

**DEVELOPMENT AND FUNCTION OF
SLEEP REGULATORY CIRCUITS IN ZEBRAFISH**

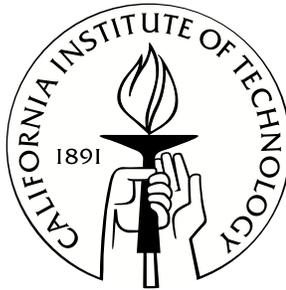
Thesis by

Justin Liu

In Partial Fulfillment of the Requirements

for the Degree

of Doctor of Philosophy



CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2015

(Defended April 2, 2015)

© 2015

Justin Liu

All Rights Reserved

For my parents

ACKNOWLEDGMENTS

I'd first like to thank my advisor, David Prober, who afforded me the opportunity to explore and grow as a scientist. As one of his first students, I witnessed how David built the lab from a virtually empty room into the bustling research facility it is today. Since we study sleep, it's actually very quiet, but make no mistake: our videotrackers are always running. Through it all, David has kept his door open for his students, but I suspect he may not know just how much we appreciate his availability and concern. So thank you, David, for investing so much time and energy on our behalf.

I also thank the members my committee: Paul Sternberg, who combines an obvious passion for science with pragmatism and humor; Angelike Stathopoulos, who generously shared her knowledge and was quick to suggest “smoking gun” experiments; and Henry Lester, who has unerringly given excellent advice not just on science but on how to succeed as a scientist.

Briefly, I would also like to acknowledge the people who mentored me before my arrival at Caltech: Grace Rosenquist, who first taught me to be meticulous and thorough; Kristin Ostrow, who showed me that some of the most productive research is collaborative; Liqun Luo, who once said, “Sometimes, the correct strategy in biology is brute force”; and Maria Spletter, who patiently introduced me to countless useful techniques, including confocal microscopy.

During my doctoral studies, I have been fortunate that the terms “colleague” and “friend” were often interchangeable. Every member of the Prober lab, without exception, has helped in some way with the research in this thesis, for which I am extremely grateful. In particular, I want to thank Daniel Lee, who has been a collaborator,

cheerleader, and friend all at once. I would especially like to thank Wendy Chen and Avni Gandhi for countless impromptu discussions of confusing experimental results, moral support during late nights in lab, and our shared belief that difficult days are best remedied with desserts and puppies. We began the journey of graduate school together and I am certain that I would not have finished it without their friendship.

Finally, I'd like to thank Alberto Stochino and Erica Cherry-Kemmerling, who gave their unconditional love and support despite long distances. And I want to thank my parents, Kojam and Tai-Ying, and my sister, June, who have always encouraged me to pursue my love of science and have done everything in their power to help me succeed.

PREFACE

“Utnapishtim said,
‘As for you, Gilgamesh,
who will assemble the gods for your sake,
so that you may find that life for which you are searching?
But if you wish,
come and put it to test:
only prevail against sleep
for six days and seven nights.’

But while Gilgamesh sat there resting on his haunches,
a mist of sleep
like soft wool teased from the fleece
drifted over him,
and Utnapishtim said to his wife,
‘Look at him now,
the strong man who would have everlasting life,
even now the mists of sleep are drifting over him.’”

-from the *Epic of Gilgamesh*

The Sumerian poem, *Epic of Gilgamesh*, was set in clay tablets over 4,000 years ago and is considered the first book ever written. In it, Gilgamesh, the mighty warrior king of Uruk, seeks to learn the secret of immortality from the mystic Utnapishtim, who reluctantly agrees to teach Gilgamesh the secret if he can first pass a strange challenge: avoid sleeping for one week. Despite his god-like strength and unparalleled accomplishments, Gilgamesh ultimately fails the test and is denied immortality. The moral of this ancient tale is simple: sleep, like death, is an inescapable part of life. And no matter how strong or resolved the hero, he will still succumb to sleep.

For a behavioral biologist like myself, the allure of sleep is both physical and emotional. Like everyone else, I experience an uncontrollable urge to sleep, but the mystery of why and how this urge arises is just as irresistible. Grappling with that mystery has been a grand adventure – an epic, you might say. I have enjoyed the journey.

ABSTRACT

Sleep is widely accepted as an essential behavior for optimum mental and physical health, yet the genetic and neural circuits that govern sleep remain poorly understood. In this thesis, I briefly introduce the behavioral criteria that define sleep, currently known sleep regulatory mechanisms, and the distinct advantages of the zebrafish, *Danio rerio*, as a simple animal model for studying sleep. I then investigate two factors previously implicated in sleep behavior: epidermal growth factor receptor and hypocretin. First, I show that epidermal growth factor receptor signaling is both necessary and sufficient for normal sleep behavior in zebrafish, just as it is in invertebrates. This demonstrates that sleep regulatory mechanisms can be conserved over large evolutionary distances, and is the first genetic study showing that the epidermal growth factor receptor signaling is necessary for normal sleep behavior in a vertebrate. Second, I capitalize upon the rapid external development of zebrafish embryos to screen for developmental factors that specify hypocretin neurons, which are known to promote arousal and consolidate sleep/wake bouts. I identify the LIM homeobox 9 transcription factor as necessary for hypocretin neuronal development in zebrafish and sufficient to specify additional hypocretin neurons in both zebrafish and mice. This is the first time any factor has been shown to induce hypocretin neurons *in vivo* and may be an important step towards curing narcolepsy, a debilitating sleep disorder caused by the selective loss of hypocretin neurons. These studies deepen our understanding of how sleep is regulated at a genetic and cellular level and underscore the potential for zebrafish to make future contributions to sleep research.

TABLE OF CONTENTS

Acknowledgements.....	iv
Preface.....	vi
Abstract.....	vii
Table of Contents.....	viii
Chapter 1: Introduction.....	1
1.1 Defining sleep behavior.....	2
1.2 Known mechanisms of sleep behavior.....	3
1.3 Zebrafish as a simple animal model of sleep.....	5
References.....	9
Chapter 2: Epidermal Growth Factor Receptor Signaling Promotes Sleep in Zebrafish .	12
2.1 Abstract.....	13
2.2 Introduction.....	14
2.3 TGF- α overexpression increases sleep.....	15
2.4 Disruption of EGFR signaling decreases sleep.....	17
2.5 EGFR signaling modulates arousal state.....	21
2.6 EGFR inhibition suppresses TGF- α -mediated sleep.....	22
2.7 TGF- α -mediated sleep requires the MAPK/ERK pathway.....	24
2.8 TGF- α overexpression activates EGFR+ cells along the brain ventricle.....	27
2.9 <i>tgfa</i> expression does not cycle in a circadian manner.....	32
2.10 TGF- α -mediated sleep does not require overt circadian rhythms and is light-dependent.....	34
2.11 Discussion.....	37
2.12 Experimental Procedures.....	39
References.....	44
Supplemental Figures.....	47
Chapter 3: Hypocretin Neuronal Specification.....	49
3.1 Abstract.....	50
3.2 Introduction.....	51
3.3 Microarray analysis identifies transcripts enriched in Hcrt neurons.....	52
3.4 Expression patterns of candidate genes validate the microarray results.....	58
3.5 <i>Lhx9</i> is sufficient to specify Hcrt neurons.....	60
3.6 <i>Lhx9</i> is necessary for Hcrt neuron specification.....	65
3.7 <i>Lhx9</i> directly promotes <i>hcrt</i> expression.....	69
3.8 <i>Lhx9</i> overexpression in mouse embryos induces Hcrt neuron specification.....	72
3.9 Discussion.....	74
3.10 Experimental Procedures.....	79
References.....	86
Supplemental Figures.....	91
Chapter 4: Conclusions and Future Directions.....	108
4.1 Conclusions.....	109
4.2 Future directions.....	109
References.....	111

CHAPTER 1:

Introduction

1.1 Defining sleep behavior

All humans, to some extent, have an intuitive understanding of sleep. However, sleep superficially resembles many other forms of behavioral inactivity, such as quiet wakefulness, hibernation, or coma. To distinguish between these states, sleep is defined by a set of behavioral criteria (Campbell and Tobler, 1984). First, sleep is a period of behavioral quiescence associated with reduced locomotor activity. This quiescent state is rapidly reversible and commonly occurs during a particular phase of the circadian cycle. Second, sleep is accompanied by reduced sensory responsiveness. That is, animals require a stronger stimulus to elicit a behavioral response when asleep than when awake. Finally, sleep behavior is subject to homeostatic regulation, whereby animals prevented from sleeping must compensate for this disturbance with an additional period of sleep rebound later.

By this definition, sleep behavior has been observed in almost all animals (Siegel, 2008). The evolutionary conservation of sleep suggests that it serves some highly adaptive function, such as metabolic homeostasis (Porkka-Heiskanen and Kalinchuk, 2011; Xie et al., 2013), synaptic plasticity (Tononi and Cirelli, 2014), or immune system regulation (Imeri and Opp, 2009). It also suggests that the mechanisms governing sleep behavior may be conserved and can therefore be identified using simple model organisms.

1.2 Known mechanisms of sleep behavior

Indeed, studies across animal species have identified a common set of genes and molecules that modulate sleep behavior. Some of these genes encode core components of the circadian clock, such as *Period* or *Casein kinase I*, which were first described in *Drosophila* (Kloss et al., 1998; Konopka and Benzer, 1971), but play conserved roles in the circadian timing of sleep in humans (Xu et al., 2005). Genes that encode potassium ion channels, such as *Shaker* and *Kv3.1*, appear to modulate neuronal excitability and cause dramatic reductions in sleep when mutated in *Drosophila* and mice (Cirelli et al., 2005; Espinosa et al., 2004). In addition, many neurotransmitter systems have been implicated in sleep/wake behavior. These include norepinephrine, dopamine, histamine, serotonin and acetylcholine, which promote wakefulness, and GABA, which promotes sleep by inhibiting the neural pathways that express the aforementioned wake-promoting neurotransmitters (Cirelli, 2009; Saper et al., 2005).

Neuropeptides form a particularly interesting group of sleep regulators. Unlike neurotransmitters, which are commonly derived from amino acid and metabolic precursors, neuropeptides are directly encoded in the genome and are therefore highly amenable to genetic manipulation. Perhaps the best-characterized sleep regulator is the neuropeptide hypocretin, also known as orexin. In rodents, hypocretin overexpression strongly inhibits sleep and optogenetic stimulation of hypocretin neurons increases the probability of sleep to wake transitions (Adamantidis et al., 2007; Tsujino and Sakurai, 2009). Loss of hypocretin neurons is also the underlying cause of narcolepsy, a debilitating human sleep disorder

characterized by excessive sleepiness, fragmented sleep-wake bouts, and cataplexy (Dauvilliers et al., 2007). The role of hypocretin is therefore to promote and maintain periods of wakefulness. In addition to hypocretin, many other neuropeptides modulate sleep behavior, including melanin-concentrating hormone, leptin, corticotropin-releasing hormone, and tumor necrosis factor, which all promote sleep in specific contexts (Richter et al., 2014).

Most neuropeptides bind to G-protein coupled receptors to mediate behavioral changes. Two notable exceptions are epidermal growth factor (EGF) and transforming growth factor alpha (TGF- α), which bind exclusively to the epidermal growth factor receptor (EGFR), a tyrosine kinase receptor. In *C. elegans*, overexpression of *lin-3*, which encodes the EGF homolog, induces a primordial sleep-like state known as lethargus (Van Buskirk and Sternberg, 2007). Similarly, genetic overexpression of EGFR ligands in *Drosophila* increases sleep (Foltenyi et al., 2007). Worms that express an EGFR hypomorph behave similarly to flies that knockdown EGFR ligand expression by RNAi: both have decreased sleep. It has also been demonstrated that EGF infused into the brains of rabbits increases NREM sleep (Kushikata et al., 1998). Taken together, studies in invertebrates and vertebrates indicate that neuropeptidergic signaling through EGFR promotes sleep behavior.

Comparative studies across different animal species have clearly made invaluable contributions to our understanding of sleep. However, further characterization and discovery of sleep regulatory mechanisms could be made using complementary model organisms, such as the zebrafish.

1.3 Zebrafish as a simple animal model of sleep

The zebrafish *Danio rerio* presents a unique platform to study sleep behavior. Zebrafish satisfy the aforementioned criteria for sleep (Zhdanova, 2006), but unlike other model organisms are prolific, diurnal vertebrates whose offspring develop externally. A single mating pair can produce hundreds of transparent embryos, which hatch into larvae after only a few days. Larvae exhibit robust sleep behavior by four days post-fertilization (Prober et al., 2006) and are sustained by a yolk sac for the first week of development, which obviates behavioral variability introduced by feeding. Pharmacological treatment is straightforward, as small molecules added to the water are rapidly absorbed through the gills and skin. Larval zebrafish are thus superbly suited for modern genetic techniques (Jao et al., 2013), *in vivo* imaging experiments (Ahrens et al., 2013; Naumann et al., 2010), and high-throughput behavioral analysis (Rihel et al., 2010a; Rihel et al., 2010b).

Importantly, zebrafish larvae possess many homologous neural structures involved in mammalian sleep (Chiu and Prober, 2013), including the noradrenergic locus ceruleus (Prober et al., 2006), histaminergic tuberomammillary nucleus (Kaslin and Panula, 2001), serotonergic dorsal raphe, dopaminergic ventral tegmental area and substantia nigra (McLean and Fetcho, 2004; Ryu et al., 2007), and the cholinergic basal forebrain (Guo et al.). The general neuroanatomy and ventricular system of larval zebrafish are summarized in Figure 1 and Figure 2, respectively, to orient the reader and provide a framework for the data presented in subsequent chapters.

In this thesis, I capitalize upon the advantages of the zebrafish model to expand our understanding of two factors previously implicated in sleep behavior: EGFR and hypocretin. In Chapter 2, I apply genetic and pharmacological techniques to show that

EGFR signaling is both necessary and sufficient for normal sleep behavior in zebrafish, just as it is in invertebrates. This demonstrates that sleep regulatory mechanisms are evolutionarily ancient, and is the first genetic study showing that EGFR signaling is required for normal sleep behavior in a vertebrate. Daniel Lee, who collaborated with me on this project, conducted the behavioral experiments presented in Chapter 2, while I generated all transgenic zebrafish lines, performed initial behavioral experiments, and generated the majority of the *in situ* hybridization data shown. In Chapter 3, I exploit the rapid external development of zebrafish embryos to screen for developmental factors that specify hypocretin neurons. I identify the LIM homeobox 9 transcription factor as necessary for hypocretin neuronal development in zebrafish and sufficient to specify additional hypocretin neurons in both zebrafish and mice. This is the first time any factor has been shown to induce hypocretin neurons *in vivo* and is an important step towards developing a cure for narcolepsy, which is caused by the selective loss of hypocretin neurons. Chapter 3 is published as “Evolutionarily conserved regulation of regulation of hypocretin neuron specification by Lhx9” in the journal *Development*. Microarray analysis, CRISPR/Cas9 injections, and mouse experiments were performed by my advisor, David Prober, and our collaborators Florian Merkle and Ian Woods; James Gagnon; and Tomomi Shimogori, respectively. Finally, in Chapter 4, I summarize the findings and implications of previous chapters and present avenues for future research.

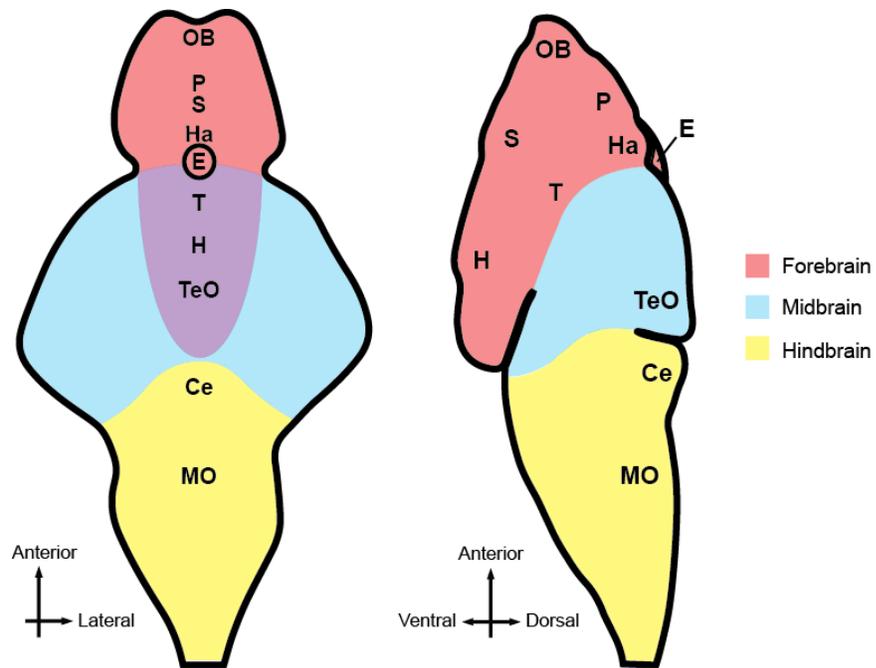


Figure 1. Zebrafish neuroanatomy at 5 days post fertilization. A simple schematic of a 5 day-old larval zebrafish brain viewed dorsally is shown on the left, while a side view is presented on the right. Abbreviations of neural structures: OB, olfactory bulb; P, pallium; S, subpallium; Ha, habenula; E, epiphysis (pineal gland); T, thalamus; H, hypothalamus; TeO, optic tectum, Ce, cerebellum; MO, medulla oblongata.

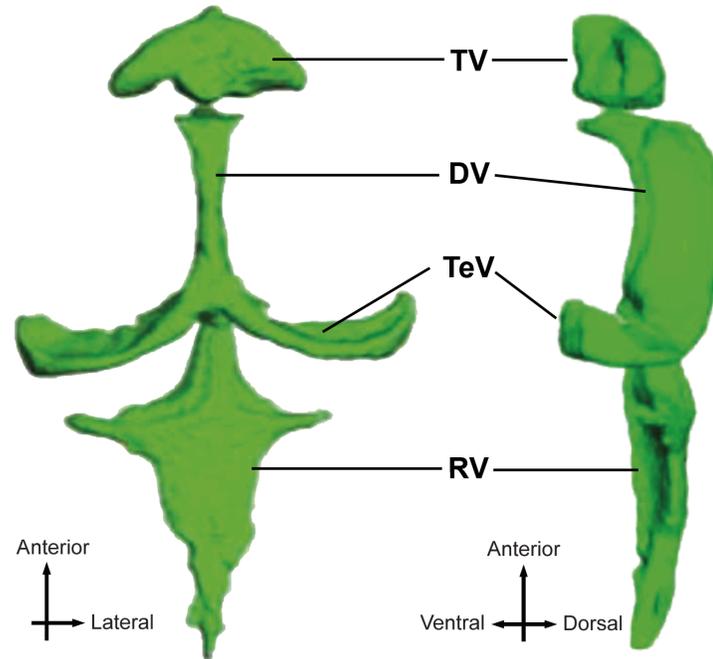


Figure 2. Ventricular system of zebrafish larvae at 6 days post fertilization. A 3D rendering of the cerebral ventricles in a 6 day-old larval zebrafish is shown on the left from a dorsal view, while a side view is presented on the right. Abbreviations: TV, telencephalic ventricle; DV, diencephalic ventricle; TeV, tectal ventricle; RV, rhombencephalic ventricle. Figure adapted from Turner et al., 2012.

References

- Adamantidis, A. R., Zhang, F., Aravanis, A. M., Deisseroth, K. and de Lecea, L. (2007). Neural substrates of awakening probed with optogenetic control of hypocretin neurons. *Nature* 450, 420–424.
- Ahrens, M. B., Orger, M. B., Robson, D. N., Li, J. M. and Keller, P. J. (2013). Whole-brain functional imaging at cellular resolution using light-sheet microscopy. *Nat Meth* 10, 413–420.
- Campbell, S. S. and Tobler, I. (1984). Animal sleep: a review of sleep duration across phylogeny. *Neurosci Biobehav Rev* 8, 269–300.
- Chiu, C. N. and Prober, D. A. (2013). Regulation of zebrafish sleep and arousal states: current and prospective approaches. *Front Neural Circuits* 7, 58.
- Cirelli, C. (2009). The genetic and molecular regulation of sleep: from fruit flies to humans. *Nat. Rev. Neurosci.* 10, 549–560.
- Cirelli, C., Bushey, D., Hill, S., Huber, R., Kreber, R., Ganetzky, B. and Tononi, G. (2005). Reduced sleep in *Drosophila* Shaker mutants. *Nature* 434, 1087–1092.
- Dauvilliers, Y., Arnulf, I. and Mignot, E. (2007). Narcolepsy with cataplexy. *The Lancet* 369, 499–511.
- Espinosa, F., Marks, G., Heintz, N. and Joho, R. H. (2004). Increased motor drive and sleep loss in mice lacking Kv3-type potassium channels. *Genes Brain Behav* 3, 90–100.
- 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*. *Nature Neuroscience* 10, 1160–1167.
- Guo, S., Wilson, S. W., Cooke, S., Chitnis, A. B., Driever, W. and Rosenthal, A. Mutations in the Zebrafish Unmask Shared Regulatory Pathways Controlling the Development of Catecholaminergic Neurons. *Developmental Biology* 208, 473–487.
- Imeri, L. and Opp, M. R. (2009). How (and why) the immune system makes us sleep. *Nat. Rev. Neurosci.* 10, 199–210.
- Jao, L.-E., Wentz, S. R. and Chen, W. (2013). Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proc. Natl. Acad. Sci. U.S.A.* 110, 13904–13909.
- Kaslin, J. and Panula, P. (2001). Comparative anatomy of the histaminergic and other aminergic systems in zebrafish (*Danio rerio*). *J. Comp. Neurol.* 440, 342–377.
- Kloss, B., Price, J. L., Saez, L., Blau, J., Rothenfluh, A., Wesley, C. S. and Young, M. W.

- (1998). The *Drosophila* clock gene double-time encodes a protein closely related to human casein kinase Iepsilon. *Cell* 94, 97–107.
- Konopka, R. J. and Benzer, S. (1971). Clock mutants of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U.S.A.* 68, 2112–2116.
- 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, R509–14.
- McLean, D. L. and Fetcho, J. R. (2004). Ontogeny and innervation patterns of dopaminergic, noradrenergic, and serotonergic neurons in larval zebrafish. *J. Comp. Neurol.* 480, 38–56.
- Naumann, E. A., Kampff, A. R., Prober, D. A., Schier, A. F. and Engert, F. (2010). Monitoring neural activity with bioluminescence during natural behavior. *Nature Publishing Group* 13, 513–520.
- Porkka-Heiskanen, T. and Kalinchuk, A. V. (2011). Sleep Medicine Reviews. *Sleep Medicine Reviews* 15, 123–135.
- 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. *Journal of Neuroscience* 26, 13400–13410.
- Richter, C., Woods, I. G. and Schier, A. F. (2014). Neuropeptidergic control of sleep and wakefulness. *Annu. Rev. Neurosci.* 37, 503–531.
- Rihel, J., Prober, D. A. and Schier, A. F. (2010a). Monitoring sleep and arousal in zebrafish. *Methods Cell Biol.* 100, 281–294.
- Rihel, J., Prober, D. A., Arvanites, A., Lam, K., Zimmerman, S., Jang, S., Haggarty, S. J., Kokel, D., Rubin, L. L., Peterson, R. T., et al. (2010b). Zebrafish behavioral profiling links drugs to biological targets and rest/wake regulation. *Science* 327, 348–351.
- Ryu, S., Mahler, J., Acampora, D., Holzschuh, J., Erhardt, S., Omodei, D., Simeone, A. and Driever, W. (2007). Orthopedia homeodomain protein is essential for diencephalic dopaminergic neuron development. *Curr. Biol.* 17, 873–880.
- Saper, C. B., Scammell, T. E. and Lu, J. (2005). Hypothalamic regulation of sleep and circadian rhythms. *Nature* 437, 1257–1263.
- Siegel, J. M. (2008). Do all animals sleep? *Trends in Neurosciences* 31, 208–213.
- Tononi, G. and Cirelli, C. (2014). Perspective. *Neuron* 81, 12–34.
- Tsujino, N. and Sakurai, T. (2009). Orexin/Hypocretin: A Neuropeptide at the Interface of Sleep, Energy Homeostasis, and Reward System. *Pharmacological Reviews* 61, 162–176.

Turner, M. H., Ullmann, J. F. P. and Kay, A. R. (2012). A method for detecting molecular transport within the cerebral ventricles of live zebrafish (*Danio rerio*) larvae. *The Journal of Physiology* 590, 2233–2240.

Van Buskirk, C. and Sternberg, P. W. (2007). Epidermal growth factor signaling induces behavioral quiescence in *Caenorhabditis elegans*. *Nature Neuroscience* 10, 1300–1307.

Xie, L., Kang, H., Xu, Q., Chen, M. J., Liao, Y., Thiyagarajan, M., O'Donnell, J., Christensen, D. J., Nicholson, C., Iliff, J. J., et al. (2013). Sleep Drives Metabolite Clearance from the Adult Brain. *Science* 342, 373–377.

Xu, Y., Padiath, Q. S., Shapiro, R. E., Jones, C. R., Wu, S. C., Saigoh, N., Saigoh, K., Ptacek, L. J. and Fu, Y.-H. (2005). Functional consequences of a CK1delta mutation causing familial advanced sleep phase syndrome. *Nature* 434, 640–644.

Zhdanova, I. V. (2006). Sleep in zebrafish. *Zebrafish* 3, 215–226.

CHAPTER 2:

Epidermal Growth Factor Receptor Signaling

Promotes Sleep in Zebrafish

2.1 Abstract

Epidermal growth factor receptor (EGFR) has many roles in the nervous system, including differentiation, proliferation, and maintenance of both neurons and glia. Genetic studies in invertebrates have also shown that EGFR signaling is necessary and sufficient for normal sleep behavior, though comparable studies in mammals have been inconclusive. We examined the EGFR system in zebrafish and found that overexpression of transforming growth factor alpha (TGF- α), an EGFR ligand, increases sleep in zebrafish larvae and activates EGFR+ cells along the brain ventricle. In contrast, TGF- α null mutants or wild-type larvae treated with an EGFR antagonist have reduced sleep. TGF- α -induced sleep is light dependent, and persists in the absence of overt circadian rhythms. We conclude that EGFR signaling is necessary and sufficient for normal sleep behavior in zebrafish. The evolutionary conservation of EGFR-mediated quiescence across phyla suggests an ancient origin for sleep and demonstrates the utility of simple genetic model organisms in uncovering the fundamental mechanisms that may underlie all sleep behavior.

2.2 Introduction

Sleep behavior is evolutionarily conserved across diverse animal species, suggesting that it performs some highly adaptive function (Siegel, 2009), or may be outright necessary for survival (Rechtschaffen et al., 1983; Shaw et al., 2002). The most parsimonious hypothesis is that the mechanisms that drive sleep behavior are also conserved, but only a handful of genes have been found to regulate sleep in both invertebrates and vertebrates (Cirelli, 2009; Crocker and Sehgal, 2010). Most of these genes encode for components of the circadian clock, neurotransmitters, or ion channels that modulate neuronal excitability. An exception is the epidermal growth factor receptor (EGFR), a tyrosine kinase with numerous roles in the nervous system, including neuronal and glial differentiation, proliferation, and maintenance (Buonanno and Fischbach, 2001; Plata-Salamán, 1991). Genetic studies in *C. elegans* and *D. melanogaster* have shown that EGFR signaling is both necessary and sufficient for normal sleep behavior (Van Buskirk and Sternberg, 2007; Foltenyi et al., 2007). There is also evidence that EGFR signaling is sufficient to induce sleep in mammals (Kushikata et al., 1998). It remains unclear, however, whether EGFR is required for normal sleep behavior in vertebrates (Kramer et al., 2001). To elucidate the role of EGFR in sleep, we turned to the zebrafish, a prolific, diurnal vertebrate that develops externally from parents into optically transparent larvae. Zebrafish are particularly amenable to modern genome-editing techniques (Huang et al., 2011; Hwang et al., 2013; Jao et al., 2013) and pharmacological treatments (Rihel et al., 2010b), which facilitated our interrogation of the EGFR system.

2.3 TGF- α overexpression increases sleep

To test whether EGFR signaling affects sleep in zebrafish, we overexpressed the zebrafish ortholog of transforming growth factor alpha (TGF- α), an EGFR ligand, using a heat shock-inducible transgene: *Tg(hs:tgfa)*. We monitored the sleep/wake behavior of heterozygous *Tg(hs:tgfa)* larvae and their wild-type (WT) siblings using a high-throughput videotracking assay (Prober et al., 2006). Prior to TGF- α overexpression, we did not observe any difference between *Tg(hs:tgfa)* and WT siblings. However, after heat shock, *Tg(hs:tgfa)* larvae exhibited a dramatic suppression of locomotor activity during the daytime (Fig 1A,C) and a commensurate increase in daytime and nighttime sleep (Fig 1B,D). The increase in sleep was due to changes in sleep architecture: TGF- α overexpression lengthened daytime sleep bout duration by 28-43% compared to WT controls (Table 1). However, the effect on sleep bout frequency was less clear. *Tg(hs:tgfa)* larvae had more frequent sleep bouts than WT siblings on day 5 post-heat shock, but a comparable number or even less frequent sleep bouts on subsequent days and nights (Table 1).

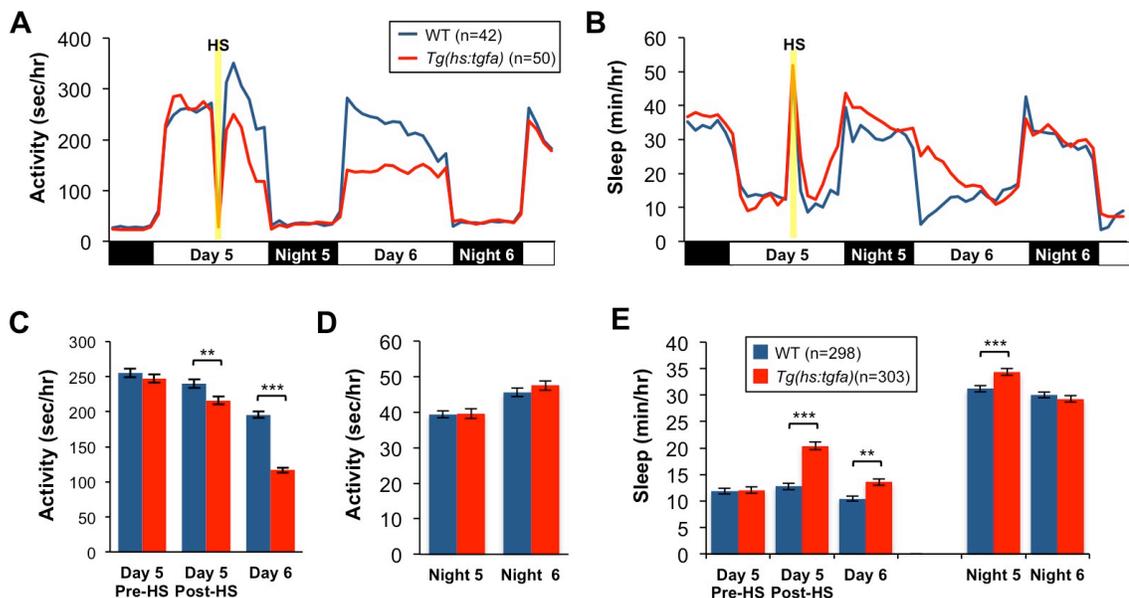


Figure 1. TGF- α overexpression decreases locomotor activity and increases sleep. (A) Compared to their WT siblings, *Tg(hs:tgfa)* larvae have significantly reduced locomotor activity after heat shock on day 5 and day 6. (B) *Tg(hs:tgfa)* larvae have a corresponding increase in sleep on day 5 post-heat shock, night 5, and day 6. Data from a single representative experiment are shown in (A-B), while the combined data from seven experiments are quantified in (C-E). Bar graphs indicate mean \pm SEM. n = number of larvae analyzed. ***, $p \leq 0.001$; **, $p \leq 0.01$ compared to WT larvae by Mann-Whitney U-test.

2.4 Disruption of EGFR signaling decreases sleep

We also generated *tgfa*^{-/-} zebrafish with a predicted null mutation in the TGF- α gene (Fig. S1) and found they had the opposite sleep phenotype to that of *Tg(hs:tgfa)* larvae. Compared to wild-type *tgfa*^{+/+} siblings, *tgfa*^{-/-} mutant larvae had increased daytime locomotor activity (Fig. 2A,C) and decreased daytime sleep (Fig. 2B,E). However, we did not detect a difference between genotypes on day 5. This observation could be due to developmental compensation in *tgfa*^{-/-} mutants, or to functional redundancy between TGF- α and other EGFR ligands, such as epidermal growth factor (EGF). To avoid these potential confounds, we explored the numerous EGFR antagonists developed as therapeutics for EGFR-mediated cancers. One antagonist, gefitinib, is a selective EGFR inhibitor with a long metabolic half-life in rodents (Barker et al., 2001). We found that WT larvae treated with 5 μ M gefitinib at 4 days post fertilization (dpf) had significantly increased locomotor activity (Fig. 3A,C,D) and decreased sleep (Fig. 3B,E) compared to DMSO-treated controls during both day and night. The daytime sleep reduction in *tgfa*^{-/-} and gefitinib-treated larvae was primarily a result of decreased sleep bout frequency (Table 1). In contrast, the nighttime sleep reduction in gefitinib-treated larvae was mostly due to decreased sleep bout duration (Table 1).

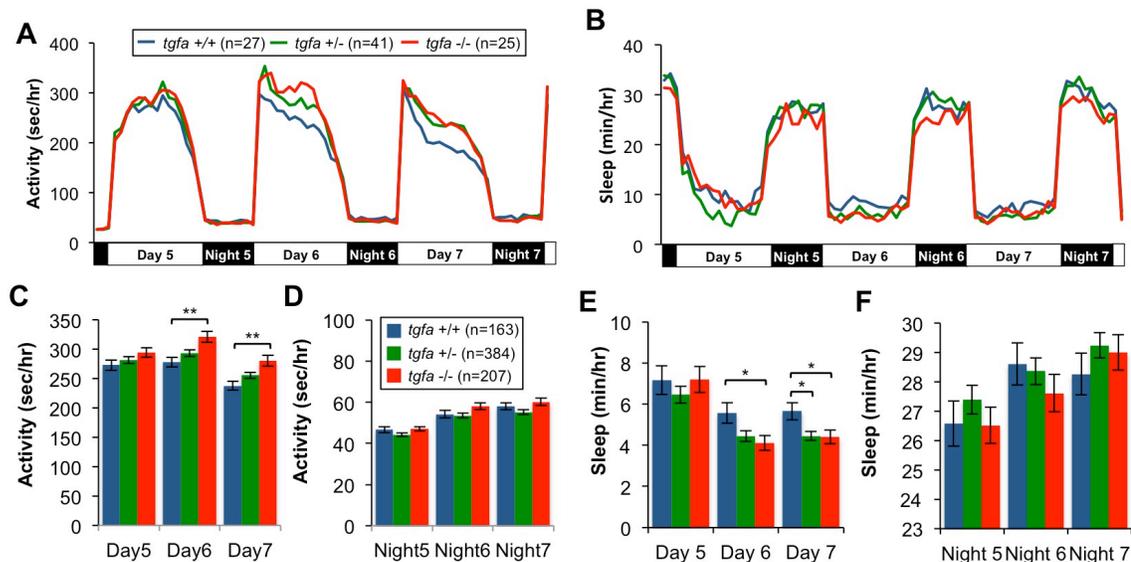


Figure 2. TGF α loss-of-function mutants have increased daytime activity and decreased sleep. (A-B) *tgfa -/-* larvae have increased locomotor activity and decreased sleep on day 6 and 7 compared to *tgfa +/+* siblings. Data from a single representative experiment are shown in (A-B), while the combined data from eight experiments are quantified in (C-F). No difference was observed between genotypes on day 5, possibly due to redundancy between TGF- α and other EGFR ligands at this developmental stage. Furthermore, the *tgfa* gene may be haploinsufficient; *tgfa +/-* larvae have significantly reduced sleep on day 7 (E) and show intermediate levels of locomotor activity compared to *tgfa -/-* and *tgfa +/+* siblings on day 6 and 7 (A, C), though this trend was not statistically significant. Bar graphs indicate mean \pm SEM. **, $p \leq 0.01$; *, $p \leq 0.05$ compared to *tgfa +/+* larvae by one-way ANOVA and Bonferroni's post hoc tests.

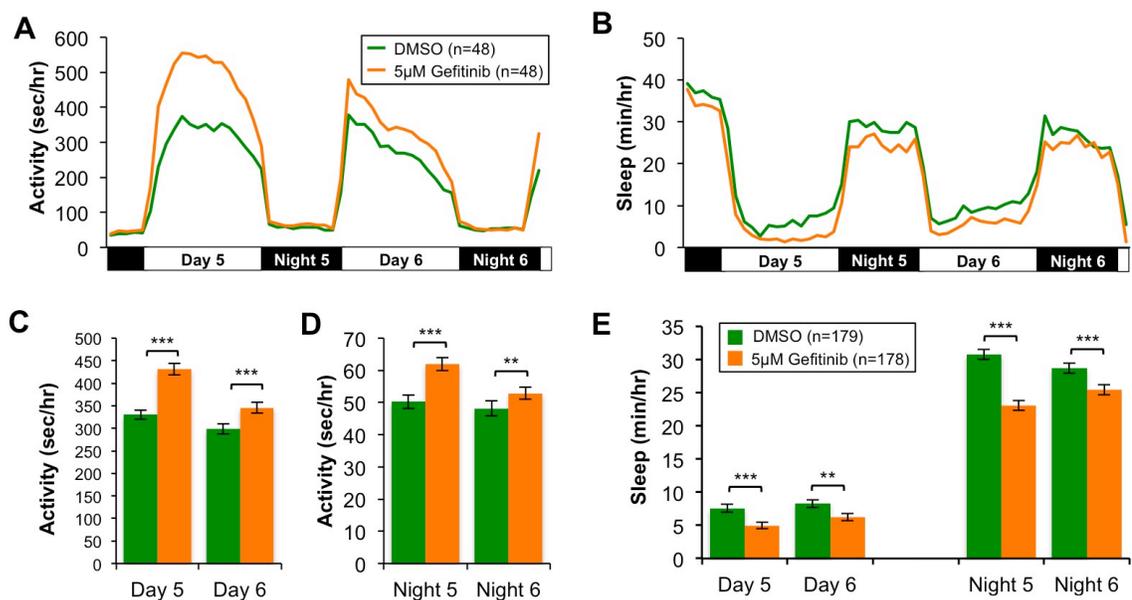


Figure 3. EGFR antagonist treatment increases activity and decreases sleep. (A-B) WT larvae treated with 5 μ M of the EGFR antagonist gefitinib had increased locomotor activity and decreased sleep on both days and nights tested, compared to larvae treated with DMSO. Data from a single representative experiment are shown in (A-B), while the combined data from six experiments are quantified in (C-E). Bar graphs indicate mean \pm SEM. ***, $p \leq 0.001$; **, $p \leq 0.01$ compared to DMSO-treated larvae by Mann-Whitney U-test.

Table 1. Sleep architecture in TGF- α GOF/LOF and gefitinib-treated larvae

		Sleep bout frequency (bouts/hr)		Sleep bout duration (min/bout)	
<i>Day 5 Pre-Heat Shock</i>		Significance		Significance	
Wild type (n=298)	2.75 \pm 0.111		Wild type	4.27 \pm 0.141	
<i>Tg(hs:tgfa)</i> (n=303)	2.80 \pm 0.107	n.s.	<i>Tg(hs:tgfa)</i>	4.28 \pm 0.139	n.s.
<i>Day 5 Post-Heat Shock</i>		Significance		Significance	
Wild type	3.26 \pm 0.130		Wild type	3.37 \pm 0.102	
<i>Tg(hs:tgfa)</i>	4.34 \pm 0.115	***	<i>Tg(hs:tgfa)</i>	4.30 \pm 0.116	***
<i>Day 6</i>		Significance		Significance	
Wild type	4.12 \pm 0.125		Wild type	2.37 \pm 0.060	
<i>Tg(hs:tgfa)</i>	3.94 \pm 0.143	n.s.	<i>Tg(hs:tgfa)</i>	3.39 \pm 0.100	***
<i>Night 5</i>		Significance		Significance	
Wild type	7.27 \pm 0.093		Wild type	4.53 \pm 0.133	
<i>Tg(hs:tgfa)</i>	7.39 \pm 0.097	n.s.	<i>Tg(hs:tgfa)</i>	4.94 \pm 0.140	*
<i>Night 6</i>		Significance		Significance	
Wild type	7.42 \pm 0.102		Wild type	4.36 \pm 0.194	
<i>Tg(hs:tgfa)</i>	7.04 \pm 0.111	**	<i>Tg(hs:tgfa)</i>	4.29 \pm 0.113	n.s.
<i>Day 5</i>		Significance		Significance	
<i>tgfa</i> +/+ (n=163)	2.18 \pm 0.178		<i>tgfa</i> +/+	2.93 \pm 0.124	
<i>tgfa</i> +/- (n=384)	1.88 \pm 0.101	n.s.	<i>tgfa</i> +/-	3.19 \pm 0.118	n.s.
<i>tgfa</i> -/- (n=207)	1.86 \pm 0.130	n.s.	<i>tgfa</i> -/-	3.58 \pm 0.175	*
<i>Day 6</i>		Significance		Significance	
<i>tgfa</i> +/+	2.38 \pm 0.177		<i>tgfa</i> +/+	2.17 \pm 0.098	
<i>tgfa</i> +/-	1.95 \pm 0.097	*	<i>tgfa</i> +/-	2.13 \pm 0.077	n.s.
<i>tgfa</i> -/-	1.68 \pm 0.112	**	<i>tgfa</i> -/-	2.45 \pm 0.164	n.s.
<i>Day 7</i>		Significance		Significance	
<i>tgfa</i> +/+	2.83 \pm 0.179		<i>tgfa</i> +/+	1.92 \pm 0.078	
<i>tgfa</i> +/-	2.24 \pm 0.098	*	<i>tgfa</i> +/-	1.90 \pm 0.050	n.s.
<i>tgfa</i> -/-	2.05 \pm 0.132	**	<i>tgfa</i> -/-	2.30 \pm 0.118	*
<i>Night 5</i>		Significance		Significance	
<i>tgfa</i> +/+	7.54 \pm 0.155		<i>tgfa</i> +/+	3.51 \pm 0.084	
<i>tgfa</i> +/-	7.48 \pm 0.092	n.s.	<i>tgfa</i> +/-	3.67 \pm 0.059	n.s.
<i>tgfa</i> -/-	7.34 \pm 0.131	n.s.	<i>tgfa</i> -/-	3.67 \pm 0.082	n.s.
<i>Night 6</i>		Significance		Significance	
<i>tgfa</i> +/+	7.08 \pm 0.129		<i>tgfa</i> +/+	4.05 \pm 0.094	
<i>tgfa</i> +/-	6.93 \pm 0.080	n.s.	<i>tgfa</i> +/-	4.15 \pm 0.068	n.s.
<i>tgfa</i> -/-	6.93 \pm 0.111	n.s.	<i>tgfa</i> -/-	4.02 \pm 0.086	n.s.
<i>Night 7</i>		Significance		Significance	
<i>tgfa</i> +/+	6.73 \pm 0.114		<i>tgfa</i> +/+	4.21 \pm 0.100	
<i>tgfa</i> +/-	6.81 \pm 0.071	n.s.	<i>tgfa</i> +/-	4.36 \pm 0.065	n.s.
<i>tgfa</i> -/-	6.88 \pm 0.106	n.s.	<i>tgfa</i> -/-	4.27 \pm 0.084	n.s.
<i>Day 5</i>		Significance		Significance	
DMSO (n=179)	2.12 \pm 0.159		DMSO	3.76 \pm 0.162	
Gefitinib (n=178)	1.40 \pm 0.136	***	Gefitinib	3.41 \pm 0.151	n.s.
<i>Day 6</i>		Significance		Significance	
DMSO	3.00 \pm 0.193		DMSO	2.82 \pm 0.101	
Gefitinib	2.30 \pm 0.180	***	Gefitinib	2.83 \pm 0.109	n.s.
<i>Night 5</i>		Significance		Significance	
DMSO	6.54 \pm 0.158		DMSO	4.93 \pm 0.123	
Gefitinib	6.15 \pm 0.185	*	Gefitinib	3.89 \pm 0.101	***
<i>Night 6</i>		Significance		Significance	
DMSO	6.76 \pm 0.173		DMSO	4.53 \pm 0.206	
Gefitinib	6.96 \pm 0.186	n.s.	Gefitinib	3.74 \pm 0.106	***

n.s. not significant, * p \leq 0.05, ** p \leq 0.01, *** p \leq 0.001

2.5 EGFR signaling modulates arousal state

To test whether the quiescence observed in *Tg(hs:tgfa)* larvae after heat shock satisfies behavioral criteria for sleep (Campbell and Tobler, 1984; Zhdanova, 2006), we delivered a mechanoacoustic tapping stimuli to *Tg(hs:tgfa)* larvae and WT siblings to measure sensory responsiveness. *Tg(hs:tgfa)* were significantly less responsive than WT siblings at all tapping intensities (Fig. 4A), while gefitinib-treated WT larvae showed the opposite phenotype compared to DMSO-treated controls (Fig. 4B). These results suggest that EGFR inhibition induces a generally heightened arousal state in zebrafish larvae, while activation of the EGFR signaling pathway leads to a decreased arousal state.

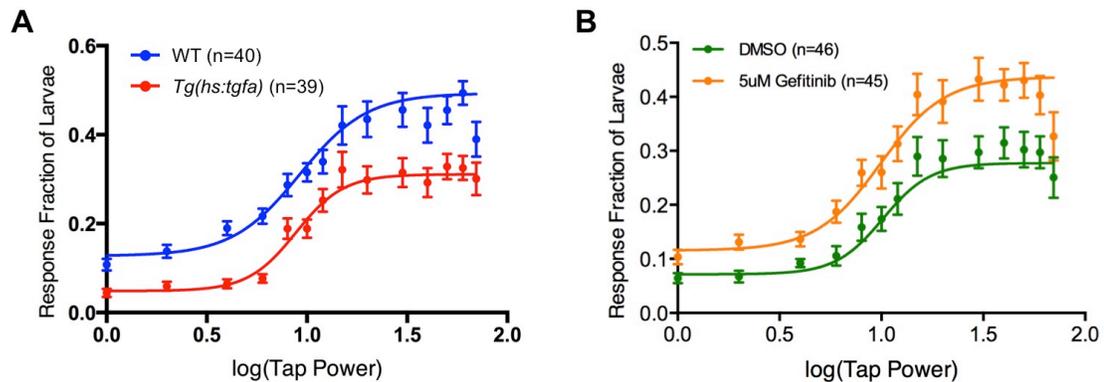


Figure 4. EGFR signaling modulates larval arousal state. (A) *Tg(hs:tgfa)* larvae were less responsive to mechanoacoustic tapping stimuli than their WT siblings after heat shock, regardless of tap power. Each data point indicates the mean response probability \pm SEM at a particular stimulus intensity. (B) In contrast, WT larvae treated with gefitinib were more responsive than DMSO-treated controls to tapping stimuli.

2.6 EGFR inhibition suppresses TGF- α -mediated sleep

Mammalian TGF- α binds exclusively to EGFR but not to other members of the EGFR tyrosine kinase family (Harris et al., 2003). We therefore hypothesized that gefitinib treatment would block the effects of TGF- α overexpression. Indeed, *Tg(hs:tgfa)* larvae treated with gefitinib were significantly more active after heat shock than DMSO-treated *Tg(hs:tgfa)* siblings (Fig. 5A,C). However gefitinib-treated *Tg(hs:tgfa)* larvae still slept significantly more than gefitinib-treated WT controls (Fig. 5B,D), perhaps because gefitinib, which binds reversibly to the EGFR ATP binding site (Ward et al., 1994; Barker et al., 2001), did not inhibit EGFR completely. We conclude that gefitinib treatment can suppress TGF- α -mediated sleep. This result suggests that TGF- α overexpression acts through EGFR, but should be corroborated with additional evidence, such as testing the *Tg(hs:tgfa)* transgene in an EGFR mutant background.

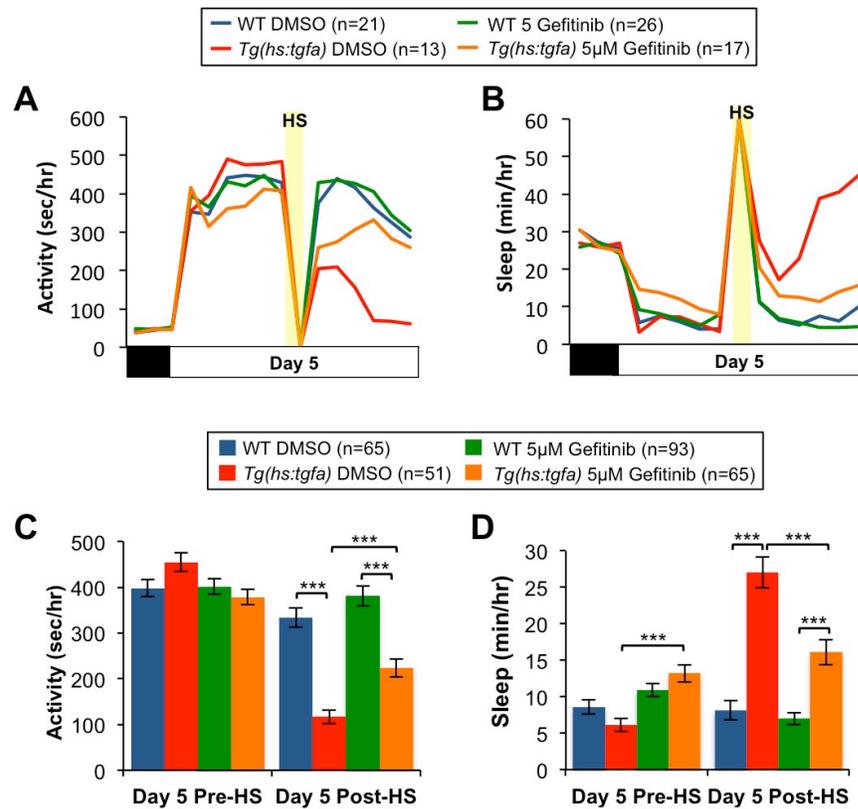


Figure 5. EGFR antagonist treatment suppresses TGF- α -mediated sleep. (A) After heat shock, gefitinib-treated *Tg(hs:tgfa)* larvae were significantly more active than DMSO-treated *Tg(hs:tgfa)* larvae, but still less active than gefitinib-treated WT siblings. (B) Gefitinib treatment also suppressed TGF- α -mediated sleep. Gefitinib was added at the start of the experiment, prior to heat shock. Data from a single representative experiment are shown in (A-B), while the combined data from four experiments are quantified in (C-E). Bar graphs indicate mean \pm SEM. ***, $p \leq 0.001$ compared to gefitinib-treated WT larvae or DMSO-treated *Tg(hs:tgfa)* larvae by two-way ANOVA followed by Tukey's post hoc tests.

2.7 TGF- α -mediated sleep requires the MAPK/ERK pathway

EGFR can interact with multiple signal transduction pathways, but genetic studies in *C. elegans* and *D. melanogaster* provide differing reports about which pathway is associated with sleep; lethargus in *C. elegans* is mediated by the phospholipase C gamma transduction pathway (Van Buskirk and Sternberg, 2007), while sleep in *D. melanogaster* requires the MAPK/ERK pathway (Foltényi et al., 2007). To test whether the MAPK/ERK pathway is required for TGF- α -mediated sleep in zebrafish, we acutely treated *Tg(hs:tgfa)* larvae with two MEK1/2 antagonists previously demonstrated to be effective *in vivo* (Hong et al., 2006). *Tg(hs:tgfa)* larvae treated with either 3 μ M SL327 (Fig. 6A-D) or 15 μ M U0126 (Fig. 6E-H) were significantly more active after heat shock than DMSO-treated *Tg(hs:tgfa)* larvae, though less active than their drug-treated WT siblings. SL327 treatment significantly reduced both sleep bout frequency and duration in *Tg(hs:tgfa)* larvae, while U0126 treatment only reduced sleep bout frequency (Table 2), possibly because SL327 had a more potent effect on sleep behavior than U0126 overall (Fig. 6D,H). Since both MEK1/2 antagonists suppressed the effect of TGF- α overexpression, we propose that TGF- α -mediated sleep acts, at least partially, through the MAPK/ERK pathway in zebrafish.

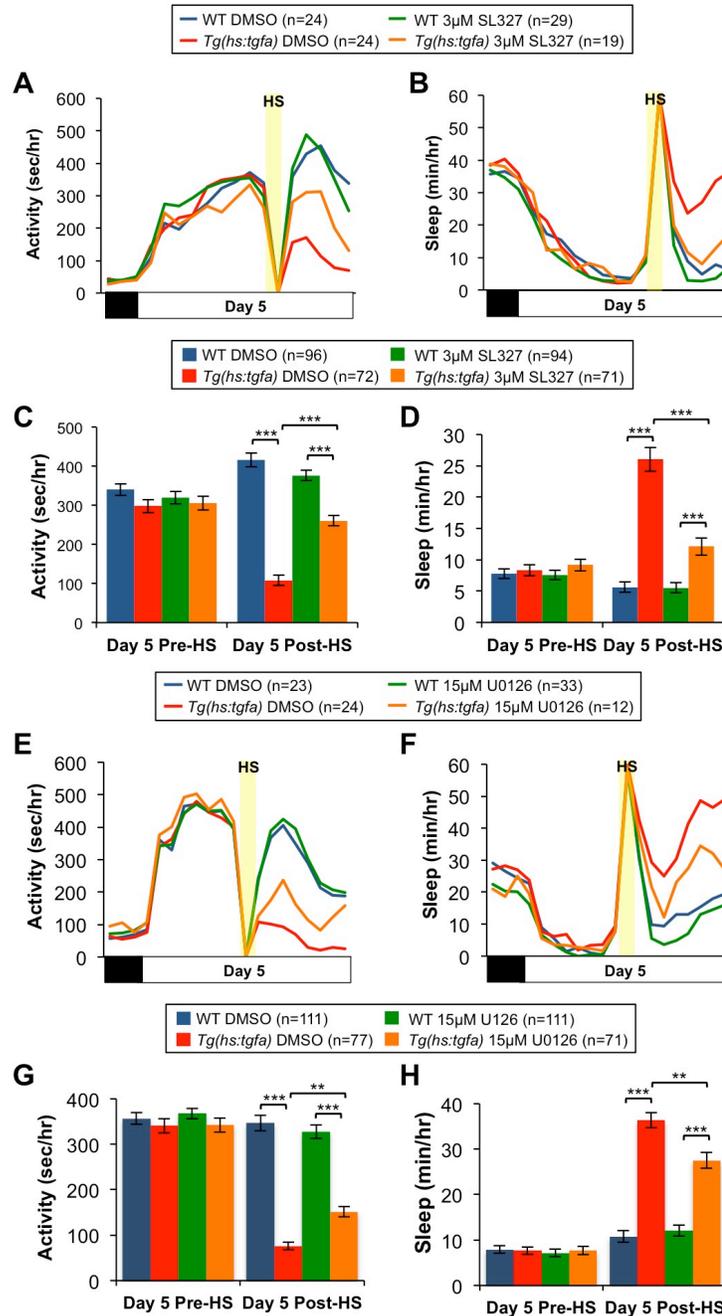


Figure 6. Acute pharmacological inhibition of the MAPK/ERK pathway suppresses TGF- α -mediated sleep. (A) After heat shock, *Tg(hs:tgfa)* larvae treated with 3 μ M of the MEK1/2 antagonist SL327 were significantly more active than DMSO-treated *Tg(hs:tgfa)* larvae, but still less active than SL327-treated WT siblings. (B) SL327 treatment also suppressed TGF- α -mediated sleep. (E-F) A second MEK1/2 antagonist, U0126, had a similar, albeit weaker, effect on locomotor activity and sleep. To minimize toxicity effects, drugs were administered immediately after heat shock. Data from a single representative experiment are shown in (A-B) and (E-F), while the combined data from four experiments are quantified in (C-D) and (G-H), respectively. Bar graphs indicate mean \pm SEM. ***, $p < 0.001$; **, $p < 0.01$ compared to MEK1/2 antagonist-treated WT larvae or DMSO-treated *Tg(hs:tgfa)* larvae by two-way ANOVA followed by Tukey's post hoc tests.

Table 2. Sleep architecture in *Tg(hs:tgfa)* larvae treated with MEK1/2 antagonists

Sleep bout frequency (bouts/hr)			Sleep bout duration (min/bout)		
<i>Day 5 Pre-Heat Shock</i>			<i>Day 5 Pre-Heat Shock</i>		
Wild type DMSO (n=96)	2.35±0.221]n.s.]	Wild type DMSO	3.83±0.409]n.s.]
<i>Tg(hs:tgfa)</i> DMSO (n=72)	2.34±0.220		<i>Tg(hs:tgfa)</i> DMSO	3.47±0.219	
Wild type 3µM SL327 (n=94)	2.32±0.198]n.s.]	Wild type 3µM SL327	3.43±0.227]n.s.]
<i>Tg(hs:tgfa)</i> 3µM SL327 (n=71)	2.78±0.311		<i>Tg(hs:tgfa)</i> 3µM SL327	3.60±0.204	
<i>Day 5 Post-Heat Shock</i>			<i>Day 5 Post-Heat Shock</i>		
Wild type DMSO	1.54±0.225]***]	Wild type DMSO	4.02±0.316]***]
<i>Tg(hs:tgfa)</i> DMSO	4.41±0.304		<i>Tg(hs:tgfa)</i> DMSO	6.37±0.489	
Wild type 3µM SL327	1.32±0.142]***]	Wild type 3µM SL327	4.49±0.494]n.s.]
<i>Tg(hs:tgfa)</i> 3µM SL327	2.88±0.267		<i>Tg(hs:tgfa)</i> 3µM SL327	3.80±0.249	
<i>Day 5 Pre-Heat Shock</i>			<i>Day 5 Pre-Heat Shock</i>		
Wild type DMSO (n=96)	1.97 ±0.167]n.s.]	Wild type DMSO	4.47 ±0.365]n.s.]
<i>Tg(hs:tgfa)</i> DMSO (n=72)	2.07 ±0.220		<i>Tg(hs:tgfa)</i> DMSO	4.28 ±0.289	
Wild type 15µM U0126 (n=94)	1.54 ±0.132]n.s.]	Wild type 15µM U0126	4.60±0.294]n.s.]
<i>Tg(hs:tgfa)</i> 15µM u0126 (n=71)	1.75 ±0.204		<i>Tg(hs:tgfa)</i> 15µM U0126	5.14 ±0.565	
<i>Day 5 Post-Heat Shock</i>			<i>Day 5 Post-Heat Shock</i>		
Wild type DMSO	2.10±0.221]***]	Wild type DMSO	4.85 ±0.290]***]
<i>Tg(hs:tgfa)</i> DMSO	5.45±0.250		<i>Tg(hs:tgfa)</i> DMSO	8.05 ±0.786	
Wild type 15µM U0126	2.16±0.205]***]	Wild type 15µM U0126	5.58 ±0.384]n.s.]
<i>Tg(hs:tgfa)</i> 15µM U0126	4.31 ±0.267		<i>Tg(hs:tgfa)</i> 15µM U0126	7.01 ±0.523	

n.s. not significant, * p≤0.05, ** p≤0.01, *** p≤0.001

2.8 TGF- α overexpression activates EGFR+ cells along the brain ventricle

We next characterized the zebrafish EGFR system using *in situ* hybridization (ISH) to elucidate the cellular basis of TGF- α -mediated sleep. At 5 dpf, *tgfa* is natively expressed in bilateral clusters in the midbrain (Fig. 7A,C arrowheads) and in cells along the diencephalic ventricle (Fig. 7B-C, arrows). *Tgfa* expression was also detected in the pineal gland (Fig. 8C-D). The bilateral *tgfa*+ cell clusters co-express *vglut2a* (Fig. 8A-A''), and are therefore likely glutamatergic in nature. In contrast, the ventricular *tgfa*+ cells did not express glutamatergic markers *vglut1*, *vglut2a*, or *vglut2b* (data not shown), but did express *sox2* (Fig. 8B-B''), a gene involved in cellular proliferation and differentiation. We detected little or no expression of the GABAergic markers *gad65* or *gad67* in any *tgfa*+ cells (data not shown).

Concurrently, *egfra* is expressed in bilateral clusters in the medial hindbrain (Fig. 7D-E, arrowheads), and along the entire brain ventricle, though expression is strongest near the rhombencephalic (fourth) ventricle (Fig. 7D-E, arrows). We were unable to detect either glutamatergic or GABAergic markers in *egfra*+ cells (Fig. 9A-D''). The *egfra*+ bilateral clusters also did not express *sox2*, *glyt2*, or glial markers *gfap* and *vim* (data not shown). However, *egfra*+ cells along the rhombencephalic ventricle expressed *sox2* (Fig. 9E-E''). Given their anatomical position and molecular profile, the *tgfa*+ and *egfra*+ cells lining the ventricle are most likely ependymal cells, which have been shown to express a *Sox2-EGFP* reporter in mice (Lee et al., 2013; Suh et al., 2007).

To determine if *egfra*+ cells are activated after TGF- α overexpression, we fixed both *Tg(hs:tgfa)* larvae and WT siblings after heat shock and examined *cfos* expression by ISH. While WT siblings showed little or no *cfos* signal (Fig. 10A), *Tg(hs:tgfa)* larvae

expressed *cfos* along the brain ventricle (Fig. 10B-C) in a pattern reminiscent of *egfra* expression (Fig. 7D, arrow). Using double fluorescent ISH, we confirmed that TGF- α overexpression induces *cfos* expression in *egfra*⁺ cells (Fig. 10D-D’’’). Taken together, these results indicate that TGF- α overexpression induces sleep by activating EGFR, which is expressed on putative ependymal cells along the brain ventricle.

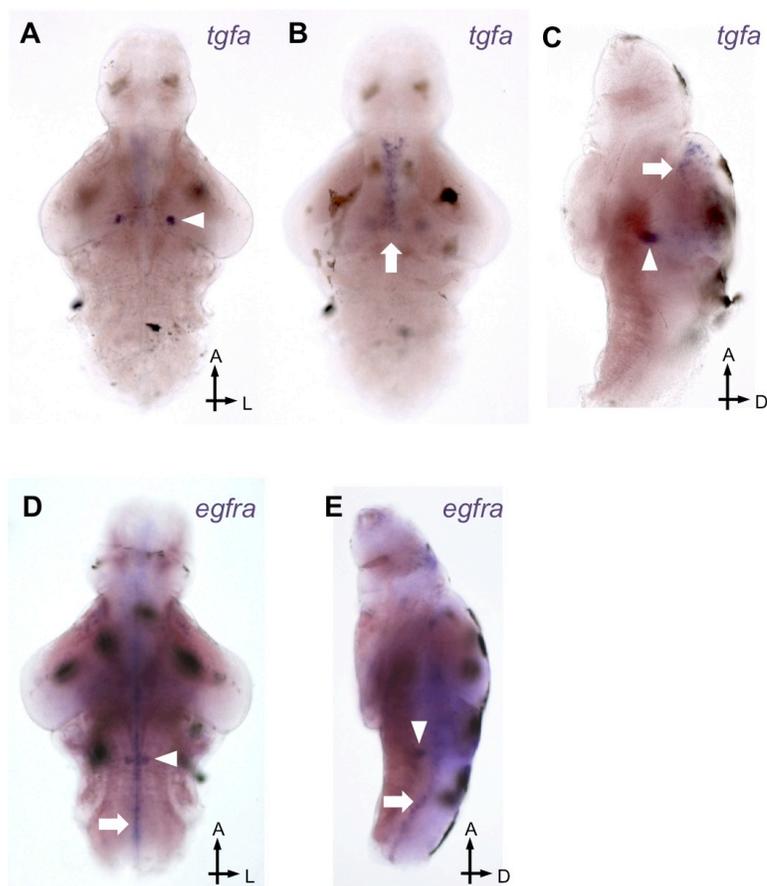


Figure 7. Wild-type *tgfa* and *egfra* expression at 5 days post fertilization. (A-C) *tgfa* is expressed in bilateral clusters in the midbrain (arrowhead), along the diencephalic ventricle (arrow), and in the pineal gland (see Fig. 7C-D). (D-E) *egfra* is expressed in bilateral clusters in the medial hindbrain (arrowhead) and along the entire brain ventricle, though expression is strongest near the rhombencephalic ventricle (arrow). Dorsal views are shown in (A-B, D), while side views are shown in (C, E). Anterior is up.

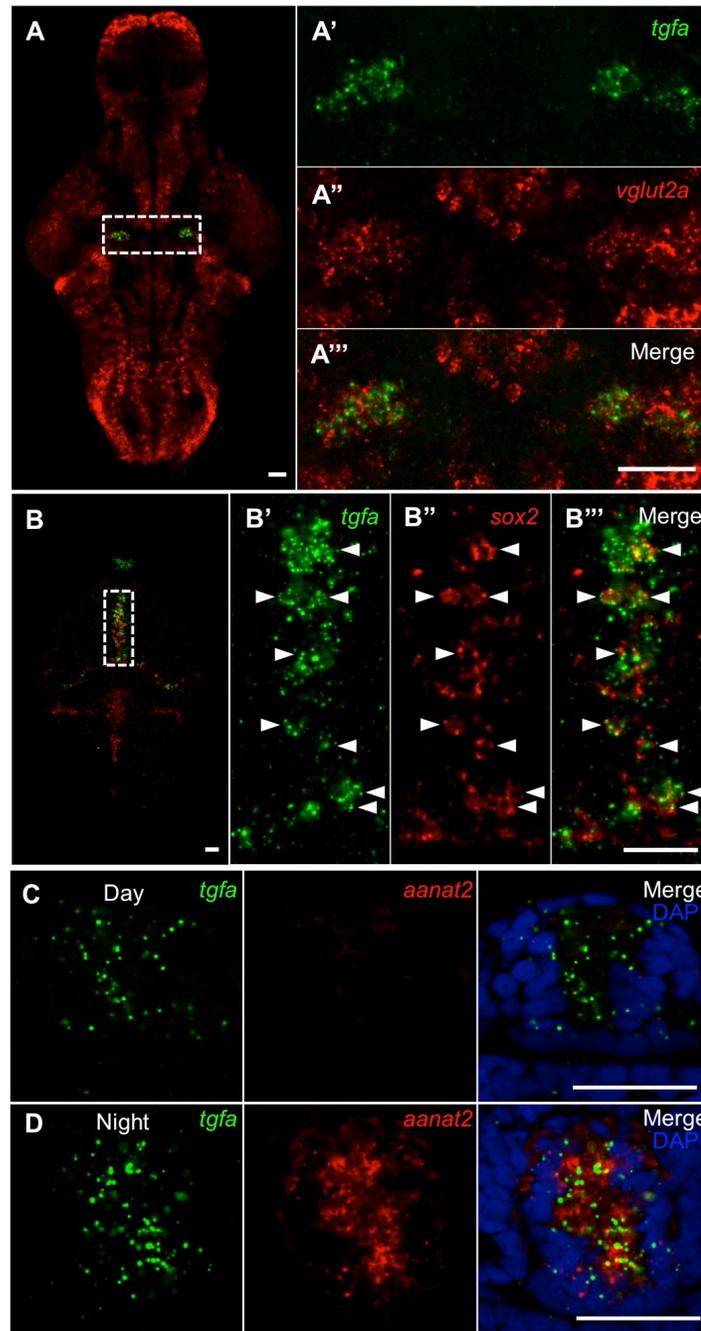


Figure 8. *tgfa* is expressed in glutamatergic bilateral clusters, *sox2*⁺ cells along the diencephalic ventricle, and the pineal gland. (A-A''') Double fluorescent ISH with probes specific for *tgfa* and *vglut2a* indicates the bilateral *tgfa*⁺ clusters are glutamatergic. (B-B''') *tgfa*⁺ cells along the diencephalic ventricle co-express *sox2* (arrowheads), suggesting these are ependymal cells. Dashed boxes in (A) and (B) indicate the approximate regions shown in (A'-A''') and (B'-B'''), respectively. (C-D) *tgfa* is also expressed in the pineal gland, which was identified anatomically by DAPI staining and by nighttime *aanat2* expression. No gross differences in the levels *tgfa* expression were observed between day and night, outside of normal sample variability. All images are single confocal sections of a WT brain at 5 days post fertilization, dorsal view. Anterior is up. Scale bars: 30 μ m.

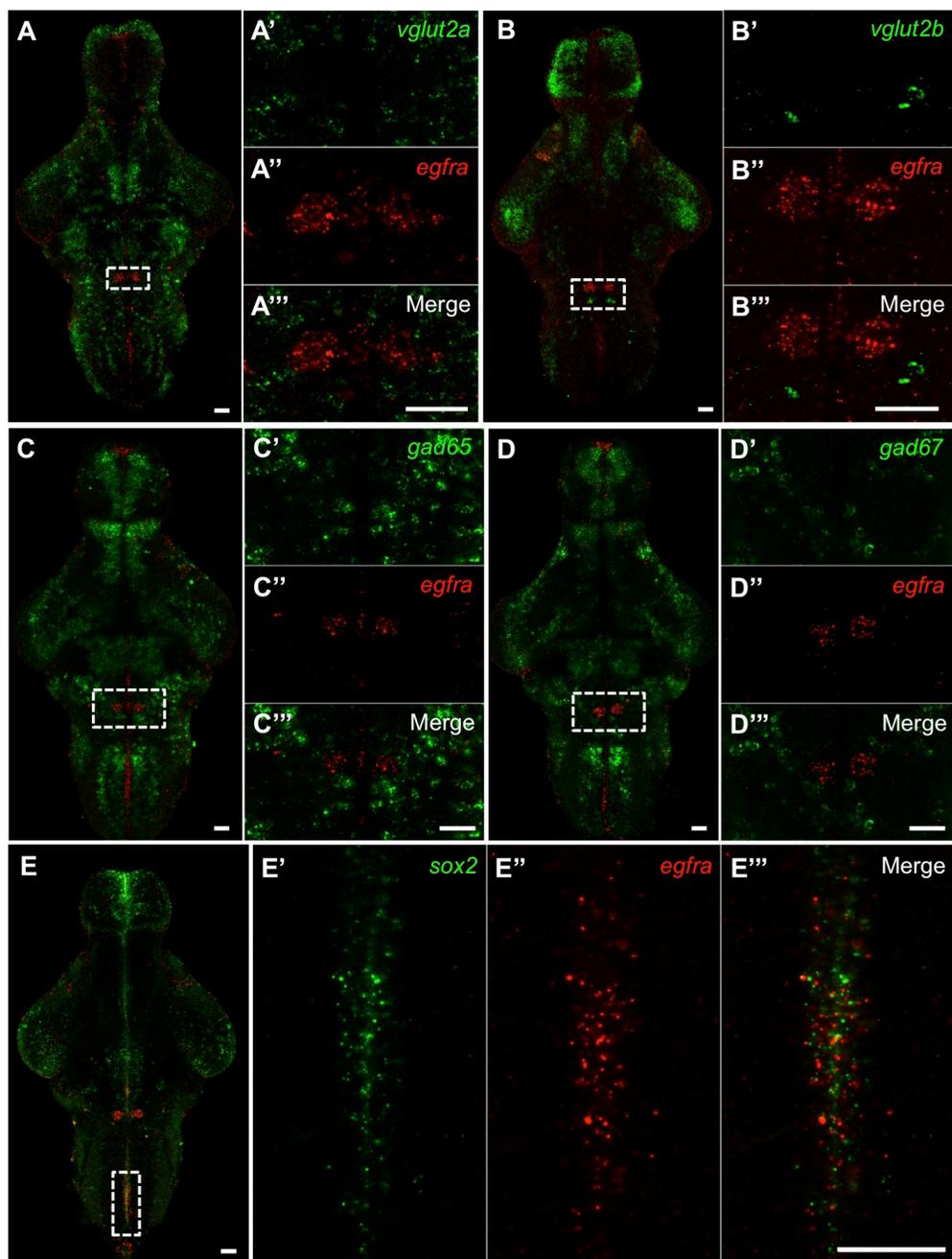


Figure 9. *egfra* is expressed in bilateral clusters in the hindbrain and *sox2*+ cells along the ventricle. (A-D''') *egfra*+ cells in the hindbrain express neither glutamatergic markers *vglut2a* and *vglut2b* nor GABAergic markers *gad65* and *gad67* by double fluorescent ISH. (E-E''') The *egfra*+ cells along the rhombencephalic ventricle express *sox2*, which suggests they might be ependymal cells. Dashed boxes in (A-E) indicate approximate regions shown at higher magnification in (A'-E'''). All images are single confocal sections of a WT brain at 5 days post fertilization, ventral view. Anterior is up. Scale bars: 30 μ m.

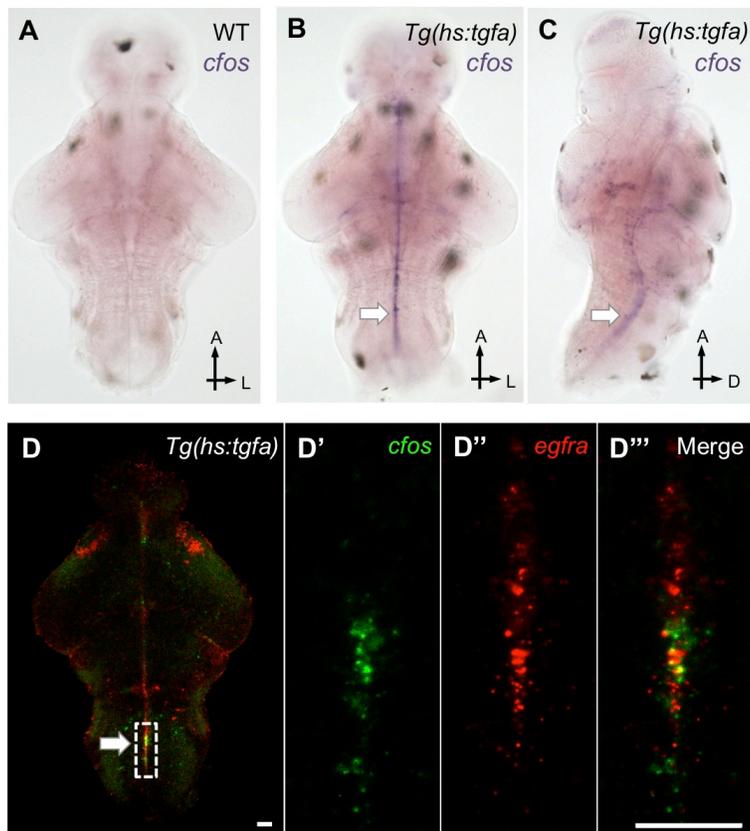


Figure 10. TGF- α overexpression activates EGFR+ cells along the rhombencephalic ventricle. (A) ISH with a *cfos*-specific probe on a WT larva fixed in the dark, 4 hours post-heat shock on day 5 shows low baseline activity in the brain. (B-C) *cfos* expression in a *Tg(hs:tgfa)* larva 4 hours post-heat shock shows activation along the ventricles, with the strongest expression near the rhombencephalic ventricle (arrow). (D-D''') Double fluorescent ISH with *cfos* and *egfra* probes on *Tg(hs:tgfa)* larvae 4 hours post-heat shock suggests that *egfra*+ cells along the rhombencephalic ventricle are activated by TGF- α overexpression. Dashed box in (D) indicates the approximate regions shown in (D'-D'''). All fluorescent images are single confocal sections. Ventral views are shown in (A-B) and (D-D'''), while (C) shows a side view. Anterior is up. Scale bars: 30 μ m.

2.9 *tgfa* expression does not cycle in a circadian manner

In mice, *Tgfa* expression in the suprachiasmatic nucleus (SCN) has been reported to cycle in a circadian manner (Kramer et al., 2001). We measured total levels of *tgfa* mRNA in 5 dpf zebrafish over a 24-hour period using quantitative reverse-transcription PCR (qRT-PCR), but did not observe a significant change in expression (Fig. 11A). However, if *tgfa* expression only cycles in a subset of all *tgfa*⁺ cells, the effect on total transcript levels might be too small to detect by qRT-PCR. As an alternative approach, we fixed WT zebrafish larvae every 6 hours for 18 hours and measured local levels of *tgfa* expression by ISH and densitometry analysis. The bilateral clusters of *tgfa*⁺ cells showed no significant change in expression over time (Fig. 11B-C). These results suggest that levels of *tgfa* mRNA in zebrafish do not cycle in a circadian manner.

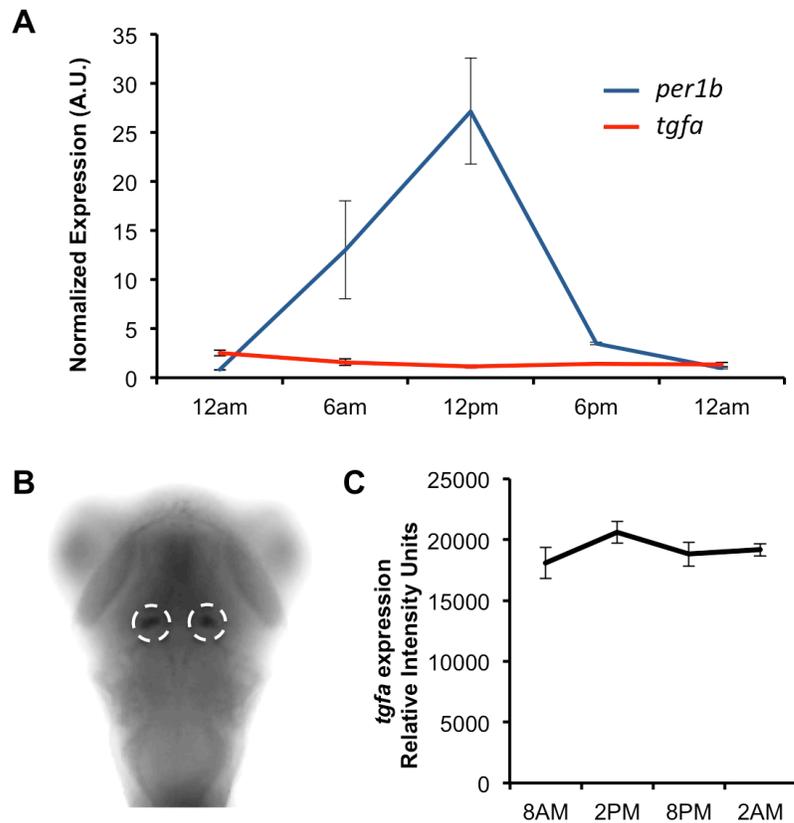


Figure 11. *tgfa* expression does not cycle in circadian manner. (A) qPCR analysis of total *tgfa* transcript levels showed no significant change in expression over 24 hours. Samples were taken every 6 hours and normalized against *eefla* expression. *per1b* included as a positive control. (B-C) Densitometry analysis after ISH showed no significant difference in bilateral *tgfa* expression over 18 hours. Dashed circles in (B) indicate regions quantified in (C). $p > 0.05$ by one-way ANOVA followed by Tukey's post hoc tests.

2.10 TGF- α -mediated sleep does not require overt circadian rhythms and is light-dependent

Because TGF- α overexpression induces greater behavioral changes during the day, we hypothesized that TGF- α -mediated sleep is influenced by either the circadian clock or by light itself. To distinguish between these possibilities, we first raised zebrafish larvae in constant light to abolish overt circadian rhythms. Following TGF- α overexpression, arrhythmic *Tg(hs:tgfa)* larvae had a significant, long-lasting reduction in locomotor activity and a commensurate increase in sleep (Fig. 12). We then raised larvae on a normal light/dark schedule until 4 dpf but shifted them to either constant light or constant dark after heat shock on day 5. *Tg(hs:tgfa)* larvae shifted to constant light after heat shock had significantly reduced locomotor activity and increased sleep, even during subjective night 5 (Fig. 13A-D). In contrast, *Tg(hs:tgfa)* larvae shifted to constant dark behaved like WT controls immediately after heat shock (Fig. 13E-H). If TGF- α -mediated sleep were dependent on circadian phase, we would expect that *Tg(hs:tgfa)* larvae shifted to constant dark would show reduced locomotor activity and increased sleep during subjective day 6, when in fact the opposite was observed (Fig. 13E-H). TGF- α -mediated sleep therefore does not require overt circadian rhythms but is dependent on light.

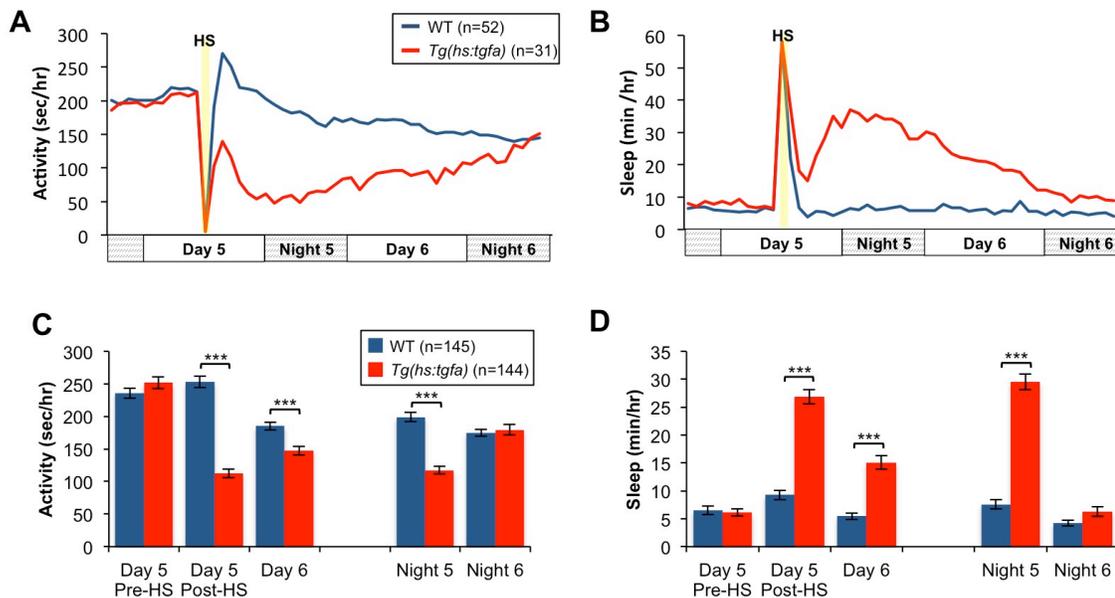


Figure 12. TGF- α -mediated sleep does not require circadian rhythms. *Tg(hs:tgfa)* larvae raised and monitored in constant light had significantly decreased locomotor activity (A) and increased sleep (B) compared to WT larvae on day 5 post-heat shock, night 5, and day 6, despite the absence of overt circadian rhythms. Data from a single representative experiment are shown in (A-B), while the combined data from four experiments are quantified in (C-D). Bar graphs indicate mean \pm SEM. ***, $p \leq 0.001$ compared to WT larvae by Mann-Whitney U-test.

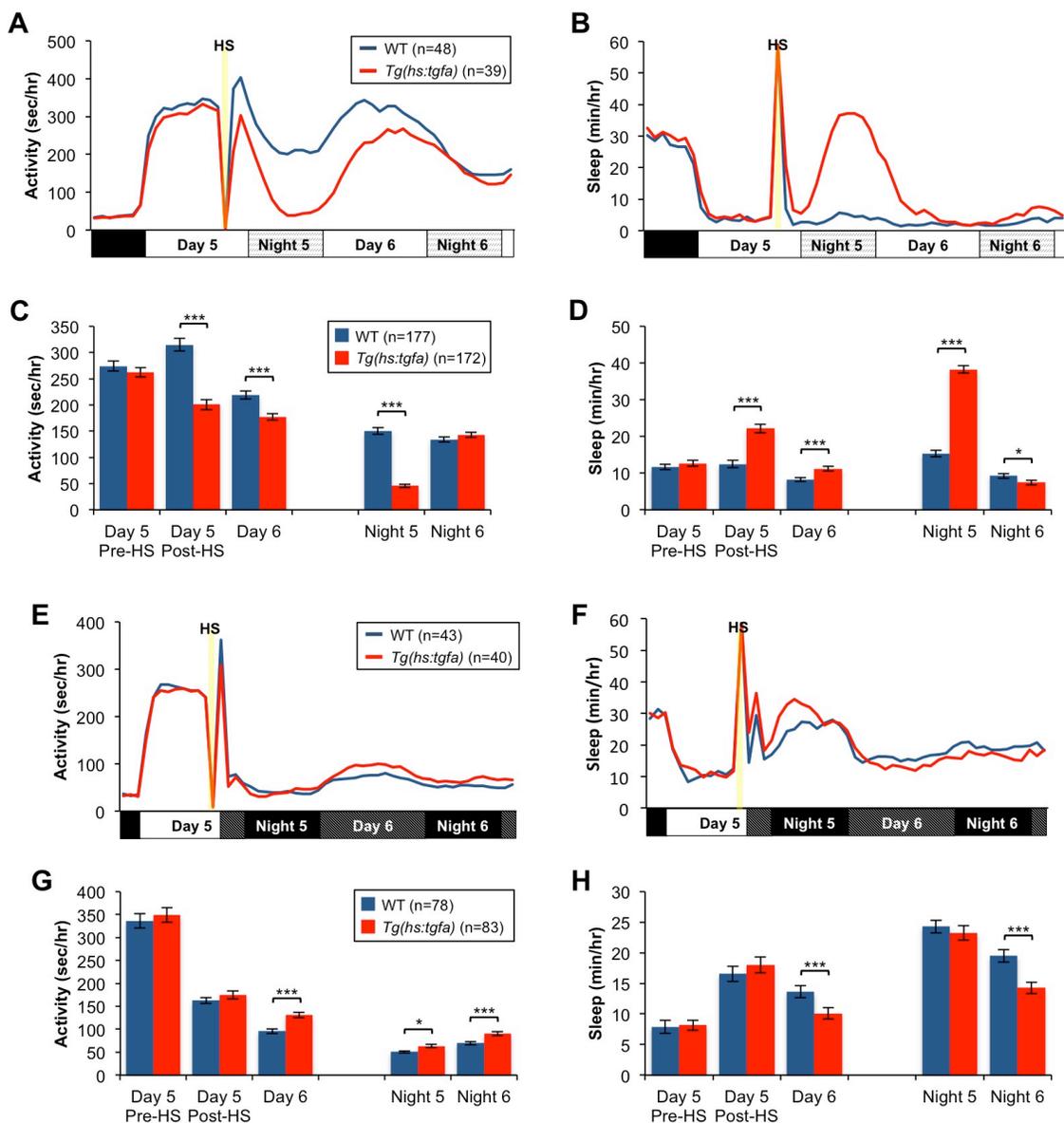


Figure 13. TGF- α -mediated sleep is light dependent. *Tg(hs:tgfa)* larvae raised in normal light/dark conditions and shifted to constant light following heat shock had significantly decreased locomotor activity (A) and increased sleep (B) compared to WT larvae on day 5 post-heat shock, night 5, day 6, and night 6. Data from a single representative experiment are shown in (A-B), while the combined data from four experiments are quantified in (C-D). However, *Tg(hs:tgfa)* larvae raised in normal light/dark conditions and shifted to constant darkness following heat shock had no difference in activity (E) and sleep levels (F) on day 5 post heat shock compared to WT larvae, and a small but significant increase in activity on night 5, day 6, and night 6 (G), and decreased sleep on day 6 and night 6 (H). Data from a single representative experiment are shown in (E-F), while the combined data from two experiments are quantified in (G-H). Bar graphs indicate mean \pm SEM. ***, $p \leq 0.001$; *, $p \leq 0.05$ compared to WT larvae by Mann-Whitney U-test.

2.11 Discussion

Using genetic and pharmacological techniques, we demonstrate that EGFR signaling is both necessary and sufficient for normal sleep behavior in zebrafish. These findings are consistent with previous studies in *C. elegans* and *D. melanogaster*, suggesting that the role of EGFR in sleep is evolutionarily ancient. We also provide insight into the cellular and molecular basis of TGF- α –mediated sleep. Specifically, we show that TGF- α overexpression activates *egfra*+ cells along the brain ventricle and requires MAPK/ERK signaling to induce sleep. We further demonstrate that TGF- α -mediated sleep is light dependent and does not require overt circadian rhythms. However, several ostensible differences between EGFR signaling in zebrafish and rodents merit further discussion.

In hamsters, *Tgfa* mRNA expression in the SCN varies over time and intracerebroventricular infusion of TGF- α affects the circadian timing, but not the overall amount of sleep of animals monitored in constant dark (Kramer et al., 2001). We found that levels of *tgfa* mRNA do not fluctuate in zebrafish, either globally or locally (Fig. 11). However, zebrafish TGF- α expression might be dynamically regulated at the protein level by proteases that cleave the membrane-bound TGF- α pro-peptide, similar to Rhomboid/Star processing of EGFR ligands in *Drosophila* (Foltényi et al., 2007). Since a TGF- α -specific antibody has not been reported in zebrafish, we were unable to confirm this by immunocytochemistry, but it may be possible to measure levels of soluble TGF- α peptide over time through CSF extraction followed by mass spectrometry (Chang and Sive, 2012). These experiments are ongoing. We also found that TGF- α overexpression in zebrafish increases total sleep in a light dependent manner (Fig. 13). While hamsters

are nocturnal, TGF- α might promote sleep in rodents when paired with light. Indeed, multiple rodent studies show an increase in behavioral quiescence after TGF- α infusion under standard LD conditions (Kramer et al., 2001; Snodgrass-Belt et al., 2005; Gilbert and Davis, 2009), but sleep after TGF- α infusion was only measured by electroencephalography (EEG) and electromyography (EMG) in constant dark (Kramer et al., 2001). Additional sleep EEG/EMG recordings of rodents infused with TGF- α during LD or constant dim light are therefore required. To further elucidate the interaction between TGF- α and circadian rhythms in zebrafish, we are currently testing whether TGF- α overexpression affects molecular components of the circadian clock using a *period3:luciferase* reporter line, and monitoring the locomotor activity of *tgfa* *-/-* larvae entrained in LD and shifted to constant light to determine whether *tgfa* null mutants have altered circadian behavior in free running conditions.

Though EGFR binds to multiple ligands in mammals (Harris et al., 2003), some ligands, such as EGF and TGF- α , bind exclusively to EGFR and not to related tyrosine kinase receptors. We demonstrated that genetic inactivation of TGF- α in zebrafish results in a modest but significant decrease in daytime sleep after day 5 (Fig. 2). We also found that treatment with the EGFR antagonist gefitinib dramatically decreased sleep during both day and night (Fig. 3). To corroborate this effect, we designed a truncated form of EGFR previously shown to act as a dominant negative in mammalian cell culture (Kashles et al., 1991), but we were unable to demonstrate that the truncated EGFR variant had any dominant negative activity *in vivo* (data not shown). Surprisingly, it is unclear whether EGFR is required for normal sleep behavior in rodents. Behavioral characterization of *waved-2* mice, which express an EGFR hypomorph, shows no

reproducible difference in locomotor activity compared to congenic littermate controls (Mrosofsky et al., 2005; Roberts et al., 2006). Furthermore, no analysis of sleep behavior in *waved-2* mice, either by behavioral assays or EEG/EMG recording, has been reported. This study is therefore the first demonstration that EGFR signaling is required for normal sleep behavior in a vertebrate. We are currently testing the behavioral phenotype of EGFR-ligand double mutant zebrafish (*tgfa* *-/-*; *egf* *-/-*), which may exhibit a more pronounced sleep defect. We are also attempting to isolate an EGFR mutant generated using CRISPR/Cas9 (Hwang et al., 2013; Jao et al., 2013) or TILLING approaches (Fig. S2).

2.12 Experimental Procedures

Ethics statement

Zebrafish experiments and husbandry followed standard protocols (Westerfield, 1993) in accordance with Caltech Institutional Animal Care and Use Committee guidelines.

Transgenic zebrafish

To generate the *Tg(hs:tgfa)* transgenic line, we amplified the TGF- α open reading frame (ORF) from 5 dpf larval zebrafish cDNA using a nested PCR approach. We used outer primers 5'-CGCGTGCCTTCATCTTTATT-3', 5'-TCCCAGTGGCCATATTGAAC-3' and inner primers 5'-ATATCCCGGGCCACCATGATGTATCGTGCTTTTTTGG-3', 5'-GGCGTCTAGATCAAACCACTGTTTCTGAGTTAC-3'. The entire TGF- α ORF was subcloned downstream of a heat shock inducible promoter (Halloran et al., 2000) in a vector backbone containing an I-SceI endonuclease recognition site. We then co-injected

the *Tg(hs:tgfa)* overexpression plasmid with I-SceI endonuclease into zebrafish embryos at the 1-cell stage, raised the injected larvae, and screened their offspring for transgenic insertions by ISH.

To generate TGF- α loss-of-function mutants, we used plasmids obtained from Addgene to create TALENs (Reyon et al., 2012) that recognize the following sites: 5'-TGATGTATCGTGCTTT-3' and 5'-TTCTCACCGGTGAGTACA-3'. We isolated a mutant line with a 7 bp deletion (ORF nucleotides 25-31: 5'-ACAATAT-3') that shifts the reading frame after the eighth amino acid and introduces an early stop codon downstream, truncating the pro-peptide from 189 to 61 amino acids in length (Fig. S1). The mutant TGF- α pro-peptide is predicted to lack essential features, including the signal peptide, protease cleavage sites, and the epidermal growth factor-like domain (Harris et al., 2003).

The EGFR dominant negative was designed to include the entire N-terminal and transmembrane domains, but a severely shortened C-terminal domain, as described previously (Kashles et al., 1991). We first amplified the complete EGFR ORF from 5 dpf larval cDNA by nested PCR, using the primers 5'-ACAAAGCCTGGAACGAAGAG-3', 5'-CCACTGGTCTAAAATAAGGTCATAAA-3', and 5'-CCGATAGCTTACAAACGCAAA-3', 5'-TAAGGTCAAATGTGAACACCTGAAT-3'. We isolated a 3756 bp band by gel purification and amplified the EGFR dominant negative fragment with the primers 5'-ATATCCCGGGCCACCATGGCAGGACCAACTGAAATC-3' and 5'-GGCGTCTAGACTACCTCCGGATGTGGCG-3'. The EGFR dominant negative coding sequence was subcloned into the same overexpression vector as the *Tg(hs:tgfa)* plasmid and co-injected with I-SceI endonuclease to generate stable transgenic lines.

Behavioral experiments

Videotracker behavioral experiments were performed as previously described (Gandhi et al., 2015; Rihel et al., 2010a). Larvae were raised at 28.5°C with 14 h of light and 10 h darkness until at least 4 days post fertilization. Individual larvae were loaded into the wells of a 96-well plate (7701-1651, Whatman) and sealed with a transparent adhesive film (4311971, Applied Biosystems) to prevent evaporation. Plates were not sealed if drug or DMSO controls were added. Each 96-well plate was continuously monitored by infrared camera in a custom Zebrabox videotracker (Viewpoint Life Sciences) illuminated with infrared and visible light LEDs and held at a temperature between 28.5-29.5°C by recirculated water. Animals were allowed to acclimate to videotracker conditions one night; behavioral activity was scored onwards. To administer heat shock, plates were removed from the videotracker and immersed in a 37°C water bath for 1 h.

For drug experiments, gefitinib (13166, Cayman Chemical), SL327 (1969, Tocris Bioscience), and U0126 (1144, Tocris Bioscience) were dissolved in DMSO and diluted in E3 embryo medium to 5 µM, 3 µM, and 15 µM concentrations, respectively. Identical concentrations of DMSO were administered as a negative vehicle control in the same 96-well plate. After dissolving in DMSO, unused U0126 or gefitinib can be aliquoted and stored in the dark at -20°C, but SL327 should be used immediately. All drugs were initially tested across a broad range of concentrations to ensure the administered dosage was below toxic levels.

Behavioral data were analyzed using custom PERL and Matlab (version R2014a, The Mathworks, Inc.) scripts and Excel (Microsoft). A sleep bout was previously defined

as a period of one or more minutes with less than 0.1 sec of movement, after which larvae have an increased arousal threshold (Prober et al., 2006). Since many behavioral parameters are not normally distributed, we used a non-parametric Mann-Whitney U test, also known as the Wilcoxon rank-sum test, to determine whether two groups were significantly different, and one- or two-way ANOVA to test for differences between three or more groups. All statistical tests were performed using Prism 6 (GraphPad).

Arousal threshold assay

The arousal threshold assay was performed as described previously (Gandhi et al., 2015). *Tg(hs:tgfa)* or gefitinib-treated WT larvae were loaded into a 96-well plate on day 5 and placed in a modified videotracker. To induce TGF- α overexpression, the plate of *Tg(hs:tgfa)* larvae was immersed in a 37°C water bath for 1 h and allowed to recover at 28.5C for 2-3 h before entering the videotracker. An automated solenoid driver delivered taps to the plate from 12:30 am to 7:30 am on night 5 at 1 min intervals. Fourteen different tapping intensities were delivered randomly, with 30 trials at each intensity. The response of larvae to stimuli was monitored using the videotracking software and subsequently analyzed in Matlab, Excel, and Prism 6.

***In situ* hybridization (ISH)**

Samples were fixed in 4% paraformaldehyde for 12-16 h at room temperature. ISH was performed using digoxigenin (DIG)- or 2,4-dinitrophenol (DNP)-labeled antisense riboprobes as previously described (Thisse and Thisse, 2008), except all samples were mounted in 50% glycerol/PBS. Brightfield images were acquired using a Zeiss Axio

ImagerM1 microscope. Fluorescent ISH was performed using the TSA Plus DNP System (PerkinElmer). Fluorescent images were acquired on an upright Zeiss LSM 780 confocal microscope and analyzed using Fiji (Schindelin et al., 2012).

Quantitative reverse-transcription PCR (qRT-PCR)

We raised WT larvae on a normal 14:10 h light:dark schedule until 6am on day 5. Total RNA was isolated using Trizol reagent (15596-026, Life Technologies) from three biological replicates (25 larvae each), collected every 6 hours for 42 hours. We then generated cDNA (Superscript III First-Strand Synthesis System, Invitrogen) and amplified transcripts using SYBR green master mix (4364346, Life Technologies). Transcripts of *tgfa* were amplified with primers 5'-GTGTGTGGTGGGCAGTGTC-3' and 5'-CCAACAGGAGAGGGTGTGAC-3', while the *eef1a* reference gene was amplified with 5'-CAGCTGATCGTTGGAGTCAA-3' and 5'-TGTATGCGCTGACTT-CCTTG-3'. Each qRT-PCR reaction was run in triplicate on an ABI PRISM 7900HT Sequence Detection System (Applied Biosystems). Relative fold-change in expression was calculated using the $2^{-\Delta\Delta C_t}$ method (Schmittgen and Livak, 2008).

References

- Barker, A. J., Gibson, K. H., Grundy, W., Godfrey, A. A., Barlow, J. J., Healy, M. P., Woodburn, J. R., Ashton, S. E., Curry, B. J., Scarlett, L., et al. (2001). Studies leading to the identification of ZD1839 (IRESSA): an orally active, selective epidermal growth factor receptor tyrosine kinase inhibitor targeted to the treatment of cancer. *Bioorg. Med. Chem. Lett.* 11, 1911–1914.
- Buonanno, A. and Fischbach, G. D. (2001). Neuregulin and ErbB receptor signaling pathways in the nervous system. *Curr. Opin. Neurobiol.* 11, 287–296.
- Campbell, S. S. and Tobler, I. (1984). Animal sleep: a review of sleep duration across phylogeny. *Neurosci Biobehav Rev* 8, 269–300.
- Chang, J. T. and Sive, H. (2012). Manual drainage of the zebrafish embryonic brain ventricles. *J Vis Exp*.
- Cirelli, C. (2009). The genetic and molecular regulation of sleep: from fruit flies to humans. *Nat. Rev. Neurosci.* 10, 549–560.
- Crocker, A. and Sehgal, A. (2010). Genetic analysis of sleep. *Genes & Development* 24, 1220–1235.
- Foltényi, 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*. *Nature Neuroscience* 10, 1160–1167.
- Gandhi, A. V., Mosser, E. A., Oikonomou, G. and Prober, D. A. (2015). Melatonin is required for the circadian regulation of sleep. *Neuron*.
- Gilbert, J. and Davis, F. C. (2009). Behavioral effects of systemic transforming growth factor- α in Syrian hamsters. *Behav. Brain Res.* 198, 440–448.
- Halloran, M. C., Sato-Maeda, M., Warren, J. T., Su, F., Lele, Z., Krone, P. H., Kuwada, J. Y. and Shoji, W. (2000). Laser-induced gene expression in specific cells of transgenic zebrafish. *Development* 127, 1953–1960.
- Harris, R. C., Chung, E. and Coffey, R. J. (2003). EGF receptor ligands. *Exp. Cell Res.* 284, 2–13.
- Hong, C. C., Peterson, Q. P., Hong, J.-Y. and Peterson, R. T. (2006). Artery/vein specification is governed by opposing phosphatidylinositol-3 kinase and MAP kinase/ERK signaling. *Curr. Biol.* 16, 1366–1372.
- Huang, P., Xiao, A., Zhou, M., Zhu, Z., Lin, S. and Zhang, B. (2011). Heritable gene targeting in zebrafish using customized TALENs. *Nature Publishing Group* 29, 699–700.
- Hwang, W. Y., Fu, Y., Reyon, D., Maeder, M. L., Tsai, S. Q., Sander, J. D., Peterson, R.

- T., Yeh, J.-R. J. and Joung, J. K. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.* 31, 227–229.
- Jao, L.-E., Wente, S. R. and Chen, W. (2013). Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proc. Natl. Acad. Sci. U.S.A.* 110, 13904–13909.
- Kashles, O., Yarden, Y., Fischer, R., Ullrich, A. and Schlessinger, J. (1991). A dominant negative mutation suppresses the function of normal epidermal growth factor receptors by heterodimerization. *Mol. Cell. Biol.* 11, 1454–1463.
- Kramer, A., Yang, F. C., Snodgrass, P., Li, X., Scammell, T. E., Davis, F. C. and Weitz, C. J. (2001). Regulation of daily locomotor activity and sleep by hypothalamic EGF receptor signaling. *Science* 294, 2511–2515.
- 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, R509–14.
- Lee, H. J., Wu, J., Chung, J. and Wrathall, J. R. (2013). SOX2 expression is upregulated in adult spinal cord after contusion injury in both oligodendrocyte lineage and ependymal cells. *J. Neurosci. Res.* 91, 196–210.
- Mrosovsky, N., Redlin, U., Roberts, R. B. and Threadgill, D. W. (2005). Masking in waved-2 mice: EGF receptor control of locomotion questioned. *Chronobiol. Int.* 22, 963–974.
- Plata-Salamán, C. R. (1991). Epidermal growth factor and the nervous system. *Peptides* 12, 653–663.
- 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. *Journal of Neuroscience* 26, 13400–13410.
- Rechtschaffen, A., Gilliland, M. A., Bergmann, B. M. and Winter, J. B. (1983). Physiological correlates of prolonged sleep deprivation in rats. *Science* 221, 182–184.
- Reyon, D., Khayter, C., Regan, M. R., Joung, J. K. and Sander, J. D. (2012). Engineering designer transcription activator-like effector nucleases (TALENs) by REAL or REAL-Fast assembly. *Curr Protoc Mol Biol* Chapter 12, Unit 12.15.
- Rihel, J., Prober, D. A. and Schier, A. F. (2010a). Monitoring sleep and arousal in zebrafish. *Methods Cell Biol.* 100, 281–294.
- Rihel, J., Prober, D. A., Arvanites, A., Lam, K., Zimmerman, S., Jang, S., Haggarty, S. J., Kokel, D., Rubin, L. L., Peterson, R. T., et al. (2010b). Zebrafish behavioral profiling links drugs to biological targets and rest/wake regulation. *Science* 327, 348–351.
- Roberts, R. B., Thompson, C. L., Lee, D., Mankinen, R. W., Sancar, A. and Threadgill,

- D. W. (2006). Wildtype epidermal growth factor receptor (Egfr) is not required for daily locomotor or masking behavior in mice. *J Circadian Rhythms* 4, 15.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat Meth* 9, 676–682.
- Schmittgen, T. D. and Livak, K. J. (2008). Analyzing real-time PCR data by the comparative CT method. *Nat Protoc* 3, 1101–1108.
- Shaw, P. J., Tononi, G., Greenspan, R. J. and Robinson, D. F. (2002). Stress response genes protect against lethal effects of sleep deprivation in *Drosophila*. *Nature* 417, 287–291.
- Siegel, J. M. (2009). Sleep viewed as a state of adaptive inactivity. *Nat. Rev. Neurosci.* 10, 747–753.
- Snodgrass-Belt, P., Gilbert, J. L. and Davis, F. C. (2005). Central administration of transforming growth factor-alpha and neuregulin-1 suppress active behaviors and cause weight loss in hamsters. *Brain Research* 1038, 171–182.
- Suh, H., Consiglio, A., Ray, J., Sawai, T., D'Amour, K. A. and Gage, F. H. (2007). In vivo fate analysis reveals the multipotent and self-renewal capacities of Sox2+ neural stem cells in the adult hippocampus. *Cell Stem Cell* 1, 515–528.
- Thisse, C. and Thisse, B. (2008). High-resolution in situ hybridization to whole-mount zebrafish embryos. *Nat Protoc* 3, 59–69.
- Van Buskirk, C. and Sternberg, P. W. (2007). Epidermal growth factor signaling induces behavioral quiescence in *Caenorhabditis elegans*. *Nature Neuroscience* 10, 1300–1307.
- Ward, W. H., Cook, P. N., Slater, A. M., Davies, D. H., Holdgate, G. A. and Green, L. R. (1994). Epidermal growth factor receptor tyrosine kinase. Investigation of catalytic mechanism, structure-based searching and discovery of a potent inhibitor. *Biochem. Pharmacol.* 48, 659–666.
- Westerfield, M. (1993). *The Zebrafish Book. A guide for the laboratory use of zebrafish (Danio rerio)*. 4th edn. Eugene, OR: University of Oregon Press.
- Zhdanova, I. V. (2006). Sleep in zebrafish. *Zebrafish* 3, 215–226.

Supplemental Figures

TGFa-001_Hs	-MVPSAGQLALFALGIVLAAEQALENSTSPLSADPP-----	35
TGFa-001_Mm	-MVPATGQLALLALGILLAVCOALENSTSPLS-DSP-----	34
TGFa-001_Dr	MMYRAFWDTIFFLLTGS_LFTYGQVQENSTSTTTIATTTTTTTTTTTTTTTS	50
TGFa-001_Dr d7	MMYRAFWDSFSPAPFLRMGRCKRILPQRRLQPRQQLLQQLPPLPPPAPR	50
TGFa-001_Hs	-----VAAA VVSHFNDCPDSHTQFCFHGTCRFLVQEDKPACV	72
TGFa-001_Mm	-----VAAA VVSHFNKCPDSHTQYCFHGTCTFLVQEEKPACV	71
TGFa-001_Dr	TTTTTTTTTRVRKFIAAA VVSHFDDCPDSHSHFCFHGTCRFLILEETPACV	100
TGFa-001_Dr d7	PPLLPHESENL	61
TGFa-001_Hs	CHSGYVGARCEHADLLAVVAASQKKQAITALVVVSIVALAVLIITCVLIH	122
TGFa-001_Mm	CHSGYVGVRCCEHADLLAVVAASQKKQAITALVVVSIVALAVLIITCVLIH	121
TGFa-001_Dr	CHPGFVGMRCCEHADLLAVVASNHRQQT VATMLVL CVVGSVMMLICTLLN	150
TGFa-001_Dr d7		61
TGFa-001_Hs	CCQVRKHCEWCRALICRHEKPSALLKGRTACCHSETVV	160
TGFa-001_Mm	CCQLRKHCEWCRALVCRHEKPSALLKGRTACCHSETVV	159
TGFa-001_Dr	CWRRRGGCGRGHTLSCWTEKPRSILTSGTSCCNSETVV	188
TGFa-001_Dr d7		61

Figure S1. TGF- α protein alignment. The amino acid sequences of human (Hs), mouse (Mm), and zebrafish (Dr) TGF- α pro-peptides are shown. The predicted TGF- α sequence in *tgfa* $-/-$ mutant larvae (Dr d7) is included for comparison. Red box indicates region between protease cleavage sites, which forms the mature peptide.

EGFR-001_Hs	-MAPSGTAGPALLLAALDPAASRALEBKVVDDGTSKLTDLGTEDHFLSLDRMANNCEAWLGNLEITWQRMVQLSFLKTIQEVAGWMLIAINTVERI	99
EGFR-001_Mm	-MAPSGTARTTLMLLTALDRAGGLEBKVVDDGTSKLTDLGTEDHFLSLDRMANNCEAWLGNLEITWQRMVQLSFLKTIQEVAGWMLIAINTVERI	99
EGFRa_Dr	MAGPTEIGLFFATLLSSSFQRTF---BKVVDDGANNKLTDLGTVEDHVVQLRWYVNGTWLSELEITHIEMVQLSFLKTIQEVAGWMLIAINTVSKI	96
EGFRa_Dr mut	MAGPTEIGLFFATLLSSSFQRTF---BKVVDDGANNKLTDLGTVEDHVVQLRWYVNGTWLSELEITHIEMVQLSFLKTIQEVAGWMLIAINTVSKI	96
EGFR-001_Hs	PLENLEIIRGHWVENSVALRWLSNYD-ANKTDLKELPARNLEILLGAMAFSNPPLCNVESIQARDIYSSDFLSNHSDFQNLHGSQKCDPSPQNGS	198
EGFR-001_Mm	PLENLEIIRGHWVENVYALFLLSNYG-TNRTDURELPARNLEILLGAMAFSNPPLCNVDTIQARDIQMVAFMSMISDQDS-PSSQKCDPSPQNGS	198
EGFRa_Dr	PLENLEIIRGHWVSLVEKFLRWLWYVNSIEDVWELPDLTSLTEILKGGVAFNPNHLCNVGTLEARDILNMGSLPTIVSKNLSVCKGKCDPSPQNGS	196
EGFRa_Dr mut	PLENLEIIRGHWVSLVEKFLRWLWYVNSIEDVWELPDLTSLTEILKGGVAFNPNHLCNVGTLEARDILNMGSLPTIVSKNLSVCKGKCDPSPQNGS	196
EGFR-001_Hs	QWGSSENDKLTIKIICADCCSGPQKQSPDCCNCDRAGCTGAPRESDCLVCKFDEFTCKDITCPPMLVNPITTWQDWPPEKVSFGATCKKCPAN	298
EGFR-001_Mm	QWGSSENDKLTIKIICADCCSGPQKQSPDCCNCDRAGCTGAPRESDCLVCKFDEFTCKDITCPPMLVNPITTWQDWPPEKVSFGATCKKCPAN	298
EGFRa_Dr	QWGTGPKDQDMTKMICADCCSGPQKQSPDCCNEHCRAGCTGAPRPTDCLACKFQDEFTCKDCPPMLVLDPTNHLAPNPVSKVSFGATCKKCPAN	296
EGFRa_Dr mut	QWGTGPKDQDMTKMICADCCSGPQKQSPDCCNEHCRAGCTGAPRPTDCLACKFQDEFTCKDCPPMLVLDPTNHLAPNPVSKVSFGATCKKCPAN	296
EGFR-001_Hs	VWTDHGSCVADGADSEVEEDGMFKKKKCEGPKKWCNGIGIGEFKDSLSINATNKKHKNDTSISCDLHILPWAFRQDSFTHTPPLDQQLDILKTV	398
EGFR-001_Mm	VWTDHGSCVADGADSEVEEDGIRKCKKCDGPKKWCNGIGIGEFKDSLSINATNKKHKNDTSISCDLHILPWAFRQDSFTHTPPLDQQLDILKTV	398
EGFRa_Dr	VWTDHGQVRTDQPTVEVDEBWKCKKCEGPKKWCNGIDGFLANVLSINATNDSSECTKISQWAILSTARGQPHDSSGDPKLSVLSIV	396
EGFRa_Dr mut	VWTDHGQVRTDQPTVEVDEBWKCKKCEGPKKWCNGIDGFLANVLSINATNDSSECTKISQWAILSTARGQPHDSSGDPKLSVLSIV	396
EGFR-001_Hs	KEITGALLIQWIPENITDLHAFENLEIIRGRTKQHGQFLAWGLNITSLGLRSLKEISGGDWISGNKINLCYANTINIKKLFGTGPKTKIISVQENH	498
EGFR-001_Mm	KEITGALLIQWIPENITDLHAFENLEIIRGRTKQHGQFLAWGLNITSLGLRSLKEISGGDWISGNKINLCYANTINIKKLFGTGPKTKIISVQENH	498
EGFRa_Dr	KEITGALLIQWIPESNCSLAFENLEIRGRTKQHGQFLAWGLNITSLGLRSLKEISGGDWISGNKINLCYSSPEHAKLPLKSKQDQSVKTIISDQAT	496
EGFRa_Dr mut	KEITGALLIQWIPESNCSLAFENLEIRGRTKQHGQFLAWGLNITSLGLRSLKEISGGDWISGNKINLCYSSPEHAKLPLKSKQDQSVKTIISDQAT	496
EGFR-001_Hs	DKATGCMVLAQSPEDGCEPRADQVSCNWSRQRECVKCNLLGEGEPREFVENSECIIQHPCELPQAINITCTGRGPDNCIQCAHWIDGPHCKVCTPFG	598
EGFR-001_Mm	DKAVHWGNPLCSSEDCGCEPRADQVSCNWSRQRECVKCNLLGEGEPREFVENSECIIQHPCELPQAINITCTGRGPDNCIQCAHWIDGPHCKVCTPFG	598
EGFRa_Dr	DFNQSTDNQCTADQCGPSPHDFGSEHVSAGKQVPSQNLLEPREVEVNTDYEQPECLDINETDCTGSPQKQTVDAWVQSGPHCVFQCPDQ	596
EGFRa_Dr mut	DFNQSTDNQCTADQCGPSPHDFGSEHVSAGKQVPSQNLLEPREVEVNTDYEQPECLDINETDCTGSPQKQTVDAWVQSGPHCVFQCPDQ	596
EGFR-001_Hs	WIGENNTLWIKVADRGVCHLCHPACTVGGCTGPGLECCPTN--SPKIPSIATGVMGRLLLWVWGLIGLAFARAHVAKRATLRLQLQRELVEPLTPSG	696
EGFR-001_Mm	WIGENNTLWIKVADRGVCHLCHPACTVGGCTGPGLECCPTN--SPKIPSIATGVMGRLLLWVWGLIGLAFARAHVAKRATLRLQLQRELVEPLTPSG	698
EGFRa_Dr	VPEKDTLWIKVADRVCHLCHPACTVGGCTGPGLECCPTN--SPKIPSIATGVMGRLLLWVWGLIGLAFARAHVAKRATLRLQLQRELVEPLTPSG	695
EGFRa_Dr mut	VPEKDTLWIKVADRVCHLCHPACTVGGCTGPGLECCPTN--SPKIPSIATGVMGRLLLWVWGLIGLAFARAHVAKRATLRLQLQRELVEPLTPSG	695
EGFR-001_Hs	BARNQALLRILKETEFKKIKVLSGAFGTWKGWVPEBQWKIPVAIKELREBTSPPKANKEILDERWVMSVDNPHMCRLLGICLTSTVQLITQLMPV	796
EGFR-001_Mm	BARNQALLRILKETEFKKIKVLSGAFGTWKGWVPEBQWKIPVAIKELREBTSPPKANKEILDERWVMSVDNPHMCRLLGICLTSTVQLITQLMPV	798
EGFRa_Dr	BARNQALLRILKETEFKKIKVLSGAFGTWKGWVPEBQWKIPVAIKELREBTSPPKANKEILDERWVMSVDNPHMCRLLGICLTSTVQLITQLMPV	795
EGFRa_Dr mut	BARNQALLRILKETEFKKIKVLSGAFGTWKGWVPEBQWKIPVAIKELREBTSPPKANKEILDERWVMSVDNPHMCRLLGICLTSTVQLITQLMPV	795
EGFR-001_Hs	CLLDVWREHKDNGISGVLLNLCVQIRKGNIMLEDRLVHDLAARMLVKTPQHWKITDFGLAKLLGAEKKEVHAFSGKVPKIMALESILHRIYTHQSD	896
EGFR-001_Mm	CLLDVWREHKDNGISGVLLNLCVQIRKGNIMLEDRLVHDLAARMLVKTPQHWKITDFGLAKLLGAEKKEVHAFSGKVPKIMALESILHRIYTHQSD	898
EGFRa_Dr	CLLDVWREHKDNGISGVLLNLCVQIRKGNIMLEDRLVHDLAARMLVKTPQHWKITDFGLAKLLGAEKKEVHAFSGKVPKIMALESILHRIYTHQSD	895
EGFRa_Dr mut	CLLDVWREHKDNGISGVLLNLCVQIRKGNIMLEDRLVHDLAARMLVKTPQHWKITDFGLAKLLGAEKKEVHAFSGKVPKIMALESILHRIYTHQSD	895
EGFR-001_Hs	MWSVGTWIELMTFGSKPYDGIIPASEISSILEKGERLPQPPICTIDVMIIMKQWIMDASRAKFAELIIFSKIMADPQRYLVIQGDERHILPSPDGN	996
EGFR-001_Mm	MWSVGTWIELMTFGSKPYDGIIPASEISSILEKGERLPQPPICTIDVMIIMKQWIMDASRAKFAELIIFSKIMADPQRYLVIQGDERHILPSPDGN	998
EGFRa_Dr	MWSVGTWIELMTFGSKPYDGIIPASEIAGVLEKGERLPQPPICTIDVMIIMKQWIMDASRAKFAELIIFSKIMADPQRYLVIQGDERHILPSPDGN	995
EGFRa_Dr mut	MWSVGTWIELMTFGSKPYDGIIPASEIAGVLEKGERLPQPPICTIDVMIIMKQWIMDASRAKFAELIIFSKIMADPQRYLVIQGDERHILPSPDGN	995
EGFR-001_Hs	FVRLVDEEDHEDWDADEVLIPOQGFNPSSTSRPTLLSSLSRATSNSTVACINRNG--SCFWKEDFLQRYSSOPTGAMTEDSIDDFLPVPEWNGS	1096
EGFR-001_Mm	FVRLVDEEDHEDWDADEVLIPOQGFNPSSTSRPTLLSSLSRATSNSTVACINRNG--SCFWKEDFLQRYSSOPTGAMTEDSIDDFLPVPEWNGS	1096
EGFRa_Dr	FVRSLSG-ELDPRVDADEVLPNHSFSSPSTSRPTLLSSLSRATSNSTVACINRNG--VWVRENSMLRWIPDTPRFQDQ--DFDPAQVNH-E	1085
EGFRa_Dr mut	FVRSLSG-ELDPRVDADEVLPNHSFSSPSTSRPTLLSSLSRATSNSTVACINRNG--VWVRENSMLRWIPDTPRFQDQ--DFDPAQVNH-E	1085
EGFR-001_Hs	WPKRPPGSDVQPVWYHQPLNAPSPDQVMDPASTAVGNPEVLY--TQPTQVSTQSPAHWAKGSHILSDNFDVQDQFFKPKVNGIRKGSITRE	1193
EGFR-001_Mm	WPKRPPGSDVQPVWYHQPLNAPSPDQVMDPASTAVGNPEVLY--TQPTQVSTQSPAHWAKGSHILSDNFDVQDQFFKPKVNGIRKGSITRE	1193
EGFRa_Dr	VMINQESSITIPWQDQ--HGPPATLHSSPALDET-EEVLYNCFKSPAPFVMEVLYTSHQLLSTKPPFSTIDNFDVQDQFFKPKVNGIRKGSITRE	1181
EGFRa_Dr mut	VMINQESSITIPWQDQ--HGPPATLHSSPALDET-EEVLYNCFKSPAPFVMEVLYTSHQLLSTKPPFSTIDNFDVQDQFFKPKVNGIRKGSITRE	1181
EGFR-001_Hs	YAEVLRVAPPSSSEFIGA	1210
EGFR-001_Mm	YAEVLRVAPPSSSEFIGA	1210
EGFRa_Dr	YAEVLRVAPPSSSEFIGA	1191
EGFRa_Dr mut	YAEVLRVAPPSSSEFIGA	1191

Figure S2. EGFR protein alignment. The amino acid sequence of human (Hs), mouse (Mm), and zebrafish (Dr) EGFR orthologs are shown. The zebrafish EGFR sequence was determined from 5 dpf larval cDNA. The effect of a nonsense mutation in the zebrafish EGFR locus (Dr mut) obtained from the Zebrafish Mutation Project is included for comparison. Red box indicates the tyrosine kinase domain, which is required for biological activity.

CHAPTER 3:
Hypocretin Neuronal Specification

3.1 Abstract

Loss of neurons that express the neuropeptide hypocretin (Hcrt) has been implicated in narcolepsy, a debilitating disorder characterized by excessive daytime sleepiness and cataplexy. Cell replacement therapy using Hcrt-expressing neurons generated *in vitro* is a potentially useful therapeutic approach, but factors sufficient to specify Hcrt neurons are unknown. Using zebrafish as a high-throughput system to screen for factors that can specify Hcrt neurons *in vivo*, we identified the LIM homeobox transcription factor Lhx9 as necessary and sufficient to specify Hcrt neurons. We found that Lhx9 can directly induce *hcrt* expression and identified two potential Lhx9 binding sites in the zebrafish *hcrt* promoter. Akin to its function in zebrafish, we found that Lhx9 is sufficient to specify *Hcrt*-expressing neurons in the developing mouse hypothalamus. Our results elucidate an evolutionarily conserved role for Lhx9 in Hcrt neuron specification that improves our understanding of Hcrt neuron development and may enable a novel therapeutic approach for narcolepsy.

3.2 Introduction

The hypocretin (Hcrt, also known as orexin) neuropeptide is conserved among vertebrates and plays key roles in regulating sleep, metabolism, feeding, anxiety, reward, and addiction (Bonnavion and Lecea, 2010; Tsujino and Sakurai, 2009). Hcrt is particularly important in promoting arousal, as the loss of Hcrt neurons is thought to cause narcolepsy (Peyron et al., 2000; Thannickal et al., 2000), a disorder characterized by daytime sleepiness, fragmented sleep-wake states, and cataplexy. Narcolepsy affects approximately 1 in 2,000 individuals, but treatments are limited to symptom management (Dauvilliers et al., 2007). Despite the importance of the Hcrt system, little is known about the developmental processes that give rise to Hcrt neurons. A recent study found that mice lacking the LIM domain homeobox transcription factor *Lhx9* had fewer Hcrt neurons (Dalal et al., 2013), suggesting that *Lhx9* is required to specify a subset of Hcrt neurons. However, overexpression of *Lhx9* in adult mice or in a mouse neuroblastoma cell line had no effect on Hcrt cell number or expression. Therefore, the role of *Lhx9* in Hcrt neuron specification remains unclear and the set of factors sufficient to specify Hcrt neurons remains unknown. Identifying these factors would help elucidate how a key neural circuit that governs sleep is established, and could lead to novel therapies for narcolepsy.

The zebrafish *Danio rerio* is a powerful genetic model of vertebrate development that provides several advantages for studying Hcrt neuron specification. First, the hypothalamus is remarkably conserved (Blackshaw et al., 2010; Machluf et al., 2011; Tessmar-Raible et al., 2007), suggesting that developmental mechanisms identified in zebrafish are likely to be relevant to mammals. Several studies have shown that the

mammalian Hcrt system is functionally and anatomically conserved in zebrafish (Chiu and Prober, 2013; Elbaz et al., 2013). But while the rodent hypothalamus contains thousands of Hcrt neurons, larval and adult zebrafish contain only approximately 10 and 40 Hcrt neurons, respectively (Faraco et al., 2006; Kaslin et al., 2004), making zebrafish a more tractable system to study Hcrt neuron development. Second, the external development and transparency of zebrafish embryos facilitate the observation of developing Hcrt neurons. Third, high-throughput genetic gain- and loss-of-function assays allows for efficient screens to identify developmental regulators. We exploited these features of zebrafish to identify genes that regulate Hcrt neuron specification.

3.3 Microarray analysis identifies transcripts enriched in Hcrt neurons

Previous studies showed that the number of Hcrt neurons in zebrafish and rodents increases as animals develop and mature to adulthood (Faraco et al., 2006; Kaslin et al., 2004; Sawai et al., 2010). We reasoned that cell-autonomous factors required to specify Hcrt neurons might still be expressed in Hcrt neurons shortly after they are specified. To identify these factors, we generated transgenic zebrafish that express monomeric red fluorescent protein (mRFP) in Hcrt neurons and enhanced green fluorescent protein (EGFP) in neurons that express the hypothalamic neuropeptide QRFP (Fig. S1). QRFP has been implicated in regulating locomotor activity (Takayasu et al., 2006), feeding (Chartrel et al., 2003; Takayasu et al., 2006), and nociception (Yamamoto et al., 2009) in rodents, and sleep/wake behaviors in zebrafish (C. Chiu, A. Chen and D. Prober, unpublished data).

Expression of *hcrt* and *qrfp* (*si:ch211-185o22.2*, incorrectly annotated as lincRNA) is first detected in zebrafish embryos at 24 hours post-fertilization (hpf) in bilateral hypothalamic nuclei of 4-6 cells (Fig. 1A,B), which expand to 10-15 cells by 120 hpf (Fig. 1C,D). Hcrt and QRFP are expressed in neighboring neurons throughout development, but are never co-expressed within the same cells (Fig. 1B,D). To identify genes with enriched expression in Hcrt neurons, we dissociated pools of 100-300 *Tg(hcrt:mRFP, qrfp:EGFP)* embryos at 26 hpf into single cells and isolated EGFP- and mRFP-expressing neurons by fluorescence activated cell sorting (FACS) (Fig. 1E, S2). FACS gates for EGFP and mRFP populations were set using wild-type embryos (0/10,000 EGFP+ or mRFP+ events). In a representative experiment, we obtained 250 EGFP+ cells and 528 mRFP+ cells from 150 *Tg(hcrt:mRFP, qrfp:EGFP)* double heterozygous embryos. To verify the fidelity of FACS, we visually screened for fluorescence in sorted cells (Fig. 1F). In the sorted *qrfp:EGFP* population, we observed EGFP in 99/117 cells (85%) but no mRFP (0/117). In the sorted *hcrt:mRFP* population, we observed mRFP in 110/146 cells (75%) but no EGFP (0/146). These values likely underestimate the purity of the sorted cells because FACS is more sensitive than visual inspection.

We extracted total mRNA from each cellular fraction and used cDNA microarrays to compare gene expression in Hcrt and QRFP neurons. We also compared gene expression in Hcrt neurons to expression in neurons labeled by a pan-neuronal marker, *Tg(elavl3:EGFP)*, and to expression in subtypes of sensory neurons: *Tg(trpa1b:EGFP)*, *Tg(isl1:Gal4VP16, 14xUAS:EGFP)*, and *Tg(p2rx3b:EGFP)*. *elavl3*

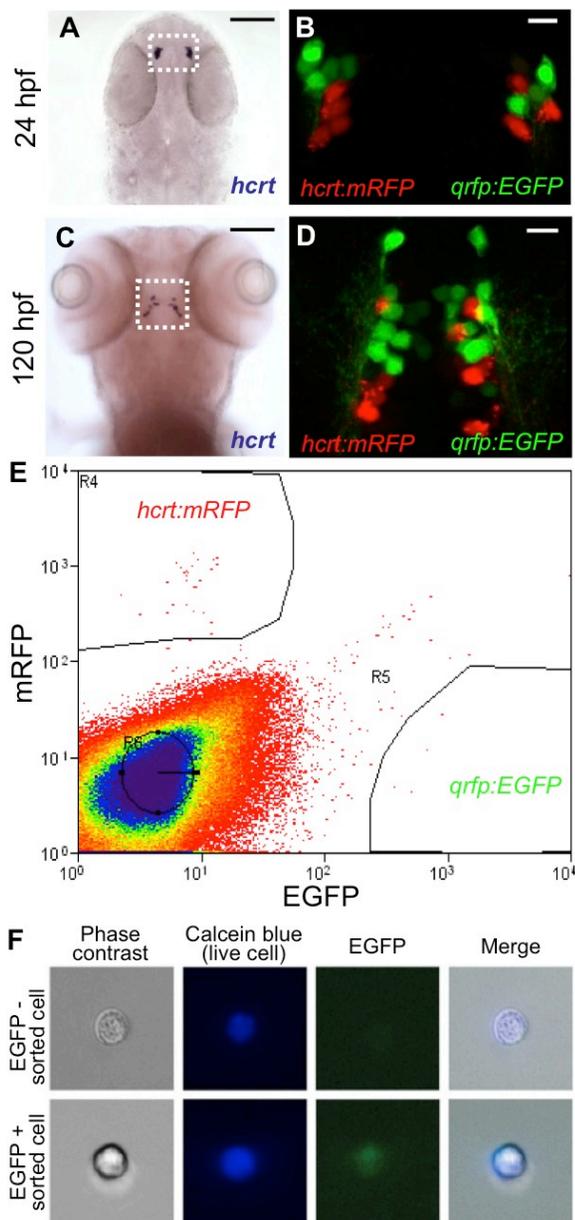


Figure 1. Isolation of *hcr*- and *qrfp*-expressing neurons from zebrafish embryos. (A-D) *hcr* is expressed in bilateral populations of 4-6 neurons at 24 hpf (A) and 10-15 neurons at 120 hpf (C). Fluorescence in *Tg(hcr:mRFP, qrfp:EGFP)* embryos is first observed at \sim 24 hpf. mRFP- and EGFP-labeled neurons are intermingled, but the markers are never co-expressed in the same cell (B, D). Boxed regions in (A, C) are shown at higher magnification in (B, D). Scale = 100 μ m (A, C) and 10 μ m (B, D). (E) *Tg(hcr:mRFP, qrfp:EGFP)* embryos were dissociated into single cells at 26 hpf and mRFP- and EGFP-expressing cells were isolated by FACS. The proportion of mRFP-positive and EGFP-positive cells was consistent with the number of Hcr and QRFP neurons, respectively, in a 26 hpf embryo. (F) EGFP is observed in a sorted *qrfp:EGFP*+ cell.

encodes an RNA binding protein that is expressed in most post-mitotic neurons (Park et al., 2000). *trpa1b* encodes a transient receptor potential (TRP) channel that is activated by chemical irritants (Prober et al., 2008). *p2rx3b* encodes an ATP-gated ion channel in non-peptidergic nociceptors (Kucenas et al., 2006). *islet1* (*isl1*) encodes a LIM/homeobox transcription factor that is expressed in sensory neurons and motoneurons (Higashijima et al., 2000). We used an *isl1* enhancer that drives expression in a subset of sensory neurons (Sagasti et al., 2005). The *Tg(trpa1b:EGFP)* and *Tg(isl1:Gal4VP16, 14xUAS:EGFP)* lines express EGFP in largely non-overlapping subsets of trigeminal and Rohon-Beard sensory neurons (Pan et al., 2012). *p2rx3b* is expressed in all cells labeled in *Tg(trpa1b:EGFP)* embryos and in a quarter of cells labeled in *Tg(isl1:Gal4VP16, 14xUAS:EGFP)* embryos. These samples allowed five separate pairwise comparisons of Hcrt neurons with different purified neuron populations (Fig. S2), which provided our study with greater statistical power than previous studies that compare Hcrt neurons to a single outgroup (Cvetkovic-Lopes et al., 2010; Dalal et al., 2013). An additional study examined changes in gene expression across multiple brain regions after the onset of narcolepsy (Honda et al., 2009); however, this approach is unlikely to identify developmentally-relevant transcripts. We focused on 19 highly ranked genes that encode transcription factors or secreted proteins (Table 1, Table 2), as both classes of proteins have well-established roles in neural development (Blackshaw et al., 2010; Wilson and Houart, 2004). As expected, *hcr1* was the most highly enriched gene in Hcrt neurons (Table 2). The complete microarray dataset is available through ArrayExpress, accession number E-MTAB-3317.

Table 1. Transcription factors enriched in Hcrt neurons.

Gene	GenBank Accession No.	Fold Increase*	ZFIN <i>in situ</i> available	Expression pattern at 24 hpf	Possible expression in Hcrt neurons
<i>hmx2</i>	NM_001115098	120	No	Hypothalamus [‡] , spinal cord, lateral line	Yes
<i>hmx3</i>	NM_131634	106	Yes	Hypothalamus, telencephalon, hindbrain, spinal cord, ear, lateral line	Yes
<i>npas4</i>	NM_001045321	106	No	Brain ubiquitous	Yes
<i>lhx9</i>	AB188254	99	No	Hypothalamus, telencephalon, hindbrain, spinal cord	Yes
<i>sox1a</i>	NM_001002483	70	Yes	Hypothalamus, telencephalon, hindbrain, lens, lateral line	Yes
<i>nr4a2</i>	NM_001002406	36	Yes	Telencephalon, hindbrain	No
<i>mybbp1a</i>	NM_001002042	25	Yes	Hypothalamus, telencephalon, hindbrain, retina, myotomes	Yes
<i>rfx2</i>	NM_001013278	22	Yes	Hypothalamus, telencephalon, spinal cord, pronephric ducts	Yes
<i>tshz2</i>	NM_173485	21	No	Hypothalamus, telencephalon, spinal cord, lateral line	Yes
<i>bhlhe40</i>	NM_212679	17	Yes	Retina, epiphysis, myotomes, neural crest	No
<i>tead1</i>	NM_212847	15	No	Hypothalamus, telencephalon, hindbrain	Yes
<i>cited2</i>	NM_001044982	14	No	Hypothalamus, telencephalon, hindbrain	Yes
<i>ets2</i>	NM_001023580	12	Yes	Hypothalamus, telencephalon, endothelium	Yes
<i>cbx4</i>	BC171352	10	No	Ventral hypothalamus, telencephalon, hindbrain	No

*Fold increase indicates the expression level of a gene in Hcrt neurons relative to its expression in the other neuronal subtypes analyzed by microarray.

[‡]Genes expressed in the hypothalamus may be expressed in a broad or restricted pattern.

Table 2. Secreted proteins enriched in Hcrt neurons.

Gene	GenBank Accession No.	Fold Increase*	ZFIN <i>in situ</i> available	Expression pattern at 24 hpf	Possible expression in Hcrt neurons
<i>hcrt</i>	NM_001077392	248	No	Hypothalamus	-
<i>cart4</i>	NM_001082932	130	No	Hypothalamus, ventral hindbrain	Yes
<i>igsf21</i>	NM_001034184	85	Yes	Hypothalamus, ventral hindbrain	Yes
<i>trh</i>	NM_001012365	45	No	Hypothalamus, ventral hindbrain	Yes
<i>endouc</i>	NM_001044974	43	No	Hypothalamus, telencephalon, lens, posterior midbrain, hindbrain, lateral line	Yes
<i>gliprlb</i>	NM_200575	19	No	Hypothalamus, telencephalon, hindbrain	Yes

*Fold increase indicates the expression level of a gene in Hcrt neurons relative to its expression in the other neuronal subtypes analyzed by microarray.

3.4 Expression patterns of candidate genes validate microarray results

High quality *in situ* hybridization (ISH) images were available in the ZFIN database for 7 candidate genes (Fig. S3). We determined the expression patterns of the remaining 12 genes using ISH on 24 hpf embryos (Fig. 2, S3). We found that 11 of 14 genes encoding transcription factors and 5 of 5 genes encoding secreted proteins are expressed in a similar or overlapping domain to Hcrt neurons. Some genes, such as the transcription factor *lhx9*, are expressed in all Hcrt cells throughout early development (Fig. S4). The microarray therefore accurately reported the expression of most candidate genes.

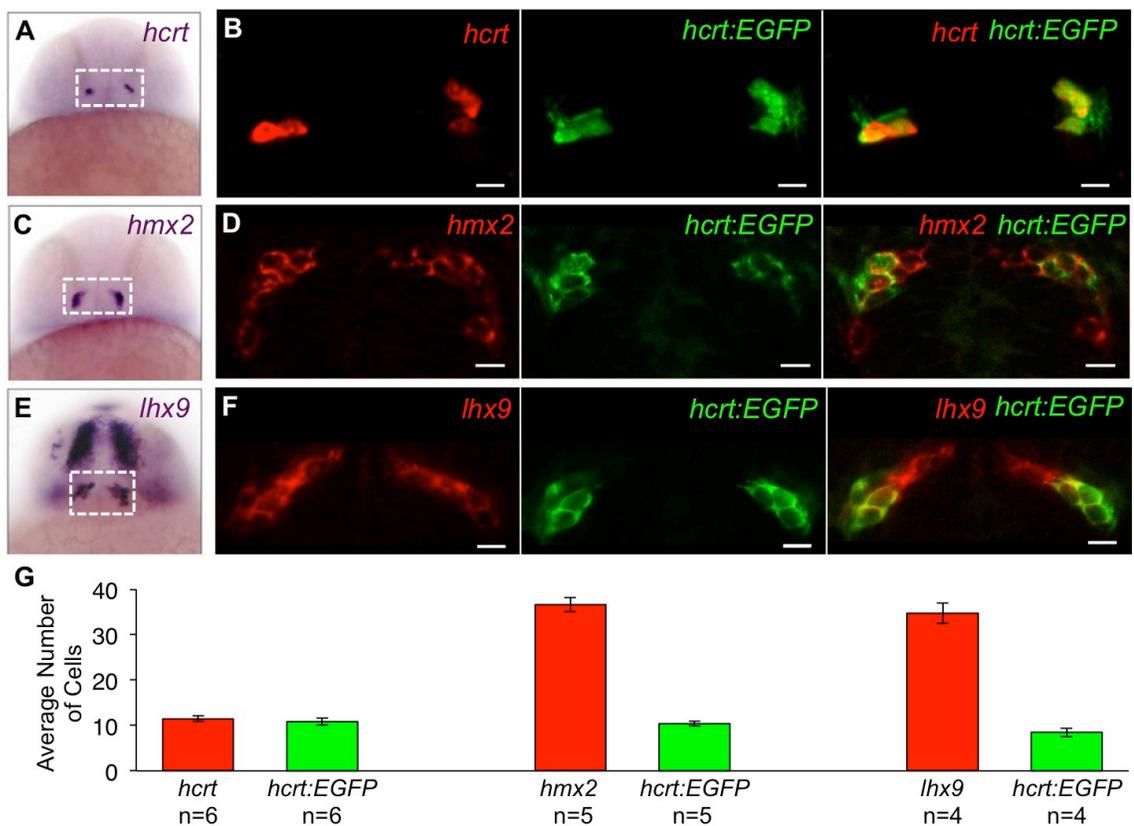


Figure 2. Examples of expression patterns of Hcrt neuron-enriched genes. (A, C, E) ISH performed on 24 hpf embryos using probes specific for *hcr*, *hmx2*, and *lhx9*. (B, D, F) A 1.25 μ m confocal section of fluorescent ISH (red) and *hcr:EGFP* immunofluorescence (green). All *hcr:EGFP* neurons express *hmx2* and *lhx9*. *hmx2* expression extends ventrally (C, D) and *lhx9* is expressed in broader domains of the telencephalon and diencephalon (E, F), as well as the hindbrain and spinal cord (not shown). (G) Mean \pm s.e.m. number of cells in white boxed regions in (A,C,E). n indicates number of brains. Scale = 10 μ m.

3.5 Lhx9 is sufficient to specify Hcrt neurons

To determine whether any candidate gene is sufficient to induce specification of Hcrt neurons, we cloned each gene downstream of a heat shock inducible promoter in a vector containing Tol2 transposase sites, and injected the plasmid together with *tol2 transposase* mRNA into zebrafish embryos at the 1-cell stage. As a result, the plasmid inserts into the genome of a random subset of cells. We transiently overexpressed each gene by performing a heat shock at 24 hpf (Fig. 3A). This approach provides an efficient, inducible method for overexpressing different candidate genes in 10-20% of cells (Fig. 3C). Quantification of hypothalamic Hcrt neurons using ISH at 120 hpf showed no significant differences between larvae overexpressing candidate genes and controls (Fig. S5). However, 14% of larvae overexpressing *lhx9* contained additional, ectopic Hcrt expression in the medial hindbrain (7/50 larvae, Fig. 3E,F). These larvae had 2.7 ectopic Hcrt cells on average, a 17% increase in the total number of Hcrt neurons. No other candidate gene was sufficient to specify ectopic Hcrt neurons. Like endogenous hypothalamic Hcrt neurons, ectopic Hcrt neurons expressed *vesicular glutamate transporter 1* (*vglut1*) and *prodynorphin* (*pdyn*) (Fig. 4). *vglut1* is widely expressed in the hypothalamus and hindbrain; however, its expression is particularly strong in Hcrt neurons (Fig. 4A,B). Hcrt neurons also exhibit intense, punctate *pdyn* labeling (Fig. 4D,E) that likely indicates sites of transcription (Hanisch et al., 2012; Kosman et al., 2004). We did not detect significant expression of *vglut2a* or *vglut2b* in Hcrt neurons (Figs S8, S9). *lhx9* overexpression generated more cells in the medial hindbrain with strong *vglut1* expression or punctate *pdyn* expression than ectopic Hcrt cells (Fig. 4C, F), suggesting that Lhx9 can specify multiple cell types. Indeed, most *lhx9*-overexpressing

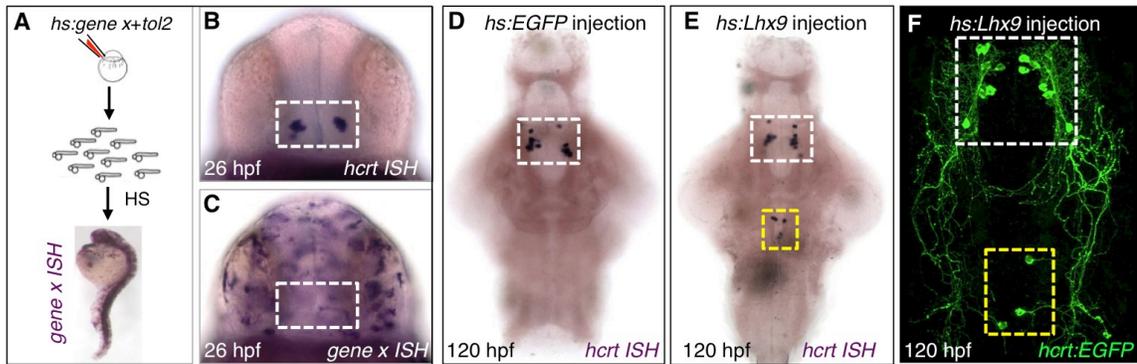


Figure 3. Transient *lhx9* overexpression induces ectopic Hcrt neurons. (A) Zebrafish embryos were injected with *tol2* transposase mRNA and a plasmid in which a heat shock (HS)-inducible promoter regulates the expression of a candidate gene (*hs:gene x*). (B, C) Anterior views of 26 hpf embryos after ISH show endogenous *hcrt* expression (B) and mosaic expression of *gene x* one hour after heat shock (C). Approximately 10-20% of cells overexpress *gene x*. White box in (B,C) indicates the hypothalamus. (D) Control embryos injected with a *hs:EGFP* plasmid exhibit normal *hcrt* expression at 120 hpf. (E) Embryos overexpressing *lhx9* contain ectopic Hcrt cells that are dorsal and caudal relative to endogenous Hcrt neurons. (F) Ectopic Hcrt cells exhibit neuronal morphology, visualized using *Tg(hcrt:EGFP)* larvae. White and yellow boxes in (D-F) indicate endogenous and ectopic Hcrt cells, respectively.

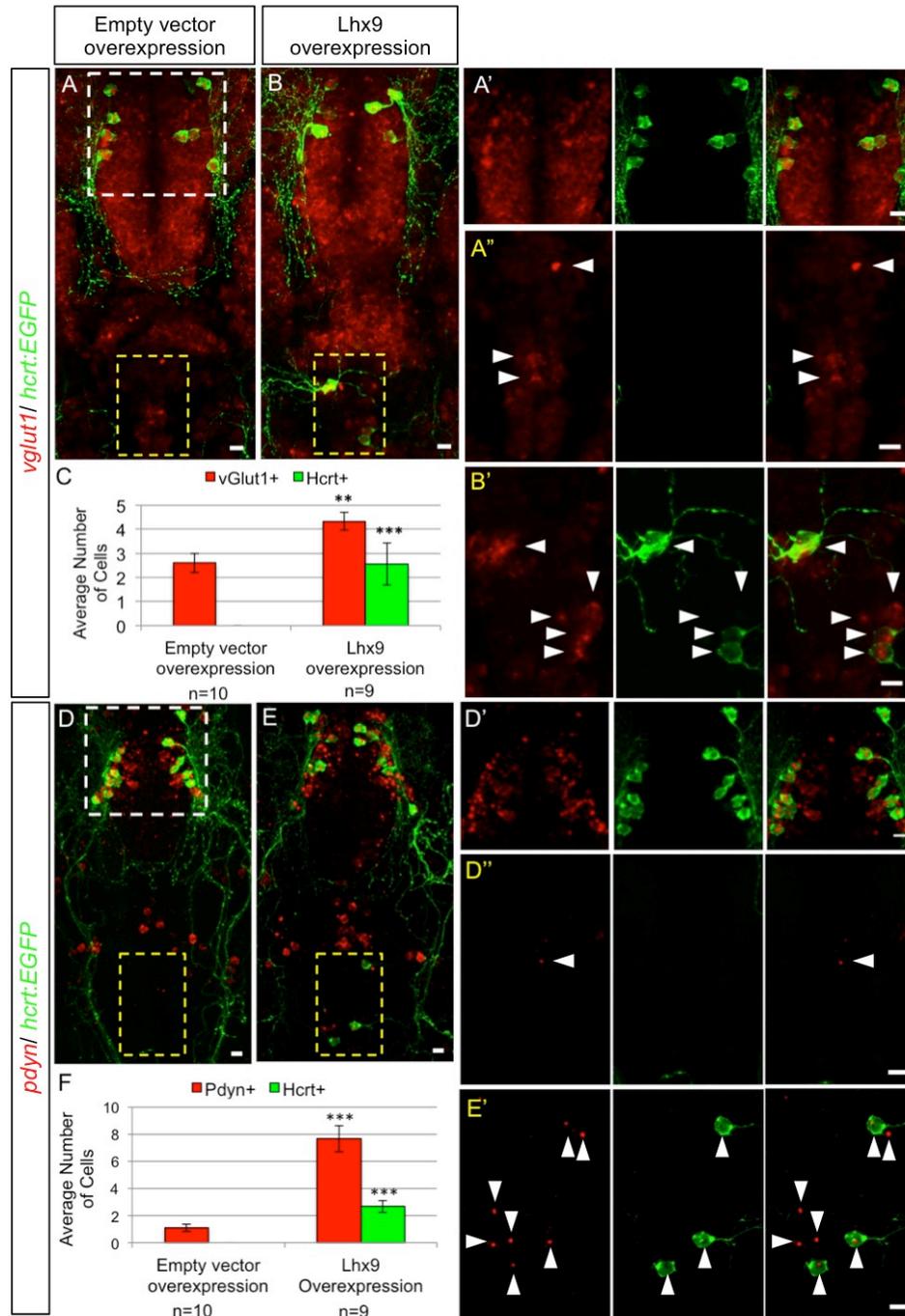


Figure 4. Hypothalamic and ectopic Hcrt neurons share biomarkers. Confocal projections of 120 hpf *Tg(hcrt:EGFP)* larval brains containing endogenous and ectopic Hcrt neurons labeled with an anti-EGFP antibody and fluorescent ISH probes specific for *vglut1* (A,B) and *pdyn* (D,E). White and yellow boxes in (A,D) indicate endogenous Hcrt neurons in the hypothalamus and endogenous *vglut1* and *pdyn* expression in the hindbrain, respectively, shown enlarged in (A',D') and in (A'',D''). Yellow boxes in (B,E) indicate ectopic Hcrt neurons, as enlarged in (B',E'). All hypothalamic and ectopic Hcrt neurons express *vglut1* and *pdyn*. Larvae injected with *hs:lhx9* and heat shocked at 24 hpf contain ectopic Hcrt neurons in the hindbrain and more cells with strong *vglut1* (B, B') and punctate *pdyn* expression (E, E'), compared to controls injected with an empty heat shock vector (A, A'', D, D''). White arrowheads indicate cells with strong *vglut1* or *pdyn* expression. Mean \pm s.e.m. number of cells in yellow boxed regions are quantified in (C) and (F). n indicates number of brains quantified. **, $p < 0.01$ and ***, $p < 0.001$.

neurons also expressed *qrfp* at 1 hour post heat shock (data not shown), and we observed an average of 3 ectopic QRFP neurons in the medial hindbrain 96 hours after *Lhx9* overexpression (8/40 larvae, Fig. S6; note that *lhx9* was enriched in *hcrt* neurons compared to *huc*, *isl1*, *trpa1b*, and *p2x3b* neurons, but not compared to *qrfp* neurons, in the microarray analysis). Similar to *Hcrt* neurons, both hypothalamic and ectopic QRFP neurons express *Lhx9* (Figs. S6, S7). Ectopic QRFP neurons did not co-express *hcrt* (0/24 QRFP neurons), suggesting that they remain developmentally distinct populations.

Both mammalian and zebrafish *Hcrt* neurons project to several brain regions, including the noradrenergic locus coeruleus (LC) (Horvath et al., 1999; Prober et al., 2006). To determine whether ectopic *Hcrt* neurons project to this endogenous *Hcrt* neuron target, we overexpressed *lhx9* in transgenic zebrafish that expressed the photoconvertible fluorescent protein Kaede in *Hcrt* neurons and EGFP in *dopamine beta-hydroxylase (dbh)*-expressing LC neurons (Fig. S1). We used a 405 nm laser to photoconvert *Hcrt* neurons in the hindbrain, but not in the hypothalamus, from green to red fluorescence. We observed that all red fluorescent ectopic *Hcrt* neurons project to the LC (15/15 neurons) (Fig. 5A,C). In contrast, stochastic labeling of neurons in the medial hindbrain with an *elavl3:Kaede* transgene indicates that only ~20% of cells in this region target the LC (9/50 neurons) (Fig. 5B,C). These experiments indicate that *Lhx9* is sufficient to specify *Hcrt* neurons *in vivo* that are genetically and anatomically similar to endogenous *Hcrt* neurons.

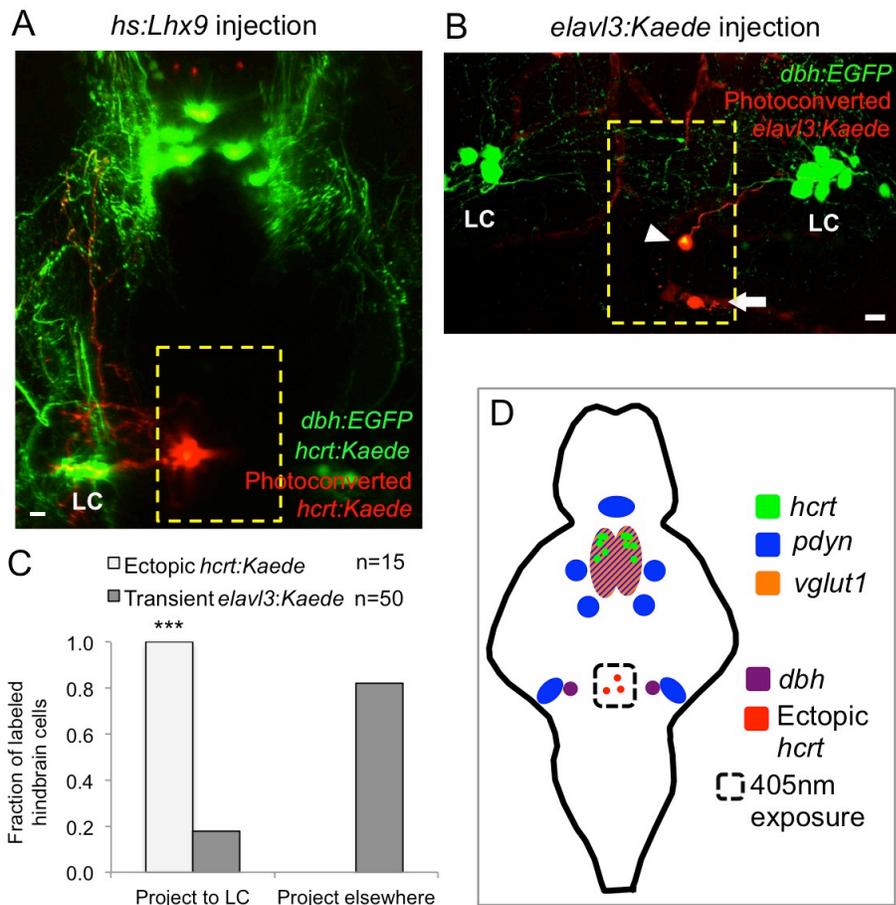


Figure 5. Ectopic Hcrt neurons project to the locus coeruleus. (A) Confocal projection of a 120 hpf *Tg(hcrt:Kaede, dbh:EGFP)* larva injected with a *hs:lhx9* plasmid shows that ectopic photoconverted Hcrt cells in the medial hindbrain (red) project to the locus coeruleus (LC). (B) Confocal projection of a 120 hpf *Tg(dbh:EGFP)* larva injected with an *elavl3:Kaede* plasmid to stochastically label neurons shows that some neurons in the medial hindbrain project to the LC (arrowhead) while others do not (arrow, axon projects orthogonally to the image). (C) All ectopic Hcrt neurons project to the LC but only 20% of *elavl3:Kaede* labeled cells in the same brain region project to the LC. n indicates number of neurons analyzed. ***, $p < 0.001$ compared to *elavl3:Kaede* by Fisher's exact test. Dashed yellow boxes in (A,B) indicate region used for quantification. Scale = 10 μ m. (D) Schematic of *hcr*, *pdyn*, and *vglut1* expression in the ventral zebrafish brain at 120 hpf. The location of ectopic Hcrt neurons, *dbh*-expressing LC neurons, and the photoconverted region are shown.

3.6 Lhx9 is necessary for Hcrt neuron specification

To test whether Lhx9 is necessary for Hcrt neuron specification, we used a splice-blocking morpholino to knock down *lhx9* expression. RT-PCR revealed robust inhibition of *lhx9* mRNA splicing, and thus production of functional Lhx9 protein, up to 72 hpf (Fig. S10). Morphants had approximately 40% fewer Hcrt neurons by ISH (Fig. 6C), and the remaining Hcrt neurons had reduced *hcrt* expression compared to embryos injected with a 5 base pair mismatch control morpholino (Fig. 6A,B). Quantitative reverse transcription-PCR revealed that *lhx9* morphants express 62% less *hcrt* transcript than control morphants (s.e.m.=5%, n=3 replicates), confirming that Hcrt neurons that persist in *lhx9* morphants contain less *hcrt* transcript than controls. To determine whether the missing Hcrt neurons in *lhx9* morphants silenced *hcrt* expression or were absent, we examined the effect of *lhx9* knockdown on *vglut1* and *pdyn* (Fig. S11). *lhx9* morphants had no gross defects in *vglut1* expression compared to controls. However, the number of cells with intense punctate *pdyn* expression, which includes all Hcrt cells, was reduced by approximately 40%. The loss of Hcrt cells was not caused by nonspecific morpholino-induced apoptosis, as embryos co-injected with the apoptosis-suppressing *p53* morpholino (Robu et al., 2007) and the *lhx9* morpholino showed the same phenotype (Fig. 6C). Furthermore, *lhx9* morpholino-injected embryos stained with acridine orange, which labels apoptotic cells, showed no increase in apoptosis (Fig. S12A-E). The morpholino phenotype was partially rescued by co-injecting the *lhx9* morpholino with the *hs:lhx9* plasmid and performing a heat shock at 24 hpf (Fig. S12F,G), indicating that reduced Hcrt cells was a specific effect of *lhx9* knockdown. *hcrt* expression was weaker in rescued cells than in endogenous Hcrt neurons, presumably because rescued cells only

received a pulse of *lhx9*, while endogenous Hcrt neurons continuously express *lhx9* (Fig. S4). We also tested morpholinos against *hmx2* or *hmx3*, which were more highly enriched in Hcrt neurons than *lhx9* in our microarray analysis. While RT-PCR confirmed that these morpholinos were effective, they had no effect on Hcrt neuron specification (data not shown).

Some Hcrt neurons persisted in *lhx9* morphants, possibly due to incomplete *lhx9* knockdown (Fig. S10), or because *Lhx9* is only necessary to specify a subset of Hcrt neurons, as in mice (Dalal et al., 2013). To distinguish between these possibilities, we used the CRISPR/Cas9 system (Jao et al., 2013; Hwang et al., 2013) to introduce mutations in *lhx9*. We co-injected Cas9 protein (Gagnon et al., 2014) with a set of short guide RNAs (sgRNAs) that target *lhx9* into embryos at the 1-cell stage to generate biallelic mutations and a loss-of-function phenotype in injected animals (Jao et al., 2013). Embryos injected with *Cas9+lhx9* sgRNAs had 90% fewer Hcrt cells than embryos injected with Cas9 alone (Fig. 6D-F). Furthermore, over half of brain hemispheres of embryos injected with *Cas9+lhx9* sgRNAs completely lacked Hcrt cells, whereas embryos injected with Cas9 alone had at least 3 Hcrt cells in each brain hemisphere (Fig. 6G). This phenotype is unlikely due to off-target effects of particular sgRNAs, as we observed a similar, albeit weaker, phenotype in embryos injected with Cas9 and independent subsets of *lhx9* sgRNAs (Fig. S13). We conclude that *lhx9* is necessary to specify all Hcrt neurons in zebrafish embryos.

Because *lhx9* overexpression was sufficient to specify QRFP neurons, we asked whether *lhx9* is also necessary for QRFP neuron specification. We observed 63% fewer

QRFP neurons in embryos injected with Cas9+*lhx9* sgRNAs compared to Cas9 alone, and over 20% of brain hemispheres lacked QRFP neurons (Fig. S13). Injection of embryos with Cas9 and subsets of *lhx9* sgRNAs produced a similar, albeit weaker, phenotype. We conclude that *lhx9* is necessary to specify all QRFP neurons.

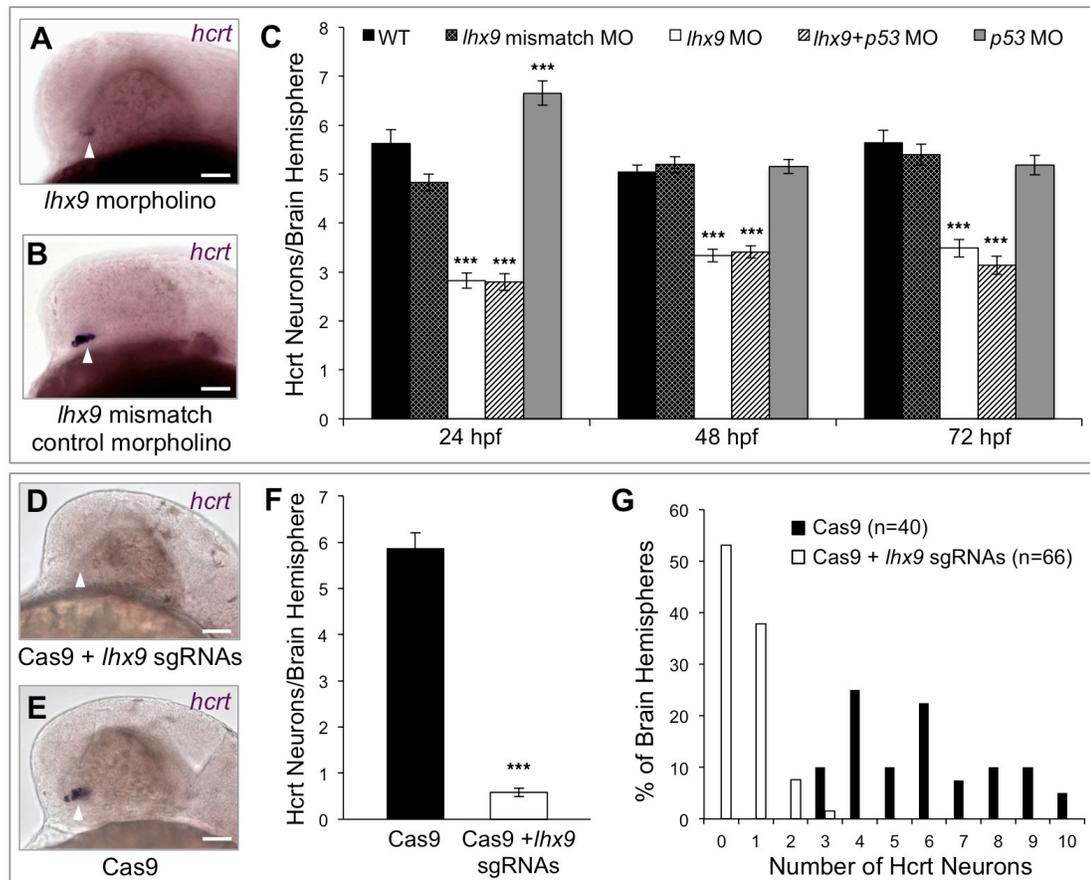


Figure 6. *Lhx9* is required for Hcrt neuron specification. (A-C) *hcrT* ISH at 24 hpf shows that morpholino-mediated knockdown of *lhx9* reduces the number of Hcrt neurons and the level of *hcrT* expression (A) compared to embryos injected with a control morpholino (B). (C) Quantification of Hcrt neurons per brain hemisphere at 24, 48, and 72 hpf confirms that *lhx9* morphants have ~40% fewer Hcrt neurons. Co-injecting a *p53* morpholino did not affect this phenotype. Mean \pm s.e.m. is shown. At least 22 embryos were quantified for each condition. ***, $p < 0.001$ compared to control morpholino by one-way ANOVA followed by Bonferroni's correction for multiple comparisons. (D-G) *hcrT* ISH at 24 hpf shows that co-injection of Cas9 protein and 10 *lhx9* sgRNAs eliminates *hcrT* expression (D) compared to embryos injected with Cas9 alone (E). (F) Quantification of Hcrt neurons per brain hemisphere at 24 hpf. Mean \pm s.e.m. is shown. ***, $p < 0.001$ compared to Cas9 alone by one-way ANOVA. (G) Histogram showing the percentage of brain hemispheres containing the indicated number of Hcrt neurons. n indicates the number of brain hemispheres analyzed in (F, G). White arrowheads indicate endogenous Hcrt neuron region. Scale = 50 μ m.

3.7 Lhx9 directly promotes *hcrt* expression

Since Lhx9 is a transcription factor, we hypothesized that it might promote *hcrt* expression directly. A previous study tested this hypothesis *in vitro*, by co-expressing *llhx9* and a mouse *hcrt* promoter-luciferase reporter in a neuroblastoma cell line, and *in vivo*, by lentiviral transduction of *llhx9* into the hypothalamus of adult mice, but observed no effect on *hcrt* expression (Dalal et al., 2013). However, when we performed double fluorescent ISH against *llhx9* and *hcrt* on zebrafish embryos fixed one hour after heat shock, we observed *hcrt* expression in almost all *llhx9*-overexpressing cells (Fig. 7A, Fig. S14). The number of ectopic *llhx9*- and *hcrt*-expressing cells is reduced at 8 hours after heat shock (Fig. S14D), and few ectopic cells are observed at 24 hours after heat shock (Fig. S14F), similar to the number observed at 120 hpf (Fig. S14H).

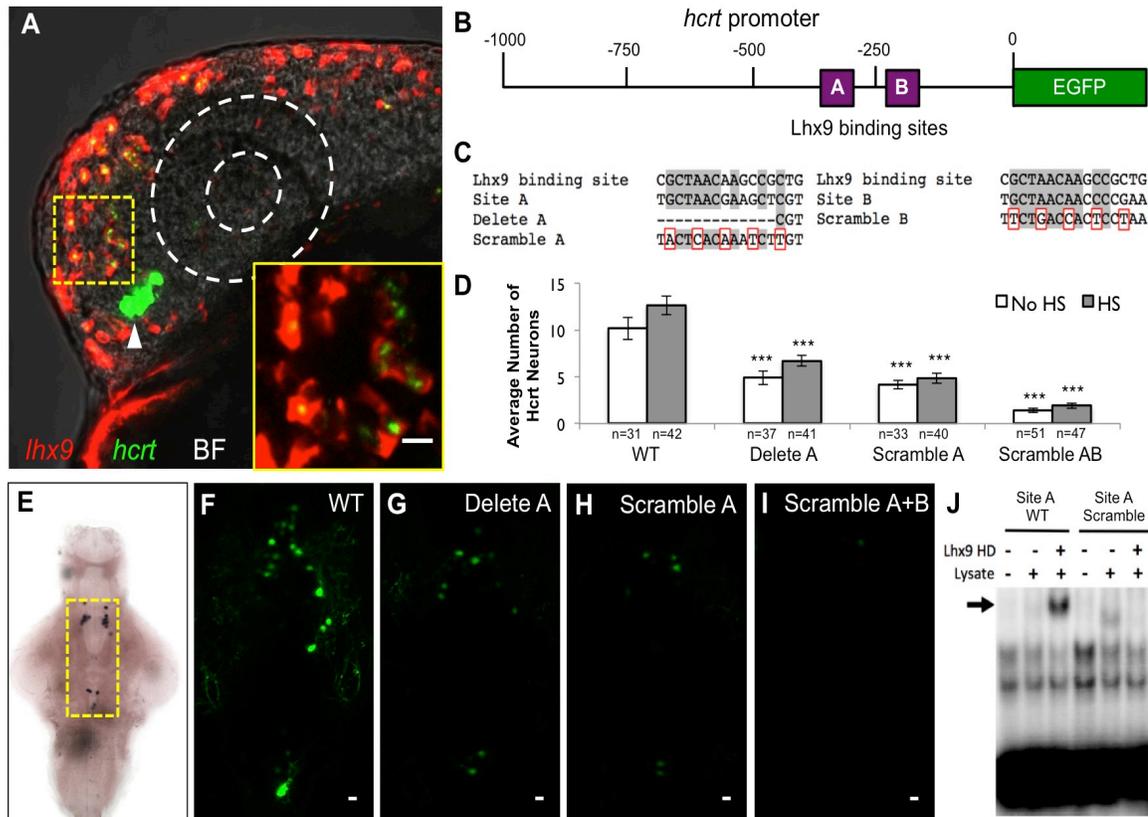


Figure 7. Lhx9 can directly induce *hcrt* expression. (A) Ectopic *lhx9*-expressing neurons also express *hcrt* in embryos injected with a *hs:lhx9* plasmid, fixed one hour after heat shock at 24 hpf, and analyzed using double fluorescent ISH with *hcrt*- and *lhx9*-specific probes. White arrowhead indicates endogenous Hcrt neurons. A single 1.5 μm confocal section is shown. The bright field (BF) overlay and dashed white circles show the position of the eye. The yellow-boxed region is shown at higher magnification in the inset. (B) Schematic diagram of the zebrafish 1 kb *hcrt* promoter, including putative Lhx9 binding sites A and B. (C) Sequence of a previously characterized mammalian Lhx9 binding site compared to sites A and B in the zebrafish *hcrt* promoter and in mutated constructs. Gray shading, dashes, and red boxes indicate conserved nucleotides, deleted nucleotides, and mutated nucleotides, respectively. (D-I) Embryos were injected with a plasmid containing both *hcrt:EGFP* and *hs:lhx9*, and some injected embryos were heat shocked at 24 hpf. Deletion or scrambling of putative Lhx9 binding sites reduced the number (D, F-I) and intensity (F-I) of EGFP-expressing cells. Cell counts indicate Hcrt cells per brain with (endogenous and ectopic Hcrt cells) and without (endogenous Hcrt cells only) heat shock. Mean \pm s.e.m. is shown. n indicates number of brains analyzed. ***, $p < 0.001$ compared to the wild-type promoter by one-way ANOVA followed by Bonferroni's correction for multiple comparisons. (E) Yellow box indicates area shown in (F-I). (J) EMSA showing that the Lhx9 homeodomain (HD) binds to the wild-type (WT) site A (arrow), but not the scrambled site A, *in vitro*. Scale bars = 10 μm .

Importantly, ectopic *hcrt*-expressing neurons also express ectopic *lhx9* at all time points. These findings support the hypothesis that Lhx9 can directly regulate *hcrt* expression.

A previous study identified an Lhx9 binding site upstream of the *wilms tumor 1* gene (Wilhelm and Englert, 2002), and we observed two similar sites in the zebrafish *hcrt* promoter (sites A and B) (Fig. 7B,C). Notably, site A corresponds to a region previously shown to be important for *hcrt* expression in zebrafish (Faraco et al., 2006). We tested whether these sites are necessary for endogenous (i.e., hypothalamic) and Lhx9-induced ectopic *hcrt* expression by injecting wild-type embryos with plasmids containing the *hcrt* promoter, in which one or both putative Lhx9-binding sites were mutated, placed upstream of EGFP. Each plasmid also contained a heat shock inducible *lhx9* transgene downstream of the EGFP reporter. Thus, any cell that contains the plasmid will have both the *hcrt*:EGFP reporter and the *hs:lhx9* transgene. Injected embryos were heat shocked at 24 hpf and analyzed at 120 hpf for hypothalamic and ectopic *hcrt*:EGFP expressing neurons (Fig. 7D, HS samples). Some injected embryos were not heat shocked (Fig. 7D, no HS samples) to determine whether the putative Lhx9 binding sites are required for EGFP expression in the endogenous Hcrt domain alone. Mutating site A, either by deletion or by scrambling every third nucleotide, reduced the number and intensity of endogenous and ectopic cells labeled with EGFP compared to the wild-type *hcrt* promoter (Fig. 7C-D, F-H). Scrambling the sequences of both sites A and B virtually abolished EGFP expression (Fig. 7D, I). Notably, ectopic *hcrt* cells were never observed for the double mutant reporter. These experiments indicate that sites A and B are crucial for *hcrt* expression *in vivo*.

To test whether Lhx9 can interact with these sites, we performed an electrophoretic mobility shift assay (EMSA) using the zebrafish Lhx9 homeodomain (Lhx9 HD) and radiolabeled oligonucleotides that include the wild-type or scrambled sequences for sites A and B. We found that Lhx9 HD binds to the wild-type site A probe, but not to the scrambled site A probe (Fig. 7J). We failed to observe an interaction between Lhx9 HD and the site B probe (data not shown), possibly due to non-optimal *in vitro* binding or electrophoresis conditions. This result indicates that Lhx9 can bind to site A *in vitro* and suggests that Lhx9 directly regulates *hcrt* expression *in vivo* by binding to this site.

3.8 *Lhx9* overexpression in mouse embryos induces Hcrt neuron specification

A previous study that overexpressed *Lhx9* in the hypothalamus of adult mice observed no effect on Hcrt neuron specification (Dalal et al., 2013). To determine whether Lhx9 can promote Hcrt neuron specification earlier in mammalian development, we used micro *in utero* electroporation (Matsui et al., 2011) to focally overexpress *EYFP* and the murine *Lhx9* ortholog, or *EYFP* alone, in the developing murine diencephalon at embryonic day 10.5 and assayed *Hcrt* expression at postnatal day 6. Embryos overexpressing *Lhx9* had significantly more *Hcrt*-expressing neurons in the lateral hypothalamus than embryos overexpressing *EYFP* alone (Fig. 8, Figs S15, S16). This effect appears to be specific, since *lhx9*-overexpression had no effect on the expression of other hypothalamic markers, including *Cartpt*, *Foxp2*, and *Gal* (data not shown). *Lhx9* overexpression in the subthalamic zona incerta (Fig. S17) or cerebral cortex (data not shown) did not induce ectopic *Hcrt* expression. We conclude that *Lhx9* is capable of promoting Hcrt neuron specification exclusively in its endogenous domain during mouse embryogenesis.

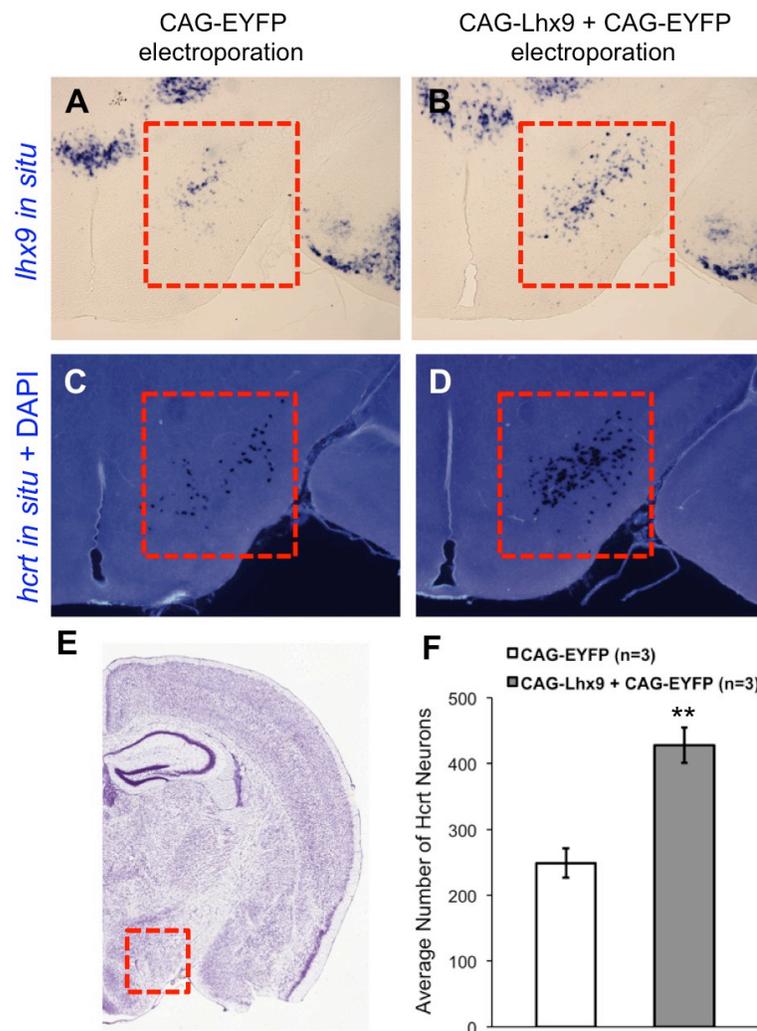


Figure 8. *Lhx9* overexpression in mouse embryos promotes *Hcrt* neuron specification.

(A-D) E10.5 mouse embryos were electroporated *in utero* into the right brain hemisphere lateral hypothalamus with either CAG-EYFP or CAG-Lhx9 + CAG-EYFP and analyzed at P6 for expression of *Lhx9* and *Hcrt* by ISH in serial coronal sections. Mice electroporated with CAG-Lhx9 + CAG-EYFP have increased *Lhx9* expression (B) and more *Hcrt* cells (D) in the lateral hypothalamus (red boxes) than controls electroporated with CAG-EYFP alone (A, C). Coronal sections shown in panels (A-D) are approximately Bregma -1.6mm. For reference, the entire right hemisphere of a comparable Nissl stained section is shown in (E), with the red box indicating the lateral hypothalamus. Image adapted from the Allen Mouse Brain Atlas (Lein et al., 2006). (F) Quantification of *Hcrt* cell number in the lateral hypothalamus of the right brain hemisphere following electroporation with CAG-Lhx9 + CAG-EYFP compared to CAG-EYFP alone. Mean \pm s.e.m. is shown for three experimental and three control brain hemispheres (see Fig. S15 for quantification details). **, $p < 0.01$ using Student's t-test.

3.9 Discussion

Using microarray gene expression analysis and high-throughput gene overexpression assays in zebrafish, we found that the LIM homeobox transcription factor *Lhx9* is both necessary and sufficient to specify Hcrt neurons in zebrafish, and is sufficient to specify Hcrt neurons in mouse embryos. We found that *Lhx9* is also necessary and sufficient to specify QRFP neurons in zebrafish, which are located adjacent to Hcrt neurons in the hypothalamus. To our knowledge, this is the first study to identify a factor that is capable of inducing the specification of these neurons, or of any terminal neural subtype in the lateral hypothalamus.

lhx9 was identified as enriched in Hcrt neurons by both our analysis of zebrafish embryos and by a previous study that used adult mice (Dalal et al., 2013). We analyzed zebrafish neurons just after the initiation of *hcr*t expression, enabling us to screen for transcripts that likely play a role in the specification of Hcrt neurons. We isolated purified cell populations by FACS, analyzed their gene expression patterns by microarray, and performed multiple pairwise comparisons of purified Hcrt neurons to a closely related cell type (QRFP-expressing neuron) and to more distantly related neurons (pan-neuronal or sensory neurons). In contrast, Dalal and colleagues used a translational profiling approach in which a tagged ribosomal subunit is expressed in Hcrt neurons. Biochemical purification of this subunit from a whole brain homogenate isolates transcripts that are actively translated in Hcrt neurons. While this approach allowed profiling of the rare Hcrt cell population, its statistical power was diminished by contamination with nonspecific transcripts, as indicated by the presence of glial transcripts. In addition to *lhx9*, the 112 genes most enriched in Hcrt neurons in our study and the 188 most enriched genes

identified by Dalal and colleagues included the definitive Hcrt neuron markers *hcrt* and *pdyn*, the transcription factors *hmx2* and *rfx4*, as well as *scg2*, *agrp*, *glipr1*, and *fam46a*. The absence of more overlapping genes can be attributed in part to ambiguity in the gene assignment of microarray probes and the imperfect annotation of the zebrafish genome. Furthermore, the stringent criteria for significance used by both studies likely underestimate the true complement of Hcrt enriched genes shared between zebrafish embryos and adult mice.

Lhx9 belongs to the LIM homeobox family of transcription factors that is conserved from invertebrates to mammals. These proteins have essential roles in tissue patterning and differentiation, particularly in the brain (Hobert and Westphal, 2000). In mice, several LIM homeobox proteins are expressed dynamically to demarcate regions of the developing hypothalamus (Shimogori et al., 2010). However, the developmental roles of specific LIM homeobox genes have been difficult to distinguish; loss-of-function phenotypes are subtle and similar LIM homeobox family members often exhibit redundancy. For instance, double knockdown of *lhx9* and *lhx2* dramatically altered thalamus and forebrain patterning in zebrafish, but knockdown of either gene alone had no gross effects (Peukert et al., 2011). Similarly, *lhx9* knockout mice survive to adulthood without gross brain defects (Birk et al., 2000).

Although we tested morpholinos against several candidate genes, only the *lhx9* morpholino decreased the number of Hcrt neurons, with an average decrease of 40%. The remaining Hcrt neurons, which expressed *hcrt* at reduced levels, likely result from incomplete *lhx9* knockdown. Indeed, co-injection of Cas9 protein with sgRNAs targeting *lhx9* completely abolished *hcrt* expression in over half of brain hemispheres analyzed,

indicating that *lhx9* is required for the specification of all Hcrt neurons. This result contrasts with *Lhx9* knockout mice, where Hcrt neurons are only reduced by 39% (Dalal et al., 2013). This discrepancy is likely due to the expression of *lhx9* in all Hcrt neurons in zebrafish but only a subset in mice (Shimogori et al., 2010).

In zebrafish, *lhx9* overexpression at 24 hpf was sufficient to produce ectopic Hcrt neurons in the medial hindbrain, but the number of Hcrt neurons in the hypothalamus remained unchanged. We characterized these ectopic Hcrt neurons at 120 hpf, when the Hcrt neuronal circuit is functional (Prober et al., 2006; Elbaz et al., 2012), and confirmed that all zebrafish Hcrt neurons in the hypothalamus and hindbrain express *pdyn* and *vglut1*, two markers of mammalian Hcrt neurons (Chou et al., 2001; Rosin et al., 2003). Unlike previous studies (Appelbaum et al., 2009; Rosin et al., 2003), we did not observe significant expression of *vglut2a* or *vglut2b* in hypothalamic or hindbrain Hcrt neurons, indicating that Hcrt neurons in larval zebrafish express different VGLUT family genes than adult zebrafish or rats. We also observed that all ectopic Hcrt neurons project to the LC, a target of Hcrt neurons in zebrafish and mammals (Horvath et al., 1999; Prober et al., 2006). Thus, despite their location in the hindbrain, *Lhx9*-induced ectopic Hcrt neurons express the same genetic markers and project to the same target as endogenous Hcrt neurons.

While it may be surprising that we did not detect an effect of *lhx9* overexpression on the number of hypothalamic Hcrt neurons, there are several possible explanations for this result. First, because *lhx9* induces few ectopic Hcrt neurons in our assay and the number of endogenous Hcrt neurons is variable, it is possible *lhx9* induced additional Hcrt neurons in the hypothalamus but the difference was not large enough to detect.

Second, endogenous *lhx9* likely acts in concert with other factors to specify Hcrt neurons because only a subset of endogenous *lhx9*-expressing cells expresses *hcrt*. If these co-factors are only present in hypothalamic neurons that express endogenous *hcrt*, *lhx9* overexpression will have no effect on the number of Hcrt neurons in the hypothalamus. Other genes identified by our microarray analysis might encode these co-factors. However, overexpressing two or more candidate genes using our assay was not feasible due to DNA toxicity. Reducing the concentration of each injected plasmid to offset this toxicity also reduces the extent of gene overexpression. We were thus unable to observe ectopic Hcrt neurons when the *hs:lhx9* plasmid was co-injected with a second plasmid, and the number of hypothalamic Hcrt neurons was unaffected (data not shown).

We detected widespread co-expression of *lhx9* and *hcrt* one hour after heat shock-induced *lhx9* overexpression. Because ectopic *hcrt* expression was observed so soon after heat shock, it seemed likely that Lhx9 was directly inducing *hcrt* expression. Consistent with this hypothesis, we identified two putative Lhx9 binding sites in the zebrafish *hcrt* promoter and found that they are necessary for both endogenous and ectopic *hcrt* expression *in vivo*. We also found that one binding site can form a complex with the Lhx9 homeodomain *in vitro*. However, this hypothesis is complicated by the fact that *lhx9* is widely expressed in the embryonic zebrafish brain, while *hcrt* is normally expressed exclusively in the hypothalamus. Because the extent of both *lhx9* overexpression and ectopic *hcrt* expression is dramatically reduced by 24 hours after heat shock (Fig. S14), we propose that the ability of Lhx9 to drive *hcrt* expression may depend on Lhx9 levels. High Lhx9 levels might directly induce *hcrt* expression, while

endogenous *Lhx9* levels require one or more co-factors in the hypothalamus and medial hindbrain to promote and maintain *hcrt* expression.

Because the expression patterns of *hcrt* and *lhx9* are conserved between zebrafish and mammals (Peukert et al., 2011; Shimogori et al., 2010), we tested whether *Lhx9* overexpression could induce *Hcrt* neuron specification in mice. Indeed, *Lhx9* overexpression by *in utero* electroporation was sufficient to specify additional *Hcrt* neurons, although they were only observed in the endogenous *Hcrt* neuron domain. This result suggests that the zone of cells competent to specify *Hcrt* neurons is more spatially restricted in mice than in zebrafish. Our findings differ from a previous study that saw no change in *Hcrt* expression after viral transduction of an *Lhx9* overexpression construct in the hypothalamus of adult mice (Dalal et al., 2013). These results suggest that *Lhx9* can induce *Hcrt* neuron specification in the embryonic, but not adult, mouse hypothalamus, possibly because cells competent to specify *Hcrt* neurons are fully differentiated in adults.

Our study demonstrates the utility of zebrafish to identify and test genes that regulate vertebrate development. Furthermore, the ability of *Lhx9* to induce *Hcrt* neuron specification suggests a therapeutic approach to compensate for the loss of *Hcrt* neurons that is thought to cause narcolepsy. This strategy would use *Lhx9* to generate *HCRT*-expressing neurons from human pluripotent stem cells *in vitro*, followed by screening and selection of *Hcrt* neurons to be transplanted into the hypothalamus. The promise of this approach is highlighted by the recent demonstration that narcoleptic-like sleep induced by lesion of *Hcrt* neurons in rats is diminished by the transplantation of *Hcrt* neurons into

the lateral hypothalamus (Arias-Carrión and Murillo-Rodríguez, 2014).

3.10 Experimental Procedures

Ethics statement

Zebrafish experiments followed standard protocols (Westerfield, 1993) in accordance with Caltech Institutional Animal Care and Use Committee guidelines. Mouse procedures were approved by the RIKEN Institutional Animal Care Committee.

Transgenic zebrafish

A 1-kb fragment of zebrafish genomic DNA upstream of *qrfp* was cloned upstream of *EGFP* using primers 5'-CTGACTCTCCCATCAGTCCT-3' and 5'-CTGAAATTTAAGGAATAATTTAAAGTTG-3'. A 1-kb zebrafish genomic fragment upstream of *hcrt* (Faraco et al., 2006) was subcloned upstream of *mRFP* and *Kaede* using primers 5'-ATAATAAATAAATCTGATGGGGTTTT-3' and 5'-GAGTTTAGCTTCTGTCCCC-TG-3'. The *dbh:EGFP* transgene was generated by cloning a 1.1-kb fragment upstream of zebrafish *dbh* using primers 5'-ACTTGAACCAGCGACCTTCT-3' and 5'-GGTTTGAAGGCCTTTCTAAGTTTTT-3'. Transgenes were co-injected with *tol2* transposase mRNA to generate stable transgenic lines. The *Tg(hcrt:EGFP)*, *Tg(elavl3:EGFP)*, *Tg(trpa1b:EGFP)*, *Tg(isl1:Gal4VP16, UAS:EGFP)*, *Tg(p2rx3b:EGFP)*, and *Tg(vglut2a:RFP)* lines have been described (Prober et al., 2006; Park et al., 2000; Pan et al., 2012; Sagasti et al., 2005; Kucenas et al., 2006; Koyama et al., 2011). Hindbrain neurons were stochastically labeled by injecting a *Tg(elavl3:Kaede)* transgene (Sato et al., 2006).

Microarray analysis

We analyzed embryos co-expressing *qrfp:EGFP* and *hcrt:mRFP*, as well as separate transgenic lines expressing EGFP in all neurons [*Tg(elavl3:EGFP)*] or in subsets of sensory neurons [*Tg(trpa1b:EGFP)*, *Tg(isll:Gal4VP16, 14xUAS:EGFP)*, *Tg(p2rx3b:EGFP)*]. Dechorionated embryos were anesthetized with tricaine at 26 hpf, incubated at room temperature in 0.25% trypsin-EDTA (Life Technologies), and manually dissociated. The cells were passed through a 40 μ m strainer, pelleted, and resuspended in cell culture medium for FACS (Manoli and Driever, 2012). Cells were incubated with Calcein Blue-AM (Life Technologies) to ensure the sorting of live cells. Sorting gates were set using wild-type embryos. Cells expressing EGFP or mRFP formed clear populations that were visually confirmed. Cells were sorted into a final volume of 100 μ l PicoPure XB lysis buffer (Arcturus) and incubated at 42°C for 30 min. At least two independent sorting experiments were performed using each transgenic line. RNA was isolated (Arcturus PicoPure RNA, Life Technologies), amplified (MessageAmp II aRNA, Ambion), and quantified by bioanalyzer. cDNA libraries were prepared in duplicate or triplicate and labeled with Cy3 or Cy5 for hybridization to NimbleGen zv7 microarrays. Fluorophores were switched between replicates to minimize labeling bias. After scanning, data were normalized and pairwise comparisons were performed in R between Hcrt neurons and the five other sorted neuron populations. For each comparison, we generated a list of the most significantly Hcrt-enriched genes by intersecting the subset of probes that were at least fourfold upregulated in Hcrt neurons with the subset of probes with the most statistically significant differential expression using Bayesian

statistics. The five pairwise comparisons were then intersected to identify probe sets consistently enriched in Hcrt neurons.

Zebrafish ISH and immunohistochemistry

Samples were fixed in 4% paraformaldehyde for 12-16h at room temperature. ISH was performed using digoxigenin (DIG)-labeled antisense riboprobes (Thisse and Thisse, 2008). Images were acquired using a Zeiss Axio ImagerM1 microscope. Fluorescent ISH used DIG- and 2,4-dinitrophenol (DNP)-labeled antisense riboprobes with the TSA Plus DNP System (PerkinElmer). Rabbit polyclonal anti-GFP (1:1000; MBL International, #598), rabbit polyclonal anti-orexin-A (Hcrt) (1:1000; Millipore, #AB3704) and goat anti-rabbit Alexa488 (1:500; Invitrogen, #A-11008) antibodies were used. Images were acquired using a Zeiss LSM 780 confocal microscope and analyzed using Fiji (Schindelin et al., 2012).

Candidate gene overexpression

The coding sequence of each candidate gene was amplified from 24 hpf zebrafish cDNA and cloned into pENTR-D-TOPO (Invitrogen). Gateway recombination (Invitrogen) was used to clone each gene downstream of a heat shock-inducible promoter (Halloran et al., 2000), and the entire cassette was flanked by Tol2 transposase sites. We co-injected individual overexpression plasmids with *tol2* transposase mRNA into zebrafish embryos at the 1-cell stage. Gene overexpression was induced by incubating embryos in a 37°C water bath for 1 h.

Morpholino-mediated knockdown

Morpholinos (GeneTools) were injected into wild-type embryos at the 1-cell stage. We used a splice-blocking morpholino to knock down *Lhx9* (5'-AGCCTCAAAGTTAATGCTTACCTGT-3'). A morpholino with a 5 bp mismatch to the target was injected as a negative control (5'-AGCGTGAAACTTAATCCTTACCTCT-3'). Potential apoptosis was suppressed by co-injecting a *p53* morpholino (5'-GCGCCATTGCTTTGCAAGAA-TTG-3'). To verify knockdown efficacy, we isolated RNA from pools of five injected embryos and used RT-PCR (Superscript III First-Strand Synthesis System, Invitrogen) to amplify a fragment of the *lhx9* transcript that spans exon 2. To detect apoptosis, 24-hpf embryos were bathed in 1 µg/ml Acridine Orange for 1 h at room temperature, followed by three 10 min washes with E3 medium. Splice-blocking morpholinos were designed for *hmx2* (5'-TGGGAACGTCACCTCACCGAGACAGA-3') and *hmx3* (5'-TGCTGCTACAGTAATAGAGGCCAAA-3') to retain the first intron of each gene, resulting in an early stop codon.

Quantitative reverse-transcription PCR (qRT-PCR)

We isolated total RNA from three biological replicates (25 embryos each) of *lhx9* morpholino-injected and control morpholino-injected embryos. TURBO DNase I was used to remove genomic DNA (TURBO DNA-free Kit, Invitrogen). We then generated cDNA (Superscript III First-Strand Synthesis System, Invitrogen) and amplified *hcrt* transcripts with primers 5'-GAGCATCAAGACTTTTCGATACA-3' and 5'-ATGAAG-ACGAGCACCTGGAG-3'. Transcripts of the *rpl13a* reference gene were amplified with primers 5'-TCTGGAGGACTGTAAGAGGTATGC-3' and 5'-AGACGCACAATCTTG-

AGAGCAG-3'. Each qRT-PCR reaction was run in triplicate on an ABI PRISM 7900HT Sequence Detection System (Applied Biosystems). Relative fold-change in expression was calculated using the $2^{-\Delta\Delta C_t}$ method (Schmittgen and Livak, 2008).

CRISPR/Cas9

We designed ten sgRNAs (sgRNAs 1-10) to target the following sites, respectively, within the *lhx9* open reading frame: 5'-GCTTTTCCACGGCATCTCTGGGG-3', 5'-AGGCTTCTCTGGGCTCATGGAGG-3', 5'-TGGCAGCATTACTGCAAAGAGG-3', 5'-CGGATATGCCGAGGTGGCAGCGG-3', 5'-GCGCATCACCATTTCGAAGCGG-3', 5'-GCTGAATTAGCGCCAAAGGCGG-3', 5'-AGAGGAAGAGTCCGGCGATGGGG-3', 5'-AGGTTGTAATGAGAACGATGCGG-3', 5'-CGCTTGGTCTTC-TGGGAAGGAGG-3', 5'-AGTCCTTGGCATCGGGGTTGTGG-3'. Cas9 protein was mixed with all ten sgRNAs and injected into embryos at the 1-cell stage (Gagnon et al., 2014). At 24 hpf, deformed embryos were removed and the remainder were fixed in 4% paraformaldehyde for ISH. To control for potential sgRNA off-target effects, we co-injected Cas9 protein with independent subsets of the sgRNAs (subset 1 comprised sgRNAs 1, 3, 5, 7, 9; subset 2 comprised sgRNAs 2, 4, 6, 8, 10).

***hcrt* enhancer constructs**

We mutated one or both putative Lhx9 binding sites in the zebrafish *hcrt* promoter by PCR and Gibson assembly (Gibson et al., 2009). Enhancer fragments were placed upstream of the *EGFP* coding sequence. An *hs:lhx9* cassette was placed downstream of *hcrt:EGFP* in a vector containing Tol2 transposase sites.

Electrophoretic mobility shift assay (EMSA)

The following oligonucleotide probes: site A wild-type probe, 5'-GTTGGTATTTGCTAACGAAGCTCGTCCTCCTGTCCA-3'; site A scramble probe, 5'-GTTGGTATTTACTCACAAATCTTGTCCTCCTGTCCA-3'; site B wild-type probe, 5'-TGACAAAGATGCTAACAACCCCGAAAAATCCTTTGT-3'; site B scramble probe, 5'-TGACAAAGATTCTGACCACTCCTAAAAATCCTTTGT-3'. Probes were radiolabeled using [γ -³²P]ATP and T4 polynucleotide kinase for 1 h at 37°C and column-purified (Illustra Microspin G-50, GE Healthcare). EMSAs were performed using a truncated form of zebrafish Lhx9 (Lhx9 HD) that contains the DNA-binding homeodomain but lacks both LIM domains (amino acids 224-396 of the 396 amino acid protein), as described (Wilhelm and Englert, 2002). Lhx9 HD was synthesized *in vitro* (TNT SP6 Quick Coupled Transcription/ Translation System, Promega). 1 μ l normalized radiolabeled oligo and 4.5 μ l TNT lysate were added to a final volume of 30 μ l binding buffer, containing 100 mM KCl, 1 mM MgCl₂, 10 μ M ZnSO₄, 10 mM Tris pH 7.5, 4% glycerol, 1 mg/ml BSA, 200 ng poly(dIdC), and 0.5 mM DTT. After 1 h at room temperature, DNA-protein complexes were resolved by electrophoresis at 4°C on a 6% polyacrylamide DNA retardation gel (Invitrogen) at 100 V for 90 min in 0.5 \times TBE buffer. The gel was dried at 60°C for 2 h using a gel dryer (Bio-Rad) and analyzed by phosphorimaging (GE Healthcare).

Mouse experiments

Outbred ICR (CD-1) timed-pregnant mice were obtained from Japan SLC. Midday of the day of vaginal plug discovery was considered embryonic (E) day 0.5. Early postnatal

mice were anesthetized with a lethal dose of pentobarbitone (100 mg/kg) and, after three failed attempts to elicit a foot withdrawal reflex, the animals were transcardially perfused with 4% paraformaldehyde in PBS. For ISH, brains were fixed overnight in 30% sucrose/4% paraformaldehyde and sectioned in the coronal plane on a Leica sledge microtome at 28 μ m. Sections were mounted on slides and processed for non-radioactive ISH as described (Grove et al., 1998). DNA for riboprobes and electroporation were obtained from FANTOM clones (Carninci et al., 2005). Mouse *Lhx9* (GenBank NM_001025565) was subcloned into a CAG vector as described (Onishi et al., 2010) to generate an overexpression plasmid. *In utero* electroporation was performed as described (Matsui et al., 2011).

References

- Appelbaum, L., Wang, G. X., Maro, G. S., Mori, R., Tovin, A., Marin, W., Yokogawa, T., Kawakami, K., Smith, S. J., Gothilf, Y., et al. (2009). Sleep-wake regulation and hypocretin-melatonin interaction in zebrafish. *Proc. Natl. Acad. Sci. USA* 106, 21942–21947.
- Arias-Carrión, O. and Murillo-Rodríguez, E. (2014). Effects of hypocretin/orexin cell transplantation on narcoleptic-like sleep behavior in rats. *PLoS ONE* 9, e95342.
- Birk, O. S., Casiano, D. E., Wassif, C. A., Cogliati, T., Zhao, L., Zhao, Y., Grinberg, A., Huang, S., Kreidberg, J. A., Parker, K. L., et al. (2000). The LIM homeobox gene *Lhx9* is essential for mouse gonad formation. *Nature* 403, 909–913.
- Blackshaw, S., Scholpp, S., Placzek, M., Ingraham, H., Simerly, R. and Shimogori, T. (2010). Molecular pathways controlling development of thalamus and hypothalamus: from neural specification to circuit formation. *J. Neurosci.* 30, 14925–14930.
- Bonnayon, P. and Lecea, L. (2010). Hypocretins in the control of sleep and wakefulness. *Curr. Neurol. Neurosci. Rep* 10, 174–179.
- Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., et al. (2005). The transcriptional landscape of the mammalian genome. *Science* 309, 1559–1563.
- Chartrel, N., Dujardin, C., Anouar, Y., Leprince, J., Decker, A., Clerens, S., Do-Régo, J.-C., Vandesande, F., Llorens-Cortes, C., Costentin, J., et al. (2003). Identification of 26RFa, a hypothalamic neuropeptide of the RFamide peptide family with orexigenic activity. *Proc. Natl. Acad. Sci. U.S.A.* 100, 15247–15252.
- Chiu, C. N. and Prober, D. A. (2013). Regulation of zebrafish sleep and arousal states: current and prospective approaches. *Front. Neural Circuits* 7, 58.
- Chou, T. C., Lee, C. E., Lu, J., Elmquist, J. K., Hara, J., Willie, J. T., Beuckmann, C. T., Chemelli, R. M., Sakurai, T., Yanagisawa, M., et al. (2001). Orexin (hypocretin) neurons contain dynorphin. *J. Neurosci.* 21, RC168.
- Cvetkovic-Lopes, V., Bayer, L., Dorsaz, S., Maret, S., Pradervand, S., Dauvilliers, Y., Lecendreux, M., Lammers, G.-J., Donjacour, C. E. H. M., Pasquier, Du, R. A., et al. (2010). Elevated Tribbles homolog 2-specific antibody levels in narcolepsy patients. *J. Clin. Invest.* 120, 713–719.
- Dalal, J., Roh, J. H., Maloney, S. E., Akuffo, A., Shah, S., Yuan, H., Wamsley, B., Jones, W. B., Strong, C. de G., Gray, P. A., et al. (2013). Translational profiling of hypocretin neurons identifies candidate molecules for sleep regulation. *Genes Dev.* 27, 565–578.
- Dauvilliers, Y., Arnulf, I. and Mignot, E. (2007). Narcolepsy with cataplexy. *The Lancet* 369, 499–511.

- Elbaz, I., Foulkes, N. S., Gothilf, Y. and Appelbaum, L. (2013). Circadian clocks, rhythmic synaptic plasticity and the sleep-wake cycle in zebrafish. *Front. Neural Circuits* 7, 9.
- Elbaz, I., Yelin-Bekerman, L., Nicenboim, J., Vatine, G. and Appelbaum, L. (2012). Genetic ablation of hypocretin neurons alters behavioral state transitions in zebrafish. *J. Neurosci.* 32, 12961–12972.
- Faraco, J. H., Appelbaum, L., Marin, W., Gaus, S. E., Mourrain, P. and Mignot, E. (2006). Regulation of hypocretin (orexin) expression in embryonic zebrafish. *J. Biol. Chem.* 281, 29753–29761.
- Fukuchi-Shimogori, T. and Grove, E. A. (2001). Neocortex patterning by the secreted signaling molecule FGF8. *Science* 294, 1071–1074.
- Gagnon, J. A., Valen, E., Thyme, S. B., Huang, P., Ahkmetova, L., Pauli, A., Montague, T. G., Zimmerman, S., Richter, C. and Schier, A. F. (2014). Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. *PLoS ONE* 9, e98186.
- Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A. and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Meth.* 6, 343–345.
- Grove, E. A., Tole, S., Limon, J., Yip, L. and Ragsdale, C. W. (1998). The hem of the embryonic cerebral cortex is defined by the expression of multiple Wnt genes and is compromised in Gli3-deficient mice. *Development* 125, 2315–2325.
- Halloran, M. C., Sato-Maeda, M., Warren, J. T., Su, F., Lele, Z., Krone, P. H., Kuwada, J. Y. and Shoji, W. (2000). Laser-induced gene expression in specific cells of transgenic zebrafish. *Development* 127, 1953–1960.
- Hanisch, A., Holder, M. V., Choorapoikayil, S., Gajewski, M., Ozbudak, E. M. and Lewis, J. (2012). The elongation rate of RNA polymerase II in zebrafish and its significance in the somite segmentation clock. *Development* 140, 444–453.
- Higashijima, S., Hotta, Y. and Okamoto, H. (2000). Visualization of cranial motor neurons in live transgenic zebrafish expressing green fluorescent protein under the control of the islet-1 promoter/enhancer. *J. Neurosci.* 20, 206–218.
- Hobert, O. and Westphal, H. (2000). Functions of LIM-homeobox genes. *Trends Genet.* 16, 75–83.
- Honda, M., Eriksson, K. S., Zhang, S., Tanaka, S., Lin, L., Salehi, A., Hesla, P. E., Maehlen, J., Gaus, S. E., Yanagisawa, M., et al. (2009). IGFBP3 colocalizes with and regulates hypocretin (orexin). *PLoS ONE* 4, e4254.
- Horvath, T. L., Peyron, C., Diano, S., Ivanov, A., Aston-Jones, G., Kilduff, T. S. and van

- Den Pol, A. N. (1999). Hypocretin (orexin) activation and synaptic innervation of the locus coeruleus noradrenergic system. *J. Comp. Neurol.* 415, 145–159.
- Hwang, W. Y., Fu, Y., Reyon, D., Maeder, M. L., Tsai, S. Q., Sander, J. D., Peterson, R. T., Yeh, J.-R. J. and Joung, J. K. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.* 31, 227–229.
- Jao, L.-E., Wentz, S. R. and Chen, W. (2013). Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proc. Natl. Acad. Sci. U.S.A.* 110, 13904–13909.
- Kaslin, J., Nystedt, J. M., Ostergård, M., Peitsaro, N. and Panula, P. (2004). The orexin/hypocretin system in zebrafish is connected to the aminergic and cholinergic systems. *J. Neurosci.* 24, 2678–2689.
- Kosman, D., Mizutani, C. M., Lemons, D., Cox, W. G., McGinnis, W. and Bier, E. (2004). Multiplex detection of RNA expression in *Drosophila* embryos. *Science* 305, 846.
- Koyama, M., Kinkhabwala, A., Satou, C., Higashijima, S.-I. and Fetcho, J. (2011). Mapping a sensory-motor network onto a structural and functional ground plan in the hindbrain. *Proc. Natl. Acad. Sci. USA* 108, 1170–1175.
- Kucenas, S., Soto, F., Cox, J. A. and Voigt, M. M. (2006). Selective labeling of central and peripheral sensory neurons in the developing zebrafish using P2X3 receptor subunit transgenes. *Neuroscience* 138, 641–652.
- Lein, E. S., Hawrylycz, M. J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A. F., Boguski, M. S., Brockway, K. S., Byrnes, E. J., et al. (2006). Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 445, 168–176.
- Liu, F. and Patient, R. (2008). Genome-wide analysis of the zebrafish ETS family identifies three genes required for hemangioblast differentiation or angiogenesis. *Circ. Res.* 103, 1147–1154.
- Machluf, Y., Gutnick, A. and Levkowitz, G. (2011). Development of the zebrafish hypothalamus. *Ann. N. Y. Acad. Sci.* 1220, 93–105.
- Manoli, M. and Driever, W. (2012). Fluorescence-activated cell sorting (FACS) of fluorescently tagged cells from zebrafish larvae for RNA isolation. *Cold Spring Harb. Protoc.* 2012.
- Matsui, A., Yoshida, A. C., Kubota, M., Ogawa, M. and Shimogori, T. (2011). Mouse in utero electroporation: controlled spatiotemporal gene transfection. *J. Vis. Exp.* e3024.
- Montague, T. G., Cruz, J. M., Gagnon, J. A., Church, G. M. and Valen, E. (2014). CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res.* 42, W401–7.

- Pan, Y. A., Choy, M., Prober, D. A. and Schier, A. F. (2012). Robo2 determines subtype-specific axonal projections of trigeminal sensory neurons. *Development* 139, 591–600.
- Park, H.-C., Kim, C.-H., Bae, Y.-K., Yeo, S.-Y., Kim, S.-H., Hong, S.-K., Shin, J., Yoo, K.-W., Hibi, M., Hirano, T., et al. (2000). Analysis of upstream elements in the HuC promoter leads to the establishment of transgenic zebrafish with fluorescent neurons. *Dev. Biol.* 227, 279–293.
- Peukert, D., Weber, S., Lumsden, A. and Scholpp, S. (2011). Lhx2 and Lhx9 determine neuronal differentiation and compartment in the caudal forebrain by regulating Wnt signaling. *PLoS Biol.* 9, e1001218.
- Peyron, C., Faraco, J., Rogers, W., Ripley, B., Overeem, S., Charnay, Y., Nevsimalova, S., Aldrich, M., Reynolds, D., Albin, R., et al. (2000). A mutation in a case of early onset narcolepsy and a generalized absence of hypocretin peptides in human narcoleptic brains. *Nat. Med.* 6, 991–997.
- 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, 13400–13410.
- Prober, D. A., Zimmerman, S., Myers, B. R., McDermott, B. M., Kim, S.-H., Caron, S., Rihel, J., Solnica-Krezel, L., Julius, D., Hudspeth, A. J., et al. (2008). Zebrafish TRPA1 channels are required for chemosensation but not for thermosensation or mechanosensory hair cell function. *J. Neurosci.* 28, 10102–10110.
- Robu, M. E., Larson, J. D., Nasevicius, A., Beiraghi, S., Brenner, C., Farber, S. A. and Ekker, S. C. (2007). p53 activation by knockdown technologies. *PLoS Genet.* 3, e78.
- Rosin, D. L., Weston, M. C., Sevigny, C. P., Stornetta, R. L. and Guyenet, P. G. (2003). Hypothalamic orexin (hypocretin) neurons express vesicular glutamate transporters VGLUT1 or VGLUT2. *J. Comp. Neurol.* 465, 593–603.
- Sagasti, A., Guido, M. R., Raible, D. W. and Schier, A. F. (2005). Repulsive interactions shape the morphologies and functional arrangement of zebrafish peripheral sensory arbors. *Curr. Biol.* 15, 804–814.
- Sato, T., Takahoko, M. and Okamoto, H. (2006). HuC:Kaede, a useful tool to label neural morphologies in networks in vivo. *Genesis* 44, 136–142.
- Sawai, N., Ueta, Y., Nakazato, M. and Ozawa, H. (2010). Developmental and aging change of orexin-A and -B immunoreactive neurons in the male rat hypothalamus. *Neurosci. Lett.* 468, 51–55.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Meth.* 9, 676–682.

- Schmittgen, T. D. and Livak, K. J. (2008). Analyzing real-time PCR data by the comparative CT method. *Nat. Protoc.* 3, 1101–1108.
- Shimogori, T. and Ogawa, M. (2008). Gene application with in utero electroporation in mouse embryonic brain. *Dev. Growth Differ.* 50, 499–506.
- Shimogori, T., Lee, D. A., Miranda-Angulo, A., Yang, Y., Wang, H., Jiang, L., Yoshida, A. C., Kataoka, A., Mashiko, H., Avetisyan, M., et al. (2010). A genomic atlas of mouse hypothalamic development. *Nat. Neurosci.* 13, 767–775.
- Takayasu, S., Sakurai, T., Iwasaki, S., Teranishi, H., Yamanaka, A., Williams, S. C., Iguchi, H., Kawasaki, Y. I., Ikeda, Y., Sakakibara, I., et al. (2006). A neuropeptide ligand of the G protein-coupled receptor GPR103 regulates feeding, behavioral arousal, and blood pressure in mice. *Proc. Natl. Acad. Sci. U.S.A.* 103, 7438–7443.
- Tessmar-Raible, K., Raible, F., Christodoulou, F., Guy, K., Rembold, M., Hausen, H. and Arendt, D. (2007). Conserved sensory-neurosecretory cell types in annelid and fish forebrain: insights into hypothalamus evolution. *Cell* 129, 1389–1400.
- Thannickal, T. C., Moore, R. Y., Nienhuis, R., Ramanathan, L., Gulyani, S., Aldrich, M., Cornford, M. and Siegel, J. M. (2000). Reduced number of hypocretin neurons in human narcolepsy. *Neuron* 27, 469–474.
- Thisse, B., Heyer, V., Lux, A., Alunni, V., Degrave, A., Seiliez, I., Kirchner, J., Parkhill, J.-P. and Thisse, C. (2004). Spatial and temporal expression of the zebrafish genome by large-scale in situ hybridization screening. *Methods Cell Biol.* 77, 505–519.
- Thisse, C. and Thisse, B. (2008). High-resolution in situ hybridization to whole-mount zebrafish embryos. *Nat. Protoc.* 3, 59–69.
- Tsujino, N. and Sakurai, T. (2009). Orexin/hypocretin: a neuropeptide at the interface of sleep, energy homeostasis, and reward system. *Pharmacol. Rev.* 61, 162–176.
- Westerfield, M. (1993). *The Zebrafish Book. A guide for the laboratory use of zebrafish (Danio rerio)*. 4th edn. Eugene, OR: University of Oregon Press.
- Wilhelm, D. and Englert, C. (2002). The Wilms tumor suppressor WT1 regulates early gonad development by activation of Sf1. *Genes Dev.* 16, 1839–1851.
- Wilson, S. W. and Houart, C. (2004). Early steps in the development of the forebrain. *Dev. Cell* 6, 167–181.
- Yamamoto, T., Miyazaki, R. and Yamada, T. (2009). Intracerebroventricular administration of 26RFa produces an analgesic effect in the rat formalin test. *Peptides* 30, 1683–1688.

Supplementary Figures

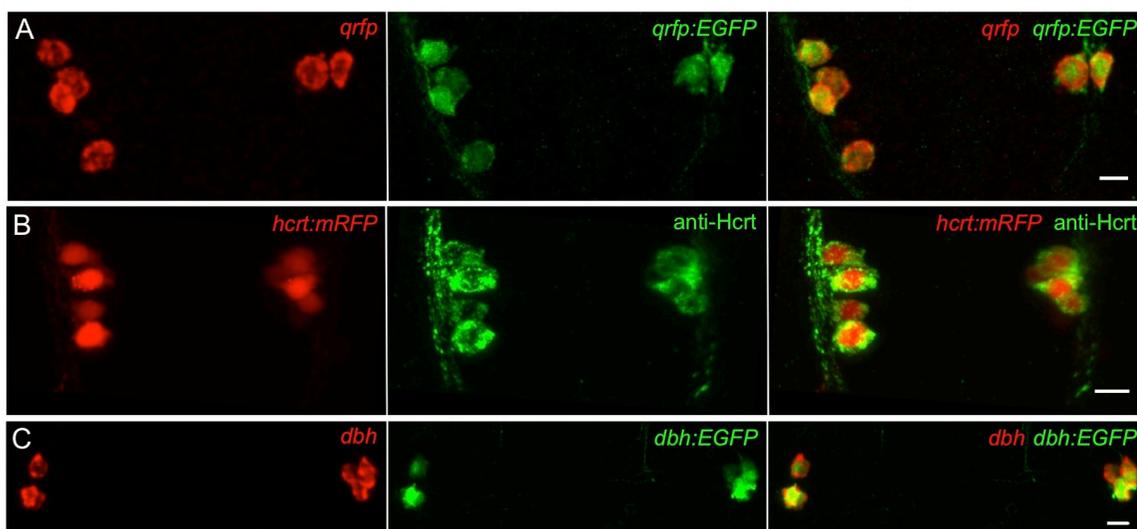


Figure S1. Validation of zebrafish QRFP, Hcrt, and DBH transgenic reporter lines. All cells labeled by fluorescent ISH for *qrfp* and *dbh* are also labeled by anti-GFP immunostaining in *Tg(qrfp:EGFP, hcrt:mRFP)* embryos at 24 hpf (A) and *Tg(dbh:EGFP)* larvae at 120 hpf (C). All mRFP expressing cells in 24 hpf *Tg(qrfp:EGFP, hcrt:mRFP)* embryos are labeled by anti-Hcrt immunostaining (B). Anterior (A,B) and dorsal (C) views are shown. Scale = 10 μ m.

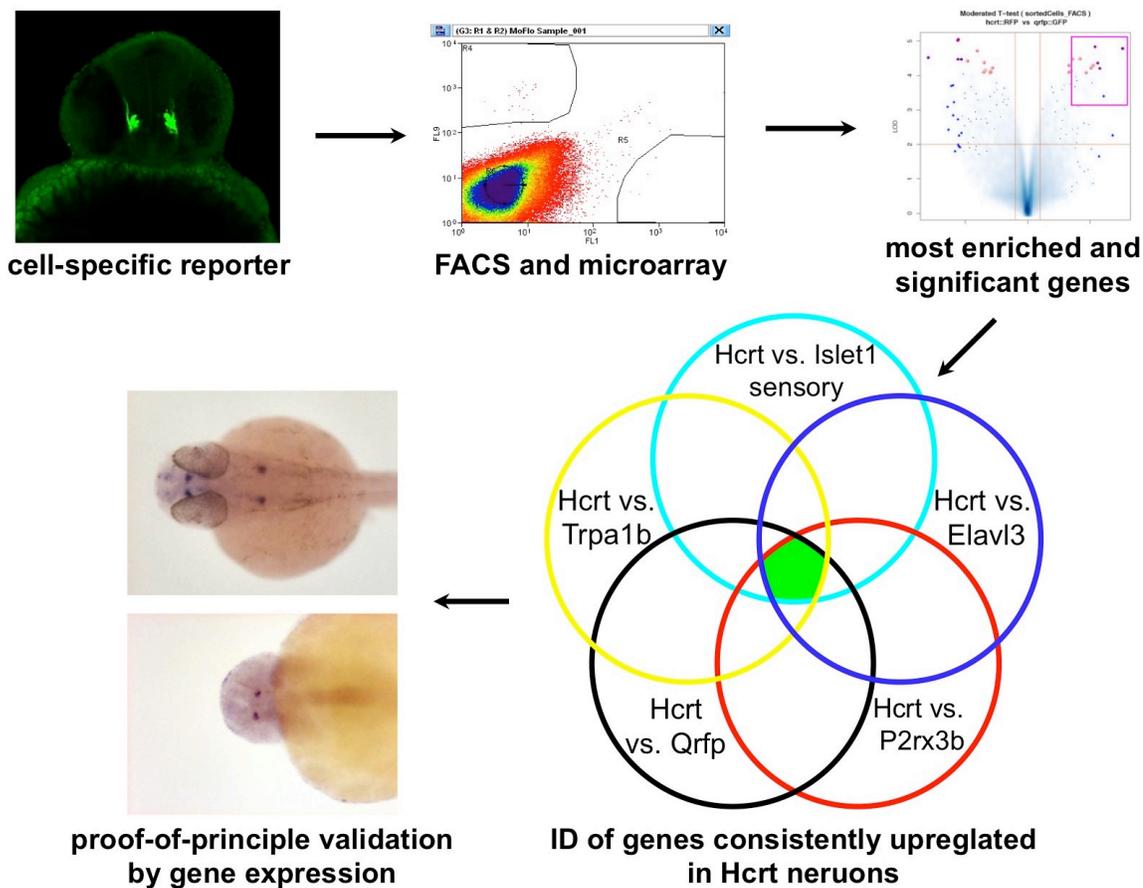


Figure S2. Strategy to identify genes enriched in embryonic zebrafish Hcrt neurons. Transgenic zebrafish embryos were dissociated into single cells at 26 hpf and cells of interest were purified by FACS. RNA from purified neurons was amplified and hybridized to microarrays. Signals from Hcrt neurons were compared to other neuron types and genes most highly enriched in Hcrt neurons were identified. Highly enriched transcription factors and secreted peptides were selected for verification by ISH and overexpression analysis in zebrafish.

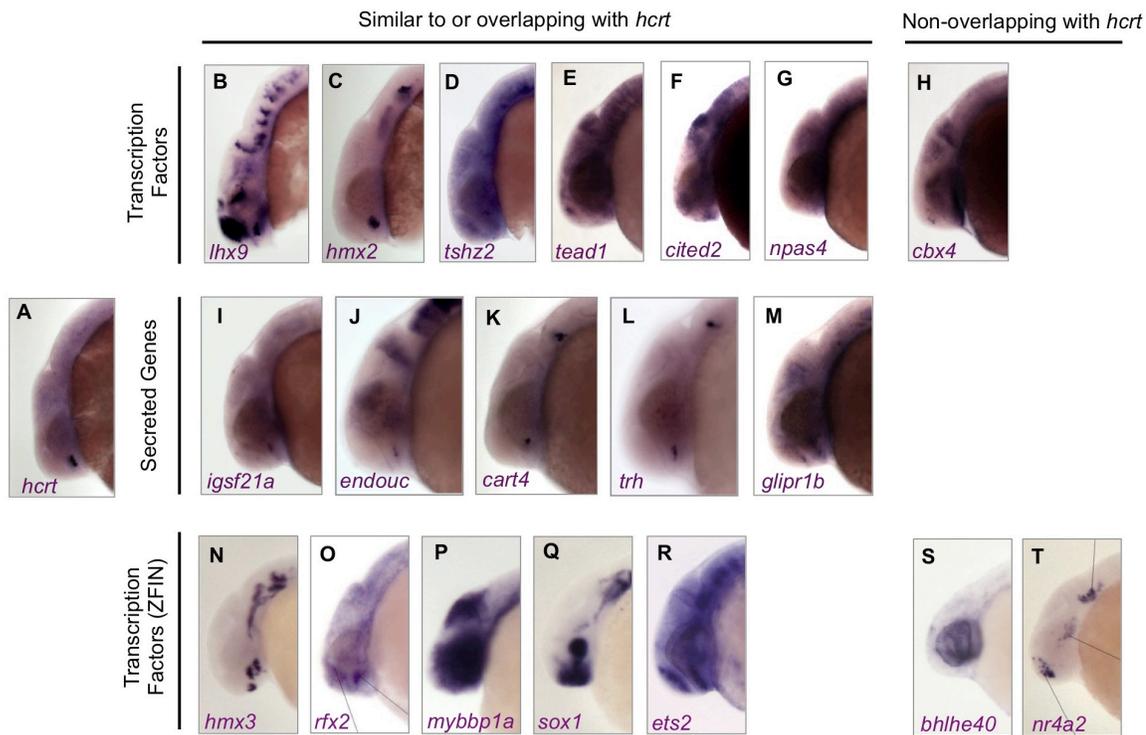


Figure S3. Expression patterns of genes identified as enriched in Hcrt neurons. ISH was performed using wild-type zebrafish embryos at 24 hpf. Anterior is left. Most candidate genes are expressed in a domain that is similar to, or overlaps with, the *hcrt* expression domain. Images in the bottom row were obtained from the ZFIN ISH database (Liu and Patient, 2008; Thisse et al., 2004).

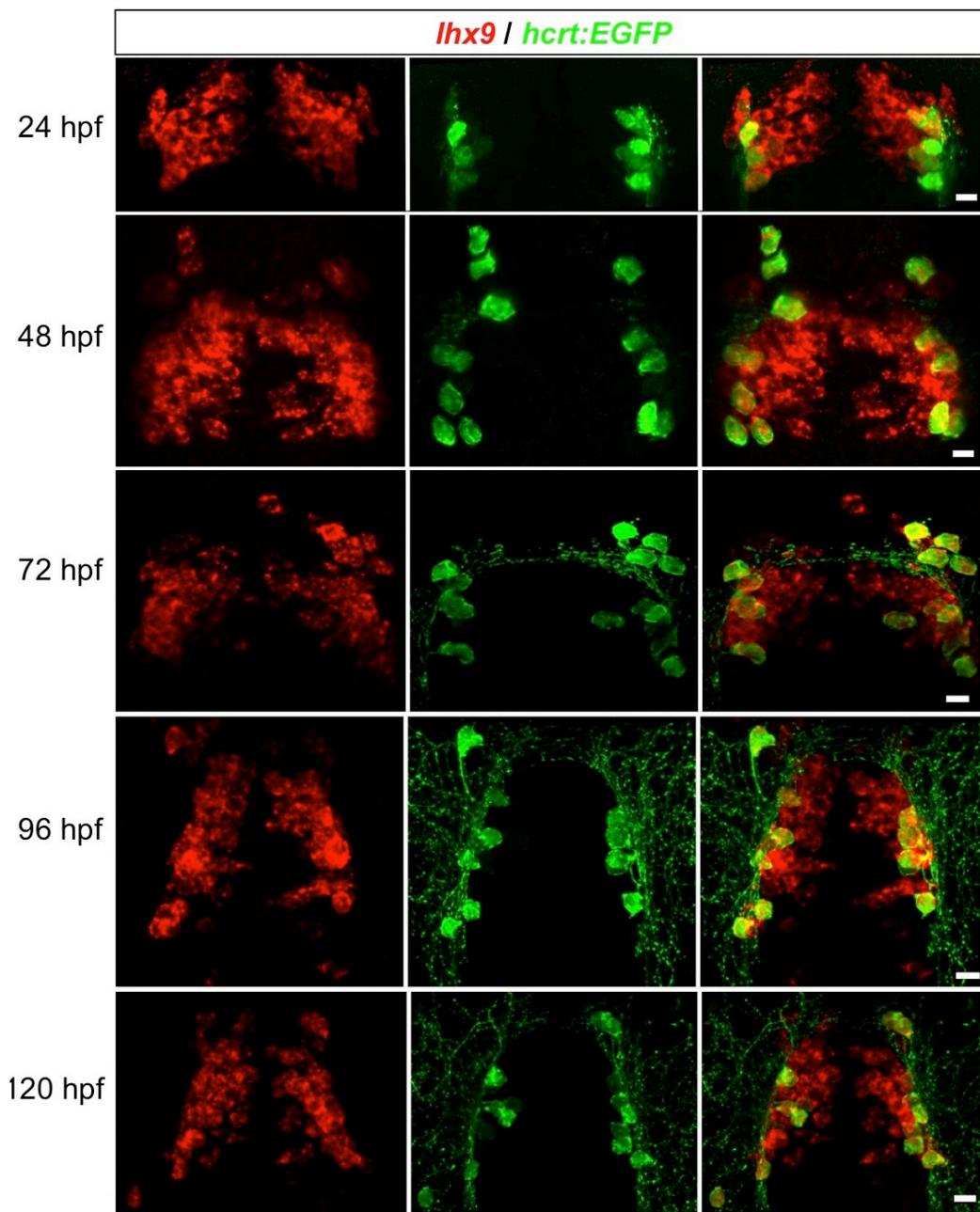


Figure S4. Time course of endogenous *hcrt* and *lhx9* expression. Confocal projections of *Tg(hcrt:EGFP)* embryos at 24, 48, 72, 96, and 120 hpf show that all Hcrt cells express *lhx9* throughout development, as determined by fluorescent ISH for *lhx9* followed by anti-GFP immunostaining. Ventral images are shown. Scale = 10 μ m.

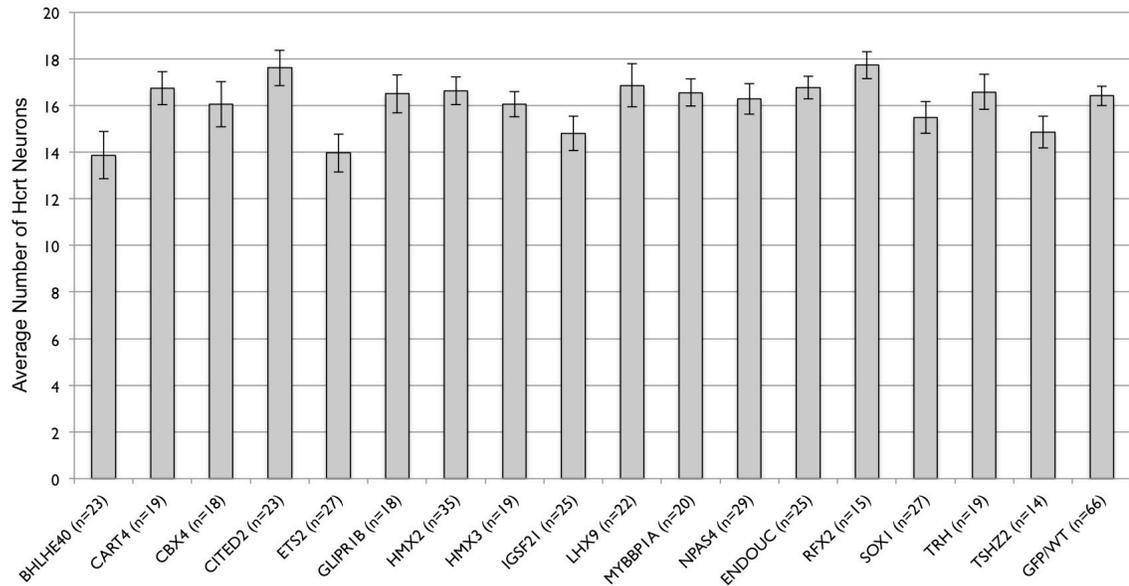


Figure S5. Overexpression of candidate genes does not affect the number of Hcrt neurons in the hypothalamus. Overexpression of each candidate gene was induced by heat shock at 24 hpf and larvae were fixed at 120 hpf for ISH with a *hcr*t probe. Mean \pm s.e.m. number of Hcrt cells in the entire hypothalamus is shown. Control larvae were injected with a HS-EGFP plasmid. n indicates number of larval brains analyzed. No significant difference was detected between any overexpressed gene and control ($p > 0.05$ by one-way ANOVA). Note that ectopic Hcrt neurons (i.e., not in the hypothalamus) were only observed following overexpression of *lhx9* (Fig. 3).

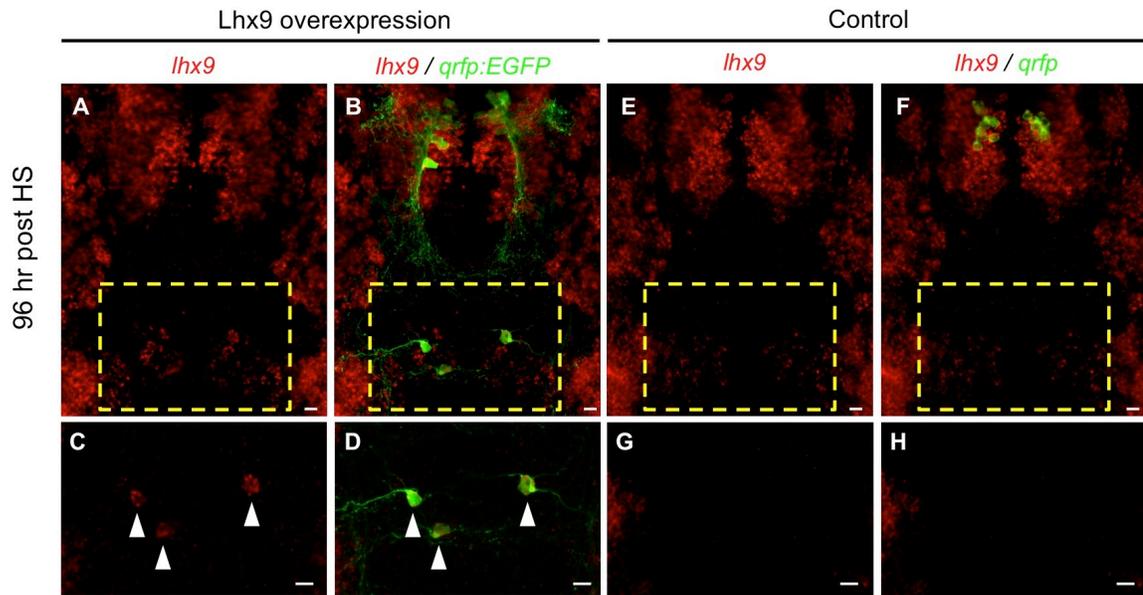


Figure S6. *lhx9* overexpression induces ectopic QRFP neurons. (A-D) Ectopic *qrfp:EGFP* expressing neurons that persist until 96 hours post HS (120 hpf) also express ectopic *lhx9* (arrowheads in C, D). (E-H) Double fluorescent ISH for *lhx9* and *qrfp* in WT 120 hpf larvae shows an absence of *lhx9* expression in the hindbrain region that contains *lhx9*-induced ectopic *qrfp* neurons in (A-D). Boxed regions in (A, B, E, F) are shown at higher magnification in (C, D, G, H). (A, B, E, F) show 95 μm thick confocal maximum intensity projections containing both endogenous and ectopic QRFP neurons. (C, D, G, H) show 43 μm thick confocal maximum intensity projections including only the region containing ectopic QRFP neurons. *lhx9*-expressing neurons that appear close to ectopic QRFP neurons in (B) are located 30 μm ventral to the ectopic QRFP neurons, and are thus not observed in (D). Scale = 10 μm .

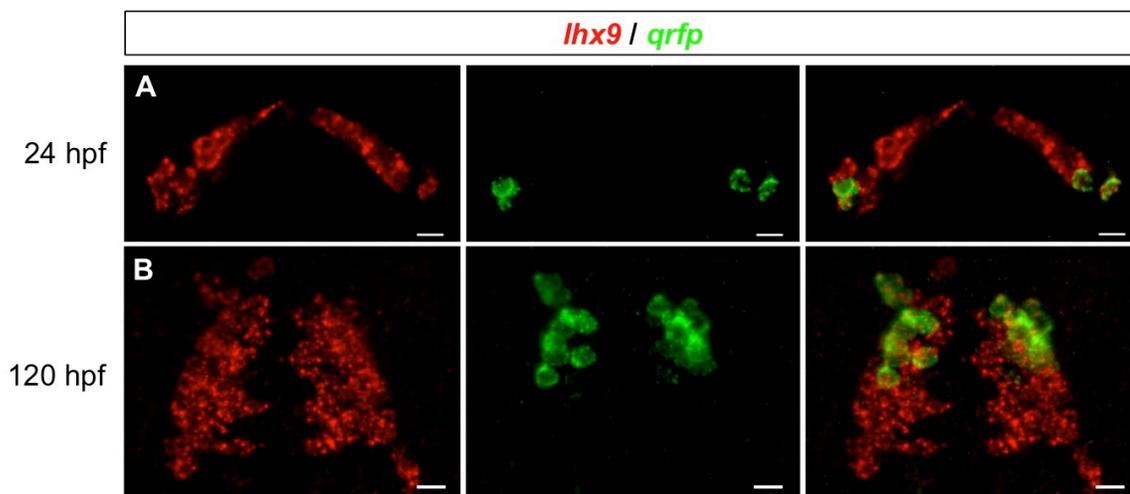


Figure S7. Expression of endogenous *qrfp* and *lh9*. Confocal projections of WT embryos at 24 and 120 hpf show that all *qrfp*-expressing cells express *lh9*, as determined by double fluorescent ISH. Ventral images are shown. Scale = 10 μ m.

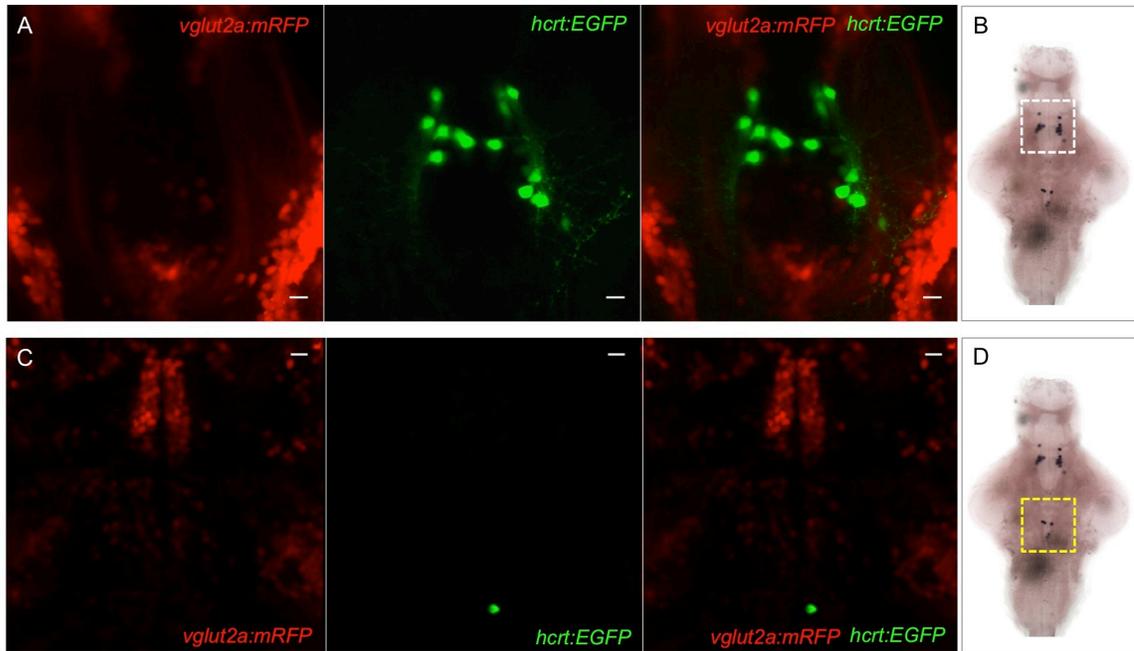


Figure S8. Hcrt neurons do not express *vglut2a* at 120 hpf. Confocal projections of a 120 hpf *Tg(vglut2a:mRFP, hcrt:EGFP)* larva shows that *vglut2a* cells labeled with mRFP do not colocalize with endogenous (A) or ectopic (C) Hcrt neurons labeled with EGFP. The regions shown in (A) and (C) are indicated with dashed boxes in (B) and (D), respectively. Scale = 10 μm.

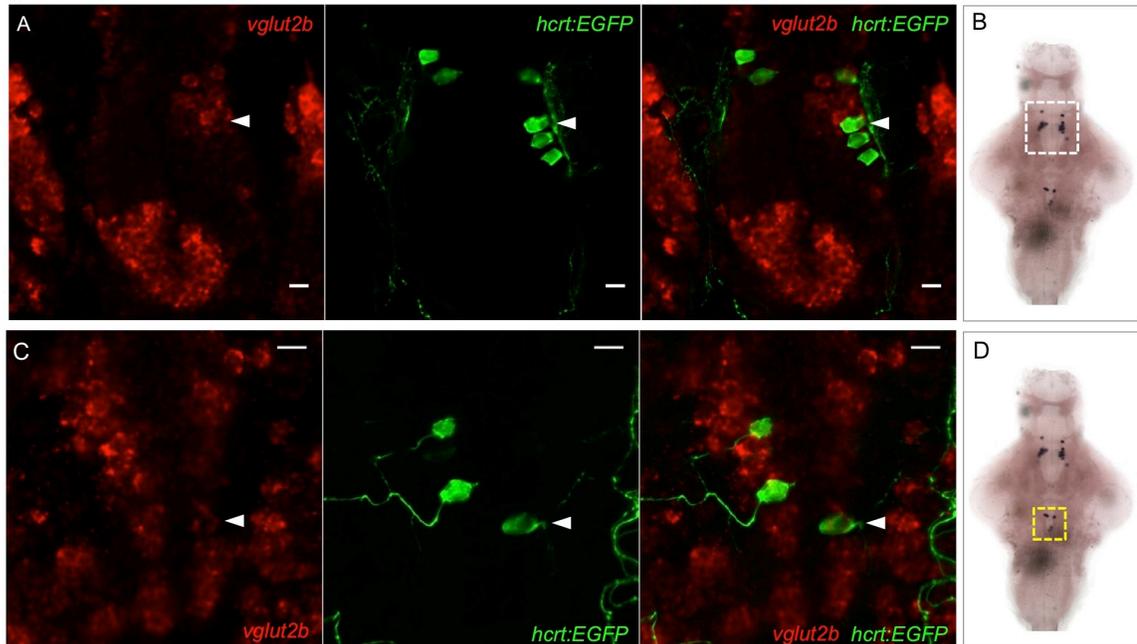


Figure S9. Few or no Hcrt neurons express *vglut2b* at 120 hpf. Confocal projections of a 120 hpf *Tg(hcrt:EGFP)* larva shows that most Hcrt neurons immunostained with a GFP-specific antibody do not co-localize with *vglut2b*-expressing cells labeled by fluorescent ISH. Weak co-labeling was occasionally observed in endogenous and ectopic Hcrt neurons (white arrowheads). The approximate regions shown in (A) and (C) are indicated with dashed boxes in (B) and (D), respectively. Scale = 10 μ m.

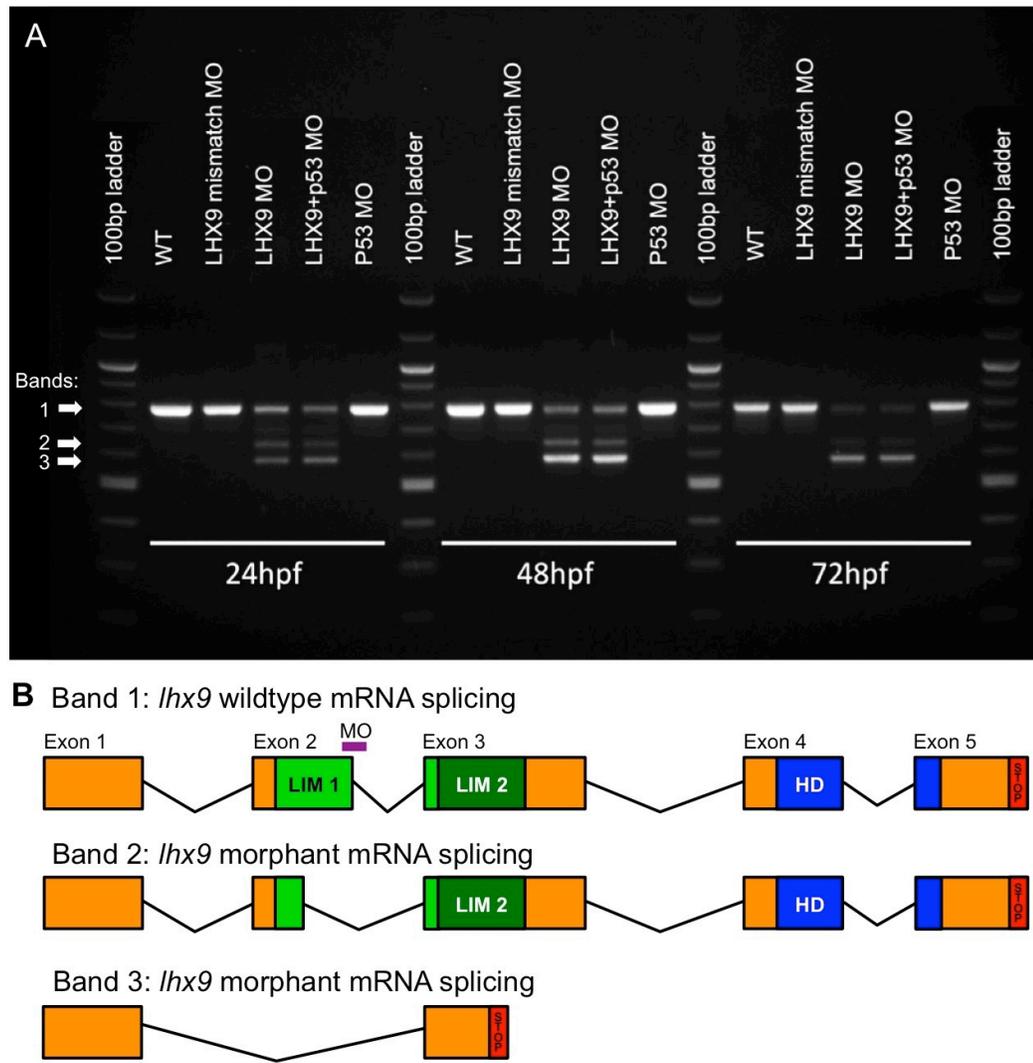


Figure S10. Molecular analysis of *lhx9* morpholino knockdown. (A) Mature *lhx9* mRNA lacks part or all of the second exon after injection with a splice blocking morpholino, reducing the PCR product size by 96 bp or 203 bp, respectively. A 5 bp mismatch control morpholino and the apoptosis suppressing p53 morpholino have no effect on *lhx9* splicing. Gene knockdown persists until at least 72 hpf. Note that a small amount of correctly spliced *lhx9* is present at all time points, indicating incomplete knockdown. (B) Diagram of the five exons of wild-type *lhx9* and the two variants caused by the *lhx9* morpholino. A cryptic splice site in exon 2 produces an in-frame 32 amino acid deletion that removes most of the LIM1 domain (Band 2). A second variant (Band 3) lacks exon 2, contains an early stop codon, and lacks both LIM domains and the DNA-binding homeodomain (HD).

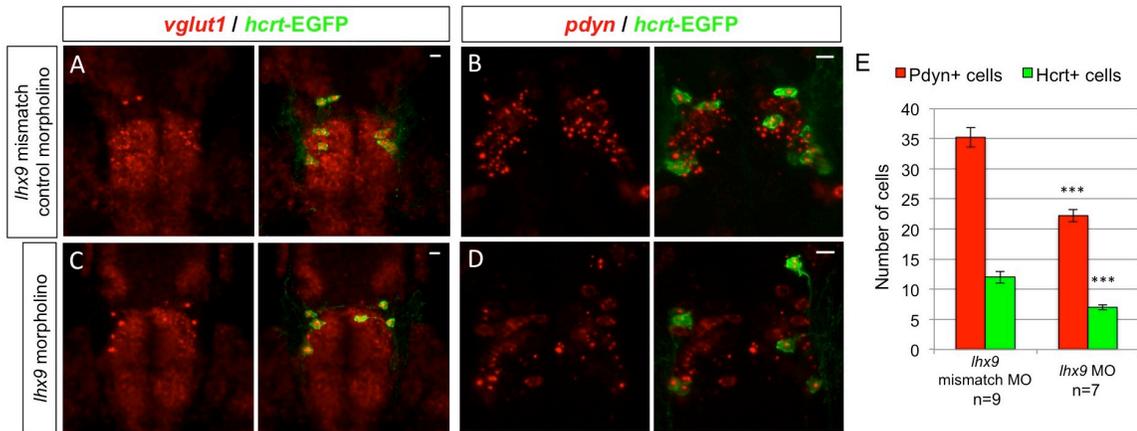


Figure S11. Endogenous *hcr*- and *pdyn*-expressing neurons are reduced in *lhx9* morphants. *Tg(hcr:EGFP)* embryos injected with either *lhx9* morpholino (C, D) or *lhx9* mismatch control morpholino (A, B) were fixed at 72 hpf and probed for *vglut1* or *pdyn* expression by fluorescent ISH. No obvious defects in *vglut1* expression were observed in *lhx9* morphants, but the total number of cells with intense, punctate *pdyn* expression was reduced by approximately 40% (E). This result suggests that the morpholino-induced reduction in Hcr neurons is caused by cell loss, rather than silencing of *hcr* expression. n indicates number of morphant brains analyzed. ***, $p < 0.001$ compared to embryos injected with the control morpholino by Student's t-test. Scale = 10 μ m.

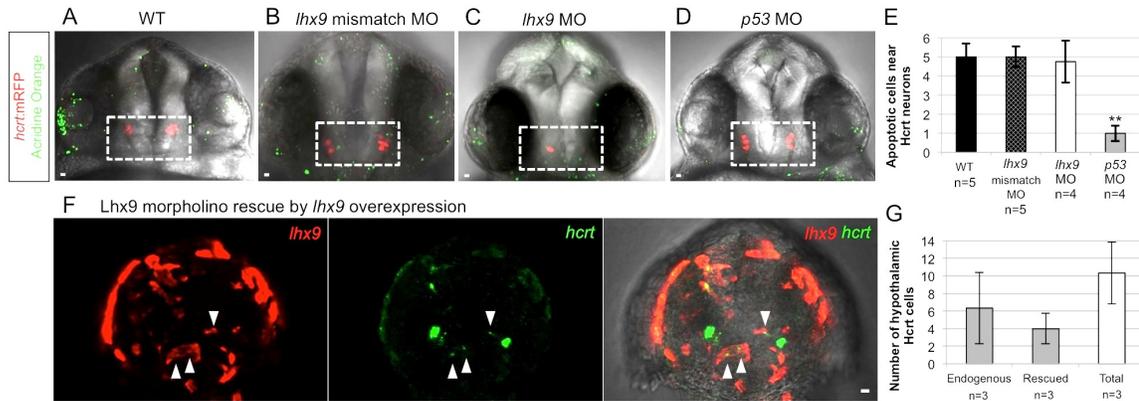


Figure S12. The *lhx9* morpholino does not induce apoptosis and can be rescued by *lhx9* overexpression. *Tg(hcrt:mRFP)* embryos were stained with acridine orange at 24 hpf to quantify apoptotic cells (A-D). There was no increase in apoptosis in embryos injected with *lhx9* morpholino (C) compared to embryos injected with the *lhx9* mismatch control morpholino (B) or wild-type embryos (A), suggesting that the reduced number of Hcrt neurons in *lhx9* morphants is not due to apoptosis. Acridine orange labeled fewer cells in embryos injected with the *p53* morpholino (D), presumably because apoptosis that normally occurs during development was suppressed. (E) Mean \pm s.e.m. number of apoptotic cells in the white boxed region. At least 4 embryo brains were analyzed for each condition. **, $p < 0.01$ compared to WT by one-way ANOVA followed by Bonferroni's correction for multiple comparisons. (F) To rescue the morpholino phenotype, *Tg(hcrt:EGFP)* embryos were co-injected with the *lhx9* morpholino and the *hs:lhx9* plasmid. Following heat shock at 24 hpf, *lhx9*-overexpressing cells located in the endogenous *hcrt* expression domain also expressed *hcrt* (white arrowheads), indicating that *lhx9* overexpression can rescue the *lhx9* morpholino phenotype. (G) Mean \pm s.e.m. number of endogenous and rescued *hcrt* cells per brain. Following rescue, the total number of *hcrt*-expressing cells is similar to the number observed in wild-type embryo brains (see Fig. 2G). *hcrt* expression in rescued cells was weaker than in endogenous Hcrt neurons, presumably because the rescued cells only received a pulse of *lhx9* while endogenous Hcrt neurons continuously express *lhx9*. Three embryo brains with ectopic *lhx9* expression in the endogenous Hcrt region were quantified. Anterior views of 26 hpf embryos are shown. Scale = 10 μ m.

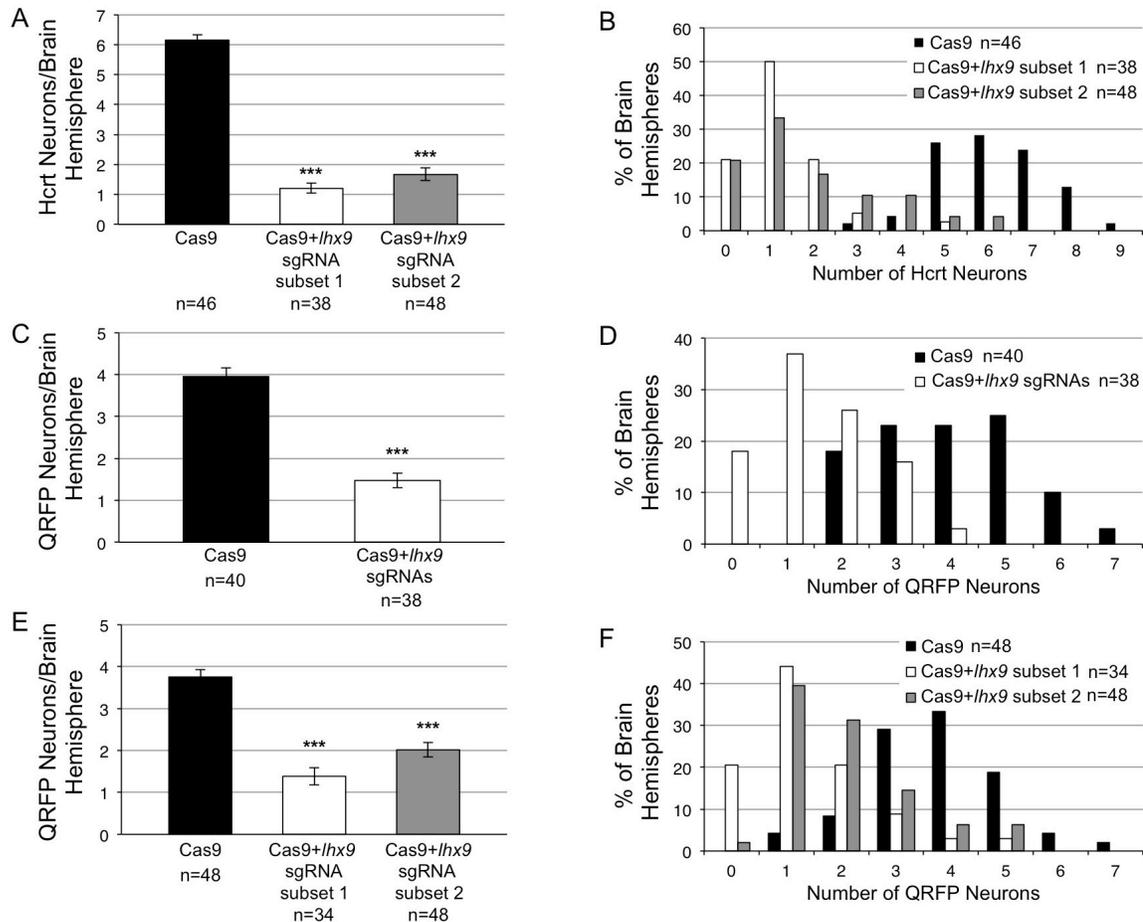


Figure S13. CRISPR/Cas9 targeting of *lhx9* affects Hcrt and QRFP neuron specification. (A) Embryos injected with Cas9+*lhx9* sgRNA subset 1 (sgRNAs 1, 3, 5, 7, 9) or Cas9+*lhx9* sgRNA subset 2 (sgRNAs 2, 4, 6, 8, 10) show a 5.1-fold and 3.7-fold reduction, respectively, in the number of Hcrt neurons per brain hemisphere compared to embryos injected with Cas9 alone. (C) Embryos injected with Cas9 and all 10 *lhx9* sgRNAs show a 2.7-fold reduction in the number of QRFP neurons per brain hemisphere. This reduction was significant, but less dramatic than the reduction observed for Hcrt neurons (Fig. 6F). (E) Embryos injected with Cas9+*lhx9* sgRNA subset 1 or Cas9+*lhx9* sgRNA subset 2 have significantly fewer QRFP neurons in each brain hemisphere. Injection of the two non-overlapping *lhx9* sgRNA subsets each causes the same phenotype as injection of all 10 sgRNAs, suggesting that loss of Hcrt and QRFP cells is due to *lhx9* knockout and not due to off-target effects of particular sgRNAs. Histograms in (B, D, F) show the percentage of brain hemispheres containing the indicated number of Hcrt neurons measured in (A, C, E). ***, $p < 0.001$ compared to control embryos injected with Cas9 alone by one-way ANOVA followed by Tukey's correction for multiple comparisons. n indicates number of brain hemispheres analyzed.

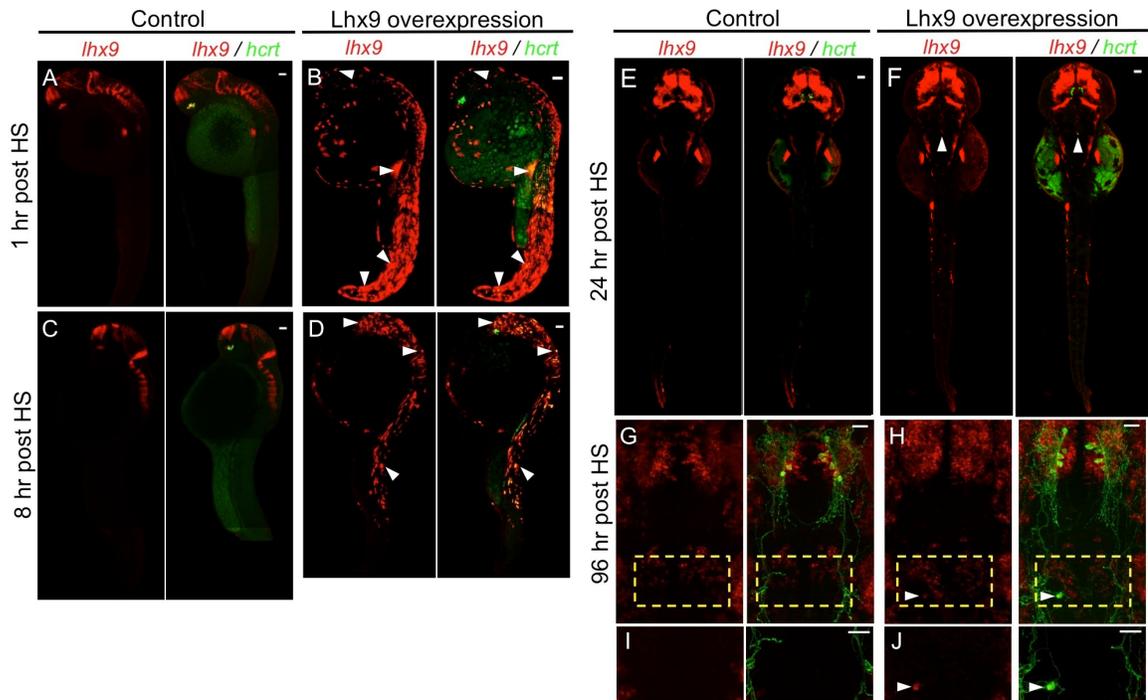


Figure S14. Time course of *lhx9* and *hcrt* expression after heat shock.

Double fluorescent ISH for *lhx9* and *hcrt* are shown at 1 hour, 8 hours, and 24 hours after heat shock (HS)-induced *lhx9* overexpression. Fluorescent ISH for *lhx9* and immunostained *hcrt:EGFP* is shown at 96 hours post HS. (B) At 1 hour post HS, widespread *lhx9* mRNA is detected throughout the embryo. Nearly all *lhx9*-expressing cells also express *hcrt*. (D) At 8 hours post HS, the number of cells with *lhx9* overexpression and ectopic *hcrt* expression is reduced by approximately four-fold. Most *lhx9*-expressing cells still express *hcrt*. (F) At 24 hours post HS, little ectopic *lhx9* or *hcrt* expression is observed, except for ectopic *lhx9*- and *hcrt*-co-expressing neurons in the medial hindbrain. Ectopic *hcrt*-expressing neurons that persist until 96 hours post HS (120 hpf) also express ectopic *lhx9* (H, J). In contrast, control embryos injected with empty heat shock vector only exhibit *hcrt* expression in the hypothalamus (A, C, E, G, I). Arrowheads indicate examples of ectopic *lhx9* and *hcrt* co-expression. (G, H) show 91 μm thick confocal maximum intensity projections containing both endogenous and ectopic Hcrt neurons. (I, J) show 45 μm thick confocal maximum intensity projections including only the region containing ectopic Hcrt neurons. *lhx9*-expressing neurons that appear close to ectopic Hcrt neurons in (H) are located 30 μm ventral to the ectopic Hcrt neurons, and are thus not observed in (J). Embryos fixed at 1 hour and 8 hours post HS are shown in side view. Embryos fixed at 24 hours and 96 hours post HS are shown dorsally and ventrally, respectively. Boxed regions in (G, H) are shown at higher magnification in (I, J). Scale indicates 50 μm (A-F) and 10 μm (G-J).

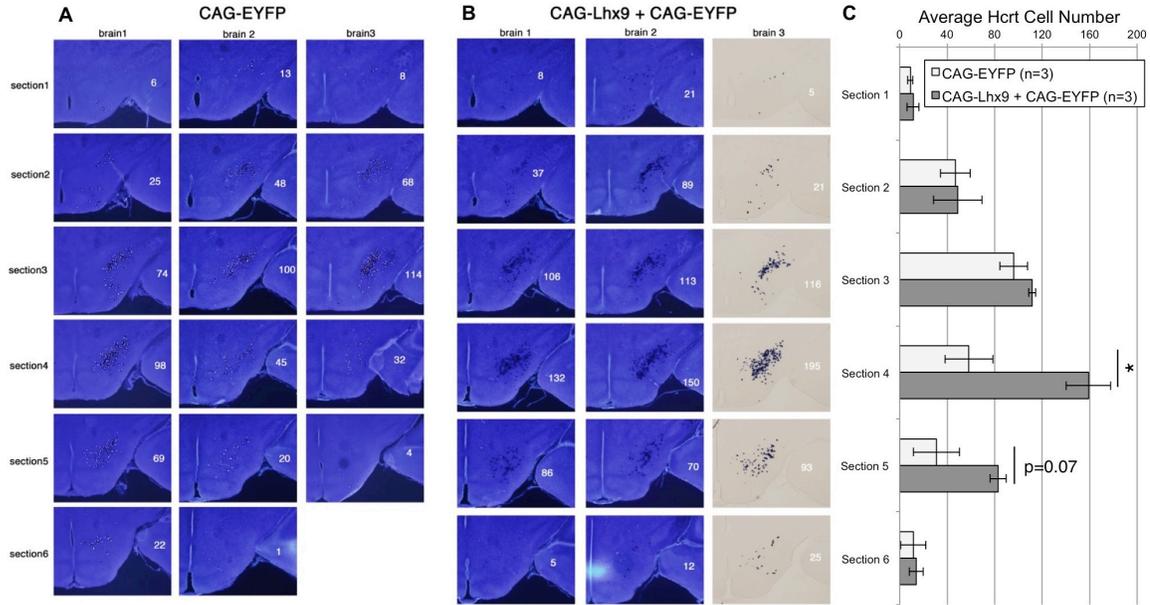


Figure S15. Quantification of Hcrt neuron specification following *in utero* overexpression of *Ihx9*. E10.5 mouse embryos were electroporated with CAG-EYFP (A) or CAG-Lhx9 + CAG-EYFP (B) *in utero* into the right brain hemisphere lateral hypothalamus and analyzed at P6 for *hcr*t expression by ISH in 6 adjacent coronal sections. Sections were co-stained with DAPI to visualize individual cells to facilitate accurate quantification of *hcr*t ISH (DAPI co-stain is shown for all three brains in (A) and for two brains in (B)). White dots in (A) are included to show how quantification was performed. White numbers indicate the number of *hcr*t-expressing neurons in a brain hemisphere in each section. (C) Mean \pm s.e.m. number of *hcr*t-expressing neurons for three experimental and three control brain hemispheres in each section. *, $p < 0.05$ compared to control by Student's t-test. These data were combined to generate the graph in Fig. 8F.

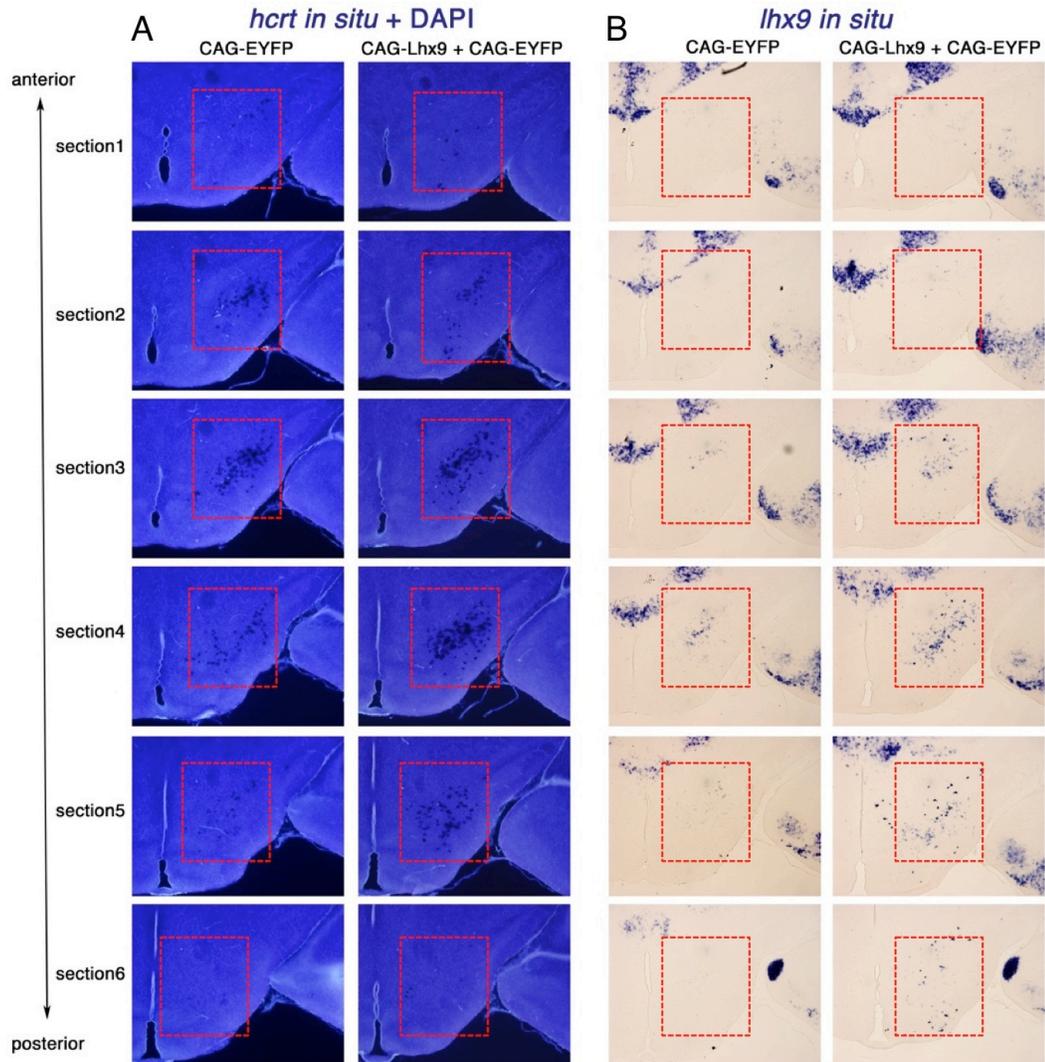


Figure S16. Hypothalamic *lhx9* overexpression coincides with increased *hcrt*-expressing neurons. Six serial coronal sections of exemplar brains electroporated with CAG-EYFP or CAG-Lhx9 + CAG-EYFP in the right brain hemisphere lateral hypothalamus at E10.5 and processed for *hcrt* (A) or *lhx9* (B) ISH at P6 are shown. *hcrt* ISH sections were co-stained with DAPI to facilitate quantification of *hcrt*-expressing neurons. The majority of ectopic *lhx9*-expressing neurons are observed in sections 4 and 5 (B, compare boxed region in right column to left column), which coincides with sections that show the greatest increase in *hcrt*-expressing neurons compared to CAG-EYFP controls (A, compare boxed regions in right column to left column). See Fig. S12 for quantification of *hcrt*-expressing neurons in each section. The CAG-EYFP and CAG-Lhx9 + CAG-EYFP images shown are the same as brain 2 and brain 1 in Fig. S12, respectively.

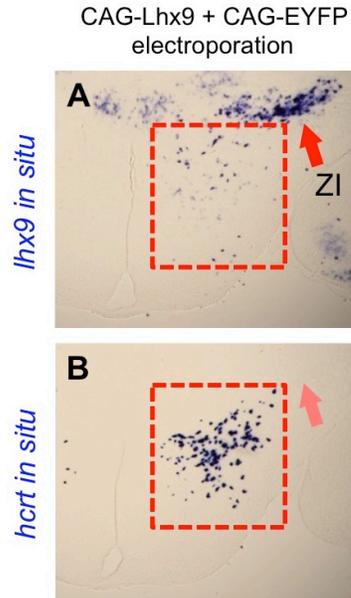


Figure S17. *Lhx9* overexpression in the zona incerta does not induce Hcrt neuron specification. Mice electroporated with CAG-Lhx9 + CAG-EYFP in the zona incerta (ZI) (A, red arrow), show no ectopic Hcrt neurons in the ZI (B). *hcrt* ISH was allowed to develop longer for these samples than those shown in Figs 8, S12 and S13 to ensure that any faint *hcrt* expression in the ZI could be detected. Dashed box indicates the lateral hypothalamus.

CHAPTER 4:

Conclusions and Future Directions

4.1 Conclusions

The findings presented in this thesis reveal a striking degree of evolutionary conservation of sleep regulatory mechanisms at both the genetic and cellular level. The EGFR signaling pathway promotes sleep behavior in zebrafish, just as it does in nematodes, fruit flies, and rabbits. Similarly, *Lhx9* specifies Hcrt neurons during early development of both zebrafish and mice. Though aspects of sleep vary widely across animal species, it is evident that sleep is an ancient behavior, regulated by processes that are preserved over time. The zebrafish, as a simple animal model of sleep, is only about a decade old and is rapidly expanding its repertoire of genetic tools. Further investigation in zebrafish and other model organisms will undoubtedly yield new insights into sleep and fundamental sleep regulatory circuits.

4.2 Future directions

The mechanisms of both EGFR signaling and Hcrt neuronal specification raise several unanswered questions. In particular, it remains unclear how TGF- α interacts with light to induce sleep. What sensory system or signaling pathway intersects with EGFR signaling to convey luminance information? Epistasis analysis with other light-sensitive sleep regulators, such as melatonin or *Period2* (Gandhi et al., 2015; Vatine et al., 2009), is required. It may also be informative to measure global changes in gene expression after light-dark transitions by RNAseq. Similar strategies might provide further insight into Hcrt neuronal specification. The *Lhx9* co-factor(s) that specify Hcrt neurons in the hypothalamus and allow Hcrt neurons to persist in the hindbrain are unknown. Alternative techniques of combinatorial gene overexpression in zebrafish, such as electroporation, might overcome technical limitations

inherent to our current approach. Furthermore, transcriptome-wide characterization of gene expression in ectopic Hcrt neurons isolated by laser capture microdissection could identify potential factors that maintain the Hcrt neuronal fate.

Recently, cellular stress in *C. elegans* has been shown to induce a sleep-like state that is mediated by the EGFR signaling pathway (Hill et al., 2014). It would be very interesting to determine whether a similar phenomenon occurs in zebrafish. Indeed, preliminary experiments suggest that administering a 37°C heat shock for one hour to wild-type zebrafish larvae during the day significantly increases sleep on the subsequent night, compared to no heat shock controls (G. Oikonomou and D. Prober, unpublished data). If this sleep increase is mediated by EGFR in response to heat-induced cellular stress, we predict that *tgfa* *-/-* or EGFR mutant zebrafish would show no increase in sleep after heat shock. These experiments are ongoing.

Finally, it would be interesting to examine EGFR signaling and Hcrt neuronal specification in humans. Overactive, mutant forms of EGFR are the underlying cause of several forms of cancer, which are accompanied by behavioral symptoms that include fatigue (Rich, 2007). EGFR inhibitors, developed to fight these cancers, might have effects on sleep in healthy human adults that could be measured during clinical trials by EEG/EMG recordings. Furthermore, loss of Hcrt neurons in humans is the cause of narcolepsy. Overexpression of Lhx9 in human induced pluripotent stem cell lines might induce or enhance the efficiency of Hcrt neuron specification *in vitro* (Merkle et al., 2015). These synthetically derived neurons could then be transplanted into narcoleptic individuals to determine whether they are capable of rescuing Hcrt neuronal function.

References

- Gandhi, A. V., Mosser, E. A., Oikonomou, G. and Prober, D. A. (2015). Melatonin Is Required for the Circadian Regulation of Sleep. *Neuron*.
- Hill, A. J., Mansfield, R., Lopez, J. M. N. G., Raizen, D. M. and Van Buskirk, C. (2014). Cellular stress induces a protective sleep-like state in *C. elegans*. *Curr. Biol.* 24, 2399–2405.
- Merkle, F. T., Maroof, A., Wataya, T., Sasai, Y., Studer, L., Eggan, K. and Schier, A. F. (2015). Generation of neuropeptidergic hypothalamic neurons from human pluripotent stem cells. *Development* 142, 633–643.
- Rich, T. A. (2007). Symptom clusters in cancer patients and their relation to EGFR ligand modulation of the circadian axis. *J Support Oncol* 5, 167–74– discussion 176–7.
- Vatine, G., Vallone, D., Appelbaum, L., Mracek, P., Ben-Moshe, Z., Lahiri, K., Gothilf, Y. and Foulkes, N. S. (2009). Light Directs Zebrafish period2 Expression via Conserved D and E Boxes. *PLoS Biol* 7, e1000223.