

Gene regulatory factors that control the
identities of specific neuron types in *Caenorhabditis elegans*

Feifan Zhang

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2014

© 2014

Feifan Zhang

All rights reserved

ABSTRACT

Gene regulatory factors that control the identities of specific neuron types in *Caenorhabditis elegans*

Feifan Zhang

The nervous system is the most complex and diverse system of the human body. And so it is in the round worm *Caenorhabditis elegans*. The easy manipulation, maintenance and visualization features of the worm have made it one of the most understood metazoans for linking genetics, anatomy, development and behavior. This thesis work focuses on two aspects during neural development in *C. elegans*: neuronal asymmetry in the ASEL/R gustatory neurons and terminal fate determination of the AIA interneuron as well as the NSM neurosecretory motor neuron. I have cloned and characterized LSY-27, a C2H2 zinc finger transcription factor, which is essential in assisting the onset of the LIM homeodomain transcription factor *lim-6* to repress ASER expressed genes in ASEL. I have also took part in characterizing LSY-12, a MYST family histone acetyltransferase, and LSY-13, a previously uncharacterized PHD finger protein, which cooperate with the bromodomain containing protein LIN-49 and form the MYST complex to both initiate and maintain the ASEL fate. I have also studied the fate determination of several distinct neuronal cell types. I dissected the *cis*-regulatory information of AIA expressed genes and identified that the LIM homeodomain transcription factor TTX-3 is required for AIA fate, possibly together with another yet unknown transcription factor. TTX-3 also acts synergistically with the POU-domain

transcription factor UNC-86 as master regulators for NSM. TTX-3 may also act as the terminal selector for ASK. This work provides extra evidence for the terminal selector concept and further demonstrates that individual neurons use unique and combinatorial codes of transcription factors to achieve their terminal identities, and that the same regulatory factor can be reused as a terminal selector in distinct cell types through cooperation with different cofactors.

TABLE OF CONTENTS

List of Tables and Figures.....	vii
Acknowledgments.....	x
Dedication.....	xii

CHAPTER 1: Introduction

Part I: Neural Development And Cell Fate Determination.....	1
1. Neural complexity.....	1
2. Proneural fate determination.....	3
2.1. Proneural genes	3
2.2. Expression and function of proneural genes.....	4
2.3. Regulation of proneural gene activity	6
3. Pan-neuronal features.....	7
4. Neuronal fate restriction	10
5. Terminal fate specification.....	11
5.1. Terminal gene batteries	11
5.2. <i>Trans</i> -acting factors that act as terminal selectors in <i>C. elegans</i>	12
5.2.1. <i>che-1</i>	12
5.2.2. <i>unc-3</i>	14
5.2.3. <i>unc-30</i>	16
5.2.4. Combinatorial codes of gene regulatory factors.....	17
5.3. Terminal selector regulation in vertebrates.....	20

5.4. Maintenance of differentiated terminal state.....	22
5.5. Parallel regulations.....	23
Part II: Neuronal Asymmetry.....	25
1. Biological asymmetry.....	25
2. Neuronal asymmetry.....	26
2.1. <i>C. elegans</i> neuronal asymmetry.....	28
2.2. Asymmetrical differentiation of the ASE.....	29
2.2.1. Lineage asymmetry.....	29
2.2.2. Functional asymmetry.....	30
2.2.3. Sub-differentiation program asymmetry.....	30
Part III: Summary of Thesis.....	35
Table and Figures.....	38
References.....	45

CHAPTER 2: The LIM and POU homeobox genes *ttx-3* and *unc-86* act as terminal selectors in distinct cholinergic and serotonergic neuron types

1. Abstract.....	60
2. Introduction.....	60
3. Results.....	61
3.1. Expression pattern of <i>ttx-3</i> in the larval and adult nervous system.....	61
3.2. <i>ttx-3</i> controls the differentiation program of the AIA interneurons.....	62
3.3. A shared cis-regulatory signature of AIA-expressed terminal identity features.....	62
3.4. <i>ttx-3</i> controls terminal differentiation of the serotonergic NSM neurons.....	64

3.5. The POU homeobox gene <i>unc-86</i> also controls NSM identity.....	64
3.6. <i>unc-86</i> cooperates with <i>ttx-3</i> to control NSM identity.....	67
3.7. <i>unc-86</i> and <i>ttx-3</i> affect axonal arborization and presynaptic specializations.....	67
3.8. <i>unc-86</i> controls terminal differentiation of the cholinergic IL2, URA and URB sensory, motor and interneurons.....	68
3.9. <i>unc-86</i> cooperates with the ARID transcription factor <i>cfi-1</i> to control IL2 and URA identity.....	69
4. Discussion.....	70
5. Materials and Methods.....	72
6. References	72
7. Supplementary Figures and Tables.....	74

CHAPTER 3: Cloning and characterization of genes required for the specification of the RMDD and RMDV motor neurons

1. Abstract.....	82
2. Results.....	83
2.1. A Genetic Screen that identified RMDD/RMDV mutants.....	83
2.2. Genetic analysis maps these five mutants to three different gene loci.....	84
2.3. Cloning of <i>ot704</i> and <i>ot705</i> using Whole genome sequencing and CloudMap....	85
2.4. Rescue experiments for <i>ot704</i> and <i>ot705</i>	86
2.5. Expression pattern of <i>unc-42</i> , <i>hlh-16</i> and <i>cnd-1</i> in RMDD/RMDV.....	86
3. Future experiments and discussion (see Chapter 6).....	135
4. References.....	88
5. Figures and tables.....	89

CHAPTER 4: A left/Right asymmetric neuronal differentiation program is controlled by the *Caenorhabditis elegans* LSY-27 Zn finger transcription factor

1. Abstract.....	99
2. Introduction.....	100
3. Results and discussion.....	101
3.1. <i>ot146</i> is an allele of the LIM homeobox gene <i>lim-6</i>	101
3.2. <i>ot108</i> affects a member of a C2H2 Zn-finger protein family.....	101
3.3. <i>ot108</i> is an altered function allele.....	102
3.4. Expression pattern and timing of action of <i>lsy-27</i>	103
4. Conclusions.....	104
5. References.....	105
6. Supplementary Information.....	107

CHAPTER 5: Maintenance of Neuronal Laterality in *Caenorhabditis elegans* Through MYST Histone Acetyltransferase Complex Components LSY-12, LSY-13 and LIN-49

1. Abstract.....	118
2. Introduction.....	119
3. Results and discussion.....	119
3.1. Laterality defects in animals lacking <i>lsy-12</i> , a histone acetyltransferase.....	119
3.2. <i>lsy-12</i> is continuously required to maintain ASE laterality.....	121
3.3. Other known components of MOZ/MORF-type HATs also display Lsy phenotypes.....	121

3.4. The <i>die-1</i> Zn finger transcription factor is also continuously required to maintain ASE laterality and is a candidate recruiter of the MYST complex.....	123
4. References.....	124
5. Supplementary Information.....	125

CHAPTER 6: Discussion and future directions

1. Combinatorial codes of terminal selector regulation.....	131
1.1. Continuous requirement of terminal selector transcription factors.....	131
1.2. Target gene specificities	132
1.3. Looking for cofactors.....	132
1.4. TTX-3 function in AIN.....	133
1.5. Considerations in other systems.....	134
2. Specification of the RMD neuron class	
2.1. Terminal selector for the RMD neuron class.....	135
2.2. Transcriptional regulation of gene expression.....	136
2.3. Proneural roles of <i>cnd-1</i> and <i>hlh-16</i>	137
2.4. The screen.....	139
3. ASE asymmetry.....	140
3.1. Considerations on ASE screens: manual clonal vs. automatic non-clonal.....	140
3.2. Whole genome sequencing.....	142
3.3. LSY-27 function in triggering <i>lim-6</i> initiation.....	143
3.4. The LSY-12/LSY-13/LIN-49 complex and its function.....	143
4. References.....	145

APPENDICES

Appendix 1.....	148
Appendix 2.....	174
Appendix 3.....	180

LIST OF TABLES AND FIGURES

CHAPTER 1

Figure 1: Terminal selectors directly control terminally differentiated genes.....	38
Figure 2: An example of parallel regulation regulons of the ASE gustatory neuron.....	40
Figure 3: The human brain is asymmetric at both morphological and functional level....	41
Figure 4: Asymmetric lineage formation of the ASE neuron pair.....	42
Figure 5: The asymmetric ASEL/R differentiation program.....	43
Table1: Examples of invertebrate and vertebrate regulons in the nervous system.....	44

CHAPTER 2

Figure 1: Expression pattern of the <i>ttx-3</i> LIM homeobox gene.....	61
Figure 2: <i>ttx-3</i> affects terminal differentiation of the AIA neurons.....	63
Figure 3: Coregulation of AIA-expressed genes by two cis-regulatory motifs.....	64
Figure 4: The effect of <i>unc-86</i> and <i>ttx-3</i> on the serotonergic identity of the NSM neurons.....	65
Figure 5: The effect of <i>unc-86</i> and <i>ttx-3</i> on other identity features of the NSM neurons.	66
Figure 6: <i>Cis</i> -regulatory analysis of NSM identity specification.....	68
Figure 7: <i>unc-86</i> and <i>ttx-3</i> affect NSM morphology.....	69
Figure 8: <i>unc-86</i> and <i>cfi-1</i> control cholinergic IL2 identity.....	70
Table 1: Summary of the effects of <i>ttx-3</i> and <i>unc-86</i> null mutants on terminal NSM identity markers.....	67
Suppl. Figure 1: Analysis of the NSM neurons.....	74
Suppl. Figure 2: AIA morphology in wildtype and <i>ttx-3</i> (<i>ot22</i>) mutant animals.....	75

Suppl. Figure 3: BH4 pathway reporters.....	76
Suppl. Figure 4: <i>unc-86</i> controls the identity of the cholinergic URA and URB neurons.....	77
Suppl. Table 1: Rescue of NSM and AIA differentiation defects of <i>ttx-3</i> mutant animals.....	78
Suppl. Table 2: Strains and transgenes used in this study.....	79
Suppl. Table 3: Probe sequences for gel shift analysis.....	81

CHAPTER 3

Figure 1: Class I and II phenotypes of RMDD/RMDV mutant animals.....	89
Figure 2: Scattered ratio plots of the <i>ot704</i> and <i>ot705</i> SNP reads.....	91
Figure 3: Overview of the <i>hlh-16</i> and <i>cnd-1</i> locus and relevant alleles.	93
Figure 4: Preliminary lineaging representation <i>cnd-1</i> expression in the RMDD and RMDV lineage.....	94
Table 1: Quantification of <i>unc-42</i> , <i>ot704</i> and <i>ot705</i> mutant phenotypes.....	96
Table 2: <i>ot704</i> rescue experiments.....	97
Table 3: <i>ot705</i> rescue experiments.....	98

CHAPTER 4

Figure 1: <i>lsy</i> genes and mutant phenotypes.	101
Figure 2: <i>lsy-27</i> is a new C2H2 Zn-finger protein.....	104
Figure 3: <i>lsy-27</i> is expressed and acts during the initiation but not during the maintenance phase of left/right asymmetry control.....	105
Table 1: <i>Lsy</i> phenotypes of <i>lim-6</i> and <i>lsy-27</i>	102
Table 2: Transformation rescue and RNAi analysis.....	103

Suppl. Figure 1: Updated <i>lim-6</i> locus and the location of the <i>ot146</i> allele.....	108
Suppl. Figure 2: Sequences of LSY-27 and its two paralogs, ZTF-25 and ZTF-28.....	109
Suppl. Figure 3: Expression pattern of the <i>lsy-27</i> paralog <i>ztf-25</i>	110
Table S1: Whole genome sequencing setting and results.....	113
Table S2: Mutant analys of <i>lsy-27</i> paralogs.....	114
Table S3: Retired and novel <i>lsy</i> gene names.....	115
Table S4: Final summary of mutant classes & genes.....	116

CHAPTER 5

Figure 1: <i>lsy-12</i> , a MYST-type histone acetyltransferase, affects ASE laterality.....	120
Figure 2: <i>lsy-12</i> is continually required in ASE but may also act early.....	121
Figure 3: The MYST-complex component <i>lsy-13</i> affects ASE laterality.....	122
Suppl. Figure 1: Mapping <i>lsy-12</i>	128
Suppl. Figure 2: The MYST complex component <i>lin-49</i> affects ASE laterality.....	129
Suppl. Table 1: Laterality phenotype of <i>lsy-12</i> transgenic animals.....	130

ACKNOWLEDGMENTS

It took me long to finish my Ph.D. thesis but to write this acknowledgement is as difficult, as I am indebted to so many people without whom I could've never made it this far. I would like to say thank you to every single one of you.

I must first thank Dr. Oliver Hobert for being such an amazing mentor, who provided me the opportunity of working in his lab and the freedom to pursue projects of my interest, supported me with endless encouragement, inspired me with his scientific enthusiasm, creativity and optimism, and has always kept his office open for discussions and suggestions.

I truly appreciate the tremendous help from past and current members of the Hobert lab. I must give thanks to Maggie O'Meara, who introduced me to worm genetics and went out of her way to lend a hand; to Sumeet Sarin, who never hesitated to share his strains and experience; and to Maria Doitsidou, Nuria Flames, and Richard Poole, who were never sick of answering my naïve questions and showed me what it takes to be a true scientist. I would like to give special thanks to Marie Gendrel, my French baymate, for exchanging ideas and sharing thoughts on everything, for organizing numerous amazing parties and outings, and for keeping life both in and out of the lab from getting tedious; to Inés Carrera, for always going above and beyond to help; and to Kelly Howell and Meital Oren for being good friends and for sharing the passion for shopping. I also thank Alex Boyanov for administrative support and Chi Chen for expert injection assistance. I was fortunate to have known you all.

I'm grateful to all collaborators involved in my projects over the years. Eileen Flowers retrieved the *lsy-27* allele, which I later worked on; Henry Biglow developed

MAQGene that I used to analyze *lsy-27* sequencing results and Gregory Minevich developed CloudMap that facilitates analysis on the whole genome sequencing; Irini Topalidou taught me yeast one-hybrid assays and troubleshot with patience. I especially thank Nuria Flames for her kindness in offering strains as well as her invaluable advice. I also thank Abhishek Bhattacharya and Pat Gordon from our lab, Jessica Nelson from the Colón-Ramos lab, and Namiko Abe from the Mann lab for help on the *ttx-3/unc-86* project.

I would like to extend my gratitude to members of the McCabe lab, the Grueber lab, the Grishok lab, the Chalfie lab and the Greenwald lab, who provided great comments and feedbacks on my projects during joint meetings, and created the wonderful and collaborative scientific community at Columbia University. I would like to thank Dr. Iva Greenwald in particular, who not only generously lent her lab reagents and equipment to the Hobertians, but also shared enormous insightful ideas and experience with us, both scientifically and personally.

I also would like to thank my thesis committee members, Dr. Iva Greenwald, Dr. Martin Chalfie, Dr. Rene Hen, Dr. Daniel Colón-Ramos for their insight and guidance. I thank Dr. Daniel Kalderon for being on my qualifying committee. I am happy to have been part of the Biological Sciences program and I thank Sarah Kim and Jaya Santosh from the Biology office for administrative assistance.

Lastly, to my mom and dad, for their unconditional love and support for me to pursue my dreams, even if it involves being thousands of miles away from home. To my husband Zhu, for always believing in me and encouraging me to follow my passion both

in science and in life. And to dear baby Hanna for being such a good little girl for the past nine months, and for making our family complete.

DEDICATION

To my family, for their endless love and support.

CHAPTER 1: INTRODUCTION

Part I: Neural Development And Cell Fate Determination

1. Neural complexity

Neurons are individual and autonomous units that constitute the nervous system. All neurons share the same pan-neuronal features at certain level, and are physically connected by cellular extensions (axons/dendrites) that transmit information through presynaptic and postsynaptic specializations. The brain is often described as the most sophisticated system in the universe. As Santiago Ramón y Cajal stated in his Nobel lecture in 1906,

“It would be very convenient and very economical from the point of view of analytical effort if all the nerve centres were made up of a continuous intermediary network between the motor nerves and the sensitive and sensory nerves. Unfortunately, nature seems unaware of our intellectual need for convenience and unity, and very often takes delight in complication and diversity.”

Cajal’s groundbreaking work opened the era of studies on the nervous system. For decades, scientists have been fascinated by the depth and intricacy of the most complex organ across almost all animal species. If you are already astounded by the infinite Milky Way galaxy that consists of hundred billions of stars, the human brain has almost as many neurons, but within only 1.5 kilograms of flaccid mass, not to mention the hundreds of trillions of processes and interconnections among individual neurons (Herculano-Houzel, 2009; Koch and Laurent, 1999).

The complexity of the nervous system does not merely lie in its vast number and there doesn’t seem to be a correlation (Herculano-Houzel, 2009). The nervous system is

highly diverse in neuronal cell types. Take the very simple round worm *C. elegans* for example, among a total of only 302 neurons in an adult hermaphrodite, there are at least 118 different neuronal classes (Hobert, 2005). At the structural level, its genome contains genes encoding 80 different types of potassium channels, 90 neurotransmitter-gated ion channels, and around 1000 G-protein coupled receptors (Bargmann, 1998). Today, despite enormous amount of efforts made over the past hundred years, it remains obscure how many distinct cell types exist in the brain, with only crude estimates available. Extensive microscopic analysis has estimated that the mammalian brain consists approximately 500 to 1000 cell groups/regions/nodes with their own unique sets of axonal outputs. If five cell types per group is considered, the total number is then around 2500 to 5000, and many more if neuronal-subtype classifications are taken into account. The overall complexity can be even more staggering if considering the fact that each neuronal cell branches and projects onto other cell types (ten to twenty according to current observations, ranging from two to hundreds) (Bota *et al.*, 2003).

How does such complexity evolve? It is not difficult to imagine that only a system with many components not essential for survival can successfully undergo the selection pressure of gene mutation and modification that potentially harms viability. This drives the brain into evolving specialized circuits, parallel pathways and redundant mechanisms. Another point to take into account is that in order to increase efficiency of the incredibly fast information-processing system, the spatial wiring network of the neurons needs to be engineered and optimized to be precise, which enforces complexity during evolution (Koch and Laurent, 1999). It has been proposed that the existence and expression of consciousness as well as higher cognitive development and sociality are

key players (Dunbar and Shultz, 2007; Tononi and Edelman, 1998) and higher cognitive functions seem to be associated with cortex expansion not only in neuron number, but also in neuronal diversity and connectivity (DeFelipe *et al.*, 2002; Rakic, 2009; Roth and Dicke, 2005). Today's brains result from 0.6 to 1.2 billion years of metazoan evolution. This vast time span allowed incredible changes and adaptations. Although what we know now is still tip of the iceberg, the appreciation of neuronal complexity should help better understand what makes us human.

2. Proneural fate determination

2.1. Proneural genes

The Basic Helix-Loop-Helix is characterized by two α -helices connected by a loop structure. bHLH genes encode proteins containing such motifs that are necessary and sufficient for promoting the generation of neuron progenitors, and are therefore considered “proneural”. Proneural genes were first identified in flies lacking a subset of the bristles back in the late 1920s, followed by later discoveries that identified a complex of genes involved in early regulation of neuronal fate development (Garcia-Bellido, 1979). Most of these genes are expressed in clusters by groups of cells in the ectoderm before neural differentiation (Campuzano and Modolell, 1992), and further analysis suggested that such defects in bristle formation is due to the initial differentiation decision instead of any defects during the differentiation process itself.

In flies, there are four components in the *achaete-scute(asc)* complex, *achaete (ac)*, *scute (sc)*, *lethal of scute (lsc)* and *asense (ase)* (Villares and Cabrera, 1987). A

more recent PCR screen identified *atonal* (*ato*), which belongs to a distinct bHLH family (Jarman *et al.*, 1993b). In vertebrates, many genes have been found related to *asc* and *ato*. The vertebrate *asc* family contains *Ash1*, *Mash2*, *Xash3* and *Cash4*, and the *ato* related genes are categorized into the neurogenin (Ngn) family, the NeuroD family and the Olig family according to their consensus in specific residues within the bHLH domain (Bertrand *et al.*, 2002).

Similarly, the *C. elegans* bHLH genes are divided into two families, the *achaete-scute* related family and the *atonal* related family, which can be further assigned to the Neurogenin group, the NeuroD group and the Atonal group (Ledent and Vervoort, 2001). There are six *achaete-scute* related genes: *hlh-3*, *hlh-4*, *hlh-6*, *hlh-12*, *hlh-14* and *hlh-19*. It was predicted that there are approximately 42 bHLH genes in *C. elegans* (Reece-Hoyes *et al.*, 2005) that may play essential roles in proneuronal fate determination and nervous system patterning. Several bHLH factors have also been indicated to be involved in later neurogenesis decisions instead of being exclusively proneural.

2.2. Expression and function of proneural genes

Rich collections of fly mutants have allowed extensive understanding of proneural gene functions. It was shown that the formation of most fly embryonic and adult external sense organs as well as a subset of neuroblasts in the CNS require *ac* and *sc* gene activity, while *lsc* is essential for neuroblast generation from the CNS primordium (Jimenez and Campos-Ortega, 1990). Instead of ectodermal cells, the fourth component of the complex, *ase*, is expressed in all progenitors of the PNS and CNS only after they have been produced (Jarman *et al.*, 1993a), and doesn't seem to be required

for the selection of the progenitors. The other family of bHLH factor *ato* is responsible for internal chordotonal organ formation and retinal founder photoreceptor development. Functions of the bHLH factors in vertebrates are highly diverse. The *asc* and *Ngn* are conserved in their proneural roles, other genes seem to have adopted divergent functions related to neuronal fate specification (Bertrand *et al.*, 2002).

The *atonal* family bHLH factor LIN-32 is the most extensively studied bHLH factor in *C. elegans*. It is required for ray formation and is expressed until the terminal division. Loss-of-function and ectopic expression analysis have shown that its function is necessary and sufficient for ray sublineage entry and is therefore a proneural gene (Zhao and Emmons, 1995). It forms heterodimer with the HLH2, the *C. elegans* ortholog of E protein/*Daughterless*, to bind to the E-box-containing elements and activate targets at multiple steps (Portman and Emmons, 2000). HLH-2 is also a candidate for dimerization with the *achaete-scute* homolog HLH-3 in neuronal precursors but not muscle cells (Krause *et al.*, 1997). Proneural genes also take part in establishing neural lineage asymmetry during development. The Neurogenin homolog NGN-1 and HLH2 are expressed in the mother cell of MI, and loss of either of them results in e3D-like fate usually generated from its counterpart lineage on the left side. This asymmetry expression of NGN-1 and HLH-2 depends on the Otx/Otd homeodomain protein, which is expressed in the MI grandmother cell and MI mother cell cell-autonomously (Nakano *et al.*, 2010).

During motor neuron specification, the *C. elegans* NeuroD homolog CND-1 plays a critical role. It is expressed very early in the embryo in many neuronal descendants of the AB lineage and becomes undetectable in most terminally

differentiated neurons. Loss of *cnd-1* results in motor neuron number reduction, position change, neuronal feature defects and terminal selector spatial expression alteration. The wide range of defects observed in *cnd-1* mutants suggest that it may have combined the roles of several vertebrate neurogenic proteins and may be an ancestral protein that combines the function of the vertebrate Neurogenin and NeuroD (Hallam *et al.*, 2000). Aside from its earlier roles in neuronal lineage generation, neuroD also takes part in regulating terminal features of differentiated neurons (Hallam *et al.*, 2000). It may also play roles in mature neurons as its expression has been seen in abundance in fully differentiated *Xenopus* adult brain structures (Lee *et al.*, 1995), and NeuroD deficient mice rescued with early and transient expression of neuroD in the embryonic brain still display severe neurological phenotypes afterwards (Miyata *et al.*, 1999).

Another example is the *C. elegans* *Achaete-Scute*-like bHLH gene *hlh-14*, which has been shown to act together with HLH-2 to specify neuroblast lineages and to promote neurogenesis (Frank *et al.*, 2003). It was also pulled out from a recent RNAi screen in search of neural specification factors of the ASE gustatory neuron. *hlh-14* is bilaterally expressed in the ASEL/R lineage despite their asymmetric lineage origins and is required for neurogenesis of the ASE as well as several other neurons in the same branch. 4D microscopy revealed that *hlh-14* mutants display hypodermal transformations and mis-positioning defects with higher frequency in more posterior descendants of the ABalpppp/ABpraap neuroblasts in the ASE lineage. Together with expression data it indicates that HLH-14 is possibly a binary switch to determine neuronal fate versus hypodermal fate (Poole *et al.*, 2011).

2.3. Regulation of proneural gene activity

Most proneural proteins act as transcriptional activators and only a few are repressors. They heterodimerize with ubiquitously expressed bHLH proteins (such as E2A/HEB/E2-2 in vertebrates, Daughterless in flies, and HLH-2 in worms) and specifically bind to the CANNTG core motif in the E-box. Subsequent upregulated Notch ligand results in lateral inhibition through activating Notch signaling pathway in neighboring cells. This leads to the expression of repressors (*Espl* genes in flies and Hes/Her/Esr in vertebrates) that down regulate proneural gene expression and usually results in epidermal fate in neighboring cells (Artavanis-Tsakonas *et al.*, 1999).

Once a neural progenitor is selected, positive-feedback mechanisms are required to increase and maintain the level of proneural genes. This can be achieved through activation of downstream transcription factors that in turn upregulate proneural gene expression. For example, the *Drosophila* zinc finger protein Senseless represses *Espl* repressor genes to further inhibit the Notch signaling pathway (Nolo *et al.*, 2000). Autoregulation may also play a role in maintaining high level of proneural gene activity. For example, the SMC enhancer of *Scute* contains functional E-boxes that mediate autoregulation in order to obtain high level of the proneural protein (Culi and Modolell, 1998). EGF (Epidermal growth factor) signaling activated by *asc* genes can also act on the SMC enhancer to elevate *Scute* level (Culi *et al.*, 2001).

3. Pan-neuronal features

The basic cellular organization of a neuron is not different from others. They all have similar organelles and subcellular components, such as Golgi apparatus,

mitochondria and a variety of vesicular structures. However, almost all neurons, regardless of the species and their origin, share certain defined features that make them unique from other cell types morphologically and functionally, such as cellular extensions like axons or dendrites that connect one neuron to another physically, and synapses made of complex pre-and post- synaptic specializations that transmit information from one neuron to another (Hobert *et al.*, 2010). One tempting theory is that there might be a defined “pan-neuronal gene battery” that is shared among all neurons versus non-neuronal cells. However evidence suggests that this may not be the case. For example, proteins that are more restricted to neuronal cells may also be preset in other cell types (Iwasaki *et al.*, 1997; Sieburth *et al.*, 2005), including synaptic vesicle-associated proteins, ion channels, cytoskeleton proteins and so on. Moreover, none of the characteristic neuronal proteins are expressed across all neuronal species in a given organism. Therefore, those genes referred to as “pan-neuronal” might be better termed “broadly expressed” neuronal genes instead.

How are the “pan-neuronal” genes regulated? Several studies indicates that the overall neuronal features seems genetically separable from a neuron’s terminal identity, as the loss of terminal selector genes that define the terminal identities of a neuron does not result in loss of its pan-neuronal features (Altun-Gultekin *et al.*, 2001; Flames and Hobert, 2009; Uchida *et al.*, 2003). Bioinformatics studies in *C. elegans* identified a ten-nucleotide *cis*-regulatory motif named “N1 box” preferentially present in the promoters of many broadly expressed genes. Replacement of this site with a LexA site almost completely abolished or diminished neuronal expression. This suggests that there might be global *trans*-acting factors that coordinately control the expression of some or even

all broadly expressed neuronal genes (Ruvinsky *et al.*, 2007). Nevertheless, other mechanisms must exist because the N1 box is not present in all broadly expressed neuronal genes. Pan-neuronal expression may also be accomplished through the “piece-meal” or “modular” manner, in which case terminal selectors for different neuron types can bind additively to different elements of the promoter of a pan-neuronal gene to assemble much broader expression. For instance, there are at least six regulatory elements (two in the 5’ upstream region and four within the first intron) that are able to drive expression of the *C. elegans* RIC-4 protein in particular neuronal cell types (Hwang and Lee, 2003). In flies, it is only the combination of the core promoter and far upstream enhancer elements as well as an intronic enhancer that could give rise to high expression level of β 1 tubulin in most neuronal cells in the CNS (Kohler *et al.*, 1996).

However, there does not seem to be a universal strategy employed to achieve broad neuronal gene expression. Hobert *et al.* proposed that parallel and diverse regulons that specify the terminal gene batteries of individual neurons might exist. There may or may not be a clear separation between the regulation of neuronal identity features, pan-neuronal features and pan-sensory features, each of which could be regulated by a distinct mechanism or by the same regulator (Etchberger *et al.*, 2007; Hwang and Lee, 2003), and multiple selector genes may act on different part of the specific identity. Moreover, neurons can respond to extrinsic signals and change their gene expression profiles, which are also proposed to be organized into regulons (Hobert *et al.*, 2010).

4. Neuronal fate restriction

After specification of the neural progenitors that are limited in developmental potential, the next question is how to terminally differentiate the progenitor cells into the mature postmitotic neurons, which is very often accompanied by the gradual restriction of cell fates. This can take place at various stages and time, either before or after the final cell-cycle exit of a progenitor cell. The neurogenesis of *C. elegans* does not seem to require much of temporal control, as they are generated from stereotypic cell divisions. In *Drosophila* and vertebrates, the specification of the neuronal fate is dependent on the interplay between two sets of determinants: extrinsic (cell non-autonomous) signals and intrinsic (cell-autonomous) mechanisms. A neuron receives extrinsic signals presented by the environment both temporally and a cell can acquire its fate by utilizing various signaling strategies (Edlund and Jessell, 1999).

These extrinsic signals are then cooperated and integrated by the nervous system. The neuron gradually loses dependence on extrinsic signals but rely more on its intrinsic signals to finally establish the identity and function of each single cell that later assembles the neuronal circuits. How does the progression/transition from extrinsic to intrinsic signaling occur? One mechanism involves persistent activation of intracellular cytoplasmic transduction proteins that are subject to post-translational regulation, such as photolytic processing and phosphorylation dependent activation of effector kinases that serve as mediators of extrinsic signaling. As a matter of fact, it has been suggested that the function of a lot of transcription factors functioning in neuronal cells are dependent on their state of phosphorylation (Fowles *et al.*, 1998; Jacobs *et al.*, 1998). Subsequent autoregulation of transcription factors that play a role in cell fate induction

and maintenance usually involves a positive feedback. Later, long term stabilization of initial state of gene expression pattern in either active or inactive state must be achieved. The final cell-cycle exit also contributes to neural fate differentiation, and may prevent neuronal cells from further receiving or responding to extrinsic signals. (Edlund and Jessell, 1999).

5. Terminal fate specification

Although much is known about the induction and specification of the developmentally restricted neuron progenitor cells, less is known about how the terminal differentiation of a mature post-mitotic neuron with a dedicated fate is achieved. The “terminal selector” concept is currently widely accepted and has been proposed to not only initiate but also maintain the terminal features of a neuron across species from invertebrates to vertebrates.

5.1. Terminal gene batteries

The last step of the neuronal cell fate specification is the expression of specific terminal gene batteries that encode proteins determining the fate and functions of a neuron. These proteins include neurotransmitters synthesis enzymes, neurotransmitter receptors, ion channels, signaling proteins, cytoskeleton proteins, adhesion molecules and *et cetera* (Figure 1A). However, it seems that none of these terminal features are exclusive to a particular neuron, and therefore it is the combination of those features that determines a neuron’s identity (Figure 1B).

The use of *C. elegans* as a model animal has greatly facilitated the study of cis-regulatory mechanisms of terminal differentiated genes *in vivo*. Regulatory elements are fused to GFP to examine expression in specific neuron types, either by molecular cloning or reporter gene fusion (Boulin *et al.*, 2006; Chalfie *et al.*, 1994; Hobert, 2002). The transparency of the worm allows fast and direct visualization of the fluorescent protein in the whole animal. Comparison between the wild-type and the mutant constructs with sites mutated can be made to assess the functionality. Current findings suggest that *cis*-regulatory motifs usually reside within 1 kilobases to the start site, but there emerging evidence indicates that they could be located farther away (a few kilobases upstream) or within introns (Doitsidou *et al.*, 2013). In the following section, I will give several examples of known terminal selectors that have come to light from recent studies in the worm.

5.2. Examples of *trans*-acting factors that act as terminal selectors in *C. elegans*

There is only a total number of 302 neurons in hermaphrodites and each single one of them has been precisely mapped with no variance (White *et al.*, 1986). The simplicity and the relative short life cycle of the worm allows fast and large-scale forward genetic screens, which have contributed to the identification of a large number of *trans*-acting factors that are involved in neuronal differentiation. In combination of *cis*-regulatory analysis, these mutants have led to discoveries that reinforce the “terminal selector” hypothesis (Table1).

5.2.1. *che-1*

The C2H2 Zinc finger-containing transcription factor CHE-1 is orthologous to the *Drosophila* GLASS protein, which is required for the differentiation of photoreceptors. The original “*che-1*” alleles were isolated in a screen seeking animals defective in chemotaxis to NaCl (Dusenbery *et al.*, 1975). GFP reporters revealed that CHE-1 is expressed in the ASE neuron but *che-1* mutants have no significant structural defects under the microscopes, and both ASEs remain at the correct position. *Che-1* mutants are chemotaxis defective to water-soluble attractants including Na⁺, Cl⁻, biotin and cAMP but not to volatile odorants (Bargmann *et al.*, 1993), and was therefore suggested to be a transcription factor mainly required for the differentiation of ASE specific terminal fates (Uchida *et al.*, 2003). *che-1* mutants lose ASE terminal gene expression, and ectopic expression of *che-1* in other sensory neurons results in ectopic expression of the ASE specific *gcy-5::gfp* reporter.

Comparison between *SAGE* library analyses on isolated *gcy-5::gfp* positive ASE neurons and the *gcy-8* positive AFD thermosensory neurons derived from embryos defined a specific transcriptome of the ASE neuron, which consists of a broad and unbiased spectrum of neuron-type-specific gene profile that distinguishes ASE from other neurons. Further promoter dissection on a subset of the ASE-expressed genes reveals a conserved ASE motif that is shared by all but one *cis*-regulatory region, with a 6-bp core at its 5' end (GAADCC) followed by an additional A/T rich sequence. Deletion or mutation of this motif in multiple *cis*-regulatory contexts completely abolished ASE expression in most cases. The ASE motif is usually within 1 kb upstream of the start codon, and is mostly present as a single copy with no obvious orientation preference on a particular strand. Fusing this ASE motif in front of a small promoter of

another gene only expressed in the AWC shows strong and consistent ASE expression in AWC. Single or multimerized ASE motifs with minimal flanking sequences are also able to drive GFP expression in the ASE (Etchberger *et al.*, 2007).

Evidence suggests that CHE-1 directly controls the ASE motif via the third and the fourth of its Zn Fingers. The predicted CHE-1 binding site by probabilistic cognition code for C2H2 zinc finger transcription factors (Benos *et al.*, 2002) shares striking similarity with the experimentally verified and derived ASE motif, and the CHE-1 *Drosophila* ortholog GLASS is 100% identical to CHE-1 in the DNA contacting residues within the Zn fingers. *In vitro* electrophoretic mobility shift assays further support this idea. Full-length CHE-1 protein expressed and purified from Bacteria is able to bind the ASE motif present in all tested genes in a sequence-specific manner. The promoter of *che-1* also contains an ASE motif that can be bound by CHE-1 itself selectively as shown *in vitro*. This ensures the persistence expression of CHE-1 throughout the life of the animal, which is likely achieved through autoregulation. In *che-1* mutants, transcriptional *gfp* reporter for *che-1* fails to express.

5.2.2. *unc-3*

Cholinergic neurons are defined by the expression of genes that are involved in the synthesis, packaging and recycling of the neurotransmitter acetylcholine. *C. elegans* has eight classes of ventral nerve cord neurons that control locomotion through communicating with target muscles, out of which six are cholinergic (DA, VA, DB, VB, AS and VC). The COE (Collier/Olf/EBF)-type transcription factor UNC-3 was first identified as a factor required for wild-type locomotion, axon guidance and proper

differentiation of ventral cord motor neurons (Brenner, 1974; Herman, 1987; Prasad et al., 1998). Recent studies have shown that it is the terminal selector for a subset of the cholinergic motor neurons in the VNC. *unc-3* mutants are defective in expression of a majority of the terminal differentiation genes (26 out of 30) in the A-, B- and/or AS-type motor neurons. These genes include the cholinergic pathway components, putative Ach autoreceptors, neurotransmitter receptors, ion channels, gap junction proteins, signaling proteins, axon path finding factors and so on. Pan-neuronally expressed genes (*rab-3*, *unc-119* and *ref-1*) are not affected in *unc-3* mutants, and molecular markers for other classes of ventral nerve cord neurons, the GABAergic neurons, are not ectopically expressed. This suggests that the neurons may still be physically present but remain in an undifferentiated state. UNC-3 is sufficient to induce cholinergic fate in other cells. Misexpression of UNC-3 in the glutamatergic sensory neurons ASE and AWC leads to ectopic expression of cholinergic markers, and misexpression of UNC-3 in D-type GABAergic motor neurons is also capable of inducing cholinergic fates in those cells (Kratsios *et al.*, 2012).

A consensus COE motif present in UNC-3 responsive genes was identified by Matinspector (Genomatrix) and Ebf1 chromatin immunoprecipitation analysis of the mouse B cells. UNC-3 directly regulates the cholinergic gene battery in the A-, B- and AS-type neurons via the phylogenetically conserved COE motifs present in the regulatory regions of those genes. Small fragments that contain these motifs are necessary and sufficient for expression in those cholinergic VNC neurons. Mutations of such sites lead to failure in correct reporter gene expression, while expression in other neuron types remain unaffected.

Kratsios *et al.* showed that UNC-3 is required not only to initiate but also to maintain the terminal features of the cholinergic fate. A fosmid-based reporter that contains about 40 kilobases of the genomic context surrounding the *unc-3* locus display persistent expression in the A-, B- and AS- type neurons throughout adulthood. The continuous requirement for UNC-3 is further validated by the fact that heat shock-induced UNC-3 activity around mid-larval stages is able to rescue *unc-3* mutant phenotypes.

The function of UNC-3 is conserved across phylogeny. *Ciona intestinalis* also contains a single UNC-3 ortholog COE that is expressed in cholinergic motor neurons, as assessed by the expression of the VAcHT-ChAT locus that contains a copy of the COE recognition motif in the regulatory region. The *C. intestinalis* COE is sufficient and necessary to induce cholinergic fate, as VAcHT expression of animals expressing a dominant negative form of COE is severely affected and cholinergic motor neurons seem to adopt a glia-like morphology other than remain cholinergic. Misexpression of the COE protein in non-cholinergic neurons is able to induce ectopic expression of the VAcHT reporter. In addition, UNC-3 orthologs are present in vertebrate species. In mouse, the majority of them are expressed in cholinergic motor neurons in the spinal cord (Kratsios *et al.*, 2012).

5.2.3. *unc-30*

GABAergic neurotransmission is widely used across species from invertebrates to vertebrates. In worms, GABAergic neurons can be categorized into several different classes, consisting of the single AVL and DVB, and RIS interneuron, four RME motor

neurons, six DD (dorsal D) neurons, and thirteen VD (Ventral D) motor neurons. These neurons are required for different functions and behaviors of the worm. The Pitx family UNC-30 homeodomain protein was first identified as required for the differentiation of two classes, the DD and VD type neurons. Mutants of *unc-30* display the shrinking defect, which phenocopies worms with DD and VD ablated, while the RME, AVL and DVB related foraging behavior as well as the defaecation cycle phenotype remain *wild type* (McIntire *et al.*, 1993). Electromotility gel shift assays as well as DNAase I footprinting analysis further demonstrates that UNC-30 directly acts on the core consensus sequence 5'-TAATCC-3' in the promoter regions of GABA synthesis/packaging pathway genes, glutamic acid decarboxylase UNC-25 and vesicular GABA transporter UNC-47 (Eastman *et al.*, 1999). Microarray analysis on cRNA library that yielded a full spectrum of gene battery of the GABAergic neurons as well as analysis of the 5' upstream regions of six of the UNC-30 target genes further revealed the consensus sequence as "WNTAATCHH", which is significantly enriched in the regulatory regions of UNC-30 target genes (Cinar *et al.*, 2005). Therefore, UNC-30 likely acts as a terminal selector in the VD- and DD- type GABAergic neurons. However, it does not seem to be required for the specification of other types of GABAergic neurons, and how the terminal fates of these neurons are regulated remain unclear.

5.2.4. Combinatorial codes of gene regulatory factors

Instead of being expressed in only one type of neurons, transcription factors are very often expressed in at least several more cell types. It is easy to imagine that it is economically more efficient to reuse the same transcription factor in different cell types.

In such cases, the terminal fate of a cell is adopted through cooperation between distinct transcription factors. In other words, how the terminal fate of a cell is determined depend on the combinatorial codes of transcription factors instead of one single master regulator (Figure 1B).

The POU homeodomain protein UNC-86 and the LIM-homeodomain protein MEC-3 are both required for the terminal differentiation of the touch neurons (ALML/R, AVM, PLML/R and PVM) that are responsible for gentle touch to the body in a cooperative manner (Finney and Ruvkun, 1990; Way and Chalfie, 1988). The transcription of MEC-3 is dependent on UNC-86 (Xue *et al.*, 1992). MEC-3 and UNC-86 then bind synergistically as a heterodimer to the promoter of *mec-3* to activate robust transcription (Lichtsteiner and Tjian, 1995). They also act in a cooperative manner to regulate downstream target genes required for the mechanosensory function of the touch cells, by direct binding to a conserved motif, AATGCAT (Duggan *et al.*, 1998; Zhang *et al.*, 2002).

Similarly, the homeodomain transcription factors TTX-10 and CEH-10, which are homologs for the mouse Lhx2/9 and Chx10, are both required for the terminal differentiation of the AIY interneuron (Figure 1B). TTX-3 is required to directly restrict *ceh-10* expression to AIY during the terminal division of the AIY mother (Bertrand and Hobert, 2009). Although expression of CEH-10 is only observed during embryogenesis and fades after hatching and TTX-3 expression is maintained throughout the life of the animal, both TTX-3 and CEH-10 are required for terminal differentiation of AIY (Altun-Gultekin *et al.*, 2001). They form a heterodimer that binds to the AIY motif, which is present in the regulatory regions of the AIY gene battery (Wenick and Hobert,

2004). It is interesting that although AIY is also cholinergic, it utilizes a completely different motif from that of the A- and B- type cholinergic neurons controlled by *unc-3*, suggesting that the *cis*-regulatory regions of terminal genes expressed in multiple cell types may contain multiple distinct binding sites for different sets of terminal selectors depending on the cellular context (Hobert, 2011).

The regulatory logic of the dopaminergic neurons in *C. elegans* is even more complex. There are eight dopaminergic neurons in the worm, namely CEPVL/R, CEPDL/R, ADEL/R and PDEL/R. The E-twenty six (ETS) transcription factor AST-1 was the first player identified to be required for both the initiation and maintenance of dopaminergic (DA) fates. Loss of AST-1 result in terminal differentiation failure in all dopaminergic neurons and ectopic expression of AST-1 is able to drive DA fate in certain other cell types. *Cis*-regulatory informational analysis reveals the core *cis*-regulatory module (CRM) present in the promoter of each dopaminergic pathway genes, which is sufficient and necessary to drive expression in all DA neurons (Flames and Hobert, 2009). Subsequent studies suggest that *ast-1* is not the sole player. The *C. elegans* Distalless/Dlx ortholog *ceh-43* is also partially responsible for the induction and maintenance of the DA fates through binding to the homeodomain sites on DA promoters (Doitsidou *et al.*, 2008). AST-1 and CEH-43 act synergistically to control the fate of the CEPVS, although CEH-43 does not seem to be able to compensate for the loss of AST-1 in other DA neurons. Two Pbx factors, CEH-20 and CEH-40 also play a role in a partially redundant and subtype-specific manner through binding to the Pbx-type homeodomain-binding sites. Yeast transcription assay (Topalidou *et al.*, 2011) with AST-1, CEH-43 and CEH-20 suggests that neither of them alone is able to induce

efficient expression of β -galactosidase driven by the promoters of two DA pathway genes, while strong and synergistic induction was observed with coexpression of all of the three. This demonstrates that all three factors act in a cooperative fashion to control terminal differentiation of the DA neurons (Doitsidou *et al.*, 2013).

5.3. Terminal selector regulation in vertebrates

Although complicated regulatory regions as well as time and cost considerations has somehow hampered studies on terminal fate differentiation in vertebrates, evidence has suggested that the terminal selector concept may also be very well extended to vertebrates.

The mouse ETS transcription factor Etv-1 (ER81) is a homolog of the *C. elegans* AST-1, and is able to rescue *C. elegans ast-1* mutant phenotype in transgenic worms. In mice lacking Etv-1, dopamine neurons in the olfactory bulb of the brain are not properly differentiated. The number of tyrosine hydroxylase-positive cells is reduced, while other periglomerular interneuron subtypes and the DA progenitor cells in the lateral ganglionic eminence seem less affected or unaffected at all. In primary cell culture, ectopic expression of Etv-1 results in an increase in the number of cells expressing tyrosine hydroxylase-positive cells, which suggests that Etv-1 is not only necessary but also sufficient to induce DA fate *in vivo*. *cis*-regulatory information analysis further revealed that activation of AST-1 is through phylogenetically conserved motifs in the mouse TH locus and other mouse dopamine pathway genes (Flames and Hobert, 2009).

Similarly, the orphan nuclear receptor Nurr1 and the homeobox transcription

factor Pitx3 are the terminal selectors for midbrain dopaminergic neurons (Smidt and Burbach, 2009). Combined lentiviral transduction of Nurr1 and Pitx3 at the neural precursor stage can synergistically induce the expression of markers for midbrain dopaminergic neurons and to promote neuron maturation in the murine and human ES cell culture. Transplantation of human or mouse ES-derived cultures that have been transduced with Pitx3 and Nurr1 together can rescue the contralateral turning behavior defects in DA neuron injured mice (Martinat *et al.*, 2006). ChIP analysis has demonstrated that they directly control terminal differentiation via binding to the same promoter regions of the target genes such as *Dlk1*, *Ptpru* and *Klhl1* (Jacobs *et al.*, 2009a; Jacobs *et al.*, 2009b)

The paired-like Orthodentical (Otx) homeoprotein CEH-36 act as a terminal selector in the *C. elegans* AWC neuron pair (Kim *et al.*, 2010; Lanjuin and Sengupta, 2004). Its ortholog, mouse Cone-rod homeobox protein Crx has also been demonstrated to have similar roles in the mouse retina by microarray, *in situ* hybridization analysis and ChIP-seq assays (Corbo *et al.*, 2010; Hsiau *et al.*, 2007). Crx is expressed in rods and cones in the retina and is required for proper photoreceptor differentiation and survival. Acting directly on the *cis*-regulatory elements distributed around each locus, it regulates hundreds of its target genes in the photoreceptor, including retinol-binding protein, rhodopsin, G protein Gnat1, phosphodiesterase, vesicular glutamate transporter, kinesin and so on. Crx also directly regulates photoreceptor transcription factors, as well as its own expression (Furukawa *et al.*, 2002). Loss of Crx results in loss of expression of these genes although evidence suggests that there might be other compensatory mechanisms under certain circumstances (Nishida *et al.*, 2003).

The ETS domain factor Pet1 is the terminal selector for all serotonergic neurons in the mouse. Pet1 is expressed in most of the serotonergic neurons but not other nonserotonergic neurons or neighboring cells in the rat brain, and its expression is about half day earlier than the onset of 5-HT. It directly regulates genes such as 5-HT synthesis pathway components as well as serotonin receptors through conserved binding sites in the regulatory regions of these genes in the mouse and human (Hendricks et al., 1999). Due to inability to express genes required for 5-HT synthesis, uptake and storage machinery (Hendricks *et al.*, 2003), *Pet1* mutant mice have severe defects in serotonergic neuron differentiation, and therefore display heightened anxiety-like and aggressive behavior at adult stages. Engineered multimerized Pet1 binding sites from mouse and human 5-HT genes are able to drive reporter gene expression *in vivo*, and ChIP assays confirmed that Pet1 directly binds to promoters of its target genes. Aside from its role in initial 5-HT neuron generation, Pet1 is also required for the maturation of serotonergic neurons, and continuous Pet1 expression is essential for axon innervation of the somatosensory cortex, expression of firing properties as well as autoreceptors (Liu *et al.*, 2010), all suggesting that Pet1 is the *bona fide* terminal selector for vertebrate serotonergic neurons.

5.4. Maintenance of differentiated terminal state

Terminal selectors are usually turned on around the terminal division of neuroblasts, possibly activated by transient regulatory signals and factors within a very short amount of time. Aside from their role in early initiation phase, terminal selectors are also required throughout the life of a neuron in both invertebrates and vertebrates. For example, in *C. elegans*, postembryonic removal of the terminal selector can result in

loss of the differentiated neuronal fate. RNAi (against *che-1*) treated L1 animals display significant defects in ASE marker expression (Etchberger *et al.*, 2009), and temporal addition or removal of *ast-1* gene activity by heat shock or utilizing a temperature sensitive allele of *ast-1* correlates with the expression of DA terminal genes (Flames and Hobert, 2009). Conditional deletion of *Pet1* at E12.5 stage after 5-HT neuron generation also confirms that it is continuously required for serotonergic function, as the expression of its target genes became diminished in conditional KO mutants (Liu *et al.*, 2010). Direct transcriptional autoregulation seems to be the most effective and efficient way of ensuring sustained expression of terminal selectors in order to lock in the initial regulatory state. For terminal selectors, there may be binding sites for themselves in their own regulatory regions in order to maintain their own expression throughout the life of an animal (Etchberger *et al.*, 2007; Way and Chalfie, 1988; Wenick and Hobert, 2004), and mutating these sites does not seem to affect initiation but the maintenance of their expression (Bertrand and Hobert, 2009).

5.5. Parallel regulations

In *C. elegans*, loss of a neuronal terminal selector leads to the loss of various terminal differentiated features of a neuron. However, other parallel regulations must exist, because at least in several demonstrated cases, such as removal of CHE-1 and AST-1, the pan-sensory features of those neurons remain intact. Loss of *Pet1*, the terminal selector for serotonergic neurons in mice, also does not affect the overall neuronal identity (Hendricks *et al.*, 2003). The term “regulon” is employed to define a regulation unit, which consists of a terminal selector and all its downstream target genes (Hobert, 2011). A good example of a separate regulon involves genes that encode cilia

structures regulated through a common *cis*-regulatory motif (the X box) by the RFX family transcription factor DAF-19 in ciliated neurons (Swoboda *et al.*, 2000). Such regulation is completely independent of neuronal terminal selectors, as the pan-sensory identities are entirely unaffected in terminal selector mutants (Etchberger *et al.*, 2007). Another parallel regulation unit is the pan-neuronal identity regulon. In terminal selector mutants, pan-neuronal features are generally not affected (Altun-Gultekin *et al.*, 2001; Etchberger *et al.*, 2007; Flames and Hobert, 2009; Uchida *et al.*, 2003) (Figure 2). Regulons may overlap with each other. In *C. elegans*, MEC-3/UNC-86 controls not only the terminal touch neuron features but also the expression of pan-neuronal genes such as *snap-25/ric-4* (Hwang and Lee, 2003). The mouse Crx also regulates genes that control cilia structures (Swoboda *et al.*, 2000), suggesting that regulons in terminally differentiated mature neurons do not act in a mutually exclusive manner.

Part II: Neuronal Asymmetry

1. Biological asymmetry

Three different axes define the animal body plan: the clearly defined anterior-posterior axis, the dorsal-ventral axis, and the left-right axis. The A/P and D/V axes are unquestionably directionally asymmetric as they could be set by exogenous cues or gravity. Although the morphological body plans of most animals appear to be bilaterally symmetric along the L/R axis, both structural and functional L/R asymmetries have been observed from invertebrates to human (Hamada *et al.*, 2002; Ramsdell and Yost, 1998; Wood, 1997). For example, the human heart, liver, stomach, pancreas and spleen are located only on one side of the body. Even paired organs such as the lungs could show structural asymmetry consisting of different numbers of lobes.

The L/R asymmetry brought about the possibility of two alternative forms of the body plan, which are of opposite handedness and mirror images of each other. Handedness can be either directional or randomized. Although in animals such as mice, rats, cats and dogs, paw preference has been observed, there doesn't seem to be a group bias. (Fabre-Thorpe *et al.*, 1993) On the other hand, almost 90% of the human population are more skillful with their right hand than with the left (Corballis, 2003). How is handed asymmetry established and maintained? Evidence from multiple organism supports the theory that the polarity of the L/R axis is an early embryonic decision after A/P and D/V polarities have been established. It seems that handedness determination is an intrinsic process and is dependent on the establishment of the other axes. It is not difficult to imagine that a consistent mechanism is required to impose consistent differences between the two sides. It was proposed that the concentration of a

handed molecule can stay higher on one side than the other through a reaction-diffusion process, and that a tissue-specific response to the difference between the two sides can result in the development of differences on the left and right side (Brown and Wolpert, 1990). The mouse lateral plate mesoderm formation is one such example. Nodal was identified as an essential signal in mesoderm formation during gastrulation and is a determinant of the left–right body axis (Collignon *et al.*, 1996). Its activity is critical for the propagation of the left positional information from the node after L/R asymmetry establishment, and is upregulated on the left side, which then leads to the subsequent asymmetric gene expression and tissue-specific laterality decisions (Brennan *et al.*, 2002; Raya and Izpisua Belmonte, 2006; Shiratori and Hamada, 2006).

2. Neuronal asymmetry

The nervous system also displays levels of asymmetry that is fundamentally and evolutionarily important. Neuronal asymmetry are thought to play important roles in enhancing information processing as well as task and behavioral performances (Rogers, 2000). The most commonplace example is that the human brain exhibits hemispheric asymmetry at both morphological and functional level. For instance, the left and the right temporal lobes display a size difference (Geschwind and Levitsky, 1968; Glick, 1981; LeMay, 1982) (Figure 3A); and the two important regions related to speech and language, Broca’s area and Wernicke’s area, are located solely on the left side (Figure 3B). As a matter of fact, such laterality in the nervous system is adopted not only by

human, but also by other vertebrates such as rodents and fish (Glick, 1981; Miklosi *et al.*, 1997).

The zebrafish provides an excellent model to study neuronal laterality with its amenability to large-scale mutagenesis and direct visualization of fluorescent proteins. The vast number of defective mutant strains of zebrafish available allows extensive testing of neuronal asymmetry at the molecular, anatomical and behavioral level. The transforming-growth-factor β (TGF- β) family member Nodal, expressed on the left side of the diencephalon is used to form the laterality of the epithalamus during embryogenesis. The Nodal-related factor Cyclops (Cyc/Ndr2), Pitx2 and the Nodal antagonist Antivin/Lefty 1 (Lft1) are all transiently expressed on the left side of the bilateral pineal anlage (Liang *et al.*, 2000). Asymmetric pineal complex formation in the epithalamus then results in the acquisition of different features of the adjacent diencephalic nuclei on the left and the right side, which may also have impact on other regions of the brain. The laterality of the epithalamus in zebrafish is very well preserved among individuals (99% of larvae develop asymmetry and 95% are left-biased) (Halpern *et al.*, 2003). As a matter of fact, directional asymmetry is thought to be advantageous for the species. For example, the right eye use is associated with decision to bite in zebrafish (Miklosi and Andrew, 1999), and vibration-stimulated animals tend to bias on right-hand startle C-bends (Heuts, 1999). Behaviors such as schooling, feeding and escape responses have also been shown mediated by neuronal laterality.

Neuronal asymmetry is also observed in invertebrates such as the round worm (Hobert *et al.*, 2002). *C. elegans* is especially suitable as an experimental organism for dissecting the effects of mutations *in vivo*, and therefore providing an exceptional tool to

study the correlations among molecular, morphological and functional asymmetries in the nervous system, which is poorly understood and difficult to study in vertebrates. In the following sections, I will discuss our current understanding of neuronal asymmetry in the nematode *C. elegans*. I will specifically focus on the ASE gustatory neuron, one of the most extensively studied neuron pair that display directional asymmetry in the head of the worm.

2.1. *C. elegans* neuronal asymmetry

One big advantage of studying the *C. elegans* nervous system is that the lineage information of all neurons has been mapped out precisely and the entire neuronal network has been revealed under their electron microscopes (Sulston, 1983; Sulston *et al.*, 1983; White *et al.*, 1982; White *et al.*, 1976, 1986). The neurons are generated invariantly through defined patterns and display different levels of asymmetry with regard to cell body position and axon placement. The nervous system of a hermaphrodite *C. elegans* consists of 302 neurons. This a much smaller population compared to that of the *Drosophila* (100,000) and the mouse (75,000,000), but they consist of a third of the number of the somatic cells (about 1000) in the worm.

Around a third of the neurons in the worm display morphological asymmetry. Seventy-five VNC neurons located on or very close to the midline do not have analogues. The retrovesicular ganglion neurons AVG and SABD are also single neurons close to the mid line. Four neurons in the head, namely AVL, RIS, RIH, RID, are unilateral, which means they are only present on one side of the animal. Not only can cell position be asymmetrical, axon placement may also display asymmetry. Most axons including those extending from some of those bilaterally symmetric neurons are solely

located on one side of the ventral nerve cord (White *et al.*, 1986). Asymmetry also plays crucial roles in neural development. Out of the ninety-eight neurons present in symmetrical pairs in *C. elegans*, almost a third are generated in an asymmetric manner (Sulston *et al.*, 1983). For example, although the amphid chemosensory AWB neurons are morphologically and functionally symmetrical, they descend from asymmetric lineages. ABalpppppap and ABpraaappap (White *et al.*, 1986).

2.2. Asymmetrical differentiation of the ASE

The ASE gustatory neuron pair is morphologically symmetric in terms of cell location, axonal and dendritic morphology, and synaptic connectivity. Yet they display laterality at different levels. Lineage, functional and neuron sub-type specification program asymmetries of ASE are discussed below.

2.2.1. Lineage asymmetry

As early as the four-cell stage, the two precursors of ASEL and ASER are directed to adopt distinct fates by a Notch signal from the P2 cell sent along the A/P axis. This signal represses expression of two T box genes in ABp but not ABa (Good *et al.*, 2004), which leads to the exclusive and transient expression of TBX-37/38 in the eight ABa great-granddaughters but not in their descendants. This somehow instructs the descents of the ABa and ABp blastomeres to adopt ASEL and ASER fate distinctly. After the eight-cell stage, the blastomere identities have been determined and ASEL/R fate develop cell autonomously, as ASE cell descending from an isolated ABalp blastomere can still correctly acquire the left fate and vice versa (Poole and Hobert, 2006) (Figure 4). This early blastomere asymmetry is memorized until later

developmental stages when ASEL and ASER are born; and later embryonic signaling events do not change this identity (Poole and Hobert, 2006). This mark of asymmetry is used to bias the miRNA *lgy-6* controlled bistable regulatory feedback loop of gene expression program in the postmitotic cells to specify the ASEL and ASER distinct fates (Johnston *et al.*, 2005), which will be discussed in 2.2.3. of this chapter.

2.2.2. Functional asymmetry

In contrast with “anti-asymmetry” in the AWC neuron pairs, the ASE neurons display “directional asymmetry”, which means that the handedness of their asymmetric functions is fixed on one side rather than random biased on either side. The left cell (ASEL) is the main sodium sensor, while the right cell (ASER) is primarily responsible for potassium and chloride detection (Bargmann and Horvitz, 1991; Suzuki *et al.*, 2008; Wes and Bargmann, 2001). A subset of *gcy* genes that encode putative chemoreceptors displays asymmetrical expression pattern in the two ASEs. Out of the eleven receptor-type guanylyl cyclase-encoding genes expressed in ASE, nine are highly biased between ASRL and ASER (Ortiz *et al.*, 2006). This leads to the functional asymmetry of the two cells and allows the animals not only to detect, but also to discriminate between different sensory cues. Disruption of this asymmetry results in chemosensory discrimination defects. For example, *gcy-22* mutants display chemotaxis defects in nearly all salt detection ability by ASER (Ortiz *et al.*, 2006).

2.2.3. Sub-differentiation program asymmetry

Previously, by looking for aberrant expression of cell fate markers that are normally expressed specifically in the left or the right ASE, several graduate students

and postdocs in the lab have identified a collection of “*lsy*” (loss of asymmetry) genes required for ASE asymmetrical specification (Chang *et al.*, 2004; Chang *et al.*, 2003; Etchberger *et al.*, 2007; Johnston and Hobert, 2003; Johnston *et al.*, 2005; Johnston *et al.*, 2006; Johnston and Hobert, 2005; Sarin *et al.*, 2007). These genes have been categorized into six different classes. In class I mutants, ASEL fate is ectopically executed in ASER; in class II mutants, ASER expression profile is ectopically expressed in the left cell; in class III mutants, no ASE fates are specified in both cells due to mutations in the master regulator *che-1*, which binds to its molecular signature, the ASE motifs in almost all ASE-specific genes; in class IV mutants, either cell adopts a mixed fate of ASEL and ASER, meaning ASEL fate is ectopically expressed in ASER with ASER fate unaffected, or vice versa; in class V mutants, ASEL fate is lost, ectopically gained or exclusively expressed in ASER, whereas ASER fate is never ectopically gained in ASEL; and in class VI mutants, the asymmetry of ASE neurons are not affected, but ASE cell fate markers are expressed in cells other than ASE.

A “bistable feedback loop” model (Figure 5) downstream if CHE-1 has been proposed to control ASE laterality, with the miRNA *lsy-6* and the C2H2 Zinc-finger transcription factor *die-1* being the input and the output respectively (Hobert, 2006; Johnston *et al.*, 2005). In ASEL, *lsy-6* represses the transcription of the homeobox gene *cog-1* through interaction with its 3' UTR, while the ASEL-specific transcription factor DIE-1 activates the transcription of *lsy-6* (Didiano and Hobert, 2008; Johnston and Hobert, 2003). This leads to the activation of downstream ASEL terminal genes such as *lim-6* and *gcy-7*, and the repression of ASER specific genes such as *gcy-5*. In ASER, the *lsy-6* switch is off and the suppression of *cog-1* is removed, which leads to the negative

regulation of *die-1* in ASER. In this case, ASER terminal differentiation markers are expressed while ASEL specific genes are now repressed.

The C2H2 Zinc finger transcription factor CHE-1 is the master regulator/terminal selector for the ASE neurons. It is exclusively expressed in the two ASEs and autoregulates its own expression (See Chapter I for details). Loss of CHE-1 results in a complete loss of the ASE terminal fate, while the two cells are still physically present and maintain pan-neuronal features (Etchberger *et al.*, 2007). An unusual *che-1* allele was recently retrieved from an extensive screen for symmetrized ASE mutants (Etchberger *et al.*, 2009; Sarin *et al.*, 2007). In this mutant, the ASEL fate marker *lim-6::gfp* is lost and the ASER maker *gcy-5::gfp* is ectopically expressed in ASEL, which reminds us of the genes that act within the downstream feedback loop. Additional mutants that display similar phenotype all harbor mutations that cluster around the second zinc finger of CHE-1. The structural features of the second zinc finger are affected in such mutants, which results in reduced DNA binding affinity to multiple tested ASE motifs. Taken together, this suggests that aside from its role in bilateral fate specification, CHE-1 also takes part in asymmetric subtype specification of ASE.

The screen (Sarin *et al.*, 2007) also revealed another unusual allele *lsy-9*, which displays a complex phenotype (Sarin *et al.*, 2009). In *lsy-9* mutants, the left cell fate marker *lim-6::gfp* is unaffected, bilaterally expressed in both ASEL and ASER, expressed in neither ASEL nor ASER, or expressed exclusively in ASER, while the right cell fate marker (*gcy-5::gfp*) is either normal or lost. What makes it unusual is that a fraction of the *lsy-9* mutants express the left fate marker *lim-6::gfp* exclusively in the

ASER rather than the ASEL neuron. This reversal defect has not been observed in any previously known mutants. Sarin *et al.* mapped *lsy-9* to the *nhr-67* locus, which encodes a tailless-related orphan nuclear receptor. A fosmid-based reporter suggests that *nhr-67* is bilaterally symmetric. It is turned on two divisions earlier in the grandmother of ASEL and ASER, and stays on until around the first larval stage. A fosmid based reporter for *che-1* is turned off by *nhr-67* in a substantial fraction of animals, while *nhr-67* expression is not affected in *che-1* mutants. This suggests that NHR-67 positively regulates CHE-1 to induce ASE fate. Epistasis analysis suggests that *nhr-67* may play another role downstream of the loop input *lsy-6* and upstream of the loop output *die-1*, activating transcription of the Nkx6 homeodomain transcription factor *cog-1* by directly binding to its promoter. This binding requires the NR2W motif for tailless-type orphan nuclear receptors (DeMeo *et al.*, 2008; Yu *et al.*, 1994). Therefore, *nhr-67* is not solely required for CHE-1 to induce ASE fate, but also takes part in the downstream bistable feedback loop for biased ASE fate determination.

As mentioned in chapter 2.2.1., the cell fates of ASEL and ASER might have already been determined very early on by the presence or absence of T-box transcription factors. This “identity mark” is somehow set but remembered until much later to take effect. How does this relate to the handedness of the bistable feedback programs in the two cells? Genetic epistasis analysis suggests that the miRNA *lsy-6* is upstream of *die-1* and *cog-1* (Johnston *et al.*, 2005). By using a more recently developed technology that generates reporters (Tursun *et al.*, 2009) in a much larger genomic context (20-40kb), it was shown that *lsy-6* expression comes on at the end of gastrulation only in the ASEL mother. *lsy-6* expression in ASEL is continuously on and maintained throughout the life

of the animal, while the asymmetric *die-1* and *cog-1* expression is initiated later at around 3-fold stage in the postmitotic ASEs (Cochella and Hobert, 2012). This implies that *lsy-6* may serve as the earliest switch as an entry point for asymmetry establishment. In fact, TBX-37/38 expression in the ABa lineage (Figure 5) primes the *lsy-6* locus through a downstream primer element very early and establishes the open chromatin conformation in the ASEL precursor. This allows a later boost of expression of LSY-6 through the C2H2 zinc finger transcription factor CHE-1, which binds to the upstream booster element. On the other hand, compacted *lsy-6* locus in ASER does not respond to the presence of CHE-1 due to absence of priming from TBX-37/38 at earlier stages. Therefore, the very early “asymmetry mark” imposed by the first Notch signal at the four cell stage is indeed “memorized” and linked to later asymmetric fate determination of ASE, through the chromatin based initiation of the miRNA *lsy-6*. A recent genome wide RNAi screen in an attempt to find more ASE mutants utilizing the reverse genetics approach identified three more genes, *ash-2*, *dpy-30* and *rbbp-5* (Poole *et al.*, 2011) that are components of the histone methyltransferase complex (COMPASS). Reporter and genetic analysis indicates that COMPASS may act upstream of the feedback loop input, the miRNA *lsy-6*, during early during embryogenesis to ensure ASE laterality, and that the COMPASS complex may possibly form a lineage specific “chromatin mark” to link the regulation between TBX-37/38 and *lsy-6*.

Part III: Summary of Thesis

The purpose of this thesis is to use the model organism *C. elegans* to study how terminally differentiated neuronal fate are achieved throughout the life of an animal, and to investigate how asymmetric neuronal sub-types are controlled and maintained by looking for additional components of the ASE specification program.

Chapter 2 takes the serotonergic neuron NSM and the cholinergic neuron AIA as a test paradigm to study terminal differentiation of the nervous system in *C. elegans*. Aside from the cholinergic interneuron AIY, TTX-3 also controls the terminal differentiation program of two additional, distinct neuronal cell types, the cholinergic AIA interneurons and the serotonergic NSM neurons. This is accomplished through collaborating with different regulatory factors in different cell types, such as UNC-86 in NSM and a yet unknown factor X in AIA. UNC-86 in turn collaborates with the ARID-type transcription factor CFI-1 to control the fate of the IL2 cholinergic inner labial neuron and the URA motor neuron. Therefore, it is the different combinations of transcription factors in distinct cell types that define the terminal fate of a cell. Aside from its role in the cholinergic interneurons AIY and AIA, the serotonergic motor neuron NSM, *ttx-3* is required for the specification of the glutamatergic amphid sensory neuron ASK. A set of terminal markers that are expressed in ASK are disrupted in *ttx-3 mutants*, suggesting that TTX-3 may act as a terminal selector for ASK that utilizes another neurotransmitter system. (See appendix 1 for ASK specification.)

Chapter 3 is a preliminary study on the fate determination based on my findings on the cholinergic ring motoneuron RMDD/RMDV. Five mutants defective in RMDD/V

were discovered by chance from a screen that looked for mutants defective in AIA specification using the marker *otIs317 (mgl-1::mcherry)*. Genetic analysis and whole genome sequencing revealed three relevant genes: the vertebrate neuroD homolog *cnd-1*, the Beta3 and Olig family related gene *hlh-16*, and the Q50 class paired-like homeobox gene *unc-42*. UNC-42 is likely a terminal selector for the RMDs as it has been shown required for *glr-1* expression in the RMDs (Baran et al., 1999). *cnd-1* and *hlh-16* are likely proneural regulators that act in earlier steps of cell fate specification.

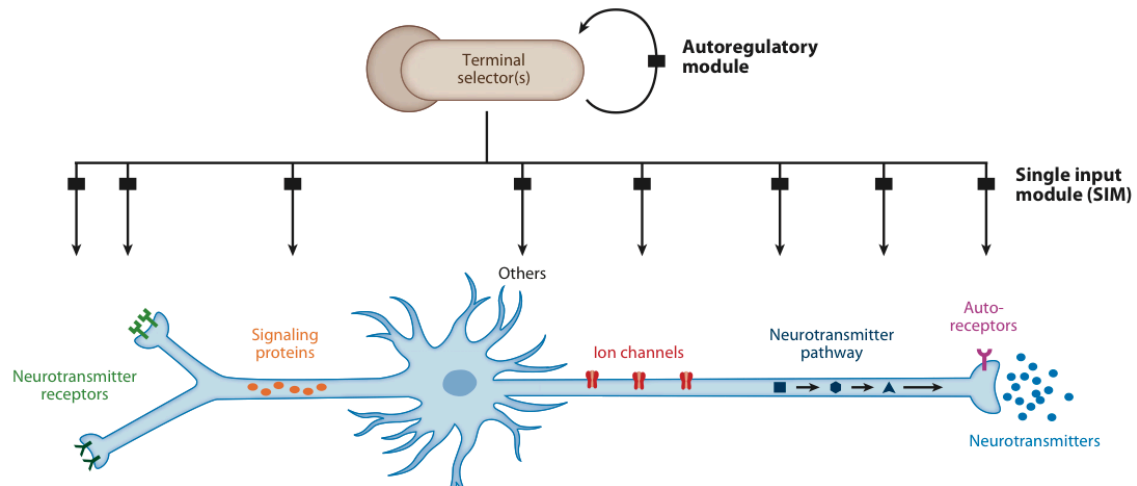
Chapter 4 describes the cloning and characterization of two phenotypically similar mutant *C. elegans* strains that are defective in asymmetric gene expression pattern of the ASE gustatory neurons with the right cell (ASER) fate depressed in the left neuron (ASEL). Classic mapping revealed that one of the mutants harbors a mutation in the LIM homeobox gene *lim-6*, while whole genome sequencing identified the other mutant to be an allele of a novel ASE fate determinant *lsy-27*, which encodes a member of a fast-evolving family of C2H2 zinc finger transcription factors. LSY-27 is broadly and exclusively expressed in the embryo. Temperature shift experiments suggest that it functions during the initiation, but not the maintenance phase of ASE laterality control, to assist the initiation of *lim-6* expression.

Chapter 5 describes the MYST-type histone acetyltransferase complex formed by the MYST-type histone acetyltransferase LSY-12, the ING-family PHD domain protein LSY-13 and the PHD/bromodomain protein LIN-49. This complex is required to not only induce, but also maintain lateralized gene expression in the ASE gustatory neurons. The defects due to mutations in the components of the complex are similar to

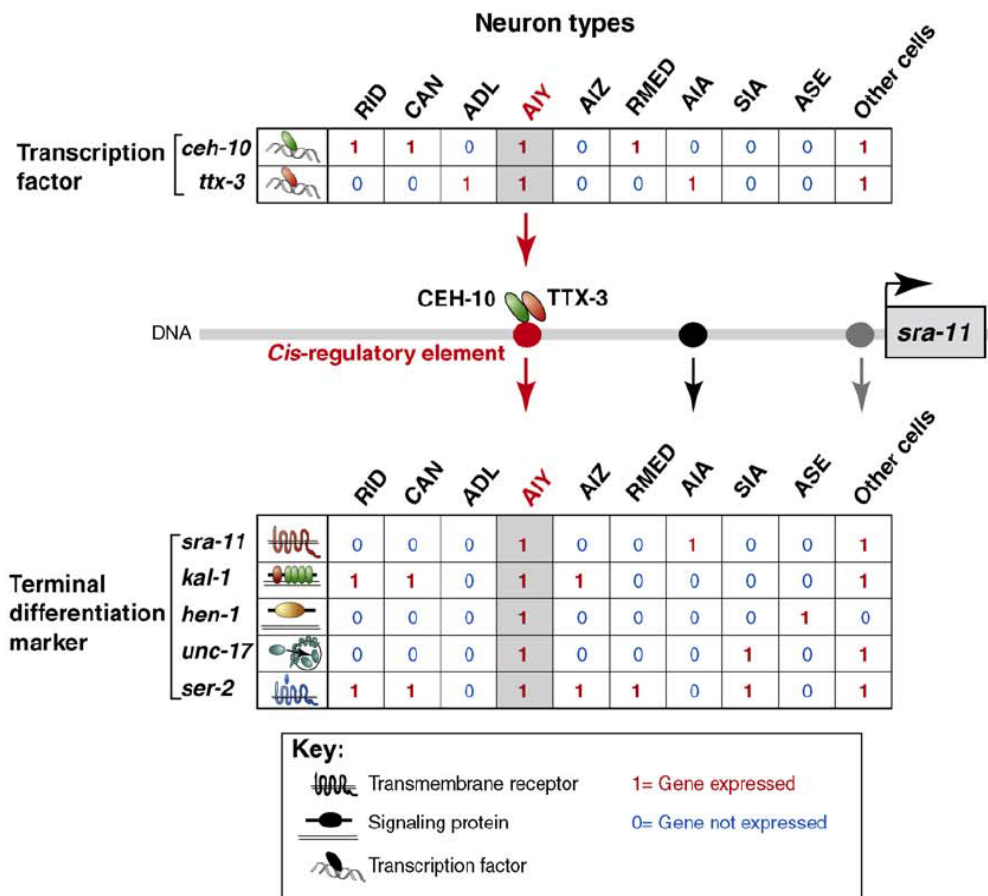
mutations in another zinc finger transcription factor DIE-1, which likely acts as a candidate recruiter for the MYST complex to the transcription activation site.

Figure 1: Terminal selectors directly control terminally differentiated genes

(A)



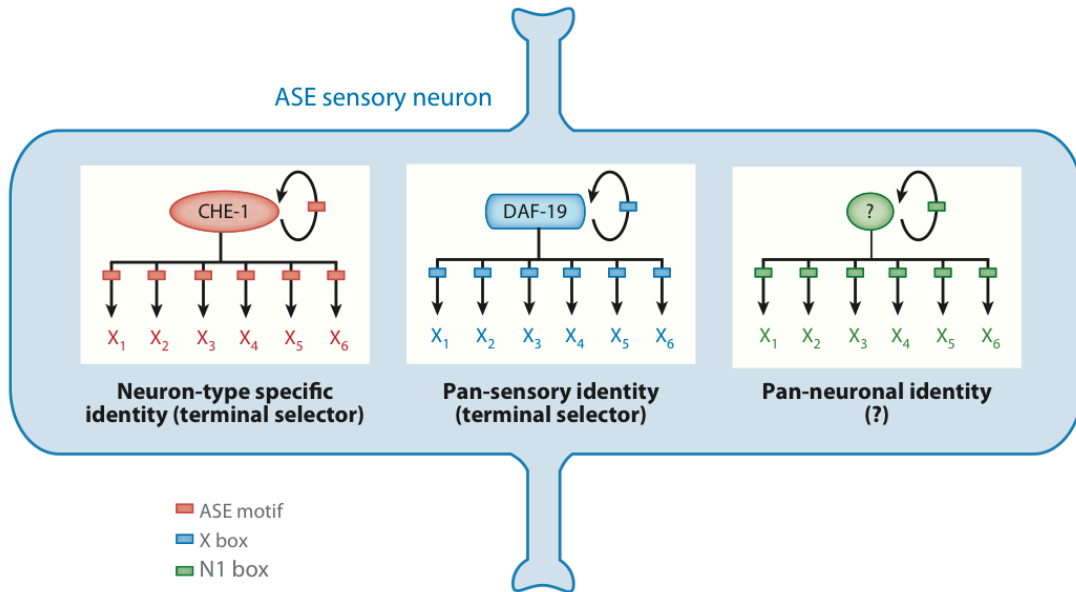
(B)



(A) Terminal selectors directly act on cis-regulatory motifs (black box) to control terminally differentiated genes that encode the terminal features of a neuron. These genes encode neurotransmitters synthesis enzymes, neurotransmitter receptors, ion channels, signaling proteins, cytoskeleton proteins, adhesion molecules and so on. (Hobert, 2011)

(B) None of the terminal differentiated markers are exclusively expressed in one particular neuron. They are very often expressed in many other cells. It is the combination of the terminal features that determines a neuron's identify. Similarly, a transcription factor is rarely only expressed in one particular cell type. It is the combinatorial code of transcription factors that determines the terminal differentiation program of a particular cell. (Hobert, 2011; Hobert *et al.*, 2010)

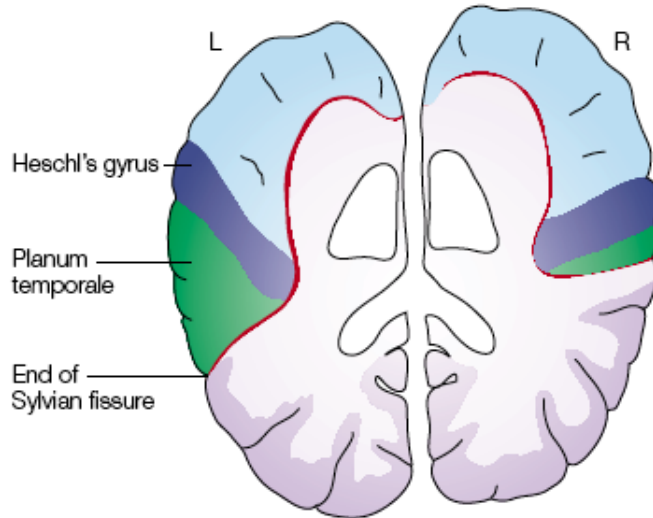
Figure 2: Example of parallel regulons of the ASE gustatory neuron (Hobert, 2011).



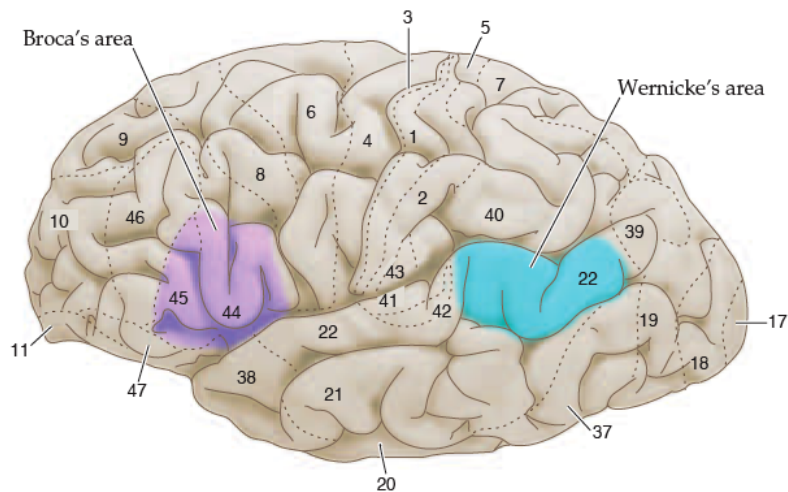
Multiple regulons act in parallel to define the terminal molecular features of the ASE neuron. The zinc-finger transcription factor CHE-1 controls the neuron-type specific identities via the ASE motif as a terminal selector, while the RFX-type transcription factor DAF-19 controls all ciliated neuronal structures via the X box. It has been proposed that there is a third regulon that consists of an unknown factor plus its target genes that encode pan-neuronal identities, with The N1 box as a shared motif for regulation. These regulons in combination determine the fate of a mature neuron.

Figure 3: The human brain is asymmetric at both morphological and functional level.

(A)



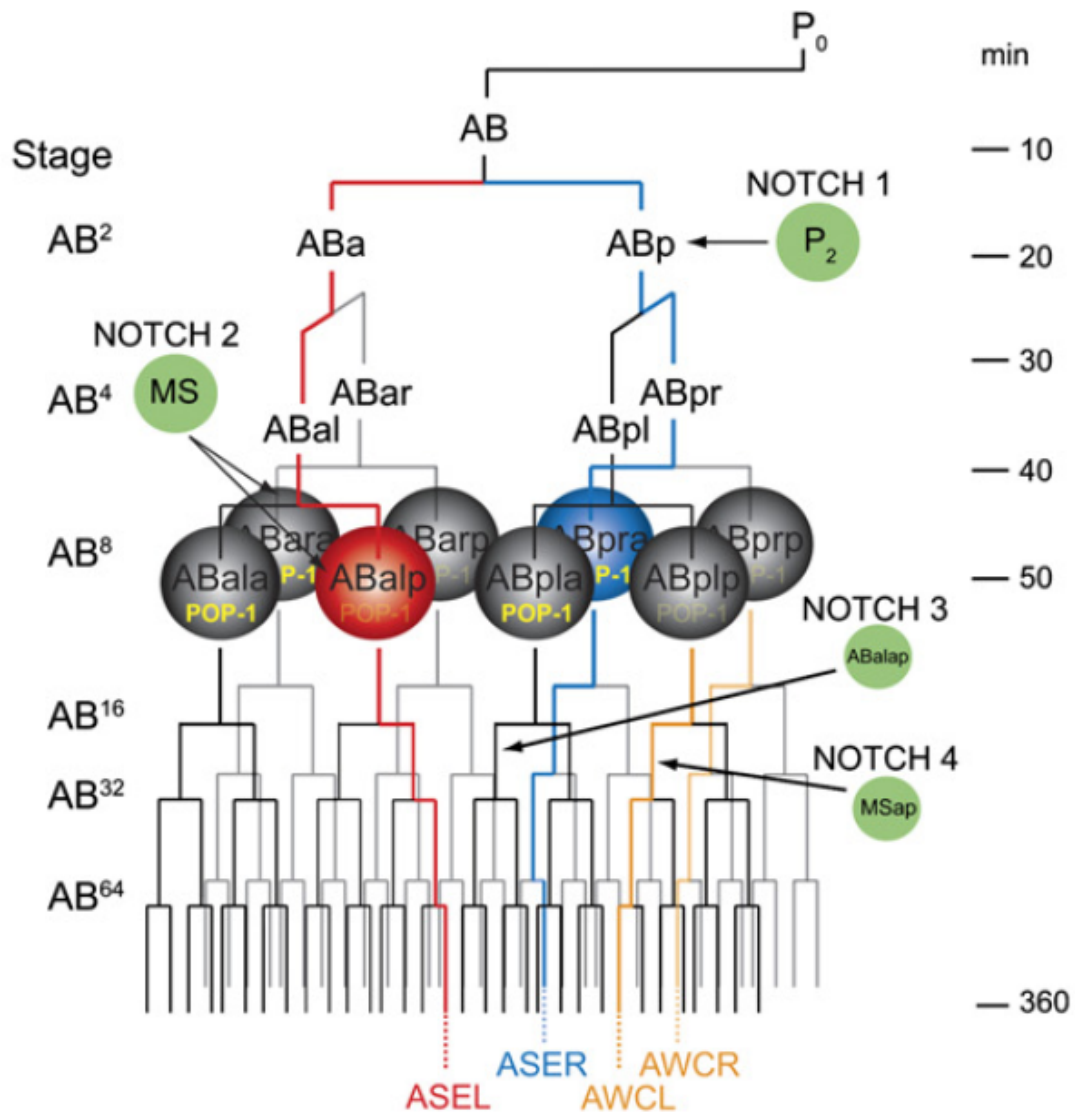
(B)



(A) A human brain cross-section that reveals the superior temporal surface. The Planum temporal lobes and Sylvian fissures are significantly different in size. Image from (Hobert *et al.*, 2002).

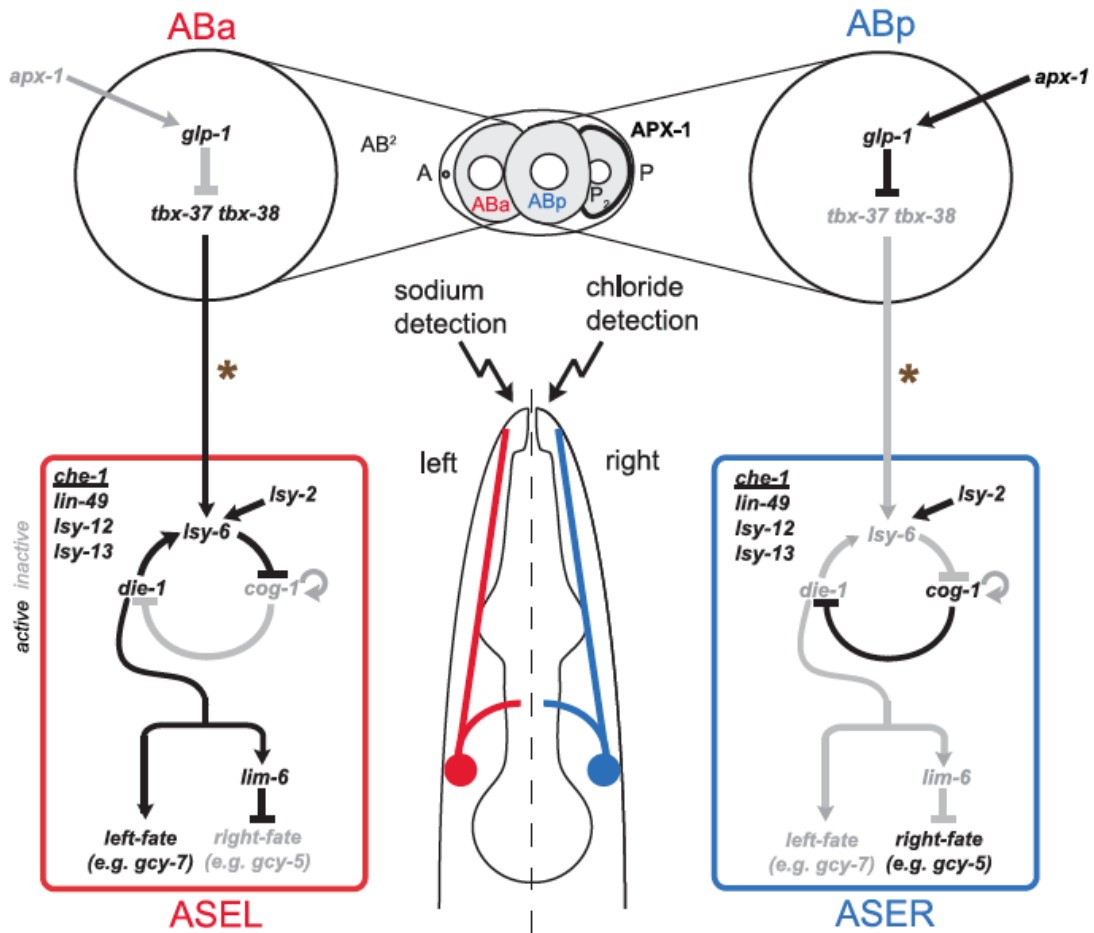
(B) Two important regions related to human higher brain function, speech and language, Broca's area and Wernicke's area, are located solely on the left side. Images from (Purves, 2008).

Figure 4: Asymmetric lineage formation of the ASE neuron pair.



The two precursors of ASEL and ASER are directed by the first NOTCH signal to adopt distinct cell fates, ABalp and ABpra. This early-established asymmetry mark is remembered until the two cells are born several divisions later. Image from (Poole and Hobert, 2006).

Figure 5: The asymmetric ASEL/R differentiation program.



The ASE asymmetry started from the four-cell stage when directed by the first NOTCH signal to adopt distinct fates. Presence of TBX-37/38 in ASEL leads to the decompacted conformation of the *lsy-6* locus and biases *lsy-6* expression in the left cell. A bistable feedback loop is utilized to confer ASE laterality with the miRNA *lsy-6* being the input and the C2H2 zinc finger transcription factor DIE-1 being the output. *Lsy-6* and DIE-1 are expressed in the left cell to activate ASEL-specific terminal genes and to repress ASER-specific genes through the LIM homeodomain protein LIM-6. Expression of COG-1 and the absence of *lsy-6* and DIE-1 result in the right cell fate in ASER. Worms defective in the handedness of this feedback loop display functional defects in sensing soluble salts. CHE-1 is the master regulator that binds to ASE motifs present in the promoters of all ASE specific genes. Image from (Ortiz *et al.*, 2009).

Table1: Examples of invertebrate and vertebrate regulons in the nervous system. From (Hobert, 2011).

Terminal selector	Type of transcription factor	Terminal neuron identity controlled by terminal selector	Selected examples of direct target genes
CHE-1	C2H2 zinc finger	<i>C. elegans</i> gustatory neuron class ASE	Putative chemoreceptors of the GCY family, neuropeptides of the FLP family, cyclic-nucleotide gated ion channels, itself (autoregulation), many others
TTX-3/CEH-10	LIM/Prd homeodomains	<i>C. elegans</i> cholinergic interneuron class AIY	Choline acetyltransferase, choline transporter, neuropeptide receptors, neuropeptides, Ig domain proteins, ion channels, itself (autoregulation), many others
UNC-86/MEC-3	POU/LIM homeodomains	<i>C. elegans</i> glutamatergic touch sensory neurons	Touch receptor channel subunits, specialized tubulin <i>mec-7</i> , extracellular matrix proteins, vesicular glutamate transporter, itself (autoregulation), many others
UNC-30	Pitx-type homeodomain	<i>C. elegans</i> GABAergic motor neurons	Glutamic acid decarboxylase, vesicular GABA transporter, GABA reuptake transporter, acetylcholine receptor, Ig domain protein, potassium channel, neuropeptides of the FLP family, itself (autoregulation), many others
AST-1/CEH-43	ETS/homeodomain	<i>C. elegans</i> dopaminergic neuron	Tyrosine hydroxylase, aromatic amino acid decarboxylase, vesicular dopamine transporter, dopamine reuptake transporter, dopamine autoreceptor, ion channels
UNC-3	EBF-type	<i>C. elegans</i> A/B-type cholinergic motor neurons	Choline acetyltransferase, choline transporter, neuropeptide receptors, neuropeptides, Ig domain proteins, signaling proteins, many others
Pet-1	ETS domain	Mouse serotonergic neurons	Tryptophan hydroxylase, 5-HT _{1a} receptor, aromatic amino acid decarboxylase, serotonin transporter, itself (autoregulation), many others
Crx	Homeodomain	Mouse retinal photoreceptors	Retinol-binding protein, rhodopsin, G protein Gnat1, phosphodiesterase, vesicular glutamate transporter, kinesin, itself (autoregulation), many others
Nurr1/Pitx-3	C4 zinc finger/homeodomain	Mouse midbrain dopaminergic neurons	Tyrosine hydroxylase, aromatic amino acid decarboxylase, vesicular dopamine transporter, dopamine reuptake transporter, dopamine autoreceptor D2R, receptor tyrosine kinase Ret, many others

References:

Altun-Gultekin, Z., Andachi, Y., Tsalik, E.L., Pilgrim, D., Kohara, Y., and Hobert, O. (2001). A regulatory cascade of three homeobox genes, *ceh-10*, *ttx-3* and *ceh-23*, controls cell fate specification of a defined interneuron class in *C. elegans*. *Development* *128*, 1951-1969.

Artavanis-Tsakonas, S., Rand, M.D., and Lake, R.J. (1999). Notch signaling: cell fate control and signal integration in development. *Science* *284*, 770-776.

Baran, R., Aronoff, R., and Garriga, G. (1999). The *C. elegans* homeodomain gene *unc-42* regulates chemosensory and glutamate receptor expression. *Development* *126*, 2241-2251.

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

Bargmann, C.I., Hartweg, E., and Horvitz, H.R. (1993). Odorant-selective genes and neurons mediate olfaction in *C. elegans*. *Cell* *74*, 515-527.

Bargmann, C.I., and Horvitz, H.R. (1991). Chemosensory neurons with overlapping functions direct chemotaxis to multiple chemicals in *C. elegans*. *Neuron* *7*, 729-742.

Benos, P.V., Lapedes, A.S., and Stormo, G.D. (2002). Probabilistic code for DNA recognition by proteins of the EGR family. *Journal of molecular biology* *323*, 701-727.

Bertrand, N., Castro, D.S., and Guillemot, F. (2002). Proneural genes and the specification of neural cell types. *Nat Rev Neurosci* *3*, 517-530.

Bertrand, V., and Hobert, O. (2009). Linking asymmetric cell division to the terminal differentiation program of postmitotic neurons in *C. elegans*. *Dev Cell* *16*, 563-575.

Bota, M., Dong, H.W., and Swanson, L.W. (2003). From gene networks to brain networks. *Nature neuroscience* *6*, 795-799.

Boulin, T., Etchberger, J.F., and Hobert, O. (2006). Reporter gene fusions. *WormBook*, 1-23.

Brennan, J., Norris, D.P., and Robertson, E.J. (2002). Nodal activity in the node governs left-right asymmetry. *Genes & development* *16*, 2339-2344.

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

Brown, N.A., and Wolpert, L. (1990). The development of handedness in left/right asymmetry. *Development* *109*, 1-9.

Campuzano, S., and Modolell, J. (1992). Patterning of the *Drosophila* nervous system: the achaete-scute gene complex. *Trends Genet* *8*, 202-208.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W., and Prasher, D.C. (1994). Green fluorescent protein as a marker for gene expression. *Science* *263*, 802-805.

Chang, S., Johnston, R.J., Jr., Frokjaer-Jensen, C., Lockery, S., and Hobert, O. (2004). MicroRNAs act sequentially and asymmetrically to control chemosensory laterality in the nematode. *Nature* *430*, 785-789.

Chang, S., Johnston, R.J., Jr., and Hobert, O. (2003). A transcriptional regulatory cascade that controls left/right asymmetry in chemosensory neurons of *C. elegans*. *Genes Dev* *17*, 2123-2137.

Cinar, H., Keles, S., and Jin, Y. (2005). Expression profiling of GABAergic motor neurons in *Caenorhabditis elegans*. *Curr Biol* *15*, 340-346.

Cochella, L., and Hobert, O. (2012). Embryonic priming of a miRNA locus predetermines postmitotic neuronal left/right asymmetry in *C. elegans*. *Cell* *151*, 1229-1242.

Collignon, J., Varlet, I., and Robertson, E.J. (1996). Relationship between asymmetric nodal expression and the direction of embryonic turning. *Nature* *381*, 155-158.

Corballis, M.C. (2003). From mouth to hand: gesture, speech, and the evolution of right-handedness. *Behav Brain Sci* *26*, 199-208; discussion 208-160.

Corbo, J.C., Lawrence, K.A., Karlstetter, M., Myers, C.A., Abdelaziz, M., Dirkes, W., Weigelt, K., Seifert, M., Benes, V., Fritsche, L.G., *et al.* (2010). CRX ChIP-seq reveals

the cis-regulatory architecture of mouse photoreceptors. *Genome research* 20, 1512-1525.

Culi, J., Martin-Blanco, E., and Modolell, J. (2001). The EGF receptor and N signalling pathways act antagonistically in *Drosophila* mesothorax bristle patterning. *Development* 128, 299-308.

Culi, J., and Modolell, J. (1998). Proneural gene self-stimulation in neural precursors: an essential mechanism for sense organ development that is regulated by Notch signaling. *Genes Dev* 12, 2036-2047.

DeFelipe, J., Alonso-Nanclares, L., and Arellano, J.I. (2002). Microstructure of the neocortex: comparative aspects. *Journal of neurocytology* 31, 299-316.

DeMeo, S.D., Lombel, R.M., Cronin, M., Smith, E.L., Snowflack, D.R., Reinert, K., Clever, S., and Wightman, B. (2008). Specificity of DNA-binding by the FAX-1 and NHR-67 nuclear receptors of *Caenorhabditis elegans* is partially mediated via a subclass-specific P-box residue. *BMC molecular biology* 9, 2.

Didiano, D., and Hobert, O. (2008). Molecular architecture of a miRNA-regulated 3' UTR. *RNA* 14, 1297-1317.

Doitsidou, M., Flames, N., Lee, A.C., Boyanov, A., and Hobert, O. (2008). Automated screening for mutants affecting dopaminergic-neuron specification in *C. elegans*. *Nat Methods* 5, 869-872.

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

Duggan, A., Ma, C., and Chalfie, M. (1998). Regulation of touch receptor differentiation by the *Caenorhabditis elegans* *mec-3* and *unc-86* genes. *Development* 125, 4107-4119.

Dunbar, R.I., and Shultz, S. (2007). Evolution in the social brain. *Science* 317, 1344-1347.

Dusenbery, D.B., Sheridan, R.E., and Russell, R.L. (1975). Chemotaxis-defective mutants of the nematode *Caenorhabditis elegans*. *Genetics* 80, 297-309.

Eastman, C., Horvitz, H.R., and Jin, Y. (1999). Coordinated transcriptional regulation of the *unc-25* glutamic acid decarboxylase and the *unc-47* GABA vesicular transporter by the *Caenorhabditis elegans* UNC-30 homeodomain protein. *The Journal of neuroscience : the official journal of the Society for Neuroscience* *19*, 6225-6234.

Edlund, T., and Jessell, T.M. (1999). Progression from extrinsic to intrinsic signaling in cell fate specification: a view from the nervous system. *Cell* *96*, 211-224.

Etchberger, J.F., Flowers, E.B., Poole, R.J., Bashllari, E., and Hobert, O. (2009). Cis-regulatory mechanisms of left/right asymmetric neuron-subtype specification in *C. elegans*. *Development* *136*, 147-160.

Etchberger, J.F., Lorch, A., Sleumer, M.C., Zapf, R., Jones, S.J., Marra, M.A., Holt, R.A., Moerman, D.G., and Hobert, O. (2007). The molecular signature and cis-regulatory architecture of a *C. elegans* gustatory neuron. *Genes Dev* *21*, 1653-1674.

Fabre-Thorpe, M., Fagot, J., Lorincz, E., Levesque, F., and Vauclair, J. (1993). Laterality in cats: paw preference and performance in a visuomotor activity. *Cortex* *29*, 15-24.

Finney, M., and Ruvkun, G. (1990). The *unc-86* gene product couples cell lineage and cell identity in *C. elegans*. *Cell* *63*, 895-905.

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

Fowles, L.F., Martin, M.L., Nelsen, L., Stacey, K.J., Redd, D., Clark, Y.M., Nagamine, Y., McMahon, M., Hume, D.A., and Ostrowski, M.C. (1998). Persistent activation of mitogen-activated protein kinases p42 and p44 and *ets-2* phosphorylation in response to colony-stimulating factor 1/c-fms signaling. *Molecular and cellular biology* *18*, 5148-5156.

Frank, C.A., Baum, P.D., and Garriga, G. (2003). HLH-14 is a *C. elegans* achaete-scute protein that promotes neurogenesis through asymmetric cell division. *Development* *130*, 6507-6518.

Furukawa, A., Koike, C., Lippincott, P., Cepko, C.L., and Furukawa, T. (2002). The mouse *Crx* 5'-upstream transgene sequence directs cell-specific and developmentally regulated expression in retinal photoreceptor cells. *The Journal of neuroscience : the official journal of the Society for Neuroscience* *22*, 1640-1647.

Garcia-Bellido, A. (1979). Genetic Analysis of the Achaete-Scute System of *DROSOPHILA MELANOGASTER*. *Genetics* 91, 491-520.

Geschwind, N., and Levitsky, W. (1968). Human brain: left-right asymmetries in temporal speech region. *Science* 161, 186-187.

Glick, S.a.R.D. (1981). Lateralization of function in the rat brain: Basic mechanisms may be operative in humans *Trends Neurosci*, 196-199.

Good, K., Ciosk, R., Nance, J., Neves, A., Hill, R.J., and Priess, J.R. (2004). The T-box transcription factors TBX-37 and TBX-38 link GLP-1/Notch signaling to mesoderm induction in *C. elegans* embryos. *Development* 131, 1967-1978.

Hallam, S., Singer, E., Waring, D., and Jin, Y. (2000). The *C. elegans* NeuroD homolog *cnd-1* functions in multiple aspects of motor neuron fate specification. *Development* 127, 4239-4252.

Halpern, M.E., Liang, J.O., and Gamse, J.T. (2003). Leaning to the left: laterality in the zebrafish forebrain. *Trends Neurosci* 26, 308-313.

Hamada, H., Meno, C., Watanabe, D., and Saijoh, Y. (2002). Establishment of vertebrate left-right asymmetry. *Nat Rev Genet* 3, 103-113.

Hendricks, T., Francis, N., Fyodorov, D., and Deneris, E.S. (1999). The ETS domain factor *Pet-1* is an early and precise marker of central serotonin neurons and interacts with a conserved element in serotonergic genes. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 19, 10348-10356.

Hendricks, T.J., Fyodorov, D.V., Wegman, L.J., Lelutiu, N.B., Pehek, E.A., Yamamoto, B., Silver, J., Weeber, E.J., Sweatt, J.D., and Deneris, E.S. (2003). *Pet-1* ETS gene plays a critical role in 5-HT neuron development and is required for normal anxiety-like and aggressive behavior. *Neuron* 37, 233-247.

Herculano-Houzel, S. (2009). The human brain in numbers: a linearly scaled-up primate brain. *Frontiers in human neuroscience* 3, 31.

Herman, R.K. (1987). Mosaic analysis of two genes that affect nervous system structure in *Caenorhabditis elegans*. *Genetics* 116, 377-388.

- Heuts, B.A. (1999). Lateralization of trunk muscle volume, and lateralization of swimming turns of fish responding to external stimuli. *Behavioural processes* 47, 113-124.
- Hobert, O. (2002). PCR fusion-based approach to create reporter gene constructs for expression analysis in transgenic *C. elegans*. *BioTechniques* 32, 728-730.
- Hobert, O. (2005). Specification of the nervous system. *WormBook*, 1-19.
- Hobert, O. (2006). Architecture of a microRNA-controlled gene regulatory network that diversifies neuronal cell fates. *Cold Spring Harb Symp Quant Biol* 71, 181-188.
- Hobert, O. (2011). Regulation of terminal differentiation programs in the nervous system. *Annu Rev Cell Dev Biol* 27, 681-696.
- Hobert, O., Carrera, I., and Stefanakis, N. (2010). The molecular and gene regulatory signature of a neuron. *Trends Neurosci* 33, 435-445.
- Hobert, O., Johnston, R.J., Jr., and Chang, S. (2002). Left-right asymmetry in the nervous system: the *Caenorhabditis elegans* model. *Nat Rev Neurosci* 3, 629-640.
- Hsiau, T.H., Diaconu, C., Myers, C.A., Lee, J., Cepko, C.L., and Corbo, J.C. (2007). The cis-regulatory logic of the mammalian photoreceptor transcriptional network. *PLoS One* 2, e643.
- Hwang, S.B., and Lee, J. (2003). Neuron cell type-specific SNAP-25 expression driven by multiple regulatory elements in the nematode *Caenorhabditis elegans*. *Journal of molecular biology* 333, 237-247.
- Iwasaki, K., Staunton, J., Saifee, O., Nonet, M., and Thomas, J.H. (1997). *aex-3* encodes a novel regulator of presynaptic activity in *C. elegans*. *Neuron* 18, 613-622.
- Jacobs, D., Beitel, G.J., Clark, S.G., Horvitz, H.R., and Kornfeld, K. (1998). Gain-of-function mutations in the *Caenorhabditis elegans* *lin-1* ETS gene identify a C-terminal regulatory domain phosphorylated by ERK MAP kinase. *Genetics* 149, 1809-1822.

Jacobs, F.M., van der Linden, A.J., Wang, Y., von Oerthel, L., Sul, H.S., Burbach, J.P., and Smidt, M.P. (2009a). Identification of *Dlk1*, *Ptpru* and *Klhl1* as novel *Nurr1* target genes in meso-diencephalic dopamine neurons. *Development* *136*, 2363-2373.

Jacobs, F.M., van Erp, S., van der Linden, A.J., von Oerthel, L., Burbach, J.P., and Smidt, M.P. (2009b). *Pitx3* potentiates *Nurr1* in dopamine neuron terminal differentiation through release of SMRT-mediated repression. *Development* *136*, 531-540.

Jarman, A.P., Brand, M., Jan, L.Y., and Jan, Y.N. (1993a). The regulation and function of the helix-loop-helix gene, *asense*, in *Drosophila* neural precursors. *Development* *119*, 19-29.

Jarman, A.P., Grau, Y., Jan, L.Y., and Jan, Y.N. (1993b). *atonal* is a proneural gene that directs chordotonal organ formation in the *Drosophila* peripheral nervous system. *Cell* *73*, 1307-1321.

Jimenez, F., and Campos-Ortega, J.A. (1990). Defective neuroblast commitment in mutants of the *achaete-scute* complex and adjacent genes of *D. melanogaster*. *Neuron* *5*, 81-89.

Johnston, R.J., and Hobert, O. (2003). A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans*. *Nature* *426*, 845-849.

Johnston, R.J., Jr., Chang, S., Etchberger, J.F., Ortiz, C.O., and Hobert, O. (2005). MicroRNAs acting in a double-negative feedback loop to control a neuronal cell fate decision. *Proc Natl Acad Sci U S A* *102*, 12449-12454.

Johnston, R.J., Jr., Copeland, J.W., Fasnacht, M., Etchberger, J.F., Liu, J., Honig, B., and Hobert, O. (2006). An unusual Zn-finger/FH2 domain protein controls a left/right asymmetric neuronal fate decision in *C. elegans*. *Development* *133*, 3317-3328.

Johnston, R.J., Jr., and Hobert, O. (2005). A novel *C. elegans* zinc finger transcription factor, *lsy-2*, required for the cell type-specific expression of the *lsy-6* microRNA. *Development* *132*, 5451-5460.

Kim, K., Kim, R., and Sengupta, P. (2010). The HMX/NKX homeodomain protein *MLS-2* specifies the identity of the AWC sensory neuron type via regulation of the *ceh-36 Otx* gene in *C. elegans*. *Development* *137*, 963-974.

Koch, C., and Laurent, G. (1999). Complexity and the nervous system. *Science* 284, 96-98.

Kohler, J., Schafer-Preuss, S., and Buttgerit, D. (1996). Related enhancers in the intron of the beta1 tubulin gene of *Drosophila melanogaster* are essential for maternal and CNS-specific expression during embryogenesis. *Nucleic acids research* 24, 2543-2550.

Kratsios, P., Stolfi, A., Levine, M., and Hobert, O. (2012). Coordinated regulation of cholinergic motor neuron traits through a conserved terminal selector gene. *Nature neuroscience* 15, 205-214.

Krause, M., Park, M., Zhang, J.M., Yuan, J., Harfe, B., Xu, S.Q., Greenwald, I., Cole, M., Paterson, B., and Fire, A. (1997). A *C. elegans* E/Daughterless bHLH protein marks neuronal but not striated muscle development. *Development* 124, 2179-2189.

Lanjuin, A., and Sengupta, P. (2004). Specification of chemosensory neuron subtype identities in *Caenorhabditis elegans*. *Current opinion in neurobiology* 14, 22-30.

Ledent, V., and Vervoort, M. (2001). The basic helix-loop-helix protein family: comparative genomics and phylogenetic analysis. *Genome research* 11, 754-770.

Lee, J.E., Hollenberg, S.M., Snider, L., Turner, D.L., Lipnick, N., and Weintraub, H. (1995). Conversion of *Xenopus* ectoderm into neurons by NeuroD, a basic helix-loop-helix protein. *Science* 268, 836-844.

LeMay, M. (1982). Morphological aspects of human brain asymmetry: An evolutionary perspective *Trends Neurosci*, 273-275.

Liang, J.O., Etheridge, A., Hantsoo, L., Rubinstein, A.L., Nowak, S.J., Izpisua Belmonte, J.C., and Halpern, M.E. (2000). Asymmetric nodal signaling in the zebrafish diencephalon positions the pineal organ. *Development* 127, 5101-5112.

Lichtsteiner, S., and Tjian, R. (1995). Synergistic activation of transcription by UNC-86 and MEC-3 in *Caenorhabditis elegans* embryo extracts. *Embo J* 14, 3937-3945.

Liu, C., Maejima, T., Wyler, S.C., Casadesus, G., Herlitze, S., and Deneris, E.S. (2010). Pet-1 is required across different stages of life to regulate serotonergic function. *Nature neuroscience* 13, 1190-1198.

Martinat, C., Bacci, J.J., Leete, T., Kim, J., Vanti, W.B., Newman, A.H., Cha, J.H., Gether, U., Wang, H., and Abeliovich, A. (2006). Cooperative transcription activation by Nurr1 and Pitx3 induces embryonic stem cell maturation to the midbrain dopamine neuron phenotype. *Proc Natl Acad Sci U S A* *103*, 2874-2879.

McIntire, S.L., Jorgensen, E., and Horvitz, H.R. (1993). Genes required for GABA function in *Caenorhabditis elegans*. *Nature* *364*, 334-337.

Miklosi, A., and Andrew, R.J. (1999). Right eye use associated with decision to bite in zebrafish. *Behavioural brain research* *105*, 199-205.

Miklosi, A., Andrew, R.J., and Savage, H. (1997). Behavioural lateralisation of the tetrapod type in the zebrafish (*Brachydanio rerio*). *Physiol Behav* *63*, 127-135.

Miyata, T., Maeda, T., and Lee, J.E. (1999). NeuroD is required for differentiation of the granule cells in the cerebellum and hippocampus. *Genes Dev* *13*, 1647-1652.

Nakano, S., Ellis, R.E., and Horvitz, H.R. (2010). Otx-dependent expression of proneural bHLH genes establishes a neuronal bilateral asymmetry in *C. elegans*. *Development* *137*, 4017-4027.

Nishida, A., Furukawa, A., Koike, C., Tano, Y., Aizawa, S., Matsuo, I., and Furukawa, T. (2003). Otx2 homeobox gene controls retinal photoreceptor cell fate and pineal gland development. *Nature neuroscience* *6*, 1255-1263.

Nolo, R., Abbott, L.A., and Bellen, H.J. (2000). Senseless, a Zn finger transcription factor, is necessary and sufficient for sensory organ development in *Drosophila*. *Cell* *102*, 349-362.

Ortiz, C.O., Etchberger, J.F., Posy, S.L., Frokjaer-Jensen, C., Lockery, S., Honig, B., and Hobert, O. (2006). Searching for neuronal left/right asymmetry: genome-wide analysis of nematode receptor-type guanylyl cyclases. *Genetics* *173*, 131-149.

Ortiz, C.O., Faumont, S., Takayama, J., Ahmed, H.K., Goldsmith, A.D., Pocock, R., McCormick, K.E., Kunimoto, H., Iino, Y., Lockery, S., *et al.* (2009). Lateralized gustatory behavior of *C. elegans* is controlled by specific receptor-type guanylyl cyclases. *Curr Biol* *19*, 996-1004.

Poole, R.J., Bashllari, E., Cochella, L., Flowers, E.B., and Hobert, O. (2011). A Genome-Wide RNAi Screen for Factors Involved in Neuronal Specification in *Caenorhabditis elegans*. *PLoS genetics* 7, e1002109.

Poole, R.J., and Hobert, O. (2006). Early embryonic programming of neuronal left/right asymmetry in *C. elegans*. *Curr Biol* 16, 2279-2292.

Portman, D.S., and Emmons, S.W. (2000). The basic helix-loop-helix transcription factors LIN-32 and HLH-2 function together in multiple steps of a *C. elegans* neuronal sublineage. *Development* 127, 5415-5426.

Prasad, B.C., Ye, B., Zackhary, R., Schrader, K., Seydoux, G., and Reed, R.R. (1998). *unc-3*, a gene required for axonal guidance in *Caenorhabditis elegans*, encodes a member of the O/E family of transcription factors. *Development* 125, 1561-1568.

Purves, D. (2008). *Neuroscience*, 4th edition. edn (Sunderland, MA.; Sinauer Associates, Inc.).

Rakic, P. (2009). Evolution of the neocortex: a perspective from developmental biology. *Nat Rev Neurosci* 10, 724-735.

Ramsdell, A.F., and Yost, H.J. (1998). Molecular mechanisms of vertebrate left-right development. *Trends Genet* 14, 459-465.

Raya, A., and Izpisua Belmonte, J.C. (2006). Left-right asymmetry in the vertebrate embryo: from early information to higher-level integration. *Nature reviews Genetics* 7, 283-293.

Reece-Hoyes, J.S., Deplancke, B., Shingles, J., Grove, C.A., Hope, I.A., and Walhout, A.J. (2005). A compendium of *Caenorhabditis elegans* regulatory transcription factors: a resource for mapping transcription regulatory networks. *Genome biology* 6, R110.

Rogers, L.J. (2000). Evolution of hemispheric specialization: advantages and disadvantages. *Brain and language* 73, 236-253.

Roth, G., and Dicke, U. (2005). Evolution of the brain and intelligence. *Trends in cognitive sciences* 9, 250-257.

Ruvinsky, I., Ohler, U., Burge, C.B., and Ruvkun, G. (2007). Detection of broadly expressed neuronal genes in *C. elegans*. *Dev Biol* 302, 617-626.

Sarin, S., Antonio, C., Tursun, B., and Hobert, O. (2009). The *C. elegans* Tailless/TLX transcription factor *nhr-67* controls neuronal identity and left/right asymmetric fate diversification. *Development* 136, 2933-2944.

Sarin, S., O'Meara, M.M., Flowers, E.B., Antonio, C., Poole, R.J., Didiano, D., Johnston, R.J., Jr., Chang, S., Narula, S., and Hobert, O. (2007). Genetic screens for *Caenorhabditis elegans* mutants defective in left/right asymmetric neuronal fate specification. *Genetics* 176, 2109-2130.

Shiratori, H., and Hamada, H. (2006). The left-right axis in the mouse: from origin to morphology. *Development* 133, 2095-2104.

Sieburth, D., Ch'ng, Q., Dybbs, M., Tavazoie, M., Kennedy, S., Wang, D., Dupuy, D., Rual, J.F., Hill, D.E., Vidal, M., *et al.* (2005). Systematic analysis of genes required for synapse structure and function. *Nature* 436, 510-517.

Smidt, M.P., and Burbach, J.P. (2009). Terminal differentiation of mesodiencephalic dopaminergic neurons: the role of *Nurr1* and *Pitx3*. *Advances in experimental medicine and biology* 651, 47-57.

Sulston, J.E. (1983). Neuronal cell lineages in the nematode *Caenorhabditis elegans*. *Cold Spring Harb Symp Quant Biol* 48 Pt 2, 443-452.

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

Suzuki, H., Thiele, T.R., Faumont, S., Ezcurra, M., Lockery, S.R., and Schafer, W.R. (2008). Functional asymmetry in *Caenorhabditis elegans* taste neurons and its computational role in chemotaxis. *Nature* 454, 114-117.

Swoboda, P., Adler, H.T., and Thomas, J.H. (2000). The RFX-type transcription factor DAF-19 regulates sensory neuron cilium formation in *C. elegans*. *Molecular cell* 5, 411-421.

Tononi, G., and Edelman, G.M. (1998). Consciousness and complexity. *Science* 282, 1846-1851.

Topalidou, I., van Oudenaarden, A., and Chalfie, M. (2011). *Caenorhabditis elegans* *aristales/Arx* gene *alr-1* restricts variable gene expression. *Proc Natl Acad Sci U S A* *108*, 4063-4068.

Tursun, B., Cochella, L., Carrera, I., and Hobert, O. (2009). A toolkit and robust pipeline for the generation of fosmid-based reporter genes in *C. elegans*. *PLoS ONE* *4*, e4625.

Uchida, O., Nakano, H., Koga, M., and Ohshima, Y. (2003). The *C. elegans* *che-1* gene encodes a zinc finger transcription factor required for specification of the ASE chemosensory neurons. *Development* *130*, 1215-1224.

Villares, R., and Cabrera, C.V. (1987). The *achaete-scute* gene complex of *D. melanogaster*: conserved domains in a subset of genes required for neurogenesis and their homology to *myc*. *Cell* *50*, 415-424.

Way, J.C., and Chalfie, M. (1988). *mec-3*, a homeobox-containing gene that specifies differentiation of the touch receptor neurons in *C. elegans*. *Cell* *54*, 5-16.

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

Wes, P.D., and Bargmann, C.I. (2001). *C. elegans* odour discrimination requires asymmetric diversity in olfactory neurons. *Nature* *410*, 698-701.

White, J.G., Horvitz, H.R., and Sulston, J.E. (1982). Neurone differentiation in cell lineage mutants of *Caenorhabditis elegans*. *Nature* *297*, 584-587.

White, J.G., Southgate, E., Thomson, J.N., and Brenner, S. (1976). The structure of the ventral nerve cord of *Caenorhabditis elegans*. *Philos Trans R Soc Lond B Biol Sci* *275*, 327-348.

White, J.G., Southgate, E., Thomson, J.N., and Brenner, S. (1986). The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Philos Trans R Soc Lond B Biol Sci* *314*, 1-340.

Wood, W.B. (1997). Left-right asymmetry in animal development. *Annu Rev Cell Dev Biol* *13*, 53-82.

Xue, D., Finney, M., Ruvkun, G., and Chalfie, M. (1992). Regulation of the *mec-3* gene by the *C.elegans* homeoproteins UNC-86 and MEC-3. *Embo J* 11, 4969-4979.

Yu, R.T., McKeown, M., Evans, R.M., and Umesono, K. (1994). Relationship between *Drosophila* gap gene *tailless* and a vertebrate nuclear receptor Tlx. *Nature* 370, 375-379.

Zhang, Y., Ma, C., Delohery, T., Nasipak, B., Foat, B.C., Bounoutas, A., Bussemaker, H.J., Kim, S.K., and Chalfie, M. (2002). Identification of genes expressed in *C. elegans* touch receptor neurons. *Nature* 418, 331-335.

Zhao, C., and Emmons, S.W. (1995). A transcription factor controlling development of peripheral sense organs in *C. elegans*. *Nature* 373, 74-78.

CHAPTER 2:

The LIM and POU homeobox genes *ttx-3* and *unc-86* act as terminal selectors in distinct cholinergic and serotonergic neuron types

Zhang F, Bhattacharya A, Nelson JC, Abe N, Gordon P, Lloret-Fernandez C, Maicas M, Flames N, Mann RS, Colón-Ramos DA and Hobert O.

The “terminal selector” theory is a widely accepted concept on how the differentiated fate of a neuron is adopted at the very end of neural development stages. A dedicated transcription factor selectively binds to a common and usually evolutionarily conserved *cis*-regulatory DNA motif present in the regulatory regions of its target genes to turn on expression of terminal features of a specific neuron. This chapter looks into the question of how the same transcription factor can be responsible for distinct cell fates in different cell types, starting with the LIM homeodomain transcription factor TTX-3, which likely acts with different cooperative partners in different cell types to induce completely different sets of terminal gene expression. The POU homeodomain transcription factor UNC-86 works in a similar manner. It synergistically collaborates with TTX-3 in the serotonergic motoneuron NSM, while it partners with the transcription CFI-1 to control the fate of the cholinergic inner labial neuron IL2 and the ring motoneuron URA. This work supports the concept that the same transcription factor can be reused in distinct cell types, and that it is the combination of different transcription factors expressed in the same cell that defines neuron-type specific terminal identities.

I performed genetic analysis of the AIA and NSM terminal fate, dissected the *cis*-regulatory information in AIA terminal genes, and discovered the cooperative activity between TTX-3 and UNC-86. Bhattacharya A and Gordon P performed genetic analysis

on IL2 and URA neurons. Nelson JC examined the morphology defects in NSM mutants. Abe N performed EMSA experiments. Lloret-Fernandez C, Maicas M and Flames N tested and assessed the requirement of the POU binding sites in UNC-86 responsive genes.

RESEARCH ARTICLE

The LIM and POU homeobox genes *ttx-3* and *unc-86* act as terminal selectors in distinct cholinergic and serotonergic neuron types

Feifan Zhang¹, Abhishek Bhattacharya¹, Jessica C. Nelson², Namiko Abe¹, Patricia Gordon¹, Carla Lloret-Fernandez³, Miren Maicas³, Nuria Flames^{1,3}, Richard S. Mann¹, Daniel A. Colón-Ramos² and Oliver Hobert^{1,4,*}

ABSTRACT

Transcription factors that drive neuron type-specific terminal differentiation programs in the developing nervous system are often expressed in several distinct neuronal cell types, but to what extent they have similar or distinct activities in individual neuronal cell types is generally not well explored. We investigate this problem using, as a starting point, the *C. elegans* LIM homeodomain transcription factor *ttx-3*, which acts as a terminal selector to drive the terminal differentiation program of the cholinergic AIY interneuron class. Using a panel of different terminal differentiation markers, including neurotransmitter synthesizing enzymes, neurotransmitter receptors and neuropeptides, we show that *ttx-3* also controls the terminal differentiation program of two additional, distinct neuron types, namely the cholinergic AIA interneurons and the serotonergic NSM neurons. We show that the type of differentiation program that is controlled by *ttx-3* in different neuron types is specified by a distinct set of collaborating transcription factors. One of the collaborating transcription factors is the POU homeobox gene *unc-86*, which collaborates with *ttx-3* to determine the identity of the serotonergic NSM neurons. *unc-86* in turn operates independently of *ttx-3* in the anterior ganglion where it collaborates with the ARID-type transcription factor *cfi-1* to determine the cholinergic identity of the IL2 sensory and URA motor neurons. In conclusion, transcription factors operate as terminal selectors in distinct combinations in different neuron types, defining neuron type-specific identity features.

KEY WORDS: *Caenorhabditis elegans*, Homeobox, Neuron differentiation

INTRODUCTION

The development of the nervous system is a multistep process that employs a series of sequentially acting regulatory factors that successively restrict and determine cellular fates. During the process of terminal differentiation, individual neuron types acquire specific, hard-wired features that are maintained by the neuron type throughout the life of the animal. A number of transcription factors have been identified that initiate and maintain specific terminal differentiation programs in the developing nervous system (Hobert, 2011). For example, in mouse, the Nurr1 (Nr4a2) transcription

factor initiates and maintains the terminal differentiation program of dopaminergic neurons in the midbrain (Smidt and Burbach, 2009), whereas the Pet1 transcription factor initiates and maintains the terminal differentiation program of serotonergic neurons (Liu et al., 2010). However, few neuronal transcription factors are expressed exclusively in only one specific neuronal cell type (Gray et al., 2004; Lein et al., 2007). For example, in addition to being expressed in midbrain dopaminergic neurons, Nurr1 is expressed in other non-dopaminergic neuronal cell types in which its function is not well understood, such as the adult olfactory bulb, specific cortical areas and the hippocampus (Zetterström et al., 1996). The expression of a given transcription factor in distinct neuronal populations poses the fundamental question of whether there are underlying common themes in the activity of the transcription factor in distinct neuronal cell types.

We have undertaken a systematic, in-depth comparison of the activity of two transcription factors in the development of several distinct neuronal cell types in the nematode *C. elegans*, examining whether there are indeed conceptual similarities in the activities of a given transcription factor in distinct neuron types. We used, as a starting point, a member of the LIM homeobox gene family, an ancient family of neuronal patterning genes that display complex expression patterns in the nervous system of many different species, from invertebrates to vertebrates (Hobert and Westphal, 2000; Simmons et al., 2012; Srivastava et al., 2010). One unifying theme is their expression in terminally differentiating neurons (Hobert and Ruvkun, 1998; Moreno et al., 2005). We focus here on the *ttx-3* LIM homeobox gene, which is the sole *C. elegans* member of the Lhx2/9 subclass of LIM homeobox genes. In vertebrates, Lhx2 is expressed in multiple neuronal cell types and is required for the differentiation of olfactory sensory neurons (Hirota and Mombaerts, 2004; Kolterud et al., 2004), the specification of cortical neuron fate (Mangale et al., 2008) and the differentiation of thalamic neurons (Peukert et al., 2011). Whether there is a common theme in the function of Lhx2 in these distinct neuronal cell types is not known.

The *C. elegans* Lhx2/9 ortholog *ttx-3* is exclusively expressed in a small number of neurons in distinct head ganglia (Altun-Gultekin et al., 2001). *ttx-3* null animals display broad differentiation defects in the cholinergic AIY interneuron class. AIY interneurons of *ttx-3* null mutants are generated and still express pan-neuronal features, but fail to express scores of terminal identity markers that define the functional properties of AIY, including genes required to synthesize and package acetylcholine, genes encoding neuropeptide receptors, various types of ion channels and many others (Altun-Gultekin et al., 2001; Hobert et al., 1997; Wenick and Hobert, 2004). TTX-3 exerts this control through direct binding to a *cis*-regulatory motif shared by all of its target genes. *ttx-3* expression is turned on in the

¹Department of Biochemistry and Molecular Biophysics, Columbia University Medical Center, New York, NY 10032, USA. ²Department of Cell Biology, Yale University School of Medicine, New Haven, CT 06520, USA. ³Instituto de Biomedicina de Valencia, IBV-CSIC, 46010 Valencia, Spain. ⁴Howard Hughes Medical Institute, Columbia University Medical Center, New York, NY 10032, USA.

*Author for correspondence (or38@columbia.edu)

Received 31 May 2013; Accepted 11 October 2013

neuroblast that generates AIY and its expression is maintained throughout the life of the neuron through an autoregulatory feedback loop (Bertrand and Hobert, 2009) to ensure persistent expression of its target genes. A number of transcription factors have been described in the *C. elegans* nervous system that display similar broad-ranging effects on the terminal differentiation programs executed by the neurons in which they are expressed. These transcription factors have been called ‘terminal selectors’ (Hobert, 2008; Hobert, 2011). It is still an open question how broadly the terminal selector concept applies throughout the nervous system; that is, how common it is that many distinct and functionally unrelated identity features of a specific neuron type are directly co-regulated by a transcription factor or a combination of transcription factors.

Here, we investigate the role of *ttx-3* in two additional neuron classes in which it is normally expressed, namely the cholinergic AIA interneuron class and the serotonergic NSM neuron class. We find in all three neuron classes that there is a common theme of *ttx-3* function in that it is broadly required to induce many distinct and functionally unrelated terminal identity features of the respective neuron class. Yet the downstream targets of *ttx-3* in these neuron classes are distinct and are determined by the cooperation of *ttx-3* with a distinct set of transcription factors in different neuron classes.

One of these factors is the POU homeobox gene *unc-86*, which is required together with *ttx-3* to control the identity of the serotonergic NSM neurons. *unc-86* in turn cooperates with the ARID-type transcription factor *cfi-1* to control many terminal identity features of the cholinergic IL2 sensory and URA motor neurons. Our studies therefore provide further support for the terminal selector concept and show that, in combination with other regulatory factors, one factor can serve as terminal selector in distinct neuronal cell types regulating distinct neuronal differentiation programs.

RESULTS

Expression pattern of *ttx-3* in the *C. elegans* larval and adult nervous system

A *ttx-3* reporter gene that contains the *ttx-3* locus together with a few kilobases upstream but no downstream sequences (*ttx-3^{promA}::gfp*; Fig. 1) was previously shown to be continuously expressed in five distinct neuronal cell types: the cholinergic AIY and AIA interneuron classes, the ASI and ADL chemosensory neuron classes and a previously uncharacterized neuronal pair in the pharyngeal nervous system (Altun-Gultekin et al., 2001). Transient expression was observed in the AIN and SMDD neurons at embryonic stages (Bertrand and Hobert, 2009). A fosmid reporter construct, which contains more than 30 kb surrounding the *ttx-3* locus and which

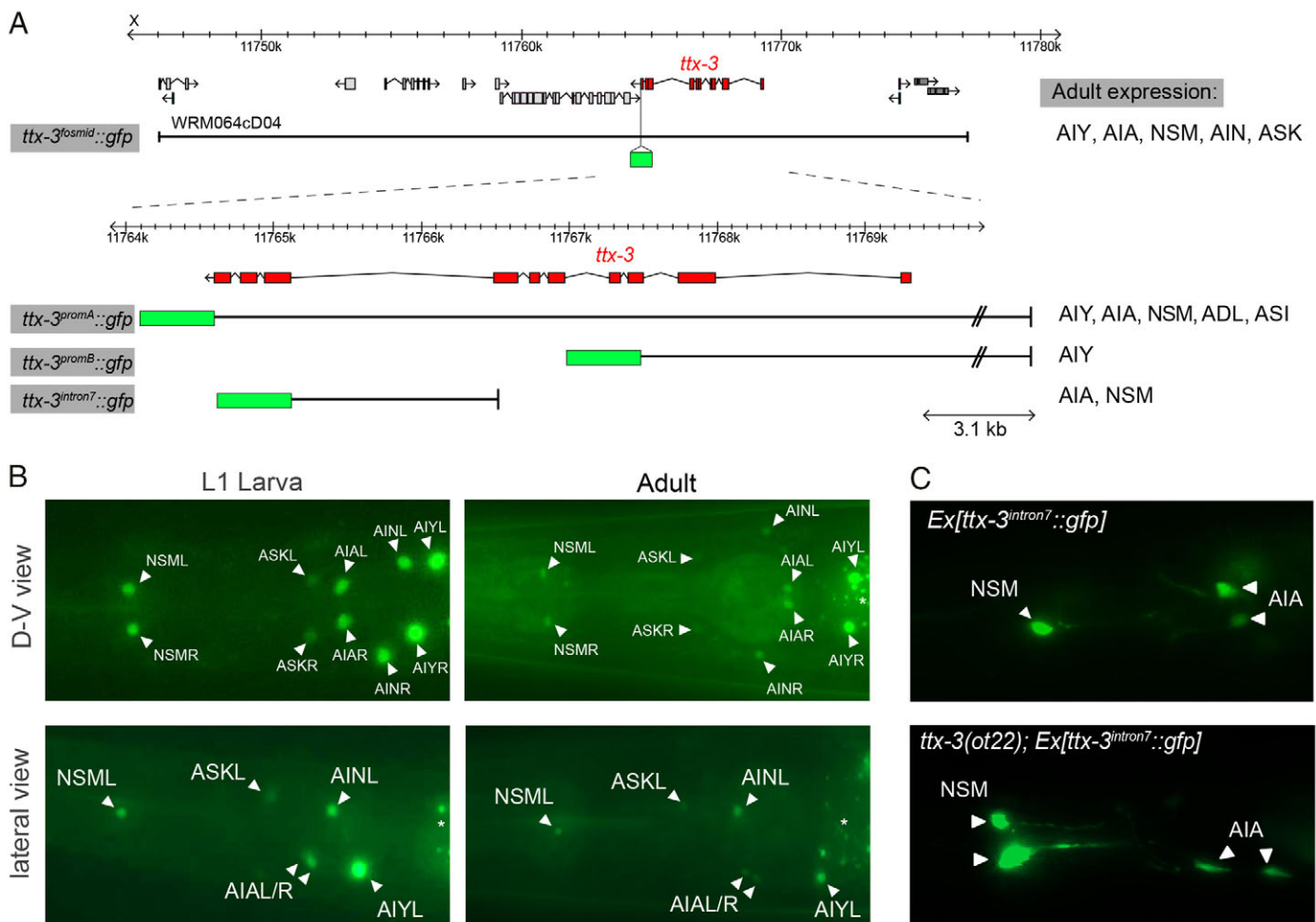


Fig. 1. Expression pattern of the *C. elegans* *ttx-3* LIM homeobox gene. (A) *ttx-3* expression constructs and summary of neuronal expression pattern. The *promA::gfp* and *promB::gfp* constructs were described previously (Altun-Gultekin et al., 2001; Wenick and Hobert, 2004) and are shown here for comparison only. (B) *ttx-3* fosmid expression (*wg/s68*) in first larval stage animals and in adult animals. D-V, dorsoventral. White asterisks indicate gut autofluorescence. (C) The seventh intron of the *ttx-3* locus contains cis-regulatory elements driving reporter gene expression in AIA and NSM neurons. These regulatory elements do not depend on *ttx-3*. Expression is shown in adult animals.

rescues the AIY differentiation defect of *ttx-3* mutant animals, mirrors the expression of the smaller, locus-restricted reporter construct in the AIY, the AIA, the AIN and the pharyngeal neuron class (Fig. 1). Based on position, morphology and colabeling with the NSM marker *mgl-1::mCherry*, we identified the pharyngeal neurons that express *ttx-3* as the NSM neuron pair. The NSM neurons are serotonergic, neurosecretory cells that are thought to be involved in sensing food (Albertson and Thomson, 1976; Harris et al., 2011; Horvitz et al., 1982).

There are also notable differences in the expression pattern of the fosmid reporter and the smaller reporters. First, expression in the AIN neurons is maintained throughout development with the fosmid reporter, whereas it is restricted to embryos with smaller reporters (Bertrand and Hobert, 2009). Second, the expression in amphid sensory neurons is markedly different. In larval and adult animals, the fosmid reporter is expressed in the ASK neuron class, whereas the smaller reporters are expressed in the ADL and ASI sensory neurons (Fig. 1).

Previous studies have shown that *ttx-3* expression in the AIY interneuron pair is controlled by a distal initiator element ~1 kb upstream of the *ttx-3* locus and a maintenance element in the second intron of the *ttx-3* locus (Bertrand and Hobert, 2009; Wenick and Hobert, 2004). We find that the expression of *ttx-3* in the NSM and AIA is controlled via regulatory elements present in the seventh intron of the *ttx-3* locus (Fig. 1). As mentioned above, *ttx-3* expression is maintained throughout the life of the AIA and NSM neurons, but maintained expression of a *ttx-3* reporter gene construct (*ttx-3^{intron7}::gfp*; Fig. 1A) in the AIA and NSM neuron types does not require *ttx-3* gene activity (Fig. 1C).

***ttx-3* controls the differentiation program of AIA interneurons**

We focused our analysis of *ttx-3* mutants on the cholinergic AIA interneurons and the serotonergic NSM neurons, which both continuously express *ttx-3* throughout their lifetime. We have previously reported that expression of the marker of cholinergic identity, *unc-17* (vesicular ACh transporter), as well as the expression of an orphan G protein-coupled receptor (GPCR), *sra-11*, is reduced in the AIA neurons of *ttx-3* mutants (Altun-Gultekin et al., 2001). We extended this analysis by examining the expression of seven additional markers of terminal AIA fate: the choline reuptake transporter encoded by *cho-1*; the metabotropic glutamate receptor *mgl-1*; the ionotropic glutamate receptor *glr-2*; the neuropeptides *flp-2* and *ins-1*; the receptor tyrosine kinase *scd-2*; and the receptor guanylyl cyclase *gcy-28d*. Each of these markers is expressed in terminally differentiated AIA interneurons and several of them have previously been implicated in AIA interneuron function (Shinkai et al., 2011; Tomioka et al., 2006). The expression of each of these seven markers is affected in the AIA neurons of *ttx-3* mutants (Fig. 2). Their expression in other neuron types is unaffected in *ttx-3* mutants, with the exception of two markers that are also downregulated in NSM neurons (*mgl-1*, *scd-2*, as described below). *ttx-3* is likely to act cell-autonomously since the AIA differentiation defects are rescued in transgenic *ttx-3* mutant animals that express *ttx-3* cDNA under control of the *ins-1* promoter (supplementary material Table S1).

AIA neurons remain present in the *ttx-3* null mutant, as assessed by the weak but recognizable expression of some terminal differentiation genes (Fig. 2). However, their normally unipolar neurite morphology appears disrupted; ectopic branches can be observed to emanate from the cell body and the main neurite appears blebbed in *ttx-3* mutants (supplementary material Fig. S2).

The AIY interneurons, which have a unipolar axon morphology similar to that of AIA interneurons in wild-type animals, display similar morphological defects in *ttx-3* mutants (Hobert et al., 1997). The expression of terminal identity markers that label several distinct neuron types that are lineally related to AIA is not altered (data not shown) (Altun-Gultekin et al., 2001), suggesting that the AIA neuron pair might remain in an undifferentiated state, rather than switching to an alternate fate. Based on a more extensive cell fate marker analysis, a similar conclusion was previously drawn about the fate of the AIY neuron class in *ttx-3* mutants (Altun-Gultekin et al., 2001). Taken together, our fate marker and morphological analyses indicate that *ttx-3* broadly affects the AIA terminal differentiation program. These effects are comparable to the previously described broad effects that loss of *ttx-3* has on the terminal differentiation of AIY interneurons.

A shared *cis*-regulatory signature of AIA-expressed terminal identity features

On a mechanistic level, *ttx-3* operates in a distinct manner in the AIA versus AIY neurons since it operates with distinct co-factors and through distinct *cis*-regulatory elements. The co-factor of *ttx-3* in AIY, the *ceh-10* homeobox gene (Altun-Gultekin et al., 2001), is not expressed in AIA neurons, and AIA neurons display no differentiation defects in *ceh-10* null mutants (two markers tested). Moreover, the *cis*-regulatory motifs through which *ttx-3* acts to control AIY versus AIA identity are distinct. In the AIY neurons, *ttx-3* acts on its many target genes through a *cis*-regulatory motif, termed the 'AIY motif', that provides a cooperative binding site for a TTX-3–CEH-10 heterodimer (Wenick and Hobert, 2004). Mutation of the AIY motif in a locus that is expressed in AIY and AIA neurons, the cholinergic *cho-1* locus, results in a severe reduction in expression in the AIY interneurons but not in the AIA interneurons (Fig. 3A).

In the AIA neurons, by contrast, *ttx-3* acts through a distinct *cis*-regulatory signature, which we deciphered through a mutational analysis of the *cis*-regulatory control regions of three AIA-expressed, *ttx-3*-dependent terminal differentiation genes: *mgl-1*, *ins-1* and *cho-1*. We generated transgenic animals that express nested, shorter versions of these three reporters and identified a 259 bp element in the *cho-1* promoter, a 74 bp element in the *mgl-1* promoter and a 68 bp element in the *ins-1* promoter that are sufficient to direct *gfp* expression to AIA neurons (Fig. 3A-C). Examining these elements for common patterns, we noted that all these elements contain a shared and phylogenetically conserved G(A/G)ATC motif (Fig. 3D). Mutating this motif in the context of any of the three promoters resulted in a reduction of AIA expression of the respective reporter (Fig. 3A-C). In the case of *mgl-1*, two G(A/G)ATC motifs are present in the minimal promoter; mutation of either causes an intermediate reduction in reporter gene expression, and mutation of both motifs results in complete loss of expression (Fig. 3A-C).

Since G(A/G)ATC does not match the consensus binding site for a LIM homeodomain transcription factor such as TTX-3, we also examined the minimal reporters for the presence of conserved TAAT motifs, which comprises the core consensus site for LIM homeodomain transcription factors (Berger et al., 2008). We indeed found several TAAT motifs in the three *cis*-regulatory modules and for each of them we identified a TAAT motif that, when mutated, affected reporter gene expression *in vivo* (Fig. 3A-C). These TAAT motifs can be assembled into a larger sequence matrix, TAATTNGA (Fig. 3D). In two cases, mutation of the TAATTNGA alone affected reporter gene expression, whereas in the third case (*cho-1*) a

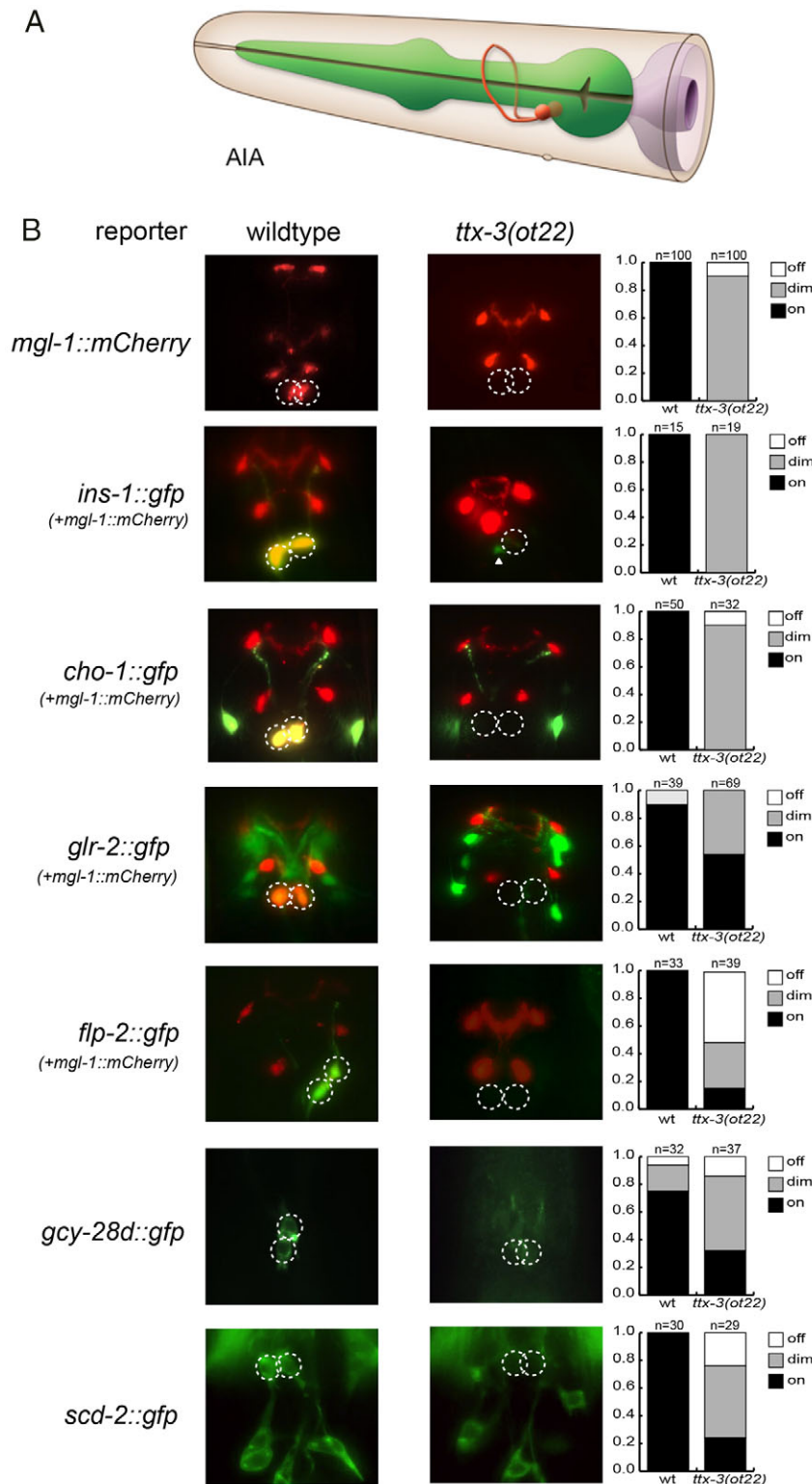


Fig. 2. *ttx-3* affects the terminal differentiation of AIA neurons. (A) Schematic representation of the AIA interneuron pair [reproduced with permission (Altun et al., 2002-2013)]. (B) The expression of terminal differentiation markers of AIA identity is affected in *ttx-3* mutants. Reporter gene arrays were crossed into *ttx-3(ot22)* null mutants. Positions of AIA neurons are outlined (dashed circles). The fraction of animals that show the indicated phenotype is presented in the bar charts. Transgenic arrays are: *otIs317* for *mgl-1*, *otIs326* for *ins-1*, *otIs379* for *cho-1*, *otEx4687* for *glr-2* and *otEx5056* for *flp-2* (see Materials and methods for more detail on the arrays; the *Ex[gcy-28d::gfp]* and *Ex[scd-2::gfp]* arrays were kindly provided by Takeshi Ishihara). Anterior is up in all panels.

complete loss of expression can only be observed upon simultaneous mutation of both the GAATC motif and the TAAT-containing motif (Fig. 3A). The residual AIA expression of a *cho-1* reporter construct in which the GAATC motif is mutated, but the TAAT motif is left intact, is abolished in *ttx-3* mutants (data not shown), consistent with *ttx-3* operating through the TAAT motif.

We examined whether the TAATTNGA motif is indeed a TTX-3 binding site using gel shift assays with bacterially produced TTX-3

protein and probes derived from the *mgl-1* and *cho-1* locus. We found that TTX-3 is able to bind these sites *in vitro* (Fig. 3E). Deletion of the TAAT site that is required for reporter gene expression *in vivo* resulted in the loss of TTX-3 binding *in vitro* (Fig. 3E).

The combination of G(A/G)ATC and TAAT motifs might define a *cis*-regulatory signature that is generally required for gene expression in AIA neurons, since we found a combination of these two motifs to be present in the *cis*-regulatory control regions of the

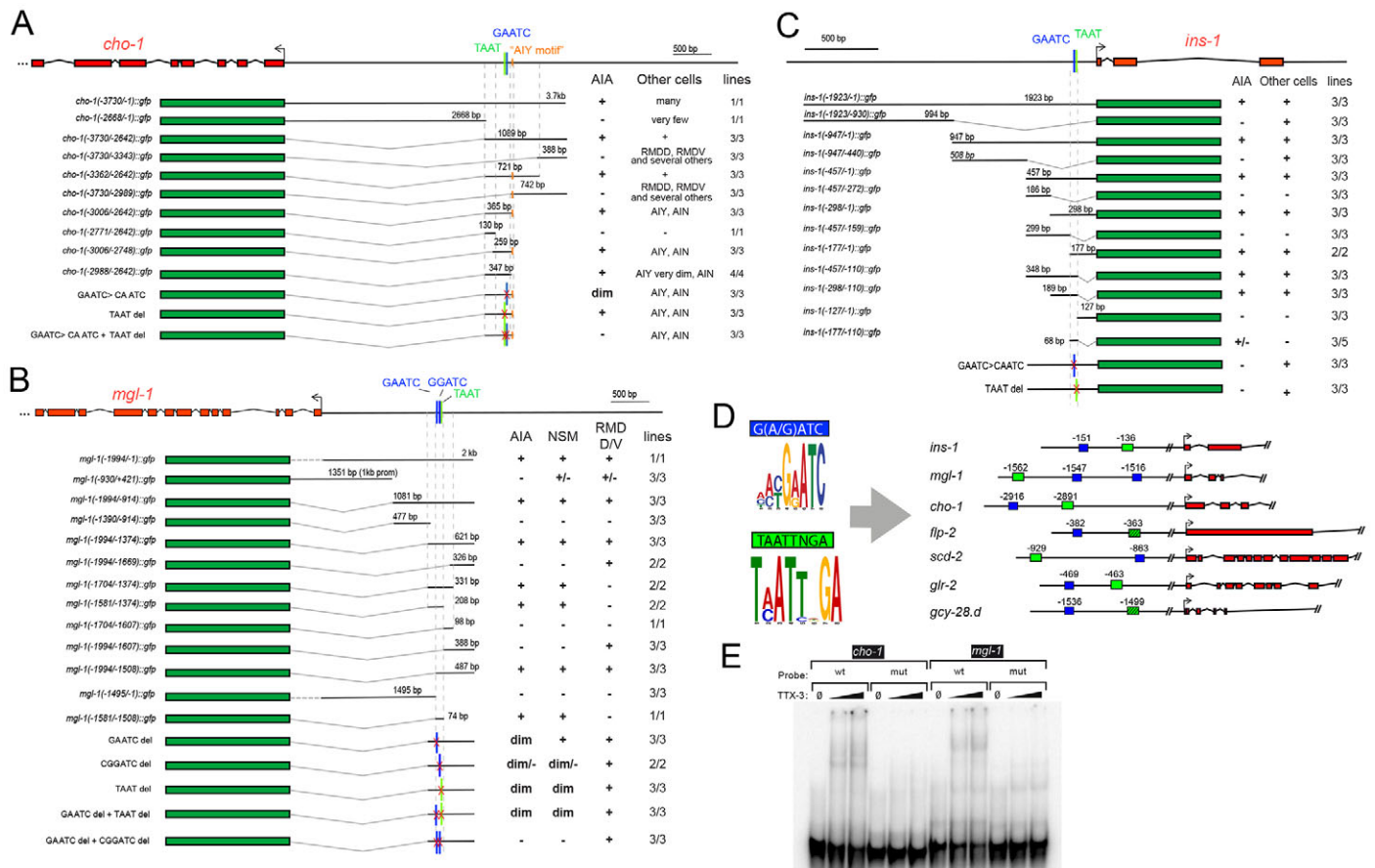


Fig. 3. Co-regulation of AIA-expressed genes by two cis-regulatory motifs. (A-C) Mutational dissection of the cis-regulatory elements of three AIA-expressed terminal identity markers. (D) Position weight matrix of the two motifs required for AIA expression, based on the motifs from *ins-1*, *cho-1* and *mgl-1* and orthologs in other nematode species. Perfect (filled box) and imperfect (stippled box) matches to the two cis-regulatory motifs [blue, G(A/G)ATC; green, TAATTNGA] in other AIA terminal identity markers are shown on the right. (E) TTX-3 binds to *cho-1* and *mgl-1* regulatory elements containing the HD (TAAT) motif. Deletion of the HD motif abolishes binding. EMSA was performed with 250 nM and 750 nM TTX-3.

other four *ttx-3*-dependent terminal AIA markers (Fig. 3D). Taken together, these data show that AIA identity features are co-regulated by a shared cis-regulatory signature that is controlled by TTX-3 and an as yet unknown co-factor.

ttx-3 controls the terminal differentiation of serotonergic NSM neurons

We next analyzed the effect of loss of *ttx-3* on the terminal differentiation program of the serotonergic NSM neurons, a neuron type that has not previously been examined in *ttx-3* mutants. Many terminal identity markers of NSM have been described, including the battery of genes that are required to synthesize, package and reuptake serotonin: *tph-1/TPH* (tryptophan hydroxylase), *cat-4/GTPCH* (GTP cyclohydrolase), *cat-1/VMAT* (vesicular monoamine transporter), *bas-1/AAAD* (aromatic amino acid decarboxylase) and *mod-5/SERT* (serotonin reuptake transporter) (Fig. 4A) (Jafari et al., 2011; Ranganathan et al., 2001; Sze et al., 2002). Previous expression analysis of a vesicular glutamate transporter, *eat-4*, suggested that NSM might use the neurotransmitter glutamate (Lee et al., 1999). However, a fosmid-based *eat-4* reporter does not show expression in NSM neurons (Serrano-Saiz et al., 2013) (supplementary material Fig. S1A).

To broaden the spectrum of available terminal markers, we analyzed the expression of other *C. elegans* orthologs of enzymes

involved in monoaminergic transmitter metabolism (Fig. 4A) and identified another NSM-expressed terminal marker, *ptps-1* (Fig. 4C; supplementary material Fig. S3). In addition to examining these serotonin (5HT)-related markers, we also examined the expression of three metabotropic neurotransmitter receptors (*mgl-1*, *mgl-3*, *dop-3*), three neuropeptides (*nlp-13*, *flp-4*, *nlp-3*), a glycoprotein hormone alpha subunit (*flr-2*) and a receptor tyrosine kinase (*scd-2*). All of these genes are expressed throughout the life of the NSM neurons. As mentioned above, *scd-2* and *mgl-1* are also expressed in AIA neurons, where their expression is affected by *ttx-3*. We find that the expression of five of these 14 NSM terminal identity markers is either partially or completely eliminated in the NSM neurons of *ttx-3* null mutants (Fig. 4C, Fig. 5, Table 1). *ttx-3* is likely to act cell-autonomously since we can rescue the NSM differentiation defects by driving *ttx-3* cDNA under the control of a *cat-1* promoter fragment, which is expressed in a subset of monoaminergic neurons of *C. elegans* (supplementary material Table S1).

The POU homeobox gene *unc-86* also controls NSM identity

We recently reported that the effects of the loss of a terminal selector type transcription factor in dopaminergic neurons can be partially compensated for by other, co-expressed terminal selectors (Doitsidou et al., 2013). Therefore, we considered the possibility that

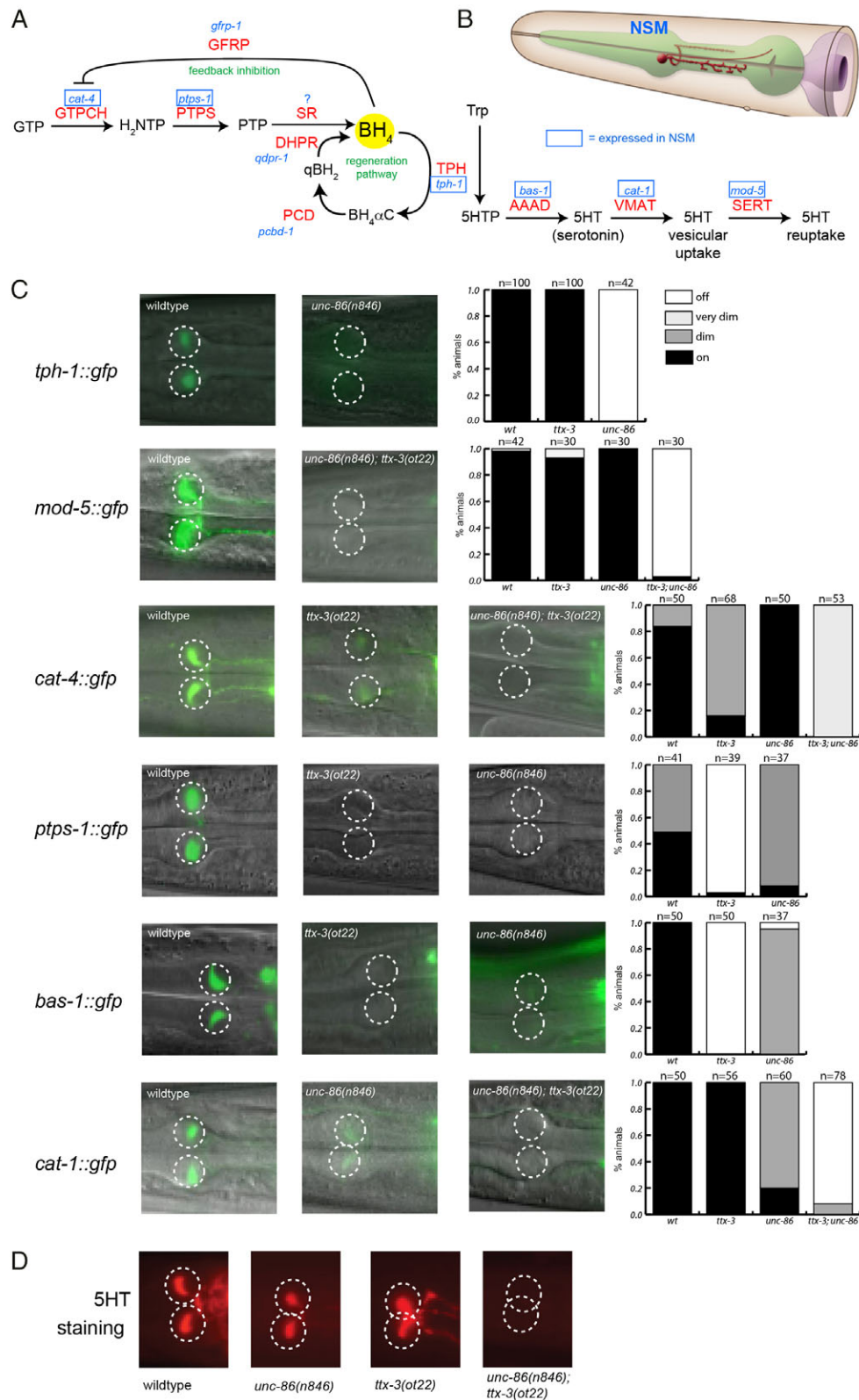


Fig. 4. The effect of *unc-86* and *ttx-3* on the serotonergic identity of NSM neurons. (A) The 5HT pathway including tetrahydrobiopterin biosynthesis genes (Deneris and Wyler, 2012). '?' indicates that a unique homolog of SR could not be identified in the worm genome. (B) Schematic representation of the NSM interneuron pair [reproduced with permission (Altun et al., 2002-2013)]. (C) The expression of serotonergic identity features of NSM (dashed circles) is affected in *unc-86(n846)*, *ttx-3(ot22)* or *unc-86(n846); ttx-3(ot22)* double-null mutants. Reporter gene arrays were crossed into the respective mutant backgrounds. Transgenic arrays are: *zdfs13* for *tph-1*; *otEx4781* for *mod-5*; *otIs225* for *cat-4*; *otEx5280* for *ptps-1*; *otIs226* for *bas-1*; and *otIs224* for *cat-1* (see Materials and methods for more detail on the arrays). Images are only shown for mutant genotypes with effects on reporter expression. (D) Serotonin antibody staining. Thirty animals were scored for each genotype. In the double mutant, no animal showed staining in NSM (circled), whereas in the other genotypes all animals showed staining.

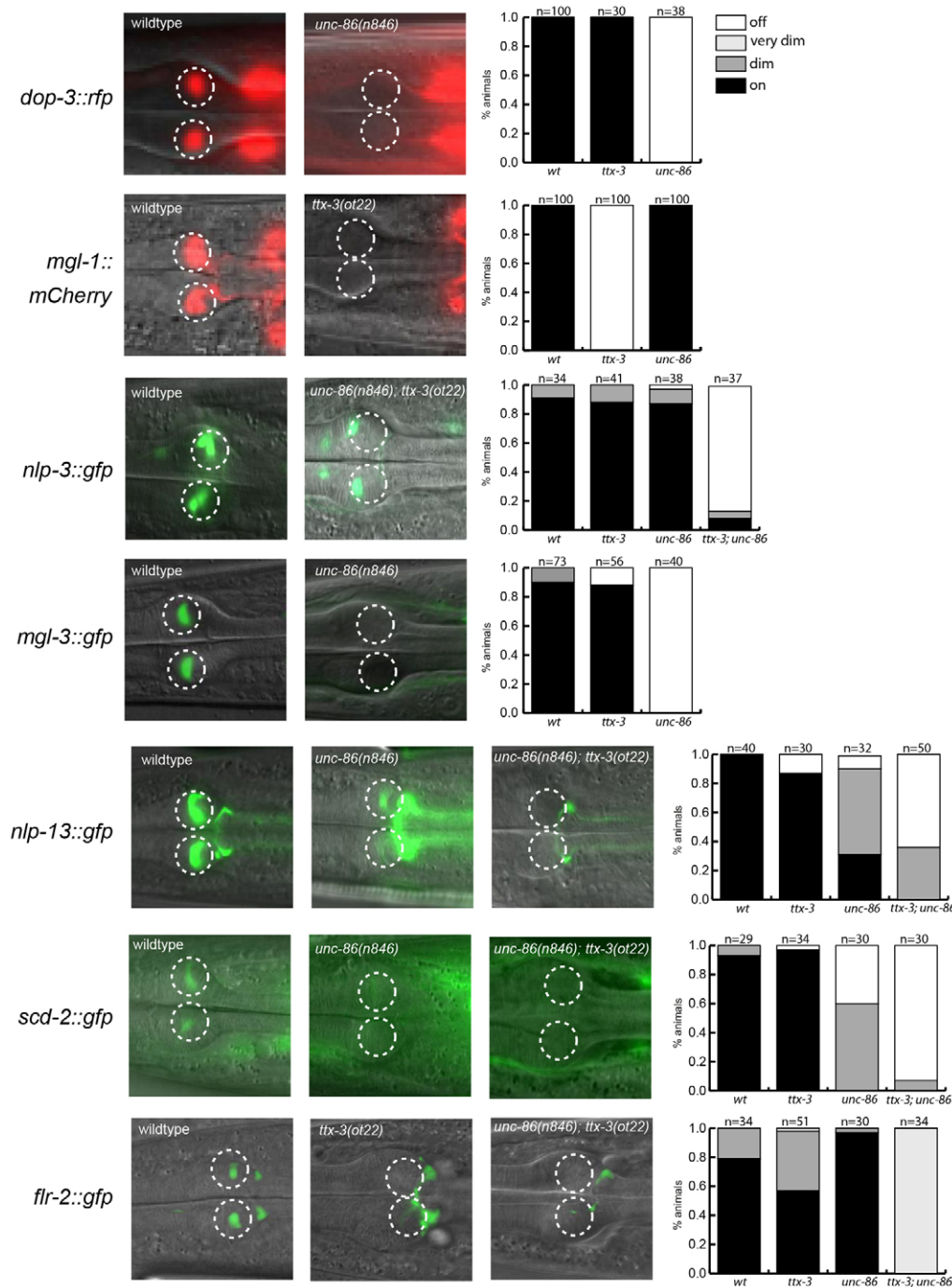


Fig. 5. The effect of *unc-86* and *ttx-3* on other identity features of NSM neurons. The expression of other identity features of NSM is also affected in *unc-86(n846)*, *ttx-3(ot22)* or *unc-86(n846); ttx-3(ot22)* double-null mutants. Reporter gene arrays were crossed into the respective mutant backgrounds. Transgenic arrays are: *vsls33* for *dop-3*; *otls317* for *mgl-1*; *otEx5163* for *nlp-3*; *otEx5364* for *mgl-3*; *otEx5163* for *nlp-13*; *otEx5055* for *scd-2*; and *otEx5363* for *flr-2* (see Materials and methods for more detail on the arrays). Micrographs are only shown for mutant genotypes with effects on reporter expression. Dashed circles indicate the position of NSM neurons. See also Table 1.

the lack of an impact of *ttx-3* loss on nine out of 14 NSM markers could be due to the activity of compensatory terminal selector type transcription factors. We sought to identify such a factor, focusing on two homeodomain transcription factors previously shown to be expressed in NSM, namely the empty spiracles homolog *ceh-2* and the POU homeobox gene *unc-86* (Aspöck et al., 2003; Finney and Ruvkun, 1990). We observed no NSM differentiation defects in *ceh-2* null animals (data not shown), but we observed striking NSM differentiation defects in *unc-86* mutants. Loss of *unc-86* was previously shown to affect the expression of *tph-1* and *cat-1* in NSM neurons, but without effect on 5HT antibody staining (Sze et al., 2002). Other differentiation features of NSM neurons had not previously been examined in *unc-86* mutants. Upon examining the expression of all 14 markers of NSM fate in *unc-86* null mutants,

we found that the expression of eight is partially or completely eliminated (Figs 4, 5, Table 1).

To examine whether *unc-86* directly affects the expression of these terminal identity features, we analyzed the *cis*-regulatory control regions of four of them: *tph-1*, *bas-1*, *cat-1* and *cat-4*. Through mutational analysis, we defined small (~200 bp) elements that still yielded expression in the NSM neurons (Fig. 6) and, within each of these elements, we identified predicted POU homeodomain binding sites (Rhee et al., 1998). We introduced mutations into these sites in the context of two loci (*tph-1* and *bas-1*) and found that these mutations resulted in a loss of reporter gene expression *in vivo* (Fig. 6A,B). Gel shift analysis further confirmed that these POU homeodomain sites indeed bind bacterially produced UNC-86 protein *in vitro* (Fig. 6E).

Table 1. Summary of the effects of *ttx-3* and *unc-86* null mutants on terminal NSM identity markers

Identity feature	Function	<i>ttx-3</i> (-)	<i>unc-86</i> (-)	<i>unc-86</i> (-); <i>ttx-3</i> (-)	Interaction
<i>cat-1</i>	5HT pathway	wt	dim	off	Synergism
<i>cat-4</i>	5HT pathway	dim	wt	very dim	Synergism
<i>mod-5</i>	5HT pathway	wt	wt	off	Synergism
<i>nlp-13</i>	Neuropeptide	wt	dim	off	Synergism
<i>nlp-3</i>	Neuropeptide	wt	wt	off	Synergism
<i>flr-2</i>	Transmembrane	wt	wt	dimmer	Synergism
<i>scd-2</i>	Kinase	wt	dim	off	Synergism
<i>flp-4</i>	Neuropeptide	dim	wt	stronger expression	Synergism
5HT antibody staining	Neurotransmitter	wt	wt	off	Synergism
<i>bas-1</i>	5HT pathway	off	dim	n.d.	?
<i>ptps-1</i>	5HT pathway	off	dim	n.d.	?
<i>tph-1</i>	5HT pathway	wt	off	n.d.	?
<i>mgl-3</i>	GPCR	wt	off	n.d.	?
<i>dop-3</i>	GPCR	wt	off	n.d.	?
<i>mgl-1</i>	GPCR	off	wt	n.d.	?

gfp reporter (or antibody staining): wt, as bright as in wild-type animals; dim, dimmer than in wild type; off, no expression observed.

n.d., not determined because single mutant already shows completely penetrant loss of expression.

Gray shading indicates presence of defect.

***unc-86* cooperates with *ttx-3* to control NSM identity**

We noted that terminal markers of NSM identity that were severely affected in *unc-86* mutants tended to be those that were weakly or unaffected in *ttx-3* mutants; vice versa, markers unaffected in *ttx-3* mutants tended to be affected in *unc-86* mutants (Table 1). Even though this observation might simply indicate that *unc-86* and *ttx-3* act completely independently of one another, we considered the possibility that *unc-86* and *ttx-3* might collaboratively control NSM identity but that their relative importance may be distinct for different target genes. To investigate this possibility, we examined *unc-86*; *ttx-3* double-null mutants and found that markers that are either partially or unaffected in *ttx-3* and *unc-86* single mutants are more strongly affected in the double mutant (Figs 4, 5, Table 1). This also holds for 5HT antibody staining, which is not affected in either single mutant but completely abrogated in the *ttx-3*; *unc-86* double mutant (Fig. 4D), probably owing to the combined effect that both genes have on the expression of the 5HT reuptake transporter *mod-5*. As summarized in Table 1, nine of the 15 tested identity features (14 reporter genes and 5HT antibody staining) are affected by both *ttx-3* and *unc-86*, with effects either visible in both single mutants, or as a non-additive, synergistic effect revealed in the double mutant. As described below, there are also synergistic effects of *ttx-3* and *unc-86* on NSM morphology. In six of the 15 cases, either *ttx-3* or *unc-86* already has completely penetrant effects (Table 1). Taken together, these data argue that *unc-86* and *ttx-3* jointly control terminal NSM differentiation. The mechanistic basis of the cooperation is unclear at present because we have so far not been able to identify functional TTX-3 binding sites in terminal NSM identity marker genes.

To further examine potential interactions of *unc-86* and *ttx-3*, we investigated whether they affect each others expression. We find that continuous expression of *unc-86* in NSM neurons depends on *unc-86* itself [autoregulation of *unc-86* was also previously noted (Baumeister et al., 1996)], but not on *ttx-3* (supplementary material Fig. S1B,C). Vice versa, *ttx-3* expression in NSM neurons is not affected in *unc-86* or in *unc-86*; *ttx-3* mutants (data not shown).

***unc-86* and *ttx-3* affect axonal arborization and presynaptic specializations**

Apart from affecting the expression of terminal identity markers, loss of *unc-86* and *ttx-3* also results in specific effects on the morphology of the NSM neurons. During embryonic stages, these

neurons normally extend a neurite posteriorly toward the nerve ring, which then bifurcates to form a ventral and a dorsal neurite (Axäng et al., 2008) (Fig. 4B). We observe that in *unc-86*(*n846*) mutants the NSM somas are correctly positioned but there are significant defects in outgrowth of the ventral neurite (61% of animals show outgrowth defects; *n*=31). By contrast, in *ttx-3*(*ot22*) mutants, the primary defect observed in ventral neurite outgrowth is the formation of aberrant bifurcations (41% of animals show such defects; *n*=39). Both *ttx-3* and *unc-86* single mutants also show defects in dorsal axon termination [25% (*n*=32) of *ttx-3* mutants and 29% (*n*=34) of *unc-86* mutants].

In early larval stages, ventral NSM neurites begin to extend elaborate arbor structures onto the nerve ring target field (Axäng et al., 2008). These axon arborizations require the NSM-expressed netrin receptor UNC-40/DCC, which is tightly localized to puncta within the main shaft of the NSM neurite and at the tips of axon arbors (Nelson and Colón-Ramos, 2013). These arbor structures persist into the adult stage and contain presynaptic sites, as assessed with a *rab-3* marker (Nelson and Colón-Ramos, 2013). In *ttx-3* mutants, these ultrastructural features are unaffected, but *unc-86* mutants display a highly penetrant defect in axon arborization (Fig. 7A,C). Furthermore, *unc-86* mutants display defects in the dynamic regulation of UNC-40 localization (Fig. 7B,D). In wild-type animals, UNC-40::GFP is diffusely distributed at the L1 stage and becomes localized to bright puncta in the NSM neurite and at the tips of axon arbors as axons are arborizing at the L4 stage. By the adult stage, UNC-40::GFP again becomes diffusely distributed. However, a significant fraction of *unc-86* mutant NSMs retain a juvenile-like pattern of UNC-40 localization during the adult stage, in which UNC-40::GFP remains localized to bright puncta (Fig. 7B).

We observe synergistic morphological defects in *unc-86*(*n846*); *ttx-3*(*ot22*) double mutants. Ventral neurites never reach the middle of the pharyngeal isthmus and are often truncated immediately following the guidance decision to turn posteriorly (100% premature ventral neurite termination, *n*=18; Fig. 7E). Furthermore, neurites contain large anterior swellings not seen in wild-type animals (33% contain additional anterior swellings, *n*=18; Fig. 7E). The morphological appearance of NSM neurites in *unc-86*; *ttx-3* mutants is reminiscent of the normal morphology of M3 neurons (Albertson and Thomson, 1976), which are lineally related to NSM (Sulston et al., 1983). M3 neurons are glutamatergic (Lee et al., 1999) and we indeed find that in *unc-86* mutants the vesicular glutamate

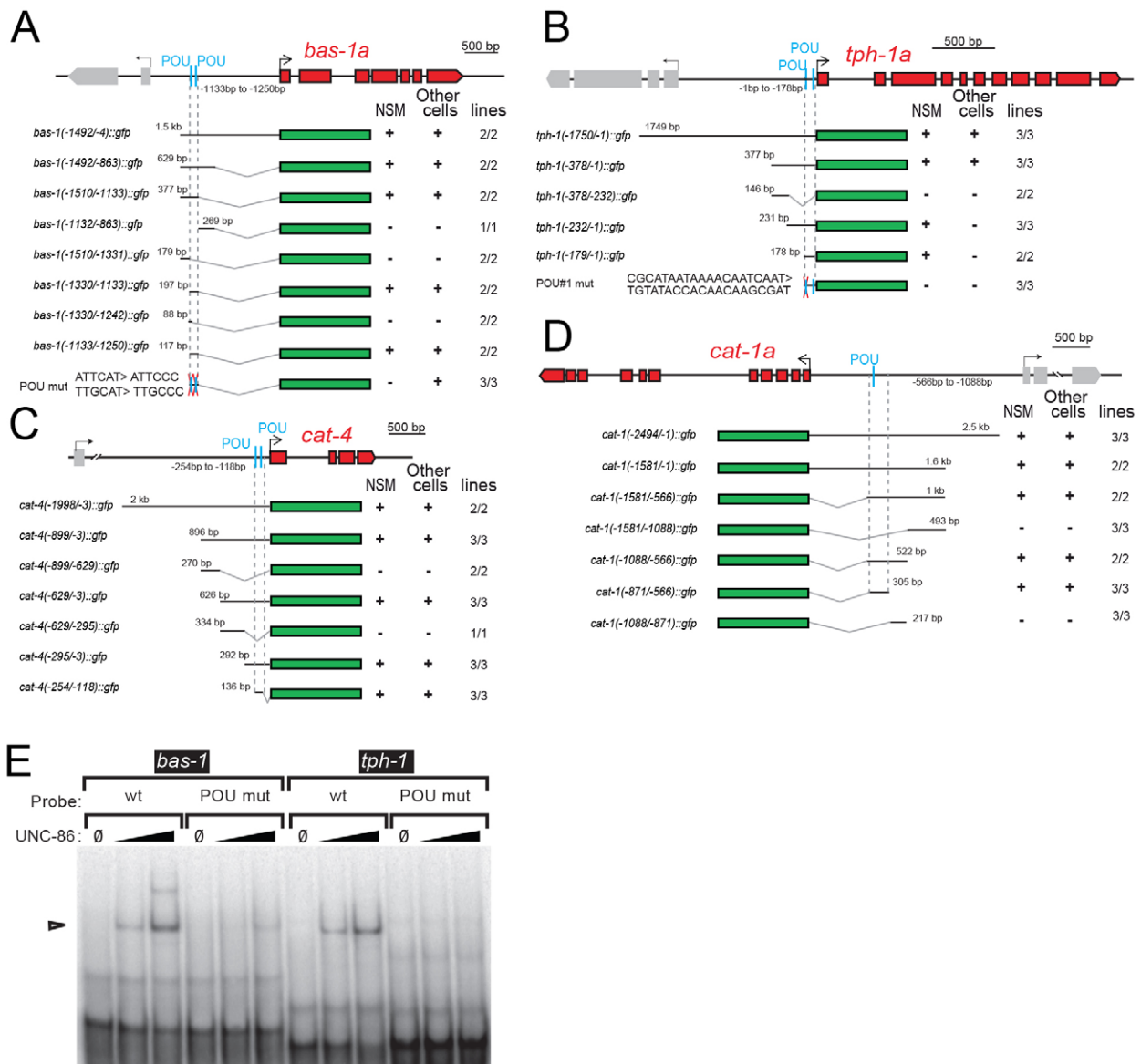


Fig. 6. Cis-regulatory analysis of NSM identity specification. (A–D) Dissection of the *cis*-regulatory elements of four NSM-expressed serotonin pathway genes. All minimal *cis*-regulatory elements contain predicted POU sites. (E) EMSAs with UNC-86 protein on *bas-1* and *tph-1* regulatory elements. Mutated POU sites are those that also disrupt reporter gene activity when deleted from the gene contexts of *bas-1* and *tph-1* (A). EMSA was performed with 10 nM and 30 nM UNC-86. Arrowhead indicates UNC86-bound DNA probe.

transporter *eat-4* is ectopically expressed in NSMs (supplementary material Fig. S1A).

***unc-86* controls terminal differentiation of the cholinergic IL2, URA and URB sensory, motor and interneurons**

Apart from our description of *unc-86* terminal selector function in the serotonergic NSM neurons, *unc-86* had previously been described to broadly affect the terminal differentiation program of other serotonergic (Sze et al., 2002) as well as glutamatergic (Duggan et al., 1998; Serrano-Saiz et al., 2013) neurons. We asked whether *unc-86* might affect the terminal differentiation program of neurons that use yet another neurotransmitter system. We turned to the six IL2 sensory neurons that are involved in nictation behavior (Lee et al., 2012). The IL2 neurons express *unc-86* throughout their lifetime and have been inferred to be cholinergic (Lee et al., 2012). We corroborated the cholinergic identity of the IL2s by finding that

reporter fusions to the *unc-17/cha-1* locus and to the choline reuptake transporter *cho-1* are expressed in IL2 neurons (Fig. 8A). The expression of these two key markers of cholinergic identity is eliminated in *unc-86* mutants (Fig. 8A). Expression of the nicotinic acetylcholine receptor subunit *des-2* is also lost in the IL2 neurons of *unc-86* mutants (Treinin et al., 1998).

In addition to these cholinergic markers, we examined the expression of other genes previously shown to be expressed in IL2 neurons, namely the *unc-5* netrin receptor, the guanylyl cyclase *gcy-19*, the kinesin *klp-6* and the Notch ligand *lag-2* (which is expressed in IL2 neurons at the dauer stage) (Leung-Hagsteijn et al., 1992; Ortiz et al., 2006; Ouellet et al., 2008; Peden and Barr, 2005). The expression of all of these terminal markers of IL2 identity is eliminated in IL2 neurons of *unc-86* mutants (Fig. 8A). IL2 neurons also fail to take up dye in *unc-86* mutants (Tong and Bürglin, 2010), suggesting morphological defects. The IL2 neurons are nevertheless

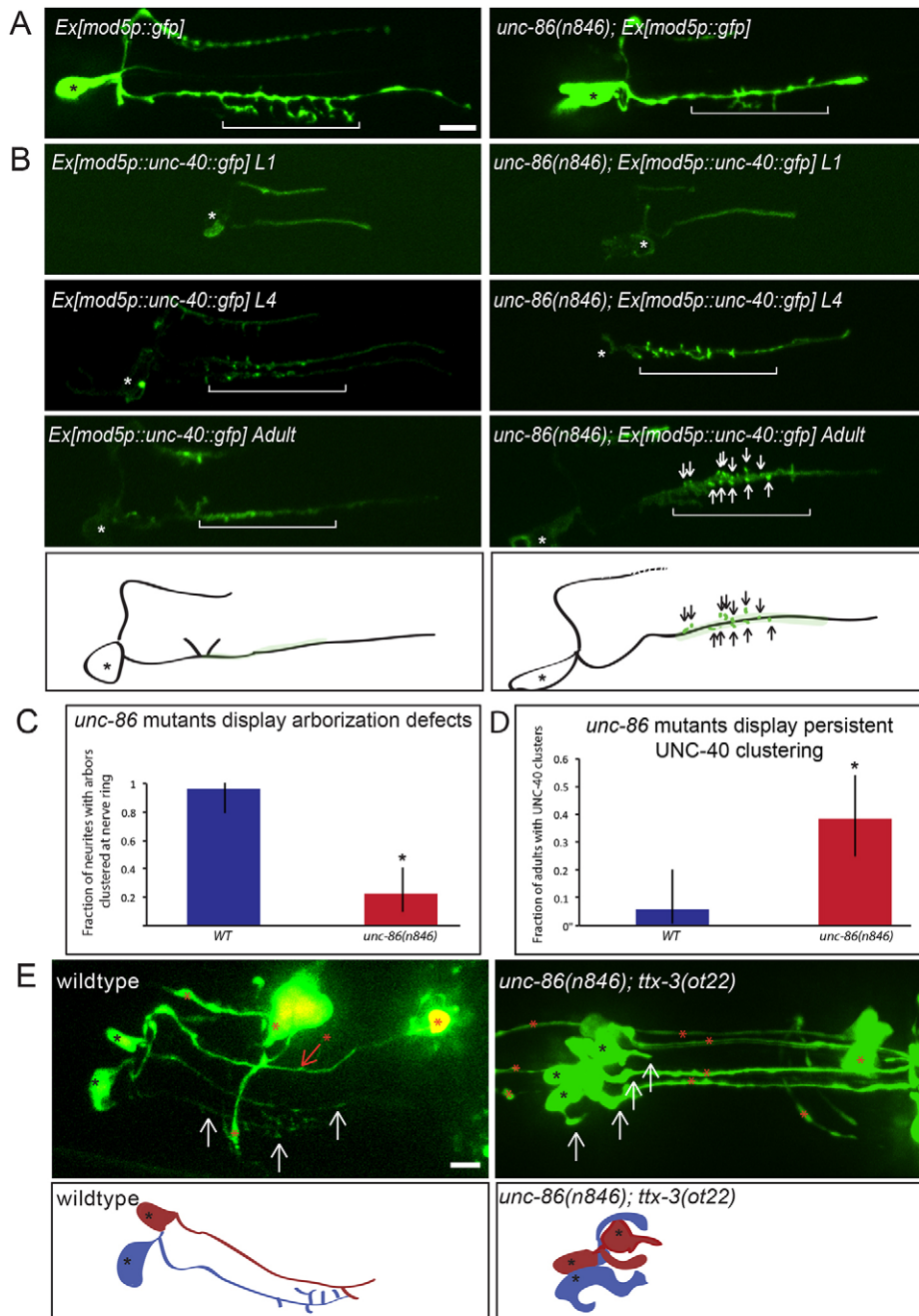


Fig. 7. *unc-86* and *ttx-3* affect NSM morphology. (A) *unc-86(n846)* mutant adults display shorter ventral neurites and fewer and shorter axon arbors. *mod-5p::gfp* (*olaEx1446*) is used to visualize NSM morphology. (B) UNC-40::GFP localization (transgene: *olaEx1448*) remains in a juvenile state in *unc-86* mutant animals. White arrows indicate UNC-40::GFP puncta. (C) Quantification of the *unc-86(n846)* arborization phenotype. Displayed is the fraction of animals with clusters of arbors in the nerve ring region in wild-type and *unc-86(n846)* animals. The difference between wild type and *unc-86* is significant ($*P < 0.0001$). (D) Quantification of the *unc-86(n846)* UNC-40::GFP localization phenotype. The fraction of animals with multiple, bright UNC-40::GFP puncta in the nerve ring region is displayed. The difference between wild type and *unc-86* is significant ($*P = 0.0017$). Error bars indicate 95% confidence intervals. (E) *unc-86(n846); ttx-3(ot22)* double mutants display numerous NSM morphology defects, as visualized with *flp-4::gfp* (*olaEx1485*). In all images, anterior is to the left and ventral down. Asterisks indicate cell bodies (A,B) or additional cell-body-like swellings (E). Brackets denote the nerve ring terminal field where arbors form. White arrows indicate NSM neurites, red arrows and asterisks denote other non-NSM structures. Fisher's *t*-test was used for statistical analysis. Scale bars: 5 μ m.

generated in *unc-86* mutants, as assessed by intact expression of the pan-neuronal marker *rab-3* and the pan-sensory marker *osm-6* (50 animals were scored for each marker).

unc-86 is expressed in two additional cholinergic neuron classes in the anterior ganglion besides the IL2 sensory neurons, namely the URA motoneurons [which are synaptically connected to the IL2 neurons (White et al., 1986)] and the URB interneurons. We found that cholinergic identity was also strongly affected in both URA and URB neurons of *unc-86(n846)* loss-of-function mutants (supplementary material Fig. S4).

***unc-86* cooperates with the ARID transcription factor *cfi-1* to control IL2 and URA identity**

Since none of the previously known co-factors of *unc-86* [*mec-3* for touch neurons (Duggan et al., 1998) and *ttx-3* for NSM neurons (this

paper)] is expressed in IL2, URA or URB neurons, *unc-86* is likely to act with another co-factor in IL2 neurons. *cfi-1* is an ARID transcription factor previously shown to be co-expressed with *unc-86* specifically in IL2 and URA neurons (Shaham and Bargmann, 2002). Loss of *cfi-1* results in ectopic expression of identity markers for the CEM neuron in IL2 and URA neurons (Shaham and Bargmann, 2002), which prompted us to investigate whether *cfi-1* might also positively control their cholinergic identity. We find that the cholinergic identity of both IL2 and URA neurons is affected in *cfi-1(ky651)* loss-of-function mutants, albeit not as strongly as in *unc-86* null mutants (Fig. 8A; supplementary material Fig. S4). To investigate whether *unc-86* and *cfi-1* genetically interact, we examined non-additive synergistic interactions of the two genes using a hypomorphic *unc-86* allele, *n848*. Animals carrying this allele show mild IL2 and URA differentiation defects, but in

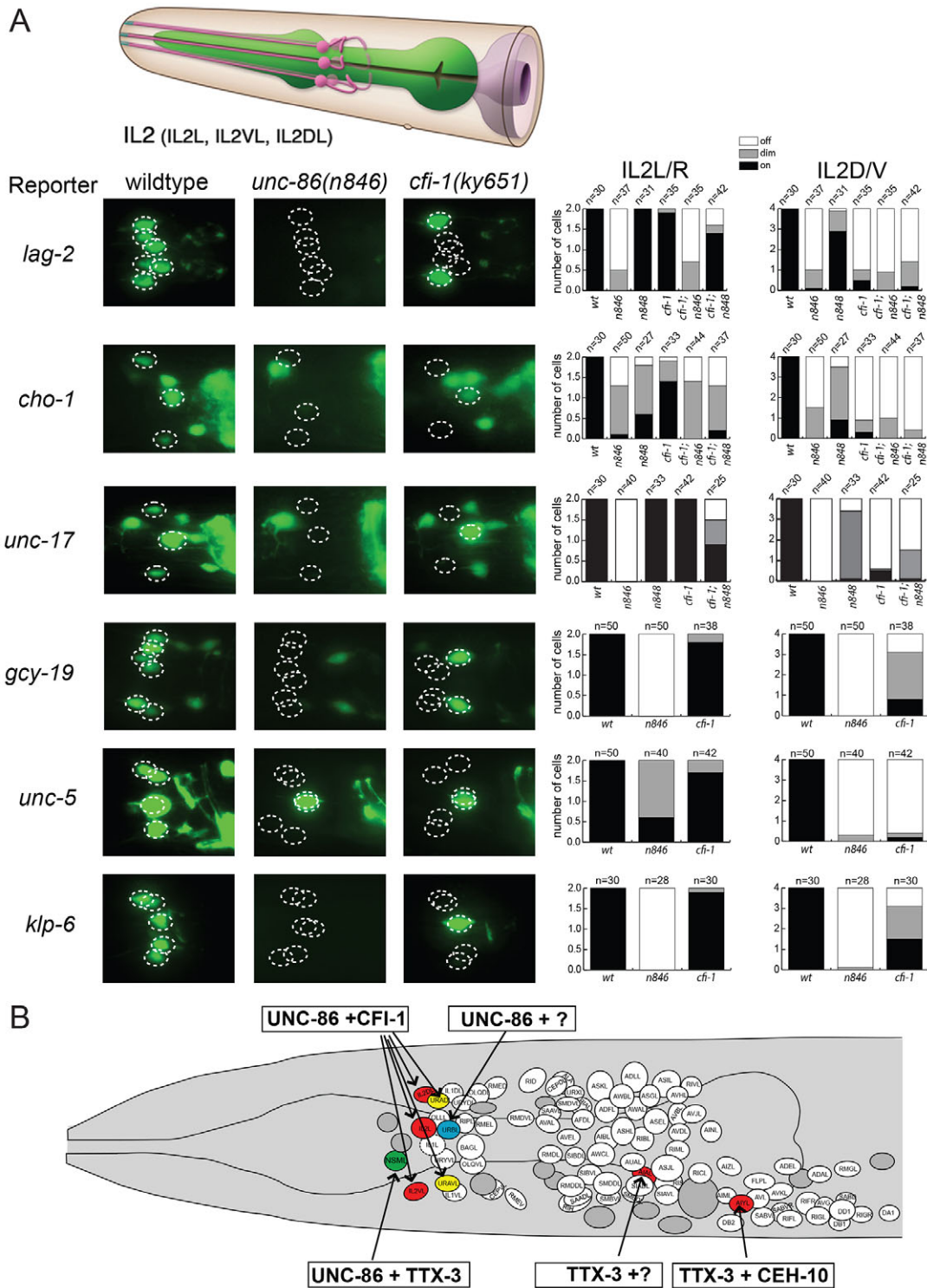


Fig. 8. *unc-86* and *cfi-1* control cholinergic IL2 neuron identity. (A) Animals are late L4 or young adults, with the exception of the *lag-2::gfp* transgenic animals which are dauers. The differential importance of *cfi-1* in the dorsal IL2DL/R and ventral IL2VL/R neurons versus the lateral IL2L/R neurons mirrors morphological differences of the ventral versus lateral neurons, with the lateral neurons having a distinct spectrum of synaptic partners (White et al., 1986). See also supplementary material Fig. S4. IL2 schematic reproduced with permission (Altun et al., 2002-2013). (B) Summary of terminal selector combinatorial codes in head ganglia of *C. elegans*. Colors refer to neurotransmitter identities: green, serotonergic; red, cholinergic; yellow, glutamatergic. Support or blast cells are in gray.

combination with the *cfi-1(ky651)* mutant allele there are strong synergistic, i.e. non-additive, defects in IL2 and URA differentiation (Fig. 8A; supplementary material Fig. S4). We conclude that *unc-86* and *cfi-1* cooperate to control IL2 and URA identity.

DISCUSSION

Two main conclusions can be drawn from the data presented in this paper. First, our data provide general support for the terminal selector concept. Second, our data show that a given transcription

factor can operate as a selector of terminal neuron identity in distinct neuronal cell types and that this is achieved through cooperation with distinct co-factors (summarized in Fig. 8B). In other words, individual neuronal cell types use distinct combinatorial codes of terminal selectors, and individual components of the code are reused in distinct combinations in different cell types.

The terminal selector concept was initially proposed based on a relatively small number of *C. elegans* transcription factor mutant phenotypes (Hobert, 2008). In each of these mutant backgrounds, a neuronal cell is born and expresses pan-neuronal features but fails to adopt neuron type-specific identity features. Importantly, terminal differentiation is very broadly affected in terminal selector mutants, such that not only functionally linked features (such as enzymes and transporter in a neurotransmitter synthesis/transport pathway), but also seemingly completely independent differentiation features that have no obvious biochemical connection (e.g. sensory receptors, neuropeptides and ionotropic neurotransmitter receptors) fail to be expressed. That the removal of an individual transcription factor results in such broad defects could not necessarily be assumed since transcriptomic approaches generally show that individual cell types expresses several dozen transcription factors (e.g. Etchberger et al., 2007). This could be interpreted to mean that the identity features of a neuron are regulated in a piecemeal manner, rather than being ‘mastered’ by a single transcription factor or a small combination thereof (Hobert, 2011). Two major questions raised by the terminal selector concept were how broadly it applies to different cell types in the *C. elegans* nervous system and how it applies to transcription factors expressed in distinct neuron types.

Here, we have shown that the terminal differentiation programs of very distinct neuron types – a cholinergic interneuron (AIA), a serotonergic sensory/motor neuron (NSM) and cholinergic sensory and motor neuron classes (IL2 and URA) – are controlled by distinct combinatorial codes of transcription factors. These factors regulate many distinct identity features of these distinct neuron types, ranging from neuropeptides to neurotransmitter synthesis pathway genes to neurotransmitter receptors and other signaling molecules.

In the case of the cholinergic AIA interneuron, we found that the expression of every tested terminal differentiation marker is affected in *ttx-3* mutants. Since the available AIA marker collection essentially represents a random snapshot of terminal markers that characterize AIA identity, one might extrapolate the regulatory impact of *ttx-3* on each one of these genes to the many hundreds, if not thousands, of genes that are expressed in AIAs, such that *ttx-3* is likely to affect a very large number of them. The estimated very broad effect of *ttx-3* on AIA identity is consistent with what we observed for the cholinergic AIY interneuron, in which *ttx-3* mutation also affects the expression of all known identity features (Altun-Gultekin et al., 2001; Wenick and Hobert, 2004). Even though both neuron types have similar morphologies, are cholinergic, and are directly postsynaptic to various sensory neurons, AIY and AIA have different functions (Hobert et al., 1997; Shinkai et al., 2011; Tomioka et al., 2006), connect to a different spectrum of synaptic partners (White et al., 1986) and express distinct gene batteries. Yet, in both cases, *ttx-3* very broadly affects the differentiation of each neuron type.

The distinct target gene specificities of *ttx-3* in AIA and AIY neurons can be explained by neuron type-specific co-factors and by *ttx-3* acting through distinct *cis*-regulatory motifs. AIY-expressed genes display a characteristic *cis*-regulatory signature that is recognized by a combination of the TTX-3 and CEH-10 homeodomain proteins (Wenick and Hobert, 2004). As we have shown here, AIA-expressed genes share a distinct *cis*-regulatory

signature that is composed of two separate motifs located in close proximity, one a TTX-3 binding site and the other a binding site for a presumptive TTX-3 co-factor. This is analogous to the situation in the AIY interneuron class, in which TTX-3 and CEH-10 operate through a bipartite motif (the ‘AIY motif’) composed of a TTX-3 and a CEH-10 binding site (Wenick and Hobert, 2004). Genes that are expressed in both AIY and AIA neurons (e.g. *cho-1*) contain a modular assembly of both the AIY and AIA *cis*-regulatory signature.

Similar to the *ttx-3*-dependent control of the central cholinergic interneurons AIY and AIA, the mouse LIM homeobox gene *Lhx7* is required for the terminal differentiation of cholinergic striatal interneurons (Lopes et al., 2012). As with other terminal selector transcription factors, *Lhx7* function appears to be continuously required to maintain cholinergic identity. Co-factors that operate together with *Lhx7* are currently not known. *Lhx7* is expressed in many other neurons in the CNS. It will be interesting to determine whether *Lhx7* also operates as a terminal selector in these other neuron types.

ttx-3 activity is not restricted to cholinergic neurons. We find that *ttx-3* is also a key regulator of serotonergic neuron identity. The activity of *ttx-3* in the serotonergic NSM neuron class is, however, distinct from that of AIA and AIY. Whereas the expression of several NSM-expressed effector genes is completely eliminated in *ttx-3* mutants, the expression of some effector genes is only partially affected or not affected at all. In cases in which only partial or no effect was observed, joint removal of another homeobox gene, *unc-86*, resulted in much stronger or complete loss of effector gene expression. Vice versa, the expression of effector genes that are unaffected in expression in *unc-86* mutants is lost in either *ttx-3* mutants or in the *ttx-3*; *unc-86* double mutant. Taken together, elimination of both of the POU/LIM homeobox genes *unc-86* and *ttx-3* has profound effects on NSM identity, paralleling the profound effect that another POU/LIM homeobox combination (*unc-86* and *mec-3*) has on touch neuron differentiation (Duggan et al., 1998). How *unc-86* and *ttx-3* interact to control NSM differentiation is currently unclear. Both genes are continuously expressed in NSM neurons, but do not regulate the expression of each other. Based on the synergistic nature of the effect of joint *ttx-3* and *unc-86* removal on the expression of some target genes (no or limited effect in single mutants, complete loss in double mutant), we propose that both transcription factors act jointly on common target gene promoters. For some target genes, the loss of one regulatory factor can be completely or partly compensated for by the other regulatory factor; in other cases, such compensation is not possible. *unc-86* and *ttx-3* might therefore not always act in a strict cooperative sense, but rather act independently on target gene promoters. There is already a notable precedent for such a mechanism, as we recently found that a combination of three different transcription factors controls dopaminergic neuron identity. For some target genes, individual transcription factor mutants display very limited effects, but double mutants strongly affect target gene expression (Doitsidou et al., 2013). In the case of NSM, we cannot however rule out the possibility that some genes are exclusively regulated by *unc-86* whereas others are exclusively regulated by *ttx-3*.

Apart from demonstrating *ttx-3* terminal selector function in distinct neuron types, we have also shown here that the POU homeobox gene *unc-86* can similarly act as a terminal selector in distinct neuron types. A role of *unc-86* in the differentiation of serotonergic and glutamatergic touch neurons has been described previously (Desai et al., 1988; Duggan et al., 1998; Sze et al., 2002; Serrano-Saiz et al., 2013). We show here that *unc-86* also controls the terminal differentiation programs of three distinct cholinergic

neuron types. Two of these cholinergic neuron types are synaptically connected and form a simple sensory-to-motor circuit (White et al., 1986). The role of *unc-86* in controlling cholinergic IL2 sensory neuron specification is reminiscent of, and might even be homologous to, the function of the POU homeobox gene *acj6* in controlling expression of the cholinergic gene locus in *Drosophila* olfactory neurons (Lee and Salvaterra, 2002). The ARID-type transcription factor *cfi-1* cooperates with *unc-86* to control the cholinergic identity of IL2 and URA neurons. Although neuronal differentiation functions have been reported for the *cfi-1* homolog *dead ringer* (*retained* – FlyBase) in *Drosophila* (Ditch et al., 2005), the functions of vertebrate orthologs (*Arid3* genes) in the nervous system remain to be explored.

MATERIALS AND METHODS

Strains and transgenes

For a list of strains and transgenes and notes on their generation see supplementary material Table S2.

Serotonin antibody staining

Young adult animals were fixed in 4% paraformaldehyde overnight and then treated with 5% β -mercaptoethanol overnight followed by 1000 units/ml collagenase (Sigma-Aldrich) treatment. Rabbit anti-serotonin whole serum (Sigma-Aldrich, S5545) was used at 1:100 dilution. Worms were then washed and incubated with Alexa Fluor 555 donkey anti-rabbit IgG (1:1000; Life Technologies, A-31571).

Cis-regulatory analysis

DNA sequences were subcloned into pPD95.75 expression vector (Addgene). For some smaller constructs, PCR products were directly amplified from subcloned constructs that have the same 3' end of the promoter sequences. DNAs for injection were PCR amplified to eliminate vector backbone, gel purified and then injected as complex arrays (10 ng/ μ l) with digested *rol-6(d)* (3 ng/ μ l) as injection marker, or plasmid mix was directly injected [50 ng/ μ l together with 100 ng/ μ l *rol-6(d)*].

Gel shift analysis

Full-length *unc-86* cDNA was cloned into the pET-21b His tag expression vector (EMD Millipore) and transformed into BL21(DE3) pLysS bacteria (Novagen). Protein expression was induced using 1 mM IPTG for 4 hours at 37°C and batch purified using Ni-NTA resin (Qiagen) under denaturing conditions as described (Wenick and Hobert, 2004). TTX-3 was purified and electrophoretic mobility shift assays (EMSAs) were performed as described (Wenick and Hobert, 2004). Probe sequences are listed in supplementary material Table S3.

Acknowledgements

We thank Q. Chen and B. Alarcon for expert assistance in generating transgenic strains, V. Reinke for providing the *ttx-3* fosmid reporter, E. Serrano for *eat-4* reporters, members of the worm community for providing reporter genes and members of the O.H. laboratory for comments on the manuscript.

Competing interests

The authors declare no competing financial interests.

Author contributions

F.Z. and O.H. initiated the study. A.B. and P.G. performed the analysis of the IL2 neurons; A.B. performed analysis of the URA and URB neurons; C.L.-F., M.M. and N.F. undertook the mutational analysis of the serotonergic pathway promoters; J.C.N. and D.A.C.-R. performed and supervised the morphological analysis of the NSM neurons; N.A. and R.S.M. performed and supervised the gel-shift analyses. F.Z. performed all other experiments. O.H. wrote the paper.

Funding

This work was funded by the National Institutes of Health [R01NS039996-05 and R01NS050266-03 to O.H.; R01NS076558 to D.A.C.-R.; and R01NS070644 to R.S.M.]; a March of Dimes Foundation Grant (to D.A.C.-R.); the Spanish Government [SAF2011-26273 to N.F.]; a Marie Curie Career Integration Grant (to

N.F.); a VAL i+d Fellowship from Generalitat Valenciana (to C.L.-F.); and a European Research Council Starting Grant (to N.F.). N.F. is a National Alliance for Research on Schizophrenia and Depression (NARSAD) Young Investigator. O.H. is an Investigator of the Howard Hughes Medical Institute. Deposited in PMC for release after 6 months.

Supplementary material

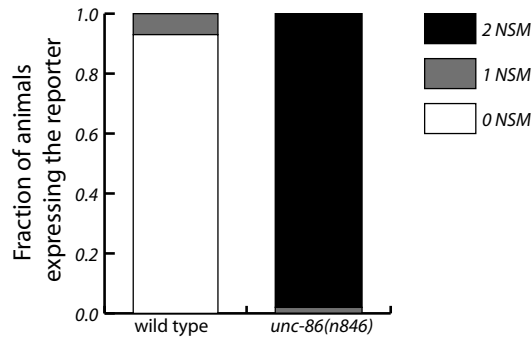
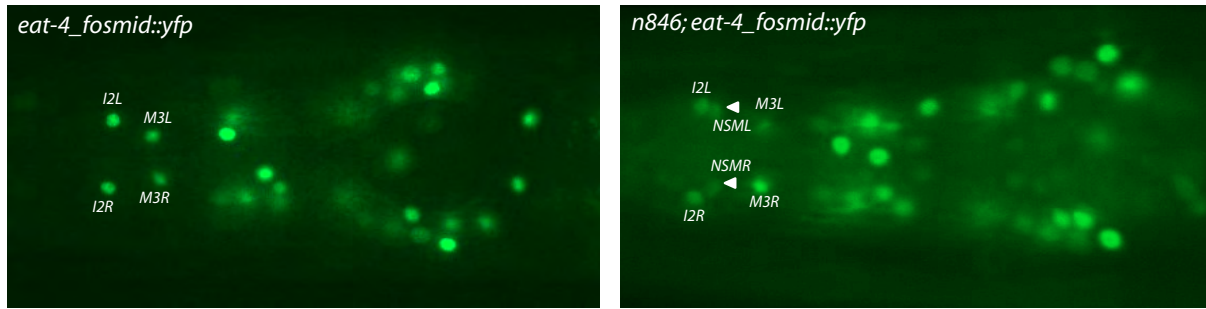
Supplementary material available online at <http://dev.biologists.org/lookup/suppl/doi:10.1242/dev.099721/-/DC1>

References

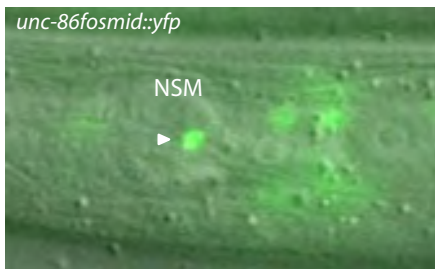
- Albertson, D. G. and Thomson, J. N. (1976). The pharynx of *Caenorhabditis elegans*. *Philos. Trans. R. Soc. B* **275**, 299-325.
- Altun, Z. F., Herndon, L. A., Crocker, C., Lints, R. and Hall, D. H. (eds) (2002-2013). WormAtlas, <http://www.wormatlas.org>.
- Altun-Gultekin, Z., Andachi, Y., Tsalik, E. L., Pilgrim, D., Kohara, Y. and Hobert, O. (2001). A regulatory cascade of three homeobox genes, *ceh-10*, *ttx-3* and *ceh-23*, controls cell fate specification of a defined interneuron class in *C. elegans*. *Development* **128**, 1951-1969.
- Aspöck, G., Ruvkun, G. and Bürglin, T. R. (2003). The *Caenorhabditis elegans* *ems* class homeobox gene *ceh-2* is required for M3 pharynx motoneuron function. *Development* **130**, 3369-3378.
- Axäng, C., Rauthan, M., Hall, D. H. and Pilon, M. (2008). Developmental genetics of the *C. elegans* pharyngeal neurons NSML and NSMR. *BMC Dev. Biol.* **8**, 38.
- Baumeister, R., Liu, Y. and Ruvkun, G. (1996). Lineage-specific regulators couple cell lineage asymmetry to the transcription of the *Caenorhabditis elegans* POU gene *unc-86* during neurogenesis. *Genes Dev.* **10**, 1395-1410.
- Berger, M. F., Badis, G., Gehrke, A. R., Talukder, S., Philippakis, A. A., Peña-Castillo, L., Alleyne, T. M., Mnaimneh, S., Botvinnik, O. B., Chan, E. T. et al. (2008). Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell* **133**, 1266-1276.
- Bertrand, V. and Hobert, O. (2009). Linking asymmetric cell division to the terminal differentiation program of postmitotic neurons in *C. elegans*. *Dev. Cell* **16**, 563-575.
- Clark, S. G. and Chiu, C. (2003). *C. elegans* ZAG-1, a Zn-finger-homeodomain protein, regulates axonal development and neuronal differentiation. *Development* **130**, 3781-3794.
- Deneris, E. S. and Wyler, S. C. (2012). Serotonergic transcriptional networks and potential importance to mental health. *Nat. Neurosci.* **15**, 519-527.
- Desai, C., Garriga, G., McIntire, S. L. and Horvitz, H. R. (1988). A genetic pathway for the development of the *Caenorhabditis elegans* HSN motor neurons. *Nature* **336**, 638-646.
- Ditch, L. M., Shirangi, T., Pitman, J. L., Latham, K. L., Finley, K. D., Edeen, P. T., Taylor, B. J. and McKeown, M. (2005). *Drosophila* *retained/dead ringer* is necessary for neuronal pathfinding, female receptivity and repression of fruitless independent male courtship behaviors. *Development* **132**, 155-164.
- Doitsidou, M., Flames, N., Topalidou, I., Abe, N., Felton, T., Remesal, L., Popovitchenko, T., Mann, R., Chalfie, M. and Hobert, O. (2013). A combinatorial regulatory signature controls terminal differentiation of the dopaminergic nervous system in *C. elegans*. *Genes Dev.* **27**, 1391-1405.
- Duggan, A., Ma, C. and Chalfie, M. (1998). Regulation of touch receptor differentiation by the *Caenorhabditis elegans* *mec-3* and *unc-86* genes. *Development* **125**, 4107-4119.
- Etchberger, J. F., Lorch, A., Sleumer, M. C., Zapf, R., Jones, S. J., Marra, M. A., Holt, R. A., Moerman, D. G. and Hobert, O. (2007). The molecular signature and cis-regulatory architecture of a *C. elegans* gustatory neuron. *Genes Dev.* **21**, 1653-1674.
- Finney, M. and Ruvkun, G. (1990). The *unc-86* gene product couples cell lineage and cell identity in *C. elegans*. *Cell* **63**, 895-905.
- Flames, N. and Hobert, O. (2009). Gene regulatory logic of dopamine neuron differentiation. *Nature* **458**, 885-889.
- Gray, P. A., Fu, H., Luo, P., Zhao, Q., Yu, J., Ferrari, A., Tenzen, T., Yuk, D. I., Tsung, E. F., Cai, Z. et al. (2004). Mouse brain organization revealed through direct genome-scale TF expression analysis. *Science* **306**, 2255-2257.
- Greer, E. R., Perez, C. L., Van Gilst, M. R., Lee, B. H. and Ashrafi, K. (2008). Neural and molecular dissection of a *C. elegans* sensory circuit that regulates fat and feeding. *Cell Metab.* **8**, 118-131.
- Harris, G., Korchnak, A., Summers, P., Hapiak, V., Law, W. J., Stein, A. M., Komuniecki, P. and Komuniecki, R. (2011). Dissecting the serotonergic food signal stimulating sensory-mediated aversive behavior in *C. elegans*. *PLoS ONE* **6**, e21897.
- Hirota, J. and Mombaerts, P. (2004). The LIM-homeodomain protein *Lhx2* is required for complete development of mouse olfactory sensory neurons. *Proc. Natl. Acad. Sci. USA* **101**, 8751-8755.
- Hobert, O. (2008). Regulatory logic of neuronal diversity: terminal selector genes and selector motifs. *Proc. Natl. Acad. Sci. USA* **105**, 20067-20071.
- Hobert, O. (2011). Regulation of terminal differentiation programs in the nervous system. *Annu. Rev. Cell Dev. Biol.* **27**, 681-696.
- Hobert, O. and Ruvkun, G. (1998). A common theme for LIM homeobox gene function across phylogeny? *Biol. Bull.* **195**, 377-380.
- Hobert, O. and Westphal, H. (2000). Functions of LIM-homeobox genes. *Trends Genet.* **16**, 75-83.

- Hobert, O., Mori, I., Yamashita, Y., Honda, H., Ohshima, Y., Liu, Y. and Ruvkun, G. (1997). Regulation of interneuron function in the *C. elegans* thermoregulatory pathway by the *txx-3* LIM homeobox gene. *Neuron* **19**, 345-357.
- Horvitz, H. R., Chalfie, M., Trent, C., Sulston, J. E. and Evans, P. D. (1982). Serotonin and octopamine in the nematode *Caenorhabditis elegans*. *Science* **216**, 1012-1014.
- Jafari, G., Xie, Y., Kullyev, A., Liang, B. and Sze, J. Y. (2011). Regulation of extrasynaptic 5-HT by serotonin reuptake transporter function in 5-HT-absorbing neurons underscores adaptation behavior in *Caenorhabditis elegans*. *J. Neurosci.* **31**, 8948-8957.
- Kolterud, A., Alenius, M., Carlsson, L. and Bohm, S. (2004). The Lim homeobox gene *Lhx2* is required for olfactory sensory neuron identity. *Development* **131**, 5319-5326.
- Lee, M. H. and Salvaterra, P. M. (2002). Abnormal chemosensory jump 6 is a positive transcriptional regulator of the cholinergic gene locus in *Drosophila* olfactory neurons. *J. Neurosci.* **22**, 5291-5299.
- Lee, R. Y., Sawin, E. R., Chalfie, M., Horvitz, H. R. and Avery, L. (1999). EAT-4, a homolog of a mammalian sodium-dependent inorganic phosphate cotransporter, is necessary for glutamatergic neurotransmission in *Caenorhabditis elegans*. *J. Neurosci.* **19**, 159-167.
- Lee, H., Choi, M. K., Lee, D., Kim, H. S., Hwang, H., Kim, H., Park, S., Paik, Y. K. and Lee, J. (2012). Nictation, a dispersal behavior of the nematode *Caenorhabditis elegans*, is regulated by IL2 neurons. *Nat. Neurosci.* **15**, 107-112.
- 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. (2007). Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168-176.
- Leung-Hagesteijn, C., Spence, A. M., Stern, B. D., Zhou, Y., Su, M. W., Hedgecock, E. M. and Culotti, J. G. (1992). UNC-5, a transmembrane protein with immunoglobulin and thrombospondin type 1 domains, guides cell and pioneer axon migrations in *C. elegans*. *Cell* **71**, 289-299.
- Liu, C., Maejima, T., Wylter, S. C., Casadesus, G., Herlitz, S. and Deneris, E. S. (2010). Pet-1 is required across different stages of life to regulate serotonergic function. *Nat. Neurosci.* **13**, 1190-1198.
- Lopes, R., Verhey van Wijk, N., Neves, G. and Pachnis, V. (2012). Transcription factor LIM homeobox 7 (*Lhx7*) maintains subtype identity of cholinergic interneurons in the mammalian striatum. *Proc. Natl. Acad. Sci. USA* **109**, 3119-3124.
- Mangale, V. S., Hirokawa, K. E., Satyaki, P. R., Gokulchandran, N., Chikbire, S., Subramanian, L., Shetty, A. S., Martynoga, B., Paul, J., Mai, M. V. et al. (2008). *Lhx2* selector activity specifies cortical identity and suppresses hippocampal organizer fate. *Science* **319**, 304-309.
- Moreno, N., Bachy, I., Rétaux, S. and González, A. (2005). LIM-homeodomain genes as territory markers in the brainstem of adult and developing *Xenopus laevis*. *J. Comp. Neurol.* **485**, 240-254.
- Nathoo, A. N., Moeller, R. A., Westlund, B. A. and Hart, A. C. (2001). Identification of neuropeptide-like protein gene families in *Caenorhabditis elegans* and other species. *Proc. Natl. Acad. Sci. USA* **98**, 14000-14005.
- Nelson, J. C. and Colón-Ramos, D. A. (2013). Serotonergic neurosecretory synapse targeting is controlled by netrin-releasing guidepost neurons in *Caenorhabditis elegans*. *J. Neurosci.* **33**, 1366-1376.
- Ortiz, C. O., Etchberger, J. F., Posy, S. L., Frøkjær-Jensen, C., Lockery, S., Honig, B. and Hobert, O. (2006). Searching for neuronal left/right asymmetry: genome-wide analysis of nematode receptor-type guanylyl cyclases. *Genetics* **173**, 131-149.
- Ouellet, J., Li, S. and Roy, R. (2008). Notch signalling is required for both dauer maintenance and recovery in *C. elegans*. *Development* **135**, 2583-2592.
- Peden, E. M. and Barr, M. M. (2005). The KLP-6 kinesin is required for male mating behaviors and polycystin localization in *Caenorhabditis elegans*. *Curr. Biol.* **15**, 394-404.
- 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.
- Ranganathan, R., Sawin, E. R., Trent, C. and Horvitz, H. R. (2001). Mutations in the *Caenorhabditis elegans* serotonin reuptake transporter MOD-5 reveal serotonin-independent and -independent activities of fluoxetine. *J. Neurosci.* **21**, 5871-5884.
- Rhee, J. M., Gruber, C. A., Brodie, T. B., Trieu, M. and Turner, E. E. (1998). Highly cooperative homodimerization is a conserved property of neural POU proteins. *J. Biol. Chem.* **273**, 34196-34205.
- Röhrig, S. (2000). *Modulation of UNC-86 Activity During Caenorhabditis elegans Neurogenesis*. München, Germany: Herbert Utz Verlag.
- Serrano-Saiz, E., Poole, R. J., Felton, T., Zhang, F., De La Cruz, E. and Hobert, O. (2013). Modular control of glutamatergic neuronal identity in *C. elegans* by distinct homeodomain proteins. *Cell* **155**, 659-673.
- Shaham, S. and Bargmann, C. I. (2002). Control of neuronal subtype identity by the *C. elegans* ARID protein CFI-1. *Genes Dev.* **16**, 972-983.
- Shinkai, Y., Yamamoto, Y., Fujiwara, M., Tabata, T., Murayama, T., Hirotsu, T., Ikeda, D. D., Tsunozaki, M., Iino, Y., Bargmann, C. I. et al. (2011). Behavioral choice between conflicting alternatives is regulated by a receptor guanylyl cyclase, GCY-28, and a receptor tyrosine kinase, SCD-2, in AIA interneurons of *Caenorhabditis elegans*. *J. Neurosci.* **31**, 3007-3015.
- Simmons, D. K., Pang, K. and Martindale, M. Q. (2012). Lim homeobox genes in the Ctenophore *Mnemiopsis leidyi*: the evolution of neural cell type specification. *EvoDevo* **3**, 2.
- Smidt, M. P. and Burbach, J. P. (2009). Terminal differentiation of mesodiencephalic dopaminergic neurons: the role of *Nurr1* and *Pitx3*. *Adv. Exp. Med. Biol.* **651**, 47-57.
- Srivastava, M., Larroux, C., Lu, D. R., Mohanty, K., Chapman, J., Degnan, B. M. and Rokhsar, D. S. (2010). Early evolution of the LIM homeobox gene family. *BMC Biol.* **8**, 4.
- Sulston, J. E., Schierenberg, E., White, J. G. and Thomson, J. N. (1983). The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev. Biol.* **100**, 64-119.
- Sze, J. Y., Zhang, S., Li, J. and Ruvkun, G. (2002). The *C. elegans* POU-domain transcription factor UNC-86 regulates the *tph-1* tryptophan hydroxylase gene and neurite outgrowth in specific serotonergic neurons. *Development* **129**, 3901-3911.
- Tomioka, M., Adachi, T., Suzuki, H., Kunitomo, H., Schafer, W. R. and Iino, Y. (2006). The insulin/PI 3-kinase pathway regulates salt chemotaxis learning in *Caenorhabditis elegans*. *Neuron* **51**, 613-625.
- Tong, Y. G. and Bürglin, T. R. (2010). Conditions for dye-filling of sensory neurons in *Caenorhabditis elegans*. *J. Neurosci. Methods* **188**, 58-61.
- Treinin, M., Gillo, B., Liebman, L. and Chalfie, M. (1998). Two functionally dependent acetylcholine subunits are encoded in a single *Caenorhabditis elegans* operon. *Proc. Natl. Acad. Sci. USA* **95**, 15492-15495.
- Tursun, B., Cochella, L., Carrera, I. and Hobert, O. (2009). A toolkit and robust pipeline for the generation of fosmid-based reporter genes in *C. elegans*. *PLoS ONE* **4**, e4625.
- Wenick, A. S. and Hobert, O. (2004). Genomic cis-regulatory architecture and trans-acting regulators of a single interneuron-specific gene battery in *C. elegans*. *Dev. Cell* **6**, 757-770.
- White, J. G., Southgate, E., Thomson, J. N. and Brenner, S. (1986). The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Philos. Trans. R. Soc. B* **314**, 1-340.
- Zetterström, R. H., Williams, R., Perlmann, T. and Olson, L. (1996). Cellular expression of the immediate early transcription factors *Nurr1* and *NGFI-B* suggests a gene regulatory role in several brain regions including the nigrostriatal dopamine system. *Brain Res. Mol. Brain Res.* **41**, 111-120.

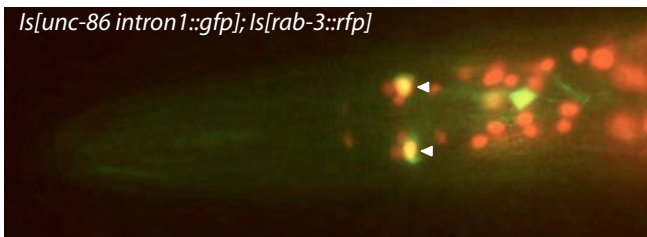
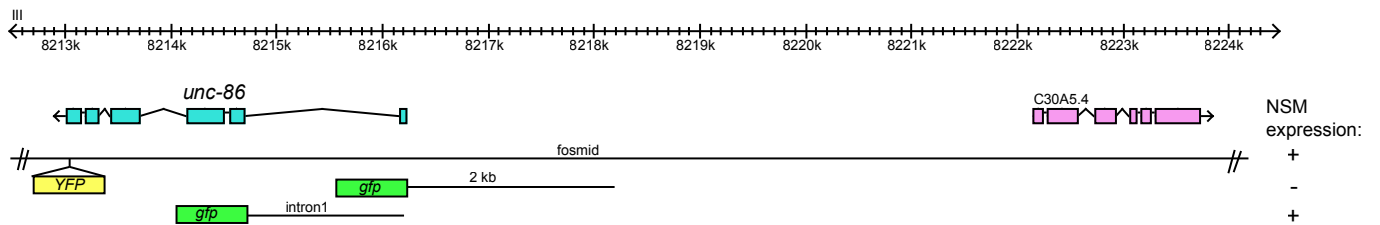
A



B



C



	% NSM on	n
<i>ls[unc-86 intron1::gfp]</i>	93%	28
<i>n846; ls[unc-86 intron1::gfp]</i>	0%	56

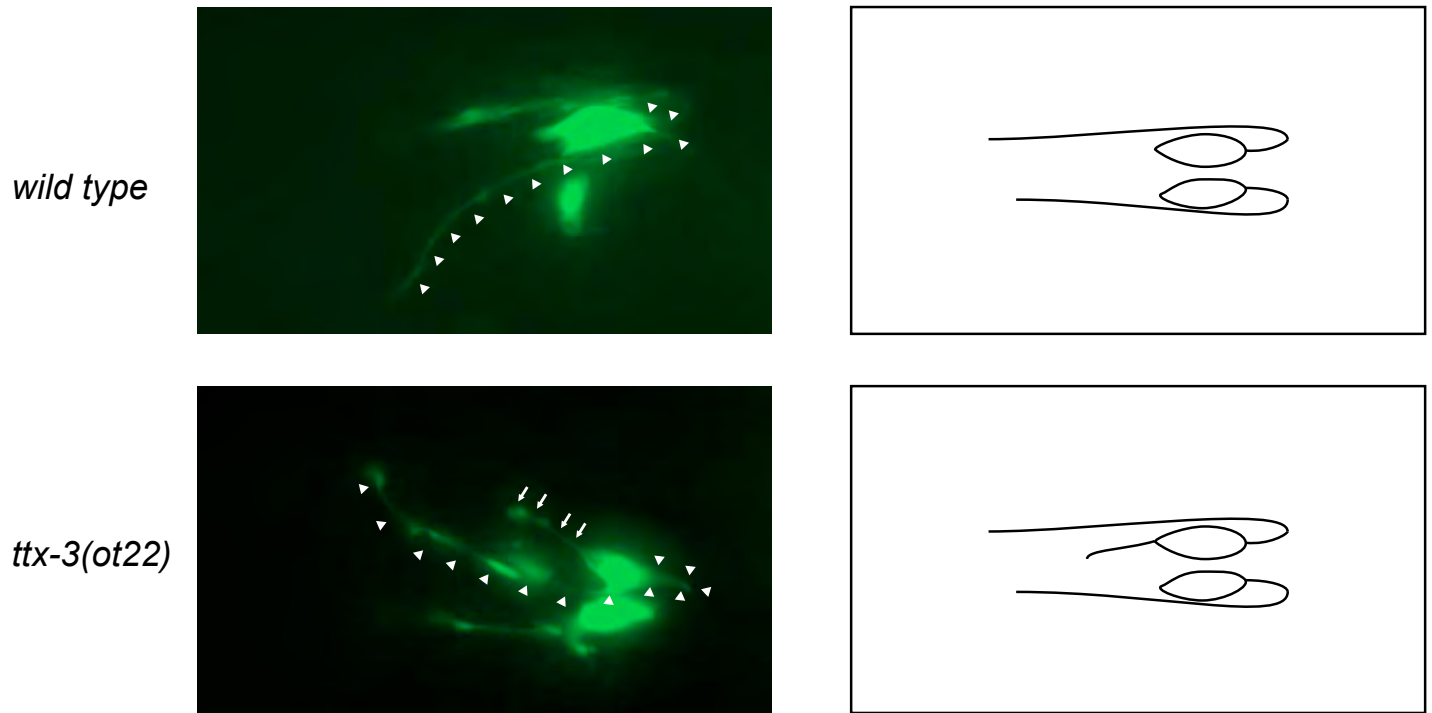
Suppl. Figure 1: Analysis of the NSM neurons.

A: A fosmid reporter of the *eat-4* locus was kindly provided by E. Serrano and will be published elsewhere.

B: A *unc-86* fosmid reporter construct is expressed in the NSM neurons of adult animals.

C: A *cis*-regulatory element from the *unc-86* locus drives expression in NSM and this expression depends on *unc-86*.

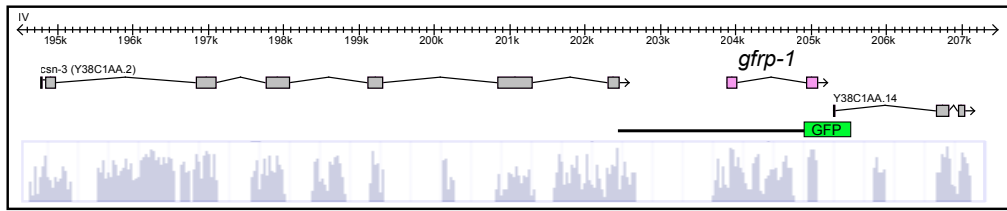
AIA neuron morphology



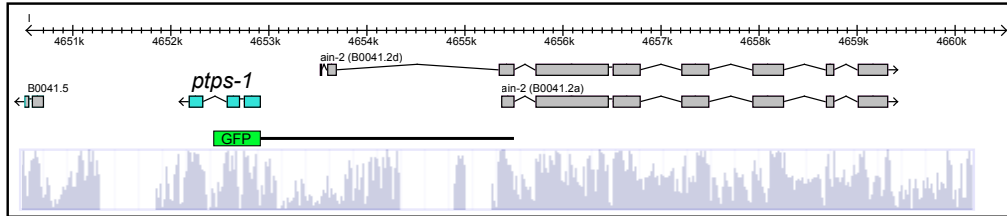
Suppl. Figure 2: AIA morphology in wildtype and *ttx-3(ot22)* mutant animals.

AIA morphology was visualized with *ins-1::gfp (otIs326)*. Left panels show AIA morphology schematically. White triangles in the *gfp* images indicate one of the main axons (normal axon). White arrows indicate the ectopic branching from the cell body in *ttx-3* mutants. Note also the blebbing of the main axon.

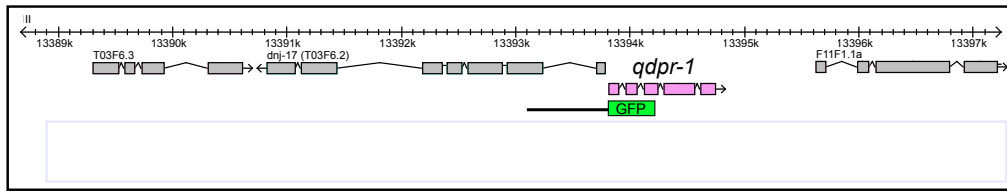
Expression pattern:



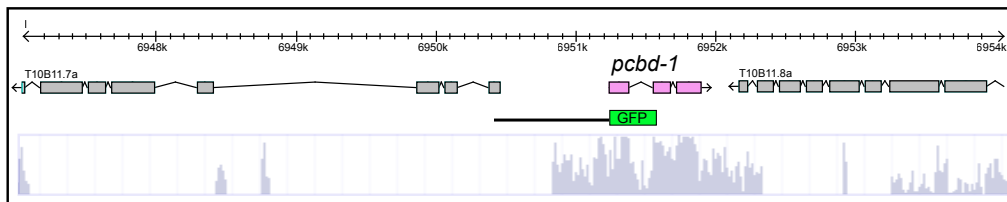
not NSM,
not HSN
a few very dim cells in the head



NSM
HSN
VC4 and VC5
no ADF
more VNC neurons
tail neuron



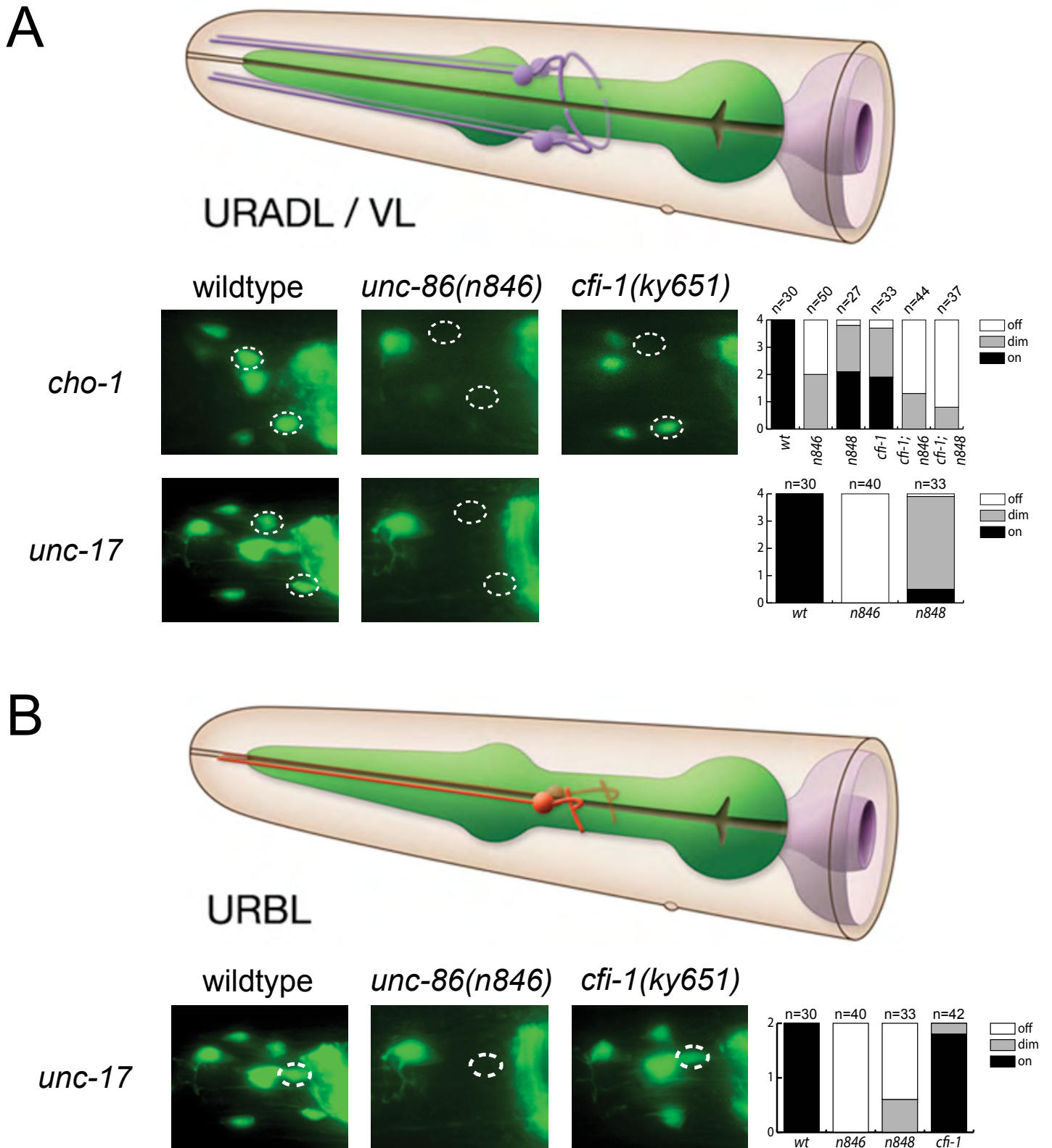
not NSM,
not HSN
several dim cells in the head (ADF?)
tail and head hypodermal cells



not NSM,
not HSN
not ADF
a couple very faint cells in head
no obvious neuronal expression
tail hypodermal cells
pharyngeal muscle expression

Suppl. Figure 3: BH4 pathway reporters.

Reporter genes and overview of expression pattern of BH4 pathway genes. Expression patterns were observed with multiple lines.



Suppl. Figure 4: *unc-86* controls the identity of the cholinergic URA and URB neurons.

A: *unc-86* and *cfi-1* affect URA identity. Lateral views (anterior to left) are shown. Bar graphs indicate average number of cells expressing *gfp* in the four URA neurons.

B: *unc-86*, but not *cfi-1* affects URB identity as assessed with two different *unc-86* allele, *n846* and *n848*. Lateral views (anterior to left) are shown. Bar graphs indicate average number of cells expressing *gfp* in the two URB neurons.

Table S1. Rescue of NSM and AIA differentiation defects of *ttx-3* mutant animals

	% animals expressing <i>mgl-1::mcherry</i> in NSM	% animals expressing <i>mgl-1::mcherry</i> in AIA	<i>n</i>
wild type	100	100	>100
<i>ttx-3(ot22)</i>	0	0	>100
<i>ttx-3(ot22); Ex[cat-1^{prom}::ttx-3cDNA] line 1</i>	92	0	36
<i>ttx-3(ot22); Ex[cat-1^{prom}::ttx-3cDNA] line 2</i>	89	0	37
<i>ttx-3(ot22); Ex[ins-1^{prom}::ttx-3cDNA] line 1</i>	0	79	38
<i>ttx-3(ot22); Ex[ins-1^{prom}::ttx-3cDNA] line 2</i>	0	58	36
<i>ttx-3(ot22); Ex[ins-1^{prom}::ttx-3cDNA] line 3</i>	0	79	38

Injection marker: *rol-6(d)*. Note that the *ins-1* promoter is still weakly expressed in *ttx-3* mutants and can hence be used to drive *ttx-3* in *ttx-3* mutants.

Table S2. Strains and transgenes used in this study

Strain/Array	Notes
<i>ttx-3(ot22)</i>	Premature stop before the homeobox (Altun-Gultekin et al., 2001).
<i>unc-86(n846)</i>	A likely null allele resulting in protein loss (Röhrig, 2000). The molecular identity of this strong allele had not been previously described. We sequenced this alleles and it to harbor a G>A splice acceptor site mutation at end of the second intron of the C30H5.7a transcript in the middle of the POU domain (aataactcagGCGG to aataactcaaGCGG)
<i>unc-86(n848)</i>	The molecular nature of this temperature-sensitive allele is a GT to AT splice donor site mutation in intron 4 (Röhrig, 2000)
<i>cfi-1(ky651)</i>	A splice acceptor site mutation before the DNA binding domain (Shaham and Bargmann, 2002)
<i>otls224</i>	<i>Is[cat-1::gfp]</i> (Flames and Hobert, 2009)
<i>otls225</i>	<i>Is[cat-4::gfp]</i> (Flames and Hobert, 2009)
<i>otls226</i>	<i>Is[bas-1::gfp]</i> (Flames and Hobert, 2009)
<i>zcls13</i>	<i>Is[tph-1::gfp]</i> (Clark and Chiu, 2003)
<i>wgls68</i>	<i>Is[ttx-3fosmid::EGFP-FLAG, unc-119(+)]</i> . Kindly provided by Valerie Reinke and the ModEncode consortium. Based on fosmid WRM064cD04. The tag was TY1 EGFP 3xFLAG and was added at the C-terminus.
<i>otls337</i>	<i>Is[unc-86 fosmid^{WRM0612cF07}::NLS::YFP::H2B; ttx-3::mCherry]</i> . The <i>unc-86</i> fosmid reporter was generated bacterial recombineered as previously described (Tursun et al., 2009), fusing an SL2::NLS::YFP::H2B reporter cassette at the C-terminus of <i>unc-86</i> in fosmid WRM0612cF07.
<i>vsls33</i>	<i>Is[dop-3::dsRed]</i> . Kindly provided by Michael Koelle
<i>otls317</i>	<i>Is[mgl-1^{long prom}::mcherry, pha-1]</i> . DNA kindly provided by Kaveh Ashrafi (Greer et al., 2008)
<i>otls341</i>	<i>Is[mgl-1^{short prom}::gfp]</i> -1994 to -1374 bp upstream of ATG
<i>otls379</i>	<i>Is[cho-1^{AlA prom}::gfp; rol-6(d)]</i> -3006 to -2642 bp upstream of ATG
<i>otls326</i>	<i>Is[ins-1::gfp; rol-6(d)]</i> -289 bp upstream of ATG
<i>otEx4687</i>	<i>Ex[glr-2::gfp; rol-6(d)]</i> -1798 bp upstream of ATG
<i>otEx4886</i>	<i>Ex[ttx-3 intron7::gfp; rol-6(d)]</i> GGAAG+intron7+CGTCTACCGATGAAGATG cloned into pPD95.75
<i>otEx5056</i>	<i>Ex[flp-2::gfp; rol-6(d)]</i> -2002 bp upstream of ATG
<i>otEx4781</i>	<i>Ex[mod-5^{NSM prom}::gfp; elt-2::gfp]</i> . First intron of mod-5 cloned into pPD95.75(CACCAGCAGCTGCAAG+ intron1+ CTGAACTCTCC driving GFP)
<i>otEx5280</i>	<i>Ex[ptps-1::gfp; rol-6(d)]</i> -2600 bp upstream of ATG
<i>otEx5163</i>	<i>Ex[nlp-3::gfp; rol-6(d)]</i> . DNA kindly provided by Hart lab (Nathoo et al., 2001)
<i>otEx5364</i>	<i>Ex[mgl-3::gfp; rol-6(d)]</i> .

	DNA kindly provided by Ashrafi lab (Greer et al., 2008)
<i>otEx5163</i>	<i>[nlp-13::gfp; rol-6(d)]</i> -1967 bp upstream of ATG
<i>otEx5055</i>	<i>[scd-2::gfp; rol-6(d)]</i> -2045 bp upstream of ATG
<i>otEx5363</i>	<i>[flr-2::gfp; rol-6(d)]</i> DNA kindly provided by Takeshi Ishihara
<i>otEx4917</i>	<i>Ex[unc-86 intron1::gfp; rol-6(d)]</i> GACGACAACCGCTTCAAAAATGCAACCT+intron1+TTCAACAAC AGTTTATTTGGATCATTTCGATGACCC cloned into pPD95.75
<i>otEx4969,</i> <i>otEx4970</i>	2 independent lines of <i>Ex[cat-1^{prom14}::ttx-3; rol-6(d)]</i>
<i>otEx5073,</i> <i>otEx5074,</i> <i>otEx5075</i>	3 independent lines of <i>Ex[ins-1^{457bp_prom}::ttx-3; rol-6(d)]</i> (-457 bp upstream of ATG)
<i>Ex[gcy-28.d::gfp]</i>	the complete genotype of this array is <i>Ex[gcy-28.dp::gcy-28.d::GFP, AIA-specific ins-1p::SNB-1::mRFP, rol-6(+)]</i> (transgene kindly provided by Takeshi Ishihara)
<i>Ex[scd-2::gfp]</i>	<i>Ex[scd-2::gfp]</i> : the complete genotype of this array is <i>Ex[scd-2p::scd-2::GFP, AIA-specific ins-1p::mRFP, lin-44::gfp]</i> (transgene kindly provided by Takeshi Ishihara)
<i>olaEx1446</i>	<i>Ex [mod-5p::egfp (2ng/ul)/unc-122p::gfp (20ng/ul)]</i>
<i>olaEx1485</i>	<i>Ex[flp-4p::egfp (30ng/ul)/unc-122p::DSRED (20ng/ul)]</i>
<i>nuls9</i>	<i>Is[unc-5::gfp]</i> (transgene kindly provided by Josh Kaplan)
<i>otEx2310</i>	<i>Ex[gcy-19::gfp; unc-122::gfp]</i> (Ortiz et al., 2006)
<i>lqls3</i>	<i>Is[osm-6::gfp]</i> (transgene kindly provided by Erik Lundquist)
<i>qls56</i>	<i>Is[lag-2::gfp]</i> (transgene kindly provided by Judith Kimble)
<i>vsIs48</i>	<i>Is[unc-17::gfp]</i> (transgene kindly provided by Michael Koelle)
<i>otIs323</i>	<i>Is[cho-1_fosmid::gfp; elt-2::dsRed]</i> (transgene kindly provided by Paschalis Kratsios)

Table S3. Probe sequences for gel shift analysis

Probe	Sequence
cho-1 wt:	5'tacacacacatcgaaatatgaatcttctcttaaaaagaagggtgtccaattagttcccctattcaGCTTTCGTTTCGTCGCCT
cho-1 TAAT del	5'tacacacacatcgaaatatgaatcttctcttaaaaagaagggtgtccagttcccctattcaGCTTTCGTTTCGTCGCCT
mgl-1 wt	5'gttccatactcatagtgctcattagaatagcacggatcgtgttcgcctctcgccttgtaaccgaa tctgccGCTTTCGTTTCGTCGCCT
mgl-1 TAAT del	5'gttccatactcatagtgctcgaatagcacggatcgtgttcgcctctcgccttgtaaccgaa tctgccGCTTTCGTTTCGTCGCCT
bas-1 wt	5'cccaacaccacattattcatgtatttcctccaaaccactgaaccatctcattctcaaaccagtttct atccgtttgttgcattcaattaaat GCTTTCGTTTCGTCGCCT
bas-1 HD mut	5'cccaacaccacgttattcatgtatttcctccaaaccactgaaccatctcattctcaaaccagtttct atccgtttgttgcattcagttgaat GCTTTCGTTTCGTCGCCT
bas-1 POU mut	5'cccaacaccacattattcccgtatttcctccaaaccactgaaccatctcattctcaaaccagtttct atccgtttgttgccctcaattaaat GCTTTCGTTTCGTCGCCT
tph-1 wt	5'tctttgttgCGcataataaaacaatcaatcaacacagcaaagaccctctcaacctcattcatg attttcttGCTTTCGTTTCGTCGCCT
tph-1 HD mut	5'tctttgttgCGcatagtaaaacaatcaatcaacacagcaaagaccctctcaacctcattcatg attttcttGCTTTCGTTTCGTCGCCT
tph-1 POU mut	5'tctttgttgTgTataCCaCaacaaGcGatcaacacagcaaagaccctctcaacctcatt cCCgatttcttGCTTTCGTTTCGTCGCCT

CHAPTER 3:

Cloning and characterization of genes required for the specification of the RMDD and RMDV motor neurons

This chapter describes the cloning and characterization of three genes required for the specification of the RMD motor neurons. These mutants were initially isolated from a screen looking for mutants that are defective in the AIA interneurons with the transgene also expressed in the RMDD and RMDV neurons. Conventional genetic approaches and whole genome sequencing were both employed to identify the mutations. Three genes were identified: the vertebrate neuroD homolog *cnd-1*, the Beta3 and Olig family related gene *hlh-16* and the Q50 class paired-like homeobox gene *unc-42*. Fosmid reporters of these factors suggest that *cnd-1* and *hlh-16* likely play proneural roles, while *unc-42* may act as a terminal selector for the RMD neuron class. Many questions are left open, and further characterization of these genes will shed more light on the fate specification of the RMD motor neurons.

Results

A genetic screen that identified RMDD/V mutants

The metabotropic glutamate receptor *mgl-1* is expressed in eight neurons: NSML/R, AIAL/R, RMDDL/R and RMDVL/R (Greer et al., 2008). The transgene *otIs341 (mgl-1::gfp)* was initially used in a non-clonal screen to search for mutants that are defective in AIA expression because we speculate that there is another transcription factor that cooperates with TTX-3 to specify the AIA cell fate (See Chapter 2). The transgene *vsIs33 (dop-3::dsRed)* was used as a background reference for future automatic screens using the worm sorter (Doitsidou et al., 2008). In short, synchronized P0s at late L4 stage were EMS mutagenized and egg prepped. F1 worms grown on 10 mm plates at a density of about 10000/plate were then egg prepped and 10 plates of F2 worms plated at a density of 10000/plate were collected (Density can be increased depending on the abundance of OP50.) F2s are then screened manually under a fluorescent dissecting scope (or they could be passed through the worm sorter given proper reference transgene is present in the background).

Unexpectedly, five additional mutants that are deficient in RMDD/RMDV expression were identified. AIA and NSM fates remain intact in these mutants. The RMDs are ring motor neurons that utilize the neurotransmitter acetylcholine (Ach) (Duerr et al., 2008). There are 6 cells in total in the RMD neuron class: RMDD, RMD and RMDV. They have been shown to mediate head withdrawal responses to touch along the side of the nose, as well as foraging behaviors (the head of the worm moves in

a rhythmic motion). Both behaviors depend on the AMPA-type ionotropic glutamate receptor *glr-1* (Hart et al., 1995).

Mutants are then assigned to two categories (class I and II) according to their phenotypes. *ot705* and *ot712* are class I mutants; *ot704* and *ot711* (*mut5F* and *mut 5G* turned out to carry the same mutation) are class II mutants. All class I mutants lose *mgl-1* expression in RMDD and RMDV completely (Figure 1A, C), while class II mutants display a mix of phenotypes, with none, one or both RMDVs affected. RMDDL/R are always absent in class II mutants (Figure 1B, C).

Genetic analysis maps these five mutants to three different gene loci

Complementation tests assigned the five mutants to three complementation groups. *Mut 5F*, *mut 5G* and *ot704* failed to complement each other, while *ot705* and *ot712* belong to separate complementation groups. (See Table 1 for quantification of mutant phenotypes.)

Aside from their defects in RMDD/V expression, *ot705* and *ot712* display severe *unc* phenotype. Moreover, *ot712* does not seem to be separable from the transgene *vsIs33* that has been mapped to chromosome V. The Q50 class paired-like homeobox gene *unc-42* has been reported to affect gene expression in the RMDs. The expression of the ionotropic glutamate receptor GLR-1, GLR-4 and GLR-5 is disrupted in the six RMDs in *unc-42* mutants (Baran et al., 1999; Brockie et al., 2001). Complementation test between *ot712* and two alleles of *unc-42* (*e419* and *e270*) confirmed that *ot712* is an allele of *unc-42*. The *unc-42* allele *e419* was then crossed to *otIs341* and phenocopied *ot712*. Sanger sequencing revealed that *ot712* harbors a late nonsense mutation

(W181>Stop) in exon 6. However, this mutation is not within the predicted homeodomain (Baran et al., 1999).

Cloning of *ot704* and *ot705* using Whole genome sequencing and CloudMap

ot704 and *ot705* were cloned using the more recently developed one-step whole-genome-sequencing approach, taking advantage of the polymorphism between two isolated *C. elegans* strains (Doitsidou et al., 2010). In each experiment the mutant strain (in N2 bristol background) was crossed to the Hawaiian isolate and around 50-60 mutant F2 progeny were picked, amplified, and pooled for subsequent sequencing. The ratio of N2 versus Hawaiian SNP was then calculated based on the sequence pile up. The closer the SNPs are to the causal mutation, the higher the N2/Hawaiian ratio is. Genome sequences were then analyzed by CloudMap, a Cloud-based pipeline that allows rapid candidate variant mapping (Minevich et al., 2012).

ot704 was mapped to the 7.75-9MB region on linkage group I (Figure 2A), within which only one transcription factor-encoding gene, *hlh-16* has a mutation in its coding region (Q118 > Stop). Sanger sequencing further revealed that *mut5F* and *mut5G* harbors a mutation in the same loci within the first exon of *hlh-16* (R21>Stop) (Figure 3), and are likely siblings from the same heterozygous mother as the screen was non-clonal. Allele name *ot711* was assigned to this mutation.

ot705 was mapped to the 3-6 MB region on chromosome III (Figure 2B). Within this interval are two candidate transcription factors with mutations in their protein coding region, the fork head transcription factor encoding gene *fkh-5* and the neuroD

homolog *cnd-1*. Complementation against the deletion allele *gk781* further confirmed that *ot705* is an allele of *cnd-1* (Table 3).

Rescue experiments for *ot704* and *ot705*

Two different constructs were used to rescue *ot704* animals. A PCR fragment of the genomic locus from 1206 bp upstream to 526 bp downstream of the *hlh-16* locus was directly injected at 10 ng/ul into mutant animals using linearized *rol-6(d)* (1 ng/ul) as injection marker (*otEx4943*). This rescued 86% of the animals back to wild type. An available transgene (Bertrand and Hobert, 2009) that spans 514 bp upstream of *hlh-16* to the next gene (~1.8kb downstream) rescued similarly when introduced to the mutant animals (Table 2).

Two separate experiments were performed to rescue *ot705*. A PCR fragment of the genomic locus of *cnd-1* from 3127 bp upstream to 500 bp downstream was injected as simple arrays in experiment 1. In both lines obtained a majority of the animals were partially rescued to intermediate phenotypes as apposed to wild type (2RMDD and 2 RMDV). In experiment 2 the same PCR fragment was injected in complex arrays at 3ng/ul, which was able to rescue 58% of the animals back to wild type (Table 3).

Expression pattern of *unc-42*, *hlh-16* and *cnd-1* in RMDD/RMDV

A fosmid *gfp* reporter was used to visualize *unc-42* expression. UNC-42 is expressed in thirty or so cells in the head of the worm. Colocalization with a red reporter gene *otIs317(mgl-1::mcherry)* expressed in RMDD and RMDV confirms *unc-42* expression in these four cells, which persists through adulthood.

No postembryonic expression was observed with a translational reporter of *hlh-16*, which rescues the defects in the left/right asymmetry in the interneuron AIY and the motoneuron SMDD (Bertrand and Hobert, 2009). This is consistent with the notion that *hlh-16* as a proneural gene probably works at an earlier step before terminal differentiation takes place.

cnd-1 expression was assessed using a transcriptional reporter as well as a fosmid based reporter. Based on a transcriptional and a semi-translational reporter expression, Hallam et al. reported that *cnd-1* expression starts from the 14-cell stage in the embryo and persists throughout gastrulation and epidermal enclosure until hatching, but is completely gone by the end of the first larval stage. In my experiments, the transcription reporter *stIs10055* (*cnd-1*^{3.2kb} *prom*::*HIS-24*::*mCherry*) is expressed throughout adulthood, with expression still seen in 7 day old adults. The fosmid reporter is not expressed post-embryonically. Preliminary lineage tracing data (n=1) suggests that *cnd-1* expression in the RMDDs and RMDVs start as early as two to three divisions earlier. For RMDDL and RMDVL, the expression starts in the grandmother cell, and persists until the last division, except that in RMDVL expression fades away soon after the birth of the cell (Figure 4A). In RMDDR and RMDVR, *cnd-1* expression begins even earlier, at the great grandmother stage, and remains on until after the cell has terminally differentiated (Figure 4B).

For future experiments and discussion see Chapter 6.

References:

Baran, R., Aronoff, R., and Garriga, G. (1999). The *C. elegans* homeodomain gene *unc-42* regulates chemosensory and glutamate receptor expression. *Development* *126*, 2241-2251.

Bertrand, V., and Hobert, O. (2009). Linking asymmetric cell division to the terminal differentiation program of postmitotic neurons in *C. elegans*. *Dev Cell* *16*, 563-575.

Brockie, P.J., Madsen, D.M., Zheng, Y., Mellem, J., and Maricq, A.V. (2001). Differential expression of glutamate receptor subunits in the nervous system of *Caenorhabditis elegans* and their regulation by the homeodomain protein *UNC-42*. *The Journal of neuroscience : the official journal of the Society for Neuroscience* *21*, 1510-1522.

Doitsidou, M., Flames, N., Lee, A.C., Boyanov, A., and Hobert, O. (2008). Automated screening for mutants affecting dopaminergic-neuron specification in *C. elegans*. *Nat Methods* *5*, 869-872.

Doitsidou, M., Poole, R.J., Sarin, S., Bigelow, H., and Hobert, O. (2010). *C. elegans* mutant identification with a one-step whole-genome-sequencing and SNP mapping strategy. *PLoS One* *5*, e15435.

Duerr, J.S., Han, H.P., Fields, S.D., and Rand, J.B. (2008). Identification of major classes of cholinergic neurons in the nematode *Caenorhabditis elegans*. *The Journal of comparative neurology* *506*, 398-408.

Greer, E.R., Perez, C.L., Van Gilst, M.R., Lee, B.H., and Ashrafi, K. (2008). Neural and molecular dissection of a *C. elegans* sensory circuit that regulates fat and feeding. *Cell metabolism* *8*, 118-131.

Hart, A.C., Sims, S., and Kaplan, J.M. (1995). Synaptic code for sensory modalities revealed by *C. elegans* *GLR-1* glutamate receptor. *Nature* *378*, 82-85.

Minevich, G., Park, D.S., Blankenberg, D., Poole, R.J., and Hobert, O. (2012). CloudMap: a cloud-based pipeline for analysis of mutant genome sequences. *Genetics* *192*, 1249-1269.

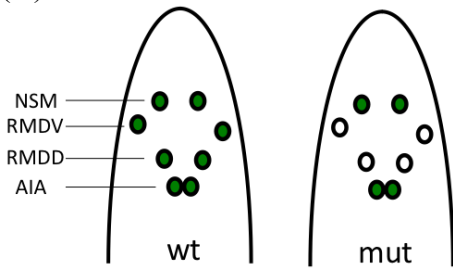
Figure 1: Class I and II phenotypes of RMDD/RMDV mutant animals.

(A) Mutant phenotype of Class I mutants are 100% penetrant in RMDD/RMDV loss.

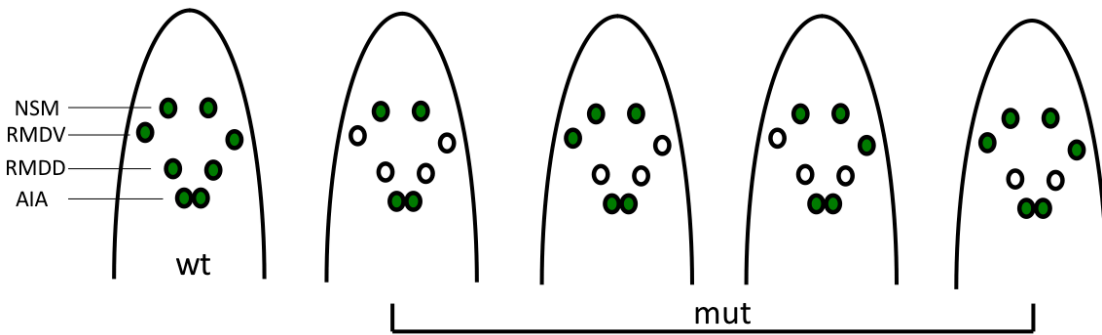
(B) Class II mutants phenotypes. Animals lose RMDDL/R, but can have zero, one or two RMDVs.

(C) Effects of *unc-42*, *ot704* and *ot705* on *otIs341 (mgl-1::gfp)*.

(A)



(B)



(C)

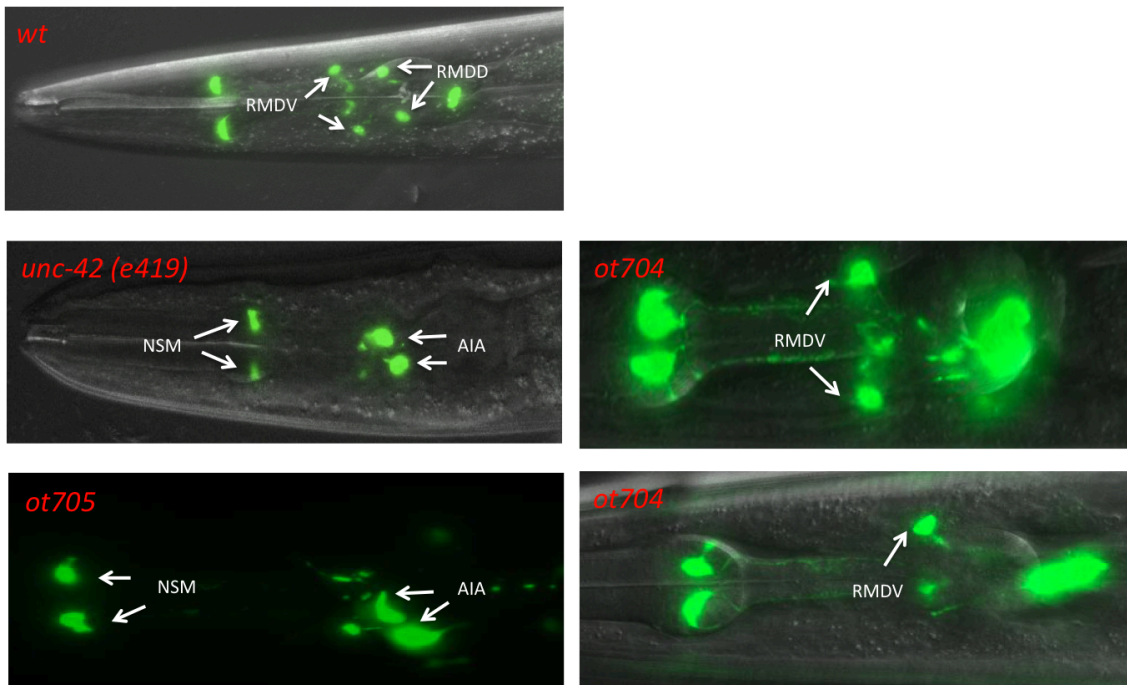
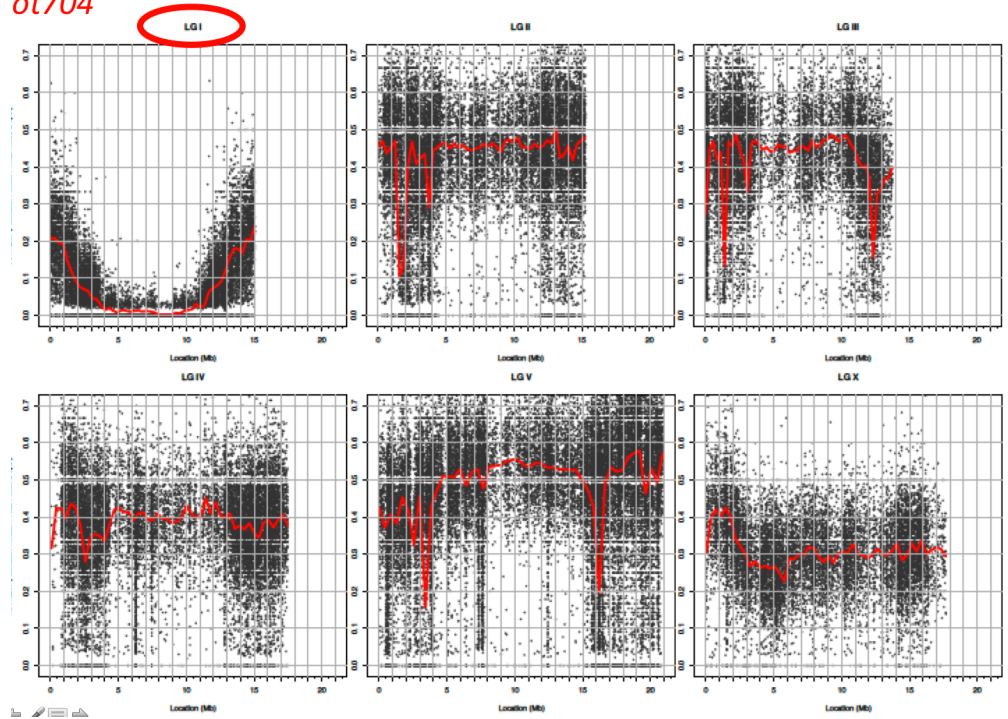


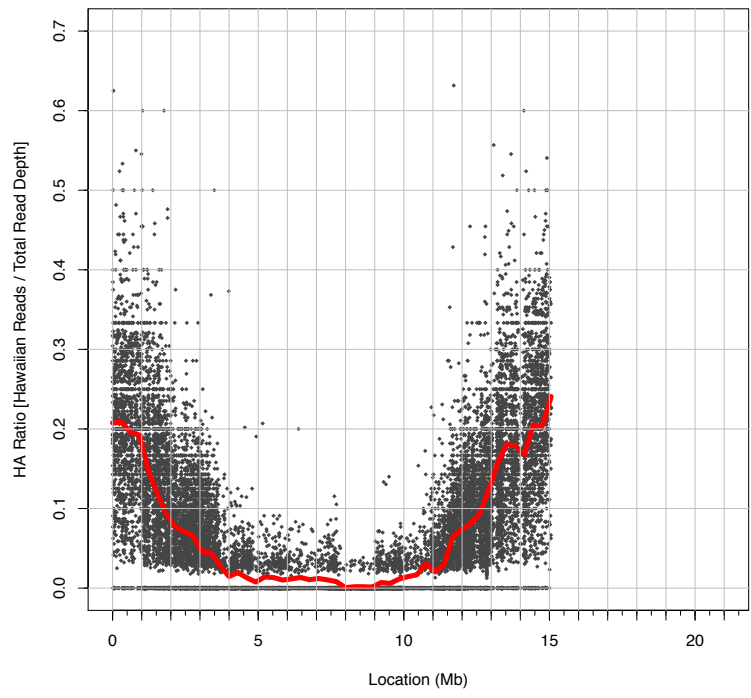
Figure 2: Scattered ratio plots of the *ot704* and *ot705* SNP reads. Y-axis: Hawaiian/total reads; X-axis: Location on the linkage group (Mb).

(A)

ot704

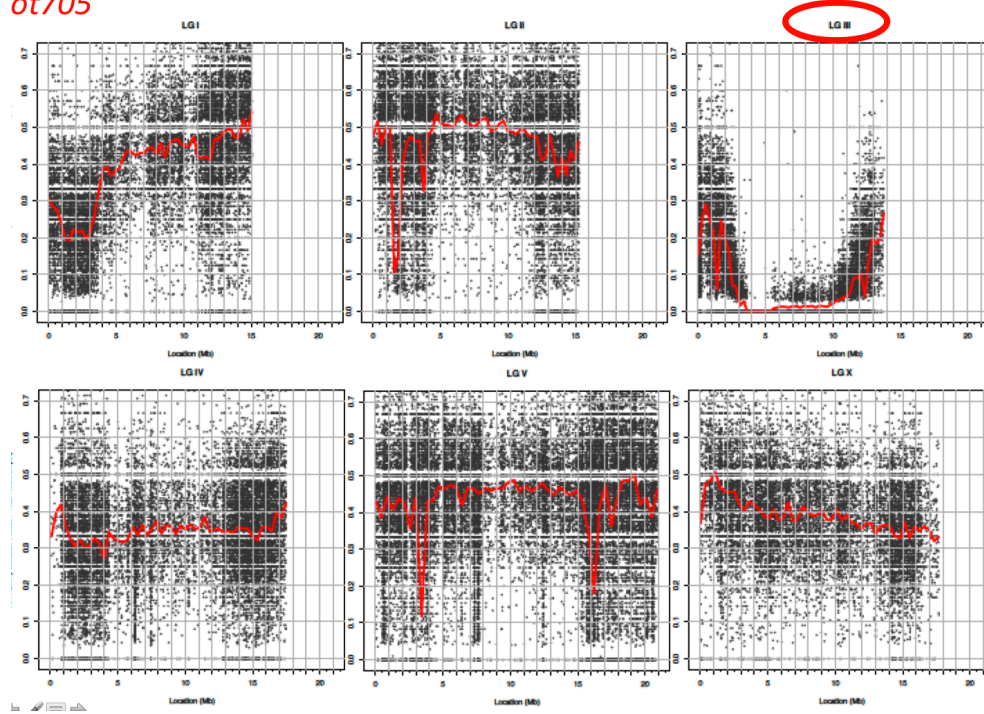


LG I



(B)

ot705



LG III

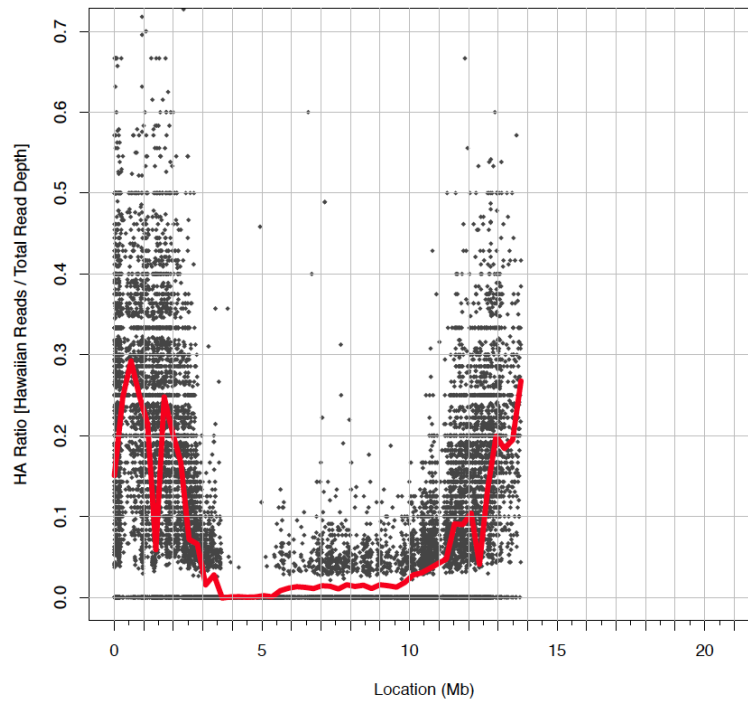
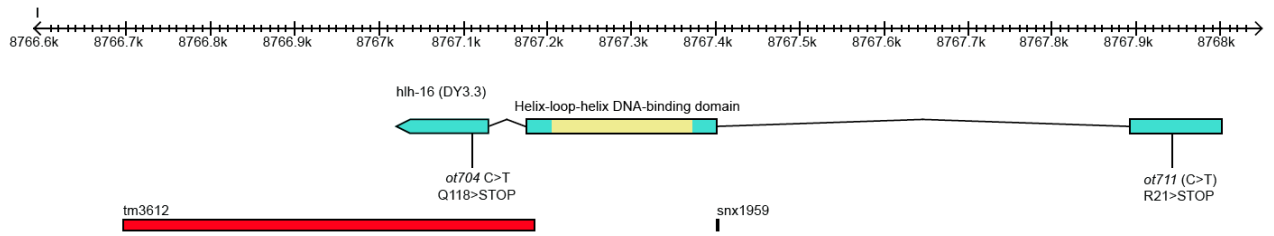


Figure 3: Overview of the *hlh-16* and *cnd-1* locus and relevant alleles.

(A)



(B)

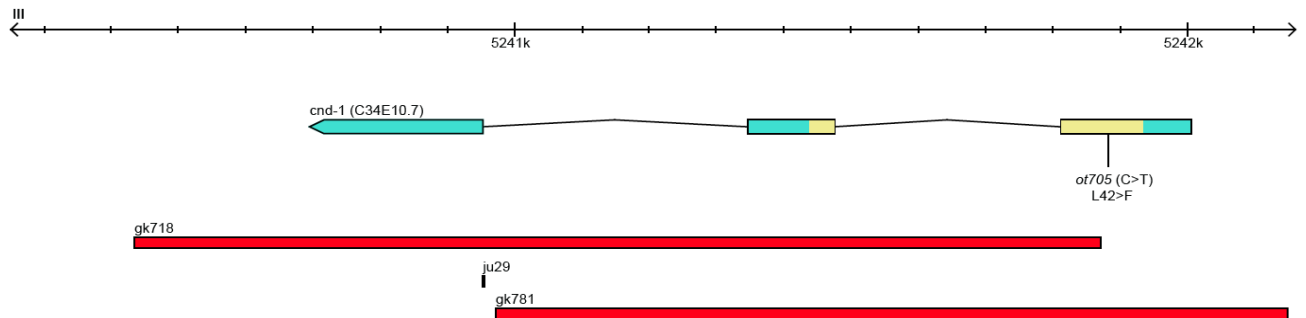


Figure 4: Preliminary lineage representation of *cmd-1* expression in the RMDD and RMDV lineage (n=1).

Note: only the RMDD and RMDV lineages were traced. Other lineages are not looked at and lack of green highlights in other lineages is not an indication of absence of expression.

(A) Left side

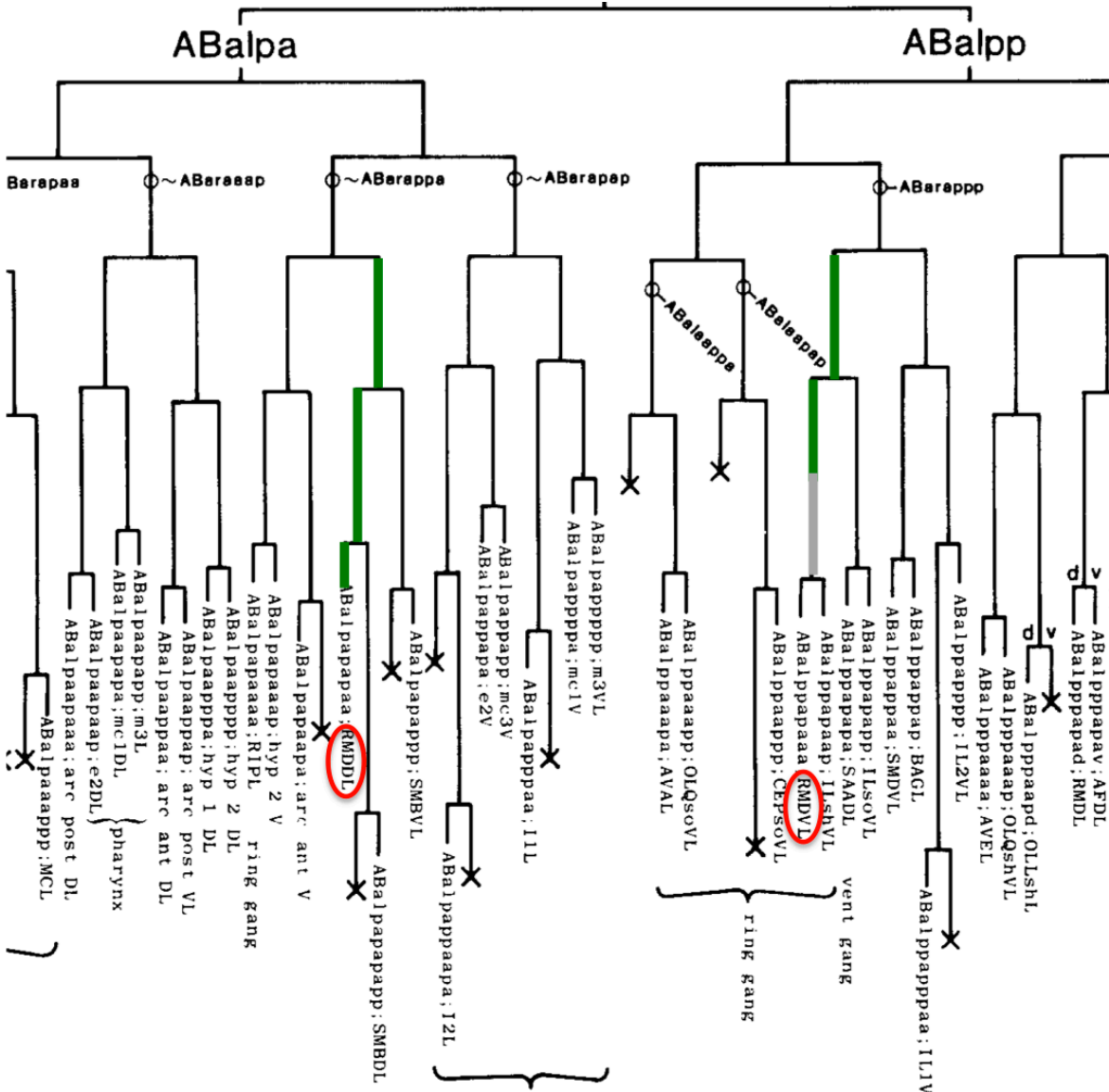


Table 1: Quantification of *unc-42*, *ot704* and *ot705* mutant phenotypes

	RMDD=0 RMDV=0	RMDD=0 RMDV=1	RMDD=0 RMDV=2	RMDD=2 RMDV=2	n
<i>wt</i>	-	-	-	100%	>100
<i>ot704</i>	14%	23%	63%	-	65
<i>ot705</i>	100%	-	-	-	30
<i>e419(unc-42)</i>	100%	-	-	-	29

Table 2: *ot704* rescue experiments

	RMDD= 0 RMDV= 0	RMDD= 2 RMDV= 2	RMDD= 0 RMDV= 1	RMDD= 0 RMDV= 2	RMDD= 1 RMDV= 2	RMDD= 1 RMDV= 1	RMDD= 1 RMDV= 0	RMDD= 2 RMDV= 1	n
<i>wt</i>	-	100%	-	-	-	-	-	-	>100
<i>ot704</i>	14%	-	23%	63%	-	-	-	-	65
<i>ot704: otEx4943</i>	-	86%	-	5%	7%	-	-	2%	44
<i>ot704: no Ex4943</i>	7%	25%	7%	57%	4%	-	-	-	44
<i>ot704: otEx4503</i>	-	85%	-	2%	11%	2%	-	-	47
<i>ot704: no Ex4503</i>	42%	17%	36%				5%	-	36

Table 3: *ot705* rescue experiments

	RMDD =0 RMDV =0	RMDD =2 RMDV =2	RMDD =0 RMDV =1	RMDD =0 RMDV =2	RMDD =1 RMDV =2	RMDD =2 RMDV =0	RMDD =1 RMDV =1	RMDD =1 RMDV =0	RMDD =2 RMDV =1	n
<i>wt</i>	-	100%	-	-	-	-	-	-	-	>100
<i>ot705</i>	100%	-	-	-	-	-	-	-	-	30
Simple 30ng/ul line1:										
<i>ot705: otEx4952</i>	17%	-	23%	30%	4%	13%	11%	2%	-	47
<i>ot705: no Ex4952</i>	97%	-	3%	-	-	-	-	-	-	31
Simple 30ng/ul line2:										
<i>ot705: otEx4953</i>	9%	14%	6%	23%	23%	14%	3%	-	9%	35
<i>ot705: no Ex4953</i>	97%	-	3%	-	-	-	-	-	-	34
Complex 3 ng/ul line1:										
<i>ot705: otEx4954</i>	-	58%	2%	9%	17%	2%	2%	-	9%	43
<i>ot705; no Ex4954</i>	Did not score because “ <i>unc</i> ” phenotype is not well rescued. (Animals that do not roll may carry the transgene.)									

CHAPTER 4:

A left/Right asymmetric neuronal differentiation program is controlled by the *Caenorhabditis elegans* LSY-27 Zn finger transcription factor

Zhang F, O'Meara MM, Hobert O. Genetics. 2011 Jul;188(3):753-9.

ASEL and ASER are a pair of gustatory neurons in the head of the worm that express different sets of terminal genes and therefore carry out distinct functions. Unbiased forward genetic screens have been vigorously performed in search of mutants that disrupt ASE laterality. This chapter characterizes two previously uncloned mutants, in which ASEL adopts a “mixed” state. ASER terminal genes are ectopically expressed in ASEL, while ASEL terminal features remain intact. One mutant turns out to be the first allele of the LIM homeobox gene *lim-6* that is pulled out from our screens. The other mutant has a lesion in the C2H2 zinc finger transcription factor *lsy-27* and displays similar phenotypes to *lim-6*. *lsy-27* also affects *lim-6* expression. Based on expression pattern analysis and temperature-shift experiments we propose that LSY-27 function is restricted to assisting the initial onset of LIM-6 expression in the embryos but not the maintenance phase at later stages.

I cloned and characterized *lsy-27*, and conducted all related genetic and molecular analysis as well as temperature shift experiments. Maggie O'Meara cloned *ot146*, and updated the *lim-6* gene structure.

A Left/Right Asymmetric Neuronal Differentiation Program Is Controlled by the *Caenorhabditis elegans* LSY-27 Zinc-Finger Transcription Factor

Feifan Zhang, M. Maggie O'Meara, and Oliver Hobert¹

Howard Hughes Medical Institute, Department of Biochemistry and Molecular Biophysics, Columbia University Medical Center, New York, New York 10032

ABSTRACT Functional diversification across the left/right axis is a common feature of many nervous systems. The genetic programs that control left/right asymmetric neuron function and gene expression in the nervous system are, however, poorly understood. We describe here the molecular characterization of two phenotypically similar mutant *Caenorhabditis elegans* strains in which left/right asymmetric gene expression programs of two gustatory neurons, called ASEL and ASER, are disrupted such that the differentiation program of the ASER neuron is derepressed in the ASEL neuron. We show that in one mutant strain the LIM homeobox gene *lim-6* is defective whereas in another strain a novel member of a nematode-specific, fast-evolving family of C2H2 zinc-finger transcription factors, *lsy-27*, is mutated, as revealed by whole-genome sequencing. *lsy-27* is broadly and exclusively expressed in the embryo and acts during the initiation, but not during the maintenance phase of ASE asymmetry control to assist in the initiation of *lim-6* expression.

LEFT/RIGHT asymmetric gene expression patterns in the nervous system of invertebrate and vertebrates species have been described and are generally thought to be the foundation of the striking functional lateralization of many nervous systems (Hobert *et al.* 2002; Sun *et al.* 2005; Sun and Walsh 2006; Taylor *et al.* 2010). Yet it is not well understood how left/right gene expression patterns are regulated. In the nematode *Caenorhabditis elegans*, a class of putative chemoreceptors of the GCY family are expressed in a left/right asymmetric manner in a bilateral pair of functionally lateralized gustatory neurons, called ASEL and ASER (Yu *et al.* 1997; Ortiz *et al.* 2006). These *gcy* genes are required for the left/right asymmetric processing of chemosensory information by the two ASE neurons (Ortiz *et al.* 2009). Genetic mutant screens have revealed a number of genes (called “*lsy* genes” for laterally symmetric) that control the left/right asymmetric expression of *gcy* genes (Sarin

et al. 2007). Phenotypic analysis of these mutants has revealed several distinct types of asymmetry mutants. In class I mutants, the *gcy* expression profile of the ASER neuron completely converts to that of the ASEL neuron (“2 ASEL” mutants). In class II mutants, the opposite occurs (“2 ASER” mutants; *e.g.*, *die-1* as shown in Figure 1A). In class III mutants, both ASEL and ASER *gcy* receptors are lost. In class IV mutants, the ASER-specific *gcy* genes are derepressed in ASEL, but the ASEL-specific *gcy* genes remain unaffected; or vice versa, ASEL-specific *gcy* genes are derepressed in ASER, but ASER-specific *gcy* genes remain unaffected (Sarin *et al.* 2007). Either the ASEL or ASER neurons therefore exist in a “mixed” state in class IV mutants (Figure 1A). Due to their more limited phenotypic effects, class IV genes would be expected to work downstream of class I and class II genes, and indeed, the analysis of the expression of class IV genes in class I or II mutant backgrounds confirmed this notion (Johnston *et al.* 2005, 2006) (Figure 1A).

Class IV genes are essential for the appropriate function of the ASE neurons. This was first demonstrated through a detailed phenotypic analysis of animals that lack the ASEL-expressed *lim-6* LIM homeobox gene and that therefore display a class IV phenotype in which ASEL-expressed *gcy* genes are unaffected, but ASER-expressed *gcy* genes are

Copyright © 2011 by the Genetics Society of America

doi: 10.1534/genetics.111.129064

Manuscript received March 28, 2011; accepted for publication May 5, 2011

Available freely online through the author-supported open access option.

Supporting information is available online at <http://www.genetics.org/content/suppl/2011/05/09/genetics.111.129064.DC1>.

¹Corresponding author: Columbia University, 701 W. 168th St., HHSC 724, New York, NY 10032. E-mail: or38@columbia.edu

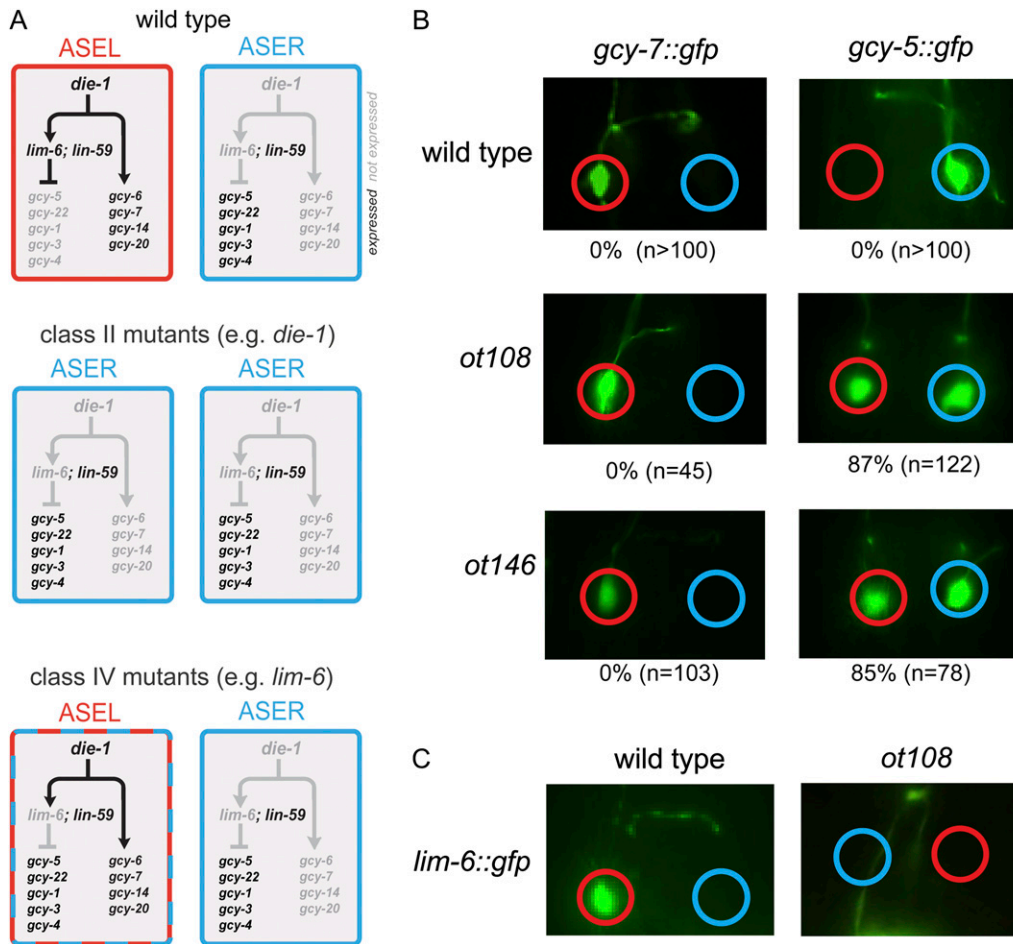


Figure 1 *Lsy* genes and mutant phenotypes. (A) A simplified version of the genetic pathway that controls left/right asymmetry in the ASE neurons. Loss of *die-1*, a Zn-finger transcription factor, results in a class II *Lsy* phenotype (in which ASEL fate markers are lost and ASER fate markers are gained in ASEL), and loss of *lim-6*, a LIM homeobox gene, results in a class IV *Lsy* phenotype (in which ASER fate markers are gained in ASEL, but ASEL fate markers unaffected) (Hobert *et al.* 1999; Chang *et al.* 2004). Loss of broadly expressed *lin-59*, a histone methyltransferase, also results in a class IV *Lsy* phenotype (Sarin *et al.* 2007, 2010). (B) Effect of *Lsy-27(ot108)* and *lim-6(ot146)* mutant alleles on ASEL/ASER asymmetry markers. *ots3(gcy-7::gfp)* labels ASEL and *ntls1(gcy-5::gfp)* labels ASER. The phenotype is quantified in Table 1. (C) *ot108* also affects *lim-6::gfp(ots114)* expression. In 57.4% of animals, *lim-6::gfp* fails to be expressed, and in 27.7% of animals, expression is visible but weaker than in wild type ($n = 47$).

derepressed in ASER (Figure 1A) (Hobert *et al.* 1999). Such mutant animals are unable to discriminate between ASEL- and ASER-sensed chemosensory cues (Pierce-Shimomura *et al.* 2001).

lim-6 is not the only gene with such a function. Three mutants retrieved from a previous large-scale mutagenesis screen for the asymmetry mutants *ot104*, *ot108*, and *ot146* (Sarin *et al.* 2007) display a phenotype similar to *lim-6* (Figure 1B and Table 1). *ot104* was found to be an allele of the ubiquitously expressed ASH1-type histone methyltransferase *lin-59* (Sarin *et al.* 2010), but the *ot108* and *ot146* alleles had not previously been molecularly characterized. We present their characterization in this Note.

ot146 is an allele of the LIM homeobox gene *lim-6*

ot146 mutant animals are viable and fertile and display no obvious morphological abnormalities. Their class IV *Lsy* phenotype is recessive. Due to its failure to complement what turned out to be a very unusual allele, called *ot101*, of the zinc (Zn)-finger transcription factor *che-1*, a terminal selector of ASEL and ASER neuron fate (Etchberger *et al.* 2009), we had assumed that *ot146* was located on chromosome I, where *che-1* is located (Sarin *et al.* 2007). However, subse-

quent mapping placed *ot146* on chromosome X, where the *lim-6* locus resides. We find that *ot146* contains a C83Y change in the second LIM domain of *lim-6* (supporting information, Figure S1). The mutated cysteine residue is 100% conserved in all LIM domains and is essential for the structural integrity of a LIM domain through the coordination of a Zn ion (Kadrmaz and Beckerle 2004). The *ot146* allele fails to complement the *lim-6* null allele *nr2073*, and its *Lsy* phenotype is rescued by a genomic piece of DNA that contains the *lim-6* locus (Table 2). We conclude that *ot146* is an allele of *lim-6*. This is the first *lim-6* allele retrieved from our mutant screen [the only previously characterized *lim-6* allele, *nr2073*, is a reverse engineered allele (Hobert *et al.* 1999)].

ot108 affects a member of a C2H2 Zn-finger protein family

Like *lim-6* mutant animals, *ot108* mutant animals show depression of the ASER marker *gcy-5* in ASEL, while *gcy-7* expression in ASEL is unaffected (Figure 1B and Table 1). Other than the *Lsy* phenotype, *ot108* mutants animals are viable and fertile and display no obvious morphological abnormalities. Aside from the effect of *ot108* on *gcy-5*

Table 1 Lsy phenotypes of *lim-6* and *lsy-27*

	% animals with the following phenotypes (at 25°):						<i>n</i>	% Lsy
	ASEL only (%)	ASEL > ASER (%)	No expression (%)	ASEL = ASER (%)	ASEL < ASER (%)	ASER only (%)		
ASEL marker (<i>gcy-7::gfp; otIs3</i>)								
Wild type	100	0	0	0	0	0	>100	0
<i>lim-6(nr2073)</i>	100	0	0	0	0	0	35	0
<i>lim-6(ot146)</i>	100	0	0	0	0	0	103	0
<i>lsy-27(ot108)</i>	100	0	0	0	0	0	45	0
<i>lsy-27(tm593)</i>	100	0	0	0	0	0	66	0
ASER marker (<i>gcy-5::gfp; ntlS1</i>)								
Wild type	0	0	0	0	0	100	>100	0
<i>lim-6(nr2073)</i>	0	0	0	89	5	6	82	94
<i>lim-6(ot146)</i>	0	0	0	8	77	15	78	85
<i>ot146/nr2073</i>	0	0	0	0	56	44	50	56
<i>lsy-27(ot108)</i>	0	0	0	39	48	13	122	87
<i>lsy-27(tm593)</i>	0	0	0	0	62	38	117	62
<i>ot108/tm593</i>	0	0	0	3	6	91	31	9
<i>ot108/+</i>	0	0	0	0	0	100	56	0

expression, *ot108* animals also show a significant loss of *lim-6* expression in ASEL, thereby providing an explanation of the *lim-6*-like phenotype of *ot108* mutant animals (Figure 1C).

Upon isolation of *ot108* mutant animals in our original Lsy screen (Sarin *et al.* 2007), we noted that *ot108* fails to complement the derepression of ASER fate in the ASEL phenotype of a mutation in the *die-1* Zn-finger transcription factor, an inducer of *lim-6* expression in ASEL (a class II gene that also results in the loss of ASEL fate) (Figure 1A). Due to this lack of complementation, we had therefore initially considered *ot108* to be an allele of *die-1* (Sarin *et al.* 2007). However, our subsequent analysis revealed no mutation in the *die-1* locus of *ot108* mutant animals and, moreover, the *ot108* mutant phenotype could not be rescued with a genomic piece of DNA that rescues a canonical *die-1* allele (data not shown). Subsequent chromosomal linkage analysis showed that *ot108* is linked to chromosome V, while *die-1* maps to chromosome II. After mapping *ot108* to the right arm of chromosome V using conventional SNP mapping (Wicks *et al.* 2001), we subjected the strain to whole-genome sequencing using an Illumina GAI genome analyzer (Sarin *et al.* 2008) and analyzed the data with MAQGene (Bigelow *et al.* 2009). Sequencing parameters and results are summarized in Table S1. In brief, within the genetically defined interval, we detected 22 sequence variants predicted to affect protein-coding genes (missense, non-sense or splice-site mutations). Nineteen of these variants were found in other whole-genome sequencing data sets that our lab has generated and were therefore considered background variants, leaving three protein-coding alterations. One of these alterations is a Ser-to-Leu change in the predicted C2H2 Zn-finger transcription factor F47H4.1 (Figure 2A and Figure S2). F47H4.1 is a member of C2H2 Zn-finger transcription factors with several paralogs in *Caenorhabditis*

elegans and orthologs in other nematode species, but no apparent orthologs outside nematodes (Figure 2B and Figure S2). All members of this family contain three closely clustered C2H2 Zn fingers at the N terminus of the protein, but no other recognizable domains. The serine residue that is mutated in *ot108* is phylogenetically conserved (Figure S2). The only gene in this family that had been previously characterized is the *ham-2* transcription factor, which is involved in *C. elegans* HSN motor neuron specification (Baum *et al.* 1999).

Both a fosmid spanning the entire F47H4.1 locus plus neighboring genes and a genomic piece of DNA containing 2.6 kb upstream of F47H4.1 and the F47H4.1 locus (Figure 2A) rescue the *ot108* mutant phenotype (Table 2). Animals carrying a deletion allele of F47H4.1, *tm593* (kindly provided by the *C. elegans* knockout facility at Tokyo Women's Medical University School of Medicine) (Figure 2A), also display a class IV Lsy phenotype (Table 1). Also, like *ot108* animals, *tm593* animals are viable and fertile and display no obvious morphological abnormalities. Taken together, we conclude that it is the mutation in F47H4.1 that results in the class IV Lsy phenotype of *ot108* mutant animals, and we therefore called this gene *lsy-27* (Table S3 shows an updated numbering of *lsy* genes).

***ot108* is an altered function allele**

The *tm593* deletion allele is a molecular null, as confirmed by RT-PCR analysis, which revealed that only very short (<37 amino acids), truncated forms of the protein are generated in *tm593* animals, which do not contain any of the DNA-binding Zn-fingers (see File S1). We were therefore surprised to note that the Lsy phenotype of the *tm593* deletion allele is notably milder than the *ot108* missense allele in terms of both expressivity and penetrance (Table 1). We

Table 2 Transformation rescue and RNAi analysis

Genotype	Lsy phenotype ^a (%)	Wild-type phenotype (%)	n
Wild type	0	100	>100
<i>lim-6(ot146)</i>	85	15	78
<i>ot146; otEx3859 (Ex[lim-6 fosmid::yfp; rol-6(d)])</i>	0	100	41
<i>lsy-27(ot108)</i>	86.9	13.1	122
<i>lsy-27(ot108); lsy-27(RNAi)</i>	2.5	97.5	200
<i>lsy-27(ot108); empty vector (RNAi)</i>	86.8	13.2	111
<i>lsy-27(RNAi)</i>	0	100	71
<i>lsy-27(ot108); Ex[lsy-27^{transl}::gfp], line #1^b</i>	18.2	81.8	44
<i>lsy-27(ot108); Ex[lsy-27^{transl}::gfp], line #2</i>	17.2	82.8	87
<i>lsy-27(ot108); Ex[fosmid], line #1</i>	5.6	94.4	18
<i>lsy-27(ot108); Ex[fosmid], line #2</i>	9.1	90.9	44
<i>lsy-27(ot108); Ex[lsy-27^{fosmid}::yfp], line #1^b</i>	0	100	70
Genotype as above but array not transmitted from parental generation ^c	0	100	11
<i>lsy-27(ot108); Ex[lsy-27^{fosmid}::yfp], line #2</i>	0	100	54
Genotype as above but array not transmitted from parental generation	21.1	78.9	19

The *ot108* and *ot146* control data are repeated from Table 1 for comparison purposes. RNAi experiments were done by feeding, using standard protocols with a double-stranded RNA clone obtained from Geneservice.

^a Scored as a *gcy-5* reporter (*ntl51* or *otls220*) derepressed in ASEL in first eleven rows or loss of *lim-6::gfp* (*otls114*) in remaining four rows.

^b All expression constructs are shown in Figure 2A. See File S1 for details on the generation of the reporter constructs.

^c Arrays contain the *elt-2::gfp* injection marker. Animals derived from *elt-2::gfp(+)* parents that have lost this array as assessed by lack of intestinal *gfp* expression were scored.

therefore considered the possibility that *ot108* (which is recessive) is an altered function allele (Table 1). We tested this possibility by removing *lsy-27* gene activity in *ot108* mutant animals using RNA interference (RNAi) directed against *lsy-27*. We found that RNAi treatment completely reverted the *ot108* phenotype (Table 2), suggesting that it is indeed altered *lsy-27* function that explains the *ot108* phenotype.

We noted that animals that carry one copy of the *ot108* allele and one copy of the *tm593* allele display a phenotype that is even milder than the phenotype of either allele alone (Table 1). One copy of the *ot108* allele alone is therefore not enough to induce the altered function activity, but perhaps may be enough to provide some wild-type gene activity, thereby alleviating the *tm593* phenotype. The need for sufficient *ot108* dosage is also illustrated by the fact that the phenotype of *ot108* mutant animals can be rescued through supplying wild-type copies of the locus (Table 1).

We considered the possibility that the complete removal of *lsy-27* in *tm593* animals may be mostly compensated for by *lsy-27* paralogs, while the *ot108* allele may interfere with the compensatory function of the paralogues. Through the use of deletion alleles of these loci (again kindly provided by the *C. elegans* knockout facility in Tokyo), we found that neither of the two most closely related *lsy-27* paralogs, *ztf-25* or *ztf-28*, either alone or in combination (*i.e.*, *ztf-25 ztf-28* double nulls) displayed a Lsy phenotype (Table S2). *ztf-28 lsy-27* double-null mutant animals also display no Lsy phenotype. *ztf-25 lsy-27* double mutants could not be built due to close linkage of the two loci, and we therefore needed to resort to RNAi. *lsy-27* RNAi in a *ztf-28 ztf-25* double-mutant background also did not result in a Lsy phenotype, but we note that even though *lsy-27* RNAi does suppress the *ot108* Lsy phenotype, it does not recapitulate the *lsy-27(tm593)*

phenotype (Table 2), thereby allowing no firm conclusion about a triple loss of function of all three *lsy-27* paralogs.

Expression pattern and timing of action of *lsy-27*

By recombineering *yfp* into the fosmid that contains the *lsy-27* locus and that rescues the *lsy-27* phenotype (Table 2), we generated a reporter with which we monitored *lsy-27* expression (Figure 2A). We find that *lsy-27* is expressed very broadly throughout the embryo (Figure 3A). Expression can already be observed in one-cell embryos and continues to about the comma stage, when expression starts to fade out (Figure 3A). By the comma stage, most neurons, including ASEL/ASER, have terminally divided and begun to terminally differentiate. No expression is observed after hatching in larvae or in adult animals. Through colocalizing expression of the *lsy-27* reporter with an ASE-specific mCherry reporter, we confirmed that *lsy-27* is expressed in both ASE neurons in the comma-stage embryo when ASE laterality is established. As assessed with translational *gfp* reporters that fuse the entire loci to *gfp*, the most closely related *lsy-27* paralog, *ztf-25*, displays an essentially indistinguishable broad, embryo-restricted expression pattern (Figure S3), while the more distant paralog *ztf-28* shows no expression in embryos and postembryonically is expressed only in the intestine (data not shown).

The expression pattern of *lsy-27* suggests an embryonic role for the gene. We sought to corroborate this notion by exploiting the observation that the *ot108* allele is strongly temperature sensitive (Figure 3B). At 25°, 87% of animals display a Lsy phenotype while 12% do at 15°. By altering *lsy-27* gene activity at different stages through temperature shifts, we find that *lsy-27* activity is required only during embryogenesis, but not during postembryonic stages (Figure

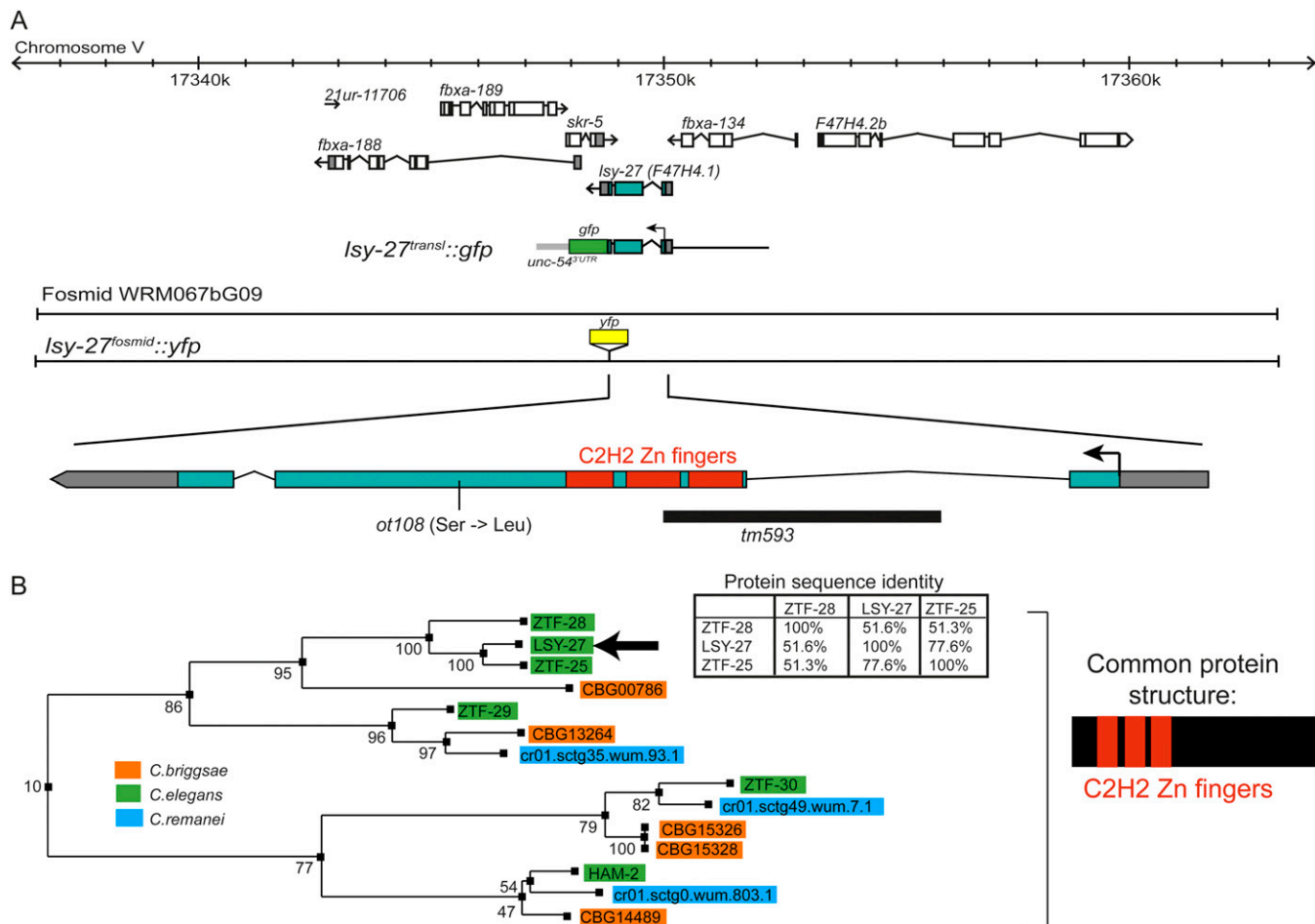


Figure 2 *lsy-27* is a new C2H2 Zn-finger protein. (A) Genomic position of *lsy-27* and rescue and reporter gene constructs. See File S1 for details on the generation of the reporter constructs. (B) The *lsy-27* gene family [modified from the TF317235 family tree generated by Treefam (<http://www.treefam.org>) (Li *et al.* 2006)].

3B). This contrasts with the continuous requirement of other *lsy* genes during postembryonic stages (O'Meara *et al.* 2010) and demonstrates that laterality control can be divided into initiation and maintenance phases.

The maternal loading of LSY-27 protein into oocytes as well as the embryonic focus of action also prompted us to ask whether *lsy-27* gene activity can be solely maternally supplied. Using transgenic *lsy-27* mutant animals that carry the germline-expressed *lsy-27* reporter fosmid, we assayed progeny that have lost the array and therefore contain only maternally supplied gene activity. In such animals, the *Lsy* phenotype is rescued (Table 2), corroborating maternal deposition of *lsy-27* gene activity.

Concluding remarks

We have described here a member of a nematode-specific C2H2 Zn-finger transcription factor family, *lsy-27*, which functions in ASE laterality control. The *lsy-27* mutant phenotype is similar to that of the ASEL-restricted LIM homeobox gene *lim-6*, as well as the ubiquitously expressed *lin-59* histone methyltransferase. We found that *lsy-27* not only

affects the terminal *gcy* gene markers in a manner similar to *lim-6*, but also affects *lim-6* expression. The embryo-restricted expression and function of *lsy-27* contrasts with the expression of *lim-6*, which is expressed continuously throughout the life of the ASEL neuron. We propose that the function of *lsy-27* is restricted to triggering the initial onset of *lim-6* expression. *lsy-27* may cooperate with ASEL-expressed *die-1* to trigger *lim-6* expression in the embryo. Once *lim-6* is turned on, *lsy-27* is no longer required to control laterality. This maintenance role is carried out by *die-1* (O'Meara *et al.* 2010) in conjunction with *lim-6*, which positively autoregulates (Johnston *et al.* 2005). Interestingly, *lsy-27* is not involved in conveying other *die-1* functions, such as the induction of ASEL fate markers (e.g., *gcy-7*), since those are affected only in *die-1*, but not in *lsy-27* mutants.

With the molecular identification of *ot108* and *ot146*, we have identified all but one gene retrieved from our large-scale screening of left/right asymmetry mutants (summarized in Table S4). Due to some adjustments in allele assignments as described here and elsewhere (Etchberger *et al.* 2009; Sarin *et al.* 2009; Flowers *et al.* 2010), we have recalculated saturation using various models (Sarin *et al.*

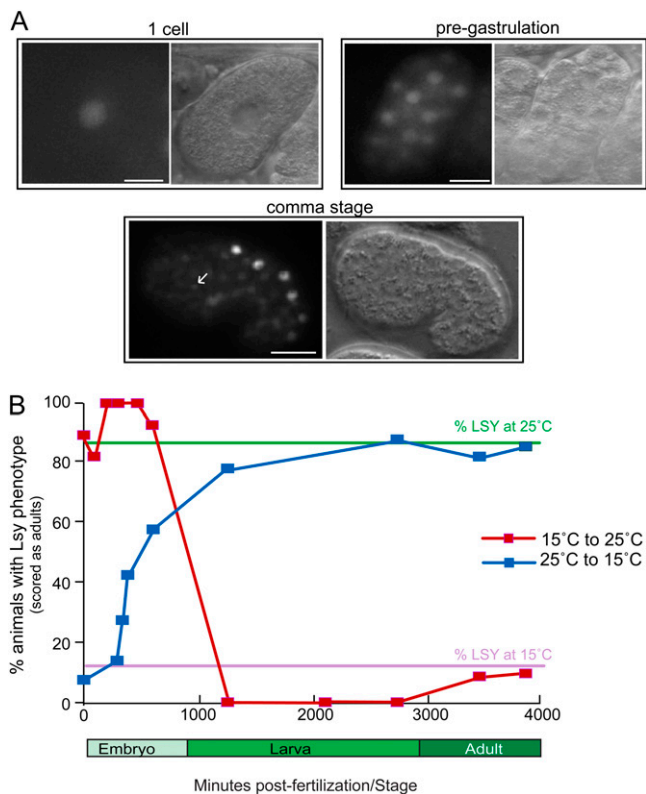


Figure 3 *Lsy-27* is expressed and acts during the initiation but not during the maintenance phase of left/right asymmetry control. (A) Expression pattern of *Lsy-27^{fosmid::yfp}* (shown in Figure 2A) at different embryonic stages. The embryos appear slightly deformed as they are squished together in the gonad of an adult animal. The white arrow indicates the ASEL neuron shortly after birth based on colocalization with a bilateral ASE-specific reporter *otIs232(che-1::mCherry)* (not shown). The *Lsy-27^{transl::gfp}* (shown in Figure 2A) shows a similar expression pattern except that, due to its failure to be expressed in the germline, we see only the onset of expression when zygotic gene expression starts in the early embryo. Bar, 10 μ m. (B) Temperature-shift experiments with *ot108; ntl51* animals indicate that *Lsy-27* activity is required only during embryogenesis, but not during postembryonic stages. Animals were cultured for at least three generations at either 15° or 25°. Animals were analyzed by isolating two- to four-cell embryos and temperature shifts were performed at various developmental stages. All animals were scored as 3-day-old adults.

2007) and retain our previous conclusion that the screen has not yet reached saturation. Future genetic screens are likely to provide further insights into the control of lateralized gene expression in the nervous system.

Acknowledgments

We thank the *C. elegans* knockout consortia directed by Shohai Mitani at Tokyo Women's Medical University School of Medicine for the *tm573*, *tm593*, and *tm630* alleles; Sumeet Sarin for help in recalculating genetic saturation; Qi Chen for expert assistance in generating transgenic strains; and members of the Hobert lab for comments on the manuscript. This work was funded by the National Institutes of Health (grants R01NS039996-05 and R01NS050266-03). O.H. is an Investigator of the Howard Hughes Medical Institute.

Literature Cited

- Baum, P. D., C. Guenther, C. A. Frank, B. V. Pham, and G. Garriga, 1999 The *Caenorhabditis elegans* gene *ham-2* links Hox patterning to migration of the HSN motor neuron. *Genes Dev.* 13: 472–483.
- Bigelow, H., M. Doitsidou, S. Sarin, and O. Hobert, 2009 MAQGene: software to facilitate *C. elegans* mutant genome sequence analysis. *Nat. Methods* 6: 549.
- Chang, S., R. J. Johnston, C. Frokjaer-Jensen, S. Lockery, and O. Hobert, 2004 MicroRNAs act sequentially and asymmetrically to control chemosensory laterality in the nematode. *Nature* 430: 785–789.
- Etchberger, J. F., E. B. Flowers, R. J. Poole, E. Bashllari, and O. Hobert, 2009 Cis-regulatory mechanisms of left/right asymmetric neuron-subtype specification in *C. elegans*. *Development* 136: 147–160.
- Flowers, E. B., R. J. Poole, B. Tursun, E. Bashllari, I. Pe'er *et al.*, 2010 The Groucho ortholog UNC-37 interacts with the short Groucho-like protein LSY-22 to control developmental decisions in *C. elegans*. *Development* 137: 1799–1805.
- Hobert, O., K. Tessmar, and G. Ruvkun, 1999 The *Caenorhabditis elegans* *lim-6* LIM homeobox gene regulates neurite outgrowth and function of particular GABAergic neurons. *Development* 126: 1547–1562.
- Hobert, O., R. J. Johnston, Jr., and S. Chang, 2002 Left-right asymmetry in the nervous system: the *Caenorhabditis elegans* model. *Nat. Rev. Neurosci.* 3: 629–640.
- Johnston, R. J., Jr., S. Chang, J. F. Etchberger, C. O. Ortiz, and O. Hobert, 2005 MicroRNAs acting in a double-negative feedback loop to control a neuronal cell fate decision. *Proc. Natl. Acad. Sci. USA* 102: 12449–12454.
- Johnston, R. J., Jr., J. W. Copeland, M. Fasnacht, J. F. Etchberger, J. Liu *et al.*, 2006 An unusual Zn-finger/FH2 domain protein controls a left/right asymmetric neuronal fate decision in *C. elegans*. *Development* 133: 3317–3328.
- Kadrmas, J. L., and M. C. Beckerle, 2004 The LIM domain: from the cytoskeleton to the nucleus. *Nat. Rev. Mol. Cell Biol.* 5: 920–931.
- Li, H., A. Coghlan, J. Ruan, L. J. Coin, J. K. Heriche *et al.*, 2006 TreeFam: a curated database of phylogenetic trees of animal gene families. *Nucleic Acids Res.* 34: D572–D580.
- O'Meara, M. M., F. Zhang, and O. Hobert, 2010 Maintenance of neuronal laterality in *Caenorhabditis elegans* through MYST histone acetyltransferase complex components LSY-12, LSY-13 and LIN-49. *Genetics* 186: 1497–1502.
- Ortiz, C. O., J. F. Etchberger, S. L. Posy, C. Frokjaer-Jensen, S. Lockery *et al.*, 2006 Searching for neuronal left/right asymmetry: genome-wide analysis of nematode receptor-type guanylyl cyclases. *Genetics* 173: 131–149.
- Ortiz, C. O., S. Faumont, J. Takayama, H. K. Ahmed, A. D. Goldsmith *et al.*, 2009 Lateralized gustatory behavior of *C. elegans* is controlled by specific receptor-type guanylyl cyclases. *Curr. Biol.* 19: 996–1004.
- Pierce-Shimomura, J. T., S. Faumont, M. R. Gaston, B. J. Pearson, and S. R. Lockery, 2001 The homeobox gene *lim-6* is required for distinct chemosensory representations in *C. elegans*. *Nature* 410: 694–698.
- Sarin, S., M. M. O'Meara, E. B. Flowers, C. Antonio, R. J. Poole *et al.*, 2007 Genetic screens for *Caenorhabditis elegans* mutants defective in left/right asymmetric neuronal fate specification. *Genetics* 176: 2109–2130.
- Sarin, S., S. Prabhu, M. M. O'Meara, I. Pe'er, and O. Hobert, 2008 *Caenorhabditis elegans* mutant allele identification by whole-genome sequencing. *Nat. Methods* 5: 865–867.
- Sarin, S., C. Antonio, B. Tursun, and O. Hobert, 2009 The *C. elegans* Tailless/TLX transcription factor *nhr-67* controls neuronal

- identity and left/right asymmetric fate diversification. *Development* 136: 2933–2944.
- Sarin, S., V. Bertrand, H. Bigelow, A. Boyanov, M. Doitsidou *et al.*, 2010 Analysis of multiple ethyl methanesulfonate-mutagenized *Caenorhabditis elegans* strains by whole-genome sequencing. *Genetics* 185: 417–430.
- Sun, T., and C. A. Walsh, 2006 Molecular approaches to brain asymmetry and handedness. *Nat. Rev. Neurosci.* 7: 655–662.
- Sun, T., C. Patoine, A. Abu-Khalil, J. Visvader, E. Sum *et al.*, 2005 Early asymmetry of gene transcription in embryonic human left and right cerebral cortex. *Science* 308: 1794–1798.
- Taylor, R. W., Y. W. Hsieh, J. T. Gamse, and C. F. Chuang, 2010 Making a difference together: reciprocal interactions in *C. elegans* and zebrafish asymmetric neural development. *Development* 137: 681–691.
- Wicks, S. R., R. T. Yeh, W. R. Gish, R. H. Waterston, and R. H. Plasterk, 2001 Rapid gene mapping in *Caenorhabditis elegans* using a high density polymorphism map. *Nat. Genet.* 28: 160–164.
- Yu, S., L. Avery, E. Baude, and D. L. Garbers, 1997 Guanylyl cyclase expression in specific sensory neurons: a new family of chemosensory receptors. *Proc. Natl. Acad. Sci. USA* 94: 3384–3387.

Communicating editor: D. I. Greenstein

GENETICS

Supporting Information

<http://www.genetics.org/content/suppl/2011/05/09/genetics.111.129064.DC1>

A Left/Right Asymmetric Neuronal Differentiation Program Is Controlled by the *Caenorhabditis elegans* LSY-27 Zinc-Finger Transcription Factor

Feifan Zhang, M. Maggie O'Meara, and Oliver Hobert

```

tctctctctctgtgttttttgaagaacccgagggcaaaagctcaaatctcATGGTGTGTATCGTTTATCCTTCTGGCGGATGGAGAGGgagcgacaatgtttttcaattaacttgacg
S L S L C F V E D R R Q K L N L M V C I V Y P S G A M E R * A T M F F Q L N L T
L S L C V L * K T G G K S S I S W C V S F I L L A R W R G E R Q C F F N * T * R
S L S V F C R R P E A K A Q S H G V Y R L S F W R D G E V S D N V F S I K L D V

ttttattcggaaattttttttccacaatattgaattaggtctttttagaagctctgtgaacatttagttttcaaatttaattgtccgaattttttaccgtgtgaaaaat
F Y S G N L F F S T N I E L G S F * K L L * T F S F Q I * M F R I F S T V V K N
F I R E I Y F F P Q I L N * V L F R S S C E H L V F K F K C S E F F L P L * K I
L F G K F I F F H K Y * I R F F L E A P V N I * F S N L N V P N F F Y R C E K S

caagggtctacaaaaatgttttctctactgtgcaacatttctgtattgtccaatttttagaaaaatccaaaagcttattgtgtgtcttaattggatggtacggtggaataacgct
Q G S T K K C F L L L G N I S A I V Q I F R K I Q K L I C V S * W M V R W N N A
K G L P K N V F Y L A T F L L S K F L E K S K S L F V C L N G W Y G G I T L
R V Y Q K M F S S T W Q H F C Y C P N F * K N P K A Y L C V L M D G T V E * R S

ctggaagctttttgaaaaatctaattttttaccgaacgtatttttaaatcaaaaaataaaagtttcaaaaaatcttcaacaattatactgatctctgaagtaccctaagaattgact
L E A F * K I * F F Y R T Y F L N Q K L K V S K N L Q Q F I L I S E V P * E L T
W K L F E K S N F F T E R I F * I K N * K F Q K I F N N L Y * S L K Y P K N * L
G S F L K N L I F L P N V F F K S K I K S F K K S S T I Y T D L * S T L R I D Y

atcttcaaaatcaatgtcttttgagcaaatctacaaaacttctcgtgaagaccgtccccatttttcagATTGGACATAGTAGATTGGACCAACCCCTCTCTCGGAACAGTGATCAGT
I F K I N A F G A N L Q T S S * D R P H F S R D C D I V D L D Q P S L G T V I S
S S K S M L L E Q I Y K L P R K T V P I F P E I A T * * I W T N P L S E Q * S V
L Q N Q C F W S K S T N F L V R P S P F F Q R L R H S R F G P T L S R N S D Q Y

ATCAAAGATGGATCAACGCCACTGATATGtaagctcaccaggaataaaaaagaatggatagagagtagaataacgaataaaaactaatatcatattgaaacaatcaaaacaagtttt
I K D G S T P T D I C K S T R N K K E M D R E * K T N K T N I I L K T I S F
S K M D Q R P L I Y V S L P G I K K K W I E S R K R I K L I S Y * K Q S K Q V L
Q R W I N A H * Y M * V Y Q E * K R N G * R V E N E * N * Y H I E N N Q N K F *

aagctcataaggttttcagggcggaagcttggtaagcaccttatagcttttcaataattttttgatgagtctgctcctcatttcagCGACACGCTCTACTACAACGGAAGACAA
K L I R F P G A E S L V K H L I S F R I I I F D E S A P H F S D N V * Y N G R Q
S S * G F Q G R K A W * S T L * A F E * L F L M S L L L I S A T T S S T T E D K
A H K V S R G G K L G E A P Y K L S N N Y F * * V C S S F Q R Q R L V Q R K T S

```

Red: previous Wormbase gene start
 ATCAAAG Exons (underlined: confirmed by RACE)
 gt splice junction
 ataag ASE motif
 aataa autoregulatory/maintenance motif
 S T P T protein coding sequence

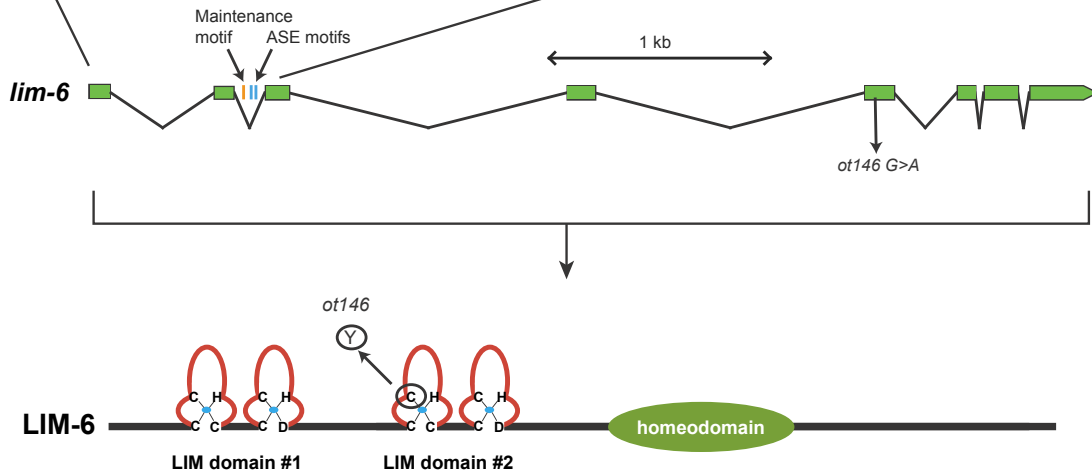
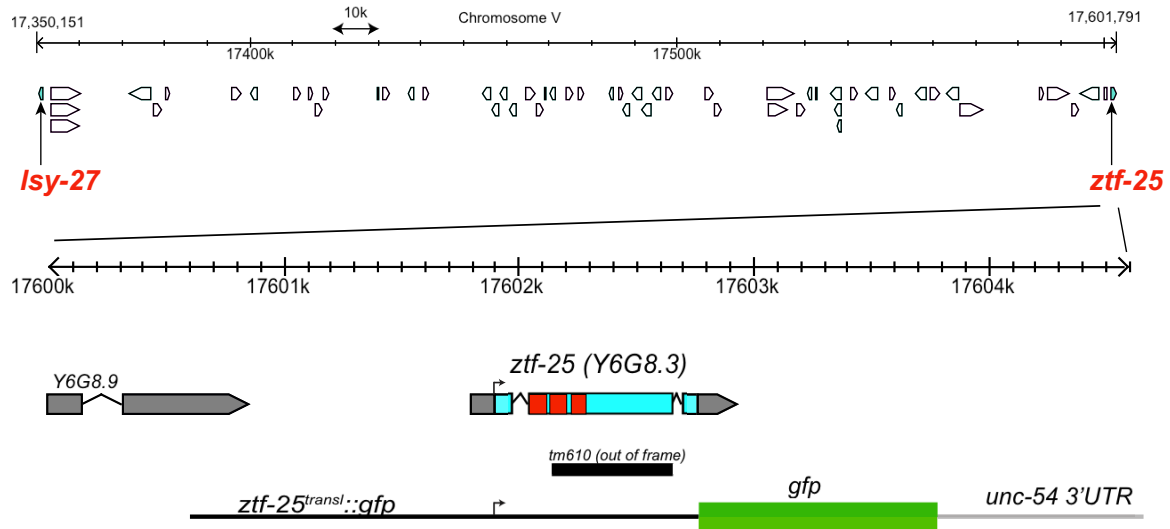


Figure S1 Updated *lim-6* locus and the location of the *ot146* allele. Based on our own 5'RACE results (using the Invitrogen GeneRacer Kit and the following *lim-6* gene specific reverse primers: *lim-6*-5'RACEup: GATCACTGTTCGAGAGAG and *lim-6*-5'RACEupnest: GGTCCAAATCTACTATGTCGC) and based on conservation to other nematode orthologs, the gene annotation shown here updates previous Wormbase annotations of *lim-6* through the addition of two new exons (first two exons). This put previously identified *cis*-regulatory sites for *lim-6* expression - two CHE-1 binding ASE motifs and a presumptive autoregulatory motif (ETCHBERGER *et al.* 2009) - into the second intron of the gene, as indicated. The Cys to Tyr mutation in *ot146* affects an invariant residue in the second LIM domain.

A



B

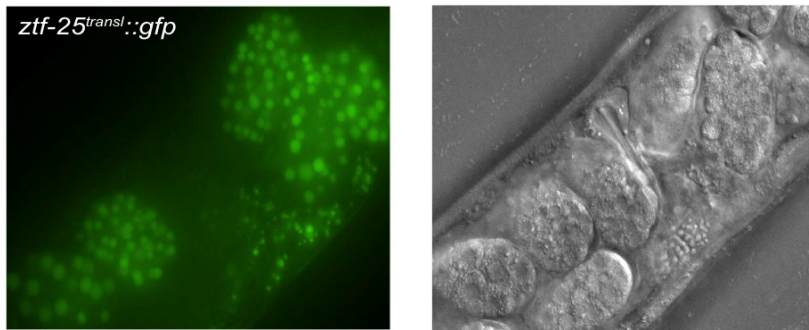


Figure S3 Expression pattern of the *lsy-27* paralog *ztf-25*. A: Gene and allele structure of *ztf-25*. A translational reporter was generated by fusing the genomic DNA containing 1.3 kb upstream of ATG to GFP and a heterologous 3'UTR. Note the proximity of *ztf-25* and its closest paralog, *lsy-27*. B: *ztf-25^{transl}::gfp* expression pattern in midembryonic stages. Embryos are contained within the gonad of the mother.

File S1
Supporting Information

Supporting Material and Methods

Genetic screen

The screen that uncovered *ot146* and *ot108*, as well as a large number of additional *Isy* genes has been described (SARIN *et al.* 2007). Since the appearance of that paper, the identities of several alleles originally characterized as novel genes have been resolved, including mutants described in this paper. We provide a summary of the updated gene designations and molecular identities in Table S3 and Table S4.

Transgenes

Transgenes that label ASEL and/or ASER fates:

otls114 = *Is[lim-6^{prom}::gfp; rol-6(d)]*

otls3 = *[gcy-7^{prom}::gfp; lin-15(+)]*

ntls1 = *[gcy-5^{prom}::gfp; lin-15(+)]*

otls220 = *Is[gcy-5^{prom}::mCherry; rol-6(d)]*

otls151 = *Is[ceh-36^{prom}::DsRed; rol-6(d)]*

otls232 = *Is[che-1^{promA}::mChopti::che-1_3'UTR; rol-6(d)]*

Rescue and expression constructs:

otEx4280, otEx4281 = two independent lines of *Ex[fosmid WRM067BG09; elt-2::gfp]*.

otEx4400, otEx4401 = two independent lines of *Ex[Isy-27^{transl}::gfp; elt-2::gfp]*

otEx4501, otEx4502 = two independent lines of *Ex[fosmid_WRM067BG09::yfp; elt-2::gfp]*

otEx4523 = *Ex[fosmid_WRM067BG09::yfp]*

otEx4337-4339 = three independent lines of *Ex[ztf-25^{transl}::gfp; rol-6(d)]*

otEx3859 = *Ex[lim-6^{fosmid}::yfp; rol-6(d)]*.

Generation of expression constructs

Isy-27^{transl}::gfp was generated by PCR-fusing the genomic locus of *Isy-27* including 2.6 kb of the upstream region to the *gfp* coding region and *unc-54_3'UTR* (30ng/μl), using a standard PCR fusion protocol (HOBERT 2002) and was coinjected with *elt-2::gfp* (50ng/μl) as injection marker. *ztf-25^{transl}::gfp* was generated by PCR-fusing the genomic locus of *ztf-25* including 1.3 kb of the upstream region to the *gfp* coding region and *unc-54_3'UTR* (30ng/μl) and was coinjected with *rol-6(d)* (50ng/μl) as injection marker. *Isy-27^{fosmid}::yfp* was generated by inserting *yfp* right before the stop codon of *Isy-27* in fosmid *WRM067BG09* (TURSUN *et al.* 2009).

The primer sequences for these constructs are as follows (from 5' to 3'):

Isy-27^{transl}::gfp:

Isy-27_translat_A: CTGATACGAGTACGGCATGGC

*Isy-27_translat_A**: GTCACGCGCAATGCATAC

Isy-27_translat_B: AGTCGACCTGCAGGCATGCAAGCTCTCAATTCCTGCTTGACGTGC

Isy-27_translat_C: AGCTTGCATGCCTGCAGGTGC

Isy-27_translat_D: AAGGGCCCGTACGGCCGACTA

*Isy-27_translat_D**: GGAAACAGTTATGTTTGGTATA

ztf-25^{transl}::gfp:

ztf-25_translat_A: GAGCCACTATCCACGCAC

*ztf-25_translat_A**: GAAATGCGTGGAGTTGC

ztf-25_translat_B: AGTCGACCTGCAGGCATGCAAGCTCTCAACTCCAGCTTGAAACGCTTC

C, D, D* same as above

lsy-27^{fosmid}::yfp:

lsy-27_fosmid_Fwd: CCGTCGAGAGCAAGATGATTCCAGAGAAGCACGTC AAGCAGGAAATTGAGatgagtaaaggagaagaactttcac

lsy-27_fosmid_Rev:

ataaaactgaacatcagttgaataggagttaaataggaataaaaatagTTAtttgtatagttcatccatgccatg

Genotyping the *ztf-25*, *ztf-28* and *lsy-27* deletion alleles

We genotyped animals for the presence of the deletion alleles:

tm593 genotyping primers:

tm593_external_fwd: CTCTCCCCTTCTCCACAAC

tm593_external_rev: GCGTTGGAGTGTGTCAGC

tm610 genotyping primers:

tm610_external_fwd: GACAGTCTGGTTACTTGGTAG

tm610_external_rev: CAGGTGGAAGTATCGTG

tm573 genotyping primers:

tm573_external_fwd: CCGAGCTGGATAGGGAG

tm573_external_rev: CAGGACAAGCAGAGGAAATC

Molecular characterization of the *lsy-27(tm593)* allele

We determined the precise nature of the *lsy-27(tm593)* deletion through Sanger sequencing and find that it is a 361 bp deletion with one T inserted instead :

Wild type: ...tttttggttaagagcccgctc [361 base pairs] ttgagaacttttcaagctt...

tm593 ...tttttggttaagagcccgctc T ttgagaacttttcaagctt...

To analyze the transcriptional product made in *tm593* animals, we performed RT-PCR analysis with the Invitrogen Superscript one-step RT-PCR System with Platinum Taq Polymerase using primers located at the 5' and 3' end of the coding sequence:

F47H4.1_RT_A: CGTCTATTCACAGTCCAGTCAC

F47H4.1_RT_B: GAATCATCTTGCTCTCGACG

Individual bands were then gel purified and Sanger sequenced. We detected three different transcripts of different length, all starting with the first, unaffected exon, but then reading in various different ways into the first intron before splicing into downstream exons. Each transcript contains premature stop codons and encode severely truncated versions of the protein (23 aa, 35 aa, and 37 aa long), none of which contain any of the DNA-binding Zn finger domains.

Unexpectedly, we detected a PCR product with two primers that are entirely located within the deletion. We suspect that the deleted DNA has inserted elsewhere in the genome but emphasize that the RT-PCR analysis described above suggests that no functional product is present.

Table S1 Whole genome sequencing setting and results

Parameter	Setting/Results
Read length	75
Number of lanes on flow cells	3
Average coverage	31.5
Size of genetically defined interval	13 Mb
Total variants in interval compared to wild-type reference genome on right arm of LGV	961
Noncoding (Intergenic/intronic/silent/ncRNA/SNP)	939
Splice junction/missense/nonsense	22
Total variants minus strain background variants *	3

Variants were considered as background if they were also found in other WGS datasets from our lab (SARIN *et al.* 2010).

Table S2 Mutant analysis of *lsy-27* paralogues

Genotype	% of animals expressing <i>gcy-5</i> (<i>ntl-1</i>) at			n
	25°C			
	ASEL=ASE	ASEL<ASE	ASER only	
wild type	0%	0%	>100	>100
<i>lsy-27(ot108)</i>	39%	48%	13%	122
<i>lsy-27(tm593)</i>	0%	62%	38%	117
<i>ztf-25(tm610)</i>	0%	0%	100%	46
<i>ztf-28(tm573)</i>	0%	0%	100%	28
<i>lsy-27(tm593); ztf-28(tm573)</i>	0%	49%	51%	168
<i>ztf-25(tm610); ztf-28(tm573)</i>	0%	0%	100%	78
<i>ztf-25(tm610); ztf-28(tm573); lsy-27(RNAi)</i> ¹	0%	0%	100%	32

¹ Triple couldn't be built due to close linkage of *lsy-27* and *ztf-25*. Wild-type, *ot108* and *tm593* data is repeated from Table 1 for comparison purposes.

Table S3 Retired and novel *lsy* gene names

Original assignment ¹	Allele	novel assignment	Reference
<i>lsy-5</i>	<i>ot37, ot240</i>	<i>unc-37</i>	(FLOWERS <i>et al.</i> 2010)
<i>lsy-14</i>	<i>ot101</i>	<i>che-1</i>	(ETCHBERGER <i>et al.</i> 2009)
	<i>ot146</i>	<i>lim-6</i>	this paper
<i>lsy-16</i>	<i>ot158</i>	<i>nhr-67</i>	(SARIN <i>et al.</i> 2009)
<i>lsy-17</i>	<i>ot190</i>	<i>nhr-67</i>	(SARIN <i>et al.</i> 2009)
<i>lsy-18</i>	<i>ot192</i>	<i>fozi-1</i>	unpubl. data ²
<i>lsy-19</i>	<i>ot177</i>	<i>lsy-12</i>	(O'MEARA <i>et al.</i> 2010)
<i>die-1</i>	<i>ot108</i>	<i>lsy-27</i>	this paper

¹ as shown in (SARIN *et al.* 2007)

² *ot192* was initially thought to be a distinct locus based on mapping results that were misleading likely due to incompatibilities between N2 Bristol and the Hawaiian mapping strain. The *ot192* mutation is a C>T change that results in a premature stop codon in *fozi-1* (Q549Stop). The same mutation is found in *ot191* animals (SARIN *et al.* 2007) and we cannot exclude the possibility that these two mutations arose from the same parent ("jackpot" mutation).

Table S4 Final summary of mutant classes & genes

Mutant class	ASE phenotype	# of genes	Gene names	Molecular identity	# of alleles	Reference
Class I	"2 ASEL"	3	<i>cog-1</i>	homeobox	19	(CHANG <i>et al.</i> 2003; SARIN <i>et al.</i> 2007)
			<i>unc-37</i>	tsk. co-factor	4	(CHANG <i>et al.</i> 2003; FLOWERS <i>et al.</i> 2010; SARIN <i>et al.</i> 2007)
			<i>lsy-22</i>	tsk. co-factor	2	(FLOWERS <i>et al.</i> 2010)
Class II	"2 ASER"	7	<i>die-1</i>	C2H2 Zn finger TF	8	(CHANG <i>et al.</i> 2004; SARIN <i>et al.</i> 2007)
			<i>lsy-2</i>	C2H2 Zn finger TF	6	(JOHNSTON and HOBERT 2005; SARIN <i>et al.</i> 2007)
			<i>lsy-12</i>	MYST HAT	6	(O'MEARA <i>et al.</i> 2010; SARIN <i>et al.</i> 2007)
			<i>lsy-6</i>	miRNA	4	(JOHNSTON and HOBERT 2003; SARIN <i>et al.</i> 2007)
			<i>lin-49</i>	tsk. co-factor	3	(CHANG <i>et al.</i> 2003; SARIN <i>et al.</i> 2007)
			<i>lsy-15</i>	WD40	1	(SARIN <i>et al.</i> 2007), Poole <i>et al.</i> submitted
			<i>ceh-36</i>	Homeobox	1	(CHANG <i>et al.</i> 2003; SARIN <i>et al.</i> 2007)
Class III	no ASEL/R fate specification	1	<i>che-1</i>	C2H2 Zn finger TF	24	(ETCHBERGER <i>et al.</i> 2009; SARIN <i>et al.</i> 2007)
Class IV	mixed fate in ASEL or ASER	5	<i>fozi-1</i>	C2H2 Zn finger TF	13	(JOHNSTON <i>et al.</i> 2006; SARIN <i>et al.</i> 2007)
			<i>lsy-20</i>	unknown	1	(SARIN <i>et al.</i> 2007)
			<i>lin-59 (prev. lsy-26)</i>	SET domain	1	(SARIN <i>et al.</i> 2010)
			<i>lim-6</i>	LIM homeobox	1	(HOBERT <i>et al.</i> 1999), this paper
			<i>lsy-27</i>	C2H2 Zn finger TF	1	this paper
Class V	heterogeneous phenotype	1	<i>nhr-67</i>	C4 Zn finger TF	7	(SARIN <i>et al.</i> 2009)

This is an updated version of the mutant summary table from (SARIN *et al.* 2007). Mutants with <10% penetrance (e.g. *lsy-21*, (SARIN *et al.* 2007)) are not shown. Only class I to V *lsy* mutants are shown. Class VI do not affect ASE per se, but affect other cells. See also Table S3 for altered gene assignments, as compared to (SARIN *et al.* 2007).

References

- CHANG, S., R. J. JOHNSTON, C. FROKJAER-JENSEN, S. LOCKERY and O. HOBERT, 2004 MicroRNAs act sequentially and asymmetrically to control chemosensory laterality in the nematode. *Nature* **430**: 785-789.
- CHANG, S., R. J. JOHNSTON, JR. and O. HOBERT, 2003 A transcriptional regulatory cascade that controls left/right asymmetry in chemosensory neurons of *C. elegans*. *Genes Dev* **17**: 2123-2137.
- ETCHBERGER, J. F., E. B. FLOWERS, R. J. POOLE, E. BASHLLARI and O. HOBERT, 2009 Cis-regulatory mechanisms of left/right asymmetric neuron-subtype specification in *C. elegans*. *Development* **136**: 147-160.
- FLOWERS, E. B., R. J. POOLE, B. TURSUN, E. BASHLLARI, I. PE'ER *et al.*, 2010 The Groucho ortholog UNC-37 interacts with the short Groucho-like protein LSY-22 to control developmental decisions in *C. elegans*. *Development* **137**: 1799-1805.
- HOBERT, O., 2002 PCR fusion-based approach to create reporter gene constructs for expression analysis in transgenic *C. elegans*. *Biotechniques* **32**: 728-730.
- HOBERT, O., K. TESSMAR and G. RUVKUN, 1999 The *Caenorhabditis elegans* *lim-6* LIM homeobox gene regulates neurite outgrowth and function of particular GABAergic neurons. *Development* **126**: 1547-1562.
- JOHNSTON, R. J., and O. HOBERT, 2003 A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans*. *Nature* **426**: 845-849.
- JOHNSTON, R. J., JR., J. W. COPELAND, M. FASNACHT, J. F. ETCHBERGER, J. LIU *et al.*, 2006 An unusual Zn-finger/FH2 domain protein controls a left/right asymmetric neuronal fate decision in *C. elegans*. *Development* **133**: 3317-3328.
- JOHNSTON, R. J., JR., and O. HOBERT, 2005 A novel *C. elegans* zinc finger transcription factor, *Isy-2*, required for the cell type-specific expression of the *Isy-6* microRNA. *Development* **132**: 5451-5460.
- O'MEARA, M. M., F. ZHANG and O. HOBERT, 2010 Maintenance of neuronal laterality in *Caenorhabditis elegans* through MYST histone acetyltransferase complex components LSY-12, LSY-13 and LIN-49. *Genetics* **186**: 1497-1502.
- PAVLETICH, N. P., and C. O. PABO, 1991 Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* **252**: 809-817.
- SARIN, S., C. ANTONIO, B. TURSUN and O. HOBERT, 2009 The *C. elegans* Tailless/TLX transcription factor *nhr-67* controls neuronal identity and left/right asymmetric fate diversification. *Development* **136**: 2933-2944.
- SARIN, S., V. BERTRAND, H. BIGELOW, A. BOYANOV, M. DOITSIDOU *et al.*, 2010 Analysis of multiple ethyl methanesulfonate-mutagenized *caenorhabditis elegans* strains by whole-genome sequencing. *Genetics* **185**: 417-430.
- SARIN, S., M. O'MEARA M, E. B. FLOWERS, C. ANTONIO, R. J. POOLE *et al.*, 2007 Genetic Screens for *Caenorhabditis elegans* Mutants Defective in Left/Right Asymmetric Neuronal Fate Specification. *Genetics* **176**: 2109-2130.
- TURSUN, B., L. COCHELLA, I. CARRERA and O. HOBERT, 2009 A toolkit and robust pipeline for the generation of fosmid-based reporter genes in *C. elegans*. *PLoS ONE* **4**: e4625.

CHAPTER 5:

Maintenance of neuronal laterality in *Caenorhabditis elegans* through MYST histone acetyltransferase complex components LSY-12, LSY-13 and LIN-49

O'Meara MM, **Zhang F**, Hobert O. Genetics. 2010 Dec;186(4):1497-502.

This chapter describes the cloning and characterization of the MYST family histone acetyltransferase *lsy-12*. Loss of *lsy-12* leads to complete conversion of the ASEL cell fate to the ASER fate without affecting any bilateral features. LSY-12 forms a complex with the PHD/bromodomain protein LIN-49 and the ING-family PHD domain protein LSY-13, which also affect ASE laterality when mutated. Post-embryonic temperature shift experiments with a *ts* allele of *lsy-12*, *ot563*, demonstrates that LSY-12 is required continuously. The ASE master regulator CHE-1 likely cooperates with DNA-binding proteins such as DIE-1 to control asymmetric gene expression and is also required throughout development. Post-developmental dsRNA treatment directed against *die-1* results in ASEL to ASER conversion. Therefore, it is likely that CHE-1 and DIE-1 recruit the MYST-HAT complex to regulatory sites to maintain terminal differentiated neuronal features of ASEL.

In this paper, I conducted genetic analysis, rescue experiments related to the ING-like genes *lsy-13* and *ing-3*, and constructed the GFP reporter for *lsy-13*. I also analyzed the effects of *lsy-12* and *lin-49* mutants on *die-1*. Maggie O'Meara cloned and characterized *lsy-12* and performed temperature-shift experiments.

Note

Maintenance of Neuronal Laterality in *Caenorhabditis elegans* Through MYST Histone Acetyltransferase Complex Components LSY-12, LSY-13 and LIN-49

M. Maggie O'Meara, Feifan Zhang and Oliver Hobert¹

Howard Hughes Medical Institute, Department of Biochemistry and Molecular Biophysics,
Columbia University Medical Center, New York, New York 10032

Manuscript received August 10, 2010
Accepted for publication October 2, 2010

ABSTRACT

Left/right asymmetrically expressed genes permit an animal to perform distinct tasks with the right *vs.* left side of its brain. Once established during development, lateralized gene expression patterns need to be maintained during the life of the animal. We show here that a histone modifying complex, composed of the LSY-12 MYST-type histone acetyltransferase, the ING-family PHD domain protein LSY-13, and PHD/bromodomain protein LIN-49, is required to first initiate and then actively maintain lateralized gene expression in the gustatory system of the nematode *Caenorhabditis elegans*. Similar defects are observed upon postembryonic removal of two C2H2 zinc finger transcription factors, *die-1* and *che-1*, demonstrating that a combination of transcription factors, which recognize DNA in a sequence-specific manner, and a histone modifying enzyme complex are responsible for inducing and maintaining neuronal laterality.

FEATURES of terminally differentiated cells not only need to be initiated through cell type-specific regulatory programs but also need to be continuously maintained throughout the life of the cell (BLAU 1992). The mechanisms that maintain the identity of post-mitotic, fully differentiated cells are, however, incompletely understood. We provide here some insights into these mechanisms, using a specific differentiation event in the nervous system of *Caenorhabditis elegans*. This neuronal differentiation event occurs differentially across the left/right axis of the animal and results in the left/right asymmetric expression of putative chemoreceptors (*gcy* genes) in the left *vs.* right ASE gustatory neuron (Figure 1A). This laterality, *i.e.*, the left/right asymmetric expression of the *gcy* genes, is controlled by several gene regulatory factors that act in a bistable feedback loop (Figure 1, A and B) (JOHNSTON and HOBERT 2003; JOHNSTON *et al.* 2005; HOBERT 2006). Components of the bistable feedback loop are only transiently required around the time the ASE neurons are born in the embryo (SARIN *et al.* 2007). Postembryonic mechanisms that robustly and repro-

ducibly maintain laterality throughout the life of the animal are unknown, yet must exist, considering the importance of ASE laterality for adult nervous system function (SUZUKI *et al.* 2008). We describe here a locus, *lsy-12*, that executes such a maintenance function.

Laterality defects in animals lacking *lsy-12*, a histone acetyltransferase: *lsy-12* mutants were initially identified in a screen for ASEL/R laterality defects (*Lsy* phenotype) (SARIN *et al.* 2007). We characterized the *lsy-12* defects in more detail by examining a panel of left/right asymmetrically expressed genes. We find that *lsy-12* broadly affects lateralized gene expression; terminal markers (such as the *gcy* chemoreceptors), as well as upstream regulators, such as the miRNA *lsy-6* or the *die-1* zinc finger transcription factors, are affected (Figure 1C). All the defects sum up to a complete conversion of the ASEL state to the ASER state (“2-ASER” phenotype) (Figure 1B). Bilateral ASE fate is unaffected, as assessed by correct, bilateral expression of the Otx-type *ceh-36* homeobox gene (Figure 1C).

lsy-12 was cloned in parallel by two independent strategies, one being classic fine mapping and ensuing transformation rescue (supporting information, Figure S1). The other strategy involved whole genome sequencing (SARIN *et al.* 2008). Both approaches showed that *lsy-12* corresponds to the previously uncharacter-

Supporting information is available online at <http://www.genetics.org/cgi/content/full/genetics.110.123661/DC1>.

¹Corresponding author: Columbia University, 701 W. 168th St., HHSC 724, New York, NY 10032. E-mail: or38@columbia.edu

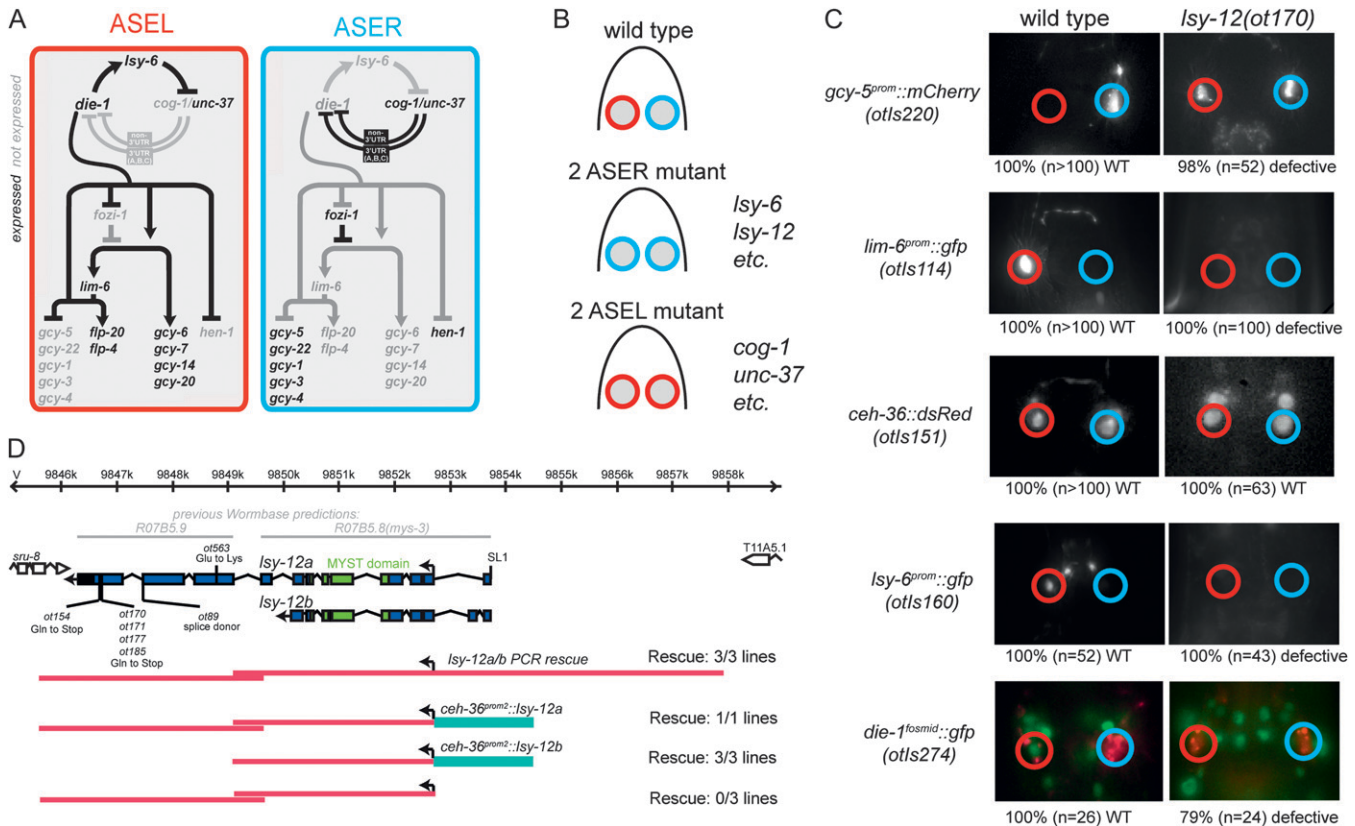


FIGURE 1.—*lsy-12*, a MYST-type histone acetyltransferase, affects ASE laterality. (A) Genes known to be involved in controlling ASEL/R laterality (HOBERT 2006; DIDIANO *et al.* 2010). (B) Schematic representation of phenotype of representative “2 ASEL” and “2 ASER” mutants. (C) Effects of *lsy-12* on ASEL/R laterality markers. A subset of the defects were already reported, upon the initial identification of the *lsy-12* locus (SARIN *et al.* 2007). Numbers below the panels indicate the penetrance of the phenotype, *i.e.*, the fraction of animals that display the phenotype shown in the fluorescent image above. Data on other *lsy-12* alleles were reported in SARIN *et al.* (2007, 2008, 2010). Animals that express the *die-1* reporter fosmid also contain a ASEL/R-expressed red fluorescent reporter (*che-1::mCherry*) for cell identification. A list of transgenes used in the study is provided in the File S1. Animals were scored as adults. (D) *lsy-12* encodes a MYST-type histone acetyltransferase. Analysis of ESTs, expression tiling array clones, PCR specific rescue, and RT-PCR (see File S1 for more details) revealed that R07B5.8 and R07B5.9 are one genetic locus encoding at least two major splice isoforms. The 3', polyadenylated end of yk82d06 and EX1785569 provide evidence for the existence of the *lsy-12b*, while other clones provide evidence of the *lsy-12a* isoform. RT-PCR data suggest that additional splice variants may be produced by the *lsy-12* locus (some possibly even including the more upstream predicted gene T11A5.1), but we have not been able to conclusively identify the start and end of such alternative transcripts (data not shown; see also www.wormbase.org). The arrow indicates the predicted translational start site. *lsy-12* expression constructs are shown in the lower part of the panel, with the bottom one being a negative control. Staggered red lines indicate that these constructs were generated by *in vivo* recombineering of co-injected, overlapping PCR fragments (BOULIN *et al.* 2006), some of which were generated by an *in vitro* PCR fusion approach (HOBERT 2002). The generation of constructs is described in File S1. Rescuing data is quantified in Table S1.

ized R07B5.9 locus, a notion confirmed by multiple alleles of *lsy-12* each showing molecular lesions in R07B5.9, rescue experiments, and phenocopy of the *lsy-12* defects by RNAi against R07B5.9 (Figure 1D, Table S1) (SARIN *et al.* 2008).

The R07B5.9 protein, as originally predicted in WormBase (WS215), codes for a 657-amino-acid protein with some limited homology to the C-terminal half of MYST-type histone acetyltransferases, yet it contains no predicted acetyltransferase domain or any other protein domain present in the databases. However, we noted that the gene predicted by WormBase to reside upstream of *lsy-12* (R07B5.8) encodes a protein with a

MYST-type histone acetyltransferase domain (previously called *mys-3*; CEOL and HORVITZ 2004), which was not previously characterized with mutant alleles. Sequencing of available EST clones (kindly provided by Yuji Kohara) as well as our own RT-PCR analysis reveals that R07B5.9 and R07B5.8 were mispredicted as separate genes and that the downstream predicted gene R07B5.9 is always fused to the upstream R07B5.8 gene (we now refer to this gene as the “a” form of the *lsy-12* locus) (Figure 1D). R07B5.8 in turn, is not always fused to R07B5.9 and can produce a smaller splice form with an alternative 3'-UTR (we refer to this variant now as the “b” form of the *lsy-12* locus) (Figure 1D). The start site

of *lsy-12a* and *lsy-12b* is confirmed by the presence of SL1 splice leader sequences. All six mutant alleles of *lsy-12* locate to the long variant, *lsy-12a* (Figure 1D). None of the available *lsy-12* alleles are unambiguous molecular nulls (Figure 1D). Attempts to retrieve such alleles by transposon mobilization have failed.

lsy-12 encodes one of four MYST-type histone acetyltransferase (HAT) proteins in the *C. elegans* genome (CEOL and HORVITZ 2004). Other MYST family members have previously been implicated in vulval and ectodermal patterning (CEOL and HORVITZ 2004; SHIBATA *et al.* 2010). On the basis of overall sequence homology, *lsy-12*/*mys-3* (from here on referred to as *lsy-12*) and *mys-4* are both members of the MOZ/MORF subfamily of MYSTs (LEE and WORKMAN 2007). A deletion allele of *mys-4* does not display a *Lsy* phenotype (data not shown).

Reporters of the *lsy-12* locus that only contain upstream regulatory information showed restricted expression patterns and no expression in ASE (data not shown), likely because not all regulatory elements of this large locus are located in the 5' upstream region. A fosmid-based *gfp* reporter rescues the *lsy-12* mutant phenotype, and is expressed broadly throughout the animal; yet its expression was too weak to unambiguously assess expression in ASEL/R. To assess whether *lsy-12* indeed acts in the ASE neurons, we expressed the *lsy-12* coding region under control of the bilateral ASE promoter from the *ceh-36* locus and found that this construct rescued the *Lsy* phenotype, demonstrating that *lsy-12* acts cell autonomously within the ASE neuron class (Figure 1D; Table S1).

***lsy-12* is continuously required to maintain ASE laterality:** Examining the onset of left/right asymmetric gene expression in the ASE neurons in the embryo, we find that in *lsy-12* mutants the normally ASER-specific *gcy-5* gene is expressed bilaterally from the onset of its expression in threefold embryos (Figure 2A). To address whether *lsy-12* may not be involved only in the initial establishment of asymmetry but also in maintaining ASE asymmetry, we used an allele of *lsy-12*, *ot563*, which is strongly temperature sensitive. This allele was retrieved as a modifier of the *lin-59* locus but was not yet characterized (SARIN *et al.* 2010). We find that animals continuously raised at 15° show a very lowly penetrant *Lsy* phenotype (~10%), while animals continuously raised at 25° show an almost completely penetrant *Lsy* phenotype (Figure 2B). Animals shifted from the permissive temperature (15°) to the nonpermissive temperature (25°) at the postembryonic L4 stage or even the adult stage (*i.e.*, long after ASE laterality has been established in the embryo) show an ~50% penetrant ASEL-to-ASER conversion (Figure 2B). Similarly, animals grown at 25° until late larval or adult stages (at which animals would normally display an almost completely penetrant ASEL-to-ASER conversion), show a partial rescue of the mutant phenotype when shifted to 15° (Figure 2B). These findings underscore the plasticity and bipotentiality of

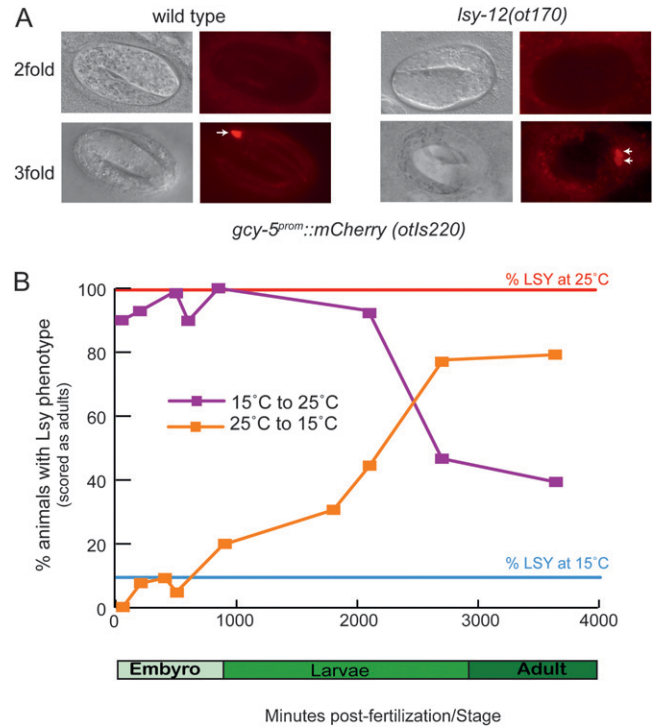


FIGURE 2.—*lsy-12* is continually required in ASE but may also act early. (A) *lsy-12* is required for the initial manifestation of asymmetry, as assessed by *gcy-5::gfp* expression (*otIs220* transgene). In wild-type animals, *gcy-5* expression is first observed exclusively in ASER in threefold-stage embryos; in *lsy-12(ot563)* animals, expression of *gcy-5::gfp* is bilateral from the onset. (B) Temperature-shift experiments indicate a sustained requirement for *lsy-12* activity. *lsy-12(ot563)* animals that had been grown for several generations at either 15° or 25° were plated and temperature shifted up to 25° or down to 15° at the following timepoints: Embryo, 2-cell; pre-comma; two- to threefold; threefold (those embryos were collected from dissected adult) and postembryonic, L1; L2; L3; L4; and 2-day-old adult. All animals were then scored as 3-day-old adults.

the system in that it can revert from one state to the other even after the initial choice has been made.

Other known components of MOZ/MORF-type HATs also display *Lsy* phenotypes: Work in other systems has revealed three proteins that functionally associate with MYST/LSY-12-type HATs namely the EAF6 protein, the BRPF1 bromodomain protein, and a PHD finger protein of the ING family (DOYON *et al.* 2006; LEE and WORKMAN 2007; ULLAH *et al.* 2008). There are *C. elegans* orthologs for each of the three proteins. Mutant alleles are available for the BRPF1 ortholog *lin-49* and the ING-like genes *T06A10.4* and *ing-3*. We had previously reported that mutant alleles in one of them, the bromodomain-encoding *lin-49* locus, display *Lsy* defects (CHANG *et al.* 2003). These defects are the same as in *lsy-12* mutants: the fate of the ASEL neuron converts to that of the ASER neuron (CHANG *et al.* 2003). We extended this previous observation by showing that not just terminal fate (as shown in CHANG *et al.* 2003) is affected in *lin-49* mutants, but that *lin-49*, like *lsy-12*, also affects the activity of the bistable feedback loop: ASEL-specific

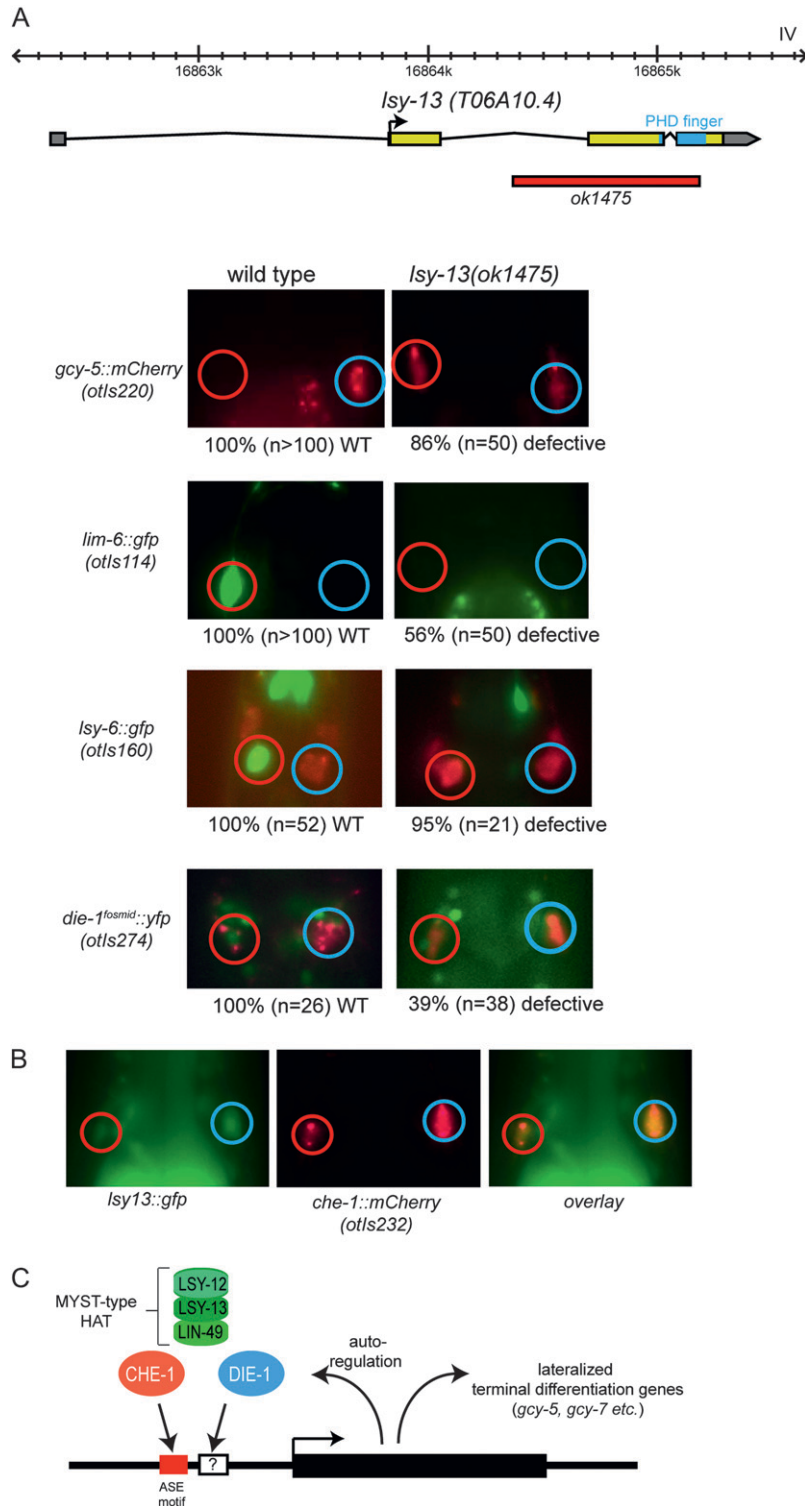


FIGURE 3.—The MYST-complex component *lsy-13* affects ASE laterality. (A) *lsy-13* controls ASE laterality. The upper panel shows the structure of the *lsy-13* locus and the *lsy-13* null allele. The gene structure is confirmed by EST clones (www.wormbase.org). All other panels show the head regions of adult animals. Numbers below the panels indicate the penetrance of the phenotype, *i.e.* the fraction of animals that display the phenotype shown in the fluorescent image above. The *die-1* and *lsy-6* wild-type control images are the same as in other figures and shown for comparison only. The red fluorescent marker in the *lsy-6* panel (*ceh-36^{prom}::dsRed2*) allows identification of the ASE neurons. Animals that express the *die-1* reporter fosmid also contain a *ASEL/R*-expressed red fluorescent reporter (*che-1::mCherry*). A list of transgenes used in the study is provided in File S1. We note that *lsy-13* function can be maternally supplied (homozygous offspring of a heterozygous *lsy-13* parent does not display a mutant phenotype). In contrast to the *lin-49* null mutant animals, *lsy-13(ok1475)* null mutant animals are viable and display no obvious morphological abnormalities. (B) A reporter gene which contains 2.8 kb of 5' sequences to the first exon of *lsy-13*, generated by PCR fusion (HOBERT 2002), is broadly expressed, including in the two ASE neurons, marked with a red fluorescent reporter gene. Primer sequences for the construct are provided in File S1. (C) Model for *lsy-12/lsy-13/lin-49* function, based on the phenotypic similarities between the genes shown here. *che-1* directly regulates expression of terminal differentiation genes—both symmetrically and asymmetrically expressed ones—as well as regulators of the bistable feedback loop (ETCHBERGER *et al.* 2007, 2009). Left/right asymmetrically expressed genes, contain *cis*-regulatory elements (indicated by “?”) in addition to the CHE-1 binding site (the ASE motif) that restrict CHE-1 activity to *ASEL* or *ASER* (ETCHBERGER *et al.* 2009). Genetically, *die-1* controls the activity of the factors that restrict *che-1* activity, but it is not known whether *die-1* fulfills this function directly (through binding to these additional motifs) or indirectly through the regulation of other factors. Since *die-1* auto-regulates its own transcription (L. Cochella and O. Hobert, unpublished data), the HAT complex also impinges on *die-1* expression itself.

lsy-6 miRNA and *die-1* expression is lost in *lin-49* mutants (Figure S2A). Moreover, the laterality defects of a hypomorphic allele of *lsy-12* allele are enhanced by *lin-49* (Figure S2B; since stronger alleles of *lsy-12* are completely penetrant, other interaction tests of this sort could not be performed).

T06A10.4 and *ing-3* encode PHD domain-containing proteins related to the ING subunit of MYST-type HAT

complexes, with *T06A10.4* being closer to the *ING1/2/4/5* genes (<http://www.treefam.org/cgi-bin/TFinfo.pl?ac=TF352014>) and *ing-3* being closer to the *ING3* gene (<http://www.treefam.org/cgi-bin/TFinfo.pl?ac=TF106497>). The PHD domain is thought to recognize histone methylation marks and thereby recruit HAT activity to methylated histone substrates (DOYON *et al.* 2006; ULLAH *et al.* 2008). Even though ING-like

proteins have been shown to be involved in cell growth, apoptosis, and tumorigenesis (COLES and JONES 2009), their role in pattern formation and cell fate specification have been little explored. Deletion alleles of *T06A10.4* and *ing-3* were generated by the *C. elegans* knockout consortia. *ing-3(tm2530)* mutants display no Lsy phenotype (data not shown). In contrast, *T06A10.4(ok1475)* mutant animals display a “2 ASER” Lsy phenotype, such that expression of a terminal ASER marker is lost while there is ectopic gain of a terminal ASER marker in ASER (Figure 3A). This phenotype is indistinguishable from *lsy-12/HAT* and *lin-49/BRPF* mutants. Again similar to *lsy-12* and *lin-49* mutants, the expression of the ASER inducers *lsy-6* and *die-1* is also lost in *T06A10.4(ok1475)* mutants, albeit at lower penetrance as observed in *lsy-12* mutants. The Lsy phenotype of *T06A10.4(ok1475)* mutant animals can be rescued by a 4.4-kb PCR fragment spanning the entire locus [2.3 kb upstream of start codon to 0.7 kb downstream of stop; the 86% penetrant *gcy-5* misexpression ($n = 50$) is rescued to 0% misexpression ($n = 28$)]. We therefore named the *T06A10.4* locus *lsy-13*. A reporter construct for the *lsy-13* locus displays a broad expression pattern throughout many tissue types, including expression in the two ASE neurons (Figure 3B), therefore resembling the broad expression of *lin-49* (CHAMBERLIN *et al.* 1999).

Taken together, the phenotypic similarity of *lsy-12*, *lsy-13*, and *lin-49* in controlling terminal ASER/R fate, together with the reported physical and functional interactions of their vertebrate homologs (DOYON *et al.* 2006; ULLAH *et al.* 2008), suggest that LSY-12, LSY-13, and LIN-49 proteins act together in a complex to control ASER/R lateralization. We note that *lsy-13* is the only ING-like gene with a reported role in nervous system development and our studies provide the first phenotypic side-by-side comparisons of HAT and ING gene activities in a metazoan organism, thereby providing *in vivo* support for the biochemical studies that link these two proteins (DOYON *et al.* 2006; ULLAH *et al.* 2008).

The *die-1* Zn finger transcription factor is also continuously required to maintain ASE laterality and is a candidate recruiter of the MYST complex: How could the maintenance function of the MYST complex be explained? Generally, the phenotypic specificity of histone-modifying enzymes must be conferred by transcription factors that recognize DNA in a sequence-specific manner (STRUHL 1998). For example, the MYST-type HAT Tip60 is recruited to DNA via diverse transcription factors, such as nuclear hormone receptors or c-Myc (SAPOUNTZI *et al.* 2006). In the context of ASE laterality control, the C2H2 *die-1* Zn finger transcription factor may be such a recruiter. This is because, first, the *lsy-12*, *lsy-13*, and *lin-49* phenotypes described here resemble those of the *die-1* Zn finger transcription factor in that mutant alleles in all these loci display an ASER-to-ASER

fate conversion (“2 ASER” phenotype) (CHANG *et al.* 2004). Second, both *die-1* and *lsy-12* are continuously required to maintain ASE laterality. In the case of *lsy-12*, this is demonstrated by the temperature-shift experiments described above; in the case of *die-1*, we uncovered such requirement through postdevelopmental treatment of animals with dsRNA directed against *die-1*. Using a *nre-1 lin-15b* RNAi hypersensitive background (SCHMITZ *et al.* 2007) and the ASER-expressed *gcy-5^{prom}::gfp* transgene *otIs186*, we observed an ASER-to-ASER conversion in the P0 generation of dsRNA treated animals [45% ($n = 62$) of animals showed such a conversion]. The maintained requirement of *die-1* is also illustrated by the maintained expression of *die-1* in ASER throughout larval and adult stages (Figure 1C).

The *che-1* C2H2 Zn finger transcription factor, a terminal selector for ASE fate, which acts through a *cis*-regulatory motif, the ASE motif, present in bilaterally and left/right asymmetrically expressed terminal differentiation genes, is, like *die-1*, also continuously required to maintain the differentiated state of the ASE neurons (ETCHBERGER *et al.* 2007, 2009). In the case of asymmetrically expressed genes, CHE-1 cooperates with additional DNA-binding proteins—possibly DIE-1—to ensure left/right asymmetric expression (ETCHBERGER *et al.* 2009). We therefore propose that continuously required CHE-1 and DIE-1 are the sequence-specific DNA binding proteins that recruit the LSY-12/LSY-13/LIN-49 MYST–HAT complex to maintain terminal differentiation features (Figure 3C).

A role in maintaining differentiated cellular states has also been recently reported for another MYST–HAT complex, composed of the *C. elegans* Bromodomain protein BET-1 (a paralog of LIN-49) and its associated MYST-type histone acetyltransferase *myst-2* (a paralog of *lsy-12*) (SHIBATA *et al.* 2010). The vertebrate LIN-49 homolog BRPF1 and the histone acetyltransferase Myst3, as well as the fly MYST family member Chameau, are required to maintain HOX gene expression during development (GRIENENBERGER *et al.* 2002; LAUE *et al.* 2008), and the vertebrate MYST family member MOZ is required for the generation and maintenance of hematopoietic stem cells (KATSUMOTO *et al.* 2008). Together, these findings suggest that MYST function in maintenance of gene expression patterns has been broadly conserved during evolution and is employed in many different cell types.

We thank Baris Tursun for the *die-1* fosmid reporter line, Sumeet Sarin for the *ot563* allele, Richard J. Poole and Enkelejda Bashllari for sharing the first evidence of a *die-1* maintenance role, Luisa Cochella for communicating unpublished results, the *C. elegans* Gene Knockout Consortia in Oklahoma and Tokyo for the *ok1475* and *tm2530* alleles, Yuji Kohara for EST clones, and Qi Chen for expert assistance in generating transgenic strains. We thank members of the Hobert lab for comments on the manuscript. This work was funded by the National Institutes of Health (R01NS039996-05; R01NS050266-03). O.H. is an investigator of the Howard Hughes Medical Institute.

LITERATURE CITED

- BLAU, H. M., 1992 Differentiation requires continuous active control. *Annu. Rev. Biochem.* **61**: 1213–1230.
- BOULIN, T., J. F. ETCHBERGER and O. HOBERT, 2006 Reporter gene fusions. *WormBook*, 1–23.
- CEOL, C. J., and H. R. HORVITZ, 2004 A new class of *C. elegans* synMuv genes implicates a Tip60/NuA4-like HAT complex as a negative regulator of Ras signaling. *Dev. Cell* **6**: 563–576.
- CHAMBERLIN, H. M., K. B. BROWN, P. W. STERNBERG and J. H. THOMAS, 1999 Characterization of seven genes affecting *Caenorhabditis elegans* hindgut development. *Genetics* **153**: 731–742.
- CHANG, S., R. J. JOHNSTON, C. FROKJAER-JENSEN, S. LOCKERY and O. HOBERT, 2004 MicroRNAs act sequentially and asymmetrically to control chemosensory laterality in the nematode. *Nature* **430**: 785–789.
- CHANG, S., R. J. JOHNSTON, JR. and O. HOBERT, 2003 A transcriptional regulatory cascade that controls left/right asymmetry in chemosensory neurons of *C. elegans*. *Genes Dev.* **17**: 2123–2137.
- COLES, A. H., and S. N. JONES, 2009 The ING gene family in the regulation of cell growth and tumorigenesis. *J. Cell Physiol.* **218**: 45–57.
- DIDIANO, D., L. COCHELLA, B. TURSUN and O. HOBERT, 2010 Neuron-type specific regulation of a 3'UTR through redundant and combinatorially acting *cis*-regulatory elements. *RNA* **16**: 349–363.
- DOYON, Y., C. CAYROU, M. ULLAH, A. J. LANDRY, V. COTE *et al.*, 2006 ING tumor suppressor proteins are critical regulators of chromatin acetylation required for genome expression and perpetuation. *Mol. Cell* **21**: 51–64.
- ETCHBERGER, J. F., A. LORCH, M. C. SLEUMER, R. ZAPP, S. J. JONES *et al.*, 2007 The molecular signature and *cis*-regulatory architecture of a *C. elegans* gustatory neuron. *Genes Dev.* **21**: 1653–1674.
- ETCHBERGER, J. F., E. B. FLOWERS, R. J. POOLE, E. BASHLLARI and O. HOBERT, 2009 *Cis*-regulatory mechanisms of left/right asymmetric neuron-subtype specification in *C. elegans*. *Development* **136**: 147–160.
- GRIENENBERGER, A., B. MIOTTO, T. SAGNIER, G. CAVALLI, V. SCHRAMKE *et al.*, 2002 The MYST domain acetyltransferase Chameau functions in epigenetic mechanisms of transcriptional repression. *Curr. Biol.* **12**: 762–766.
- HOBERT, O., 2002 PCR fusion-based approach to create reporter gene constructs for expression analysis in transgenic *C. elegans*. *Biotechniques* **32**: 728–730.
- HOBERT, O., 2006 Architecture of a microRNA-controlled gene regulatory network that diversifies neuronal cell fates. *Cold Spring Harb. Symp. Quant. Biol.* **71**: 181–188.
- JOHNSTON, R. J., and O. HOBERT, 2003 A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans*. *Nature* **426**: 845–849.
- JOHNSTON, JR., R. J., S. CHANG, J. F. ETCHBERGER, C. O. ORTIZ and O. HOBERT, 2005 MicroRNAs acting in a double-negative feedback loop to control a neuronal cell fate decision. *Proc. Natl. Acad. Sci. USA* **102**: 12449–12454.
- KATSUMOTO, T., N. YOSHIDA and I. KITABAYASHI, 2008 Roles of the histone acetyltransferase monocytic leukemia zinc finger protein in normal and malignant hematopoiesis. *Cancer Sci.* **99**: 1523–1527.
- LAUE, K., S. DAUJAT, J. G. CRUMP, N. PLASTER, H. H. ROEHL *et al.*, 2008 The multidomain protein Brpf1 binds histones and is required for Hox gene expression and segmental identity. *Development* **135**: 1935–1946.
- LEE, K. K., and J. L. WORKMAN, 2007 Histone acetyltransferase complexes: one size doesn't fit all. *Nat. Rev. Mol. Cell Biol.* **8**: 284–295.
- SAPOUNTZI, V., I. R. LOGAN and C. N. ROBSON, 2006 Cellular functions of TIP60. *Int. J. Biochem. Cell Biol.* **38**: 1496–1509.
- SARIN, S., M. O'MEARA, E. B. FLOWERS, C. ANTONIO, R. J. POOLE *et al.*, 2007 Genetic screens for *Caenorhabditis elegans* mutants defective in left/right asymmetric neuronal fate specification. *Genetics* **176**: 2109–2130.
- SARIN, S., S. PRABHU, M. M. O'MEARA, I. PE'ER and O. HOBERT, 2008 *Caenorhabditis elegans* mutant allele identification by whole-genome sequencing. *Nat. Methods* **5**: 865–867.
- SARIN, S., V. BERTRAND, H. BIGELOW, A. BOYANOV, M. DOITSIDOU *et al.*, 2010 Analysis of multiple ethyl methanesulfonate-mutagenized *Caenorhabditis elegans* strains by whole-genome sequencing. *Genetics* **185**: 417–430.
- SCHMITZ, C., P. KINGE and H. HUTTER, 2007 Axon guidance genes identified in a large-scale RNAi screen using the RNAi-hypersensitive *Caenorhabditis elegans* strain nre-1(hd20) lin-15b(hd126). *Proc. Natl. Acad. Sci. USA* **104**: 834–839.
- SHIBATA, Y., H. TAKESHITA, N. SASAKAWA and H. SAWA, 2010 Double-bromodomain protein BET-1 and MYST HATs establish and maintain stable cell fates in *C. elegans*. *Development* **137**: 1045–1053.
- STRUHL, K., 1998 Histone acetylation and transcriptional regulatory mechanisms. *Genes Dev.* **12**: 599–606.
- SUZUKI, H., T. R. THIELE, S. FAUMONT, M. EZCURRA, S. R. LOCKERY *et al.*, 2008 Functional asymmetry in *Caenorhabditis elegans* taste neurons and its computational role in chemotaxis. *Nature* **454**: 114–117.
- ULLAH, M., N. PELLETIER, L. XIAO, S. P. ZHAO, K. WANG *et al.*, 2008 Molecular architecture of quartet MOZ/MORF histone acetyltransferase complexes. *Mol. Cell. Biol.* **28**: 6828–6843.

Communicating editor: B. J. MEYER

GENETICS

Supporting Information

<http://www.genetics.org/cgi/content/full/genetics.110.123661/DC1>

Maintenance of Neuronal Laterality in *Caenorhabditis elegans* Through MYST Histone Acetyltransferase Complex Components LSY-12, LSY-13 and LIN-49

M. Maggie O'Meara, Feifan Zhang and Oliver Hobert

Copyright © 2010 by the Genetics Society of America
DOI: 10.1534/genetics.110.123661

FILE S1

Supporting Material and Methods

Transgenes

Transgenes that label ASEL and ASER fates include ASEL markers *otIs114 = Is[lim-6^{prom}::gfp; rol-6(d)]*, *otIs160 = Is[lsy-6^{prom}::gfp; unc122^{prom}::gfp]*, ASER markers include *otIs186 = Is[gcy-5^{prom}::gfp; rol-6(d)]*, *otIs274 = [die-1::yfp recombineered fosmid; che-1::mCherry:che-1_3'UTR; rol-6(d)]* (kindly provided by Baris Tursun), *ntIs1 = [gcy-5^{prom}::gfp; lin-15(+)]*, *otIs220 = Is[gcy-5^{prom}::mCherry; rol-6(d)]*; ASEL/R markers *otIs151 = Is[ceh-36^{prom}::DsRed2; rol-6(d)]*, *otIs232 = Is[che-1^{promA}::mCherry::che-1_3'UTR; rol-6(d)]*. *otIs186 = Is[gcy-5^{prom}::gfp; rol-6(d)]*. *lsy-12*-containing fosmid rescuing lines *otEx3510-2 = three independent lines of Ex[fosmid WRM061aF10; elt-2::gfp]* (SARIN *et al.* 2008). Additional transgenes generated in this study: *otEx3676; otEx3677; otEx3678 = three independent rescue lines of Ex[lsy-12(R07B5.8 & R07B5.9 overlapping PCR genomic regions); elt-2::gfp]*; *otEx4330, otEx4330 = Ex[ceh-36^{prom2}::lsy-12a; elt-2::dsRed2]*; *otEx4317-21 = five independent lines of Ex[ceh-36^{prom2}::lsy-12b; elt-2::dsRed2]*. *otEx4366-68 = three independent lines of Ex[nonprom_lsy-12a; elt-2::gfp]*; *otEx4418 = Ex[lsy-13 4.4 kb genomic locus; elt-2::gfp]*; *otEx4417 = Ex[lsy-13^{prom}::gfp:unc-54^{3'UTR}; elt-2::gfp]*.

Expression constructs

ceh-36^{prom2}::lsy-12, *nonprom_lsy-12a*, and *lsy-13^{prom}::gfp* were constructed by a combination of an *in vitro* PCR fusion approach (HOBERT 2002) and an *in vivo* recombineering approach, in which overlapping PCR fragments recombine after injection into worms (BOULIN *et al.* 2006). Rescue constructs are shown in **Fig. 2**. *ceh-36^{prom2}::lsy-12a* was generated by PCR fusing a 1.8kb promoter region of *ceh-36* with the 3.6kb genomic region of R07B5.8 (12ng/ul) and was coinjected with a 3.8 PCR fragment including the entire R07B5.9 genomic locus with endogenous 3'UTR with 110bp overlap with the R07B5.8 fragment (12ng/μl). *elt-2::dsRed2* (3ng/μl) was used as an injection marker. *ceh-36^{prom2}::lsy-12b* is the same construct injected without the R07B5.9 fragment. *nonprom_lsy-12a* construct was made by coinjecting a 3.6 kb promoterless fragment of the *R07B5.8* genomic locus (12.5 ng/μl), the entire *R07B5.9* genomic locus as described above (12.5 ng/μl) and *elt-2::gfp* (50ng/μl) as injection marker.

lsy-13^{prom}::gfp was generated by PCR-fusing 2.8kb of sequences upstream of the first predicted exon of the *lsy-13* locus to the *gfp* coding region and *unc-54^{3'UTR}* (20ng/μl) and was coinjected with *elt-2::gfp* (50ng/μl) as injection marker.

Primer sequences for each construct are as follows (primer name-sequence 5' to 3'):

R07B5.8 PCR rescue fragment:

R07B5.8RescueFwd – ggctcgcttcattagac
R07B5.8RescueRev – ggtgcggattgatgtgagg

R07B5.9 PCR rescue fragment:

R07B5.9RescueFwd – gtcacattccccggttatgc
R07B5.9RescueRev – cgtctatgatgcctattgcc

ceh-36^{prom2}::lsy-12a

ceh36p2_5A – ttAAGCTTATCCGATAAGGCTG
ceh36R07B58pB – cttgagaaggaacacatagGGATCcgcaaatggcggagggtg
R07B5.8tnsIC - ctatgtgtccctctcaag
R07B5.8rescRevout (use as D) - gacgacatagaatacgtgc
R07B5.8RescueRev (use as D*) – ggtgcggattgatgtgagg
coinject with R07B5.9 overlapping PCR rescue fragment

ceh-36^{prom2}::lsy-12b

Same as *ceh-36^{prom2}::lsy-12a* but no R07B5.9 overlapping rescue fragment

nonprom_lsy-12a

R07B5.9 genomic region primers:

R07B5.9RescueFwd: *cgtctatgatgcctattgcc*

R07B5.9RescueRev: *cgtctatgatgcctattgcc*

R07B5.8 genomic region without promoter:

7R07B5.8tns1C: *ctatgtgtcccttccaag*

R07B5.8RescueRev: *ggtagcggattgatgagg*

lsy-13^{prom}::gfp

T06A10.4_A: *cacagtgacttttccccg*

T06A10.4_A*: *gagatgagtgccgatgg*

T06A10.4_B: *agtcgacctgcagcatgcaagcttcacttcttcttccaatcccccttcg*

C – *agcttgcctgcctgcaggtcgact*

D – *aagggccctacggccgacta*

D* – *ggaaacagttatgttggatattggg*

***lsy-12* transcript analysis**

RT-PCR analysis was performed with the Invitrogen Superscript one-step RT-PCR System with Platinum Taq Polymerase.

RT-PCR primers include:

lsy12RTfullp1: *cggtcagtgatagaaacg*

lsy12RTfullp2: *caaccattatcggaactcgg*

lsy12RTfullp3: *gctgctagagatctcactg*

lsy12RTfullp4: *cttgagagtaagcctggac*

lsy12RTfullp5: *gccgtgattgctccaattg*

lsy12RTfullp6: *cctccaattccacctgcac*

R07B5.9end: *gaactgattggtggcagttcc*

R07B5.9endNest: *ggtggcagttccattgtttg*

***lsy-12* cloning**

The *lsy-12* alleles *ot89* and *ot170* alleles were mapped to one map unit on LGV by traditional three factor mapping using physical markers *dpy-11* and *unc-76*. Hawaiian SNP mapping was performed as described (WICKS *et al.* 2001), which narrowed the interval to 0.07 map units on LGV. *lsy-12* was rescued by transgenic lines containing fosmid WRM0610aF10 with 39.3 kb of genomic sequences including R07B5.8 and R07B5.9 (SARIN *et al.* 2008). Two overlapping PCR fragments containing the entire R07B5.8 (8.3kb) and R07B5.9 (3.9kb) loci also rescued the *lsy-12* phenotype (**Fig. 2**, primers listed above). This procedure happened in parallel to a whole genome sequencing approach for mutant identification on *lsy-12(ot177)*, previously mischaracterized as being in a distinct complementation group (SARIN *et al.* 2008).

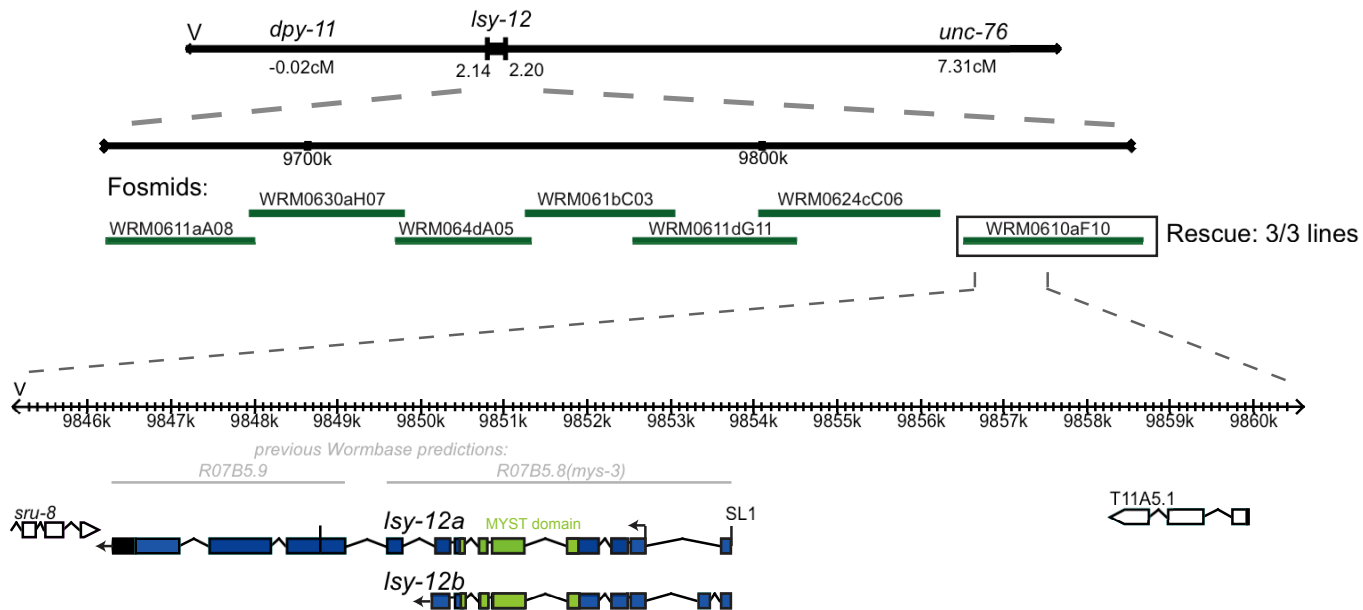
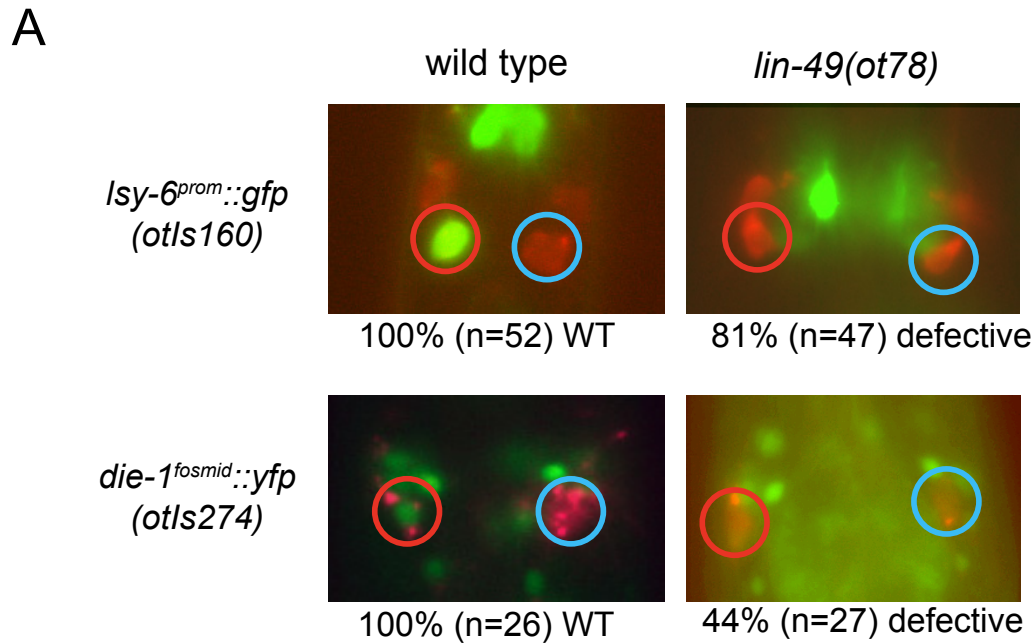


FIGURE S1.—Mapping *lsy-12*. The *lsy-12* alleles *ot89* and *ot170* alleles were both mapped to one map unit on LGV by traditional three factor mapping using physical markers *dpy-11* and *unc-76*. SNP mapping with the polymorphic Hawaiian *C.elegans* isolate CB4856 narrowed the interval to 0.07 map units on LGV. *lsy-12* was rescued by transgenic lines containing fosmid WRM0610aF10 with 39.3 kb of genomic sequences including R07B5.8 and R07B5.9.

**B**

Genotype	<i>lim-6::gfp (otIs114)</i> expression			n
	ASEL only	ASEL=ASER	no express.	
wild type	100%	0%	0%	>100
<i>lsy-12(ot154)</i>	58%	0%	42%	50
<i>lin-49(ot78)</i>	84%	0%	16%	52
<i>lsy-12(ot154); lin-49(ot78)</i>	12%	0%	88%	33

FIGURE S2.—The MYST complex component *lin-49* affects ASE laterality. A: *lin-49* controls *lsy-6* and *die-1* expression. The effect of *lin-49* on other laterality markers was previously reported in (CHANG *et al.* 2003). Numbers below the panels indicate the penetrance of the phenotype, i.e. the fraction of animals that display the phenotype shown in the fluorescent image above. The *die-1* wild-type control images are the same as in Fig.1 and shown for comparison only. B: Partial loss of function alleles of *lsy-12* and *lin-49* enhance one another. Since stronger alleles of *lsy-12* are completely penetrant, other interaction tests of this sort could not be performed.

TABLE S1

Laterality phenotype of *lsy-12* transgenic animals

genotype	<i>lim-6^{prom}::gfp (otIs114)</i> expression			n
	ASEL only	ASEL=ASER	no express.	
wild type	100%	0%	0%	>100
<i>lsy-12(ot170)</i>	0%	0%	100%	>100
<i>fosmid rescue</i>				
<i>lsy-12(ot170); otEx3510</i>	100%	0%	0%	52
<i>lsy-12(ot170); otEx3511</i>	90%	0%	10%	30
<i>lsy-12(ot170); otEx3512</i>	94%	0%	4%	26
<i>lsy-12a/b PCR rescue</i>				
<i>lsy-12(ot170); otEx3676</i>	97%	3%	0%	33
<i>lsy-12(ot170); otEx3677</i>	78%	22%	0%	36
<i>lsy-12(ot170); otEx3678</i>	84%	16%	0%	32
<i>ceh-36^{prom2}::lsy-12a</i>				
<i>lsy-12(ot170); otEx4330</i>	84%	10%	6%	31
<i>ceh-36^{prom2}::lsy-12b</i>				
<i>lsy-12(ot170); otEx4318</i>	62%	0%	38%	26
<i>lsy-12(ot170); otEx4319</i>	36%	0%	64%	25
<i>lsy-12(ot170); otEx4320</i>	18%	0%	82%	33
<i>no prom lsy-12a</i>				
<i>lsy-12(ot170); otEx4366</i>	2.6%	0%	97.4%	39
<i>lsy-12(ot170); otEx4367</i>	0%	0%	100%	21
<i>lsy-12(ot170); otEx4368</i>	0%	0%	100%	26

See Fig.1 for schematic representation of constructs. Note that animals that express the genomic *lsy-12* locus from a multicopy transgenic array not only rescue the loss of ASEL fate in ASEL, but also show a partially penetrant and partially expressive conversion of ASER to ASEL. Similar ASER to ASEL conversions can be observed in transgenic animals expressing the *lsy-12* locus under control of the bilaterally expressed *ceh-36* promoter. Within the appropriate cellular context, *lsy-12* is therefore not only required but, at least to some extent, sufficient to induce ASEL fate. The potential link of *lsy-12* with *die-1* and *che-1* may explain the overexpression effect of *lsy-12*. We have previously shown that *die-1* or *che-1* gene action is dosage sensitive such that overexpression of either gene can induce ASER to ASEL cell fate alterations (ETCHBERGER *et al.* 2009; JOHNSTON *et al.* 2005). Perhaps the overexpression of *lsy-12* raises the activity of *die-1* and/or *che-1* gene activity in ASER above a certain threshold, resulting in an ASER to ASEL conversion. Overexpression of the mouse homolog of *lsy-12*, Querkopf, also results in effects that are reciprocal to the loss of function phenotype (MERSON *et al.* 2006). We also find that both the *lsy-12a* and the *lsy-12b* isoform can rescue the laterality defect, even though mutant alleles are restricted to the *lsy-12a* isoform. Perhaps *lsy-12a* and *lsy-12b* are functionally equivalent, but *lsy-12a* is the only isoform that is expressed in ASEL.

CHAPTER 6: DISSCUSSION AND FUTURE DIRECTIONS

Combinatorial codes of terminal selector regulation:

The nervous system employs a series of sequentially acting factors that successively restrict and determine cellular fates in multiple steps. Despite the fact that much is known about the early steps, it remains obscure how a mature neuron achieves its terminally differentiated state and how this state is maintained at later stages. The same transcription factor is very often expressed in distinct cell types, yet it is unclear how these cells acquire different differentiated fates and whether the same transcription factor carries out the same function in different cell types. Chapter 2 takes the LIM homeodomain transcription factor TTX-3 as a starting point to test the proposed concept that combinatorial codes of transcription factors control the adoption of distinct neuronal cell types. This notion can also be expanded to other transcription factors such as UNC-86, which not only cooperates with TTX-3 but also collaborates with other regulatory factors in other cell types in a synergistic manner. Therefore, it is the combinatorial codes of transcription factors that restrict and define the terminal differentiated fate of a neuron.

Continuous requirement of terminal selector transcription factors

Another key feature of terminal selectors is that they are continuously required throughout the life of an animal, suggesting that they are not only involved in initiating expression of terminal differentiated genes, but also responsible for maintaining such identities (Etchberger et al., 2009; Flames and Hobert, 2009; Kratsios et al., 2012). Utilizing fosmid based reporters, maintained expression of both UNC-86 and TTX-3 at

later stages of the worm has been observed. Further experiments could be done to address the question. For example, temperature-sensitive alleles, if available, can be used to temporally remove or provide gene activities to assess its roles in neuronal maintenance. Other experiments such as heat-shock induced expression of proteins at post-embryonic stages may also provide some insight into this question.

Target gene specificities

Our promoter bashing analysis suggests that although the same transcription factor TTX-3 is expressed in both AIY and AIA, it binds to different *cis*-regulatory motifs to achieve specificity. In AIY, together with the homeodomain protein CEH-10, TTX-3 recognizes the “AIY motif” present in its target genes (Wenick and Hobert, 2004). In AIA, TTX-3 utilizes a different signature that is shared by AIA-specific genes. In the case of *cho-1*, which is expressed in both AIA and AIY, both the “AIY motif” and the “AIA motif” are present in the regulatory region, but deleting either of the two motifs only affects expression in one of the two neurons, while the other remain present. Therefore, we propose that aside from cooperating with a cofactor, the specificities of the same transcription factor in different cell types is achieved by utilizing different and dedicated regulatory motifs that respond to this transcription factor.

Looking for cofactors

We have shown that TTX-3 is required for the expression of the terminal gene battery in the AIA interneuron. However, evidence suggests that it is likely that *ttx-3* acts with another yet unknown factor to jointly control AIA fate, as majorities of the

AIA-expressed genes assessed with fluorescent reporters are not completely abolished in *ttx-3* mutants. I have performed a preliminary screen using two reporters, *mgl-1::mcherry* (or *mgl-1::gfp*) and *ins-1::gfp*. Several mutants were retrieved but no mutants aside from *ttx-3* alleles have been found. It is possible that the locus encoding such a factor is small or difficult to hit. One can also argue that the other factor when mutated may lead to sterility or lethality. This problem can be overcome by performing manual clonal screens instead of automatic screens utilizing the Biometrica worm sorter, which pools the F2 generation and does not allow recovery of the heterozygous mother. It is also possible that the other factor has much weaker or no effect alone when mutated, and only when *ttx-3* is also mutated, a much stronger effect could be observed. Further screens in the *ttx-3* mutant background shall help identify such mutants. This requires a reporter that is not expressed in too many cells (easy for AIA identification), and is weakly affected by mutations in *ttx-3*. Appendix 1 of this thesis also describes TTX-3 as a terminal selector required for the specification of the glutamatergic chemosensory neuron ASK. Similar to AIA, the expression of several ASK terminal identities are not completely eliminated by the absence of *ttx-3* activity, which indicates that TTX-3 may as well collaborate with a cofactor in ASK. Screens with a fluorescent reporter in which ASK is weakly affected by *ttx-3* may provide some clues for identification of the unknown factor.

TTX-3 function in AIN

AIN has been previously thought to be glutamatergic, but our data suggests that it may instead be cholinergic, because we observe GFP expression in AIN with a short

version of transcriptional GFP reporter for the choline transporter *cho-1* (*otIs379*). This could be easily confirmed by looking at a red fosmid-based reporter of *cho-1* to see if there is colocalization or by performing antibody staining against the synaptic vesicle acetylcholine transporter UNC-17.

Aside from its expression in NSM, AIA, AIY and ASK, our observation with the fosmid reporter *wgIs68* suggests that TTX-3 is also expressed in the motoneuron AIN in the head of the worm. The question is then whether TTX-3 has similar functions in AIN, and if so, whether it collaborates with another unknown factor to determine AIN fate. With our short *cho-1* reporter that is only expressed in AIY, AIA and AIN, we know that although AIA and AIY are affected in *ttx-3* mutants, AIN remain intact. Therefore, TTX-3 is at least not the sole terminal selector for AIN, but may still partner with another factor to determine the fate of AIN in synergy. Baran et al. has reported that AIN express a high level of the homeodomain protein UNC-42. Then is *unc-42* the cofactor? What is the role of *unc-42*? Future genetic screens for AIN mutants in the *ttx-3* mutant background may also shed some light on this question. Another possibility is that TTX-3 plays a different role other than a transcription activator in AIN. For example, it may act as a repressor, similarly to what *unc-3* does in the ASI sensory neuron, to repress genes specific to other neuron types (Kim et al., 2005).

Considerations in other systems

We have shown that in the nematode *C. elegans*, the POU homeodomain transcription factor UNC-86 is required for the fate specification of the cholinergic IL2 neurons, which likely have sensory roles in response to chemical cues such as salts. This

is achieved through cooperation with the ARID-type transcription factor CFI-1. In *Drosophila*, the POU homeobox protein abnormal chemosensory jump 6 (Acj6) is a positive transcription regulator of the cholinergic gene locus in primary olfactory neurons (Lee and Salvaterra, 2002). The homolog for CFI-1 in flies, Dead Ringer (Ditch et al., 2005) has been reported to function in neuronal differentiation as well. It is interesting to examine whether these two proteins in flies have any overlap in their expression pattern and whether there is any cooperativity between them. Moreover, it may be worthy of looking into the vertebrate homolog of CFI-1, the ARID3 genes, whose role in neural development hasn't been very well explored yet.

Specification of the RMD neuron class

Terminal selector for the RMD neuron class

unc-42 encodes a paired-like homeodomain protein of the Q50 class related to the mammalian protein Prop1. It is required for the proper specification of the terminal features of the ASH sensory neurons, AVA, AVD, and AVE interneurons, two RME neurons and all six RMD neurons based on expression of the ionotropic glutamate receptors GLR-1, GLR-4 and GLR-5 (Baran et al., 1999; Brockie et al., 2001; Serrano-Saiz E, 2013). It is likely that *unc-42* serves as the terminal selector for the RMD neuron class especially when considering that *unc-42* is expressed throughout adult hood, since terminal selectors are generally thought to be continuously required at later stages. All six RMD neuron utilize acetylcholine. It is yet unknown whether cholinergic fate of these neurons depend on *unc-42*, which could be tested with reporters for the choline

transporter *cho-1* and for the synaptic vesicle acetylcholine transporter *unc-17*. *unc-42* has been reported to severely affect its own expression when mutated (autoregulation) (Baran et al., 1999), but whether this holds true in all six RMDs remain unclear and needs to be further examined.

Transcriptional regulation of gene expression

There are six neurons in total in the RMD neuron class. According to previous studies, it seems that *unc-42* could potentially be the terminal selector for all six RMD class neurons. Although terminal genes such as *glr-1*, *glr-4* and *glr-5* are expressed in all six neurons, the transgene *otIs317 (mgl-1::cherry)* only labels RMDDL/R and RMDV/R, which suggests that gene expression profiles of distinct subtypes of neurons within this class are not entirely identical. It has also been reported that the LIM homeobox gene *lim-4* is expressed in RMDL and RMDR (Sagasti et al., 1999) and that *glr-2* expression in a subset of RMD neurons is not dependent on *unc-42* (Brockie et al., 2001). This prompted the question whether there are different layers of terminal fate regulation. For example, although all AIY terminal genes are under the control of the CHE-10/TTX-3 heterodimer, another transcription factor, *ceh-23* also plays a role, although only required to maintain the expression of one defined AIY terminal feature, the orphan serpentine receptor *sra-11* (Altun-Gultekin et al., 2001). Similarly, Does *lim-4* have roles only in RMD specification? Where does it act and is it dependent on *unc-42*? Moreover, does *lim-4* contribute to distinguishing RMD from other neurons in the same class? It would also be interesting to look into the gene expression profiles of different RMD neurons, and try to distinguish features that are shared by all six neurons from

those only expressed in a subset, in order to study whether there is any differential gene expression and regulation that exist.

*Proneural roles of *cnd-1* and *hlh-16**

As mentioned previously, *cnd-1* and *hlh-16* mutants were retrieved by utilizing a reporter gene that is only expressed in the RMDV and RMDD neuron class. The roles of *cnd-1* and *hlh-16* in the RMD lineages remain unknown, which could be assessed by using reporter genes that labels all six neurons, namely *glr-1*, *glr-4* or *glr-5*, or by looking at a terminal marker that is expressed only in RMD if available in future experiments.

According to their expression patterns and timing, neural bHLH factors are often considered to have roles in the generation of neuron progenitors and therefore considered “proneural” during metazoan development. In *C. elegans*, for example, the *achaete-scute* family member *hlh-14* is expressed in neuron precursors and is required for neurogenesis. *hlh-14* mutants do not generate three lineally related neurons, the PVQ interneuron, the HSN motor neuron and the PHB sensory neuron, and possibly transformed the neuroblast of the whole branch into a hypodermal blast cell (Frank et al., 2003). Similar phenotype was observed in *hlh-14* mutants within the neuronal lineage branch that generates ASE, OLL and AFD, with neuron-to-hypodermal transformation observed (Poole et al., 2011). It is possible that *hlh-16* as well as *cnd-1* exert similar functions on promoting neuroblast formation and therefore have roles in neuronal lineage specification of the RMD neuron class. Lineaging tracing experiments with *hlh-16* and *cnd-1* mutants may help answer the question of whether or not similar neuronal

to hypodermal lineage transformation is observed. It would be interesting to examine lineally related neurons in the RMD lineages as well, such as SMBD and SMBV in the RMDD lineage, SAAD in the RMDV lineage, and AFD and ASK in the RMD lineage (See Chapter 3, Figure 4). Lastly, looking at terminal markers for these neurons might provide further information on whether the whole lineage is specified in *cnd-1* and *hlh-16* mutants, and where and how early do they act.

Studies have shown that in VNC motor neurons, expression of neuronal specific transcription factors/terminal selectors *unc-3*, *unc-4* and *unc-30* is altered in *cnd-1* mutants (Hallam et al., 2000). If *unc-42* is the terminal selector for the RMD class, then one should examine whether the expression of *unc-42* is affected in *cnd-1* or *hlh-16* mutants. In a broader view, all neurons that are most closely related to RMDD and RMDV are cholinergic (SMBV, SMBD, SAA, SMDV). One bold but legitimate speculation in the bigger picture would be that *cnd-1* and/or *hlh-16* are required for generating cholinergic features of these lineages. This can be tested by building reporters for the choline transporter *cho-1* and the synaptic vesicle acetylcholine transporter *unc-17* into *cnd-1* and *hlh-16* mutants.

If both *cnd-1* and *hlh-16* are involved in the specification of the RMD lineages, another interesting question is what the relationship between the two factors is? Evidence has suggested that NeuroD can be activated by neurogenin, another member of the bHLH transcription factor family (Huang et al., 2000). In vertebrates, the NeuroD family genes have been reported to have roles in promoting differentiation as well as acting in neuroepithelial cells. Studies in the amacrine interneurons have demonstrated that, instead of proneural functions, genes in the NeuroD family can have differentiation

roles in specifying neuronal identity (Morrow et al., 1999). Studies in *Xenopus* and the mouse also reported that NeuroD can act downstream of or subsequent to other bHLH factors. Then does *hlh-16* act upstream of *cnd-1* in *C. elegans*? Or do they cooperate with each other synergistically or work in a complementary manner?

It has also been pointed out that non-proneural bHLH factors need to work together with homeobox genes to specify retinal neuron fates (Bertrand et al., 2002). NeuroD or Math3 can induce amacrine cell fate with the homeobox protein Chx10 or induce bipolar cell fate with Pax6 (Hatakeyama et al., 2001; Inoue et al., 2002). If *cnd-1* is involved in cell fate restriction, then it would be curious to ask whether in *C. elegans* it requires a cofactor as well.

Another question worth considering is that RMDD, RMDV and RMD are not closely and lineally related, and they descend from quite distinct lineages. However they seem to be under the control of the same set of transcription factor *cnd-1* and *hlh-16*. How is the lineage specificity achieved? Is there any other cofactor that is also involved? Future experiments using available fosmid reporters for these genes might help address these questions.

The screen

All RMD mutants that are alleles of *unc-42*, *cnd-1* and *hlh-16* are retrieved quite unexpectedly from a screen for another purpose (for AIA mutants). Due to the nature of the screen (non-clonal), the number of genomes cannot be properly calculated. However, all mutants were found from the same round of preliminary screen (See Chapter 3 for details). Based on the allele frequency and the number of loci hit from limited numbers

of genomes screened, it is very likely that more alleles and more novel genes will be uncovered in future screens. One could also take advantage of using the worm sorter to automate and facilitate the sorting process. Another advantage of continuing doing a screen using the same strain *otIs317 (mgl-1::mcherry)* is that one could potentially retrieve mutants that are affected in the NSM and AIA from the same screen.

ASE asymmetry

Considerations on ASE screens: manual clonal screen vs. automatic non-clonal

Recently, the COPAS Biosort System (Union Biometrica), also called the “worm sorter” was made available for high-throughput mutant screening (Doitsidou et al., 2008). It is a special flow cytometry machine that is designed for sorting larger particles including living organisms like *C. elegans*, based on parameters such as size, particle density and fluorescent intensity. This can markedly speed up the screening process and alleviate tedious and laborious manual screening. All can be done within a significant shorter amount of time. Doitsidou et al. demonstrated that they were able to retrieve a variety of mutants that display various abnormal neuronal phenotypes, some which were not picked up in the previous manual screen with the same reporter gene.

Although the use of the worm sorter has been quite successful in looking for dopaminergic mutants, it is not always the case for all screens. A previous large-scale screen that went through approximately 120, 000 haploid genomes manually reported around 120 alleles with disrupted ASE asymmetry (Sarin et al., 2007). Aside from low-penetrance and multiple-loci mutants, a total of 14 regulatory genes have been identified, 12 of them hit multiple times. However, in a subsequent automatic screen that sought to

identify more mutants involved in the ASE regulatory network, out of the 49 alleles identified (across 15 different runs of experiments, non-clonal), 48 most of them are alleles of either *fozi-1* or *cog-1* alleles (See appendix 3), suggesting that automated screens are probably more biased than manual clonal screens in at least some cases.

There are also certain drawbacks with the automated approach. First, in order to sort more efficiently and to grow worms in larger quantity, all screens need to be non-clonal. This biases on F2 homozygous animals that are viable, and therefore lethal or sterile mutants might be selected against because they are not able to grow or to propagate. Although not entirely impossible, it is extremely difficult to recover viable but sterile mutants from a pooled population consisting of complex genetic makeups. One could argue that the lethality or sterility issue may be compensated by the much larger number of genomes screened in the hope of obtaining hypomorphic alleles that are healthier and viable, but a full spectrum of mutants would still be favorable, and genes with severe pleiotropies are less likely to be picked up by the machine, because a worm would have to go through multiple steps before it can be finally sorted onto a plate. Moreover, conducting a non-clonal screen leaves the number of genomes screened incalculable, which makes the estimation of mutation rate and saturation degree difficult. Second, although it is much faster for the machine to sort out mutant candidates, the rate of getting false positive hits is relatively high. This is inevitable because of the variability among individual animals, even if they haven been synchronized. Third, compared to the machine, the human eyes are more capable of performing multiple sophisticated tasks at the same time without complicated parameter setups. Just by looking at a worm, one could

easily identify gain/loss of expression, cell mispositioning, or any visible phenotypes, while the machine can only handle one task at a time. Therefore, one could probably get a better spectrum of mutants by screening manually, although the automatic screen can increase the chance of hitting rare alleles.

Whole genome sequencing:

lsy-27 was cloned taking advantage of the newly developed high-throughput whole genome sequencing approach. Mutant animals were analyzed by direct sequencing instead of the conventional and somewhat tedious mapping and rescuing approach. More recently, methods that utilizing either polymorphism (Doitsidou et al., 2010) or mutagen induced nucleotide changes (Zuryn et al., 2010) have been introduced. In the old days, characterization of a particular mutant can be extremely time consuming, which could take up to several years and is very often the time-limiting step of a project. With whole genome sequencing (WGS), all can be done within a much shorter amount of time, on the scale of weeks. The cloning of *lsy-12* is another good example of how WGS can be extremely beneficial. It was mapped independently by two parallel strategies. Both conventional genetic mapping methods followed by rescue experiments and the WGS approach were able to pinpoint the mutation to the same locus but the difference is striking. It only took a couple of months for WGS as opposed to years spent on traditional mapping by a former graduate student. Now the only limiting step is to retrieve mutants. This is extremely useful when dealing with an organism such as *C. elegans*, with which large collections of mutants are relatively easy to obtain. It is even possible to streamline the mutant cloning process, which can potentially lead to much

greater productivity in a time and cost efficient manner.

LSY-27 function in triggering lim-6 initiation

Based on reporter analysis, *l sy-27* not only affects terminal markers for ASE neurons, but is also required for *lim-6* expression. The embryonic-restricted LSY-27 expression provides further support for the speculation that LSY-27 only has an earlier role in ASE fate specification, and is only required for the onset but not the maintenance phase of *lim-6* expression, while *lim-6* expression persists throughout the life of ASEL. The C2H2 zinc finger transcription factor DIE-1 is also required for triggering the expression of *lim-6* expression. It is possible that LSY-27 assists DIE-1 to exert such a role only in the initiation phase, as DIE-1 is also expressed throughout adulthood.

The LSY-12/LSY-13/LIN-49 complex and its function

The phenotypic similarity and the reported functions in their vertebrate homologs of LSY-12, LSY-13 and LIN-49 suggests that these three factors may act in a complex to control the establishment and maintenance of ASE laterality In *C. elegans*. Yet the interactions among the components of this MYST histone acetyltransferase complex haven't been fully explored. In zebrafish, the LIN-49 homolog BRPF1 (bromodomain-PHD finger protein 1) recruits the LSY-12 homolog MOZ (monocytic leukemia zinc finger protein), and is able to directly bind to acetylated histone through the bromodomain (Kim et al., 2005). It has also been shown that mammalian BRPF1 serves as a scaffold to bridge the formation of the complex, which also involves MOZ/MORF and the *l sy-13* homolog ING-5 (inhibitor of growth protein 5) (Morrow et al., 1999). We

have demonstrated the enhancement between a partial loss of function allele of *lsy-12* and a *lin-49* allele. This suggests that they may play similar roles in *C. elegans*. Other genetic tests could be performed to further assess the interactions between *lsy-12* and *lsy-13*, and between *lin-49* and *lsy-13*, and among all of the three if possible. Biochemical approaches such as co-immunoprecipitation and chromatin-immunoprecipitation experiments may be a great tool to determine how closely these proteins are associated and what the specific targets of this complex are.

Another unanswered question is that what recruits the complex to specific sites and how this is achieved. The ASEL-expressed C2H2 zinc finger protein DIE-1, and the master regulator CHE-1, are candidate transcription factors for such activities, as transcription factors that recognize specific DNA sequences are generally considered to be required for the specificity and activity of histone-modifying proteins (Inoue et al., 2002). In *die-1* mutants, ASEL are converted to ASER, and an unusual allele of *che-1*, *ot101*, has revealed an additional role of CHE-1 in establishing L/R asymmetry (Etchberger et al., 2009). Defects resulted from mutations in either *die-1* and *che-1(ot101)* are similar to the phenotypes induced by mutations in the components of the HAT complex. Another piece of evidence is that DIE-1 and CHE-1 activities are required throughout the life of an animal, which is similar to the continuous expression pattern of *lsy-12*. Taken together, it is suggested that DIE-1 and CHE-1 may act as sequence-specific adaptors that recruit the LSY-12/LSY-13/LIN-49 DNA-modifying complex to the DNA to maintain terminal differentiated features of the ASE.

References:

Altun-Gultekin, Z., Andachi, Y., Tsalik, E.L., Pilgrim, D., Kohara, Y., and Hobert, O. (2001). A regulatory cascade of three homeobox genes, *ceh-10*, *ttx-3* and *ceh-23*, controls cell fate specification of a defined interneuron class in *C. elegans*. *Development* *128*, 1951-1969.

Baran, R., Aronoff, R., and Garriga, G. (1999). The *C. elegans* homeodomain gene *unc-42* regulates chemosensory and glutamate receptor expression. *Development* *126*, 2241-2251.

Bertrand, N., Castro, D.S., and Guillemot, F. (2002). Proneural genes and the specification of neural cell types. *Nat Rev Neurosci* *3*, 517-530.

Brockie, P.J., Madsen, D.M., Zheng, Y., Mellem, J., and Maricq, A.V. (2001). Differential expression of glutamate receptor subunits in the nervous system of *Caenorhabditis elegans* and their regulation by the homeodomain protein UNC-42. *The Journal of neuroscience : the official journal of the Society for Neuroscience* *21*, 1510-1522.

Ditch, L.M., Shirangi, T., Pitman, J.L., Latham, K.L., Finley, K.D., Edeen, P.T., Taylor, B.J., and McKeown, M. (2005). *Drosophila* retained/dead ringer is necessary for neuronal pathfinding, female receptivity and repression of fruitless independent male courtship behaviors. *Development* *132*, 155-164.

Doitsidou, M., Flames, N., Lee, A.C., Boyanov, A., and Hobert, O. (2008). Automated screening for mutants affecting dopaminergic-neuron specification in *C. elegans*. *Nat Methods* *5*, 869-872.

Doitsidou, M., Poole, R.J., Sarin, S., Bigelow, H., and Hobert, O. (2010). *C. elegans* mutant identification with a one-step whole-genome-sequencing and SNP mapping strategy. *PLoS ONE* *5*, e15435.

Etchberger, J.F., Flowers, E.B., Poole, R.J., Bashllari, E., and Hobert, O. (2009). Cis-regulatory mechanisms of left/right asymmetric neuron-subtype specification in *C. elegans*. *Development* *136*, 147-160.

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

Frank, C.A., Baum, P.D., and Garriga, G. (2003). HLH-14 is a *C. elegans* achaete-scute protein that promotes neurogenesis through asymmetric cell division. *Development* *130*, 6507-6518.

Hallam, S., Singer, E., Waring, D., and Jin, Y. (2000). The *C. elegans* NeuroD homolog *cnd-1* functions in multiple aspects of motor neuron fate specification. *Development* *127*, 4239-4252.

Hatakeyama, J., Tomita, K., Inoue, T., and Kageyama, R. (2001). Roles of homeobox and bHLH genes in specification of a retinal cell type. *Development* *128*, 1313-1322.

Huang, H.P., Liu, M., El-Hodiri, H.M., Chu, K., Jamrich, M., and Tsai, M.J. (2000). Regulation of the pancreatic islet-specific gene BETA2 (neuroD) by neurogenin 3. *Molecular and cellular biology* *20*, 3292-3307.

Inoue, T., Hojo, M., Bessho, Y., Tano, Y., Lee, J.E., and Kageyama, R. (2002). Math3 and NeuroD regulate amacrine cell fate specification in the retina. *Development* *129*, 831-842.

Kim, K., Colosimo, M.E., Yeung, H., and Sengupta, P. (2005). The UNC-3 Olf/EBF protein represses alternate neuronal programs to specify chemosensory neuron identity. *Dev Biol* *286*, 136-148.

Kratsios, P., Stolfi, A., Levine, M., and Hobert, O. (2012). Coordinated regulation of cholinergic motor neuron traits through a conserved terminal selector gene. *Nature neuroscience* *15*, 205-214.

Lee, M.H., and Salvaterra, P.M. (2002). Abnormal chemosensory jump 6 is a positive transcriptional regulator of the cholinergic gene locus in *Drosophila* olfactory neurons. *J Neurosci* *22*, 5291-5299.

Morrow, E.M., Furukawa, T., Lee, J.E., and Cepko, C.L. (1999). NeuroD regulates multiple functions in the developing neural retina in rodent. *Development* *126*, 23-36.

Poole, R.J., Bashllari, E., Cochella, L., Flowers, E.B., and Hobert, O. (2011). A Genome-Wide RNAi Screen for Factors Involved in Neuronal Specification in *Caenorhabditis elegans*. *PLoS genetics* *7*, e1002109.

Sagasti, A., Hobert, O., Troemel, E.R., Ruvkun, G., and Bargmann, C.I. (1999). Alternative olfactory neuron fates are specified by the LIM homeobox gene *lim-4*. *Genes & development* *13*, 1794-1806.

Sarin, S., O'Meara, M.M., Flowers, E.B., Antonio, C., Poole, R.J., Didiano, D., Johnston, R.J., Jr., Chang, S., Narula, S., and Hobert, O. (2007). Genetic screens for *Caenorhabditis elegans* mutants defective in left/right asymmetric neuronal fate specification. *Genetics* *176*, 2109-2130.

Serrano-Saiz E, P.R., Felto T, Zhang F, De La Cruz ED, and Hobert O (2013). Modular Control of Glutamatergic Neuronal Identity in *C. elegans*

by Distinct Homeodomain Proteins. *Cell*.

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

Zuryn, S., Le Gras, S., Jamet, K., and Jarriault, S. (2010). A strategy for direct mapping and identification of mutations by whole-genome sequencing. *Genetics* *186*, 427-430.

APPENDIX 1:

Modular control of glutamatergic neuronal identity in *C.elegans* by distinct homeodomain proteins

Esther Serrano-Saiz, Richard J. Poole, Terry Felton, **Feifan Zhang**, Estanislá Daniel De La Cruz and Oliver Hobert. Cell. 2013 Oct 24;155(3):659-73.

In this paper, Serrano-Saiz et al. mapped out all the glutamatergic neurons that are categorized into 38 neuron classes by examining the expression of EAT-4/VGLUT, the vesicular glutamate transporter. The expression of EAT-4 is controlled in a modular fashion, with different regulatory modules responsible for expression in distinct glutamatergic neuron classes. Based on observation made in *C. elegans*, the vertebrate ortholog, Lim homeodomain protein LHX1 was identified as a regulator of glutamatergic neurons in the brainstem of the mouse.

I identified expression of TTX-3 in the ASK sensory neuron by diI staining, and performed genetic analysis on *ttx-3* mutants with GFP reporters for *eat-4* as well as three additional markers for ASK.

Modular Control of Glutamatergic Neuronal Identity in *C. elegans* by Distinct Homeodomain Proteins

Esther Serrano-Saiz,^{1,*} Richard J. Poole,^{1,2} Terry Felton,^{1,2} Feifan Zhang,¹ Estanislá Daniel De La Cruz,¹ and Oliver Hobert^{1,*}

¹Department of Biochemistry and Molecular Biophysics, HHMI, Columbia University Medical Center, New York, NY 10032, USA

²Department of Cell and Developmental Biology, University College London, London WC1E 6BT, UK

*Correspondence: es2754@columbia.edu (E.S.-S.), or or38@columbia.edu (O.H.)

<http://dx.doi.org/10.1016/j.cell.2013.09.052>

SUMMARY

The choice of using one of many possible neurotransmitter systems is a critical step in defining the identity of an individual neuron type. We show here that the key defining feature of glutamatergic neurons, the vesicular glutamate transporter EAT-4/VGLUT, is expressed in 38 of the 118 anatomically defined neuron classes of the *C. elegans* nervous system. We show that distinct *cis*-regulatory modules drive expression of *eat-4/VGLUT* in distinct glutamatergic neuron classes. We identify 13 different transcription factors, 11 of them homeodomain proteins, that act in distinct combinations in 25 different glutamatergic neuron classes to initiate and maintain *eat-4/VGLUT* expression. We show that the adoption of a glutamatergic phenotype is linked to the adoption of other terminal identity features of a neuron, including cotransmitter phenotypes. Examination of mouse orthologs of these homeodomain proteins resulted in the identification of mouse LHX1 as a regulator of glutamatergic neurons in the brainstem.

INTRODUCTION

A key identity feature of an individual neuron type is its neurotransmitter phenotype. Most classic neurotransmitters are synthesized by specialized enzymes, loaded by specific transporter proteins into synaptic vesicles and taken back into the neuron by specialized plasma membrane transporters. In many cases, the neurotransmitter identity of a specific neuron type is therefore defined by the coordinated expression of genes coding for specific enzymes and transporters. Understanding the regulatory mechanisms that control expression of these enzymes and transporters presents a fruitful “bottom-up” approach that will help explain how a specific neuronal identity is imposed onto a neuron type during development and how this identity is maintained throughout the life of a neuron.

Glutamate is the most broadly employed excitatory neurotransmitter in most vertebrate and invertebrate nervous systems.

In contrast to other neurotransmitter systems, the identity of glutamatergic neurons is not defined by the expression of specific biosynthetic enzymes and reuptake transporters. Since glutamate is present in all cells, its utilization as a neurotransmitter critically depends on the ability of a neuron to load glutamate into synaptic vesicles. This is achieved by a vesicular transporter for glutamate of the SLC17 family of solute carriers, called VGLUT (Takamori et al., 2001). Ectopic expression of VGLUT is sufficient to confer the glutamatergic phenotype (i.e., synaptic release of glutamate) onto heterologous neurons (Takamori et al., 2000, 2001). Consistent with the sufficiency of VGLUT to determine the glutamatergic phenotype, there are no pan-glutamatergic markers other than the VGLUT genes (see Supplemental Information).

Given the importance of VGLUT genes in defining the glutamatergic phenotype of a neuron, it is perhaps surprising that very little is known about how VGLUT expression is regulated in the nervous system of any vertebrate or invertebrate species, including mouse, *Drosophila* and *C. elegans*. In the mouse, several transcription factors have been described to be involved in the generation of glutamatergic neurons in different areas of the developing central nervous system (Brill et al., 2009; Cheng et al., 2004; Englund et al., 2005; Lou et al., 2013; Ma and Cheng, 2006), but it is not clear whether any of these factors directly initiates and maintains VGLUT expression or whether they act transiently at earlier stages of differentiation and operate through intermediary factors.

The nematode *C. elegans* contains one well-characterized VGLUT-encoding gene, *eat-4* (Lee et al., 1999). *eat-4/VGLUT* enables glutamatergic transmission in various neuronal circuits that control distinct behaviors (e.g., Chalasani et al., 2007; Lee et al., 1999) and the *eat-4* mutant phenotype can be rescued by human VGLUT (Lee et al., 2008). How *C. elegans eat-4/VGLUT* expression is regulated in distinct neuronal cell types has not previously been investigated, mirroring the absence of insight into the regulation of *Drosophila* or vertebrate VGLUT gene expression.

In principle, one could imagine several distinct scenarios by which VGLUT gene expression is controlled in different neuronal cell types. A dedicated regulatory factor (or combination thereof) could exist to control VGLUT expression in all different glutamatergic neuron types (model #1 in Figure 1A). This dedicated factor could be turned on by distinct sets of earlier acting factors in

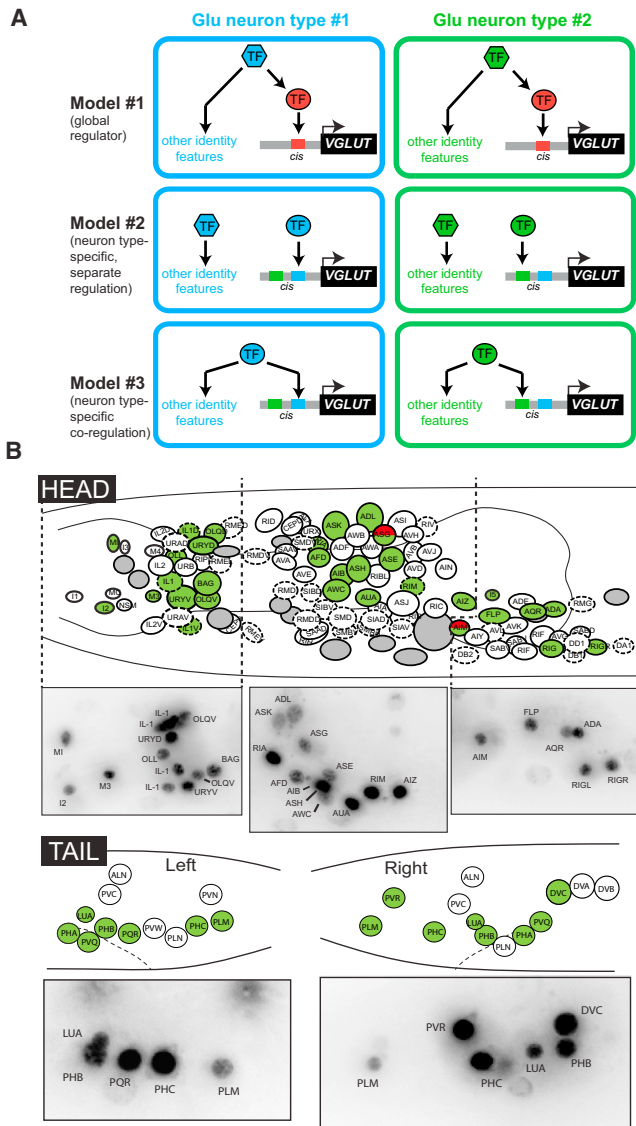


Figure 1. *eat-4/VGLUT* Expression

(A) Three different models for regulation of *VGLUT* expression in different glutamatergic neuron types.

(B) Expression of a fosmid-based *eat-4* reporter (*ots388*) in head and tail ganglia. *yfp*-based images are lateral projections inverted to sharpen the signal. Green circles indicate *eat-4* expressing neurons. Green/red circles indicate coexpression of *eat-4* and serotonin. Dashed lines indicate motor neuron class. See also Figure S1.

distinct glutamatergic neuron populations. Alternatively, *VGLUT* gene expression could be regulated in different manners in distinct glutamatergic neuron types (Figure 1A, model #2 and #3). Furthermore, *VGLUT* expression may be controlled separately from the expression of other identity features of a neuron (Figure 1A, model #2) or its expression could be tightly linked with the expression of other cell-specific identity features of particular glutamatergic neuron types (Figure 1A, model #3). Whether distinct terminal identity features of a neuron are sepa-

rately regulated or whether they are coregulated via a common (set of) *trans*-acting factor(s) (model #2 versus #3) are fundamental but little understood neurodevelopmental problems.

Here, we approach these questions using two conceptually distinct but converging approaches. First, we elucidate the *cis*-regulatory logic of *eat-4/VGLUT* expression by defining *cis*-regulatory regions in the *eat-4* locus that are required for expression of *eat-4* in distinct glutamatergic neurons. If one common regulatory mechanism exists (model #1) that controls *eat-4/VGLUT* expression in all glutamatergic neurons, such *cis*-regulatory analysis should reveal a specific *cis*-regulatory element required for expression in all *eat-4/VGLUT* expressing neurons. Alternatively, if distinct glutamatergic neuron types employ distinct regulatory mechanisms (model #2 and #3), the *eat-4/VGLUT* regulatory elements should be complex and modular in nature, with different elements driving expression in different neurons types. Second, we analyzed the effect of removal of a number of transcription factors on the expression of not only *eat-4/VGLUT* but also on the expression of other identity markers of the respective glutamatergic neuron type (hence distinguishing model #2 and #3) and we report evidence of pervasive coregulation. Furthermore, we provide evidence of the conservation of regulatory mechanisms in the mouse. Our analysis reveals a comprehensive picture of the regulation of glutamatergic neuronal identity in the nervous system.

RESULTS

eat-4/VGLUT Expression Defines 38 Glutamatergic Neuron Classes

We defined the glutamatergic nervous system of *C. elegans* by examining the expression of a fosmid-based *eat-4* reporter construct (Figures 1B and 2). This reporter is expressed in 78 of the 302 neurons of the adult hermaphrodite, which fall into 38 neuron classes (out of a total of 118 anatomically defined neuron classes in the hermaphrodite; Table 1 and Figures 1B and 2). Most of these neurons are either sensory- or interneurons. Only two motorneurons utilize glutamate; both are located in the pharynx.

If the *eat-4/VGLUT* expressing neurons that we describe here indeed use glutamate as neurotransmitter, one would expect that synaptically connected neurons should express ionotropic glutamate receptors. Based on the complete synaptic connectivity diagram of the *C. elegans* hermaphrodite (White et al., 1986) and previously described expression patterns of all known glutamate-gated ion channels (Brockie and Maricq, 2006), we infer that each of the *eat-4/VGLUT*-expressing cells is presynaptic to at least one neuron (or pharyngeal muscle in the case of the pharyngeal motor neurons) that expresses glutamate-gated ion channels (Table 1). This corroborates the glutamatergic identity of the *eat-4/VGLUT* expressing neurons. Similar to vertebrates, we found that the expression of *C. elegans* glutaminases does not track with glutamatergic neuronal identity (Supplemental Information and Figure S1 available online).

The identity of other neurotransmitter systems (cholinergic, GABAergic, dopaminergic, serotonergic, tyraminerpic, octopaminergic) has been described in great detail in *C. elegans* (<http://www.wormatlas.org>). The pattern of *eat-4/VGLUT* expression that we describe here is complementary and not

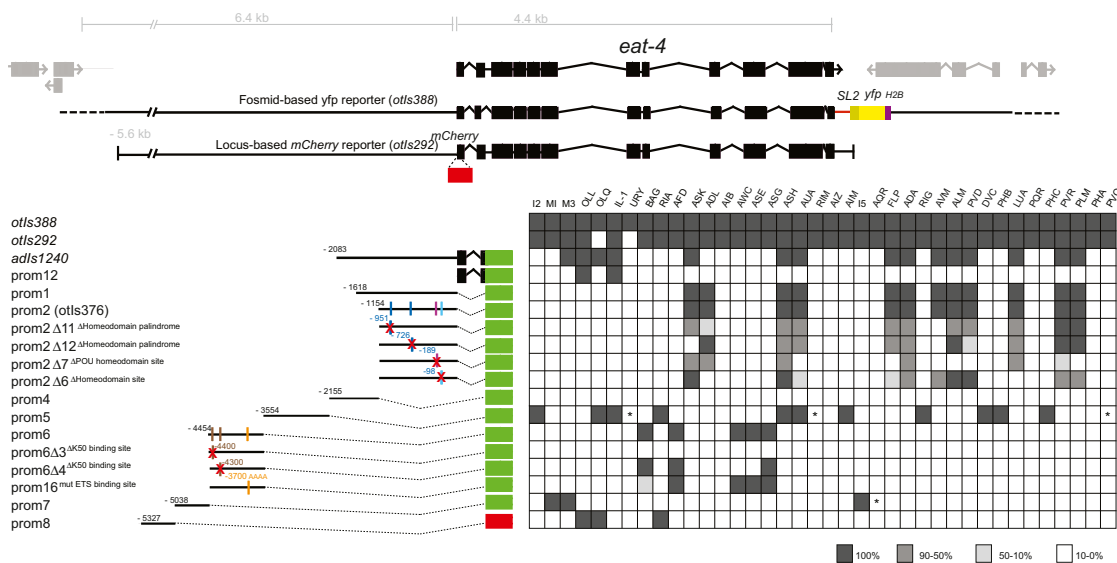


Figure 2. Modular Composition of the *eat-4* cis-Regulatory Control Regions

Overview of expression of different *eat-4* reporter gene fusions, as indicated. The *adIs1240* transgene contains the reporter from the previous Lee et al. (1999) study. For the dissection of 5' regulatory information between two and three lines ($n \geq 10$) were scored for expression and lines commonly showed very similar expression pattern and penetrance of expression (the penetrance of expression shown in different shades of gray was derived by averaging multiple lines). Cell identifications were done based on either characteristic cell body position and/or through labeling specific, known neuron types with a red fluorescent marker (Dil staining or reporter lines). "*" indicates: DIC identification, not confirmed with reporter. For a list of transgenic strains, see Supplemental Information. See also Figure S4.

overlapping with the expression of markers for these other, previously described neurotransmitter identities, with four exceptions (Table 1). The glutamatergic AIM and I5 interneurons are also serotonergic (Jafari et al., 2011; Sawin et al., 2000) and so is the glutamatergic ASG sensory neuron under hypoxic conditions (Pocock and Hobert, 2010). The *eat-4/VGLUT*-expressing RIM motoneurons also employ the monoamine tyramine (Alkema et al., 2005).

Dissecting cis-Regulatory Control Regions of *eat-4/VGLUT* Reveals a Modular Logic of Expression

To dissect the cis-regulatory information content of the *eat-4* locus, we first compared expression of the *eat-4* fosmid reporter and a reporter containing most of the 5' intergenic region upstream of *eat-4*, the *eat-4* locus, and 500 bp of downstream sequences (Figure 2). This reporter is still expressed in all but two neurons classes compared to the fosmid reporter (Figure 2). A previously described, much smaller transcriptional reporter (*adIs1240*) is expressed in a much more restricted manner (Figure 2) (Lee et al., 1999).

We generated a series of reporter genes that encompasses various overlapping and nonoverlapping pieces of the *eat-4* locus and examined their expression pattern in transgenic animals. We find that the broad expression generated from the upstream cis-regulatory region can be broken into much smaller elements that direct expression to very small numbers of specific glutamatergic neuron classes (Figure 2). The modularity of the cis-regulatory control logic is further underscored by a more fine-grained mutational dissection in which we mutated conserved small motifs that constitute predicted binding sites

for transcription factors whose identity we will describe further below. Mutations of such motifs abrogate expression in even smaller numbers of neuron classes, in some cases single neuron classes (Figure 2). The modular structure of cis-regulatory control regions of the *eat-4* locus, with individual cis-regulatory elements driving expression in distinct glutamatergic neuron types, rules out the master regulatory model #1 (Figure 1A) and argues for neuron-type specific control mechanisms (model #2 or model #3 in Figure 1A).

We furthermore note that our mutational analysis did not reveal derepression in other neuronal or nonneuronal cell types, indicating that *eat-4* expression is sculpted by activating rather than repressive regulatory inputs. Our previous analysis of the regulation of other neurotransmitter systems (e.g., genes controlling dopamine or acetylcholine biosynthesis) derived similar conclusions (Flames and Hobert, 2009; Kratsios et al., 2012), thereby corroborating the previously proposed concept that gene activation, rather than gene repression, is the predominant mode of controlling terminal identity features of a neuron (Hobert, 2011).

Known Terminal Selector-Type Transcription Factors Control *eat-4/VGLUT* Expression

To identify the trans-acting factors that operate through the modular cis-regulatory elements in the *eat-4* locus, we first turned to eight distinct, terminal selector-type transcription factors that have previously been shown to define the identity of distinct neuron types that we determine here to be glutamatergic: the *che-1* Zn finger transcription factor (controls ASE gustatory neuron differentiation; [Etchberger et al., 2007]), the *unc-86* POU homeodomain and *mec-3* LIM homeobox genes (light

Table 1. *eat-4* Expressing Neurons, Their Regulators, and Postsynaptic Targets

<i>eat-4</i> -Expressing Neuron Class	<i>eat-4</i> Regulator Identified in This Study	Postsynaptic Target Based on EM Analysis ^{a,b}	Cotransmitter
ADA interneuron	^c	AVB, AVJ, RIM, SMD , RIP	
ADL sensory neuron	<i>lin-11</i> Lim homeodomain gene	AIA, AIB, AVD, AVB, AVA	
AFD sensory neuron	<i>ttx-1/ceh-14</i> Otx-type/LIM homeodomain	AIY	
AIB interneuron		RIM, RIB, AVB, SAAD	
AIM interneuron	<i>unc-86</i> POU homeodomain	AIA, ASG, ASK, ASJ, AVF	Serotonin
AIZ interneuron	<i>unc-86</i> POU homeodomain ^d	RIA, SMB, AIB, RIM, AIY, AVE	
ALM sensory neuron	<i>unc-86/mec-3</i> POU/LIM homeodomain	BDU, PVC , CEP	
AQR sensory neuron	<i>unc-86</i> POU homeodomain	AVA, AVB, RIA, BAG, PVC, AVD	
ASE sensory neuron	<i>che-1</i> Zn finger & <i>ceh-36</i> Otx-type homeodomain	AIY, AIA, AIB, RIA	
ASG sensory neuron	<i>lin-11</i> LIM homeodomain + <i>ceh-37</i> Otx-type homeodomain	AIA, AIB	Serotonin ^e
ASH sensory neuron	<i>unc-42</i> Prd-type homeodomain	AIA, AIB, RIA, AVA, AVB, AVD	
ASK sensory neuron	<i>ttx-3</i> LIM homeodomain	AIA, AIB, AIM	
AUA interneuron	<i>ceh-6</i> POU homeodomain	RIA, RIB, AVA, AVE	
AVM sensory neuron	<i>unc-86/mec-3</i> POU/LIM homeodomain	AVB, PVC , BDU, ADE, PVR	
AWC sensory neuron	<i>ceh-36</i> Otx-type homeodomain	AIY, AIA, AIB	
BAG sensory neuron	<i>ets-5</i> Ets + <i>ceh-37</i> Otx-type homeodomain	RIA, RIB, AVE, RIG	
DVC interneuron	<i>ceh-14</i> LIM homeodomain	RIG, AVA, AIB, RMF	^f
FLP sensory neuron	<i>mec-3</i> LIM homeodomain ^d	AVA, AVD, AVB, AIB, ADE	
IL1 sensory neuron		RMD, RIP	
LUA interneuron	Unknown homeodomain protein ^g	AVA, AVD, PVC, AVJ	
OLL sensory neuron	<i>vab-3</i> Prd homeodomain	SMD, AVE, RIB, RMD, CEP	
OLQ sensory neuron		RMD, RIC, SIB, RIH	
PHA sensory neuron	<i>ceh-14</i> LIM homeodomain	PHB, AVG, PVQ, DVA, AVF, AVH	
PHB sensory neuron	<i>ceh-14</i> LIM homeodomain	AVA, PVC	
PHC sensory neuron	<i>ceh-14</i> LIM homeodomain ^d	DVA, PVC, DA9	
PLM sensory neuron	<i>unc-86/mec-3</i> POU/LIM homeodomain	AVA, AVD, DVA, PDE	
PQR sensory neuron	<i>unc-86</i> POU homeodomain	AVA, AVD	
PVD sensory neuron	<i>mec-3</i> LIM homeodomain	AVA, PVC	
PVQ interneuron	^c	AIA	
PVR interneuron	<i>unc-86/ceh-14</i> POU/LIM homeodomain	AVB, RIP, AVJ	
RIA interneuron		SMD, RMD, RIV	
RIG interneuron		AVE, AIZ, AVK, RIB, BAG, RIR, RMH	
RIM motor neuron		head muscle, SMD, RMD, SAA, AVB	Tyramine
URY sensory neuron	<i>vab-3</i> Pax homeodomain	SMD, RMD, RIB, AVE	
<i>Pharyngeal neurons:</i>			
M3 motor neuron	^c	pharyngeal muscle, M3	
I2 interneuron		NSM, I4, I6, M1	
MI motor neuron		pharyngeal muscle, NSM, M1, M2, M3, MC, I4, I5	
I5 interneuron		M1, M3, M4	Serotonin

In addition to the neurons listed in this table, males show *eat-4* expression in about 20 unidentified male-specific tail neurons. A recent report using a different, smaller *eat-4* reporter also identified AVA, AVE, SIB, RMD and ASJ neurons as *eat-4*-expressing (Ohnishi et al., 2011). We found that expression of this reporter perfectly overlaps with our fosmid-based reporter, but we could not confirm expression in any of these cells.

^aBased on White et al., 1986.

^bBold indicates postsynaptic target expresses ionotropic GluR. *glc/avr* (glutamate-gated ion channel), *glr* (AMPA/Kainate-type), *nmr* (NMDA-type glutamate receptor) (Brockie and Maricq, 2006).

^cCandidate regulators expressed in this cell, but found to have no effect on *eat-4* expression: *unc-86* (ADA), *ceh-2* (M3), *ceh-14* (PVQ).

(legend continued on next page)

touch receptor differentiation; [Duggan et al., 1998]), the *ceh-36* Otx-type homeobox gene (ASE and AWC chemosensory neuron differentiation [Chang et al., 2003; Lanjuin et al., 2003]), the *ttx-1* Otx-type and *ceh-14* LIM homeobox transcription factors (AFD thermosensory neuron identity; [Cassata et al., 2000; Satterlee et al., 2001; H. Kagoshima, personal communication]), the *lin-11* LIM homeodomain transcription factor (ASG chemosensory neuron identity; [Sarafi-Reinach et al., 2001]) and the *ets-5* ETS domain transcription factor (BAG CO₂/O₂ sensory neurons [Brandt et al., 2012; Guillermin et al., 2011]). Each of these transcription factors is continuously expressed in mature neuron types to control the expression of various terminal identity features of these individual neurons (e.g., neurotransmitter receptors, neuropeptides, ion channels, sensory receptors, etc.), but in none of these cases is it known whether the glutamatergic phenotype, i.e., expression of *eat-4/VGLUT*, is affected. We crossed *eat-4* reporter genes into the respective mutant backgrounds and found that each of the eight terminal selector transcription factors is required for *eat-4* expression in the neuron types in which these factors were known to act as terminal selectors (Figure 3; Table 1). Using *mec-3* as an example, we confirmed through temporally controlled addition and removal of *mec-3* that *mec-3* is required continuously to maintain *eat-4* expression (Figure S2A). These findings demonstrate that the induction and maintenance of the glutamatergic phenotype is linked to the induction and maintenance of other terminal identity features of specific glutamatergic neuron types.

Ectopic misexpression of terminal selector-type transcription factors has been shown to impose specific neuronal identities on other cell types (e.g., Flames and Hobert, 2009; Kratsios et al., 2012). Using *che-1*, *mec-3*, and *ceh-36* as examples, we confirmed that misexpression of these terminal selectors is also able to induce ectopic *eat-4/VGLUT* expression (Figures S2D–S2F).

Dual Neurotransmitter Identity Is Coregulated via Common *trans*-Acting Factors

The glutamatergic ASG neuron pair displays the intriguing property of upregulating an additional neurotransmitter system under specific environmental conditions in order to improve chemosensory acuity (Pocock and Hobert, 2010). Specifically, in hypoxic conditions, 5HT antibody staining and expression of tryptophan hydroxylase (*tph-1*), the rate-limiting enzyme of 5HT biosynthesis, are significantly upregulated in ASG. We find that in addition to regulating *eat-4/VGLUT* and other terminal features of ASG, *lin-11* is also required for the upregulation of *tph-1* in ASG under hypoxic conditions (Figure 4A). Therefore, both glutamatergic and serotonergic identity of the ASG neurons are coregulated by a common *trans*-acting factor.

A similar coregulation of dual neurotransmitter identities is observed in the AIM neurons. These neurons were previously

reported to be serotonergic and to require the POU homeobox gene *unc-86* to acquire their serotonergic identity (Jafari et al., 2011) (Figure 4B). We find that *eat-4/VGLUT* expression in AIM is also abolished in *unc-86* mutant animals (Figure 4B). Additionally, we observe loss of the *flp-10* neuropeptide-encoding gene in AIM in *unc-86* mutants, mirroring the previously reported impact of *unc-86* on specific morphological features of AIM (Kage et al., 2005) (Figure 4B). As in the case of *lin-11* and ASG, these findings indicate coregulation of distinct neurotransmitter identities within a single neuron class by a common *trans*-acting factor.

Identification of New Regulators of Glutamatergic Neuronal Identity

We next examined the function of three transcription factors that previous studies had found to be expressed postmitotically and continually in what we define here as *eat-4(+)*, glutamatergic neurons but whose impact on the identity of these neurons has either not been studied in detail or not studied at all.

unc-42 Controls the Identity of the ASH Sensory Neurons

unc-42 encodes a Paired-type homeodomain transcription factor related to mammalian PROP1 (Baran et al., 1999). *unc-42* is expressed in the ASH nociceptive sensory neurons and its expression persists throughout adulthood due to autoregulation (Baran et al., 1999). *unc-42* is required for the expression of two orphan seven transmembrane receptors of the odorant receptor family, *sra-6* and *srb-6*, in ASH (Baran et al., 1999) but since the expression of odorant receptor family members in sensory neurons is strongly activity dependent (Nolan et al., 2002; Peckol et al., 2001), it was unclear if *unc-42* broadly affects neuronal identity or whether it has a narrower role in controlling sensory responsiveness or neuronal activity. We find that *unc-42* controls *eat-4/VGLUT* expression in ASH (Figure 3F). Through postembryonic removal of *unc-42*, we found that *unc-42* is continuously required to maintain *eat-4* expression (Figure S2B). *unc-42* also affects the expression of all other terminal identity markers tested, including the three G α -encoding genes *gpa-11*, *gpa-13*, *gpa-15* and the neuropeptide *flp-21* (Figure 4C). The loss of terminal identity features is not indicative of a loss of these neurons since ASH neurons still take up dye in *unc-42* mutants (Baran et al., 1999).

ceh-6 Controls the Identity of the AUA Interneurons

ceh-6 is a POU homeobox gene expressed in a small number of head neuron classes (Bürglin and Ruvkun, 2001), one of them the AUA interneuron class, which regulates aggregation behavior (Coates and de Bono, 2002). *ceh-6* expression in AUA and other head neurons is maintained throughout adulthood (Bürglin and Ruvkun, 2001). We find that in *ceh-6* null mutant animals *eat-4/VGLUT* expression in the AUA neurons is abrogated (Figure 3G). We tested a number of additional terminal markers of AUA identity (neuropeptide-encoding genes *flp-8*, *flp-10*, and *flp-11*) and

^dApart from earlier functions of *unc-86* in this lineage, *unc-86* may also have late roles in terminal differentiation. See Figure S6 for more information on this subject.

^eUnder hypoxic conditions (Pocock and Hobert, 2010).

^fReported to be cholinergic (Duerr et al., 2008) but based on *unc-17* and *cho-1* reporter expression, we believe that DVA and not DVC is cholinergic.

^gInferred from requirement of TAAT homeodomain consensus site for *eat-4* expression in LUA.

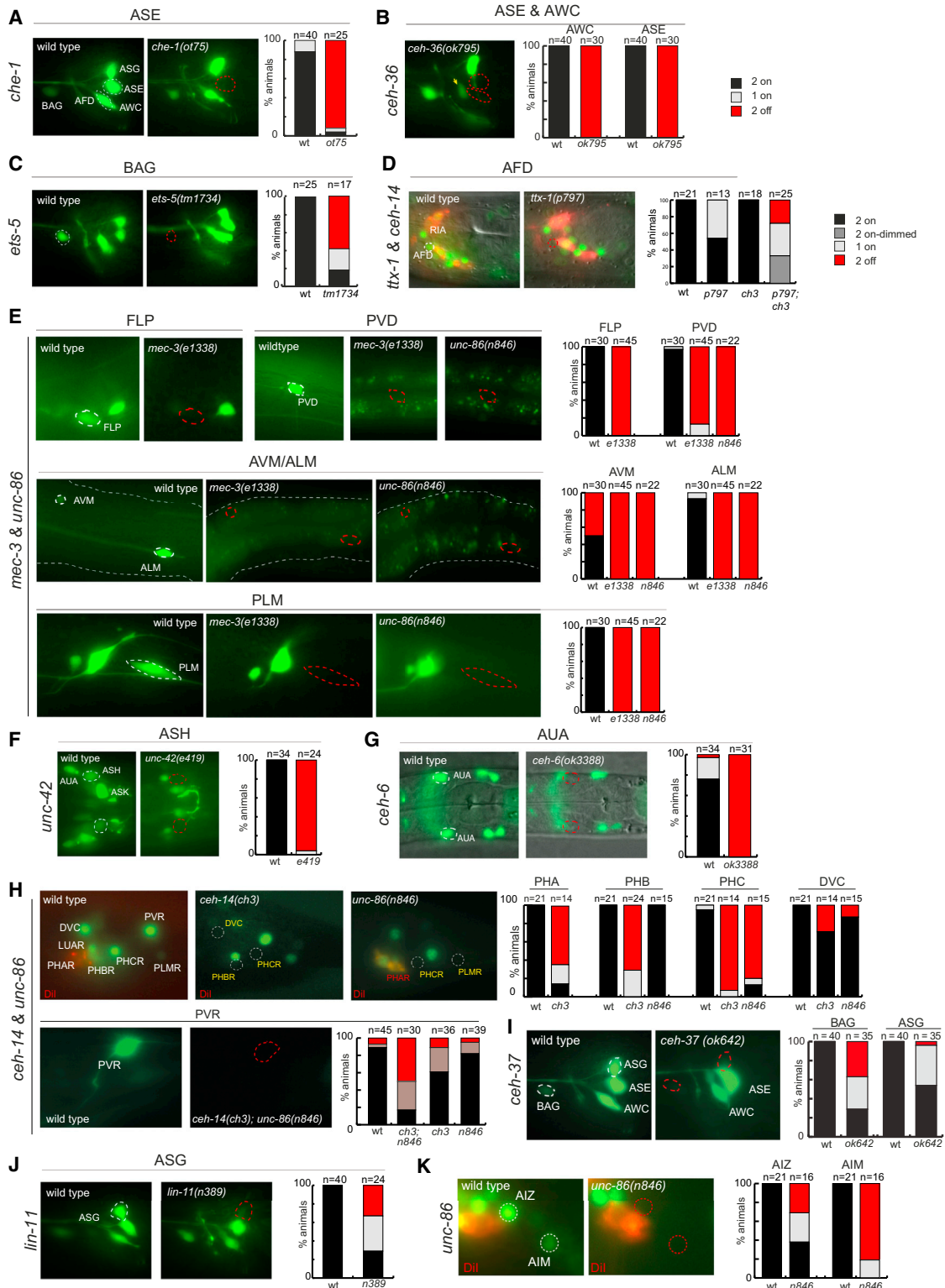


Figure 3. Regulators of eat-4/VGLUT Expression

eat-4 reporter lines were crossed into the null mutant backgrounds indicated. Percentage (%) of animals (n = 20–40) that express the reporter in both cells of the respective left/right neuron pair (“2 on”) or in one of the two neurons of a neuron pair (“1 on”) or in neither (“0 on”) are indicated.

(A–E) Known terminal selectors of individual neuronal identities control *eat-4* expression. See also Figure S2. (A) *che-1* affects *eat-4* expression (assayed with *eat-4^{prom6}* reporter transgene *otIs392*) in the ASE neurons. (B) *ceh-36* affects *eat-4* expression (assayed with *eat-4^{prom6}* reporter transgene *otIs392*) in the ASE and (legend continued on next page)

found that the expression of each of them was also strongly affected in the AUA neuron class of *ceh-6* mutants (Figure 4D). The AUA neuron is not absent in *ceh-6* mutants, since we find that expression of the panneuronal marker *rab-3* is unaffected in AUA (Figure 2G). Like other terminal selectors (Hobert, 2011), *ceh-6* therefore affects the adoption of a specific identity while panneuronal identity is unaffected.

***ceh-14* Controls the Identity of Phasmid Sensory Neurons**

ceh-14 is a LIM homeobox gene that is expressed in a small number of head and tail neurons, including the PHA, PHB, and PHC tail phasmid sensory neurons (Cassata et al., 2000). Using a fosmid-based reporter, we confirmed that *ceh-14* expression is maintained throughout the life of these neurons (data not shown). We find that *eat-4* expression is affected in all three phasmid sensory neurons of *ceh-14* null mutants (Figure 3H). The expression of additional identity markers for PHA and PHB (GPCR *srb-6* and neuropeptide *flp-4*) and PHC (tyrosine phosphatase *ida-1* and the dopamine receptor *dop-1*) is also affected in *ceh-14* mutants (Figures 4E and 4F). PHA and PHB also display dye filling defects in *ceh-14* mutants (Kagoshima et al., 2013). Panneuronal features (*rab-3* expression) of the phasmid neurons are unaffected in *ceh-14* null mutants (Figure S2H).

We also observed an effect of loss of *ceh-14* on *eat-4* expression in the DVC interneurons (Figure 4H). Another DVC cell fate marker, the GPCR *srab-12*, also displays defective expression in DVC (Figure 5G). We observed no defects in *eat-4* expression in the normally *ceh-14* expressing PVQ neurons of *ceh-14* mutant animals. Joint removal of *ceh-14* and *ttx-1*, a previously described regulatory of AFD neuron identity (Satterlee et al., 2001), results in loss of *eat-4* expression in AFD (Figure 3D). *ceh-14* and *ttx-1* also collaborate in the activation of other AFD-expressed terminal effector genes (H. Kagoshima, personal communication).

***vab-3* Controls the Identity of the OLL and URYV Neurons**

In addition to pursuing candidate genes, we identified an additional regulator of glutamatergic neuron identity through an unbiased screen for EMS-induced mutants in which the glutamatergic OLL neurons, sensors of bacterial pathogens (Chang et al., 2011), do not adopt their identity. This screen identified the *vab-3* Paired homeobox gene (Figure 5; see Extended Experimental Procedures). In *vab-3* mutants, the expression of *eat-4*, as well as the tyramine receptor *ser-2*, the acetylcholinesterase *ace-1*, and the groundhog gene *grd-8* fail is affected in the OLL neurons (Figure 5A). The OLL neurons are generated in these animals as assessed by unaffected expression of both *ift-20* and

the panneuronal *rab-3* marker in the anterior ganglion (Figure S2I).

A fosmid-based *vab-3* reporter is expressed in the OLL neurons, and this expression is maintained throughout the life of these neurons, consistent with a role of *vab-3* in not only initiating but also maintaining the OLL differentiation program (Figures 5B and 5C). Apart from OLL, we also observe expression of *vab-3* in the glutamatergic ventral URY neuron class, which are lineally unrelated sensory neurons of unknown function. Expression of *eat-4/VGLUT* expression, as well as the expression of two additional markers of ventral URY fate, the *pdf-1* neuropeptide receptor and the Toll receptor *tol-1* is lost in *vab-3* mutants (Figure 5A).

Taken together, our candidate as well as genetic screening approaches have revealed four additional regulators of the identity of multiple distinct glutamatergic neuron types. Each factor is expressed in the respective neuron class during its entire post-mitotic life span. Each of these factors not only controls *eat-4* expression but also several additional identity features of the respective glutamatergic neuron type. Due to their broad effect on distinct terminal identity features, each of these factors can therefore be considered a terminal selector-type transcription factor.

Redundancy of Glutamatergic Identity Regulators

The expression of a number of the transcriptional regulators of glutamatergic identity that we have identified here is restricted to one class of glutamatergic neurons (*che-1* in ASE, *ets-5* in BAG, *ttx-1* in AFD, and *mec-3* in touch neurons). In contrast, *unc-86*, *lin-11* and *ceh-14* are expressed in multiple very distinct glutamatergic neuron classes, and we examined whether these regulators generally affect glutamatergic identity in all neurons in which they are expressed. Genetic elimination of *unc-86* results in no effect on *eat-4/VGLUT* expression in the URY sensory neurons, the I2 pharyngeal interneurons or the tail sensory neuron PVR (Figure S3). Similarly, the LIM homeobox gene *ceh-14* affects glutamatergic identity in PHA, PHB, and PHC, but not in PVQ or PVR neurons (Figure 3H and data not shown). Likewise, the LIM homeobox gene *lin-11* affects ASG glutamatergic identity, but not AIZ glutamatergic identity (data not shown).

We considered the possibility that potential functions of these regulators may be masked by redundancies with other glutamatergic regulators. To address this possibility, we specifically focused on the PVR tail sensory neuron because it coexpresses *unc-86* and *ceh-14*, each of which acts as a glutamatergic

AWC neurons. The control for the *ceh-36* panel is shown in (A). (C) *ets-5* affects *eat-4* expression (assayed with *eat-4^{prom6}* reporter transgene *otls392*) in the BAG neurons. (D) *ttx-1* affects *eat-4* expression (assayed with *eat-4* fosmid reporter transgene *otls388*) in the AFD neurons. The animals shown were also stained with Dil, which labels other sensory neurons to facilitate cell identification. (E) The collaborating *mec-3/unc-86* genes affect *eat-4* expression (assayed with the reporter transgene *adls1240*) in different types of touch receptor neurons. The effect of *unc-86* was not examined in FLP neurons due to a previously described lineage transformation that eliminates the FLP neuron. See also Figure S3.

(F–K) Novel regulators of specific neuronal identities. (F) *unc-42* controls *eat-4* expression (assayed with *eat-4^{prom2}* reporter transgene *otls376*). (G) *ceh-6* controls *eat-4* expression (assayed with *eat-4^{prom2}* reporter transgene *otls376*). (H) The LIM homeobox gene *ceh-14* and the POU homeobox gene *unc-86* regulate glutamatergic identity of several tail sensory neurons. The effect of *unc-86* in PHC could be the result of *unc-86* function at earlier stages in the PHC-producing lineage (see Figure S3). The effect of *ceh-14* and *unc-86* either alone or in combination on *eat-4* expression in PVR is assayed using the *eat-4^{prom10}* reporter transgene *otEx5301*. (I) *ceh-37* affects *eat-4* expression (assayed with *eat-4^{prom6}* reporter transgene *otls392*) in the BAG and ASG neurons. (J) *lin-11* affects *eat-4* expression (*eat-4^{prom6}* transgene *otls392*) in the ASG neurons. (K) *unc-86* affects *eat-4* expression (*eat-4* fosmid reporter transgene *otls388*) in the AIZ and AIM neurons. See Figure S3 for more notes on *unc-86* function in AIZ.

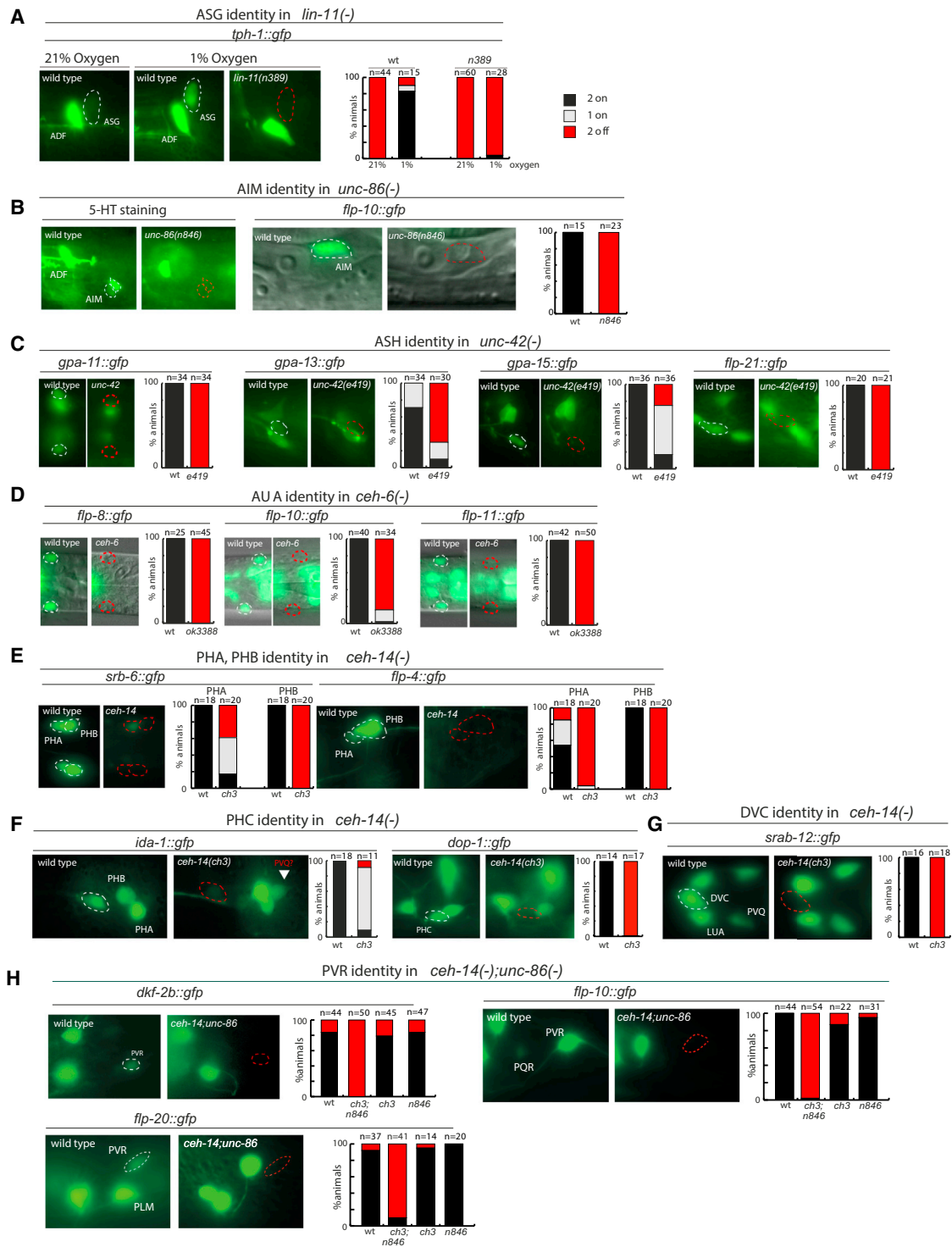


Figure 4. Adoption of Glutamatergic Identity Is Linked to the Adoption of other Neuronal Identity Features

(A) In the ASG sensory neurons, *lin-11* affects expression of the serotonergic marker *tph-1* that is induced under hypoxic conditions in ASG (*tph-1::gfp* transgene *zds13*).

(B) *unc-86* affects serotonin staining and another identity marker of AIM (*flp-10::gfp* transgene *otIs92*).

(C) *unc-42* controls several features of ASH identity (*gpa-11^{prom2}::gfp* transgene *otEx5336*, *gpa-13::gfp* transgene *otEx213*, *gpa-15::gfp* transgene *pkIs591*, *flp-21::gfp* transgene *ynIs80*). See also Figure S2.

(D) *ceh-6* controls several features of AUA identity (*flp-8::gfp* transgene *ynIs2022*, *flp-10::gfp* transgene *otIs92*, *flp-11::gfp* transgene *ynIs40*).

(legend continued on next page)

identity regulator in other neuron types, as described above. PVR coexpresses *unc-86* and *ceh-14* throughout its life, consistent with a role for these factors in controlling the terminal differentiation state of PVR (data not shown). While neither single *unc-86* or *ceh-14* null mutations affects *eat-4* expression, *unc-86;ceh-14* double mutants show a dramatic loss of *eat-4* expression in PVR (Figure 3H). Three additional terminal markers of PVR identity (*flp-10*, *flp-20*, *dkf-2*) are also redundantly controlled by *ceh-14* and *unc-86* (Figure 3H). The loss of marker gene expression in *unc-86; ceh-14* double mutants is not due to loss of the neurons, since we were able to detect generation of the cell by lineage analysis using Nomarski optics. Taken together, the redundancy of *unc-86* and *ceh-14* function in PVR suggests that glutamatergic neurons not affected by a transcription factor that has a glutamatergic control function in other neuron types may be masked by redundantly acting factors.

Requirement for Specific *cis*-Regulatory Motifs in the *eat-4/VGLUT* Locus Argues for Direct Regulation by Specific Transcription Factors

We next asked whether any of the glutamatergic identity regulators may control *eat-4/VGLUT* expression directly. To this end, we made use of the knowledge of the binding sites for six of the above mentioned transcription factors, ETS-5, UNC-86, CEH-6, CEH-36, TTX-1, and CEH-14 (Baird-Titus et al., 2006; Brandt et al., 2012; Duggan et al., 1998; Etchberger et al., 2007; Kim et al., 2010; Wingender et al., 1996). We indeed found binding sites for each of these factors in the *eat-4* modular elements that drive expression in the respective neuron types, and we found that deletion of these binding sites resulted in a loss of expression in the respective neuron types (Figure 2), thereby providing a strong indication that these transcription factors directly control *eat-4* expression. For example, deletion of an Otx/K50-type homeodomain site (TAATCC) affects expression exclusively in the ASE and AWC neuron classes (Figure 2), which both require *ceh-36/Otx* for correct *eat-4* expression (Figures 3A and 3B). Mutation of another K50-type homeodomain binding also affected *eat-4* expression in the ASE and AWC neurons, but additionally in the AFD neurons, which require the K50-type *ttx-1/Otx* homeobox gene for correct *eat-4* expression (Figures 2 and 3). Mutation of an ETS domain binding site affected expression exclusively in the BAG neurons that require the ETS-5 transcription factor for *eat-4* expression (Figure 2). Mutation of a predicted POU homeodomain sites affected *unc-86*- and *ceh-6*-dependent *eat-4* expression in touch neurons, PVR, and AUA (Figure 2). Lastly, *ceh-14*-dependent *eat-4* expression in PHA, PHB, PHC, and DVC expression is controlled by a *cis*-regulatory module, *eat-4^{prom5}*, and ModEncode data (Niu et al., 2011) reveals several CEH-14 binding peaks that map onto this module (Figure S4), providing additional support for *ceh-14* directly regulating *eat-4/VGLUT*.

Redeployment of the Same *cis*-Regulatory Motif in Distinct Neuron Types

Our *cis*-regulatory analysis also revealed that distinct transcription factor family members apparently use the same *cis*-regulatory sites to control *eat-4* expression in distinct neuron types. For example, deletion of the putative POU homeobox site in the *eat-4^{prom2}* module affects expression of *eat-4/VGLUT* in the light and harsh touch receptor neurons that require the POU homeobox gene *unc-86* for correct *eat-4* expression. This deletion also affects *eat-4/VGLUT* expression in the AUA neurons that require the POU homeobox gene *ceh-6* for correct *eat-4* expression (Figures 2 and 3). The same scenario applies for TAATCC K50-homeodomain sites in the *eat-4^{prom6}* module: as mentioned above, deletion of one of the two sites abrogates *ceh-36^{K50-homeodomain}*-dependent *eat-4* expression in the ASE and AWC neurons, while a mutation in the other site also abrogates *ttx-1^{K50-homeodomain}*-dependent expression in AFD (Figure 2). Intriguingly, mutation of this second TAATCC site also abrogates *eat-4* expression in the BAG and ASG neurons, suggesting that these neurons utilize another K50 homeodomain protein.

We tested whether *ceh-37*, the third Otx/K50 ortholog in *C. elegans* besides *ceh-36* and *ttx-1*, could be involved in controlling *eat-4* expression in ASG and BAG. *ceh-37* is expressed in the nonglutamatergic AWB neurons as well as ASG and BAG (Lanjuin et al., 2003; Y.G. Tong and T. Bürglin, personal communication). We find that *ceh-37* is indeed required for *eat-4* expression in both the ASG and BAG neuron classes (Figure 3I), likely through its cognate binding site TAATCC. Hence, the glutamatergic identity of five different neuron classes is controlled by three distinct Otx-type homeodomain transcription factors (*ceh-36*: AWC and ASE, *ttx-1*: AFD, *ceh-37*: ASG and BAG), operating through at least partially shared TAATCC sites. Each of these Otx factors may collaborate with distinct cofactors in these distinct cell types (e.g., *ceh-37* with *ets-5* in BAG and with *lin-11* in ASG; Figure S5).

Additional Homeodomain Regulators of *eat-4* Expression

The striking preponderance of homeodomain transcription factors in the collection of transcription factors that we found to control *eat-4* expression prompted us to search for additional predicted homeodomain binding sites that may be required for *eat-4* expression in neurons for which we had not yet identified regulatory factors. We focused on the *eat-4^{prom2}* module, which drives expression in a number of distinct neuron classes (Figure 2). This module contains two sets of palindromic, predicted homeodomain binding sequences (ATTAN₂₋₃TAAT). Mutation of one of these homeodomain binding palindromes resulted in complete elimination of *eat-4* reporter expression in the ASK neurons, while mutation of the other homeodomain binding palindrome (located ~200 bp away) resulted in very strong reduction of *eat-4* expression in the ADL sensory neurons (Figure 2). ASK

(E–G) *ceh-14* controls several identity features of the PHB, PHC, and DVC neurons (*ida-1::gfp* transgene *inls179*, *dop-1::gfp* transgene *vsls28*, *srab-12::gfp* transgene *sEx12012*).

(H) *ceh-14* and *unc-86* redundantly control the identity of the PVR neuron (*flp-10::gfp* transgene *otIs92*, *flp-20::gfp* transgene *ynIs54*, *dkf-2b::gfp* transgene *otEx5323*).

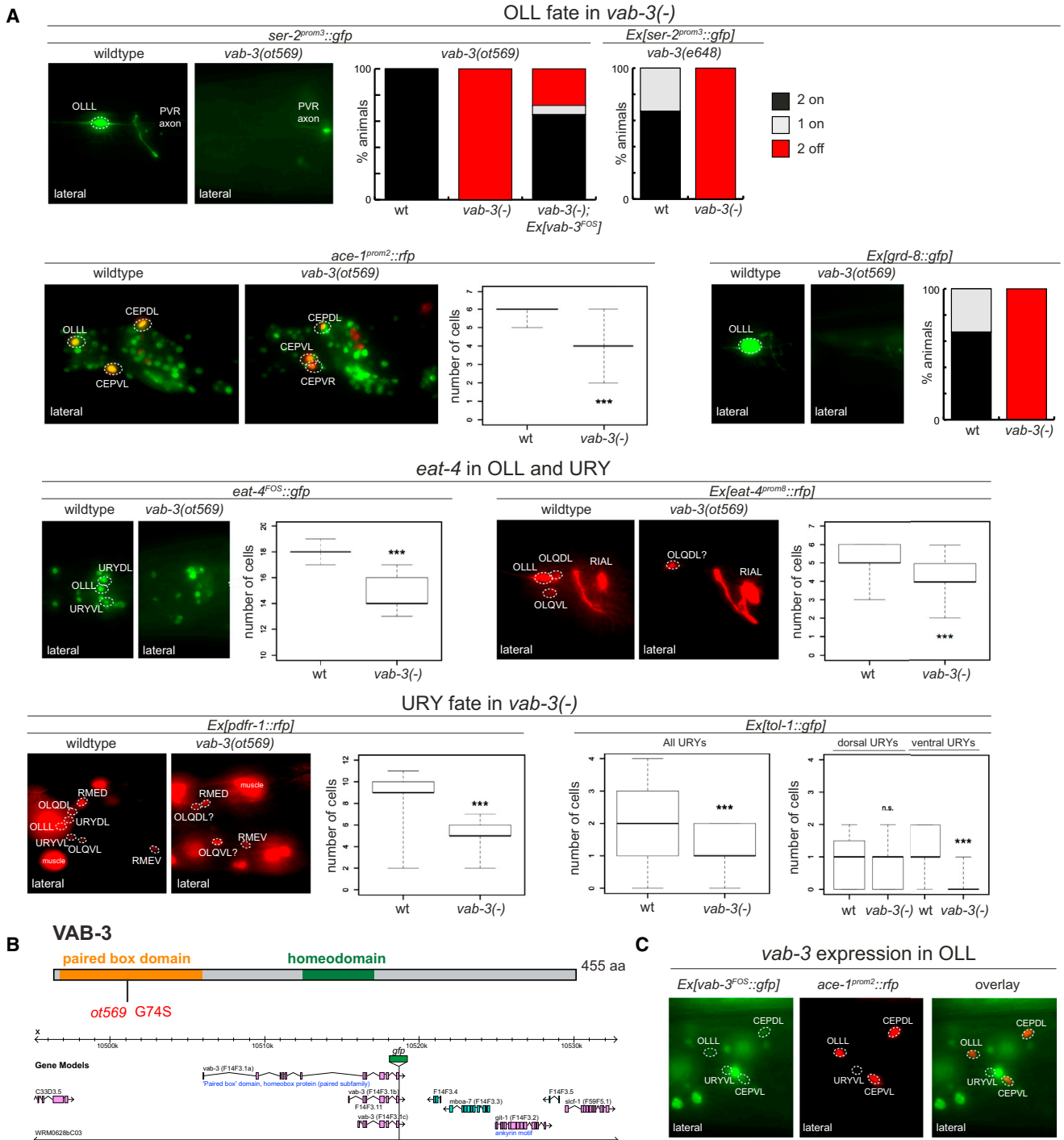


Figure 5. *vab-3* Is a Novel Candidate Terminal Selector for the Glutamatergic OLL and URY Neurons

(A) *vab-3(ot569)* affects *eat-4* expression in the OLL and URY neurons, as well as other identity markers of OLL and URY. Standard whisker-and-box plots of total counts of *gfp*-expressing cells in the anterior ganglia are presented. The *vab-3(ot569)* allele was used except where indicated otherwise.

(B) Diagram of VAB-3 protein structure and the genomic locus encompassed by the rescuing fosmid WRM0628bC03.

(C) Expression of *vab-3* in adult OLL and URY neurons. The OLL neurons were identified by overlap with *ace-1::tagrpf* and the URY neurons were identified by position and cell body morphology.

and ADL express distinct LIM homeobox genes that may operate through these sites. Based on a fosmid reporter gene provided by the ModEncode consortium, the ASK neurons express and maintain the LIM homeobox gene *ttx-3*, the worm ortholog of the vertebrate LHX2/9 gene (F.Z. and O.H., unpublished data). *eat-4/VGLUT* expression is abolished in the ASK neurons of *ttx-3* mutants (Figure S6). The ADL neurons express the LHX1/5 ortholog *lin-11* (Hobert et al., 1998) and we find *eat-4/VGLUT* expression in this neuron class to be reduced in *lin-11* mutants (Figure S6). In the case of ASK and *ttx-3*, we examined four additional terminal identity features of ASK in *ttx-3* mutants (neuropeptides *flp-13* and *nlp-14*, guanylyl cyclase *gcy-27* and dye filling of the neuron) and found expression of each identity feature to be also defective (Figure S6). Hence, neurotransmitter identity is also linked in this case to the adoption of other identity features through coregulation of these distinct features by *ttx-3*.

Mutation of yet another predicted homeodomain binding site in the *eat-4^{prom2}* cis-regulatory module eliminated expression of *eat-4* exclusively in the LUA tail interneurons (Figure 2). It is not yet known which homeobox genes are expressed in LUA.

Potential Conservation of Homeodomain Regulation of Glutamatergic Identity in Mouse

The preponderance of homeodomain transcription factors among the glutamatergic identity regulators in *C. elegans* is remarkable. Every single glutamatergic neuron differentiation program that we have described here depends on at least one homeodomain protein. Eleven out of the 13 factors (85%) described here to control terminal neuron identity are homeodomain transcription factors. This proportion far exceeds the proportions (~11%) of homeodomain transcription factors (~100) relative to other types of transcription factors (~900) in the *C. elegans* genome. This preponderance of homeodomain regulators of glutamatergic identity in *C. elegans* prompted us to ask whether homeodomain proteins may also control glutamatergic identity in vertebrates. Indeed, the Otx-type protein CRX is known to control the identity of glutamatergic photoreceptors in the mouse retina (Furukawa et al., 1997). To show that homeodomain proteins may be more commonly employed as regulators of glutamatergic neurons, we examined the expression of mouse homologs of the two most prominent homeodomain subtypes that we identified as glutamatergic regulators in *C. elegans*, Brn-type POU homeodomain proteins (*unc-86* and *ceh-6* orthologs) and LIM homeodomain proteins (*ceh-14*, *mec-3*, *lin-11*, and *ttx-3* orthologs). We focused on examining expression in adult neurons because we aimed to avoid identifying factors with only transient roles in glutamatergic neuron development and because our *C. elegans* studies suggested that glutamatergic identity regulators are continuously expressed throughout the life of the neuron to maintain their identity.

We stained adult mice with antibodies directed against three Brn-type POU (BRN3a,b,c) and four LIM-type homeodomain proteins (LHX1, LHX2, LHX3, and LHX5). We found expression of five of these seven proteins in adult CNS neurons (data not shown). We then made use of the observation made in *C. elegans* that in several cases LIM and POU homeodomain work together to determine glutamatergic identity (*unc-86/mec-3* in touch neurons; *unc-86/ceh-14* in PVR neurons) and

sought to identify regions of overlap of POU and LIM homeodomain expression in adult glutamatergic CNS neurons. Antibody costaining revealed that the BRN3a and LHX1 are coexpressed in two sets of adult glutamatergic neuron types, one in the hippocampus (data not shown) and one in the largest nucleus of the olivary body, the inferior olive (Figure 6A). To examine whether one of these genes may indeed have a role in these glutamatergic neurons, we used a floxed, tamoxifen-inducible LHX1 knockout allele (Kwan and Behringer, 2002) to conditionally remove LHX1 in 8-week-old mice. We find that 10 days after tamoxifen treatment, BRN3a and VGLUT2-expressing glutamatergic neurons have disappeared from the inferior olive (Figure 6B). Whether these cells have died because LHX1 is directly involved in controlling survival or whether they have died due to a neuronal identity loss is not clear at present. In either case, this data strongly implies an important function of LHX1 in defining the existence of this glutamatergic neuron type.

As it is the case for *C. elegans* LIM homeobox genes, there are glutamatergic CNS neurons that do not express LHX1 and, vice versa, LHX1 is also expressed in nonglutamatergic neurons (Figure S7) (Lein et al., 2007; Zhao et al., 2007). This supports the existence of neuron-type specific combinatorial codes for the regulation of glutamatergic identity in the vertebrate nervous system akin to what we observe in the *C. elegans* system.

DISCUSSION

We have revealed insights into how glutamatergic neuron identity is controlled. Through mutational analysis of the cis-regulatory control regions of *eat-4/VGLUT* as well as through genetic loss of function analysis of trans-acting factors, we have ruled out the possibility that *eat-4/VGLUT* expression is controlled by one global, glutamatergic regulator in *C. elegans* (see model #1 in Figure 1A). Instead, the *eat-4/VGLUT* locus operates as an integrator device that samples distinct regulatory inputs in distinct cellular contexts through a modular arrangement of cis-regulatory elements.

We have described here more than a dozen terminal selector-type transcription factors that act through these modular cis-regulatory elements to control VGLUT expression in two thirds of all glutamatergic neurons (Table 1). All of these transcription factors are continuously expressed throughout the life of the respective neuron type suggesting they do not only initiate but also maintain VGLUT expression. Moreover, all the transcription factors described here affect not only VGLUT expression, and hence the glutamatergic phenotype, but also a number of additional terminal differentiation genes that define the specific identity features of distinct glutamatergic neurons. This is in line with model #3 described in the introductory Figure 1A and further supports the terminal selector concept that posits the existence of master regulatory-type transcription factors that coregulate a multitude of terminal identity features of a mature neuron (Hobert, 2011). In the absence of these terminal selectors, neurons appear to remain in an undifferentiated neuronal ground state and do not usually display obvious switches in identity (see Supplemental Information).

Coregulation of neuronal identity features also extends to multiple transmitter identities of an individual neuron type, as

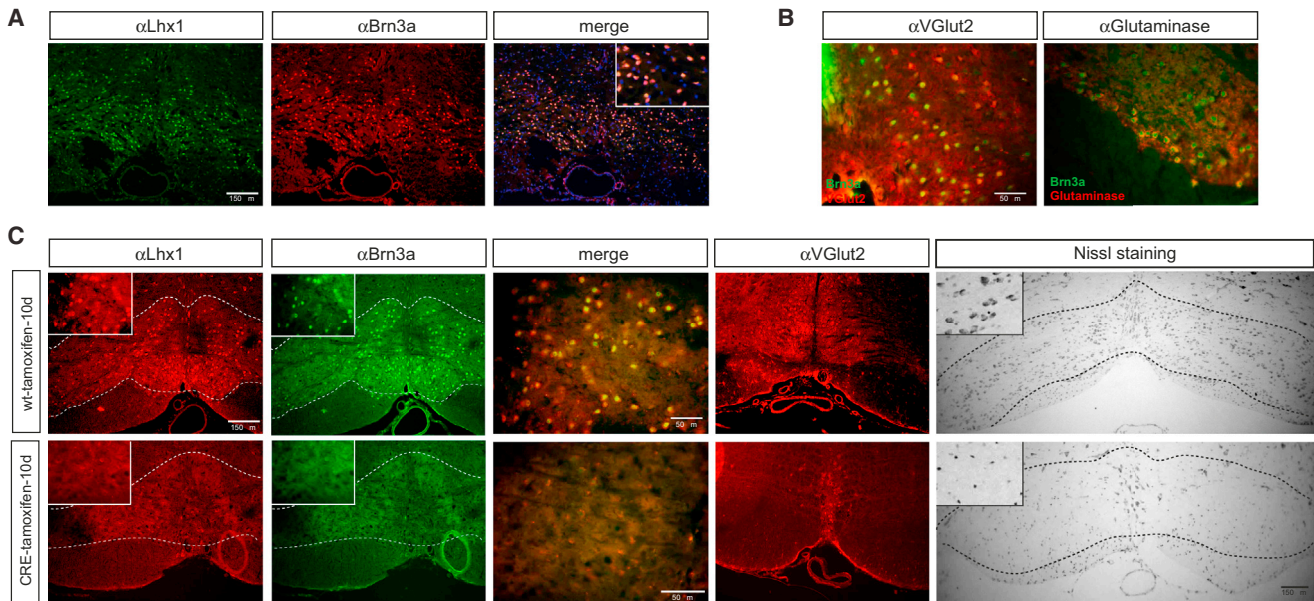


Figure 6. The Loss of Lhx1 Gene in the Adult Mouse Brain Affects the Survival of Glutamatergic Neurons in the Inferior Olive

(A) LHX1 and BRN3A colocalize in neurons of the inferior olive in the brain stem. See also Figure S7.

(B) Double immunostaining of VGLUT2 and glutaminase with BRN3A demonstrates the glutamatergic identity of the BRN3A/LHX1-positive neurons in the inferior olive (yellow overlap).

(C) Loss of glutamatergic inferior olive neurons upon postdevelopmental removal of *LHX1*. Immunostaining analysis and Nissl staining of the inferior olive nucleus in *LHX1^{flox};R26CreER* animals compared to *LHX1^{flox}* siblings. Eight- to ten-week-old adult animals were administered tamoxifen for 10 days and analyzed shortly thereafter. In the wild-type, tamoxifen-treated animals the neurons show LHX1, BRN3a, and VGLUT2 staining (upper), but *LHX1^{flox};R26CreER* animals do not. Nissl staining reveals that in these animals there is a massive loss of cells in the inferior olive nucleus compared with the wild-type brains. Five animals were analyzed for each group.

exemplified by *lin-11* and *unc-86*, which control the glutamatergic/serotonergic identity of the ASG sensory and AIM interneurons, respectively. Neurons have long thought to be only employing one neurotransmitter system (“Dale’s principle”), but over the last several years more than one transmitter system has been found to be employed simultaneously in a number of distinct neuron types in vertebrates and invertebrates, including the cousage of glutamate and serotonin in some regions of the vertebrate CNS (Seal and Edwards, 2006). However, in none of these cotransmitter cases has it been reported whether the two neurotransmitter phenotypes of a given neuron type are independently regulated or controlled by the same *trans*-acting factor. We have provided here two examples of coregulation of two distinct neurotransmitter identities by a common *trans*-acting factor.

Our analysis of *trans*-acting factors also illustrates the combinatorial coding nature of neuronal identity control. This is illustrated with the POU homeobox gene *unc-86* and the LIM homeobox genes *mec-3* and *ceh-14*. *unc-86* and *mec-3* cooperate to control glutamatergic touch neurons, such as the ALM neurons. In the PVR neuron and possibly also in the PHC neurons, *unc-86* cooperates with another LIM homeobox gene, *ceh-14*. In yet other neurons (phasmid neurons), *ceh-14* controls glutamatergic identity independently of *unc-86* (which is not expressed in phasmid neurons).

The observation that mutation of the same *cis*-regulatory motif can affect *eat-4* expression in multiple distinct neuron types indi-

cates that there are limits to modularity. That is, *cis*-regulatory information is not always encoded by distinct *cis*-regulatory modules but may be encoded by a similar grammar to be read out by different *trans*-acting factors in different neuron types. For example, the same POU homeobox site is apparently recognized by *unc-86* in light touch receptor neurons and by *ceh-6* in the AUA neurons. Three distinct Otx-type transcription factors read out the same *cis*-regulatory motif to control *eat-4* expression in distinct sensory neuron classes. Each of these Otx genes appear to cooperate with distinct cofactors in different neuron types (Figure S5). Otx genes are expressed in several distinct sensory neuron structures in the mouse (Acampano et al., 2001), most of which likely use glutamate as neurotransmitter.

Every single *C. elegans* glutamatergic neuron differentiation program that we have described here depends on at least one homeodomain protein. This remarkable preponderance of homeodomain transcription factors led us to explore the expression and function of homeodomain transcription factors in terminal differentiation of mouse glutamatergic neurons. Following the *C. elegans* lead of POU and LIM homeodomain proteins working together in distinct glutamatergic neuron types, we identified a glutamatergic neuron type in the inferior olive of the brainstem that coexpresses POU and LIM homeobox genes and requires the LIM homeobox genes LHX1 for its continuous presence. Even though our mouse findings do not yet prove that vertebrate POU/LIM homeodomain proteins directly activate VGLUT

expression, they are consistent with such a notion. The selective loss of inferior olive glutamatergic neurons and the restricted expression of LHX1 in only some glutamatergic neuron types in the CNS is also consistent with a conservation of the modular, piece-meal regulatory logic of *VGLUT* regulation.

Previous work on terminal selector-type transcription factors, further extended here, has shown that they operate through simple *cis*-regulatory motifs (Hobert, 2011). Terminal differentiation genes that are expressed in multiple distinct neuron types, such as the *eat-4/VGLUT* gene described here, contain a modular assembly of simple terminal selector motifs that are read out in individual neuron types by specific terminal selectors. Gene expression profiles may be able to rapidly evolve through the gain and loss of terminal selector motifs. In the context of *eat-4/VGLUT* this means that on an evolutionary time scale the glutamatergic phenotype of a neuron can be rapidly gained (or lost) through the acquisition (or loss) of terminal selector motifs in *eat-4/VGLUT*. Compared to other neurotransmitters, glutamate is different because its employment as a neurotransmitter does not require the presence of a specialized synthesis and recycling machinery; rather, the only determinant of the glutamatergic phenotype is the expression of *VGLUT* (Takamori et al., 2000, 2001). Hence, to gain a glutamatergic phenotype, only the *VGLUT* locus rather than an entire pathway of neurotransmitter synthesizing enzymes and transporters needs to acquire responsiveness to a neuron-type specific terminal selector. Since glutamate is a very widely employed neurotransmitter in many different nervous systems, our findings—and the terminal selector gene concept in general—provide a straight-forward conceptual framework for how neurotransmitter phenotypes and neuronal gene expression patterns in general can rapidly evolve to generate the enormous diversity of neuronal cell types.

EXPERIMENTAL PROCEDURES

C. elegans Strains and Transgenes

A list of strains and transgenes can be found in the [Supplemental Information](#).

eat-4/VGLUT Reporter Transgenes

The *eat-4* fosmid reporter was generated by fosmid recombineering using fosmid WRM0623aF12 and an SL2-based, nuclear localized *yfp* reporter (Tursun et al., 2009). The *eat-4* locus reporter was generated by *in vivo* recombination (Boulin et al., 2006), using two overlapping fragments of the *eat-4* locus (see [Extended Experimental Procedures](#)).

The 5' *eat-4* reporter constructs were generated by PCR and subcloning into the standard pPD95.75-based expression vectors and mutagenized with the QuickChange XL Site-Directed Mutagenesis Kit (Stratagene). Constructs were injected at 50 ng/μl with *rol-6(su1006)* or *ttx-3::dsRed* as coinjection marker. The resulting transgenic arrays are listed in the [Supplemental Information](#). All strains were scored as young adults.

Genetic Screen

otIs138 transgenic animals (*ser-2^{prom}::gfp*) were EMS-mutagenized and animals with loss of expression in OLL were isolated a Copas Biosort machine (Doitsidou et al., 2008). Whole-genome sequencing followed by data analysis with MAQGene (Bigelow et al., 2009) was used to determine the molecular identity of *ot569*, a new *vab-3* allele. See [Supplemental Information](#).

Analysis of Serotonergic Fate of ASG Neurons

Young adult worms were incubated at 1% oxygen for 24 hr at 25°C in a hypoxic semisealed chamber (oxygen levels were controlled by a ProOx P110 compact oxygen controller [BioSpherix]) and compared to 21% oxygen incubated

worms at 25°C. Antibody staining was performed using a tube fixation protocol described in more detail in the [Supplemental Information](#).

Mouse Genetics

We used two previously generated mouse lines, LHX1^{fllox} (Kwan and Behringer, 2002) and ROSA26CreER (Badea et al., 2003). Tamoxifen was administered orally in the diet at a dose of 80 mg/kg/day (Harlan tamoxifen diet). Animals were treated when they reached 8–10 weeks of age. After 10 days they were perfused and analyzed. For each condition, five animals were used.

Mouse brain sections were stained with antibodies, *in situ* hybridization probes, or other strains using standard procedure, described in the [Supplemental Information](#), which also contain details on antibodies and probes.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Results, Extended Experimental Procedures and six figures and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2013.09.052>.

ACKNOWLEDGMENTS

We thank Q. Chen for expert assistance in generating transgenic strains, A. Goldsmith for generating the *eat-4* locus reporter, members of the *C. elegans* community for providing strains, A. Boyanov for expert assistance in whole-genome sequencing, G. Minevich for initial identification of three mutants that affect *ser-2* expression in OLL, J. Crissman, L. Khirman, M. Peterson, J. Hou, B. Wu and U. Aghayeva for help with generating strains, M. Oren for examining *eat-4* expression in the male tail, J. Dodd for providing LHX antibody, several colleagues cited in the text for communicating unpublished results and members of the Hobert lab, Iva Greenwald and Piali Sengupta for comments on the manuscript. This work was funded by the NIH (R01NS039996-05; R01NS050266-03). O.H. is an Investigator of the Howard Hughes Medical Institute.

Received: May 30, 2013

Revised: August 14, 2013

Accepted: September 23, 2013

Published: October 24, 2013

REFERENCES

- Acampora, D., Gulisano, M., Broccoli, V., and Simeone, A. (2001). Otx genes in brain morphogenesis. *Prog. Neurobiol.* 64, 69–95.
- Alkema, M.J., Hunter-Ensor, M., Ringstad, N., and Horvitz, H.R. (2005). Tyramine Functions independently of octopamine in the *Caenorhabditis elegans* nervous system. *Neuron* 46, 247–260.
- Badea, T.C., Wang, Y., and Nathans, J. (2003). A noninvasive genetic/pharmacologic strategy for visualizing cell morphology and clonal relationships in the mouse. *J. Neurosci.* 23, 2314–2322.
- Baird-Titus, J.M., Clark-Baldwin, K., Dave, V., Caperelli, C.A., Ma, J., and Rance, M. (2006). The solution structure of the native K50 Bicoid homeodomain bound to the consensus TAATCC DNA-binding site. *J. Mol. Biol.* 356, 1137–1151.
- Baran, R., Aronoff, R., and Garriga, G. (1999). The *C. elegans* homeodomain gene *unc-42* regulates chemosensory and glutamate receptor expression. *Development* 126, 2241–2251.
- Bigelow, H., Doitsidou, M., Sarin, S., and Hobert, O. (2009). MAQGene: software to facilitate *C. elegans* mutant genome sequence analysis. *Nat. Methods* 6, 549.
- Boulin, T., Etchberger, J.F., and Hobert, O. (2006). Reporter gene fusions. *WormBook*, 1–23.
- Brandt, J.P., Aziz-Zaman, S., Juozaityte, V., Martinez-Velazquez, L.A., Petersen, J.G., Pocock, R., and Ringstad, N. (2012). A single gene target of an ETS-family transcription factor determines neuronal CO₂-chemosensitivity. *PLoS ONE* 7, e34014.

- Brill, M.S., Ninkovic, J., Winpenny, E., Hodge, R.D., Ozen, I., Yang, R., Lepier, A., Gascón, S., Erdelyi, F., Szabo, G., et al. (2009). Adult generation of glutamatergic olfactory bulb interneurons. *Nat. Neurosci.* *12*, 1524–1533.
- Brockie, P.J., and Maricq, A.V. (2006). Ionotropic glutamate receptors: genetics, behavior and electrophysiology. *WormBook*, 1–16.
- Bürglin, T.R., and Ruvkun, G. (2001). Regulation of ectodermal and excretory function by the *C. elegans* POU homeobox gene *ceh-6*. *Development* *128*, 779–790.
- Cassata, G., Kagoshima, H., Andachi, Y., Kohara, Y., Dürrenberger, M.B., Hall, D.H., and Bürglin, T.R. (2000). The LIM homeobox gene *ceh-14* confers thermosensory function to the AFD neurons in *Caenorhabditis elegans*. *Neuron* *25*, 587–597.
- Chalasan, S.H., Chronis, N., Tsunozaki, M., Gray, J.M., Ramot, D., Goodman, M.B., and Bargmann, C.I. (2007). Dissecting a circuit for olfactory behaviour in *Caenorhabditis elegans*. *Nature* *450*, 63–70.
- Chang, S., Johnston, R.J., Jr., and Hobert, O. (2003). A transcriptional regulatory cascade that controls left/right asymmetry in chemosensory neurons of *C. elegans*. *Genes Dev.* *17*, 2123–2137.
- Chang, H.C., Paek, J., and Kim, D.H. (2011). Natural polymorphisms in *C. elegans* HECW-1 E3 ligase affect pathogen avoidance behaviour. *Nature* *480*, 525–529.
- Cheng, L., Arata, A., Mizuguchi, R., Qian, Y., Karunaratne, A., Gray, P.A., Arata, S., Shirasawa, S., Bouchard, M., Luo, P., et al. (2004). *Tlx3* and *Tlx1* are postmitotic selector genes determining glutamatergic over GABAergic cell fates. *Nat. Neurosci.* *7*, 510–517.
- Coates, J.C., and de Bono, M. (2002). Antagonistic pathways in neurons exposed to body fluid regulate social feeding in *Caenorhabditis elegans*. *Nature* *419*, 925–929.
- Doitsidou, M., Flames, N., Lee, A.C., Boyanov, A., and Hobert, O. (2008). Automated screening for mutants affecting dopaminergic-neuron specification in *C. elegans*. *Nat. Methods* *5*, 869–872.
- Duerr, J.S., Han, H.P., Fields, S.D., and Rand, J.B. (2008). Identification of major classes of cholinergic neurons in the nematode *Caenorhabditis elegans*. *J. Comp. Neurol.* *506*, 398–408.
- Duggan, A., Ma, C., and Chalfie, M. (1998). Regulation of touch receptor differentiation by the *Caenorhabditis elegans* *mec-3* and *unc-86* genes. *Development* *125*, 4107–4119.
- Englund, C., Fink, A., Lau, C., Pham, D., Daza, R.A., Bulfone, A., Kowalczyk, T., and Hevner, R.F. (2005). *Pax6*, *Tbr2*, and *Tbr1* are expressed sequentially by radial glia, intermediate progenitor cells, and postmitotic neurons in developing neocortex. *J. Neurosci.* *25*, 247–251.
- Etchberger, J.F., Lorch, A., Sleumer, M.C., Zapf, R., Jones, S.J., Marra, M.A., Holt, R.A., Moerman, D.G., and Hobert, O. (2007). The molecular signature and cis-regulatory architecture of a *C. elegans* gustatory neuron. *Genes Dev.* *21*, 1653–1674.
- Flames, N., and Hobert, O. (2009). Gene regulatory logic of dopamine neuron differentiation. *Nature* *458*, 885–889.
- Furukawa, T., Morrow, E.M., and Cepko, C.L. (1997). *Crx*, a novel *otx*-like homeobox gene, shows photoreceptor-specific expression and regulates photoreceptor differentiation. *Cell* *91*, 531–541.
- Guillermin, M.L., Castelletto, M.L., and Hallem, E.A. (2011). Differentiation of carbon dioxide-sensing neurons in *Caenorhabditis elegans* requires the ETS-5 transcription factor. *Genetics* *189*, 1327–1339.
- Hobert, O. (2011). Regulation of terminal differentiation programs in the nervous system. *Annu. Rev. Cell Dev. Biol.* *27*, 681–696.
- Hobert, O., D'Alberti, T., Liu, Y., and Ruvkun, G. (1998). Control of neural development and function in a thermoregulatory network by the LIM homeobox gene *lin-11*. *J. Neurosci.* *18*, 2084–2096.
- Jafari, G., Xie, Y., Kullyev, A., Liang, B., and Sze, J.Y. (2011). Regulation of extrasynaptic 5-HT by serotonin reuptake transporter function in 5-HT-absorbing neurons underscores adaptation behavior in *Caenorhabditis elegans*. *J. Neurosci.* *31*, 8948–8957.
- Kage, E., Hayashi, Y., Takeuchi, H., Hirotsu, T., Kunitomo, H., Inoue, T., Arai, H., Iino, Y., and Kubo, T. (2005). MBR-1, a novel helix-turn-helix transcription factor, is required for pruning excessive neurites in *Caenorhabditis elegans*. *Curr. Biol.* *15*, 1554–1559.
- Kagoshima, H., Cassata, G., Tong, Y.G., Pujol, N., Niklaus, G., and Bürglin, T.R. (2013). The LIM homeobox gene *ceh-14* is required for phasmid function and neurite outgrowth. *Dev. Biol.* *380*, 314–323.
- Kim, K., Kim, R., and Sengupta, P. (2010). The HMX/NKX homeodomain protein MLS-2 specifies the identity of the AWC sensory neuron type via regulation of the *ceh-36 Otx* gene in *C. elegans*. *Development* *137*, 963–974.
- Kratsios, P., Stolfi, A., Levine, M., and Hobert, O. (2012). Coordinated regulation of cholinergic motor neuron traits through a conserved terminal selector gene. *Nat. Neurosci.* *15*, 205–214.
- Kwan, K.M., and Behringer, R.R. (2002). Conditional inactivation of *Lim1* function. *Genesis* *32*, 118–120.
- Lanjuin, A., VanHoven, M.K., Bargmann, C.I., Thompson, J.K., and Sengupta, P. (2003). *Otx/otd* homeobox genes specify distinct sensory neuron identities in *C. elegans*. *Dev. Cell* *5*, 621–633.
- Lee, R.Y., Sawin, E.R., Chalfie, M., Horvitz, H.R., and Avery, L. (1999). *EAT-4*, a homolog of a mammalian sodium-dependent inorganic phosphate cotransporter, is necessary for glutamatergic neurotransmission in *Caenorhabditis elegans*. *J. Neurosci.* *19*, 159–167.
- Lee, D., Jung, S., Ryu, J., Ahnn, J., and Ha, I. (2008). Human vesicular glutamate transporters functionally complement *EAT-4* in *C. elegans*. *Mol. Cells* *25*, 50–54.
- 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. (2007). Genome-wide atlas of gene expression in the adult mouse brain. *Nature* *445*, 168–176.
- Lou, S., Duan, B., Vong, L., Lowell, B.B., and Ma, Q. (2013). *Runx1* controls terminal morphology and mechanosensitivity of VGLUT3-expressing C-mechanoreceptors. *J. Neurosci.* *33*, 870–882.
- Ma, Q., and Cheng, L. (2006). Roles of *Tlx1* and *Tlx3* and Neuronal Activity in Controlling Glutamatergic over GABAergic Cell Fates. In *Transcription Factors in the Nervous System*, G. Thiel, ed. (Weinheim: Wiley-VCH).
- Niu, W., Lu, Z.J., Zhong, M., Sarov, M., Murray, J.I., Brdlik, C.M., Janette, J., Chen, C., Alves, P., Preston, E., et al. (2011). Diverse transcription factor binding features revealed by genome-wide ChIP-seq in *C. elegans*. *Genome Res.* *21*, 245–254.
- Nolan, K.M., Sarafi-Reinach, T.R., Horne, J.G., Saffer, A.M., and Sengupta, P. (2002). The DAF-7 TGF-beta signaling pathway regulates chemosensory receptor gene expression in *C. elegans*. *Genes Dev.* *16*, 3061–3073.
- Ohnishi, N., Kuhara, A., Nakamura, F., Okochi, Y., and Mori, I. (2011). Bidirectional regulation of thermotaxis by glutamate transmissions in *Caenorhabditis elegans*. *EMBO J.* *30*, 1376–1388.
- Peckol, E.L., Troemel, E.R., and Bargmann, C.I. (2001). Sensory experience and sensory activity regulate chemosensory receptor gene expression in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* *98*, 11032–11038.
- Pocock, R., and Hobert, O. (2010). Hypoxia activates a latent circuit for processing gustatory information in *C. elegans*. *Nat. Neurosci.* *13*, 610–614.
- Sarafi-Reinach, T.R., Melkman, T., Hobert, O., and Sengupta, P. (2001). The *lin-11* LIM homeobox gene specifies olfactory and chemosensory neuron fates in *C. elegans*. *Development* *128*, 3269–3281.
- Satterlee, J.S., Sasakura, H., Kuhara, A., Berkeley, M., Mori, I., and Sengupta, P. (2001). Specification of thermosensory neuron fate in *C. elegans* requires *ttx-1*, a homolog of *otd/Otx*. *Neuron* *31*, 943–956.
- Sawin, E.R., Ranganathan, R., and Horvitz, H.R. (2000). *C. elegans* locomotory rate is modulated by the environment through a dopaminergic pathway and by experience through a serotonergic pathway. *Neuron* *26*, 619–631.
- Seal, R.P., and Edwards, R.H. (2006). Functional implications of neurotransmitter co-release: glutamate and GABA share the load. *Curr. Opin. Pharmacol.* *6*, 114–119.

- Takamori, S., Rhee, J.S., Rosenmund, C., and Jahn, R. (2000). Identification of a vesicular glutamate transporter that defines a glutamatergic phenotype in neurons. *Nature* *407*, 189–194.
- Takamori, S., Rhee, J.S., Rosenmund, C., and Jahn, R. (2001). Identification of differentiation-associated brain-specific phosphate transporter as a second vesicular glutamate transporter (VGLUT2). *J. Neurosci.* *21*, RC182.
- Tursun, B., Cochella, L., Carrera, I., and Hobert, O. (2009). A toolkit and robust pipeline for the generation of fosmid-based reporter genes in *C. elegans*. *PLoS ONE* *4*, e4625.
- White, J.G., Southgate, E., Thomson, J.N., and Brenner, S. (1986). The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* *314*, 1–340.
- Wingender, E., Dietze, P., Karas, H., and Knüppel, R. (1996). TRANSFAC: a database on transcription factors and their DNA binding sites. *Nucleic Acids Res.* *24*, 238–241.
- Zhao, Y., Kwan, K.M., Mailloux, C.M., Lee, W.K., Grinberg, A., Wurst, W., Behringer, R.R., and Westphal, H. (2007). LIM-homeodomain proteins Lhx1 and Lhx5, and their cofactor Ldb1, control Purkinje cell differentiation in the developing cerebellum. *Proc. Natl. Acad. Sci. USA* *104*, 13182–13186.

SUPPLEMENTAL RESULTS

Markers of Glutamatergic Identity

The enzyme glutaminase generates 70% of all neuronal glutamate in vertebrates (Hertz, 2004). However, the expression of glutaminase is not restricted to glutamatergic neurons within the vertebrate nervous system (e.g., Kaneko et al., 1990; Kaneko and Mizuno, 1992). Glutaminase also does not serve as a selective marker for *C. elegans* glutamatergic neurons. Sequence homolog searches identify three glutaminase-encoding genes in the *C. elegans* genome, *glna-1*, *glna-2* and *glna-3*. *glna-1* and *glna-2* reporter genes are expressed outside the nervous system and in very small (<5) number of neurons, while a *glna-3* reporter is expressed in many neuron types (Figure S1). However, colabeling transgenic *glna-3::gfp* worms with markers that label other neurotransmitter populations show that *glna-3* is not restricted to glutamatergic neurons (Figure S1). The nonselectivity of glutaminase expression for glutamatergic neurons mirrors the situation in the vertebrate nervous system and argues that like in vertebrates, the VGLUT gene *eat-4* is the only currently available identity marker of glutamatergic neurons.

Glutamate plasma membrane reuptake transporters are also not good markers of glutamatergic neurons since, unlike the reuptake transporters for other neurotransmitters, glutamate reuptake transporters are primarily expressed in cells surrounding glutamatergic neurons, rather than in the glutamatergic neurons themselves (Huang and Bergles, 2004; Mano et al., 2007).

Examination of the Antagonism of Glutamatergic and GABAergic Identity

In several distinct regions of the vertebrates CNS, excitatory glutamatergic and inhibitory GABAergic identity of neuron types can be controlled in an antagonistic, binary switch-type manner, such that loss of glutamatergic identity is accompanied by gain of GABAergic fate (for example in the dorsal spinal horn of Tlx mutant mice; Ma and Cheng, 2006). A similar excitatory/inhibitory antagonism is also observed in the cholinergic system of the basal ganglia, where loss of cholinergic identity of a neuron type can result in a switch to a GABAergic identity (Fragkouli et al., 2009). We examined whether such antagonism is built into the glutamatergic differentiation programs of *C. elegans* neurons as well. Aside from the 38 classes of glutamatergic neurons described above, the *C. elegans* nervous system contains 26 GABAergic neurons (defining six different classes) (McIntire et al., 1993). We examined potential switches of glutamatergic to GABAergic identity in eight mutant strains that lack distinct regulators of 18 distinct glutamatergic neuron classes (*ceh-36*, *ttx-1*, *ceh-37*, *vab-3*, *unc-42*, *ets-5*, *che-1*, *ceh-14*) using anti-GABA immunostaining. We observed no ectopic GABA staining (data not shown). Considering the number of cells and regulators we examined, we can conclude that GABAergic and glutamatergic identity are not commonly executed in an antagonistic, binary switch-type manner in *C. elegans*. This is consistent with previous studies of terminal selector transcription factors, whose disruption does not usually result in cell identity switches, but mere failures to differentiate into any specific state (Hobert, 2011).

EXTENDED EXPERIMENTAL PROCEDURES

eat-4 Reporter Genes

The *eat-4* fosmid reporter was generated by fosmid recombineering using fosmid WRM0623aF12 and an SL2-based, nuclear localized *yfp* reporter (Tursun et al., 2009). The reporter was injected at 15 ng/uL into *pha-1(e2123)* mutant animals with *pBX* as injection marker (2 ng/uL) [100 ng/uL OP50 gDNA as filling DNA]. One extrachromosomal array was integrated to yield *otIs388* III.

The *eat-4* locus reporter (shown in Figure 2) was generated by in vivo recombination (Boulin et al., 2006), using two overlapping fragments of the *eat-4* locus. Briefly, a 5'-*eat-4*prom-*mChOpt*^[1-516] fragment was generated by PCR fusing an upstream *eat-4* region from position -5.6 kb to the *eat-4* start codon (using the primer 5'-GGATTGAAGTAGCTCACTGATGGATCG -3') to the first two-thirds of the codon-optimized *mCherry*. A second fragment, *mChOpt*^[116-861Δstop codon]-*EAT-4::eat-4* 3'-UTR, was generated PCR fusing a C terminus fragment (+116-861 bp) of codon-optimized *mCherry* to the 5' end of the *eat-4* locus plus 560 bp of *eat-4* 3'UTR (using the primer 5'-GAACATCCTTGATTTCTTCTTGCTCA -3'). Both fragments contained a 400 bp overlap within the *mCherry* sequence, so only upon successful in vivo recombination an intact fluorescent protein can be generated. Both fragments were injected at equal molar ratios with *rol-6* (*su1006*) (50 ng/uL) as a coinjection marker. One extrachromosomal array was integrated to yield *otIs292*.

We identified neurons that express *eat-4* reporter constructs based on their cell position and, in many cases, by colabeling individual neurons with *mCherry*-based reporters.

The *eat-4* reporter transgenes are as follows:

otIs388 III: *eat-4*^{FO5}::*sl2::yfp::H2B*; injected in *pha-1(e2123)*; *pBX*

otIs292: *eat-4* in vivo recombineered reporter; *rol-6*

adIs1230: *eat-4* reporter construct from Lee et al., 1999.

otIs376: *eat-4*^{prom2}::*gfp*; *rol-6*

otIs392: *eat-4*^{prom6}::*gfp*; *ttx-3::dsRed*

otEx4478, *otEx4479*, *otEx4480*: 3 lines for *eat-4*^{prom1}::*gfp*; *rol-6*

otEx4492, *otEx4432*, *otEx4494*: 3 lines for *eat-4*^{prom12}::*gfp*; *rol-6*

otEx5292, *otEx5293*, *otEx5294*: 3 lines for *eat-4*^{prom5}::*gfp*; *ttx-3::dsRed*

otEx5298, *otEx5299*, *otEx5300*: 3 lines for *eat-4*^{prom6Δ3}::*gfp*; *ttx-3::dsRed*

otEx5311, otEx5312: 2 lines for *eat-4^{prom6Δ4}::gfp; rol-6*
otEx5295, otEx5296, otEx5297: 3 lines for *eat-4^{prom7}::gfp; ttx-3::dsRed*
otEx4488, otEx4489, otEx4490: 3 lines for *eat-4^{prom8}::tagRFP; rol-6*
otEx5301, otEx5310: 2 lines for *eat-4^{prom10}::gfp; rol-6*
otEx5098, otEx5099, otEx5100: 3 lines for *eat-4^{prom2Δ7}::gfp; rol-6*
otEx5313, otEx5314, otEx5315: 3 lines for *eat-4^{prom16}; ttx-3::dsRed*
otEx5316, otEx5316: 2 lines for *eat-4^{prom2Δ11}; rol-6*
otEx5318, otEx5319, otEx5320: 3 lines for *eat-4^{prom2Δ12}; rol-6*
otEx5345, otEx5346, otEx5347: 3 lines for *eat-4^{prom2Δ6}; rol-6*
otEx5330, otEx5331, otEx5332: 3 lines for *eat-4^{prom10Δ6}; rol-6*
otEx5333, otEx5334, otEx5335: 3 lines *eat-4^{prom10Δ7}; rol-6*

Serotonin and GABA Antibody Staining

Antibody staining was performed using a tube fixation protocol (adapted from McIntire et al., 1992). Briefly, young adult worms well fed were fixed with paraformaldehyde (PFA) 4% for 24 hr at 4°C; for GABA staining animals were fixed with PHA 4% - Glutaraldehyde 1%. The next day they were washed with 1% PBS – 0.5% Triton X-100 three times and incubated for 18 hr at 37°C in a nutator mixer with 5% β-mercapto-ethanol-1% Triton X-100 - 0.1 M Tris (pH 7.5). The third day the worms were rinsed three or four times with 1% PBS – 0.5% Triton X-100 and treated with collagenase type IV (Sigma Aldrich, C-5138) in collagenase buffer (1% Triton X-100/0.1 M Tris, pH 7.5/1 mM CaCl₂) for 1 hr at 37°C/700 rpm. Worms were washed with 1% PBS – 0.5% Triton X-100 and proceeded to stain. Blocking solution (PBS 1X – 0.2% Gelatin - 0.25% Triton X-100) was added to the worms for 30 min at room temperature and then they were incubated for 24 hr at 4°C in primary antibody [anti-5HT antibody 1/100 (Sigma Aldrich, S-5545); anti-GABA antibody 1/500 (AbCam, ab17413)] in PBS 1X - 0.1% Gelatin - 0.25% Triton X-100. The worms were washed three times and incubated with secondary antibody [anti-rabbit Alexa Fluor 1/1,000 (BD Biosciences) for anti-5HT and anti-guinea pig Alexa Fluoro 1/100 (BD Biosciences) for anti-GABA] for 2 hr at room temperature. Finally worms were washed three times and mounted on Fluoro-Gel II with DAPI (EMS).

Isolation, Identification, and Characterization of *vab-3*

We utilized a transgenic strain that expresses a reporter gene for the tyramine receptor *ser-2* (*otIs138* transgenic animals), which is exclusively expressed in the OLL and PVD sensory neuron classes (Tsalik et al., 2003). After EMS mutagenesis, we identified with a Copas Biosort machine (Doitsidou et al., 2008) three nonallelic mutant strains in which expression of *ser-2::gfp* is lost specifically in the OLL neurons. The mutation *ot569* completely abolishes *ser-2::gfp* expression in the OLL but not the PVD neurons (Figure 5A). Whole-genome sequencing was used to determine the molecular identity of *ot569*. DNA from *ot569* mutants was sequenced to an average depth of 8x in an Illumina GA2 sequencer as previously described (Doitsidou et al., 2010). Analysis of the WGS data with MAQGene (Bigelow et al., 2009) resulted in the identification of 280 variants on the X chromosome, to which the mutant had been previously mapped genetically. A list of background variants was compiled by combining variants present in the whole-genome sequencing data of two other mutants isolated in the same screen and subtracted from the *ot569* data set. Of the resulting 85 variants only 9 are predicted to be splice site or protein coding variants and only one of these affected a transcription factor, namely the *vab-3* locus, which codes for the *C. elegans* ortholog of the vertebrate Pax6/*Drosophila* Eyeless gene (Chisholm and Horvitz, 1995). The mutation converts a highly conserved glycine residue in the linker region of the paired domain to a serine. The linker region makes extensive contacts with the minor groove of DNA suggesting this mutation affects the ability of VAB-3 to bind DNA (Cohen and Melton, 2011). A canonical allele of *vab-3*, *e648* showed a similar OLL differentiation defect as *ot569* and the *ot569* allele can be rescued with a fosmid that contains the *vab-3* locus (Figures 5A and 5B).

The *vab-3* fosmid reporter was generated by fosmid recombineering based in fosmid WRM0623aF12 (Tursun et al., 2009) using a *gfp* reporter fused to the N terminus of the protein and replacing the stop codon. It was injected at a concentration of 15 ng/μl, in combination with 2 ng/μl of *elt-2::DsRed* and 100 ng/μl OP50 gDNA (as filling DNA) to generate 3 extrachromosomal arrays (*otEx5057, otEx5058, otEx5059*) none of which rescue the *vab-3*(*ot569*) phenotype. An untagged fosmid array does rescue the mutant phenotype (Figure 5A).

Neuronal Identity Markers

The following neuronal identity markers were used: *otIs92 [flp-10::gfp]*, *inIs179 [ida-1::gfp]*, *ynIs30 [flp-4::gfp]*, *gmls12 [srb-6::gfp]*, *otIs33 [kal-1::gfp]*, *otIs358 [ser-2^{prom2}::gfp]*, *gmls21 [nlp-1::gfp]*, *ynIs80 [flp-21::gfp]*, *ynIs2022 [flp-8::gfp]*, *ynIs40 [flp-11::gfp]*, *vsIs28 [dop-1::gfp]*, *sEx12012 [srab-12::gfp]*, *zdIs13 [tph-1::gfp]*, *otEx5336 [gpa-11^{prom2}::gfp]*, *pkIs589 [gpa-13::gfp]*, *pkIs591 [gpa-15::gfp]*, *otEx5323 [dkf-2b::gfp]*, *ynIs54 [flp-20::gfp]*, *otIs138* and *otEx449 [ser-2^{prom3}::gfp]*, *otIs396 [ace-1^{prom2}::tagRFP]*, *sEx15238 [grd-8::gfp]*, *myEx741 [pdf-1::NLS::rfp]*, *vdEx078 [tol-1::gfp]*, *otEx2540 [gcy-23::gfp]*, *ynIs37 [flp-13::gfp]*, *rtEx247 [nlp-14::gfp]*, *otEx5428 [glna-1::NLS::gfp]*, *sEx10131 [glna-2::gfp]*, *otEx5429 [glna-3^{prom1}::NLS::gfp]*.

The *ace-1* reporter was generated by PCR fusion of 697 bp promoter sequence to NLS::tagRFP. This construct was injected in *pha-1*(*e2123*) and in combination with wild-type *pha-1* (pBX plasmid). One extrachromosomal array line spontaneously integrated to generate *otIs396 [ace-1^{prom2}::NLS::tagRFP; pha-1(+)]*.

The *gpa-11* reporter was generated by amplifying 1,603 bp of the promoter cloned into pPD95.77. The PCR was injected in N2 at 5 ng/ μ l with *rol-6* (*su1006*) and 100 ng/ μ l OP50 gDNA as filling DNA.

The *glna-1* reporter was generated by PCR fusion of 2,103 bp promoter sequence to *NLS::gfp*. The PCR was injected in N2 at 15 ng/ μ l with *rol-6* (*su1006*) and 100 ng/ μ l OP50 gDNA as filling DNA.

The *glna-3^{brom1}* reporter was generated by PCR fusion of 1,300 bp of the promoter sequence, the first exon of the gene and the first intron to *NLS::gfp*. The PCR was injected in N2 at 15 ng/ μ l with *rol-6* (*su1006*) and 100 ng/ μ l OP50 gDNA as filling DNA.

Staining of Mouse Brain Sections

Primary antibodies used in this study are anti-BRN3a (1:50, mouse monoclonal; Santa Cruz Biotechnology), anti-VGLUT2 (1:100, guinea pig polyclonal) (Brumovsky et al., 2007), anti-glutaminase (1:600, rabbit polyclonal) (Kaneko and Mizuno, 1992) and LHX1 (1:20,000, kindly provided by Jane Dodd). Secondary antibodies were donkey antisera coupled with Alexa dyes (Invitrogen). Sections were counter-stained with DAPI (1:1,000) (Invitrogen).

Animals were perfused intracardially with 4% paraformaldehyde. Brains were cryoprotected in Optimal Cutting Temperature compound (Tissue-Tek) and sectioned in 12 μ m sections.

In situ hybridization was performed as previously described (Wallén-Mackenzie et al., 2006). The VGLUT2 probe was kindly provided by Dr. Kullander, Uppsala University, Sweden (Wallén-Mackenzie et al., 2006). Neurons in the inferior olive were visualized using the Nissl staining with cresyl violet (Sigma).

SUPPLEMENTAL REFERENCES

Brumovsky, P., Watanabe, M., and Hökfelt, T. (2007). Expression of the vesicular glutamate transporters-1 and -2 in adult mouse dorsal root ganglia and spinal cord and their regulation by nerve injury. *Neuroscience* 147, 469–490.

Chalfie, M., Horvitz, H.R., and Sulston, J.E. (1981). Mutations that lead to reiterations in the cell lineages of *C. elegans*. *Cell* 24, 59–69.

Chisholm, A.D., and Horvitz, H.R. (1995). Patterning of the *Caenorhabditis elegans* head region by the Pax-6 family member *vab-3*. *Nature* 377, 52–55.

Cohen, D.E., and Melton, D. (2011). Turning straw into gold: directing cell fate for regenerative medicine. *Nat. Rev. Genet.* 12, 243–252.

Doitsidou, M., Poole, R.J., Sarin, S., Bigelow, H., and Hobert, O. (2010). *C. elegans* mutant identification with a one-step whole-genome-sequencing and SNP mapping strategy. *PLoS ONE* 5, e15435.

Finney, M., and Ruvkun, G. (1990). The *unc-86* gene product couples cell lineage and cell identity in *C. elegans*. *Cell* 63, 895–905.

Fragkouli, A., van Wijk, N.V., Lopes, R., Kessar, N., and Pachnis, V. (2009). LIM homeodomain transcription factor-dependent specification of bipotential MGE progenitors into cholinergic and GABAergic striatal interneurons. *Development* 136, 3841–3851.

Hertz, L. (2004). Intercellular metabolic compartmentation in the brain: past, present and future. *Neurochem. Int.* 45, 285–296.

Huang, Y.H., and Bergles, D.E. (2004). Glutamate transporters bring competition to the synapse. *Curr. Opin. Neurobiol.* 14, 346–352.

Kaneko, T., and Mizuno, N. (1992). Mosaic distribution of phosphate-activated glutaminase-like immunoreactivity in the rat striatum. *Neuroscience* 49, 329–345.

Kaneko, T., Akiyama, H., Nagatsu, I., and Mizuno, N. (1990). Immunohistochemical demonstration of glutaminase in catecholaminergic and serotonergic neurons of rat brain. *Brain Res.* 507, 151–154.

Mano, I., Straud, S., and Driscoll, M. (2007). *Caenorhabditis elegans* glutamate transporters influence synaptic function and behavior at sites distant from the synapse. *J. Biol. Chem.* 282, 34412–34419.

McIntire, S.L., Garriga, G., White, J., Jacobson, D., and Horvitz, H.R. (1992). Genes necessary for directed axonal elongation or fasciculation in *C. elegans*. *Neuron* 8, 307–322.

McIntire, S.L., Jorgensen, E., Kaplan, J., and Horvitz, H.R. (1993). The GABAergic nervous system of *Caenorhabditis elegans*. *Nature* 364, 337–341.

Tsalik, E.L., Niacaris, T., Wenick, A.S., Pau, K., Avery, L., and Hobert, O. (2003). LIM homeobox gene-dependent expression of biogenic amine receptors in restricted regions of the *C. elegans* nervous system. *Dev. Biol.* 263, 81–102.

Wallén-Mackenzie, A., Gezelius, H., Thoby-Brisson, M., Nygård, A., Enjin, A., Fujiyama, F., Fortin, G., and Kullander, K. (2006). Vesicular glutamate transporter 2 is required for central respiratory rhythm generation but not for locomotor central pattern generation. *J. Neurosci.* 26, 12294–12307.

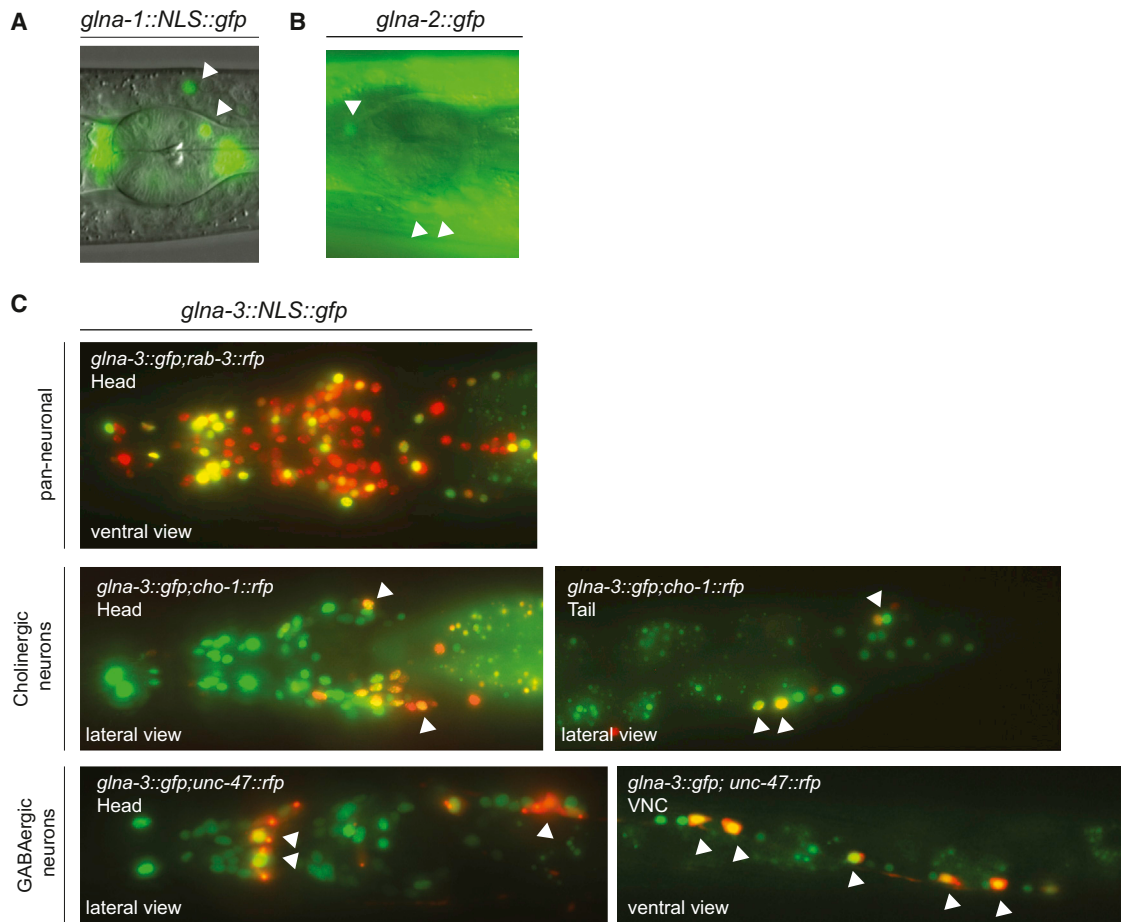


Figure S1. Expression Pattern of Glutaminase Genes, Related to Figure 1 and 2

We examined whether glutaminase could perhaps serve as a marker for *C. elegans* glutamatergic neurons. Sequence homolog searches identified three glutaminase-encoding genes in the *C. elegans* genome, *glna-1*, *glna-2* and *glna-3*.

(A) Expression pattern of a *glna-1* reporter that contains 2.1 kb sequences upstream of the first exon of the gene. Expression is observed in a very small number of neurons in the head (white arrowheads).

(B) Expression pattern of a *glna-2* reporter that contains 3 kb sequences upstream of the first exon of the gene. Expression is observed in a very small number of neurons in the head (white arrowheads).

(C) Expression pattern of a *glna-3* reporter that contains 1.3 kb sequences upstream of the first exon of the gene and the first intron. Expression of the *gfp* reporter is seen in many neuron types in the head (as seen in the yellow overlap of the green *gfp* reporter and a red fluorescent marker that labels all neurons). Middle and lower: the overlap of the *gfp* reporter with cholinergic neurons (*cho-1::rfp*) and GABAergic neurons (*unc-47::rfp*) is shown. Some examples of clear overlaps in expression are indicated with white arrowheads. The nonselectivity of glutaminase expression for glutamatergic neurons mirrors the situation in the vertebrate nervous system and argues that like in vertebrates, the *VGLUT* gene *eat-4* is the only currently available identity marker of glutamatergic neurons.

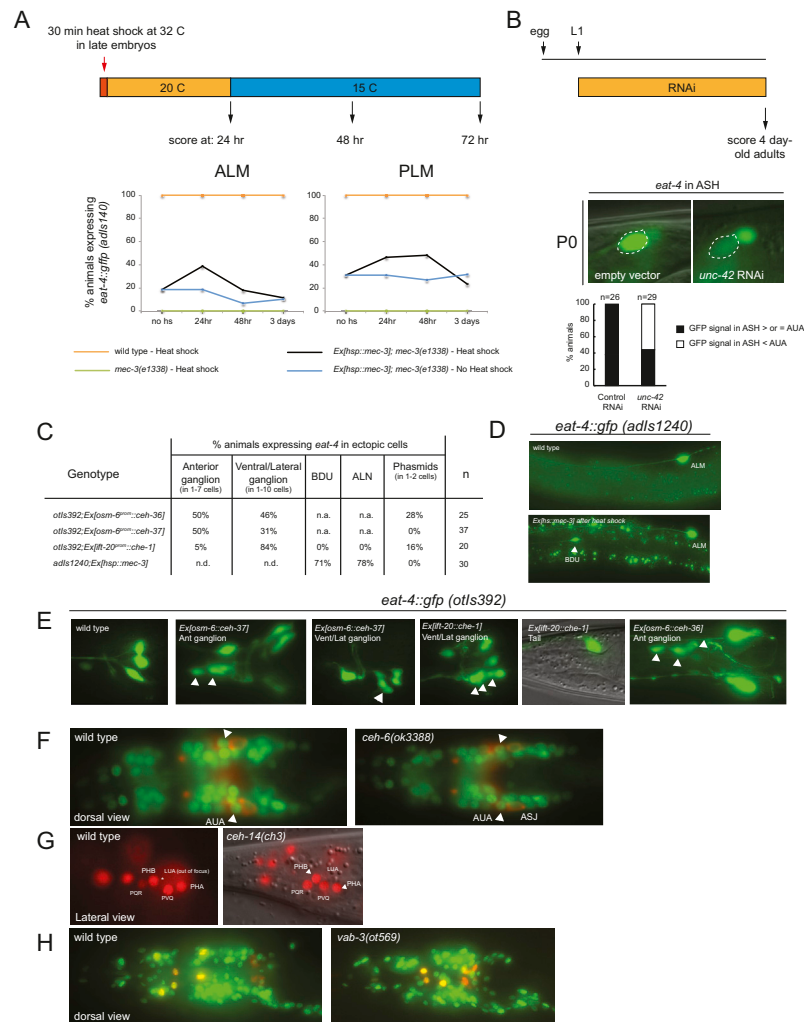


Figure S2. The Role of Terminal Selectors in Maintaining and Ectopically Inducing Neuronal Identity and Their Effect on Panneuronal Identity Features, Related to Figure 3, 4, and 5

(A and B) *mec-3* and *unc-42* are required to maintain the expression of *eat-4/VGLUT*. (A) Temporally controlled *mec-3* expression was achieved through driving *mec-3* expression under the control of the heat shock promoter in a *mec-3(e1338)* mutant background. A brief pulse of *mec-3* expression, achieved through 30 min of heat shock, is able to partially restore the expression of *eat-4* (assayed with the *adIs1240* transgene) in the touch receptor neurons 24 hr after heat shock (comparisons between heat-shocked and non-heat-shocked animals performed using student's t test, in both cases p values < 0.05). Three days after the transient, 30 min-pulse of *mec-3* expression (when *mec-3* expression has presumably faded away), *eat-4* expression is reduced compared with the expression at 24 hr after heat shock (comparisons between values 24 hr after heat shock and 3 days after heat shock were performed using student's t test, both * p values are < 0.05). (B) Postembryonic removal of *unc-42* results in a decrease of *eat-4* expression (assayed with the *otIs376* array) in the ASH neurons. RNAi-sensitized *nre-1 lin-15; otIs376* animals were fed with *unc-42(dsRNA)* from the L1 stage onward and assayed 4 days later. As an internal control, *gfp* expression in ASH was compared to *gfp* expression in the AUA neurons. RNAi against *unc-42* showed a reduction in the expression of *eat-4* in ASH compared to AUA fluorescence. In animals fed with control (empty vector) RNAi the GFP intensity in ASH is always higher or similar to that of AUA.

(C–E) Ectopic expression of *eat-4/VGLUT* regulators induces ectopic *eat-4/VGLUT* expression. *ceh-36*, *ceh-37* and *che-1* were misexpressed with the pansensory promoters *osm-6* and *ift-20*, and *mec-3* was misexpressed using the heat-shock promoter. Heat shock was induced at the L1 stage for 30 min and animals were scored 48 hr later at the L4 stage. (C) Quantification of effects of misexpression of glutamatergic regulators. We ascribe the cellular context-dependency of ectopic *eat-4* expression to the limited availability of cofactors with which these factors act in their normal cellular context. n.a.: not applicable because promoter is not expressed in these cells. n.d.: not determined. (D and E) Representative examples of the effects of misexpression of glutamatergic regulators.

(F–H) Expression of the panneuronal marker *rab-3* is unaffected in terminal selector mutants. (F) *rab-3* expression (monitored with *otIs356 = rab-3::NLS::tagRFP*) is unaffected in phasmid sensory neurons in 3/3 *ceh-14* null mutant animals. (G) *rab-3* expression (monitored with *otIs291 = rab-3::NLS::yfp*) is unaffected in the AUA neurons of 6/6 *ceh-6* null mutant animals. Identification of AUA was facilitated by Dil staining (red), which labels a closely neighboring neuron. (H) *rab-3* expression (monitored with *otIs291(rab-3::NLS::yfp); otIs396(ace-1^{prom2}::NLS::tagRFP)*) is unaffected. Since the position of neuronal cell bodies is somewhat variable in *vab-3* mutants, it is difficult to unambiguously identify OLL in these animals. We therefore counted overall *rab-3::yfp(+)* neuron number in the anterior ganglion. Adult wild-type animals (n = 22) have an average of 37.8 neurons in the anterior ganglion and *vab-3* mutants have an average of 36.9 (n = 51), which is not statistically significantly different.

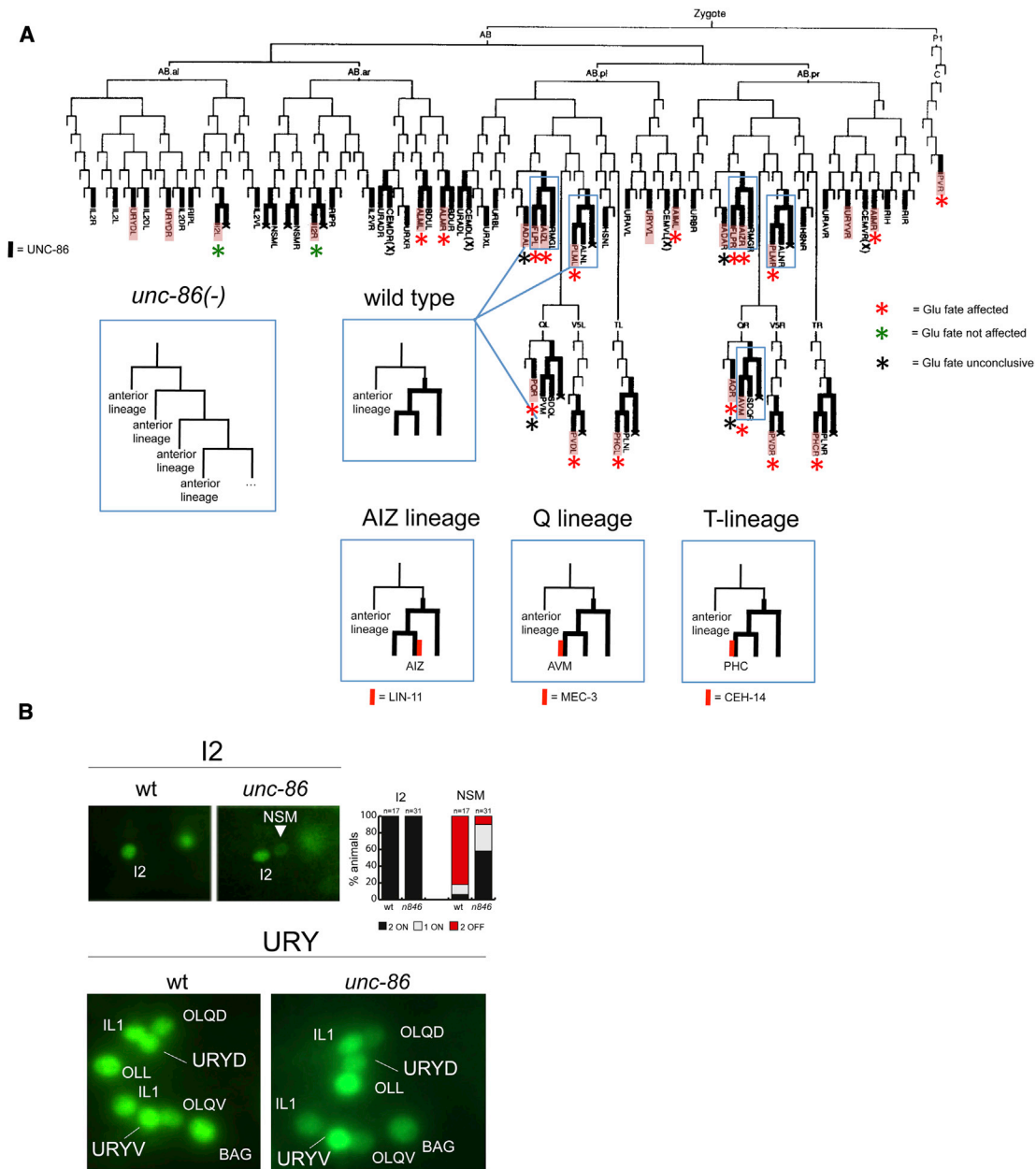


Figure S3. *unc-86* and Neuroblast Identity of Glutamatergic Lineages, Related to Figure 3

(A) Expression of *unc-86* (diagram taken from Finney and Ruvkun, 1990) and its overlap with *eat-4* expression. While *unc-86* affects the terminal differentiation programs of specific neuron types (e.g., ALM and PLM), in a small number of cases, it is expressed and acts earlier in the lineage to affect neuroblast identity (Chalfie et al., 1981; Finney and Ruvkun, 1990). Considering the expression pattern analysis of UNC-86, the common theme emerges that whenever UNC-86 is expressed through 3 cell generations, its loss results in a reiteration of the fate of the mother cell. To corroborate the loss of glutamatergic neuron identity in these lineage-defective mutants with a molecular marker, we examined *eat-4/VGLUT* expression in these lineages in *unc-86* mutants. We indeed observed a loss of *eat-4/VGLUT* expression in the Q neuroblast-derived AVM touch receptors neuron, in the T lineage-derived PHC tail sensory neurons and in the AIZ interneurons (Figure 3). In the Q neuroblast-derived AVM and PVM neurons, *unc-86* is known to not only affect neuroblast identity, but to also act later during terminal differentiation (Duggan et al., 1998). It is possible that in the PHC and AIZ lineages *unc-86* may also have late roles in controlling terminal differentiation via regulating *eat-4* expression in addition to defining neuroblast identity. Notable, *unc-86* appears to cooperate with distinct LIM homeobox genes in distinct neuron types. In the case of the Q lineage that produces the AVM/PVM neuron, *unc-86* has been found to cooperate with the LIM homeobox gene *mec-3* to control the terminal identity state of the neuron. In the lineage that produces PHC, a similar cooperation with *ceh-14* may occur. In the lineage that produces AIZ, such a cooperation with the resident LIM homeobox gene *lin-11* appears, however, less likely since AIZ identity is not affected in *lin-11* mutants (data not shown) (Tsai et al., 2003).

(B) *unc-86* does not affect expression of *eat-4* in I2 or URY. The effect of *unc-86* on *eat-4* expression in AQR, PQR and ADA was inconclusive.

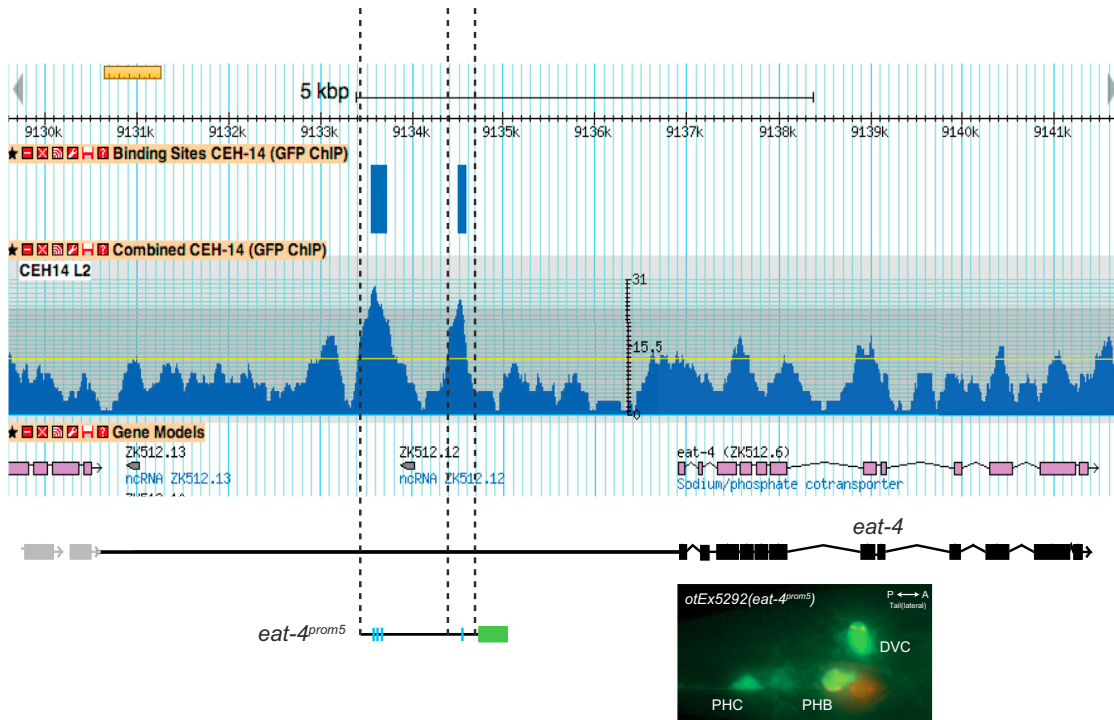


Figure S4. ModEncode Data Reveals CEH-14 Binding Sites in the *eat-4* Locus

Related to Figure 2. Wormbase genome browser representation of *ceh-14* binding sites (dark blue boxes in genome browser image) and their specific location in *eat-4^{prom5}*. The picture shows a representative projection of the expression of *eat-4^{prom5}* in the tail in PHC, PHB and DVC. The expression of *eat-4 (otIs388)* was abolished in these neurons in a *ceh-14* null mutant. The light blue bars in the promoter represent TAAT homeodomain binding sequences corresponding to the region within the blue boxes.

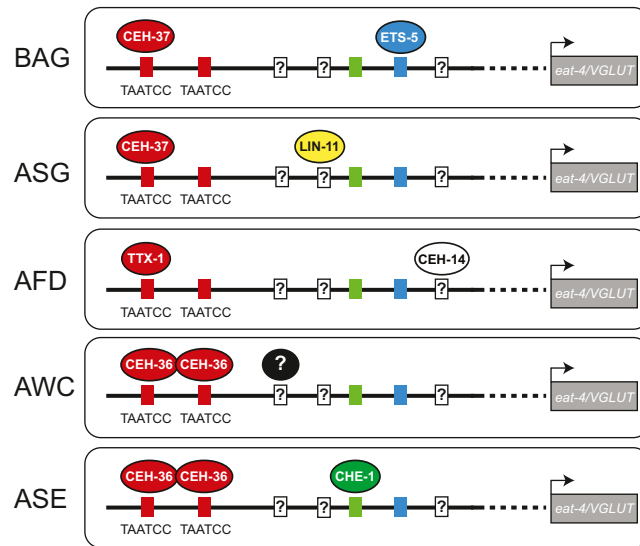


Figure S5. Summary of the Effect of Regulation of *eat-4* Expression in Distinct Neuron Types by Distinct Transcription Factor Combinations, Each Involving an Otx-Type Homeodomain Transcription Factor

Related to Figure 2 and 3. Figure 2 shows the data for the involvement of the TAATCC and the ETS domain binding sites. We hypothesize that other, as yet unidentified TFs act in parallel. The cofactor of *ceh-36* in the AWC neurons is likely *sox-2* (B. Vidal-Iglesias, personal communication).

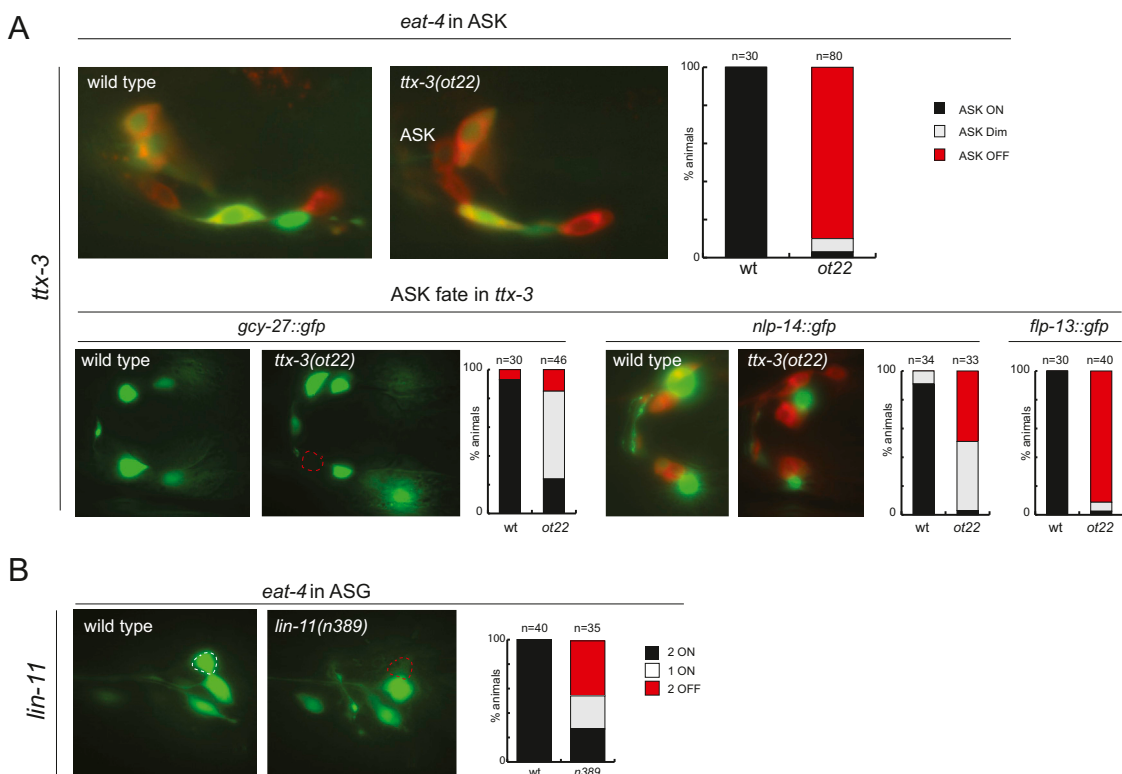


Figure S6. The LIM Homeobox Genes *lin-11/LHX1/5* and *ttx-3/LHX2/9* Control the Identity of the ADL and ASK Sensory Neurons, Related to Figure 2 and 3

(A) *ttx-3* affects expression of *eat-4* (assayed with the *otIs376* transgene) and several other terminal features (*gcy-27::gfp* transgene *otEx2540*, *nlp-14::gfp* transgene *rtEx247* and *flp-13::gfp* transgene *ynIs37*) of the ASK neurons.

(B) *lin-11* affects *eat-4* expression (assayed with the *otIs392* transgene) in the ASK neurons.

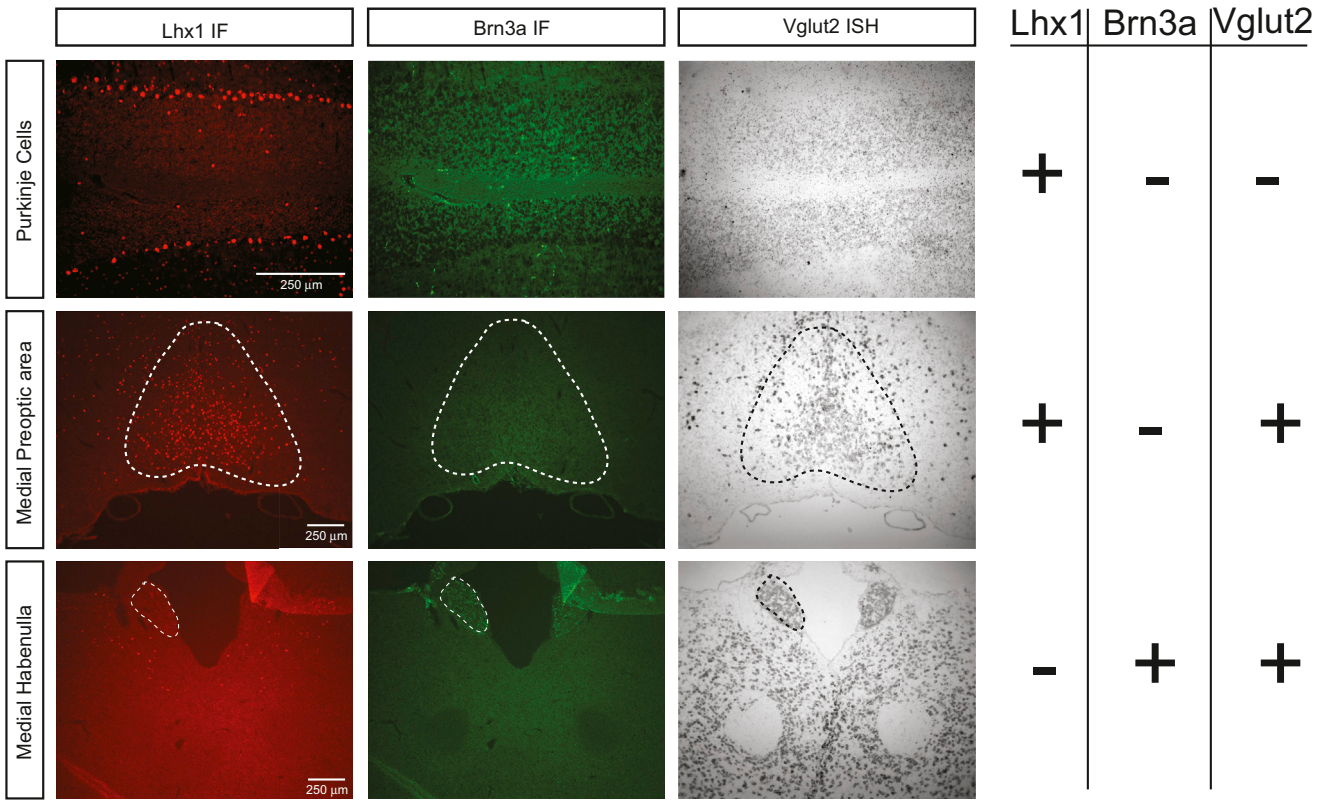


Figure S7. LHX1 Is Not Expressed in All Glutamatergic Neurons and Not All Glutamatergic Neurons Are LHX1-Positive, Related to Figure 6 Immunostaining analysis for LHX1 and BRN3A and VGLUT2 in situ hybridization in sequential coronal sections of different regions of the adult mouse brain.

APPENDIX 2:

Table 1: List of 2ASEL mutants retrieved from *lsy-6(ot71); otIs4(gcy-7::gfp); vsIs33(dop-3::dsRed)* using the COPAS worm sorter.

strain name	allele	genetic trait	<i>cog-1</i> allele	<i>fozi-1</i> allele	Notes
OH8114	<i>ot424</i>	recessive		+	
OH8115	<i>ot425</i>	recessive	+		
OH8149	<i>ot426</i>	recessive, not on X	+		
OH8150	<i>ot427</i>	recessive, not on X		+	
OH8151	<i>ot428</i>	recessive, not on X		+	
OH8152	<i>ot429</i>	recessive, not on X	+		
OH8153	<i>ot430</i>	recessive, not on X	+		
OH8180	<i>ot440</i>	recessive, not on X		+	
OH8181	<i>ot441</i>	recessive, not on X	+		
OH8182	<i>ot442</i>	recessive, not on X		+	
OH8226	<i>ot444</i>	recessive, not on X		+	
OH8227	<i>ot445</i>	recessive, not on X		+	
OH8228	<i>ot446</i>	recessive, not on X		+	
OH8229	<i>ot447</i>	recessive, not on X		+	
OH8230	<i>ot448</i>	recessive, not on X		+	
OH8231	<i>ot449</i>	recessive, not on X		+	
OH8232	<i>ot450</i>	recessive, not on X		+	
OH8407	<i>ot459</i>				sick, lost
OH8324	<i>ot451</i>	recessive, not on X		+	
OH8325	<i>ot452</i>	recessive, not on X	+		
OH8326	<i>ot453</i>	recessive, not on X		+	
OH8361	<i>ot457</i>			+	
OH8330	<i>ot454</i>	recessive, not on X		+	
OH8331	<i>ot455</i>	recessive, not on X		+	
OH8332	<i>ot456</i>	recessive, not on X		+	
OH8377	<i>ot458</i>	recessive, not on X	+		
OH8410	<i>ot460</i>	recessive, not on X	+		
OH8411	<i>ot461</i>	recessive, not on X		+	
OH8412	<i>ot462</i>	recessive, not on X		+	
OH8413	<i>ot463</i>	recessive, not on X		+	
OH8437	<i>ot464</i>	recessive, not on X	+		
OH8438	<i>ot465</i>	recessive, not on X		+	
OH8468	<i>ot466</i>	recessive, not on X	+		
OH8469	<i>ot467</i>	recessive, not on X		+	
OH8485	<i>ot468</i>	recessive, not on X	+		
OH8486	<i>ot469</i>	recessive, not on X		+	
OH8516	<i>ot470</i>	recessive, not on X		+	
OH8517	<i>ot471</i>	recessive, not on X	+		
OH8518	<i>ot472</i>	recessive, not on X		+	
OH8519	<i>ot473</i>	recessive, not on X		+	
OH8536	<i>ot474</i>	recessive, not on X		+	

OH8537	<i>ot475</i>		+		
OH8573	<i>ot483</i>	recessive, not on X		+	
OH8574	<i>ot484</i>	recessive, not on X			
OH8575	<i>ot485</i>	recessive, not on X		+	
OH8576	<i>ot486</i>	recessive, not on X		+	
OH8577	<i>ot487</i>	recessive, not on X		+	
OH8578	<i>ot488</i>	recessive, not on X		+	
OH8579	<i>ot489</i>	Semi-dominant			<i>che-1</i> allele

Notes:

1. Mutant animals were first mated with N2 males to assess whether a mutation is recessive or on X chromosome. Complementation tests were then performed by mating *fozi-1/+ (ot131)* or *cog-1/+ (ot28)* males to mutant hermaphrodites. F1 male progeny were scored. Duplicate tests were made for each mutant.
2. “+” means that a mutant is an allele of either *fozi-1* or *cog-1*.

Table 2: List of ASEL off *che-1* mutants retrieved from *otIs4; vsIs33* using the COPAS worm sorter.

strain name	genotype	phenotype	n=	% off	Notes
OH8111	<i>ot421; otIs4</i>	ASEL off	56	100	
OH8112	<i>ot422; otIs4</i>	ASEL/R off	67	100	
OH8113	<i>ot423; otIs4</i>	ASEL/R off	54	100	
OH8159	<i>ot431; otIs4; vsIs33</i>	ASEL off	22	100	
OH8160	<i>ot432; otIs4; vsIs33</i>	ASEL off	21	95	
OH8161	<i>ot433; otIs4; vsIs33</i>	ASEL off	27	93	
OH8174	<i>ot435; otIs4; vsIs33</i>	ASEL off			likely <i>che-1</i>
OH8176	<i>ot436; otIs4; vsIs33</i>	ASEL off			likely <i>che-1</i>
OH8177	<i>ot437; otIs4; vsIs33</i>	ASEL off			likely <i>che-1</i>
OH8178	<i>ot438; otIs4; vsIs33</i>	ASEL off			did not grow
OH8179	<i>ot439; otIs4; vsIs33</i>	ASEL off			likely <i>che-1</i>

Notes:

Complementation tests were performed by mating *che-1(ot66)* males to mutant hermaphrodites. F1 male progeny were scored. Duplicate tests were made for each mutant. “Likely *che-1*” means complementation tests were not performed but based on the observation of high penetrance and lack of sickness, these mutants are likely *che-1* alleles. “+” means that a mutant is an allele of *che-1*.

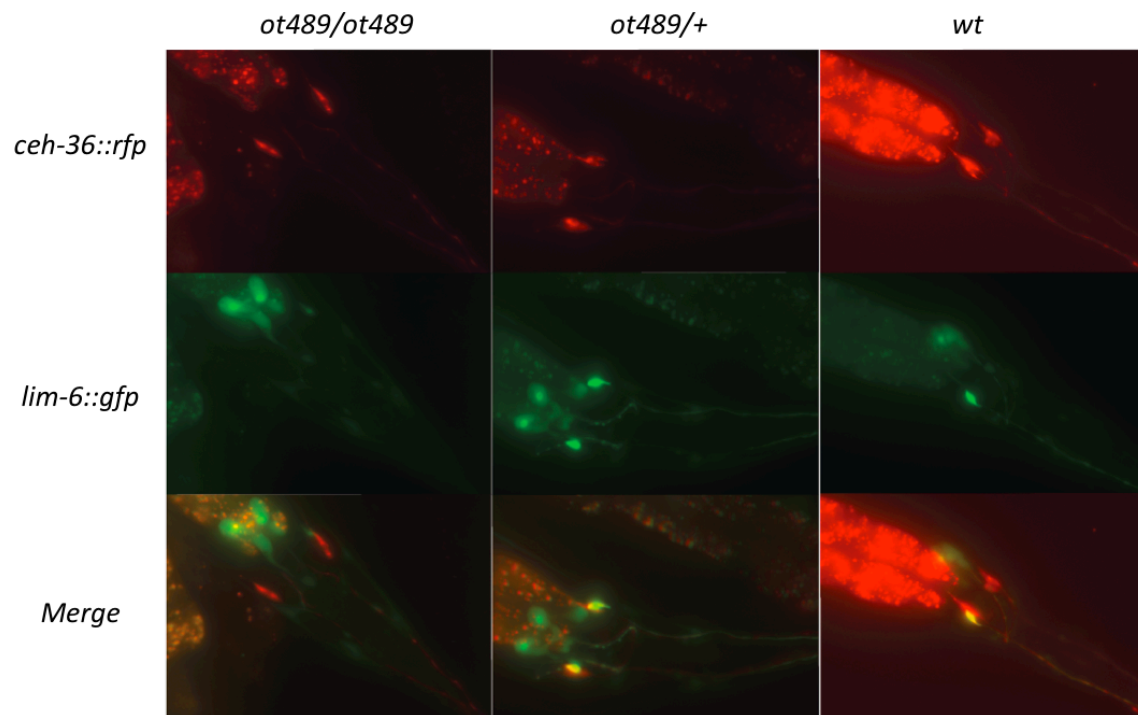
Table 3: List of ASEL off *che-1* mutants retrieved from *otIs243*.

Strain Name	Genotype	Chromosome	<i>che-1</i> allele	Notes
EMS treated				
OH8661	<i>ot490; otIs114</i>	not X	+	
OH8662	<i>ot491; otIs114</i>	not X	+	
OH8714	<i>ot504; otIs114; otIs243</i>	not X	+	
OH8715	<i>ot505; otIs114</i>	not X	+	
OH8760	<i>ot512; otIs114; otIs243</i>	not X	+	
OH8761	<i>ot513; otIs114</i>	not X	+	
OH8763	<i>ot514; otIs114; otIs243</i>	not X	+	Him?
OH8764	<i>ot515; otIs114</i>	not X	?	did not grow
OH8781	<i>ot516; otIs114</i>	not X	+	
OH8782	<i>ot517; otIs114</i>	not X	+	
OH8783	<i>ot518; otIs114</i>	not X	+	
OH8784	<i>ot519; otIs114</i>	not X	+	
OH8799	<i>ot521; otIs114</i>	not X	+	
OH8800	<i>ot522; otIs114</i>	not X	+	
OH8801	<i>ot523; otIs114; otIs243</i>	not X	+	
OH8809	<i>ot524; otIs114</i>	not X	+	
OH8815	<i>ot525; otIs243</i>	not X	+	
OH8825	<i>ot527; otIs243</i>		+	
OH8826	<i>ot528; otIs243</i>		+	
OH8827	<i>ot529; otIs243</i>	not X	+	
OH8828	<i>ot530; otIs243</i>		+	
OH8829	<i>ot531; otIs243</i>		+	
OH8830	<i>ot532; otIs243</i>	not X	+	
OH8831	<i>ot533; otIs243</i>	not X	+	
OH8832	<i>ot534; otIs243</i>	not X		
OH8841	<i>ot535; otIs243</i>	not X	+	
ENU treated				
OH8842	<i>ot536; otIs243</i>			likely <i>che-1</i>
OH8843	<i>ot537; otIs243</i>			likely <i>che-1</i>
OH8844	<i>ot538; otIs243</i>			likely <i>che-1</i>
OH8845	<i>ot538; otIs243</i>			likely <i>che-1</i>

Notes:

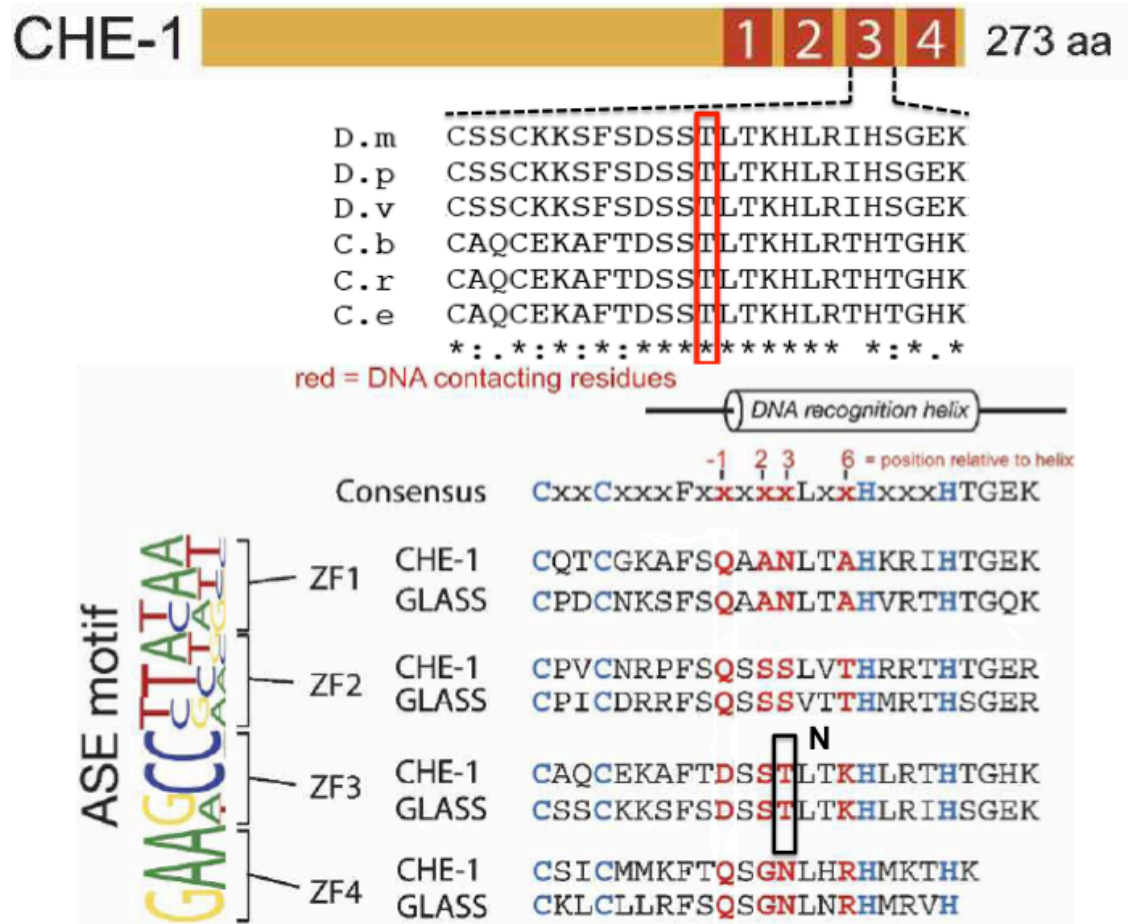
1. *otIs243* = *che-1_fosmid::venus gcy-7::gfp rgef-1::dsred2*
2. Number of genomes screened: 48,000
3. Some strains have been outcrossed (*otIs114* used).
4. Complementation tests were performed by mating *che-1*(*ot66*) males to mutant hermaphrodites. F1 male progeny were scored. Duplicate tests were made for each mutant. "Likely *che-1*" means complementation tests were not performed but based on the observation of high penetrance and lack of sickness, these mutants are likely *che-1* alleles.

Figure 1: *ot489* is a semi-dominant allele of *che-1*.



ot489 was originally uncovered from a screen looking for 2ASEL mutants using *lsy-6*(*ot71*); *otIs4*(*gcy-7::gfp*); *vsIs33*(*dop-3::dsRed*). Preliminary analysis indicated that this mutation is not on LG X. The strain displays “2ASEL” and “ASEL off” phenotypes and is semi-dominant. The phenotype of *ot489/ot489* homozygous animals is “ASEL off”, while *ot489/+* worms are “2 ASEL” or “1 ASEL”. Sanger sequencing of the *che-1* locus reveals a mutation in one of the highly conserved DNA contacting residues (T, Threonine to N, Asparagine) in the 3rd zinc finger of *che-1*.

Figure 2: The molecular nature of *ot489*. Gene structure adapted from (Etchberger et al., 2007).



Reference:

Etchberger, J.F., Lorch, A., Sleumer, M.C., Zapf, R., Jones, S.J., Marra, M.A., Holt, R.A., Moerman, D.G., and Hobert, O. (2007). The molecular signature and cis-regulatory architecture of a *C. elegans* gustatory neuron. *Genes Dev* 21, 1653-1674.

APPENDIX 3:

Table 1: List of variants in *ot219* from CloudMap (WS235 genome release).

Position	Reference	Change	Gene_name	Bio_type	Trancrypt_ID	old/new_AA	Old/New codon	Codon_Number
12587247	C	T	E02A10.3	protein_coding	E02A10.3a	Q/*	Cag/Tag	56
12587247	C	T	E02A10.3	protein_coding	E02A10.3b	Q/*	Cag/Tag	60
12587247	C	T	E02A10.3	protein_coding	E02A10.3c	Q/*	Cag/Tag	56
12729626	C	T	R11D1.10	protein_coding	R11D1.10a.2	A/V	gCc/gTc	21
12729626	C	T	R11D1.10	protein_coding	R11D1.10a.1	A/V	gCc/gTc	21

Note: All variants are between 12.5- 12.75 on LG V (100-recombinant experiment).

Table 2: List of uncovered regions within mapped region on LG V in *ot219*.
(WS235 genome release).

Start	End	Size	Gene_ID	Gene_name	Bio_type	Trancrypt_I D	Effect
12576011	12576082	71	E02A10.8	E02A10.8	ncRNA	E02A10.8	UPSTREAM: 872 bases
12576011	12576082	71	E02A10.4	E02A10.4	protein_coding	E02A10.4	UPSTREAM: 8060 bases
12576011	12576082	71	E02A10.10	E02A10.10	ncRNA	E02A10.10	UPSTREAM: 2200 bases
12576011	12576082	71	E02A10.9	E02A10.9	ncRNA	E02A10.9	UPSTREAM: 2588 bases
12576011	12576082	71	E02A10.5	E02A10.5	protein_coding	E02A10.5	UPSTREAM: 3663 bases
12576011	12576082	71	E02A10.7	E02A10.7	ncRNA	E02A10.7	UPSTREAM: 7397 bases
12576011	12576082	71	E02A10.3	E02A10.3	protein_coding	E02A10.3b	UPSTREAM: 7825 bases
12576011	12576082	71	E02A10.3	E02A10.3	protein_coding	E02A10.3c	UPSTREAM: 8819 bases
12576011	12576082	71	E02A10.3	E02A10.3	protein_coding	E02A10.3a	UPSTREAM: 9088 bases
12576011	12576082	71	E02A10.2	grl-23	protein_coding	E02A10.2a	Exon_V_12575852_12576114
12576011	12576082	71	E02A10.2	grl-23	protein_coding	E02A10.2b	INTRON
12579753	12579771	18	E02A10.7	E02A10.7	ncRNA	E02A10.7	UPSTREAM: 3655 bases
12579753	12579771	18	E02A10.3	E02A10.3	protein_coding	E02A10.3b	UPSTREAM: 4083 bases
12579753	12579771	18	E02A10.3	E02A10.3	protein_coding	E02A10.3c	UPSTREAM: 5077 bases
12579753	12579771	18	E02A10.3	E02A10.3	protein_coding	E02A10.3a	UPSTREAM: 5346 bases
12579753	12579771	18	E02A10.10	E02A10.10	ncRNA	E02A10.10	DOWNSTREAM: 1405 bases
12579753	12579771	18	C14C10.1	C14C10.1	protein_coding	C14C10.1	UPSTREAM: 8675 bases
12579753	12579771	18	E02A10.9	E02A10.9	ncRNA	E02A10.9	DOWNSTREAM: 1009 bases
12579753	12579771	18	C14C10.7	ttr-43	protein_coding	C14C10.7	DOWNSTREAM: 9959 bases
12579753	12579771	18	E02A10.6	E02A10.6	ncRNA	E02A10.6	DOWNSTREAM: 6775 bases
12579753	12579771	18	E02A10.2	grl-23	protein_coding	E02A10.2a	UPSTREAM: 2858 bases
12579753	12579771	18	E02A10.2	grl-23	protein_coding	E02A10.2b	UPSTREAM: 2858 bases
12579753	12579771	18	E02A10.8	E02A10.8	ncRNA	E02A10.8	DOWNSTREAM: 2656 bases
12579753	12579771	18	E02A10.5	E02A10.5	protein_coding	E02A10.5	Exon_V_12579674_12579979
12647810	12647882	72	F56H9.2	F56H9.2	protein_coding	F56H9.2b	DOWNSTREAM: 9573 bases
12647810	12647882	72	F56H9.2	F56H9.2	protein_coding	F56H9.2a	DOWNSTREAM: 9572 bases
12647810	12647882	72	W05B10.6	W05B10.6	protein_coding	W05B10.6	DOWNSTREAM: 9064 bases
12647810	12647882	72	F56H9.5	lin-25	protein_coding	F56H9.5	UPSTREAM: 2048 bases
12647810	12647882	72	F56H9.3	gpa-8	protein_coding	F56H9.3	UPSTREAM: 7340 bases
12647810	12647882	72	F56H9.8	F56H9.8	protein_coding	F56H9.8a	DOWNSTREAM: 1178

							bases
12647810	12647882	72	F56H9.8	F56H9.8	protein_coding	F56H9.8b	DOWNSTREAM: 1178 bases
12647810	12647882	72	F56H9.4	gpa-9	protein_coding	F56H9.4a	INTRON
12647810	12647882	72	F56H9.4	gpa-9	protein_coding	F56H9.4b	UTR_5_PRIME: 526 bases from TSS
12866663	12866732	69	T07F10.3	T07F10.3	protein_coding	T07F10.3	DOWNSTREAM: 1363 bases
12866663	12866732	69	T07F10.6	T07F10.6	protein_coding	T07F10.6	INTRON
12866663	12866732	69	T07F10.1	T07F10.1	protein_coding	T07F10.1a	DOWNSTREAM: 2830 bases
12866663	12866732	69	T07F10.1	T07F10.1	protein_coding	T07F10.1b	DOWNSTREAM: 2830 bases
12866663	12866732	69	T07F10.12	T07F10.12	ncRNA	T07F10.12	UPSTREAM: 8233 bases
12866663	12866732	69	T07F10.10	T07F10.10	ncRNA	T07F10.10	DOWNSTREAM: 8235 bases
12866663	12866732	69	T07F10.5	T07F10.5	protein_coding	T07F10.5.2	DOWNSTREAM: 116 bases
12866663	12866732	69	T07F10.5	T07F10.5	protein_coding	T07F10.5.1	DOWNSTREAM: 107 bases
12975514	12975570	56	R186.7	R186.7	protein_coding	R186.7	DOWNSTREAM: 2231 bases
12975514	12975570	56	R186.14	R186.14	ncRNA	R186.14	DOWNSTREAM: 8085 bases
12975514	12975570	56	R186.10	R186.10	ncRNA	R186.10	DOWNSTREAM: 1205 bases
12975514	12975570	56	R186.13	R186.13	ncRNA	R186.13	UPSTREAM: 69 bases
12975514	12975570	56	R186.2	srd-35	protein_coding	R186.2a	DOWNSTREAM: 9492 bases
12975514	12975570	56	R186.2	srd-35	protein_coding	R186.2b	DOWNSTREAM: 9492 bases
12975514	12975570	56	R186.3	R186.3	protein_coding	R186.3.1	UPSTREAM: 8304 bases
12975514	12975570	56	R186.3	R186.3	protein_coding	R186.3.2	UPSTREAM: 8302 bases
12975514	12975570	56	R186.8	R186.8	protein_coding	R186.8.2	UPSTREAM: 7625 bases
12975514	12975570	56	R186.8	R186.8	protein_coding	R186.8.1	UPSTREAM: 7532 bases
12975514	12975570	56	R186.4	lin-46	protein_coding	R186.4	DOWNSTREAM: 4782 bases
12975514	12975570	56	R186.5	shw-3	protein_coding	R186.5	INTRON