsRed(10ng/ul), KS+(90ng/ul)]*
PS6573 *syEx1296[Pflp-24::GFP(10ng/ul), PmyosdsRed(10ng/ul), KS+(90ng/ul)]*

Expression pattern of *ceh-14* is illuminated by the transgenic line TB513 available at CGC, while expression patterns for *flp-13* and *flp-24* were examined by generating a fusion of the conserved non-coding sequence in the vicinity of coding sequence to a basal promoter, *pes-10*, attached with GFP coding sequences and an *unc-54* 3'-UTR (Stringham et al., 1992; Hobert, 2002). MUSSA (multiple species sequence analysis) software was used to identify conserved non-coding sequences (Kuntz et al., 2008). Genomic sequences of *C. elegans*, *C. briggsae*, *C. remanei*, and *C. brenneri* were obtained from WormBase (www.wormbase.org) and compared for conservation at or higher than 67% sequence identity. All conserved sequences were amplified from wild-

type mixed stage worm genomic lysates. The *pes-10*, GFP and *unc-54* 3'-UTR sequences were amplified from pD97.78 (Kuntz et al., 2008).

Transgenic strains of *C. elegans* were generated by injecting fusion DNA constructs into the gonads of young adult wild-type hermaphrodites along with *Pmyo2::dsRed* as a coinjection marker causing fluorescence in the pharynx and pBluescript KS+ as carrier DNA (Van Buskirk and Sternberg, 2007). In order to control for variation between transgenes, at least two independent lines from each injection were used for expression patterns.

Primer sequences used in the generation of *promoter*::GFP transgenic lines.

DNA fragment	Forward primer 5' to 3'	Reverse primer 5' to 3'
P1-P65 *Ppes-10::*GFP	CTAGCAAAAATGCATAAGG	GTGTCAGAGGTTTTCACCGTCA
P157-P158 *Ppes10* promoter	CTAGCAAAAATGCATAAGGTTTTGCTG	TTTTTCTACCGGTACCTTACGCTTC
P253-P254 *Pflp-13*	CATCGTCGTAAAAACAAATTCAA	CCTTATGCATTTTTGCTAGTTTGACACAAAATGCCGACT
P242-P243 Pflp-*24*	CATCCAATATGGTGAGTTTCTCTG	CCTTATGCATTTTTGCTAGCGTCTGAAATTTCGAAAAGTAATAAT

Supplemental References

Brenner, S. (1974). The genetics of *Caenorhabditis elegans*. Genetics *77,* 71-94.

Halloran, M.C., Sato-Maeda, M., Warren, J.T., Su, F., Lele, Z., Krone, P.H., Kuwada, J.Y., Shoji, W. (2000). Laser-induced gene expression in specific cells of transgenic zebrafish. Development *127,* 1953-1960.

Hobert, O. (2002). PCR fusion-based approach to create reporter gene constructs for expression analysis in transgenic *C. elegans*. BioTechniques *32,* 728-730.

Kuntz, S.G., Schwarz, E.M., DeModena, J.A., De Buysscher, T., Trout, D, Shizuya, H., Sternberg, P.W., and Wold, B.J. (2008). Multigenome DNA sequence conservation identifies Hox cis-regulatory elements. Genome Res *18(12),* 1955-1968.

Li, C., Kim, K., Nelson, L.S. (1999). FMRFamide-related neuropeptide gene family in *Caeborahbditis elegans*. Brain Res. *848(1-2),* 26-34.

Mark, N.J., Maule, A.G., Geary, T.G., Thompson, D.P., Davis, J.P., Halton, D.W., Verhaert, P., and Shaw, C. (1997). APEASPFIRFamide, a novel FMRFamide-related decapeptide from Caenorhabditis elegans: structure and myoactivity. Biochem. Biophys. Res. Commun. *231(3),* 591-595.

Marks, N.J., Shaw, C., Halton, D.W., Thompson, D.P., Geary, T.G., Li, C., and Maule, A.G. (2001). Isolation and preliminary biological assessment of AADGAPLIRFamide and SVPGVLRFamide from Caenorhabditis elegans. Biochem. Biophys. Res. Commun. *286(5),* 1170-1176.

Raizen, D.M., Lee, R.Y., and Avery L. (1995). Interacting genes required for pharyngeal excitation by motor neuron MC in *Caenorhabditis elegans*. Genetics *141(4),* 1365-1382.

Schwarz, E.M., Kato,M., Sternberg, P.W. (2012). Functional transcriptomics of a migrating cell in *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. U.S.A. *109(40),* 16246-16251.

Stringham, E.G., Dixon, D.K., Jones, D., Candido, E.P. (1992). Temporal and spatial expression patterns of the small heat shock (hsp16) genes in transgenic *Caenorhabditis elegans*. Mol Biol Cell. *3(2),* 221-233.

Ubuka, T., McGuire, N.L., Calisi, R.M., Perfito, N., and Bentley, G.E. (2008). The control of reproductive physiology and behavior by gonadotropin-inhibitory hormone. Integr Comp Biol. *48,* 560-569.

Westerfiled, M. (1994). The zebrafish book. A guide for the laboratory use of zebrafish (*Danio rerio*). 4th Edition. (Eugene, University of Oregon Press).

Table S1. Protein-encoding genes in detected in ALA at late-L4 stage larvae. This lists the expression of 8,133 genes in poly(A)$^+$ RNA isolated from two pools of ALA neurons, compared to previously determined gene expression levels in poly(A)$^+$ RNA isolated from mixed-stage whole larvae[20]. The data columns are as follows. "Gene" gives the full gene identifier (WormBase name, sequence name, and CGC name) of a gene; gene names were taken from WormBase release WS220. "ALA" and "larvae" denote the gene expression values (measured in RPKM) for a given gene observed either in pooled ALA neurons (this study) or in wild-type larvae[27]. "ALA/larvae" gives the ratio of gene expression in pooled ALA neurons to gene expression in whole larvae, with an empirical pseudominimum for larval expression of 0.03 RPKM (the smallest non-zero RPKM value observed in the larval data set), used when no larval expression was actually observed (to avoid division by zero). "OMIM" denotes orthology to a human disease gene in the Online Mendelian Inheritance in Man database. Disease orthologies were computed by means of human-*C. elegans* orthologies in either WormBase release WS230 or in eggNOG 3.0's metazoan orthology groups (meNOGs]), along with human gene-disease links downloaded from Ensembl (via EnsMart, using "Homo sapiens genes" data set GRCh37.p6) on 4/28/2012). "Protein size(s)" lists the sizes of protein products. "Protein feature" lists predicted features such as signal, transmembrane, coiled-coil, or low-complexity sequences, predicted respectively by the programs SignalP, TMHMM, Ncoils, and SEG. "TF" indicates whether a gene's product was predicted to be a transcription factor by J. Thomas, the Walhout laboratory, or the Gupta laboratory. "Bork KOG" lists orthology annotations by eggNOG 2.0. "PFAM domain" lists any such protein domains annotated in WormBase WS220. "WBPhenotype" lists any RNAi or mutant phenotypes

annotated for a gene in WS220, with most phenotypes coming from mass RNAi screens. "NOT WBPheno" indicates that a gene was annotated as negative for such phenotypes in WS220. References for RNA-seq and annotation analyses are previously reported (Schwarz et al., 2012); references for phenotypic data are given in the WS220 release of WormBase.

Table S2. ALA expresses 23 FMRFamide-like neuropeptide-encoding genes. Six are ALA-enriched and have more than 17-fold higher expression in ALA neurons (shaded in yellow, as measured in reads per kilobase of exon model per million mapped reads, or RPKM) than their expression in whole larvae.

Gene	Expression in ALA (RPKM)	Normalized expression in whole larvae (RPKM)	Fold enrichment in ALA vs. whole larvae
flp-24	10176.34	1.59	6400.21
flp-7	4518.55	13.51	334.46
flp-13	2063.11	7.52	274.35
flp-6	8.35	0.16	51.5
flp-28	169.94	8.76	19.4
flp-10	74.42	4.25	17.51
flp-19	234.35	29.84	7.85
flp-5	258.56	40.43	6.4
flp-8	83.15	15.37	5.41
flp-22	14.77	3.62	4.08
flp-9	50.46	13.42	3.76
flp-32	0.41	0.12	3.42
flp-27	40.13	17.3	2.32
flp-25	12.65	5.53	2.29
flp-14	24.07	13.22	1.82
flp-16	51.3	32.38	1.58
flp-3	5.05	3.49	1.45
flp-4	1.51	2.35	0.64
flp-21	1.78	4.61	0.39
flp-2	4.28	12.02	0.36
flp-18	7.29	41.87	0.17
flp-15	0.1	1.26	0.08
flp-12	0.09	2.56	0.04

Table S3. Mutant strains tested for feeding quiescence at 20 minutes post heat-shock.

Strain	Number of animals	Fraction quiescent (Feeding)	Fold change normalized to wild type
Wild type	200	75%	N/A
ceh-14 (ch3)	50	3%	3.9
flp-5 (gk3123)	120	65%	1.4
flp-7 (ok2625)	50	54%	1.8
flp-13 (tm2427)	120	18%	3.3
flp-10 (ok2624)	50	64%	1.4
flp-19 (ok2460)	50	58%	1.7
flp-24 (gk3109)	100	42%	2.3
flp-28 (gk1075)	100	56%	1.8

Table S4. Shown are *C. elegans* FMRFamide peptides previously shown to be capable of activating calcium response via the G-protein coupled receptor NPR-22 *in vitro* (Mertens, et al., 2006) and are ALA-synthesized (Supplementary Table 2).

Gene	Peptide sequence	Activation of NPR-22
flp-7	SPMQRSSMVRF	+
	TPMQRSSMVRF	+
	SPMERSAMVRF	+
	SPMDRSKMVRF	+
flp-9	KPSFVRF	+
flp-13	AMDSPLIRF	-
	AADGAPLIRF	+
	APEASPLIRF	-
	ASPSAPLIRF	+
	SPSAVPLIRF	+
	ASSAPLIRF	+
	SAAAPLIRF	+
flp-22	SPSAKWMRF	+
flp-24	VPSAGDMMVRF	N/A

Table S5. Comparison of amino acid identity of *C. elegans* GPCRs, NPR-7 and NPR-22, with zebrafish (*Danio rerio*) and human GPCRs. Sequence alignment was performed with www.uniprot.org.

	NPR-7 (Q20067)	NPR-22a (Q9N324)	NPR-22b (Q59E83)
Human FF1 (Q9GZQ6)	23.1%	23.6%	24.6%
Human FF2 (Q9Y5X5)	18.4%	20%	20.2%
Zebrafish NPFFR1	20.4%	17.6%	16.9%
Zebrafish NPFFR-1l1 (F1RB60)	23.1%	21.2%	20.9%
Zebrafish NPFFR-1l2 (F1QP76)	20.9%	23.7%	23.7%
Zebrafish NPFFR-1l3 (F1QCV6)	21.7%	24.5%	25.1%
Zebrafish NPFFR-2.1 (A2AV71)	22.1%	23.8%	23.7%
Zebrafish NPFFR-2.2 (F1Q764)	21.3%	22.9%	22.3%

Video S1.

Shown is a 4-second video converted from a 240-second recording of young adult hs:FLP-13 animals 2 hours after heat-shock of animals, at 1 frame per minute (fpm) speed. hs:FLP-13 animals exhibit dramatically reduced locomotion with occasional body movements.

Video S2.

Shown is a 24-second video converted from a 240-second recording, at 1 frame per minute (fpm) speed, of young adult hs:FLP-24 animals 2 hours after heat-shock of animals. hs:FLP-24 animals ceased all body movement during the 240-second recording.

Chapter 6

The gene regulatory network of a *C. elegans* sleep-inducing neuron

Gene regulatory network of a *Caenorhabditis elegans* sleep-inducing neuron.

Elly Suk Chow[1,2], Erich M. Schwarz[3], Cheryl Van Buskirk[4], Paul W. Sternberg[1,2,5]

1. Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, U.S.A.
2. Howard Hughes Medical Institute
3. Department of Molecular Biology and Genetics, Biotechnology 351, Cornell University, Ithaca, NY 14853, U.S.A.
4. Department of Biology, California State University Northridge, 18111 Nordhoff Street, Northridge, CA 91330. U.S.A.
5. Corresponding author: pws@caltech.edu

Abstract

Specific characteristics and functionalities of an individual cell, the fundamental building blocks of the physiological appearance and behavior of the animals as a whole, are governed by the regulatory logics encoded in the genome. Despite the increasing knowledge in genome sequences, little is known about the genomic information and the organization and function of the regulatory apparatus that control animal behavior at the single-cell resolution. Here we used the *Caenorhabditis elegans* sleep-inducing neuron, ALA, as a model to decipher the genomic regulatory codes that convey a specific behavior. We identified 1,056 ALA-expressed genes by genome-wide search using a newly dissected ALA motif. This motif is dependent on the transcription factors *ceh-10* and *ceh-14*. In parallel, we performed RNA-seq on micro-dissected individual ALA neurons and observed 8,133 protein-coding genes in wild-type ALA, of which 57 are neuropeptide encoding genes. Majority of these neuropeptide genes are novel transcripts found in ALA and have vertebrate orthologs that were not previously described in sleep regulation. Comparing with mutant ALA transcriptomes and in conjunction with CEH-14 ChIP-seq data on modENCODE, we constructed a gene regulatory network to illustrate

the direct genomic interplays of sleep promotion associated genes. Our integrated analysis from multiple analyses provides an entry point to expanding our understanding of the pipelines of genomic information transaction and animal behavioral outputs.

Introduction

Combinatorial action of transcription factors on *cis*-regulatory modules (CRMs) or enhancers drives robust and precise patterns of gene expression, hence determining the signature and function of a given cell (Davidson 2010). Several lines of evidence from studies in metazoans suggest that genes with overlapped spatiotemporal expression patterns often share common regulatory codes embedded in CRMs or enhancers (Michelson 2002). CRMs are composed of one or more sets of motifs encoding one or more binding sites for transcription factors that read these codes. Interconnection of these genomic regulatory codes establishes a stable gene regulatory network (GRN) that illustrates the blueprints of developmental and physiological functions (Davidson 2010).

Terminal differentiation GRN guards the final process of transcriptional regulatory hierarchy in controlling differentiation gene battery that deploys specific cellular functions (Hobert 2008, Davidson 2010). The nematode *Caenorhabditis elegans* presents an elegant model for deciphering terminal differentiation GRN for its simple body plan, compact and well-described genome, the availability of multispecies sibling genomes for genomic comparison, easy amenability to genetic manipulation, and the availability of green fluorescent protein (*gfp*) reporter genes that reveals spatiotemporal expression of individual cells (Sulston and Horvitz 1977; Hobert 2008; Chalfie et al. 1994; Sudhaus

and Kiontke 1996, 2007; Baldwin et al. 1997; Hillier et al. 2005, Spencer et al. 2011). Although composed of only 302 neurons in a mature hermaphrodite, the *C. elegans* nervous system has a vast array of neuron classes (Hobert 2013, WormBook). Exploiting the abundant genomic information and resources, several studies in *C. elegans* identified cell type-specific motif (Wenick and Hobert, 2004), neuronal subtype-specific motif (Kratsios et al, 2012; Zhang et al, 2014) and neuron function-specific motif (Doitsidou et al, 2013) that terminal differentiation regulators, known as terminal selectors (Hobert 2008), recognize and activate, thereby driving terminal neuronal cell fates. These terminal selectors sample the output of the controlling specification GRN, and in turn determine transcription of terminal effector genes specifically in the given cell (Hobert 2008; Davidson 2010). Together these notions support the hypothesis that terminal differentiated neurons are co-regulated via a common strategy throughout the nervous system specification (Hobert 2008; Hobert et al, 2010; Michelson 2002), and sub-circuits of differentiation GRNs are utilized for genomic information transaction between the initial inputs to the terminal selectors and the activation of effector genes that generate cell specific outputs (Davidson 2010). An in-depth analysis on the functional differentiation GRN of post-mitotic neurons at the single-neuron resolution is yet to be discovered.

One challenge for functional genomics is the understanding in fine details of how encoded gene regulatory programs drive functionality of a particular cell. The traditional approach to decoding the genomic information is to use *in vitro* or *in vivo* methods to dissect individual CRM, a rather laborious and expensive approach (Michelson 2002,

Kuntz et al, 2008). A more efficient strategy is the bottom-up approach in which one integrates bioinformatics and experimental evaluation to first identify coexpressed genes and then analyze their CRMs that codes for *trans*-acting factors (Michelson 2002; Hobert 2008). This approach was elegantly exemplified by the identification of the *Drosophila* Dorsal target genes in the dorso-ventral patterning GRN (Markstein et al, 2002), and by defining the regulatory hierarchy of transcriptional controls in anteroposterior patterning of the *Drosophila* embryo (Berman et al, 2002). A more recent study in *C. elegans* took this approach beyond the single genome analysis to multispecies sibling genome analysis (Kuntz et al, 2008), based on the assumption that phylogenetically conserved sequences are more likely to be functional. Using four *Caenorhabditis* species, Kuntz et al (2008) identified the *cis*-regulatory architecture of a *Hox* cluster in *C. elegans*. Moreover, this approach is not restricted to searching for developmental CRMs but also for differentiation CRMs, as demonstrated by the identification of the aforementioned motifs in the *C. elegans* nervous system (Wenick and Hobert 2004; Kratsios et al, 2012; Doitsidou et al, 2013; Zhang et al, 2014), suggesting its global and reliable application in identifying functional CRMs.

Another barrier in understanding genomic information and regulatory logic is the availability of genomic information, especially at the single cell-resolution. *gfp* reporter gene technology developed in *C. elegans* (Chalfie et al, 1994) has been commonly used to identify expression of protein-coding genes in the animal (Dupuy et al, 2007; Hunt-Newbury et al. 2007; Murray et al, 2008). However, this approach relies heavily on the promoter region, which may not include all necessary regulatory elements, and thus may

not fully recapitulate endogenous expression pattern (Hunt-Newbury et al, 2007). Building on this approach, isolation of *gfp* labeled specific cell type and tissues by fluorescence-activated cell sorting (FACS) and RNA-seq successfully measured native transcripts in collection of a given cell type (Spencer et al, 2011). But the presence of unmarked cells isolated by FACS may reduce the accuracy of detected gene expression in a tissue type (Spencer et al, 2011). We previously adopted an alternative strategy to first isolate a *gfp*-labeled individual cell by micro-dissection and then perform RNA-seq to identify and quantify the genomic information in the migrating linker cell (Schwarz et al, 2012). Such an approach identified an extensive catalog of genes that are involved in cell migration (Schwarz et al, 2012). Here, we applied this strategy to uncover the genomic information of the mature ALA interneuron that plays important roles in mediating the sleep-like behavior in *C. elegans* (Van Buskirk and Sternberg 2007).

C. elegans exhibit a sleep-like state that is induced by activation of the epidermal growth factor receptor (EGFR) signaling pathway (Van Buskirk and Sternberg 2007). Transgenic reporter analysis and imaging showed that ALA expresses three transcription factors, including *ceh-10* (Wu et al, 2011) and *ceh-17* (Pujol et al, 2000) in the Paired-like homeodomain class, and *ceh-14* (Cassata et al, 2000) in the LIM homeodomain class families. Our previous genetic analysis revealed that these transcription factors are essential for the generation and differentiation of ALA in a combinatorial and temporal fashion (Van Buskirk and Sternberg 2010). Moreover, absence or reduction of these regulatory genes strikingly abolished expression of component genes in the EGFR pathway, as well as other known ALA-expressed differentiation genes. As a result, these

mutants lost their ability to respond to ALA-induced sleep effect (Van Buskirk and Sternberg 2010). But how might such genetic information convey into functional readout, i.e., what are the local genomic information and gene regulatory logics employed to elicit the sleep-inducing ability of the ALA neuron?

We first identified the terminal selectors of ALA as two temporal modules, CEH-14/CEH-10 and CEH-14/CEH-17, regulating early initiation and late maintenance state of the ALA-expressed genes, respectively. Next, using the published promoter regions of ALA coexpressed genes that are also dependent on these transcription factors, we performed multispecies genome computational comparisons and identified a collection of phylogenetically conserved intergenic sequences that harbors putative ALA-specific CRMs. Based on these CRMs, we generated an ALA motif that drives *gfp* expression exclusively in the ALA neuron. We validated the *trans*-inputs of the ALA motif to be an assembly of CEH-10 and CEH-14. This motif identified 1,056 ALA-expressed genes in genome-wide motif search analysis. Concurrently, we performed an unbiased single-cell RNA-seq to explore the transcriptome in wild-type (8,133 genes) and *ceh-14* loss-of-function mutant (7,078 genes) ALA. With these parallel approaches, we identified a catalog of genes that are CEH-14 dependent and ALA-expressed. In conjunction with published chromatin immuno-precipitation followed by sequencing (ChIP-seq) data of CEH-14 on modENCODE (http://www.modENCODE.org/), we further identified CEH-14 directly regulated genes in ALA. Based on our assimilated analysis of these distinct but related studies, we constructed a gene regulatory network that elucidates the mechanism of transcriptional regulation in a *C. elegans* sleep-like behavior.

Results

Combinatorial transcriptional regulation for ALA differentiation genes

We previously shown that cooperative inputs of one LIM (CEH-14) and two Paired-like HD class (CEH-10 and CEH-17) transcription factors govern differentiation of the ALA neuron (Van Buskirk and Sternberg 2010). Regulation of ALA differentiation genes relies on the combinatorial inputs from CEH-14 and CEH-10 during embryogenesis, and from CEH-14 and CEH-17 from late embryonic stage onward. Expression of CEH-14 and CEH-17 in ALA are interdependent and once activated, these two proteins are apparently locked-down in an autoregulatory loop to maintain expression throughout larval stages and adulthood, hence establishing a stable regulatory circuit for the ALA differentiation genes (Van Buskirk and Sternberg 2010). Both CEH-14 and CEH-17 are required for the expression of differentiation genes encoding *let-23* and *plc-3*, two members of the EGFR signaling pathway for mediating EGFR-induced sleep behavior in *C. elegans* (Van Buskirk and Sternberg 2007, 2010). These observations suggest that CEH-14 and CEH-17 are the ALA terminal selectors.

Genetic analysis identified CEH-14 as an activator of CEH-17 in the early larval stage; however, the source of its own initial transcriptional inputs remains elusive. We previously reported that reduction of CEH-10, as a result of reduction-of-function mutation, did not affect *gfp* reporter expression of CEH-14 or CEH-17. Given the necessity of CEH-10 in proper ALA development, here we revisited this issue by expressing a different *ceh-14::gfp* transgenic line in the *ceh-10(ct78)* reduction-of-function mutant. We observed dramatic and specific reduction of *ceh-14::gfp* expression

in ALA throughout larval stages (Fig. 1A-C). Using the PHB neurons in the tail and AFD neuron in the head as internal expression signal control, we quantified the relative fluorescence intensity (see Methods) in wild-type and *ceh-10(ct78)* mutant ALA neurons. We found that 48% of L1 *ceh-10(ct78)* animals displayed normal expression, while the remaining animals either had reduced (33%) or abolished (19%) *ceh-14::gfp* expression in ALA (Fig. 1D). Consistent with this view, we found that 52% of L4 *ceh-10(ct78)* animals had normal expression, while 48% had reduced *ceh-14::gfp* expression in ALA. Interestingly, we did not observe abolishment of *ceh-14::gfp* expression in L4 mutant animals, suggesting the presence of temporal-spatial regulatory modules in the *ceh-14* promoter region, and that activation of a temporal regulatory module at later stage is sufficient to initiate *ceh-14* transcription. Indeed, this transgenic line is driven by a 3.7 kb intergenic sequence of *ceh-14* (Cassata et al, 2000), which harbors two CEH-14 binding domains revealed by ChIP-seq analysis available on modENCODE (http://www.modencode.org/). We postulate that the 3.7kb *ceh-14* promoter is composed of two *ceh-14 cis*-regulatory modules lying in the CEH-14 ChIP-seq binding domains that take part in cell specification and cell differentiation: the early *cis*-regulatory module consists of CEH-10/CEH-14 combinatorial inputs that initiate expression of terminal selectors, while the later *cis*-regulatory modules consists of CEH-14/CEH-17 that maintains expression of terminal selectors and differentiation genes in ALA (Supplementary Fig. 1). The second CEH-14 binding domain is truncated in the 2 kb *ceh-14* promoter that we previously tested (Van Buskirk and Sternberg 2010). Based on these data and the evidence below (Fig 3A, G), we infer that the early CEH-10/CEH-14 module is removed in the 2 kb promoter and its activity is independent of CEH-10, hence driving

normal expression in the absence of CEH-10. Taken together, we showed that CEH-14 is the central modulator that regulates the ALA-expressed genes, and its combinatorial inputs with CEH-10 or CEH-17 form stable terminal selectors for the generation and differentiation of the ALA neuron (Fig. 1E).

Conserved cis-regulatory modules (CRM) of ALA-expressed genes

Terminal selectors directly regulate differentiation genes whose expression and function give identifying features of a neuron (Hobert 2008). Expression of differentiation genes is controlled via a coregulatory strategy through shared *cis*-regulatory motifs that typically contain an assembly of *trans*-acting factors known as terminal selector transcription factors (Davidson 2001; Hobert 2011). These *cis*-regulatory motifs are phylogenetically conserved. One example is the *C. elegans* interneuron AIY motif that executes cooperative regulatory inputs from TTX-3 (LIM-HD) and CEH-10 (Paired-like HD) transcription factors (Wenick and Hobert 2004). The peptidergic interneuron ALA (White et al, 1986) shares similar terminal selectors with the AIY neuron: a LIM-HD and Paired-like HD assembly. Yet these two neurons seem to function differently. Understanding the *cis*-regulatory modulation of ALA and AIY may shed light on how individual neurons in subgroups differentiate and maintain its unique functionality. To this end, we sought to identify the ALA cell-specific motif.

We first defined the conserved regulatory regions driving expression in ALA. We next searched for a functional *cis*-regulatory motif or *cis*-regulatory module (CRM) that is commonly shared in these conserved regions. We used a differentiation gene belonging

to the vascular endothelial growth factor receptor family, *ver-3,* as a starting point for its simple expression pattern. Reporter construct expression revealed that a 2.94kb promoter region driving *ver-3::gfp* is expressed exclusively in one neuron, the ALA neuron (Popovici et al, 2002). Outside of the nervous system, *ver-3* is found in the pharynx muscles and anus sphincter muscles (Popovici et al, 2002). The *C. elegans* genome is compact and given that most *cis*-regulatory motifs are located in close proximity (<1 kb to several kb) to the gene that it regulates (Hobert 2011), it is plausible that the roughly 3 kb *ver-3* promoter fragment harbors necessary and sufficient regulatory information for proper expression of *ver-3*. We identified similar-sized DNA fragments from the intergenic sequences of *ver-3* coding sequences from *C. remanei* and *C. briggsae*, assuming that the *cis*-regulatory architecture is conserved through homologous genes across species. Mussa (http://mussa.caltech.edu/mussa) is a bioinformatics tool that performs transitive pairwise comparison to obtain conservation between DNA sequences from multiple species and is, hence an excellent tool for identifying putative functional *cis*-regulatory modules (Kuntz et al, 2008). Using Mussa comparison, we found four conserved regions (CR) among *ver-3* promoter sequences from multiple species (Fig. 2A). Each CR ranges from 100-400 bp in length. We systematically analyzed their functionality in driving *ver-3* expression in ALA, using a *gfp* reporter as readout (Fig. 2B). This reporter is controlled by a basal promoter, *pes-10,* which does not drive *gfp* expression without the presence of a *cis*-regulatory module (Mello and Fire 1995; Kuntz et al, 2008). We found that CR4 (*ver-3*CR4::*gfp*), a conserved region proximal to the *ver-3* transcription start site, is sufficient to drive the specific reporter signal in ALA (Fig. 2C, D). The ALA neuron is the only cell that expresses *ver-3* CR4::*gfp* in the head. Similar to

the 3 kb promoter fragment (Popovici et al, 2002), we observed *ver-3*CR4::*gfp* expression in the anus sphincter muscles (not shown). Regardless, the success of identifying a functional and cell-specific CRM using computational approach suggests that a shared cell-specific *cis*-regulatory motif that is controlled by common *trans*-acting factors may be predicted based on sequence conservation.

We therefore used *ver-3* CR4 as reference to compare with the promoter regions of four known ALA-expressed genes (Fig. 2E), including the terminal selectors, *ceh-14* and *ceh-17,* a direct target of activated EGFR signaling, phospholipase C-γ encoded by *plc-3* (Van Buskirk and Sternberg, 2007, WormBase, http://www.wormbase.org/), and a terminal differentiation gene that encodes protein tyrosine phosphatase-like receptor, *ida-1* (Zahn et al, 2008). We observed high identity among these promoter sequences, and that a threshold >60% pairwise identity (in 30 bp window) is reliable to isolate functional CR. The *ceh-17* CR2 is a 238 bp fragment that is phylogenetically conserved (Supplementary Fig. 2) and is capable of driving *gfp* expression in ALA only (Fig. F-H), but not in the SIA neurons as controlled by the full promoter (Pujol et al, 2000). In all cases, *gfp* expression in ALA is stable throughout larval stages and adulthood (not shown), suggesting that these CRMs contain functional *cis*-regulatory motifs for ALA-specific expression.

ALA cell-specific cis-regulatory motif

We systematically examined functionality of all conversed regions from known ALA-expressed genes (Fig. 3A-D). Most genes are regulated via multiple active CRMs (Fig.

3A-C). In some cases, a CRM may drive broad expression that partially or fully recapitulates the gene expression controlled by reported promoters, suggesting the presence of multiple *cis*-regulatory motifs for multiple cell-specific terminal selectors. Indeed, transcription factors often have multiple binding sites in the *cis*-regulatory modules of target genes (Davidson 2001). It is possible that a transcription factor, in combination with a cofactor, may serve as terminal selectors for multiple cell type expression or temporal expression of a target gene. For the purpose of this study, we focused on CRMs that drive exclusive expression in ALA. We compared the ALA functional CRMs for *de novo* motif discovery using the MEME tool (Bailey and Elkan 1994). As a negative control, we used CRMs that do not drive *gfp* expression at all, or drive *gfp* expression elsewhere but not in ALA, to filter out irrelevant phylogenetically conserved sequences (Fig 3E). This approach allowed us to isolate a 13 bp AT-rich motif (Fig. 3G), hereafter we named the ALA motif. Further motif analysis using TOMTOM (Gupta et al, 2007) revealed that it has high similarity to the mouse Lhx3 transcription factor binding site, the mouse Vsx1 transcription factor binding site, and the mouse Phox2b transcription factor (Fig. 3G). *ceh-14* encodes the sole *C. elegans* orthologs of the vertebrate Lhx3 and Lhx4 (Cassata et al., 2000), *ceh-10* encodes a homeodomain protein close to the vertebrate Chx10 in mouse and Vsx1 in goldfish (Svendsen and McGhee 1995), and *ceh-17* encodes a *C. elegans* ortholog of the vertebrate Phox (Pujol et al, 2000).

In summary, we show that ALA-expressed genes, whether at the terminal selector level or at the terminal differentiation level, share a common regulatory motif composed of

putative binding sites for a LIM and a Paired homeodomain transcription factors. The LIM-HD binding site appears to occupy the entire motif, while the two Paired HD binding sites appear to occupy the 5' end of the motif. Genetic analysis showed that CEH-10 and CEH-17 interchange their role in regulating ALA development, with CEH-10 being the early initiating factor and CEH-17 being the maintenance factor in later stage (Van Buskirk and Sternberg 2010). The newly revealed ALA motif architecture seems to fit into this model, and we postulate that CEH-14 might cooperate with a temporal cofactor for the selection of CEH-10 or CEH-17 to coregulate ALA terminal differentiation genes at a given time.

CEH-10 and CEH-14 are the cooperative trans-acting factors of the ALA motif

To identify the upstream transcriptional inputs of the ALA motif, we analyzed CRMs that drive exclusive expression in ALA. We first performed a series of deletion mutation analysis to identify the core element in two functional ALA CRMs, *ver-3*CR4 and *ceh-17*CR2 (Fig. 4). We found that the 3' proximal 39 bp of *ver-3*CR4 (Fig. 4A-H), or a 45 bp within *ceh-17*CR2, is sufficient to drive ALA-specific expression (Fig. 4I-P). We then focused on the 39 bp *ver-3*CR4 core element, *ver-3*CR4.2 (Fig. 4E), to further characterize the necessity and sufficiency of this CRM in driving stable and specific *gfp* expression. Replacing 28 bp in the 5' region of *ver-3*CR4.2 with non-transcription factor-coding GFP sequence did not affect *ver-3*CR4.2::*gfp* expression in ALA (Fig. 5E-G), suggesting that the 3' proximal 11 bp are necessary for driving reporter expression. Next, we replaced these 11 bp with random GFP sequence and found no reporter expression in ALA (Fig. 5H-J). Moreover, these 11 bp lie in the region where the ALA motif was

predicted by the MEME tool (Fig.3F, Fig 5K). We therefore name *ver-3*CR4.2 as the *ver-3* ALA motif for simple reference. Our data demonstrates that functional *cis*-regulatory elements can be uncovered by combined efforts of bioinformatics prediction and in-depth motif characterization.

To verify the *trans*-acting inputs of the *ver-3* ALA motif, we expressed *ver-3* ALA motif::*gfp* in *ceh-10(ct78)* reduction-of-function mutants (Fig. 6A-C) and *ceh-14(ch3)* mutants (Fig. 6D-F). We found that the presence of either transcription factor is required for driving *ver-3* ALA motif reporter expression in ALA (Fig. 6) in L1 larvae. We found dim but detectable *ver-3* ALA motif expression in L4 *ceh-10(ct78)* mutant larvae (not shown). This observation is consistent with the abolished *ceh-14::gfp* reporter expression in *ceh-10(ct78)* mutants in L1 larvae (Fig. 1A-C). The weak detection of *ceh-14::gfp* and *ver*-3 ALA motif::*gfp* in *ceh-10(ct78)* reduction-of-function L4 larvae suggests that the absence of early expression was due to lack of transcriptional activation, but not the absence of the ALA neuron (Van Buskirk and Sternberg 2010). Our data provides further evidence that, in cooperation with the central *cis*-regulatory modulator CEH-14, CEH-10 is the early coregulator whose role in activating ALA terminal differentiation genes in later stages is substituted by other terminal selectors such as CEH-17.

Genome wide motif targeted search and unbiased single cell transcriptome profiling of ALA-expressed genes

We have defined the terminal selectors and identified a common ALA motif for ALA differentiation genes. Next, we wish to test the ability of the ALA motif to predict ALA-

expressed genes. Using the ALA motif as input, we performed a genome-wide search and looked for the presence of a similar motif in the non-coding regions of protein-coding genes. These regions include intronic sequences and intergenic sequences ranged up to 3 kb from the transcription start site or the 3'UTR. We detected 1,056 protein-coding genes in the genome that possess at least one ALA motif in the non-coding regions vicinity to their coding sequences (Fig. 7A, Supplementary Table 1). Among the motif-searched genes, two are in fact the ALA terminal selectors, *ceh-14* and *ceh-17*. Moreover, 285 (27%; Fig. 7B) of these genes are directly regulated by CEH-14 based on the ChIP-seq analysis data available on modENCODE (http://www.modencode.org/). We searched for genes that have at least one active CEH-14 binding site in the intergenic region ranged up to 5 kb both ways from their coding sequences. Other CEH-14 direct regulated genes include the gap junction (*unc-7*) and the neurotransmitter regulation (*unc-13*) encoding genes involved in the EGFR signaling-induced sleep pathway (Van Buskirk and Sternberg 2007). A FMRFamide-like neuropeptide, *flp-19*, is detected in our motif search (Supplementary Table 1) and ChIP-seq analysis on modENCODE. However, previous reporter gene expression analysis did not detect *flp-19* expression in ALA (Kim and Li 2004). We therefore verified *flp-19* expression in ALA by identifying the *flp-19* ALA motif and tested its capability to drive *gfp* expression in ALA (Supplementary Fig. 3). Expression of *flp-19* ALA motif::*gfp* is detected throughout larvae stages and adulthood. A common feature of terminal differentiation genes is their lack of repressors in the *cis*-regulatory modules (Davidson 2001). In fact, the lack of repressor binding site has been reported in numerous *C. elegans* motifs that participate in multiple steps throughout development of the nervous system. These include the cell specific AIY motif (Wenick

and Hobert 2004), the neuron subgroup specific cholinergic neuron motif (Kratsios et al, 2012), and the function specific dopamine pathway motif (Doitsidou et al, 2013). One may argue that in the absence of a repressor binding site, motif driven expression may not fully recapitulate endogenous expression of a gene. Alternatively, terminal differentiation gene expression could likely be a reflection of combined *cis*-regulatory modulation of their terminal selectors, whose *cis*-regulatory modules receive activation and repression inputs to define precise spatial and temporal expression (Wenick and Hobert 2004). Having been pre-regulated via their terminal selector *cis*-regulatory modules, it is therefore possible that the presence of repressors in a terminal differentiation gene motif is dispensable.

To verify the reliability of motif-searched genes at the cell-specific level, we turned to a cell-targeted approach to uncover transcriptomes in ALA. We performed single cell transcriptome profiling of the ALA neuron using procedures we reported previously (Schwarz et al, 2012). We detected transcriptome of 8,133 protein-coding genes in wild-type ALA (Fig. 7A) at late L4 larval stage (Supplementary Table 2). In conjunction with ChIP-seq data generated by modENCODE, we found that 2,242 (27.6%) of ALA-expressed genes are CEH-14 directly regulated (Fig. 7B). Further analysis revealed that both motif-search and profiling approach found a similar percentage of CEH-14 direct regulated genes in the genome (Fig. 7B). More interestingly, due to its targeted approach, we found that motif-search is more efficient in identifying CEH-14 direct regulated and ALA-expressed genes. Our results suggest that a cell-specific motif can serve as a quick

tool to sort out direct targets in the genome that are regulated by common terminal selectors.

Gene regulatory network of a sleep-inducing neuron

Our ultimate goal is to construct a gene regulatory network that provides a road map of molecular mechanistic interaction of genes that gives the feature of a cell. The ALA motif, in conjunction with ChIP-seq data on modENCODE, can successfully identify direct target genes of a terminal selector and cell-specific genes in the ALA neuron, yet ChIP-seq data did not reveal the regulatory nature of a *trans*-acting factor upon the *cis*-regulatory elements on the motif that it reads. Whether the regulatory output is activation or repression remains unknown. To break the gridlock, we profiled transcriptomes in late L4 *ceh-14(ch3)* mutant ALA and detected 7,078 protein-coding genes (Supplementary Table 2). We infer that about 1,055 (13%) of ALA-expressed genes are positively regulated by CEH-14 (Supplementary Table 3). And of those detected ALA-expressed genes, 3,171 (39%) are 10-fold less in the mutant ALA, suggesting that these genes are CEH-14-dependent. Incorporating the ChIP-seq data into our analysis, we found that 729 (23%) of ALA-expressed and CEH-14-dependent genes are directly regulated by CEH-14. As expected, a majority of them are differentiation genes such as neuropeptides. ALA expresses the EGF ligand receptor ortholog, LET-23/EGFR, and is the sole neuron that mediates EGFR-induced sleep-like behavior in *C. elegans* (Van Buskirk and Sternberg 2007). Previous studies suggest that neuropeptides are the downstream effectors to induce sleep upon activation of the EGFR signaling (Van Buskirk and Sternberg 2007; Chow et al, unpublished data). In addition, CEH-14 is required for the expression of LET-23 in

ALA, without which the EGFR-induced sleep-like behavior is completely abolished (Chow et al, unpublished data). But how exactly CEH-14 *cis*-regulation is translated into the sleep-inducing property of ALA is mysterious. Here, we illustrate the molecular mechanistic interactions that lead to the functional readouts of the ALA neuron in a gene regulatory network (Fig. 8). Similar to the characteristics of the hypothesized gene regulatory network of terminal selectors and their targeted terminal differentiation genes (Hobert 2008), this network (Fig. 8, Supplementary Table 1) possesses: (i) positive autoregulatory motif between the terminal selector genes, e.g. CEH-14 and CEH-17; (ii) single input modules indicating that the coregulation of terminal differentiation genes, e.g. *flp-19,* is coregulated by CEH-14, CEH-26, SKN-1, and ELT-3; (iii) feed-forward loop motifs in which a terminal selector gene regulates the expression of a transcription factor, which in turn controls downstream terminal differentiation genes, e.g. CEH-14 activates SKN-1 that activates *flp-13* and feeds back to *ceh-14*. In summary, we found that the gene regulation of ALA-expressed genes is a multi-step process involving multiple transcription factors and autoregulatory feedback loops. Moreover, a terminal selector gene can regulate a targeted terminal differentiation gene directly or indirectly via sub-transcriptional circuits in the network.

Discussion

We sought to uncover the molecules whose expression gives unique function and feature of a cell. We also aimed to decipher the mechanisms of transcriptional regulation and extend our knowledge in understanding the molecular underpinnings of cell development and function. Here, we integrated computational predictive, experimental validation, and

high-throughput explorative approaches to analyze the molecular information of a *C. elegans* sleep-inducing neuron, ALA. We identified the terminal selectors essential for transcriptional activation of the ALA terminal differentiation genes as assemblies of CEH-14/CEH-10 and CEH-14/CEH-17, in which CEH-10 and CEH-17 play pivotal roles in regulating early and late gene expression, respectively. Using the *cis*-regulatory modules of terminal differentiation genes, we identified a cell-specific ALA motif that harbors *cis*-regulatory modules recognized by the ALA terminal selectors. A genome wide search for genes with the ALA motif revealed ALA cell-specific and CEH-14 regulated genes in the genome. In addition, profiling the ALA transcriptome in wild type and a loss-of-function *ceh-14* mutant provides validation for the ALA motif cell-specificity and evidence for CEH-14 dependency, as well as the sign of the regulatory output, i.e., activation or repression of target genes. Moreover, in conjunction with ChIP-seq data generated by modENCODE, we further characterized CEH-14 direct regulation on ALA-expressed genes found by motif search or by transcriptome profiling. By incorporating findings from these different approaches, we generated an ALA gene regulatory network to illustrate the translation of *cis*-regulatory inputs into the signature cell functional readout. This network provides a road map for the molecular mechanistic interactions of ALA-expressed genes that ultimately give the sleep-inducing property of the ALA neuron. Lastly, findings in this study also yield insights into setting strategic and high throughput analysis of regulatory code and functional readout in a broad context.

Terminal selectors are transcription factors that read the shared *cis*-regulatory code of the nuts-and-bolts gene batteries that specify the structure and functional properties of a

mature neuron (Hobert 2008). These *trans*-acting factors initiate and maintain terminal differentiation programs often through direct regulation of corresponding terminal differentiation genes (Hobert 2011). Our previous analysis identified CEH-14 as the central regulator of ALA-expressed genes, and that partnering with CEH-17; it controls expression of their own as well as ALA differentiation genes (Van Buskirk and Sternberg 2010). We therefore proposed that CEH-14 and CEH-17 are terminal selectors for the ALA neuron. However, the requirement of CEH-17 regulation is concentrated on later larval stages, leaving the early input unknown. The transcription factor *ceh-10* belongs to the same homeodomain class of *ceh-17* and is critical for ALA neuron generation during embryogenesis and for functional maintenance in adulthood (Van Buskirk and Sternberg 2010). Moreover, the ALA neuron failed to elicit sleep response to activated EGFR signaling in *ceh-10(ct78)* reduction-of-function mutants, but expressed normal pan-neuronal genes such as *unc-119* and *rab-3* (Van Buskirk and Sternberg 2010). These observations fit into the characteristics of a terminal selector and lead us to believe that CEH-10, partnering with CEH-14, is the early input to initiate the ALA gene battery. Indeed, we found that CEH-10 is required for the ALA cell-specific expression of *ceh-14* in early larvae. Further analysis of an ALA motif suggests that CEH-10 also likely activate ALA terminal differentiation genes (Fig 6A-C). Together these findings indicate that the ALA terminal differentiation genes contain a modular terminal differentiation motif controlled by temporal regulatory inputs throughout the lifetime of the neuron.

Combinatorial regulation of CEH-14/CEH-10 in behavioral output is not restricted in the ALA neuron or in *C. elegans*. The CEH-14 mouse ortholog, Lhx3 (Cassata et al., 2000),

and the CEH-10 mouse ortholog, Chx10 (Svendsen and McGhee 1995), are coexpressed in the medullary reticular formation neurons that mediate motor behavior and are involved in locomotion (Bretzner and Brownstone 2013). In addition, the brainstem reticular formation known as the ascending reticular activating system plays important roles in regulating sleep/wake cycle (Haas and Lin 2012). Activation of EGFR signaling in *C. elegans* induced a sleep-like behavior where wild-type animals ceased locomotion and feeding (Van Buskirk and Sternberg 2007), along with dampened sleep neuronal circuit (Cho and Sternberg 2014) and hence increased arousal to sensory stimulation (Chow et al, unpublished data). Mutants of both *C. elegans* CEH-14 and CEH-10 are resistant to the EGFR sleep inducing effect due to their necessity in activating EGFR signaling pathway member genes (Van Buskirk and Sternberg, 2010). In rats, activation of EGFR signaling inhibits locomotion activity (Kramer et al, 2001). Whether Lhx3 and Chx10 are required to mediate EGFR induced locomotor effect, or whether EGFR signaling is involved in the medullar reticular formation neuron mediated locomotor activity, remains to be tested. Nonetheless, these observations suggest that the regulatory code of CEH-14/Lhx3 and CEH-10/Chx10 in locomotion could, at least in part, be evolutionarily conserved in metazoans. The lack of anatomical resemblance of the *C. elegans* and mammalian nervous systems has been a challenge for correlating data from model organism behavioral outputs and molecular controls to human, deciphering conserved sleep associated regulatory codes may connect the dots and reveal novel players in sleep regulation in mammals.

The composition of the ALA motif and the cooperative manner it applies to activate target genes is reminiscent of the AIY motif that confers specific properties to the AIY interneuron. In both cases, a combination of LIM and Paired-like HD class transcription factors are recruited to regulate and maintain functionality of a mature interneuron. The ALA motif contains *cis*-regulatory binding sites for multiple transcription factors whose combined inputs give temporal regulatory modulation. Moreover, these two assemblies appear to play distinguished roles during development, in which CEH-14/CEH-10 initiate early expression and CEH-14/CEH-17 maintain the established regulatory state for differentiation genes. The AIY (CEH-10/TTX-3) motif, on the other hand, responds to two transcription factors whose cooperative inputs regulate AIY-expressed genes (Wenick and Hobert 2004). Though both are interneurons and recruit CEH-10 as a terminal selector, these neurons display distinct features and functions. Such discrepancy is likely determined by cellular context with the presence of a transcription factor that it interacts (Hobert 2008). Here, the LIM-HD transcription factors *ceh-14* and *ttx-3* are present in ALA and AIY, respectively. Their precise temporal and spatial expression in ALA or AIY is a result of an orchestrated multiple outputs and multiple inputs in the local genomic information during development, hence establishing different terminal differentiation gene networks and readouts between neurons.

Transcriptome profiling and functional genomic analysis from *C. elegans* single cells or pools of mammalian cells have yield valuable insights into identifying genes specifically expressed in a domain or a cell type at a given time, or in response to cellular changes through development or upon treatments (Shapiro el al, 2011, Wu et al, 2011, Schwarz et

al, 2012). We previously profiled the transcriptome of the *C. elegans* linker cell and found highly dynamic gene regulation over the course of larval development (Schwarz et al, 2012). The linker cell is attached to the tip of the male gonad, whose migration leads and shapes the mature gonad (Schwarz et al, 2012). As it migrates and develops, the linker cell undergoes dynamic cellular context changes, as reflected by upregulation or down-regulation of more than half of the genes expressed in the cell between larval stages (Schwarz et al, 2012). With this note in mind, one would expect to find more transcripts in the developing linker cell than a mature cell such as the ALA neuron. Here we captured the dynamics of gene expression in ALA using similar experimental and analysis tools. Surprisingly, we found similar number of protein-coding genes in the linker cell (8,011 genes) and in ALA (8,133 genes), which is about 40% of protein-coding genes detected in the *C. elegans* genome. Other RNA-seq analysis using GFP-labeled embryonic cells isolated by FACS revealed that about 22.5% to 35.5% (4,572 genes to 7,199 genes) of *C. elegans* protein-coding genes are cell type specific (Spencer et al, 2011). We therefore conclude that dissected single cell RNA-seq allows deep transcriptome profiling analysis at the single-cell resolution and, that it is efficient in detecting cell-specific genes.

Of the protein-coding genes detected in ALA, 57 are neuropeptide-coding genes (Supplementary Table 1). Given that there is a total of 113 neuropeptide-coding genes in *C. elegans* (Li and Kim 2008, WormBook), nearly half of all *C. elegans* neuropeptide-coding genes are expressed in ALA. This observation suggests that communication between ALA and other neurons relies on discharged neuropeptides. This is consistent

with the morphology and functional analysis of the ALA neuron. The synaptic connectivity diagram showed that ALA has just a few synaptic connections and only one pair of axons originating from the neuron in the head and extending along lateral nerve bundles to the tail (White et al, 1986). The axons are dispensable for mediating ALA-mediated sleep behavior (Van Buskirk and Sternberg 2007). By contrast, genes encoding synaptic vesicle release protein or vesicle docking protein is required to mediate sleep induction (Van Buskirk and Sternberg 2007), suggesting that neuropeptides are involved in ALA-mediated sleep. In fact, the most strongly expressed gene in ALA is a FMRFamide-like neuropeptide gene, *flp-24*, whose expression is also CEH-14 dependent (Supplementary Table 1). Two other ALA-highly expressed and CEH-14 dependent neuropeptide genes are *flp-7* and *flp-13*. Previous *promoter*::*gfp* construct analysis did not report expression of *flp-13* or flp-*24* in ALA (Li and Kim 2004). Strikingly, mutations of these neuropeptide genes severely affect ALA's capability to mediate sleep behavior (Chow et al, unpublished data). With these observations, we suggest that transcriptome profiling can detect novel genes, and that quantification of gene expression accomplished with experimental verification can be useful for predicting cell function.

Our previous analysis established a connection between transcriptional inputs and functional outputs of the ALA neuron (Fig. 1E, Van Buskirk and Sternberg 2010). We now provide more fine details of how genomic controls are organized and translated into behavioral outputs in the form of a gene regulatory network of terminal selectors and neuropeptides expressed in ALA (Fig. 8). This network illustrates the blueprint of sleep regulation within a single, precisely timed neuron, and provides an entry point for further

functional analysis and prediction of genomic interplays in the organism, and may yield further insights into the mechanism of sleep regulation in metazoans.

Materials and Methods

Strains

Strains were grown, unless indicated otherwise, on nematode growth medium (NGM) 2% agar Petri plate seeded with *E. coli* strain OP50 and maintained at 20°C under standard conditions (Brenner, 1974).

Wild-type worm strain was N2 (Bristol). Mutant strains used in this study include, BW506 *ceh-10ct78)* and TB528 *ceh-14ch3)*, are obtained from the *Caenorhabditis* Genetic Center.

Single ALA neuron dissection and transcriptome profiling

Individual wild-type larvae of strain TB513 at the mid-L4 larval stage were hand picked and glued on an agar pad for microdissection as previously described (Schwarz et al., 2012). GFP-tagged ALA neurons were individually collected with an unpolished patch-clamp tube that served as a pipette, transferred to a prelubricated microcentrifuge tube, and snap-frozen with liquid nitrogen. Frozen tubes containing individual ALA neurons were kept at -70°C until their RNA was amplified.

RT-PCR, RNA-seq, and computational analysis of individual neurons were done essentially as in Schwarz et al., 2012. To obtain RNA-seq data by Illumina sequencing, aliquots of RT-PCR from individual cells were collected into two pools (4 cells and 5

cells). All RNA-seq reads were single-end, and originally 50 nt in length. Raw reads were quality-filtered as in Schwarz et al., 2012. They were then truncated *in silico* from 50 nt to 38 nt, the read length for previously published control data from mixed whole larvae Schwarz et al., 2012. This truncation allowed the ALA reads to be mapped and quantitated using exactly the same pipeline that had been used for larval data, and thus allowed more exact comparisons between ALA and larvae. After quality filtering and truncation but before mapping, RNA-seq data from the two pools of wild-type ALA comprised 1,164,892,280 nt in 30,655,060 reads and 1,520,526,262 nt in 40,013,849 reads. Of these, 25.2% could be mapped to WS190 protein-coding gene models (i.e., 17,798,207 out of 70,668,909 reads). We used existing whole wild-type larval RNA-seq data (Schwarz et al., 2012) as controls for housekeeping versus ALA-enriched genes. Expression values for genes were computed as in Schwarz et al., 2012. They were defined by pooling reads from both wild-type mid-L4 ALA neuron sets into a single set of expression values, doing likewise for both wild-type larval RNA-seq sets from Schwarz et al., 2012, and computing ALA/larval ratios of gene activity. We detected expression of 7,698 and 4,068 genes in the two ALA pools separately, and 8,133 genes collectively.

Reporter genes and expression analysis

Transgenic strains of *C. elegans* were generated by injecting fusion DNA constructs into the gonads of young adult wild-type hermaphrodites along with *Pmyo2::dsRed* or *Pmyo3::mCherry* as a coinjection marker, causing fluorescence in the pharynx and pBluescript KS+ as carrier DNA (Van Buskirk and Sternberg, 2007). In

order to control for variation between transgenes, at least two independent lines from each injection were used for expression patterns. Quantification of the intensity of fluorescence in ALA was done using Openlab software (Improvision) as follows. Fluorescence images of ALA in wild-type and mutant were taken under identical exposure. The mean pixel intensity (m.p.i.) over the brightest spot of the cell was measured, and compared with the m.p.i. of the control neuron in the same animal, expressing the same *promoter::gfp* to calculate the relative fluorescence intensity (r.f.i.). For each phenotype, approximately 20 animals were scored.

Primer used for generation of reporter gene expression transgenic lines.

DNA fragment	Forward primer 5' to 3'	Reverse primer 5' to 3'
P65-P1 *Ppes-10*::GFP	CTAGCAAAAATGCATAAGG	GTGTCAGAGGTTTTCACCGTCA
P85-P86 mCherry	CTGTAATTTTTAACTTTCAGATGGTCTCAAAGGGTGAAGAA	CTACTTATACAATTCATCCATGCCACCT
P157-P158 *Ppes10* promoter	CTAGCAAAAATGCATAAGG	TTTTTCTACCGGTACCTTACGCTTC
P89-P90 *Pver-3*::mCherry	TGTTTTCAAAGTGTTGGAATCAAT	AAAAATCGATCCTGCAGGcGAACCGAACCGAATGAAACA
P4-P6 *ceh-14* CR2	GAAAAAGAAACTCACACAGAAATTG	CCTTATGCATTTTTGCTAGGAAAAAGTGCAAAATTTGAAAAA
P9-10 *ceh-14* CR3	CCATTCATGAAAAGTTCAATAGTTT	CCTTATGCATTTTTGCTAGGAAAAAGTGCAAAATTTGAAAAA
P20-P24 *ceh-14* CR1	CGATGTCCCAAATTTCATACC	CCTATGCATTTTTGCTAGCCCGTTTAATTTTGTTTTGC
P47-P48 *ceh-17* CR1	GTCCTGCTACGGGCAAAATA	CCTTATGCATTTTTGCTAGTTTCCAAAAGCGAACTTCTGT
P44-P45 *ceh-17* CR2	TTCCGGTGCCGTTATAAAAT	CCTTATGCATTTTTGCTAGTCCTCTATGCTCTATGCTCAAATTC
P59-P60 *ida-1*	CCGAACGGAAACTTTAATTTTGT	CCTTATGCATTTTTGCTAGTTTTTCCCAAAACTTTTTCAAAC
P62-P63 *plc-3*	AATTTTTCTGTGACTTTGTTGCAT	CCTTATGCATTTTTGCTAGTGAAAAACGAAAGCTGTGATACAT
P89-P90 *ver-3* CR4	TGTTTTCAAAGTGTTGGAATCAAT	CCTTATGCATTTTTGCTAGTTCATTCGGTTCGGTTCTGT
P89-P91 *ver-3*CR4.1	TGTTTTCAAAGTGTTGGAATCAAT	CCTTATGCATTTTTGCTAGACGAGAAGAATTTAGTAGCGTCCGTA
P102-P1 *ver-3* CR4.2	GTCTATTTAGAAAATTAAAATCTATTCTAAATTAATTTACTAGCAAAAATGCATAAGG	GTGTCAGAGGTTTTCACCGTCA
P101-P44 *ceh-17* CR2.1	GATGAATTCAGATAAATTTCAAACCGAATCA	CCTTATGCATTTTTGCTAGTCCTCTATGCTCTATGCTCAAATTC
P100-P44 *ceh-17* CR2.2	AAAATTGCACACGAGCAATATATTCTGTTTTAATG	CCTTATGCATTTTTGCTAGTCCTCTATGCTCTATGCTCAAATTC

Figure 1. Combinatorial transcriptional input is required for ALA differentiation genes. (A-F) Expression of *ceh-14* is dependent on CEH-10. The ALA neuron (arrow) is present in wild-type animals (A) and *ceh-10(ct78)* reduction of function mutants (B,C). Merged DIC and fluorescent images show that *ceh-14::gfp* transgenic expression is found in ALA in the head (arrow) and a cluster of neurons in the tail (asterisk) in wild-type animals (A) but greatly reduced (B) or completely abolished in *ceh-10* mutants (C). (D) *ceh-10* reduction of function reduced average *ceh-14::gfp* intensity in the ALA neuron in L1 (n=21) and L4 larvae (n=21). ***$p<0.001$. Reduction of *gfp* in L1 and L4 are comparable. n. s.: not significant. Student's *t*-test with unequal variance. Error bars, mean s.e.m. Anterior is left. Dorsal is up. Scale bar represents 20μm. (E) A model for the gene regulatory circuit that regulates ALA differentiation genes.

Figure 2. Conserved *cis*-regulatory module (CRM) of ALA-expressed genes. (A) Comparison of the non-coding intergenic sequences 5' to the *ver-3* protein-coding sequences among multiple *Caenorhabditis* species revealed evolutionarily conserved regions (CR) with >70% pair wise identity in every 30 bp window. (B-D) A conserved region of 151 bp (*ver3CR4*) proximal to the transcription start site (black arrow) is sufficient to drive exclusive transgenic GFP expression in the ALA neuron (white arrow, C-D). Red boxes represent functional CRM. Grey boxes represent non-functional CRM. (E) Comparison of the *ver-3* ALA cell-specific CRM with previously reported promoter regions revealed functional CRM of other known ALA-expressed genes. (F-H) A 238 bp (*ceh-17CR2)* in the published *ceh-17*::*gfp* transgenic promoter is sufficient to drive exclusive expression in ALA (white arrow). The pharynx is illuminated by *myo-2::dsRed,*

which was co-injected to serve as a selection marker for transgenic lines. Anterior is left. Dorsal is up. Scale bar represents 20μm.

Figure 3. ALA cell-specific *cis*-regulatory modules. (A-D) Functional analysis of conserved *cis*-regulatory modules (CRM) of four ALA-expressed genes, *ceh-14, ceh-17, plc-3,* and *ida-1*. Red blocks represent functional CRMs that drive specific reporter expression in ALA and grey blocks represent non-functional CRMs fail to drive expression in ALA. (E-G) Identification of an ALA-motif. (E-F) Functional ALA CRMs of ALA-expressed genes were filtered with non-functional ALA CRMs before used to search for similar CRMs in the genome. GFP activity = reporter gene expression. +: presence of reporter gene expression. -: absence of reporter gene expression. (G) An AT-rich ALA-motif. This motif includes binding sites for homeodomain class transcription factors. TOMTOM search revealed high similarity of the ALA-motif with Lhx3 (MA0135), a vertebrate ortholog of CEH-14; Vsx1 (UP00141_1), a vertebrate ortholog of CEH-10; Phox2b (UP00149_1), a vertebrate ortholog of CEH-17.

Figure 4. The core elements of ALA cell-specific CRMs. Deletion mutation analysis identified minimal functional sequences in *ver-3CR4* and *ceh-17CR2*. Transgenic GFP expression constructs are shown as in schematic diagrams (A and E for *ver-3CR4*; I and M for *ceh-17CR2*). (A-H) A 39 bp (*ver-3CR4.2*) of *ver-3CR4* proximal to the transcription start site of *ver-3* are sufficient to drive exclusive GFP expression in ALA (arrow), while the distal 100 bp (*ver-3CR4.1*) are dispensable. (I-P) A minimal of 110 bp (*ceh-17CR2.*1) on the 3'end of *ceh-17CR2* is required to drive ALA expression, and

further deletion of 45 bp (blue box) completely abolished GFP expression in ALA (P). A body wall marker *myo-3::mCherry* construct (D) or a pharynx marker *myo-2::dsRed* construct (H, L, P) was introduced for selection of transgenic lines. Anterior is left. Dorsal is up. Scale bar represents 20μm.

Figure 5. An AT-rich and ALA-specific motif. Modular characterization of *ver-3CR4.2* revealed an AT-rich and cell-specific *ver-3* ALA-motif. (A) Schematic diagram of expression reporter constructs for *ver-3CR4* and the minimal element *ver-3CR4.2*. (B-D). The 39 bp minimal CRM *ver-3CR4.2::gfp* (green) recapitulates specific reporter expression in ALA driven by the full CRM *ver-3CR4::mCherry* (red). (E-J) A core AT-rich 11 bp element of *ver-3CR4.2* is sufficient and necessary for driving reporter expression in ALA (arrow). Replacing the first 28 bp of *ver-3CR4.2* with random non-transcription factor binding sequence had no effect on reporter expression pattern (E-G). Replacing the AT-rich 11 bp element completely abolished *ver-3CR4.2::gfp* reporter expression (green) in ALA, while the full CRM (*ver-3CR4*, red) expressed normally (H-J). (K) *ver-3* CR4 compared with ALA motif. Images from *ver-3CR4::mCherry* and *ver-3CR4.2::gfp* were superimposed for comparison of expression patterns. A pharynx marker *myo-2::dsRed* construct was introduced for selection of transgenic lines. Anterior is left. Dorsal is up. Scale bar represents 20μm.

Figure 6. The ALA-motif is dependent on CEH-10 and CEH-14. DIC images indicated the presence of ALA neuron (arrows) in *ceh-10(ct78)* reduction of function mutant (A) and *ceh-14(ch3)* null mutant (D). Expression of *ver-3 ALA-motif* is abolished in ALA in

the reduction of CEH-10 (B) or absence of CEH-14 (E). Expression of the transgenic line selection marker *myo-2::dsRed* in the pharynx remains intact (C, F). Anterior is left. Dorsal is up. Scale bar represents 20μm.

Figure 7. Comparison of ALA-motif-searched genes, ALA transcriptome profiling and CEH-14 ChIP-Seq analysis from modENCODE.

Figure 8. The gene regulatory network of ALA-expressed genes in the EGF-induced sleep pathway. CEH-14 regulates sleep-inducing genes in the ALA both directly and indirectly through activation of downstream transcription factors. Arrows represent activation. Bars represent repression. Direct interaction was based on ChIP-Seq data available on modENCODE. All inputs originating from CEH-14 were validated with ALA transcriptome profiling in wild-type and *ceh-14(ch3)* null mutant. Inputs originating from *ceh-17, gei-11, ceh-26, skn-1,* and *elt-3* were inferred from their interaction with CEH-14 and ALA transcriptome profiling analysis.

Supplementary Fig. 1

Schematic diagram of *ceh-14::gfp* promoters. The 3.7 kb promoter contains the early module comprised of CEH-10/CEH-14 and the late module CEH-14/CEH-17 is capable of driving *gfp* expression in the ALA from early embryogenesis till adulthood. The 2 kb promoter has the truncated early module comprised of CEH-10/CEH-14 and drives *gfp* expression in ALA mainly in later larval stages and in adulthood.

Supplementary Fig. 2

The promoter sequence of *ceh-17* is phylogenetically conserved. Comparison of the non-coding intergenic sequences 5' to the *ceh-17* protein-coding sequences among *C. elegans, C. remanei,* and *C. briggsae*. Conserved regions (red box) have more than 67% pairwise identity in every 30 bp window.

Supplementary Fig. 3

(A) Schematic diagram showing a 227 bp intergenic DNA fragment contains two *flp-19* ALA motifs and the *gfp* expression construct. (B-E) *flp-19* ALA motif::*gfp* is capable of driving *gfp* expression in ALA (arrows) in L2 (B-C) and in L4 (D-E) stages. A pharynx marker *myo-2::dsRed* construct was introduced for selection of transgenic lines. Anterior is left. Dorsal is up. Scale bar represents 20μm.

Supplementary Table 1

Genome wide ALA motif searched genes.

Supplementary Table 2

Transcriptome profiling genes in wild type and *ceh-14* loss-of-function mutant ALA.

Supplementary Table 3

ALA-expressed and CEH-14 dependent genes.

References

Bailey TL, Elkan C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* **2:** 28-36.

Baldwin JG, Frisse LM, Vida JT, Eddleman CD, Thomas WK. 1997. An evolutionary framework for the study of developmental evolution in a set of nematodes related to *Caenorhabditis elegans*. *Mol Phylogenet Evol* **8:** 249-259.

Berman BP, Nibu Y, Pfeiffer BD, Tomancak P, Celniker SE, Levine M, Rubin GM, Eisen MB. 2002. Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the Drosophila genome. *Proc Natl Acad Sci U.S.A.* **99(2):** 757-762.

Brenner, S. (1974). The genetics of *Caenorhabditis elegans*. Genetics **77:** 71-94.

Bretzner F, Brownstone RM. 2013. Lhx-Chx10 reticulospinal neurons in locomotor circuits. *J Neurosci* **33(37):** 14681-14692.

Cassata G, Kagoshima H, Andachi Y, Kohara Y, Durrenberger MB, Hall DH, Burglin TR. 2000. The LIM homeobox gene *ceh-14* confers thermosensory function to the AFD neurons in *Caenorhabditis elegans*. *Neuron* **25:** 587-597.

Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC. 1994. Green fluorescent protein as a marker for gene expression. *Science*. **263:** 802-805.

Cho JY, Sternberg PW. 2014. Multilevel modulation of a sensory motor circuit during *C. elegans* sleep and arousal. *Cell* **156(1-2):** 249-260.

Davidson EH. 2001. *Genomic Regulatory Systems*. Academic Press, San Diego, CA.

Davidson EH. 2010. Emerging properties of animal gene regulatory networks. *Nature* **468(7326):** 911–920.

Doitsidou M, Flames N, Topalidou I, Abe N, Felton T, Remesal L, Popovitchenko T, Mann R, Chalfie M, Hobert O. 2013. A combinatorial regulatory signature controls terminal differentiation of the dopaminergic nervous system in *C. elegans*. *Genes Dev* **27(12):** 1391-1405.

Dupuy D, Bertin N, Hidalgo CA, Venkatesan K, Tu D, Lee D, Rosenberg J, Svrzikapa N, Blanc A, Carnec A, Carvunis AR, Pulak R, Shingles J, Reece-Hoyes J, Hunt-Newbury R, Viveiros R, Mohler WA, Tasan M, Roth FP, Le Peuch C, Hope IA, Johnsen R, Moerman DG, Barabási AL, Baillie D, Vidal M. 2007. Genome-scale analysis of in vivo spatiotemporal promoter activity in Caenorhabditis elegans. *Nat Biotechnol* **25(6):** 663-668.

Gupta S, Stamatoyannopolous JA, Bailey T, Noble WS. 2007. Quantifying similarity between motifs. *Genome Biol* **8(2):** R24.

Hillier LW, Coulson A, Murray JI. Bao Z, Sulston JE, Waterston RH. 2005. Genomics in *C. elegans*: So many genes, such a little worm. *Genome Res.* **15:** 1651-1660.

Hobert O. 2008. Regulatory logic of neuronal diversity: Terminal selector genes and selector motifs. *Proc Natl Acad Sci U.S.A.* **105:** 20067-20071.

Hobert O. 2011. Regulation of the terminal differentiation programs in the nervous system. *Annu Rev Cell Dev Biol* **27:** 681-696.

Hobert O. 2013. The neuronal genome of *Caenorhabditis elegans*. WormBook.

Hunt-Newbury R, Viveiros R, Johnsen R, Mah A, Anastas D, Fang L, Halfnight E, Lee D, Lin J, Lorch A, McKay S, Okada HM, Pan J, Schulz AK, Tu D, Wong K, Zhao Z, Alexeyenko A, Burglin T, Sonnhammer E, Schnabel R, Jones SJ, Marra MA, Baillie DL, Moerman DG. 2007. High-throughput in vivo analysis of gene expression in Caenorhabditis elegans. *Plos Biol* **5(9):** e237.

Kim K, Li C. 2004. Expression and regulation of an FMRFamide-related neuropeptide gene family in *Caenorhabditis elegans*. *J Comp Neurol* **475(4):** 540-550.

Kramer A, Yang FC, Snodgrass P, Li X, scammell TE, Davis FC, Weitz CJ. 2001. Regulation of daily locomotor acitivity and sleep by hypothalamic EGF receptor signaling. *Science* **294(5551):** 2511-2515.

Kratsios P, Stolfi A, Levine M, Hobert O. 2011. Coordinated regulation of cholinergic motor neuron traits through a conserved terminal selector gene. *Nat Neurosci* **15(2):** 205-214.

Kuntz SG, Schwarz EM, DeModena JA, De Buysscher T, Trout D, Shizuya H, Sternberg PW, Wold BJ. 2008. Multigenome DNA sequence conservation identifies Hox cis-regulatory elements. *Genome Res* **18(12):** 1955-1968.

Li C, Kim K. 2008. Neuropeptides. *WormBook.*

Markstein M, Markstein P, Marstein V, Levine MS. 2002. Genome-wide analysis of clustered Dorsal binding sites identifies putative target genes in the *Drosophila* embryo. *Proc Natl Acad Sci U.S.A.* **99(2):** 763-768.

Mello C, Fire A. 1995. DNA transformation. *Methods Cell Biol* **48:** 451-482.

Michelson AM. 2002. Deciphering genetic regulatory codes: A challenge for functional genomics. *Proc Natl Acad Sci U.S.A.* **99(2):** 546-548.

Murray JI, Bao Z, Boyle TJ, Boeck ME, Mericle BL, Nicholas TJ, Zhao Z, Sandel MJ, Waterston RH. 2008. Automated analysis of embryonic gene expression with cellular resolution in C. elegans. *Nat Methods.* **5(8):** 703-709.

Popovici C, Isnardo D, Birnbaum D, Roubin R. 2002. *Caenorhabditis elegans* receptors related to mammalian vescular endothelial growth factor receptors are expressed in neural cells. *Neurosci Lett* **329:** 116-120.

Pujol N, Torregrossa P, Ewbank JJ, Brunet JF. 2000. The homeodomain protein CePhox2/CEH-17 controls antero-posterior axonal growth in *C. elegans. Development* **127:** 3361-3371.

Schwarz EM, Kato M, Sternberg PW. 2012. Functional transcriptomics of a migrating cell in *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A.* **109(40):** 16246-16251.

Shapiro IM. 2011. An EMT-driven alternative splicing program occurs in human breast cancer and modulates cellular phenotype. *PLos Genet* **7:**e1002218.

Spencer WC, Zeller G, Watson JD, Henz SR, Watkins KL, McWhirter RD, Petersen S, Sreedharan VT, Widmer C, Jo J, Reinke V, Petrella L, Strome S, Von Stetina SE, Katz M, Shaham S, Rätsch G, Miller DM III. 2011. A spatial and temporal map of *C. elegans* gene expression. *Genome Res.* **21:** 325-341.

Sudhaus W, Kiontke K. 1996. Phylogeny of Rhabditis subgenus *Caenorhabditis* (Rhabditidae, Nematoda). *J Zoo Sys. Evol* **34:** 217-233.

Sudhaus W, Kiontke K. 2007. Comparison of the cryptic nematode species *Caenorhabditis brenneri* sp. N and *C. remanei* (Nematoda: Rhabditidae) with the stem species pattern of the *Caenorhabditis elegans* group. *Zootaxa* **1456:** 45-62.

Sulston JE, Horvitz HR. 1977. Post-embryonic cell lineages of the nematodes, *Caenorhabditis elegans*. *Dev Biol* **56:** 110-156.

Svendsen PC, McGhee JD. 1995. The *C. elegans* nueronally expressed homeoboxgene ceh-10 is closely related to genes expressed in the vertebrate eye. *Development* **121(5):** 1253-1262.

Van Buskirk C, Sternberg PW. 2007. Epidermal growth factor signaling induces behavioral quiescence in *Caenorhabditis elegans. Nat Neurosci* **10(10):** 1300-1307.

Van Buskirk C, Sternberg PW. 2010. Paired and LIM class homeodomain proteins coordinate differentiation of the *C. elegans* ALA neuron. *Development* **137:** 2065-2074.

Wenick AS, Hobert O. 2004. Genomic cis-regulatory architecture and trans-acting regulators of a single interneuron-specific gene battery in *C. elegans. Dev Cell* **6(6):** 757-770.

White JG, Southgate E, Thomson JN, Brenner FRS. 1996. The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Philos Trans R Soc Lond B Biol Sci* **314:** 1-340.

Wu MZ, Tsai YP, Yang MH, Huang CH, Chang SY, Chang CC, Teng SC, Wu KJ. 2011. Interplay between HDAC3 and WDR5 is essential for hypoxia-induced epithelial-mesenchymal transition. *Mol Cell* **43(5):** 811-822.

Wu Y, Ghitani A, Christensen R, Santella A, Du Z, Rondeau G, Bao Z, Colón-Ramos D, Shroff H. 2011. Inverted selective plane illumination microscopy (iSPIM) enables coupled cell identity lineaging and neurodevelopmental imaging in Caenorhabditis elegans. *Proc Natl Acad Sci U S A*. **108(43):** 17708-17713.

Zahn TR, Macmorris MA, Dong W, Day R, Hutton JC. 2001. IDA-1, a *Caenorhabditis elegans* homolog of the diabetic autoantigens IA-2 and phogrin, is expressed in peptidergic neurons in the worm. *J Comp Neurol* **429:** 127-143.

Zhang F, Bhattacharya A, Nelson JC, Abe N, Gordon P, Lloret-Fernandez C, Maicas M, Flames N, Mann RS, Colón-Ramos DA, Hobert O. 2014. The LIM and POU homeobox genes ttx-3 and unc-86 act as terminal selectors in distinct cholinergic and serotonergic neuron types. *Development* **141(2):** 422-435.

Figure 1.

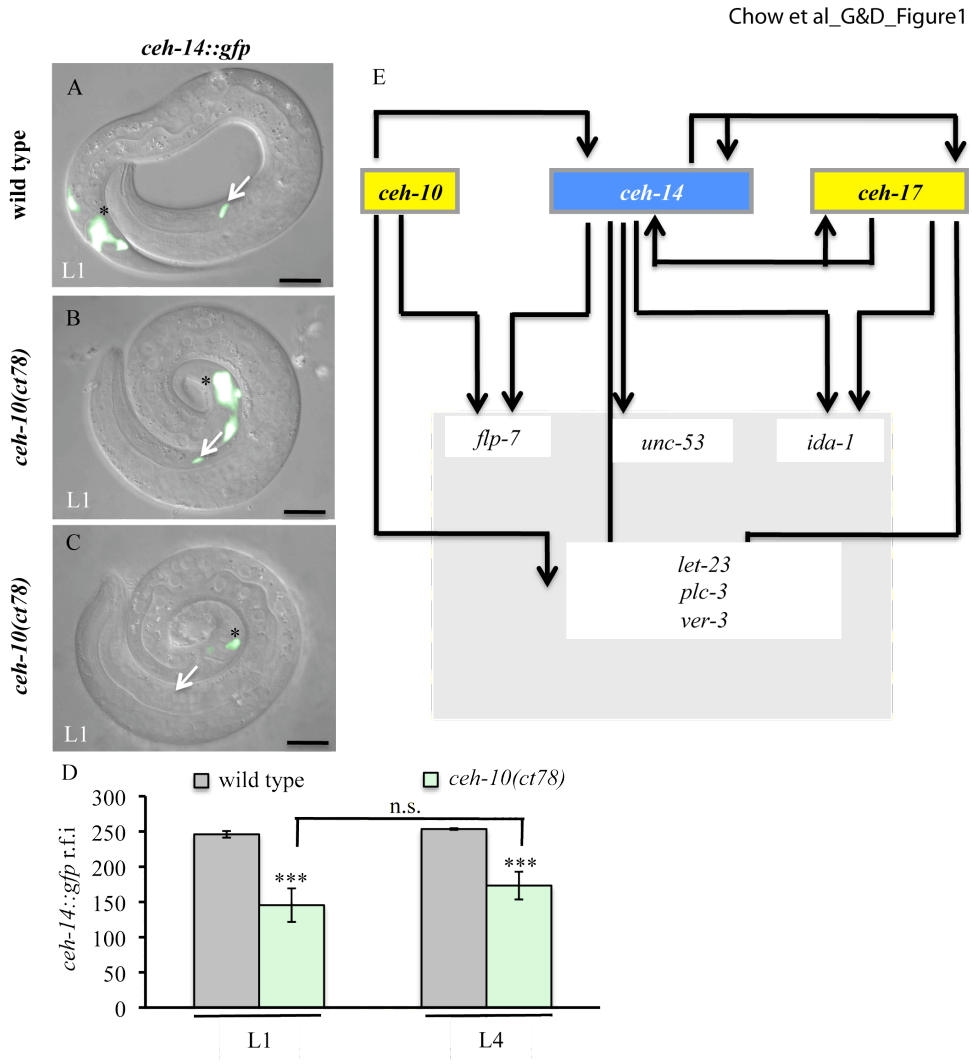

Chow et al_G&D_Figure1
ceh-14::gfp
wild type
ceh-10(ct78)
ceh-10(ct78)
A
B
C
L1
L1
L1
E
ceh-10
ceh-14
ceh-17
flp-7
unc-53
ida-1
let-23
plc-3
ver-3
D
wild type
ceh-10(ct78)
ceh-14::gfp r.f.i
300
250
200
150
100
50
0
n.s.
***
***
L1
L4

Figure 2.

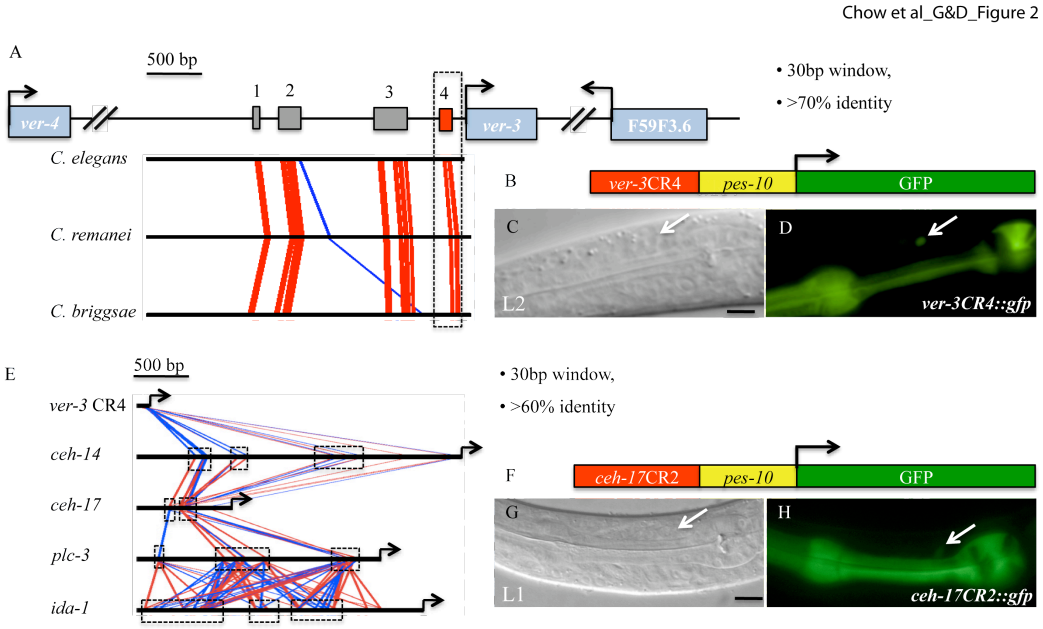

Chow et al_G&D_Figure 2
A
500 bp
1
2
3
4
ver-4
ver-3
F59F3.6
• 30bp window,
• >70% identity
C. elegans
C. remanei
C. briggsae
B
ver-3CR4
pes-10
GFP
C
L2
D
ver-3CR4::gfp
E
500 bp
ver-3 CR4
ceh-14
ceh-17
plc-3
ida-1
• 30bp window,
• >60% identity
F
ceh-17CR2
pes-10
GFP
G
L1
H
ceh-17CR2::gfp

Figure 3.

Chow et al_G&D_Figure 3

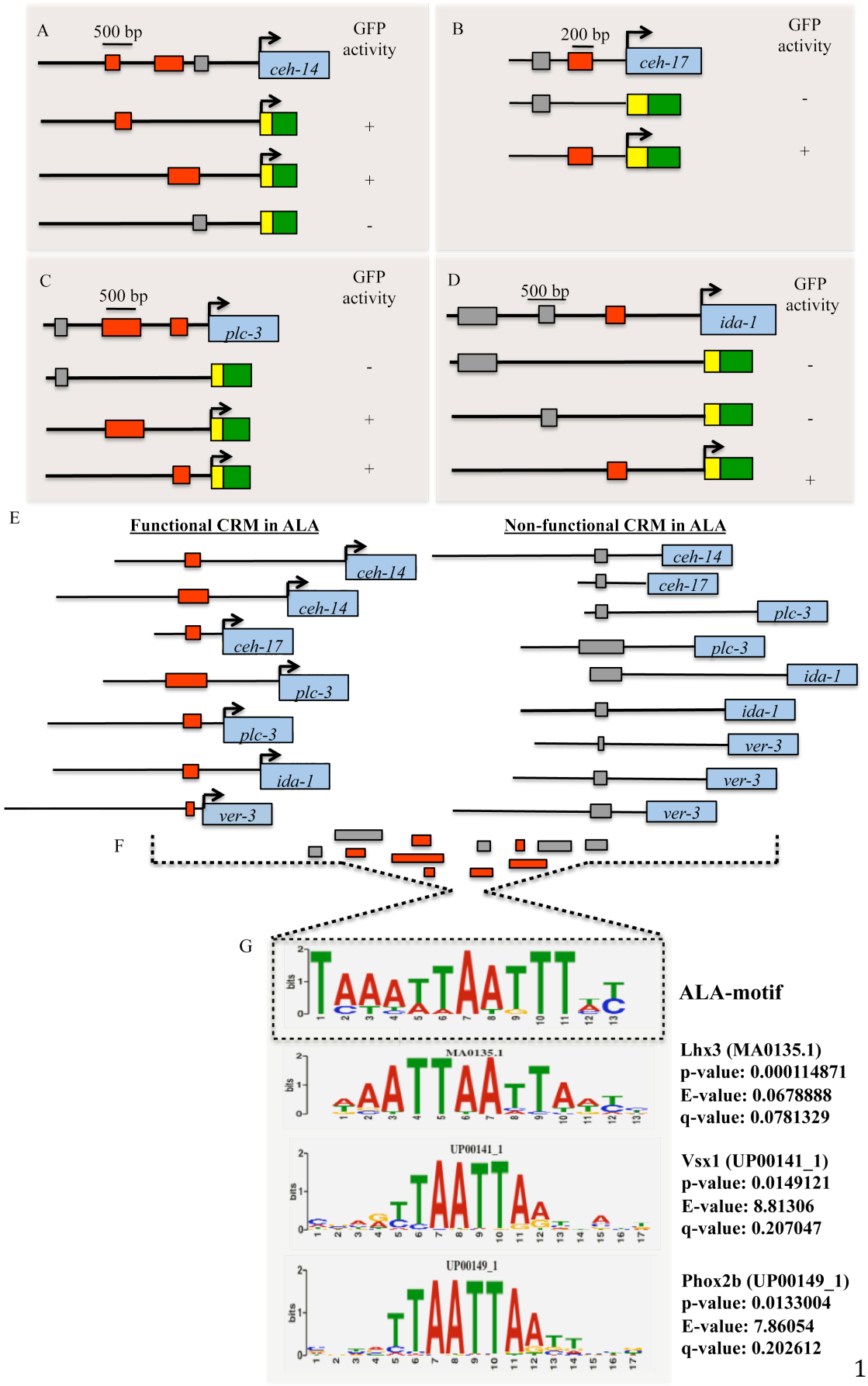


Figure 4.

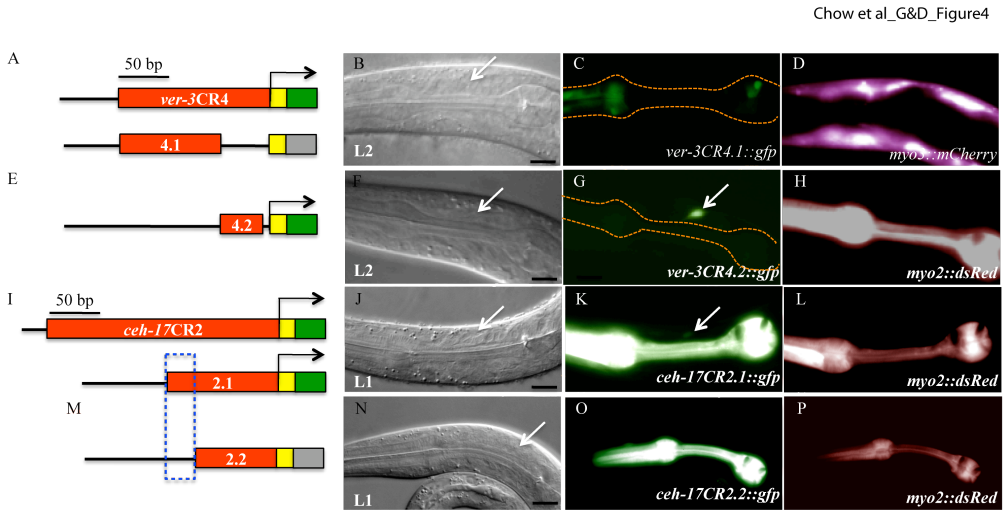

Chow et al_G&D_Figure4
A
50 bp
ver-3CR4
4.1
E
4.2
I
50 bp
ceh-17CR2
2.1
M
2.2
B
L2
C
ver-3CR4.1::gfp
D
myo2::mCherry
F
L2
G
ver-3CR4.2::gfp
H
myo2::dsRed
J
L1
K
ceh-17CR2.1::gfp
L
myo2::dsRed
N
L1
O
ceh-17CR2.2::gfp
P
myo2::dsRed

Figure 5.

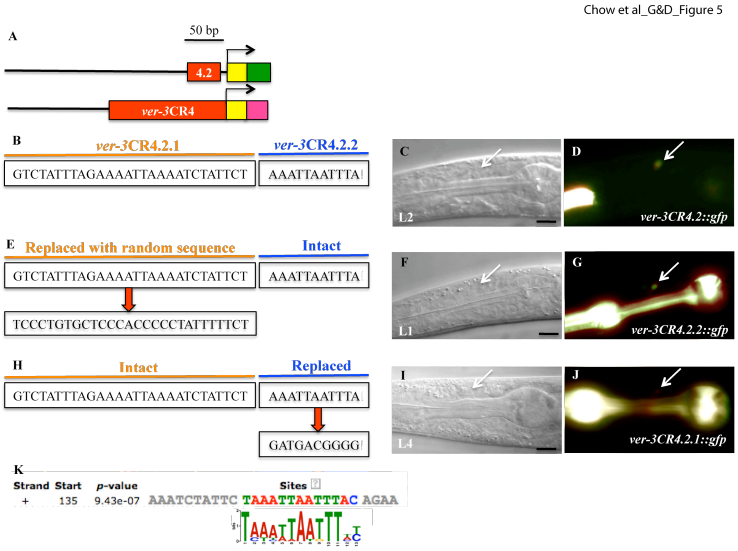

Chow et al_G&D_Figure 5
A
50 bp
4.2
ver-3CR4
B
ver-3CR4.2.1
ver-3CR4.2.2
GTCTATTTAGAAAATTAAAATCTATTCT
AAATTAATTTA
C
L2
D
ver-3CR4.2::gfp
E
Replaced with random sequence
Intact
GTCTATTTAGAAAATTAAAATCTATTCT
AAATTAATTTA
TCCCTGTGCTCCCACCCCCTATTTTTCT
F
L1
G
ver-3CR4.2.2::gfp
H
Intact
Replaced
GTCTATTTAGAAAATTAAAATCTATTCT
AAATTAATTTA
GATGACGGGG
I
L4
J
ver-3CR4.2.1::gfp
K
Strand Start p-value Sites
+ 135 9.43e-07 AAATCTATTC TAAATTAATTTAC AGAA

Figure 6

Chow et al_G&D_Figure 6

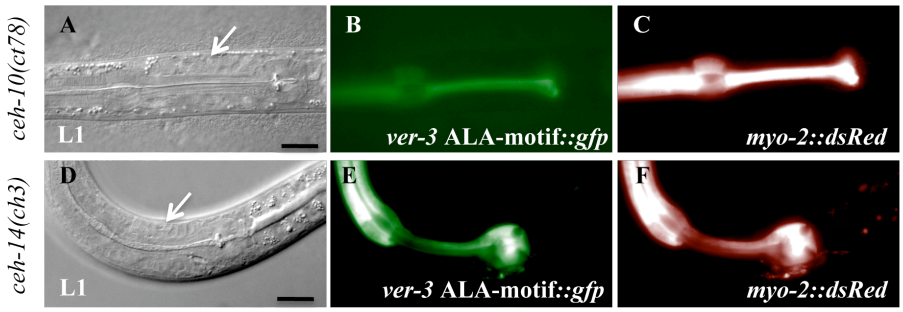
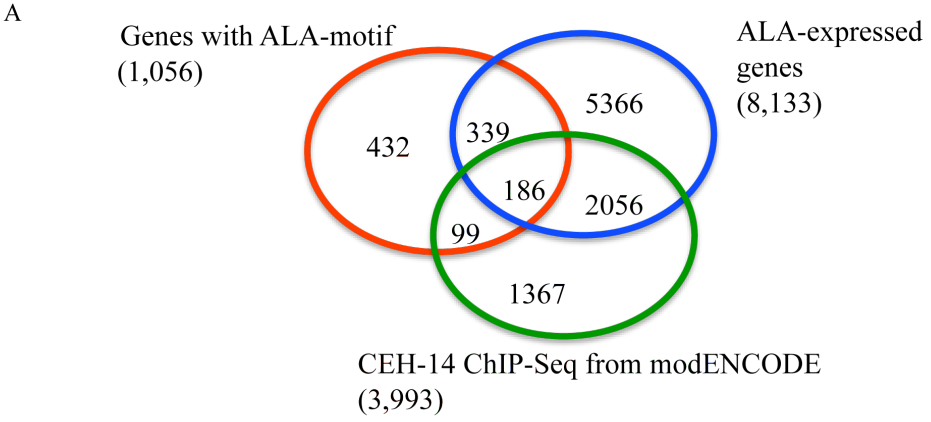


Figure 7.

Chow et al_G&D_Figure 7

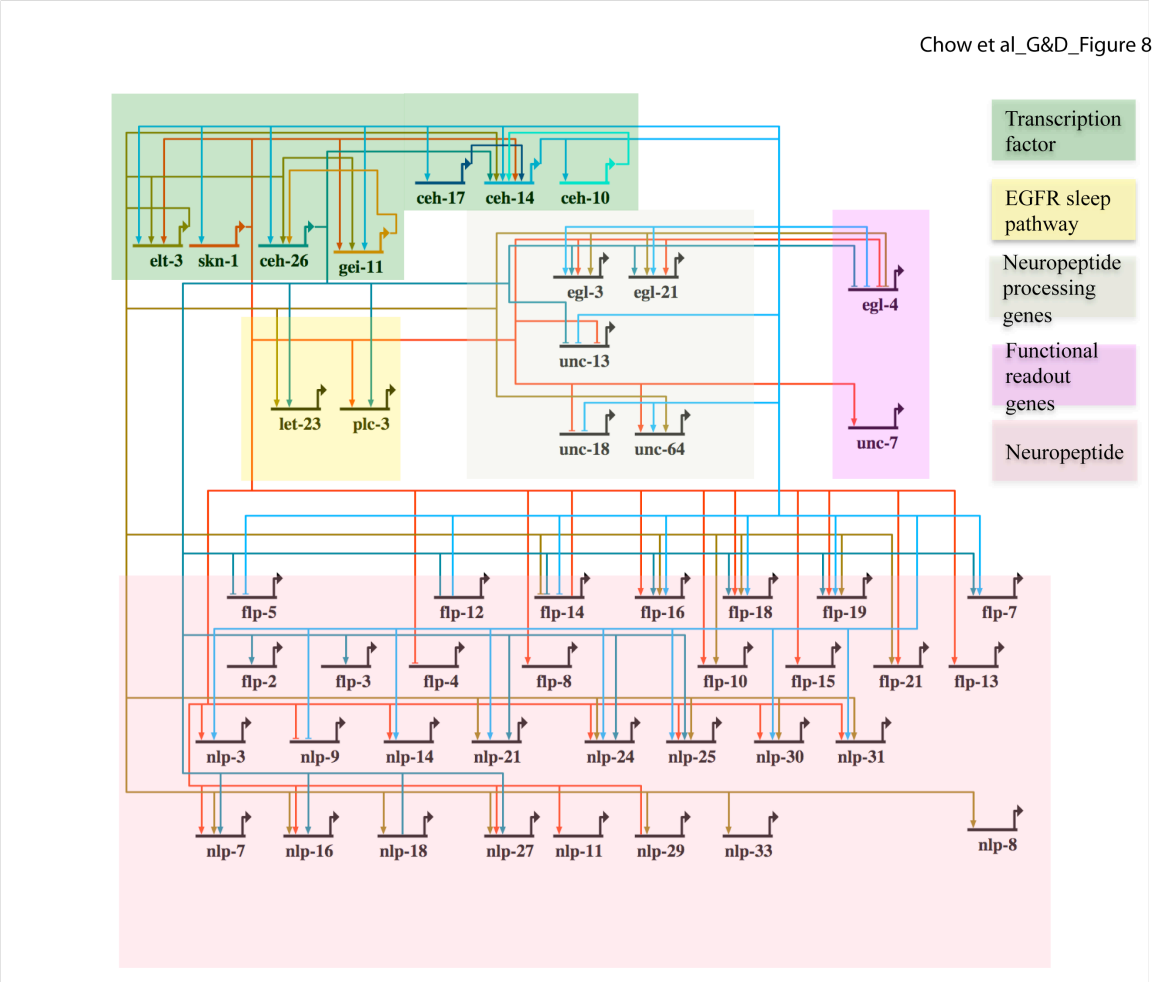
A



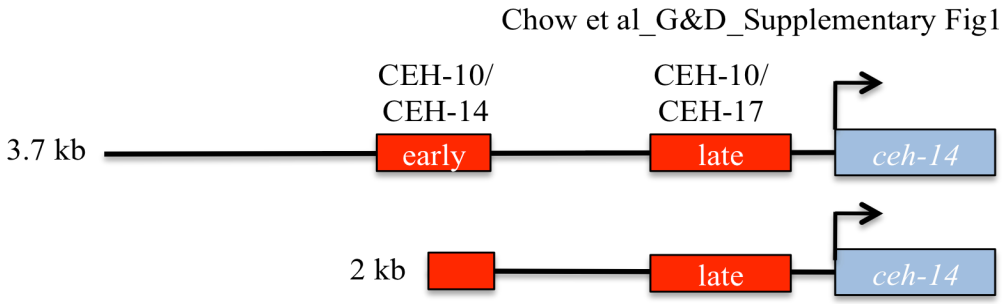
B

	Motif-searched (1,056 genes)	Transcriptome profiling (8,133 genes)
CEH-14 direct-regulated in detected genes	285 (27%)	2,242 (27.6%)
ALA-expressed	525 (49.7%)	8,133 (100%)
CEH-14 direct-regulated among ALA-expressed	186 (35.4%)	2,242 (27.6%)

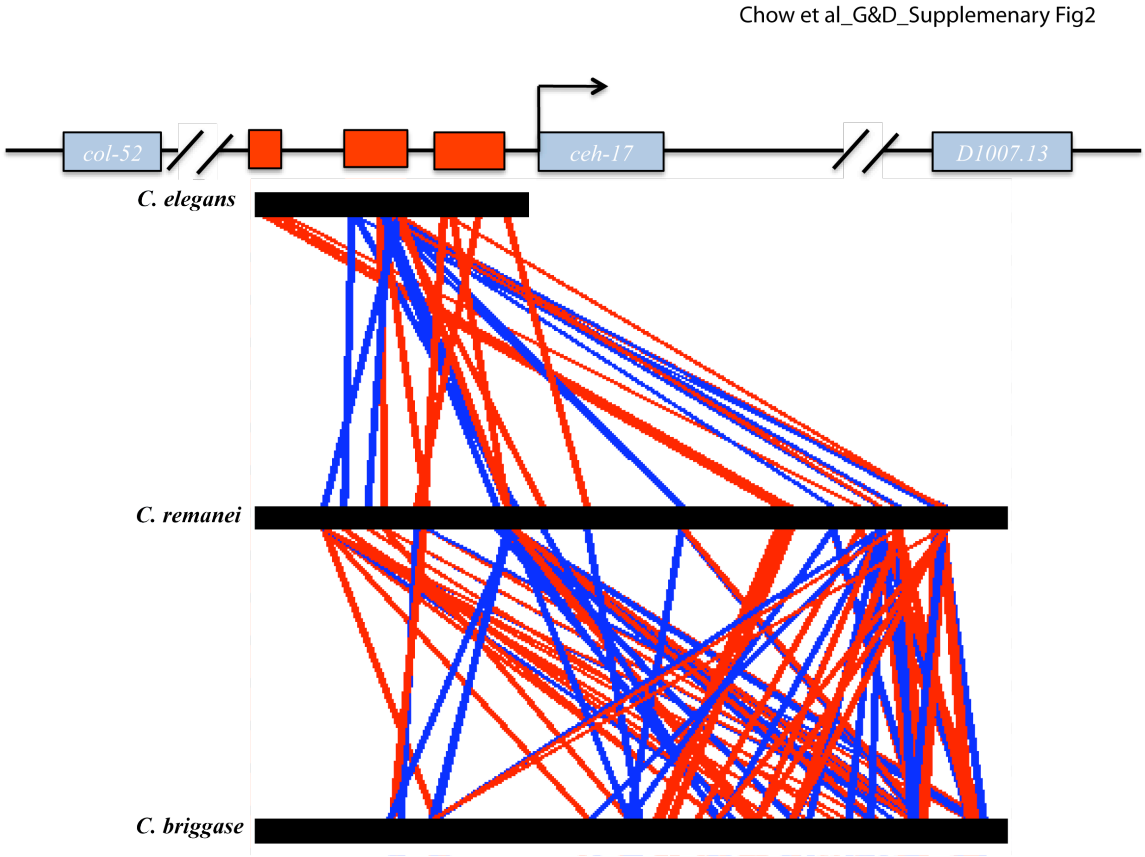
Figure 8.


Chow et al_G&D_Figure 8
Transcription factor
EGFR sleep pathway
Neuropeptide processing genes
Functional readout genes
Neuropeptide
ceh-17
ceh-14
ceh-10
elt-3
skn-1
ceh-26
gei-11
egl-3
egl-21
egl-4
unc-13
let-23
plc-3
unc-18
unc-64
unc-7
flp-5
flp-12
flp-14
flp-16
flp-18
flp-19
flp-7
flp-2
flp-3
flp-4
flp-8
flp-10
flp-15
flp-21
flp-13
nlp-3
nlp-9
nlp-14
nlp-21
nlp-24
nlp-25
nlp-30
nlp-31
nlp-7
nlp-16
nlp-18
nlp-27
nlp-11
nlp-29
nlp-33
nlp-8

Supplementary Fig. 1

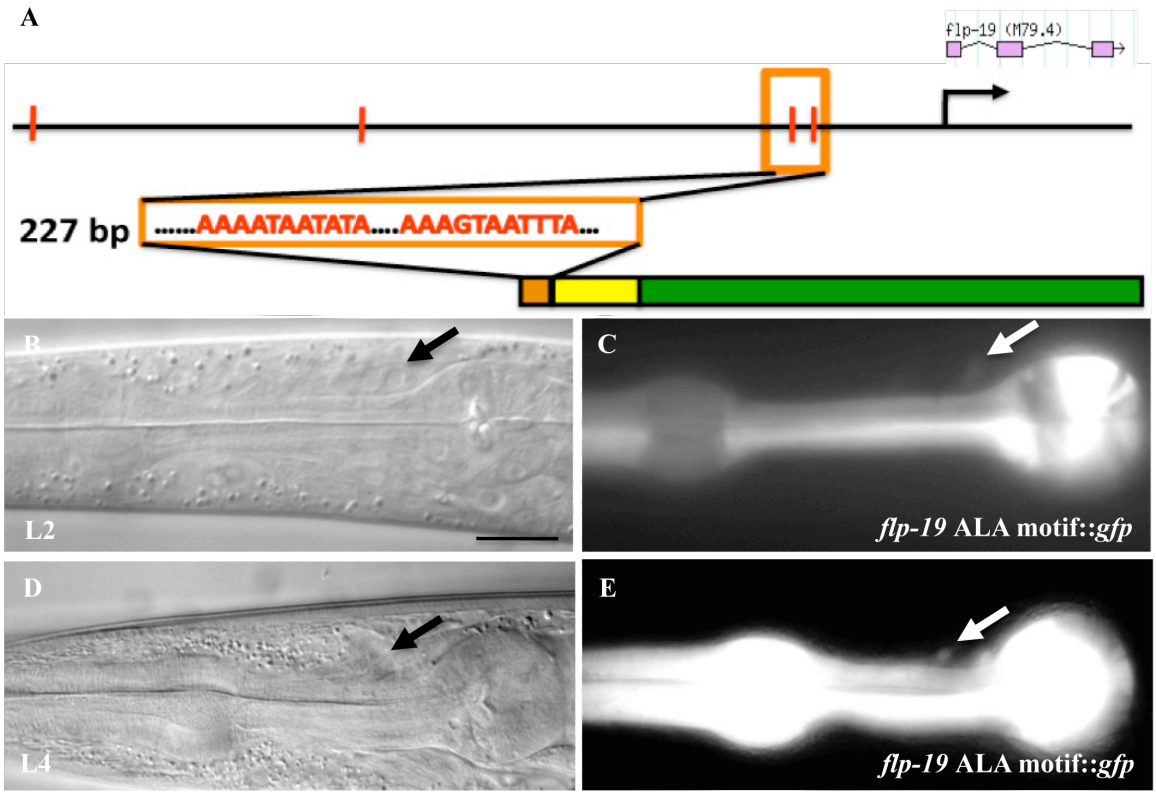


Supplementary Fig. 2



Supplementary Fig. 3

Chow et al_G&D_Supplementary Fig 3



Supplementary Table 3

	Transcriptome profiling (8,133 genes in wild type)
ALA-expressed gene positively regulated by CEH-14	1,055 (13%)
ALA-expressed and >10 folds higher in wild type	3,171 (39%)
ALA-expressed, CEH-14-dependent, CEH-14 direct regulated	729 (23%)

Chapter 7

Findings, Implications, Future Plans

Abstract

Animals are constantly interacting with each other and the environment; however, the underlying molecular mechanisms driving behavior remains a mystery. *C. elegans* is capable of executing a wide range of behavior and shows decision making in response to external stimulations. The animal has a small nervous system that is packed with 302 neurons; each possesses specific features and functions. While the delineated connectome provides a clear hard-wired map of the nervous system, it lacks the molecular explanation of decision-making. Gene expression and function are subject to local genetic information that governs transcriptional control and post-transcriptional control such as the presence of intracellular signaling pathways that integrate cell-cell molecular messages they receive into a unified action plan. It is therefore of interest and important to identify genes that are co-expressed in a neuron and to decipher the gene regulatory interactions that give rise to its functions and features. Understanding the functional genomics of individual neurons can help establishing a functional map of the *C. elegans* nervous system and potentially those of other model organisms. As a starting point, I chose to profile the transcriptome of a single ALA neuron in the *C. elegans* head. This chapter summarizes the findings, implications and discusses future plans to extend our understanding of genetic codes and animal behaviors.

Findings

1. Identified transcriptomes in a single neuron
2. Identified novel function of FMRFamide neuropeptides as sleep-promoting molecules in *C. elegans* and in zebrafish
3. Identified GPCRs that modulates FMRFamide neuropeptide in regulating sleep

4. Identified a cell-specific motif for genome-wide search of potential enriched genes in a sleep-inducing neuron
5. Constructed a gene regulatory network of *C. elegans* sleep behavior

Implications

Animals are constantly threatened by predators for survival and by competitors for resources. Behavior is complex, dynamic and subject to fluctuations of external and internal cues. With a limited number of neurons in the nervous system but seemingly unlimited challenges, how does an animal decide whether to feed, to flee for life, to roam around, or to sleep? The question remains: how are multiple factors integrated to give rise to a unified action?

The hard-wired connectome represents a set of potential connections of information flow in the nervous system. However, this information flow is shaped by genetic information and the internal state of individual neurons (Bargmann, 2012). With this note in mind, it is reasonable to think that a neuron may perform differently in response to external inputs and change of internal states. Neuromodulators are often the molecules that differentiate context and internal states of a neuron (Bargmann, 2012), and they act through GPCRs to modify neuronal dynamics, excitability and synaptic efficiency. In *C. elegans* and other animals, neuropeptides and neurotransmitters are the major neuromodulators (Bargmann, 1998).

This work provides the first evidence of neuromodulation through genetic information at the single-neuron resolution. FMRFamide neuropeptides are known to modulate pain perception (Lameh et al., 2010) and feeding behavior (Dockray, 2004) in

avian and other model organisms. Here, my data show that FMRFamide neuropeptides have a novel function in promoting sleep in worm and fish (Chapter 5). FMRFamide neuropeptides are widely expressed in *C. elegans* and one neuropeptide-coding gene can be found in multiple neurons (Li and Kim, 2008). The sleep-inducing neuropeptide-coding gene, *flp-13*, is expressed in sensory neurons (ASE, ASG, ASK) involved in water-soluble attractant chemotaxis (Bargmann, 2006), indicating that *flp-13* may play multiple roles in different neurons depending on the genomic information of individual neuron. Alternatively, the presence of *flp-13* in ASE, ASG and ASK may indicate that these neurons are involved in sleep regulation, and it remains to be tested.

The ALA sleep GRN elucidates the correlation of internal state and genomic information by connecting the dots of intracellular signaling pathway (EGFR signaling) and neuromodulators (FLP neuropeptides). Gathering the genomic information of a neuron is by all means informative, but the core messages lie at the regulatory level that governs the neuronal internal state. Here, I provide a network to illustrate the fine details of transcriptional inputs that regulate the expression of genes encoding neuropeptides and members in the EGFR signaling pathway in the ALA neuron. In addition to 23 genes coding for FMRFamide neuropeptide (*flp*), there are 25 genes coding for non-FMRFamide neuropeptides in *C. elegans* (*nlp*), and 8 genes coding for insulin-like peptides (*ins*) expressed in ALA (Chapter 6). It is possible that neuropeptides other than the FMRFamide class neuropeptides are involved in the EGFR sleep regulation and that employment of neuropeptides may be state dependent. Alternatively, other intracellular signaling molecules may be involved in the sleep pathway. A well-described regulatory

network can provide a road map for gene manipulation to single out the function of a neuropeptide in a given context at a given time.

Novel molecular players in sleep regulation can be identified using transcriptome profiling and validation in simple model organisms. Here I showed that FMRFamide neuropeptides induce sleep in *C. elegans*. However, there is no clear correlation of *C. elegans* neuropeptides with vertebrate neuropeptides (Bargmann, 1998), but it is possible to discern the equivalent of *C. elegans* neuropeptides by identifying their receptors and comparing their similarity to subgroups of vertebrate receptors. This work reports that FMRFamide neuropeptides are evolutionarily conserved sleep-promoters in worm and fish. It first shows that a zebrafish FMRFamide neuropeptide can induce sleep in *C. elegans*, and then identified the *C. elegans* receptor that it acts through. This receptor also mediates *C. elegans* FMRFamide neuropeptides to induce sleep. The easy amenability of *C. elegans* can serve as a starting point for screening potential functional genes, which can be validated in a vertebrate model like zebrafish. Furthermore, the zebrafish nervous system has neuroanatomical and neurochemical systems similar to that of mammals (Chiu and Prober, 2013), making it a media to correlate functional genes found in *C. elegans* with mammalian behavior. The conserved neuropharmacology of mammals and zebrafish (Rihel and Schier, 2012) is an added bonus for testing the response of novel functional genes to drugs and bioactive agents.

Future plans

This work accomplished the goal to identify genes expressed in a neuron and to correlate the function of these genes with the function of the neuron. Here, I used the *C.*

elegans ALA neuron and its function in regulating sleep as a behavioral readout. Taking an unbiased transcriptome profiling approach, this work identified novel neuropeptides that regulate animal sleep. Neuropeptides act through GPCRs (Bargmann 1998) and play crucial role in modulating the nervous system across phyla (Kow and Pfaff, 1988). The approach of identifying genomic information of a neuron, particularly neuropeptides, and correlating it with a known behavioral output can serve as an entry point to expand our understanding on the function of individual neuron. However, the nervous system is intrigue and behavior is the reflection of functional circuits, an amalgamated effect of the functional outputs of multiple neurons. Understanding the functional genomics of one neuron is not sufficient to explain the complex of animal behavior. We now have the tools to profile transcriptomes of a single neuron and to analysis gene function in a regulatory network scale. Together with the newly reported technology that allows simultaneous imaging of up to 70% of neuronal activity in the *C. elegans* brain (Schrödel et al., 2013), it is possible to decipher the functional genomics of a nervous system to further yield insights into the correlation of genetic codes and behavioral outputs in *C. elegans* and potentially other model organisms. To further expand the picture of behavior regulation, take sleep as an example, there are a few tasks to be accomplished:

1. Investigate sleep- or wake-promoting functions of other neuropeptides in ALA
2. Identify the expression of neuropeptide receptor genes by generating *promoter::gfp* transgenic lines of the GPCRs, *npr-7* and *npr-22*.
3. Validate GPCR function by rescuing receptor null mutants and overexpressing neuropeptides in these rescued null mutants.

4. Identify genes in the neuropeptide receptor-expressed neuron(s) by single-cell transcriptome profiling.
5. Investigate circuitry dynamic and function of the ALA neuron related sleep circuit by optogenetics image and compare it with other sleep-regulating circuits (e.g. lethargus and satiety) in *C. elegans*.

Reference

1. Bargmann, C.I. (2006). Chemosensation in *C. elegans*. WormBook ed. The C. elegans Research Community:doi/10.1895/wormbook, http://www.wormbook.org.
2. Bargmann, C.I. (2012). Beyond the connectome: How neuromodulators shape neural circuits. Bioessays *34,* 458-465.
3. Bargmann, C.I.. (1998). Neurobiology of the *Caenorhabditis elegans* genome. Science *282(5396),* 2028-2033.
4. Chiu, C.N., and Prober, D.A. (2013). Regulation of zebrafish sleep and arousal states: current and prospective approaches. Frontiers in Neural Circuits *7:58.*
5. Dockray, G.J. (2004). The expanding family of –RFamide peptides and their effects on feeding behavior. *Exp. Physiol.* **89(3)**, 229-235. (2004).
6. Lameh, J., Bertozzi, F., Kelly, N., Jacobi, P.M., Nguyen, D., Bajpai, A., Gaubert, G., Olsson, R., Gardell, L.R. (2010). Neuropeptide FF receptors have opposing modulatory effects on nociception. *J. Pharmacol. Exp. Ther. 334(1),* 244-254.
7. Li, C., and Kim, K. (2008). Neuropeptides. WormBook ed. The C. elegans Research Community:doi/10.1895/wormbook, http://www.wormbook.org.
8. Rihel, J., and Schier, A.E. (2012). Behavioral screening for neuroactive drugs in zebrafish. Dev. Neurobiol. *72,* 373-385.
9. Schrödel, T., Prevedel, R., Aumayr, K., Zimmer, M., and Vaziri, A. (2013). Brain-wide 3D imaging of neuronal activity in *Caenorhabditis elegans* with sculpted light. Nat. Methods *10(10),* 1013-1020.

Appendix

Supplementary table of ALA transcriptomes

Appendix

Supplementary table of ALA transcriptome in wild type and *ceh-14(ch3)* null mutant ALA neuron.

Gene	Wild type ALA	*ceh-14(ch3)* ALA	Larvae	*ceh-14* dependent	ALA enrichment
WBGene00016028|C24A1.1|flp-24	10176.34	3435.34	1.59	2.96	6400.21
WBGene00020171|T02H6.1	169.62	95.57	0	1.77	5654.00
WBGene00016669|C45G7.2|ilys-2	161.7	0.44	0	367.50	5390.00
WBGene00015380|C03B1.10	93.83	39.51	0	2.37	3127.67
WBGene00014848|VM106R.1	92.61	39.98	0	2.32	3087.00
WBGene00010377|H12D21.1|nspa-1	43.92	5.58	0	7.87	1464.00
WBGene00018917|F56A4.9	128.89	258.07	0.11	0.50	1171.73
WBGene00045401|T26F2.3	30.4	0.13	0	233.85	1013.33
WBGene00008658|F10F2.6|clec-152	24.49	0.06	0.03	408.17	816.33
WBGene00021334|Y34D9A.10|vps-4	124.34	155.43	0.2	0.80	621.70
WBGene00009853|F48F7.8	18.18	0.14	0	129.86	606.00
WBGene00013888|ZC412.7|nspa-3	16.08	2.64	0	6.09	536.00
WBGene00021222|Y19D10A.7	148.44	297.2	0.29	0.50	511.86
WBGene00001816|Y48G8AL.11|haf-6	14.25	4.39	0	3.25	475.00
WBGene00019842|R02F11.4	13.32	0.03	0	444.00	444.00
WBGene00011673|T10B9.3|cyp-13A6	29.82	0.07	0.07	426.00	426.00
WBGene00003474|T08G5.10|mtl-2	10.87	32.93	0	0.33	362.33
WBGene00001756|F11G11.1|gst-8	10.21	0.15	0	68.07	340.33
WBGene00001450|F49E10.3|flp-7	4518.55	2614.15	13.51	1.73	334.46
WBGene00004806|T27F2.1|skp-1	12.95	3.5	0.04	3.70	323.75
WBGene00005953|K09D9.10|srx-62	9.45	0	0	315.00	315.00
WBGene00001781|C02A12.1|gst-33	87.52	7.49	0.3	11.68	291.73
WBGene00003890|T20H4.1|osm-10	11.63	2.49	0.04	4.67	290.75
WBGene00001456|F33D4.3|flp-13	2063.11	194.4	7.52	10.61	274.35
WBGene00018562|F47D12.6	7.82	0.11	0	71.09	260.67
WBGene00005656|C13D9.2|srr-5	7.81	0	0	260.33	260.33
WBGene00009529|F38B2.2	918.67	1527.59	3.7	0.60	248.29
WBGene00044570|T22B7.8	7.42	0	0	247.33	247.33
WBGene00022448|Y110A2AL.13	388.24	503.4	1.57	0.77	247.29
WBGene00010971|R01E6.7	103.57	0.15	0.43	690.47	240.86
WBGene00009692|F44E5.5	136.37	6.32	0.6	21.58	227.28
WBGene00020269|T05H4.6	1540.36	1405.12	6.9	1.10	223.24
WBGene00012294|W06A7.5|nspa-8	6.41	1.14	0	5.62	213.67
WBGene00077682|F58B4.7	6.24	0.18	0	34.67	208.00
WBGene00013765|Y113G7B.16	8.12	0.03	0.04	270.67	203.00
WBGene00020515|T14G12.6	28.28	0.05	0.14	565.60	202.00
WBGene00044271|H12D21.12|nspa-2	5.9	0.88	0	6.70	196.67
WBGene00003843|ZK455.8|oct-2	541.03	9.16	2.79	59.06	193.92
WBGene00009174|F26H9.2	123.36	463.58	0.66	0.27	186.91
WBGene00044272|H12D21.13|nspa-4	5.55	0.91	0	6.10	185.00
WBGene00012799|Y43F4A.4	5.12	0	0	170.67	170.67
WBGene00009750|F46A8.7	150.63	204.36	0.9	0.74	167.37
WBGene00014002|ZK593.2	125.21	0.4	0.76	313.02	164.75
WBGene00010011|F53H4.6	6.47	0	0.04	215.67	161.75
WBGene00022733|ZK418.3	4.74	0	0	158.00	158.00