Gene regulatory factors that control the
to 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
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. Trans-acting factors that act as terminal selectors in C. elegans.........................................................12

5.2.1. che-1................................................................................................................................................12

5.2.2. unc-3...............................................................................................................................................14

5.2.3. unc-30............................................................................................................................................16

5.2.4. Combinatorial codes of gene regulatory factors..............................................................................17

5.3. Terminal selector regulation in vertebrates.........................................................................................20
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 *ttx-3* in the larval and adult nervous system..............61

   3.2. *ttx-3* 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. *ttx-3* controls terminal differentiation of the serotonergic NSM neurons.......64
3.5. The POU homeobox gene *unc-86* also controls NSM identity.........................64
3.6. *unc-86* cooperates with *ttx-3* to control NSM identity.............................67
3.7. *unc-86* and *ttx-3* affect axonal arborization and presynaptic specializations......67
3.8. *unc-86* controls terminal differentiation of the cholinergic IL2, URA and URB sensory, motor and interneurons.................................................................68
3.9. *unc-86* cooperates with the ARID transcription factor *cfi-1* 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 *ot704* and *ot705* using Whole genome sequencing and CloudMap...85
   2.4. Rescue experiments for *ot704* and *ot705*..................................................86
   2.5. Expression pattern of *unc-42*, *hlh-16* and *cnd-1* 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. *ot146* is an allele of the LIM homeobox gene *lim-6*.........................................................101
   3.2. *ot108* affects a member of a C2H2 Zn-finger protein family..............................................101
   3.3. *ot108* is an altered function allele.......................................................................................102
   3.4. Expression pattern and timing of action of *lsy-27*.................................................................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 *lsy-12*, a histone acetyltransferase.........................119
   3.2. *lsy-12* 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 *die-1* Zn finger transcription factor is also continuously required to maintain ASE laterality and is a candidate recruiter of the MYST complex.

4. References

5. Supplementary Information

CHAPTER 6: Discussion and future directions

1. Combinatorial codes of terminal selector regulation
   1.1. Continuous requirement of terminal selector transcription factors
   1.2. Target gene specificities
   1.3. Looking for cofactors
   1.4. TTX-3 function in AIN
   1.5. Considerations in other systems

2. Specification of the RMD neuron class
   2.1. Terminal selector for the RMD neuron class
   2.2. Transcriptional regulation of gene expression
   2.3. Proneural roles of *cnd-1* and *hlh-16*
   2.4. The screen

3. ASE asymmetry
   3.1. Considerations on ASE screens: manual clonal vs. automatic non-clonal
   3.2. Whole genome sequencing
   3.3. LSY-27 function in triggering *lim-6* initiation
   3.4. The LSY-12/LSY-13/LIN-49 complex and its function

4. References
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
Table 1: Examples of invertebrate and vertebrate regulons in the nervous system..........44

CHAPTER 2

Figure 1: Expression pattern of the ttx-3 LIM homeobox gene.................................61
Figure 2: ttx-3 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 unc-86 and ttx-3 on the serotonergic identity of the NSM neurons........................................................................................................................................65
Figure 5: The effect of unc-86 and ttx-3 on other identity features of the NSM neurons.66
Figure 6: Cis-regulatory analysis of NSM identity specification....................................68
Figure 7: unc-86 and ttx-3 affect NSM morphology.....................................................69
Figure 8: unc-86 and cfi-1 control cholinergic IL2 identity..........................................70
Table 1: Summary of the effects of ttx-3 and unc-86 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 ttx-3(ot22) mutant animals............75
Suppl. Figure 3: BH4 pathway reporters...........................................................................76
Suppl. Figure 4: *unc-86* controls the identity of the cholinergic URA and URB neurons ............................................................................................................................................77
Suppl. Table 1: Rescue of NSM and AIA differentiation defects of *ttx-3* 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 *ot704* and *ot705* SNP reads............................91
Figure 3: Overview of the *hlh-16* and *cnd-1* locus and relevant alleles. .................93
Figure 4: Preliminary lineaging representation *cnd-1* expression in the RMDD and RMDV lineage..................................................................................................................................................94
Table 1: Quantification of *unc-42*, *ot704* and *ot705* mutant phenotypes.............96
Table 2: *ot704* rescue experiments..............................................................................97
Table 3: *ot705* rescue experiments............................................................................98

**CHAPTER 4**

Figure 1: *lsy* genes and mutant phenotypes. ...............................................................101
Figure 2: *lsy-27* is a new C2H2 Zn-finger protein....................................................104
Figure 3: *lsy-27* is expressed and acts during the initiation but not during the maintenance phase of left/right asymmetry control..............................................................................................................................105
Table 1: Lsy phenotypes of *lim-6* and *lsy-27*............................................................102
Table 2: Transformation rescue and RNAi analysis.....................................................103
Suppl. Figure 1: Updated lim-6 locus and the location of the ot146 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 lsy-27 paralog ztf-25.................................110
Table S1: Whole genome sequencing setting and results..............................................113
Table S2: Mutant analysis of lsy-27 paralogs...............................................................114
Table S3: Retired and novel lsy gene names.................................................................115
Table S4: Final summary of mutant classes & genes....................................................116

CHAPTER 5

Figure 1: lsy-12, a MYST-type histone acetyltransferase, affects ASE laterality............120
Figure 2: lsy-12 is continually required in ASE but may also act early.....................121
Figure 3: The MYST-complex component lsy-13 affects ASE laterality......................122
Suppl. Figure 1: Mapping lsy-12....................................................................................128
Suppl. Figure 2: The MYST complex component lin-49 affects ASE laterality...........129
Suppl. Table 1: Laterality phenotype of lsy-12 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 troubleshoot 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).
more recent PCR screen identified *atonic* (*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 *atonic* 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 β1 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 graduate 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 reply 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, fours 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 continuous 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
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
anallogues. 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 ABpraappap (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 \textit{lcy-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 \textit{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, \textit{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 \textit{die-1} and \textit{cog-1} expression is initiated later at around 3-fold stage in the postmitotic ASEs (Cochella and Hobert, 2012). This implies that \textit{lsy-6} may serve as the earliest switch as an entry point for asymmetry establishment. In fact, TBX-37/38 expression in the A\textsubscript{B}a lineage (Figure 5) primes the \textit{lsy-6} locus through a downstream primer element very early and establishes the open chromatin conformation in the A\textsubscript{SEL} 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 \textit{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 \textit{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, \textit{ash-2}, \textit{dpy-30} and \textit{rbbp-5} (Poole \textit{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 \textit{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 \textit{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 otl317 (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 identity. 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)
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) 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).
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).
Table 1: Examples of invertebrate and vertebrate regulons in the nervous system. From (Hobert, 2011).

<table>
<thead>
<tr>
<th>Terminal selector</th>
<th>Type of transcription factor</th>
<th>Terminal neuron identity controlled by terminal selector</th>
<th>Selected examples of direct target genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHE-1</td>
<td>C2H2 zinc finger</td>
<td>C. elegans gustatory neuron class ASB</td>
<td>Putative chemoreceptors of the GCK family, neuropeptides of the FLP family, cyclic-nucleotide gated ion channels, itself (autoregulation), many others</td>
</tr>
<tr>
<td>TTX-3/CEH-10</td>
<td>LIM/Prd homeodomain</td>
<td>C. elegans cholinergic interneuron class AIB</td>
<td>Choline acetyltransferase, choline transporter, neuropeptide receptors, neuropeptides, Ig domain proteins, ion channels, itself (autoregulation), many others</td>
</tr>
<tr>
<td>UNC-86/MEC-3</td>
<td>POU/LIM homeodomain</td>
<td>C. elegans glutamatergic touch sensory neurons</td>
<td>Touch receptor channel subunits, specialized tubulin mec-7, extracellular matrix proteins, vesicular glutamate transporter, itself (autoregulation), many others</td>
</tr>
<tr>
<td>UNC-30</td>
<td>Fyb-type homeodomain</td>
<td>C. elegans GABAergic motor neurons</td>
<td>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</td>
</tr>
<tr>
<td>AST-1/CEH-41</td>
<td>ETS/homeodomain</td>
<td>C. elegans dopaminergic neuron</td>
<td>Tyrosine hydroxylase, aromatic amino acid decarboxylase, vesicular dopamine transporter, dopamine reuptake transporter, dopamine autoreceptor, ion channels</td>
</tr>
<tr>
<td>UNC-3</td>
<td>EBF-type</td>
<td>C. elegans A/B-type cholinergic motor neurons</td>
<td>Choline acetyltransferase, choline transporter, neuropeptide receptors, neuropeptides, Ig domain proteins, signaling proteins, many others</td>
</tr>
<tr>
<td>Pet-1</td>
<td>ETS domain</td>
<td>Mouse serotonergic neurons</td>
<td>Tryptophan hydroxylase, 5-HT1a receptor, aromatic amino acid decarboxylase, serotonin transporter, itself (autoregulation), many others</td>
</tr>
<tr>
<td>Csx</td>
<td>Homeodomain</td>
<td>Mouse retinal photoreceptors</td>
<td>Retinol-binding protein, rhodopsin, G protein Gnt1, phosphodiesterase, vesicular glutamate transporter, kinase, itself (autoregulation), many others</td>
</tr>
<tr>
<td>Nurr1/Pitx-3</td>
<td>C4 zinc finger/homeodomain</td>
<td>Mouse midbrain dopaminergic neurons</td>
<td>Tyrosine hydroxylase, aromatic amino acid decarboxylase, vesicular dopamine transporter, dopamine reuptake transporter, dopamine autoreceptor D2R, receptor tyrosine kinase Ret, many others</td>
</tr>
</tbody>
</table>
References:


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


CHAPTER 2:

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


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

Feifan Zhang\(^1\), Abhishek Bhattacharya\(^1\), Jessica C. Nelson\(^2\), Namiko Abe\(^1\), Patricia Gordon\(^1\), Carla Lloret-Fernandez\(^3\), Miren Maicas\(^3\), Nuria Flames\(^1,3\), Richard S. Mann\(^1\), Daniel A. Colón-Ramos\(^2\) 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 \( C. \) \textit{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: \textit{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 \( C. \) \textit{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 \( C. \) \textit{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 \( C. \) \textit{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 A\textit{IY} interneuron class. A\textit{IY} 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 A\textit{IY}, 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 \textit{cis}-regulatory motif shared by all of its target genes. \( ttx-3 \) expression is turned on in the
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 is it 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 (wgIs68) in first larval stage animals and in adult animals. D-V, dorsal-ventral. 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 AIIY differentiation defect of *ttx-3* mutant animals, mirrors the expression of the smaller, locus-restricted reporter construct in the AIIY, 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 AIIIN 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, 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*uniso::*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 neuropeptidases *flp-2* and *ins-1*; the receptor tyrosine kinase *sra-1*; 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*, *sda-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).

AI A 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 blebbled in *ttx-3* mutants (supplementary material Fig. S2).

The AIIIN 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 linearly 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 AIIYN 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 AIIYN 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 AIIYN neurons since it operates with distinct co-factors and through distinct cis-regulatory elements. The co-factor of *ttx-3* in AIIYN, 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 AIIYN versus AIA identity are distinct. In the AIIYN neurons, *ttx-3* acts on its many target genes through a cis-regulatory motif, termed the ‘AIIYN motif’, that provides a cooperative binding site for a TTX-3–CEH-10 heterodimer (Wenick and Hobert, 2004). Mutation of the AIIYN motif in a locus that is expressed in AIIYN and AIA neurons, the cholinergic *cho-1* locus, results in a severe reduction in expression in the AIIYN 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, TAAATNGA (Fig. 3D). In two cases, mutation of the TAAATNGA alone affected reporter gene expression, whereas in the third case (*cho-1*) a...
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 TAATNGA 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 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).
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. S1).

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, pips-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: zdIs13 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.
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).
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. 51B,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 linearly 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

---

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

<table>
<thead>
<tr>
<th>Identity feature</th>
<th>Function</th>
<th>ttx-3(–)</th>
<th>unc-86(–)</th>
<th>unc-86(–); ttx-3(–)</th>
<th>Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>cat-1</td>
<td>5HT pathway</td>
<td>wt</td>
<td>dim</td>
<td>off</td>
<td>Synergism</td>
</tr>
<tr>
<td>cat-4</td>
<td>5HT pathway</td>
<td>dim</td>
<td>wt</td>
<td>very dim</td>
<td>Synergism</td>
</tr>
<tr>
<td>mod-5</td>
<td>5HT pathway</td>
<td>wt</td>
<td>wt</td>
<td>off</td>
<td>Synergism</td>
</tr>
<tr>
<td>nlp-13</td>
<td>Neuropeptide</td>
<td>wt</td>
<td>dim</td>
<td>off</td>
<td>Synergism</td>
</tr>
<tr>
<td>nlp-3</td>
<td>Neuropeptide</td>
<td>wt</td>
<td>wt</td>
<td>off</td>
<td>Synergism</td>
</tr>
<tr>
<td>ftr-2</td>
<td>Transmembrane</td>
<td>wt</td>
<td>wt</td>
<td>dimmer</td>
<td>Synergism</td>
</tr>
<tr>
<td>scd-2</td>
<td>Kinase</td>
<td>wt</td>
<td>dim</td>
<td>off</td>
<td>Synergism</td>
</tr>
<tr>
<td>flp-4</td>
<td>Neuropeptide</td>
<td>dim</td>
<td>wt</td>
<td>stronger expression</td>
<td>Synergism</td>
</tr>
<tr>
<td>5HT antibody staining</td>
<td>Neurotransmitter</td>
<td>wt</td>
<td>wt</td>
<td>off</td>
<td>Synergism</td>
</tr>
<tr>
<td>bas-1</td>
<td>5HT pathway</td>
<td>off</td>
<td>dim</td>
<td>n.d.</td>
<td>?</td>
</tr>
<tr>
<td>ptpa-1</td>
<td>5HT pathway</td>
<td>off</td>
<td>dim</td>
<td>n.d.</td>
<td>?</td>
</tr>
<tr>
<td>tph-1</td>
<td>5HT pathway</td>
<td>wt</td>
<td>off</td>
<td>n.d.</td>
<td>?</td>
</tr>
<tr>
<td>mgl-3</td>
<td>GPCR</td>
<td>wt</td>
<td>off</td>
<td>n.d.</td>
<td>?</td>
</tr>
<tr>
<td>dcp-3</td>
<td>GPCR</td>
<td>wt</td>
<td>off</td>
<td>n.d.</td>
<td>?</td>
</tr>
<tr>
<td>mgl-1</td>
<td>GPCR</td>
<td>off</td>
<td>wt</td>
<td>n.d.</td>
<td>?</td>
</tr>
</tbody>
</table>

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.
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-Hagesteijn 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
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
combined with the \textit{cfl-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 \textit{unc-86} and \textit{cfl-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 express 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 AIAIs, 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 AY 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, AY 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 AY neurons can be explained by neuron type-specific co-factors and by ttx-3 acting through distinct cis-regulatory motifs. AY-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 AY interneuron class, in which TTX-3 and CEH-10 operate through a bipartite motif (the ‘AIE motif’) composed of a TTX-3 and a CEH-10 binding site (Wenick and Hobert, 2004). Genes that are expressed in both AIA and AY neurons (e.g., cho-1) contain a modular assembly of both the AIE and AIA cis-regulatory signatures.

Similar to the ttx-3-dependent control of the central cholinergic interneurons AY 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.

The 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 AY. 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 synchronically 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 Salavaterra, 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; Lifc 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/µl) with digested rol-6((d)) (3 ng/µl) as injection marker, or plasmid mix was directly injected [50 ng/µl together with 100 ng/µl 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) pLyS5 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 trx-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 R01NS052686-03 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 +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


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.
Suppl. Figure 2: AIA morphology in wildtype and ttx-3(ot22) mutant animals. AIA morphology was visualized with ins-1::gfp (otIs326). Left panels shows 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.
Suppl. Figure 3: BH4 pathway reporters.
Reporter genes and overview of expression pattern of BH4 pathway genes. Expression patterns were observed with multiple lines.

Expression pattern:
- **gfp-1**
  - not NSM, not HSN
  - a few very dim cells in the head

- **ptps-1**
  - NSM
  - HSN
  - VC4 and VC5
  - no ADF
  - more VNC neurons
tail neuron

- **odor-1**
  - not NSM, not HSN
  - several dim cells in the head (ADF?)
tail and head hypodermal cells

- **pcbd-1**
  - 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 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

<table>
<thead>
<tr>
<th></th>
<th>% animals expressing mgl-1::mcherry in NSM</th>
<th>% animals expressing mgl-1::mcherry in AIA</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>wild type</td>
<td>100</td>
<td>100</td>
<td>&gt;100</td>
</tr>
<tr>
<td>ttx-3(ot22)</td>
<td>0</td>
<td>0</td>
<td>&gt;100</td>
</tr>
<tr>
<td>ttx-3(ot22); Ex[cat-1&lt;sup&gt;prom&lt;/sup&gt;::ttx-3cDNA] line1</td>
<td>92</td>
<td>0</td>
<td>36</td>
</tr>
<tr>
<td>ttx-3(ot22); Ex[cat-1&lt;sup&gt;prom&lt;/sup&gt;::ttx-3cDNA] line2</td>
<td>89</td>
<td>0</td>
<td>37</td>
</tr>
<tr>
<td>ttx-3(ot22); Ex[ins-1&lt;sup&gt;prom&lt;/sup&gt;::ttx-3cDNA] line1</td>
<td>0</td>
<td>79</td>
<td>38</td>
</tr>
<tr>
<td>ttx-3(ot22); Ex[ins-1&lt;sup&gt;prom&lt;/sup&gt;::ttx-3cDNA] line2</td>
<td>0</td>
<td>58</td>
<td>36</td>
</tr>
<tr>
<td>ttx-3(ot22); Ex[ins-1&lt;sup&gt;prom&lt;/sup&gt;::ttx-3cDNA] line3</td>
<td>0</td>
<td>79</td>
<td>38</td>
</tr>
</tbody>
</table>

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>
<thead>
<tr>
<th>Strain/Array</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>ttx-3(ot22)</td>
<td>Premature stop before the homeobox (Altun-Gultekin et al., 2001).</td>
</tr>
<tr>
<td>unc-86(n846)</td>
<td>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&gt;A splice acceptor site mutation at end of the second intron of the C30H5.7a transcript in the middle of the POU domain (aatactcagGCGG to aatactcaaGCGG)</td>
</tr>
<tr>
<td>unc-86(n848)</td>
<td>The molecular nature of this temperature-sensitive allele is a GT to AT splice donor site mutation in intron 4 (Röhrig, 2000)</td>
</tr>
<tr>
<td>cfr-1(ky651)</td>
<td>A splice acceptor site mutation before the DNA binding domain (Shaham and Bargmann, 2002)</td>
</tr>
<tr>
<td>otIs224</td>
<td>Is[cat-1::gfp] (Flames and Hobert, 2009)</td>
</tr>
<tr>
<td>otIs225</td>
<td>Is[cat-4::gfp] (Flames and Hobert, 2009)</td>
</tr>
<tr>
<td>otIs226</td>
<td>Is[bas-1::gfp] (Flames and Hobert, 2009)</td>
</tr>
<tr>
<td>zdIs13</td>
<td>Is[tpb-1::gfp] (Clark and Chiu, 2003)</td>
</tr>
<tr>
<td>wgIs68</td>
<td>Is[ttx-3fosmid::EGFP-FLAG, unc-119(+)]. 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.</td>
</tr>
<tr>
<td>otIs337</td>
<td>Is[unc-86 fosmid WRM0612cF07::NLS::YFP::H2B; ttx-3::mCherry]. The unc-86 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 unc-86 in fosmid WRM0612cF07.</td>
</tr>
<tr>
<td>vsIs33</td>
<td>Is[dop-3::dsRed]. Kindly provided by Michael Koelle</td>
</tr>
<tr>
<td>otIs317</td>
<td>Is[mgl-1prom::mcherry, pha-1]. DNA kindly provided by Kaveh Ashrafi (Greer et al., 2008)</td>
</tr>
<tr>
<td>otIs341</td>
<td>Is[mgl-1prom::gfp] -1994 to -1374 bp upstream of ATG</td>
</tr>
<tr>
<td>otIs379</td>
<td>Is[cho-1prom::gfp; rol-6(d)] -3006 to -2642 bp upstream of ATG</td>
</tr>
<tr>
<td>otIs326</td>
<td>Is[ins-1::gfp; rol-6(d)] -289 bp upstream of ATG</td>
</tr>
<tr>
<td>otEx4687</td>
<td>Ex[glr-2::gfp; rol-6(d)] -1798 bp upstream of ATG</td>
</tr>
<tr>
<td>otEx4886</td>
<td>Ex[ttx-3 intron7::gfp; rol-6(d)] GGAAG+intron7+CGTCTACCAGATGAAGATG cloned into pPD95.75</td>
</tr>
<tr>
<td>otEx5056</td>
<td>Ex[tp-2::gfp; rol-6(d)] -2002 bp upstream of ATG</td>
</tr>
<tr>
<td>otEx4781</td>
<td>Ex[mod-5NSM prom::gfp; elt-2::gfp]. First intron of mod-5 cloned into pPD95.75(CACCAGCAGCTGCAAG+ intron1+ CTGAACTCTCC driving GFP)</td>
</tr>
<tr>
<td>otEx5280</td>
<td>Ex[ptps-1::gfp; rol-6(d)] -2600 bp upstream of ATG</td>
</tr>
<tr>
<td>otEx5163</td>
<td>Ex[nlp-3::gfp; rol-6(d)]. DNA kindly provided by Hart lab (Nathoo et al., 2001)</td>
</tr>
<tr>
<td>otEx5364</td>
<td>Ex[mgl-3::gfp; rol-6(d)].</td>
</tr>
</tbody>
</table>
DNA kindly provided by Ashrafi lab (Greer et al., 2008)

**otEx5163**
[nlp-13::gfp; rol-6(d)]
-1967 bp upstream of ATG

**otEx5055**
[scd-2::gfp; rol-6(d)]
-2045 bp upstream of ATG

**otEx5363**
[fir-2::gfp; rol-6(d)]
DNA kindly provided by Takeshi Ishihara

**otEx4917**
Ex[unc-86 intron1::gfp; rol-6(d)]
GACGACAACCGCTTCAAAAATGCAACCT+intron1+TTCAACAAC
AGTTTATTTGGATCATTCGATGACCC cloned into pPD95.75

**otEx4969, otEx4970**
2 independent lines of Ex[cat-1prom14::ttx-3; rol-6(d)]

**otEx5073, otEx5074, otEx5075**
3 independent lines of Ex[ins-1457bp_prom::ttx-3; rol-6(d)] (-457 bp upstream of ATG)

**Ex[gcy-28.d::gfp]**
the complete genotype of this array is Ex[gcy-28.dp::gcy-28.d::GFP, AIA-specific ins-1p::SNB-1::mRFP, rol-6(+)] (transgene kindly provided by Takeshi Ishihara)

**Ex[scd-2::gfp]**
Ex[scd-2p::scd-2::GFP, AIA-specific ins-1p::mRFP, lin-44::gfp] (transgene kindly provided by Takeshi Ishihara)

**olaEx1446**
Ex [mod-5p::egfp (2ng/ul)/unc-122p::gfp (20ng/ul)]

**olaEx1485**
Ex[flp-4p::egfp (30ng/ul)/unc-122p::DSRED (20ng/ul)]

**nuls9**
Is[unc-5::gfp] (transgene kindly provided by Josh Kaplan)

**otEx2310**
Ex[gcy-19::gfp; unc-122::gfp] (Ortiz et al., 2006)

**lqIs3**
Is[osm-6::gfp] (transgene kindly provided by Erik Lundquist)

**qIs56**
Is[lag-2::gfp] (transgene kindly provided by Judith Kimble)

**vls48**
Is[unc-17::gfp] (transgene kindly provided by Michael Koelle)

**otls323**
Is[cho-1_fosmid::gfp; elt-2::dsRed] (transgene kindly provided by Paschalis Kratsios)
<table>
<thead>
<tr>
<th>Probe</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>cho-1 wt:</td>
<td>5’tacacacacatcgaatctttctttttaaaaaagaagttgtccttaattatttttcccctattcaGCTTTTCGTTCGTCGCCT</td>
</tr>
<tr>
<td>cho-1 TAAT del</td>
<td>5’tacacacacatcgaatctttctttttaaaaaagaagttgtccttaattatttttcccctattcaGCTTTTCGTTCGTCGCCT</td>
</tr>
<tr>
<td>mgl-1 wt</td>
<td>5’gtttccataactcatagtgccttagatagcagcgatcgtttgtgtgccttgcttttaaccggaattaacctgcGCTTTTCGTTCGTCGCCT</td>
</tr>
<tr>
<td>mgl-1 TAAT del</td>
<td>5’gtttccataactcatagtgccttagatagcagcgatcgtttgtgtgccttgcttttaaccggaattaacctgcGCTTTTCGTTCGTCGCCT</td>
</tr>
<tr>
<td>bas-1 wt</td>
<td>5’cccaacacaccaattcttttctttcctaaccactgaaccctttatctcctcaactcagttttt attcgggtttggtgcattcataaatatatttGCTTTTCGTTCGTCGCCT</td>
</tr>
<tr>
<td>bas-1 HD mut</td>
<td>5’cccaacacaccaattcttttctttcctaaccactgaaccctttatctcctcaactcagttttt attcgggtttggtgcattcataaatatatttGCTTTTCGTTCGTCGCCT</td>
</tr>
<tr>
<td>bas-1 POU mut</td>
<td>5’cccaacacaccaattcttttctttcctaaccactgaaccctttatctcctcaactcagttttt attcgggtttggtgcattcataaatatatttGCTTTTCGTTCGTCGCCT</td>
</tr>
<tr>
<td>tph-1 wt</td>
<td>5’tctttgtttgcataataaaaaaactaatacaacacagcaaaagccgtctcaaccctttatttcattgtGCTTTTCGTTCGTCGCCT</td>
</tr>
<tr>
<td>tph-1 HD mut</td>
<td>5’tctttgtttgcataataaaaaaactaatacaacacagcaaaagccgtctcaaccctttatttcattgtGCTTTTCGTTCGTCGCCT</td>
</tr>
<tr>
<td>tph-1 POU mut</td>
<td>5’tctttgtttgcataataaaaaaactaatacaacacagcaaaagccgtctcaaccctttatttcattttGCTTTTCGTTCGTCGCCT</td>
</tr>
</tbody>
</table>
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 \textit{cnd-1}, the Beta3 and Olig family related gene \textit{hlh-16} and the Q50 class paired-like homeobox gene \textit{unc-42}. Fosmid reporters of these factors suggest that \textit{cnd-1} and \textit{hlh-16} likely play proneural roles, while \textit{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 \textit{mgl-1} is expressed in eight neurons: NSML/R, AIAL/R, RMDDL/R and RMDVL/R (Greer et al., 2008). The transgene \textit{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 \textit{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/µl into mutant animals using linearized *rol-6(d)* (1 ng/µl) 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/µl, 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 \textit{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 \textit{hlh-16} as a proneural gene probably works at an earlier step before terminal differentiation takes place.

\textit{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 \textit{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 \textit{stIs10055} \((\textit{cnd-1}^{3.2kb} \text{prom}::\text{HIS-24}::\text{mCherry})\) 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 \textit{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, \textit{cnd-1} expression begins even earlier, at the great grandmother stage, and remains on until after the cell has terminally differentiated (Figure 4B).

\textbf{For future experiments and discussion see Chapter 6.}
References:


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

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

(A)

(B)
Figure 4: Preliminary lineage representation of *cnd-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

<table>
<thead>
<tr>
<th></th>
<th>RMDD=0 RMDV=0</th>
<th>RMDD=0 RMDV=1</th>
<th>RMDD=0 RMDV=2</th>
<th>RMDD=2 RMDV=2</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>wt</em></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100%</td>
<td>&gt;100</td>
</tr>
<tr>
<td><em>ot704</em></td>
<td>14%</td>
<td>23%</td>
<td>63%</td>
<td>-</td>
<td>65</td>
</tr>
<tr>
<td><em>ot705</em></td>
<td>100%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>30</td>
</tr>
<tr>
<td><em>e419(unc-42)</em></td>
<td>100%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>29</td>
</tr>
</tbody>
</table>
Table 2: *ot704* rescue experiments

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


Table 3: *ot705* rescue experiments

<table>
<thead>
<tr>
<th></th>
<th>RMDD =0</th>
<th>RMDD =2</th>
<th>RMDD =0</th>
<th>RMDD =2</th>
<th>RMDD =1</th>
<th>RMDD =2</th>
<th>RMDD =0</th>
<th>RMDD =2</th>
<th>RMDD =1</th>
<th>RMDD =2</th>
<th>RMDD =1</th>
<th>RMDD =2</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>wt</strong></td>
<td>-</td>
<td>100%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>&gt;100</td>
<td></td>
</tr>
<tr>
<td><strong>ot705</strong></td>
<td>100%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Simple 30ng/ul line1:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>ot705: otEx4952</em></td>
<td>17%</td>
<td>-</td>
<td>23%</td>
<td>30%</td>
<td>4%</td>
<td>13%</td>
<td>11%</td>
<td>2%</td>
<td>-</td>
<td>47</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>ot705: no Ex4952</em></td>
<td>97%</td>
<td>-</td>
<td>3%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>31</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simple 30ng/ul line2:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>ot705: otEx4953</em></td>
<td>9%</td>
<td>14%</td>
<td>6%</td>
<td>23%</td>
<td>23%</td>
<td>14%</td>
<td>3%</td>
<td>-</td>
<td>9%</td>
<td>35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>ot705: no Ex4953</em></td>
<td>97%</td>
<td>-</td>
<td>3%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>34</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complex 3 ng/ul line1:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>ot705: otEx4954</em></td>
<td>-</td>
<td>58%</td>
<td>2%</td>
<td>9%</td>
<td>17%</td>
<td>2%</td>
<td>2%</td>
<td>-</td>
<td>9%</td>
<td>43</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>ot705: no Ex4954</em></td>
<td>Did not score because “unc” phenotype is not well rescued. (Animals that do not roll may carry the transgene.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A left/Right asymmetric neuronal differentiation program is controlled by the Caenorhabditis elegans LSY-27 Zn finger transcription factor


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 \textit{lim-6} that is pulled out from our screens. The other mutant has a lesion in the C2H2 zinc finger transcription factor \textit{lsy-27} and displays similar phenotypes to \textit{lim-6}. \textit{lsy-27} also affects \textit{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 \textit{lsy-27}, and conducted all related genetic and molecular analysis as well as temperature shift experiments. Maggie O'Meara cloned \textit{ot146}, and updated the \textit{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 Hobert**

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 invertebrates 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...
derepressed in ASEL (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).

*lims* 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, subsequent 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 (Kadrmas 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 derepression 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*
expression, \textit{ot108} animals also show a significant loss of \textit{lim-6} expression in ASEL, thereby providing an explanation of the \textit{lim-6}-like phenotype of \textit{ot108} mutant animals (Figure 1C).

Upon isolation of \textit{ot108} mutant animals in our original \textit{Lsy} screen (Sarin et al. 2007), we noted that \textit{ot108} fails to complement the derepression of ASER fate in the ASEL phenotype of a mutation in the \textit{die-1} Zn-finger transcription factor, an inducer of \textit{lim-6} expression in \textit{ASEL} (a class II gene that also results in the loss of \textit{ASEL} fate) (Figure 1A). Due to this lack of complementation, we had therefore initially considered \textit{ot108} to be an allele of \textit{die-1} (Sarin et al. 2007). However, our subsequent analysis revealed no mutation in the \textit{die-1} locus of \textit{ot108} mutant animals and, moreover, the \textit{ot108} mutant phenotype could not be rescued with a genomic piece of DNA that rescues a canonical \textit{die-1} allele (data not shown). Subsequent chromosomal linkage analysis showed that \textit{ot108} is linked to chromosome V, while \textit{die-1} maps to chromosome II. After mapping \textit{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 GAIAs 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 F47I4.1 (Figure 2A and Figure S2). F47H4.1 is a member of C2H2 Zn-finger transcription factors with several paralogs in \textit{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 \textit{ot108} is phylogenetically conserved (Figure S2). The only gene in this family that had been previously characterized is the \textit{ham-2} transcription factor, which is involved in \textit{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 \textit{ot108} mutant phenotype (Table 2). Animals carrying a deletion allele of F47H4.1, \textit{tm593} (kindly provided by the \textit{C. elegans} HSN knockout facility at Tokyo Women’s Medical University School of Medicine) (Figure 2A), also display a class IV \textit{Lsy} phenotype (Table 1). Also, like \textit{ot108} animals, \textit{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 \textit{Lsy} phenotype of \textit{ot108} mutant animals, and we therefore called this gene \textit{lsy-27} (Table S3 shows an updated numbering of \textit{lsy} genes).

\textbf{\textit{ot108} is an altered function allele}

The \textit{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 \textit{tm593} animals, which do not contain any of the DNA-binding Zn-fingers (see File S1). We were therefore surprised to note that the \textit{Lsy} phenotype of the \textit{tm593} deletion allele is notably milder than the \textit{ot108} missense allele in terms of both expressivity and penetrance (Table 1). We

<table>
<thead>
<tr>
<th>Table 1 \textbf{Lsy phenotypes of \textit{lim-6} and \textit{lsy-27}}</th>
</tr>
</thead>
<tbody>
<tr>
<td>% animals with the following phenotypes (at 25°C):</td>
</tr>
<tr>
<td>ASEL only (%)</td>
</tr>
<tr>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td>ASEL marker (gcy-7::gfp; ot103)</td>
</tr>
<tr>
<td>Wild type</td>
</tr>
<tr>
<td>\textit{lim-6} (nr2073)</td>
</tr>
<tr>
<td>\textit{lsy-27} (ot146)</td>
</tr>
<tr>
<td>\textit{lsy-27} (tm593)</td>
</tr>
<tr>
<td>\textit{lsy-27} (ot108)</td>
</tr>
<tr>
<td>\textit{lsy-27} (tm593)</td>
</tr>
<tr>
<td>\textit{lsy-27} (ot108)</td>
</tr>
<tr>
<td>\textit{lsy-27} (ot108)</td>
</tr>
<tr>
<td>ASER marker (gcy-5::gfp; nts1)</td>
</tr>
<tr>
<td>Wild type</td>
</tr>
<tr>
<td>\textit{lim-6} (nr2073)</td>
</tr>
<tr>
<td>\textit{lsy-27} (ot108)</td>
</tr>
<tr>
<td>\textit{lsy-27} (tm593)</td>
</tr>
<tr>
<td>\textit{lsy-27} (ot108)</td>
</tr>
</tbody>
</table>

\Note 755

102
therefore considered the possibility that \textit{ot108} (which is recessive) is an altered function allele (Table 1). We tested this possibility by removing \textit{lsy-27} gene activity in \textit{ot108} mutant animals using RNA interference (RNAi) directed against \textit{lsy-27}. We found that RNAi treatment completely reverts the \textit{ot108} phenotype (Table 2), suggesting that it is indeed altered \textit{lsy-27} function that explains the \textit{ot108} phenotype.

We noted that animals that carry one copy of the \textit{ot108} allele and one copy of the \textit{tm593} allele display a phenotype that is even milder than the phenotype of either allele alone (Table 1). One copy of the \textit{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 \textit{tm593} phenotype. The need for sufficient \textit{ot108} dosage is also illustrated by the fact that the phenotype of \textit{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 \textit{lsy-27} in \textit{tm593} animals may be mostly compensated for by \textit{lsy-27} paralogs, while the \textit{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 \textit{C. elegans} knockout facility in Tokyo), we found that neither of the two most closely related \textit{lsy-27} paralogs, \textit{ztf-25} or \textit{ztf-28}, either alone or in combination (\textit{i.e.}, \textit{ztf-25 ztf-28} double nulls) displayed a \textit{Lsy} phenotype (Table S2). \textit{ztf-25} \textit{lsy-27} double-null mutant animals also display no \textit{Lsy} phenotype. \textit{ztf-25} \textit{lsy-27} double mutants could not be built due to close linkage of the two loci, and we therefore needed to resort to RNAi. \textit{lsy-27} RNAi in a \textit{ztf-28 ztf-25} double-mutant background also did not result in a \textit{Lsy} phenotype, but we note that even though \textit{lsy-27} RNAi does suppress the \textit{ot108} \textit{Lsy} phenotype, it does not recapitulate the \textit{lsy-27(tm593)} phenotype (Table 2), thereby allowing no firm conclusion about a triple loss of function of all three \textit{lsy-27} paralogs.

**Expression pattern and timing of action of \textit{lsy-27}**

By recombining yfp into the fosmid that contains the \textit{lsy-27} locus and that rescues the \textit{lsy-27} phenotype (Table 2), we generated a reporter with which we monitored \textit{lsy-27} expression (Figure 2A). We find that \textit{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 ASE/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 \textit{lsy-27} reporter with an ASE-specific mCherry reporter, we confirmed that \textit{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 \textit{lsy-27} paralog, \textit{ztf-25}, displays an essentially indistinguishable broad, embryo-restricted expression pattern (Figure S3), while the more distant paralog \textit{ztf-28} shows no expression in embryos and postembryonically is expressed only in the intestine (data not shown).

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

**Table 2 Transformation rescue and RNAi analysis**

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

\textsuperscript{a} Genotype as above but array not transmitted from parental generation 21.1 78.9 19

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

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

The \textit{ot108} and \textit{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.

* Scored as a \textit{goy-5} reporter (\textit{mt51} or \textit{ot1220}) derepressed in ASEL in first eleven rows or loss of \textit{lim-6::}gfp (\textit{ot114}) in remaining four rows.

1. The expression pattern of \textit{lsy-27} shows no expression in ASE/ASER, displays an essentially indistinguishable broad, embryo-restricted expression pattern (Figure S3), while the more distant paralog \textit{ztf-28} shows no expression in embryos and postembryonically is expressed only in the intestine (data not shown). The expression pattern of \textit{lsy-27} suggests an embryonic role for the gene. We sought to corroborate this notion by exploiting the observation that the \textit{ot108} allele is strongly temperature sensitive (Figure 3B). At 25°C, 87% of animals display a \textit{Lsy} phenotype while 12% do at 15°C. By altering \textit{lsy-27} gene activity at different stages through temperature shifts, we find that \textit{lsy-27} activity is required only during embryogenesis, but not during postembryonic stages (Figure S3).
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. 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.)
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 Shohei Mitani at Tokyo Women’s Medical University School of Medicine for the tm573, tm593, and tm630 alleles; Sumeet Sarin for help in recalcitrating 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


Sarin, S., C. Antonio, B. Tursun, and O. Hobert, 2009 The C. elegans Tailless/TLX transcription factor nhr-67 controls neuronal...


Communicating editor: D. I. Greenstein
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
Figure S1  Updated  lim-6 locus and the location of the o146 allele. Based on our own S’RACE results (using the Invitrogen GeneRacer Kit and the following lim-6 gene specific reverse primers: lim-6-5’RACEup: GATCAGTGGTCCGAGAG and lim-6-5’RACEupset: GTGCCAATCTACTATGTCGC) 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 o146 affects an invariant residue in the second LIM domain.
Figure S2  Sequence of LSY-27 and its two paralogs, ZTF-25 and ZTF-28. The C2H2 Zn fingers and the position of the *ot108* allele are indicated. The blue F and L residues can be found in the ~50% of all C2H2 Zn fingers (http://smart.embl-heidelberg.de/). The arrows indicate positions in the C2H2 Zn fingers thought to contact DNA (PAVLETICH and PAO 1991).
Figure S3  Expression pattern of the *lsy-27* paralog *ztf-25*. A: Gene and allele structure of *ztf-25*. A translational reporter was generating 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-5trans::gfp* expression pattern in midembryonic stages. Embryos are contained within the gonad of the mother.
Supporting Material and Methods

Genetic screen

The screen that uncovered *at146* and *at108*, as well as a large number of additional *lsy* 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:

\[ otls114 = ls[lim-6prom::gfp; rol-6(d)] \]
\[ otls3 = [gcy-7prom::gfp; lin-15(+)] \]
\[ ntls1 = [gcy-5prom::gfp; lin-15(+)] \]
\[ otls220 = ls[gcy-5prom::mCherry; rol-6(d)] \]
\[ otls151 = ls[ceh-36prom::DsRed; rol-6(d)] \]
\[ otls232 = ls[che-1prom::mChopti::che-1_3'UTR; rol-6(d)] \]

Rescue and expression constructs:

\[ otEx4280, otEx4281 = two independent lines of Ex[fosmid WRM067BG09; elt-2::gfp]. \]
\[ otEx4400, otEx4401 = two independent lines of Ex[lsy-27trans::gfp; elt-2::gfp] \]
\[ otEx4501, otEx4502 = two independent lines of Ex[fosmid WRM067BG09::yfp; elt-2::gfp] \]
\[ otEx4523 = Ex[fosmid WRM067BG09::yfp] \]
\[ otEx4337-4339 = three independent lines of Ex[ztf-25trans::gfp; rol-6(d)] \]
\[ otEx3859 = Ex[lim-6prom::yfp; rol-6(d)]. \]

Generation of expression constructs

*lsy-27trans::gfp* was generated by PCR-fusing the genomic locus of *lsy-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 (Hober 2002) and was coinjected with *elt-2::gfp* (50ng/μl) as injection marker. *ztf-25trans::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. *lsy-27trans::yfp* was generated by inserting *yfp* right before the stop codon of *lsy-27* in fosmid *WRM067BG09* (Tursun et al. 2009).

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

*lsy-27trans::gfp:

*lsy-27_translat_A*: CTGATACGAGTAGCGCATGGC
*lsy-27_translat_A*: GTACCGGCAAATGCATAC
*lsy-27_translat_B*: AGTCACTGAGAACAGCTTCATTCTACAGTGAGTCG
*lsy-27_translat_C*: AGCTTTGACATTCCAGGCG
*lsy-27_translat_D*: AAGGGGCCCAGTAGGCG
*lsy-27_translat_D*: AGGAAACAGTTAGTTGGTATA
**ztf-25**

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

We genotyped animals for the presence of the deletion alleles:

**Molecular characterization of the lsy-27**

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

Wild type: ...ttttgtaagccgctc [361 base pairs] ttgagaacttttaagctt...

**tm593**

...ttttgtaagccgctc T ttgagaacttttaagctt...

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:** CGTCTACAGTCGTCAC

**F47H4.1_RT_B:** GAATCATCTTGCTCTCGACG

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

Variants were considered as background if they were also found in other WGS datasets from our lab (Sarin et al. 2010).
<table>
<thead>
<tr>
<th>Genotype</th>
<th>% of animals expressing gcy-5 <em>(ntls1)</em> at 25°C</th>
<th></th>
<th></th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ASEL=ASE</td>
<td>ASEL&lt;ASE</td>
<td>ASER only</td>
<td></td>
</tr>
<tr>
<td>wild type</td>
<td>R</td>
<td>R</td>
<td>&gt;100</td>
<td>&gt;100</td>
</tr>
<tr>
<td><em>lsp-27</em>(ot108)</td>
<td>39%</td>
<td>48%</td>
<td>13%</td>
<td>122</td>
</tr>
<tr>
<td><em>lsp-27</em>(tm593)</td>
<td>0%</td>
<td>62%</td>
<td>38%</td>
<td>117</td>
</tr>
<tr>
<td><em>ztf-25</em>(tm610)</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>46</td>
</tr>
<tr>
<td><em>ztf-28</em>(tm573)</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>28</td>
</tr>
<tr>
<td><em>lsp-27</em>(tm593); <em>ztf-28</em>(tm573)</td>
<td>0%</td>
<td>49%</td>
<td>51%</td>
<td>168</td>
</tr>
<tr>
<td><em>ztf-25</em>(tm610); <em>ztf-28</em>(tm573)</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>78</td>
</tr>
<tr>
<td><em>ztf-25</em>(tm610); <em>ztf-28</em>(tm573); <em>lsp-27</em>(RNAi) ¹</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>32</td>
</tr>
</tbody>
</table>

¹ Triple couldn’t be built due to close linkage of *lsp-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

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

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 Hawaian 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

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

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


CHAPTER 5:

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


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 MIST-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 reproducibly 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, *by-12*, that executes such a maintenance function.

**Laterality defects in animals lacking by-12, a histone acetyltransferase:** *by-12* mutants were initially identified in a screen for ASEL/R laterality defects (*Lsy* phenotype) (Sarin et al. 2007). We characterized the *by-12* defects in more detail by examining a panel of left/right asymmetrically expressed genes. We find that *by-12* defects 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).

*by-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 *by-12* corresponds to the previously uncharacter-
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). Animals that express the die-1 reporter fosmid also contain an 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 EX1785669 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 T11A3.1), 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 recombinering 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.

**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; Dimiano 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 an 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 EX1785669 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 T11A3.1), 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 recombinering 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.
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 ASE.

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 ASE-l 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 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
**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

---

**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 (**che-36prom::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** autoregulates its own transcription (L. Cochella and O. Hobert, unpublished data), the HAT complex also impinges on **die-1** expression itself.
proteins have been shown to be involved in cell growth, apoptosis, and tumorgenesis (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 ASE marker is lost while there is ectopic gain of a terminal ASER marker in ASEL (Figure 3A). This phenotype is indistinguishable from lsy-12/HAT and lin-49/BRPF mutants. Again similar to 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 goy-5 misexpression (n = 50) is rescued to 0% misexpression (n = 28)].

We therefore named the T06A10.4 locus by-13. A reporter construct for the by-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 ASEL/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 ASEL/R lateralization. We note that by-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 (Stuhr et al. 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 ASER 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 ASEL-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 geyslow::gfp transgene otIs186, we observed an ASEL-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 ASEL 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 LSV-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 myt-2 (a paralog of by-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 (Grienberger et al. 2002; Laue et al. 2008), and the vertebrate MYST family member MOZ is required for the generation and maintenance of hemapoietic 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 Cochehla 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 (R01NS03996-05; R01NS050266-05). O.H. is an investigator of the Howard Hughes Medical Institute.
LITERATURE CITED


Communicating editor: B. J. Meyer
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
Supporting Material and Methods

Transgenes

Transgenes that label ASEL and ASER fates include ASEL markers otIs114 = Is[lim-6prom::gfp; rol-6(d)], otIs160 = Is[lsy-6prom::gfp; unc122prom::gfp], ASER markers include otIs186 = Is[lsy-6prom::gfp; lin-15(+)], otIs274 = [die-1::yfp recombineered fosmid; che-1::mCherry:che-1_3’UTR; rol-6(d)] (kindly provided by Baris Tursun), ntIs1 = [gcy-5prom::gfp; lin-15(+)], otIs220 = Is[gcy-5prom::mCherry; rol-6(d)]; ASEL/R markers otIs151 = Is[ceh-36prom::DsRed2; rol-6(d)], otIs232 = Is[che-1promA::mCherry::che-1_3’UTR; rol-6(d)], otIs186 = Is[gcy-5prom::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-36prom2::lsy-12a; elt-2::DsRed2]; otEx4317-21 = five independent lines of Ex[ceh-36prom2::lsy-12b; elt-2::DsRed2]; otEx4366-68 = three independent lines of Ex[neprom_lsy-12a; elt-2::gfp]; otEx4418 = Ex[lsy-13 4.4 kb genomic locus; elt-2::gfp]; otEx4417 = Ex[lsy-13prom::gfp::unc-543’UTR; elt-2::gfp].

Expression constructs

ceh-36prom2::lsy-12, nonprom_lsy-12a, and lsy-13prom::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-36prom2::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/μl) 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) and was used as an injection marker. ceh-36prom2::lsy-12b is the same construct injected without the R07B5.9 fragment.

ceh-36prom2::lsy-12a 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-543’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’):

R07B5.9 PCR rescue fragment:

R07B5.8RescueFwd – ggctcgcttcatttagac
R07B5.8RescueRev – ggtgcggattgatgtgagg
R07B5.9PCRRecFwd – ggctcgcttcatttagac
R07B5.9PCRRecRev – ggtgcggattgatgtgagg

R07B5.9 PCR rescue fragment:

R07B5.9RescueFwd – gtcacatccccgggattagac
R07B5.9RescueRev – ggctcgcttcatttagac

R07B5.9 PCR rescue fragment:

R07B5.9RescueFwd – gtcacatccccgggattagac
R07B5.9RescueRev – ggctcgcttcatttagac

cenh36p2_5A - ttaAGCTTATCCGATAAGGCTG
cenh36R07B50pB – cttgagaagggaacacatagGGATCcgcaaatgggcggagggtg
R07B5.8tnsIC - ctagttcctcggcgagtc
R07B5.8rescRevout (use as D) - gacgacatcgacctaggg
R07B5.8RescueRev (use as D*) - ggctcgcttcatttagac

coinject with R07B5.9 overlapping PCR rescue fragment
**ceh-36prom2::lsy-12b**

Same as **ceh-36prom2::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:
  - 7R07B5.8tnsIC: caatgtttctctctctaag
  - R07B5.8RescueRev: gggtggatgtgtgtgagg

**ly-13prom::gfp**

- T06A10.4_A: cacagtgaactttccccg
- T06A10.4_A*: gagatgagtggcgatgg
- T06A10.4_B: agtcgacctgcaggcatgcaagcttcactttcttcttcaatctcttctctctctctctct
  - C – agcttgcatgcctgcaggtcgact
  - D – aagggcccgtacggccgacta
  - D* – ggaaacagttatgtttggtatattgg

**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: gccgttgattgctccaattg
- lsy12RTfullp6: cctccaattccacctgcac
- R07B5.9end: gaactgattggtggcagttcc
- R07B5.9endNest: ggtggcagttccatttgtttg

**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.
**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 lys-12 transgenic animals**

<table>
<thead>
<tr>
<th>genotype</th>
<th>lim-6prom::gfp (otIs114) expression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ASEL only</td>
</tr>
<tr>
<td>wild type</td>
<td>100%</td>
</tr>
<tr>
<td>lys-12(ot170)</td>
<td>0%</td>
</tr>
</tbody>
</table>

**fosmid rescue**

<table>
<thead>
<tr>
<th>lys-12(ot170); otEx3510</th>
<th>100%</th>
<th>0%</th>
<th>0%</th>
<th>52</th>
</tr>
</thead>
<tbody>
<tr>
<td>lys-12(ot170); otEx3511</td>
<td>90%</td>
<td>0%</td>
<td>10%</td>
<td>30</td>
</tr>
<tr>
<td>lys-12(ot170); otEx3512</td>
<td>94%</td>
<td>0%</td>
<td>4%</td>
<td>26</td>
</tr>
</tbody>
</table>

**lsy-12a/b PCR rescue**

<table>
<thead>
<tr>
<th>lys-12(ot170); otEx3676</th>
<th>97%</th>
<th>3%</th>
<th>0%</th>
<th>33</th>
</tr>
</thead>
<tbody>
<tr>
<td>lys-12(ot170); otEx3677</td>
<td>78%</td>
<td>22%</td>
<td>0%</td>
<td>36</td>
</tr>
<tr>
<td>lys-12(ot170); otEx3678</td>
<td>84%</td>
<td>16%</td>
<td>0%</td>
<td>32</td>
</tr>
</tbody>
</table>

**ceh-36prom2::lsy-12a**

| lys-12(ot170); otEx4330 | 84%  | 10% | 6%  | 31 |

**ceh-36prom2::lsy-12b**

<table>
<thead>
<tr>
<th>lys-12(ot170); otEx4318</th>
<th>62%</th>
<th>0%</th>
<th>38%</th>
<th>26</th>
</tr>
</thead>
<tbody>
<tr>
<td>lys-12(ot170); otEx4319</td>
<td>36%</td>
<td>0%</td>
<td>64%</td>
<td>25</td>
</tr>
<tr>
<td>lys-12(ot170); otEx4320</td>
<td>18%</td>
<td>0%</td>
<td>82%</td>
<td>33</td>
</tr>
</tbody>
</table>

**no prom lsy-12a**

<table>
<thead>
<tr>
<th>lys-12(ot170); otEx4366</th>
<th>2.6%</th>
<th>0%</th>
<th>97.4%</th>
<th>39</th>
</tr>
</thead>
<tbody>
<tr>
<td>lys-12(ot170); otEx4367</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>21</td>
</tr>
<tr>
<td>lys-12(ot170); otEx4368</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>26</td>
</tr>
</tbody>
</table>

See Fig.1 for schematic representation of constructs. Note that animals that express the genomic lys-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 lys-12 locus under control of the bilaterally expressed ceh-36 promoter. Within the appropriate cellular context, lys-12 is therefore not only required but, at least to some extent, sufficient to induce ASEL fate. The potential link of lys-12 with die-1 and che-1 may explain the overexpression effect of lys-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 lys-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 lys-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 lys-12a and the lys-12b isoform can rescue the laterality defect, even though mutant alleles are restricted to the lys-12a isoform. Perhaps lys-12a and lys-12b are functionally equivalent, but lys-12a is the only isoform that is expressed in ASEL.
CHAPTER 6: DISCUSSION 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 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 collaborates 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 ots317 (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 *cnl-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 *cnl-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 *cnl-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 *cnl-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 *cnl-1* and *hlh-16* mutants.

If both *cnl-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 otl317 (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 pleitropies 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, *lsy-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 *lsy-13* homolog ING-5 (inhibitor of growth protein 5) (Morrow et al., 1999).
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:


APPENDIX 1:

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


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 Estanisla Daniel De La Cruz,1 and Oliver Hobert1,3

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.), 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
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 neuron 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 motoneurons 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, tyraminergic, 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
Conserved small motifs that constitute predicted binding sites fine-grained mutational dissection in which we mutated cis-stream of eat-4 glutamatergic neuron classes (Figure 2). The modularity of the elements that direct expression to very small numbers of specific cis-upstream mals. We find that the broad expression generated from the locus and examined their expression pattern in transgenic ani-

---

**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 adls1240 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 ≥ 10) were scored for expression and lines commonly showed very similar expression pattern and penetrence 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 (DiI staining or reporter lines). *** indicates: DIC identification, not confirmed with reporter. For a list of transgenic strains, see Supplemental Information. See also Figure S4.

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 up-stream 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 (adls1240) 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 theeat-4 locus and examined their expression pattern in transgenic an-
imals. 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 fac-
tors that have previously been shown to define the identity of distinct neuron types that we determine here to be glutamater-
gic: theche-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

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

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.

| Legend | |
|--------| |
| a | Based on White et al., 1986. |
| b | Bold 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). |
| c | Candidate 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 chemosensory neuron identity; [Cassata et al., 2000; Satterlee et al., 2001; H. Kagoshima, personal communication]), the lin-11 LIM homedomain 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, SHT 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 flip-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 5F). 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 Gz-encoding genes gap-11, gap-13, gpa-15 and the neuropeptide flip-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 flip-8, flip-10, and flip-11) and

---

*Apart 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.*

*Under hypoxic conditions (Pocock and Hobert, 2010).*

*Reported 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.*

*Inferred from requirement of TAAT homeodomain consensus site for eat-4 expression in LUA.*
Figure 3. Regulators of \textit{eat-4}/VGLUT Expression

\textit{eat-4} reporter lines where 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 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 \textit{eat-4} expression. See also Figure S2. (A) \textit{che-1} affects \textit{eat-4} expression (assayed with \textit{eat-4} \textit{prom} \textit{6} reporter transgene \textit{ot}Is392) in the ASE neurons. (B) \textit{ceh-36} affects \textit{eat-4} expression (assayed with \textit{eat-4} \textit{prom} \textit{6} reporter transgene \textit{ot}Is392) in the ASE and...
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 phosphate dsa-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 4I). Another DVC cell fate marker, the GPCR srb-12, also displays defective expression in DVC (Figure 5G). We observed no defects in eat-4 expression in the normally ceh-14 expressing PVO 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 homebox 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 linearly 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 pdfr-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 postmitotic 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, ttk-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 prom transgene otis392) in the BAG neurons. (D) ttk-1 affects eat-4 expression (assayed with eat-4 fosmid reporter transgene otis388) 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 adl240) 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 prom reporter transgene otis376). (G) ceh-6 controls eat-4 expression (assayed with eat-4 prom reporter transgene otis376). (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 prom reporter transgene etEx3301. (I) ceh-37 affects eat-4 expression (assayed with eat-4 prom reporter transgene otis392) in the BAG and ASG neurons. (J) lin-11 affects eat-4 expression (eat-4 prom transgene otis392) in the ASG neurons. (K) unc-86 affects eat-4 expression (eat-4 fosmid reporter transgene otis388) 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 zdIs13).

(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::gfp transgene otEx5336, gpa-13::gfp transgene ofEx213, gpa-15::gfp transgene pkIs591, flp-21::gfp transgene ynIs80).

(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 terminal markers of PVR identity (flip-10, flip-20, dkk-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 in the respective neuron types (Figure 3), 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-4prom5, 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-4prom6 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-4prom6 module: as mentioned above, deletion of one of the two sites abrogates ceh-36K50-homeodomain-dependent eat-4 expression in the ASE and AWC neurons, while a mutation in the other site also abrogates ttx-1K50-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 3), 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: AVC 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-4prom5 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–Q) ceh-14 controls several identity features of the PHB, PHC, and DVC neurons (ida-1::gfp transgene ins179, dop-1::gfp transgene vls628, sab-12::gfp transgene sEx12012).

(H) ceh-14 and unc-86 redundantly control the identity of the PVR neuron (flip-10::gfp transgene otIs92, flip-20::gfp transgene ynl54, dkk-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*:tagrfp 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 tx-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 *tx-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 *tx-3*, we examined four additional terminal identity features of ASK in *tx-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 *tx-3*.

Mutation of yet another predicted homeodomain binding site in the *eat-4* promotor 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 *tx-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 co-staining 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 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 feature 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
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 indicates 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 (Acampora 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

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<sup>flam</sup>;R26CreER animals compared to LHX1<sup>flam</sup> 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<sup>flam</sup>;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.
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 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 WRM0823af12 and an SL2-based, nuclear localized yfp reporter (Tursun et al., 2008). 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 pDP95.75-based expression vectors and mutagenized with the QuickChangeII 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::m38::gfp) were EMS-mutagenized and animals with loss of expression in OLL were isolated a Copas Biosort machine. All strains were scored as young adults.

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<sup>fl<sup>ox</sup></sup> (Kwan and Behringer, 2002) and ROSA26CreER (Badea et al., 2003). Tamoxifen was administrated 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 stains 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. Khrimian, M. Peterson, J. Hou, B. Wu and U. Aghaye 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 Piai 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


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, gln-1, gln-2 and gln-3. gln-1 and gln-2 reporter genes are expressed outside the nervous system and in very small (<5) number of neurons, while a gln-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 gln-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) (Mclntire 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 otsl388 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-4prom-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[516-861,stop 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’-GAACATCCTTGATTTCCTTCTTGCTCA –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 otsl292.

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:

otsl388 III: eat-4D::sl2::yfp::h2B; injected in pha-1(e2123); pbX
otsl292: eat-4 in vivo recombineered reporter; rol-6
adsl1230: eat-4 reporter construct from Lee et al., 1999.
otsl376: eat-4prom1::gfp; rol-6
otsl392: eat-4prom2::gfp; ttx-3::dsRed
otEx4478, otEx4479, otEx4480: 3 lines for eat-4prom1::gfp; rol-6
otEx4492, otEx4493, otEx4494: 3 lines for eat-4prom12::gfp; rol-6
otEx5292, otEx5293, otEx5294:: 3 lines for eat-4prom3::gfp; ttx-3::dsRed
otEx5298, otEx5299, otEx5300: 3 lines for eat-4prom61::gfp; ttx-3::dsRed
an average depth of 8x in an Illumina GA2 sequencer as previously described (Doitsidou et al., 2010). Analysis of the WGS data
Whole-genome sequencing was used to determine the molecular identity of
otEx5428[glna-1::NLS::gfp], sEx10131[glna-2::gfp], otEx5429[glna-3prom1::NLS::gfp],
[grd-8::gfp] vsIs28 [dop-1::gfp], sEx12012 [srab-12::gfp], zdIs13 [tph-1::gfp], otEx5336 [gpa-11prom2::gfp], pkIs589 (gpa-13::gfp), pkIs591 (gpa-
PBS with 5%

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
(AbcCam, 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).

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).

The following neuronal identity markers were used: otls92 [flip-10::gfp], insl179 [ida-1::gfp], ynl30 [flip-4::gfp], gmls12 [srb-6::gfp],
ots33 [kal-1::gfp], otls358 [ser-2prom2::gfp], gmls21 [nlp-1::gfp], ynl80 [flip-21::gfp], ynls2022 [flip-8::gfp], ynl34 [flip-11::gfp],
vsIs28 [dop-1::gfp], sEx12012 [srab-12::gfp], zdls13 [fph-1::gfp], otEx5336 [gpa-11prom::gfp], pkls589 [gpa-13::gfp], pkls591 [gpa-
15::gfp], otEx5323 (dkf-2b::gfp), ynl54 [flip-20::gfp], otls138 and otEx449 [ser-2prom3::gfp], otls396 [ace-1prom2::tagRFP], sEx15238
[grd-8::gfp], myEx741[pdr-1::NLS::rfp], vdEx078 [tol-1::gfp], otEx2540 [gcy-23::gfp], ynl37 [flip-13::gfp], rtEx247 [nlp-14::gfp],
otEx5428[glna-1::NLS::gfp], sEx10131[glna-2::gfp], otEx5429[glna-3prom1::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 otls396 [ace-1prom2::NLS::tagRFP; pha-1(+)].

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.

Neuronal Identity Markers
The following neuron identity markers were used: otEx5311, otEx5312; 2 lines for eat-4prom6::gfp; rol-6
otEx5295, otEx5296, otEx5297; 3 lines for eat-4prom7::gfp; tbx-3::dsRed
otEx4488, otEx449, otEx4490; 3 lines for eat-4prom8::tagRFP; rol-6
otEx5301, otEx5310; 2 lines for eat-4prom2::gfp; rol-6
otEx5098, otEx5099, otEx5100; 3 lines for eat-4prom27::gfp; rol-6
otEx5313, otEx5314, otEx5315; 3 lines for eat-4prom16; tbx-3::dsRed
otEx5316, otEx5318; 2 lines for eat-4prom2a11; rol-6
otEx5319, otEx5320; 3 lines for eat-4prom2a12, rol-6
otEx5345, otEx5346, otEx5347; 3 lines for eat-4prom2a6; rol-6
otEx5330, otEx5331, otEx5332; 3 lines for eat-4prom2a6; rol-6
otEx5333, otEx5334, otEx5335; 3 lines for eat-4prom2a17; rol-6

Isolation, Identification, and Characterization of vab-3
We utilized a transgenic strain that expresses a reporter gene for the tyramine receptor ser-2 (ots138 transgenic animals), which is
exclusively expressed in the OLL and PVD sensory neuron classes (Tsakalou 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 shows 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).
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/μl with rol-6 (su1006) and 100 ng/μl 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/μl with rol-6 (su1006) and 100 ng/μl OP50 gDNA as filling DNA.

The glna-3prom1 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/μl with rol-6 (su1006) and 100 ng/μl 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 μm 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


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 adf1240 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 ots376 array) in the ASH neurons. RNAi-sensitized nre-1 lin-15; ots376 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 expression. ceh-36, ceh-37 and ceh-1 were misexpressed with the pansensory promoters osm-6 and ifr-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 ots356 = rab-3::NLS::tagRFP) is unaffected in phasmid sensory neurons in 3/3 ceh-14 null mutant animals. (G) rab-3 expression (monitored with ots291 = rab-3::NLS::yfp) is unaffected in the AUA neurons of 6/6 ceh-6 null mutant animals. Identification of AUA was facilitated by Di staining (red), which labels a closely neighboring neuron. (H) rab-3 expression (monitored with ots291/rab-3::NLS::yfp; ots396/ace-1prom::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 37.6 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-4prom5. The picture shows a representative projection of the expression of eat-4prom5 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 ymls37) of the ASK neurons.

(B) lin-11 affects eat-4 expression (assayed with the otIs392 transgene) in the ASK neurons.
<table>
<thead>
<tr>
<th>Region</th>
<th>Lhx1 IF</th>
<th>Brn3a IF</th>
<th>Vglut2 ISH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purkinje Cells</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Medial Preoptic area</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Medial Habenula</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>strain name</th>
<th>allele</th>
<th>genetic trait</th>
<th>cog-1 allele</th>
<th>fozi-1 allele</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>OH8114</td>
<td>ot424</td>
<td>recessive</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8115</td>
<td>ot425</td>
<td>recessive</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8149</td>
<td>ot426</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8150</td>
<td>ot427</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8151</td>
<td>ot428</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8152</td>
<td>ot429</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8153</td>
<td>ot430</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8180</td>
<td>ot440</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8181</td>
<td>ot441</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8182</td>
<td>ot442</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8226</td>
<td>ot444</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8227</td>
<td>ot445</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8228</td>
<td>ot446</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8229</td>
<td>ot447</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8230</td>
<td>ot448</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8231</td>
<td>ot449</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8232</td>
<td>ot450</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8407</td>
<td>ot459</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8324</td>
<td>ot451</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8325</td>
<td>ot452</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8326</td>
<td>ot453</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8361</td>
<td>ot457</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8330</td>
<td>ot454</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8331</td>
<td>ot455</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8332</td>
<td>ot456</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8377</td>
<td>ot458</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8410</td>
<td>ot460</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8411</td>
<td>ot461</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8412</td>
<td>ot462</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8413</td>
<td>ot463</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8437</td>
<td>ot464</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8438</td>
<td>ot465</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8468</td>
<td>ot466</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8469</td>
<td>ot467</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8485</td>
<td>ot468</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8486</td>
<td>ot469</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8516</td>
<td>ot470</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8517</td>
<td>ot471</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8518</td>
<td>ot472</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8519</td>
<td>ot473</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8536</td>
<td>ot474</td>
<td>recessive, not on X</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8537</td>
<td>ot475</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>-------</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OH8573</td>
<td>ot483</td>
<td>recessive, not on X</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OH8574</td>
<td>ot484</td>
<td>recessive, not on X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OH8575</td>
<td>ot485</td>
<td>recessive, not on X</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OH8576</td>
<td>ot486</td>
<td>recessive, not on X</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OH8577</td>
<td>ot487</td>
<td>recessive, not on X</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OH8578</td>
<td>ot488</td>
<td>recessive, not on X</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OH8579</td>
<td>ot489</td>
<td>Semi-dominant</td>
<td>che-1 allele</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 \textit{fozi-1/+ (ot131)} or \textit{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 \textit{fozi-1} or \textit{cog-1}. 
Table 2: List of ASEL off *che-1* mutants retrieved from *otIs4; vsIs33* using the COPAS worm sorter.

<table>
<thead>
<tr>
<th>strain name</th>
<th>genotype</th>
<th>phenotype</th>
<th>n=</th>
<th>% off</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>OH8111</td>
<td><em>ot421; otIs4</em></td>
<td>ASEL off</td>
<td>56</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>OH8112</td>
<td><em>ot422; otIs4</em></td>
<td>ASEL/R off</td>
<td>67</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>OH8113</td>
<td><em>ot423; otIs4</em></td>
<td>ASEL/R off</td>
<td>54</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>OH8159</td>
<td><em>ot431; otIs4; vsIs33</em></td>
<td>ASEL off</td>
<td>22</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>OH8160</td>
<td><em>ot432; otIs4; vsIs33</em></td>
<td>ASEL off</td>
<td>21</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td>OH8161</td>
<td><em>ot433; otIs4; vsIs33</em></td>
<td>ASEL off</td>
<td>27</td>
<td>93</td>
<td></td>
</tr>
<tr>
<td>OH8174</td>
<td><em>ot435; otIs4; vsIs33</em></td>
<td>ASEL off</td>
<td></td>
<td></td>
<td>likely <em>che-1</em></td>
</tr>
<tr>
<td>OH8176</td>
<td><em>ot436; otIs4; vsIs33</em></td>
<td>ASEL off</td>
<td></td>
<td></td>
<td>likely <em>che-1</em></td>
</tr>
<tr>
<td>OH8177</td>
<td><em>ot437; otIs4; vsIs33</em></td>
<td>ASEL off</td>
<td></td>
<td></td>
<td>likely <em>che-1</em></td>
</tr>
<tr>
<td>OH8178</td>
<td><em>ot438; otIs4; vsIs33</em></td>
<td>ASEL off</td>
<td></td>
<td></td>
<td>did not grow</td>
</tr>
<tr>
<td>OH8179</td>
<td><em>ot439; otIs4; vsIs33</em></td>
<td>ASEL off</td>
<td></td>
<td></td>
<td>likely <em>che-1</em></td>
</tr>
</tbody>
</table>

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 *otls243*.

<table>
<thead>
<tr>
<th>Strain Name</th>
<th><em>Genotype</em></th>
<th>Chromosome</th>
<th><em>che-1</em> allele</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>OH8661</td>
<td><em>ot490; otls114</em></td>
<td>not X</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8662</td>
<td><em>ot491; otls114</em></td>
<td>not X</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8714</td>
<td><em>ot504; otls114; otls243</em></td>
<td>not X</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8715</td>
<td><em>ot505; otls114</em></td>
<td>not X</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8760</td>
<td><em>ot512; otls114; otls243</em></td>
<td>not X</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8761</td>
<td><em>ot513; otls114</em></td>
<td>not X</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8763</td>
<td><em>ot514; otls114; otls243</em></td>
<td>not X</td>
<td>+</td>
<td>Him?</td>
</tr>
<tr>
<td>OH8764</td>
<td><em>ot515; otls114</em></td>
<td>not X</td>
<td>?</td>
<td>did not grow</td>
</tr>
<tr>
<td>OH8781</td>
<td><em>ot516; otls114</em></td>
<td>not X</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8782</td>
<td><em>ot517; otls114</em></td>
<td>not X</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8783</td>
<td><em>ot518; otls114</em></td>
<td>not X</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8784</td>
<td><em>ot519; otls114</em></td>
<td>not X</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8799</td>
<td><em>ot521; otls114</em></td>
<td>not X</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8800</td>
<td><em>ot522; otls114</em></td>
<td>not X</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8801</td>
<td><em>ot523; otls114; otls243</em></td>
<td>not X</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8809</td>
<td><em>ot524; otls114</em></td>
<td>not X</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8815</td>
<td><em>ot525; otls243</em></td>
<td>not X</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8825</td>
<td><em>ot527; otls243</em></td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OH8826</td>
<td><em>ot528; otls243</em></td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OH8827</td>
<td><em>ot529; otls243</em></td>
<td>not X</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8828</td>
<td><em>ot530; otls243</em></td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OH8829</td>
<td><em>ot531; otls243</em></td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OH8830</td>
<td><em>ot532; otls243</em></td>
<td>not X</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8831</td>
<td><em>ot533; otls243</em></td>
<td>not X</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>OH8832</td>
<td><em>ot534; otls243</em></td>
<td>not X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OH8841</td>
<td><em>ot535; otls243</em></td>
<td>not X</td>
<td>+</td>
<td></td>
</tr>
</tbody>
</table>

**ENU treated**

| OH8842      | *ot536; otls243* | likely *che-1* | |
| OH8843      | *ot537; otls243* | likely *che-1* | |
| OH8844      | *ot538; otls243* | likely *che-1* | |
| OH8845      | *ot538; otls243* | likely *che-1* | |

**Notes:**
1. *otls243* = *che-1* _fosmid::venus gcy-7::gfp rgef-1::dsred2*
2. Number of genomes screened: 48,000
3. Some strains have been outcrossed (*otls114* 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:
**APPENDIX 3:**

Table 1: List of variants in *ot219* from CloudMap (WS235 genome release).

<table>
<thead>
<tr>
<th>Position</th>
<th>Reference</th>
<th>Change</th>
<th>Gene_name</th>
<th>Bio_type</th>
<th>Transcript_ID</th>
<th>old/new_AA</th>
<th>Old/New_codon</th>
<th>Codon_Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>12587247</td>
<td>C</td>
<td>T</td>
<td>E02A10.3</td>
<td>protein_coding</td>
<td>E02A10.3a</td>
<td>Q/*</td>
<td>Cag/Tag</td>
<td>56</td>
</tr>
<tr>
<td>12587247</td>
<td>C</td>
<td>T</td>
<td>E02A10.3</td>
<td>protein_coding</td>
<td>E02A10.3b</td>
<td>Q/*</td>
<td>Cag/Tag</td>
<td>60</td>
</tr>
<tr>
<td>12587247</td>
<td>C</td>
<td>T</td>
<td>E02A10.3</td>
<td>protein_coding</td>
<td>E02A10.3c</td>
<td>Q/*</td>
<td>Cag/Tag</td>
<td>56</td>
</tr>
<tr>
<td>12729626</td>
<td>C</td>
<td>T</td>
<td>R11D1.10</td>
<td>protein_coding</td>
<td>R11D1.10a.2</td>
<td>A/V</td>
<td>gCc/gTc</td>
<td>21</td>
</tr>
<tr>
<td>12729626</td>
<td>C</td>
<td>T</td>
<td>R11D1.10</td>
<td>protein_coding</td>
<td>R11D1.10a.1</td>
<td>A/V</td>
<td>gCc/gTc</td>
<td>21</td>
</tr>
</tbody>
</table>

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).

<table>
<thead>
<tr>
<th>Start</th>
<th>End</th>
<th>Size</th>
<th>Gene_ID</th>
<th>Gene_name</th>
<th>Bio_type</th>
<th>Transcript_ID</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>12576011</td>
<td>12576082</td>
<td>71</td>
<td>E02A10.8</td>
<td>E02A10.8</td>
<td>ncRNA</td>
<td>E02A10.8</td>
<td>UPSTREAM: 872 bases</td>
</tr>
<tr>
<td>12576011</td>
<td>12576082</td>
<td>71</td>
<td>E02A10.4</td>
<td>E02A10.4</td>
<td>protein_coding</td>
<td>E02A10.4</td>
<td>UPSTREAM: 8060 bases</td>
</tr>
<tr>
<td>12576011</td>
<td>12576082</td>
<td>71</td>
<td>E02A10.10</td>
<td>E02A10.10</td>
<td>ncRNA</td>
<td>E02A10.10</td>
<td>UPSTREAM: 2200 bases</td>
</tr>
<tr>
<td>12576011</td>
<td>12576082</td>
<td>71</td>
<td>E02A10.9</td>
<td>E02A10.9</td>
<td>ncRNA</td>
<td>E02A10.9</td>
<td>UPSTREAM: 2588 bases</td>
</tr>
<tr>
<td>12576011</td>
<td>12576082</td>
<td>71</td>
<td>E02A10.5</td>
<td>E02A10.5</td>
<td>protein_coding</td>
<td>E02A10.5</td>
<td>UPSTREAM: 3663 bases</td>
</tr>
<tr>
<td>12576011</td>
<td>12576082</td>
<td>71</td>
<td>E02A10.7</td>
<td>E02A10.7</td>
<td>ncRNA</td>
<td>E02A10.7</td>
<td>UPSTREAM: 7397 bases</td>
</tr>
<tr>
<td>12576011</td>
<td>12576082</td>
<td>71</td>
<td>E02A10.3</td>
<td>E02A10.3</td>
<td>protein_coding</td>
<td>E02A10.3b</td>
<td>UPSTREAM: 7825 bases</td>
</tr>
<tr>
<td>12576011</td>
<td>12576082</td>
<td>71</td>
<td>E02A10.3</td>
<td>E02A10.3</td>
<td>protein_coding</td>
<td>E02A10.3c</td>
<td>UPSTREAM: 8819 bases</td>
</tr>
<tr>
<td>12576011</td>
<td>12576082</td>
<td>71</td>
<td>E02A10.3</td>
<td>E02A10.3</td>
<td>protein_coding</td>
<td>E02A10.3a</td>
<td>UPSTREAM: 9088 bases</td>
</tr>
<tr>
<td>12576011</td>
<td>12576082</td>
<td>71</td>
<td>E02A10.2</td>
<td>grl-23</td>
<td>protein_coding</td>
<td>E02A10.2a</td>
<td>Exon_V_12575852_12576114</td>
</tr>
<tr>
<td>12576011</td>
<td>12576082</td>
<td>71</td>
<td>E02A10.2</td>
<td>grl-23</td>
<td>protein_coding</td>
<td>E02A10.2b</td>
<td>INTRON</td>
</tr>
<tr>
<td>12579753</td>
<td>12579771</td>
<td>18</td>
<td>E02A10.7</td>
<td>E02A10.7</td>
<td>ncRNA</td>
<td>E02A10.7</td>
<td>UPSTREAM: 3655 bases</td>
</tr>
<tr>
<td>12579753</td>
<td>12579771</td>
<td>18</td>
<td>E02A10.3</td>
<td>E02A10.3</td>
<td>protein_coding</td>
<td>E02A10.3b</td>
<td>UPSTREAM: 4083 bases</td>
</tr>
<tr>
<td>12579753</td>
<td>12579771</td>
<td>18</td>
<td>E02A10.3</td>
<td>E02A10.3</td>
<td>protein_coding</td>
<td>E02A10.3c</td>
<td>UPSTREAM: 5077 bases</td>
</tr>
<tr>
<td>12579753</td>
<td>12579771</td>
<td>18</td>
<td>E02A10.3</td>
<td>E02A10.3</td>
<td>protein_coding</td>
<td>E02A10.3a</td>
<td>UPSTREAM: 5346 bases</td>
</tr>
<tr>
<td>12579753</td>
<td>12579771</td>
<td>18</td>
<td>E02A10.10</td>
<td>E02A10.10</td>
<td>ncRNA</td>
<td>E02A10.10</td>
<td>DOWNSTREAM: 1405 bases</td>
</tr>
<tr>
<td>12579753</td>
<td>12579771</td>
<td>18</td>
<td>C14C10.1</td>
<td>C14C10.1</td>
<td>protein_coding</td>
<td>C14C10.1</td>
<td>UPSTREAM: 8675 bases</td>
</tr>
<tr>
<td>12579753</td>
<td>12579771</td>
<td>18</td>
<td>E02A10.9</td>
<td>E02A10.9</td>
<td>ncRNA</td>
<td>E02A10.9</td>
<td>DOWNSTREAM: 1009 bases</td>
</tr>
<tr>
<td>12579753</td>
<td>12579771</td>
<td>18</td>
<td>C14C10.7</td>
<td>tr-43</td>
<td>protein_coding</td>
<td>C14C10.7</td>
<td>DOWNSTREAM: 9959 bases</td>
</tr>
<tr>
<td>12579753</td>
<td>12579771</td>
<td>18</td>
<td>E02A10.6</td>
<td>E02A10.6</td>
<td>ncRNA</td>
<td>E02A10.6</td>
<td>DOWNSTREAM: 6775 bases</td>
</tr>
<tr>
<td>12579753</td>
<td>12579771</td>
<td>18</td>
<td>E02A10.2</td>
<td>grl-23</td>
<td>protein_coding</td>
<td>E02A10.2a</td>
<td>UPSTREAM: 2858 bases</td>
</tr>
<tr>
<td>12579753</td>
<td>12579771</td>
<td>18</td>
<td>E02A10.2</td>
<td>grl-23</td>
<td>protein_coding</td>
<td>E02A10.2b</td>
<td>UPSTREAM: 2858 bases</td>
</tr>
<tr>
<td>12579753</td>
<td>12579771</td>
<td>18</td>
<td>E02A10.8</td>
<td>E02A10.8</td>
<td>ncRNA</td>
<td>E02A10.8</td>
<td>DOWNSTREAM: 2656 bases</td>
</tr>
<tr>
<td>12579753</td>
<td>12579771</td>
<td>18</td>
<td>E02A10.5</td>
<td>E02A10.5</td>
<td>protein_coding</td>
<td>E02A10.5</td>
<td>Exon_V_12579674_12579979</td>
</tr>
<tr>
<td>12647810</td>
<td>12647882</td>
<td>72</td>
<td>F56H9.2</td>
<td>F56H9.2</td>
<td>protein_coding</td>
<td>F56H9.2b</td>
<td>DOWNSTREAM: 9573 bases</td>
</tr>
<tr>
<td>12647810</td>
<td>12647882</td>
<td>72</td>
<td>F56H9.2</td>
<td>F56H9.2</td>
<td>protein_coding</td>
<td>F56H9.2a</td>
<td>DOWNSTREAM: 9572 bases</td>
</tr>
<tr>
<td>12647810</td>
<td>12647882</td>
<td>72</td>
<td>W05B10.6</td>
<td>W05B10.6</td>
<td>protein_coding</td>
<td>W05B10.6</td>
<td>DOWNSTREAM: 9064 bases</td>
</tr>
<tr>
<td>12647810</td>
<td>12647882</td>
<td>72</td>
<td>F56H9.5</td>
<td>lin-25</td>
<td>protein_coding</td>
<td>F56H9.5</td>
<td>UPSTREAM: 2048 bases</td>
</tr>
<tr>
<td>12647810</td>
<td>12647882</td>
<td>72</td>
<td>F56H9.3</td>
<td>gpa-8</td>
<td>protein_coding</td>
<td>F56H9.3</td>
<td>UPSTREAM: 7340 bases</td>
</tr>
<tr>
<td>12647810</td>
<td>12647882</td>
<td>72</td>
<td>F56H9.8</td>
<td>F56H9.8</td>
<td>protein_coding</td>
<td>F56H9.8a</td>
<td>DOWNSTREAM: 1178 bases</td>
</tr>
<tr>
<td>Chromosome</td>
<td>Start</td>
<td>End</td>
<td>Strand</td>
<td>Type</td>
<td>Description</td>
<td>Length</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>-------</td>
<td>-----</td>
<td>--------</td>
<td>------------</td>
<td>--------------------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td>12647810</td>
<td>12647882</td>
<td>72</td>
<td>F56H9.8</td>
<td>protein_coding</td>
<td>F56H9.8b</td>
<td>DOWNSTREAM: 1178 bases</td>
<td></td>
</tr>
<tr>
<td>12647810</td>
<td>12647882</td>
<td>72</td>
<td>F56H9.4</td>
<td>protein_coding</td>
<td>F56H9.4a</td>
<td>INTRON</td>
<td></td>
</tr>
<tr>
<td>12647810</td>
<td>12647882</td>
<td>72</td>
<td>F56H9.4</td>
<td>protein_coding</td>
<td>F56H9.4b</td>
<td>UTR_5_PRIME: 526 bases from TSS</td>
<td></td>
</tr>
<tr>
<td>1286663</td>
<td>12866732</td>
<td>69</td>
<td>T07F10.3</td>
<td>protein_coding</td>
<td>T07F10.3</td>
<td>DOWNSTREAM: 1363 bases</td>
<td></td>
</tr>
<tr>
<td>1286663</td>
<td>12866732</td>
<td>69</td>
<td>T07F10.6</td>
<td>protein_coding</td>
<td>T07F10.6</td>
<td>INTRON</td>
<td></td>
</tr>
<tr>
<td>1286663</td>
<td>12866732</td>
<td>69</td>
<td>T07F10.1</td>
<td>protein_coding</td>
<td>T07F10.1a</td>
<td>DOWNSTREAM: 2830 bases</td>
<td></td>
</tr>
<tr>
<td>1286663</td>
<td>12866732</td>
<td>69</td>
<td>T07F10.1</td>
<td>protein_coding</td>
<td>T07F10.1b</td>
<td>DOWNSTREAM: 2830 bases</td>
<td></td>
</tr>
<tr>
<td>1286663</td>
<td>12866732</td>
<td>69</td>
<td>T07F10.12</td>
<td>ncRNA</td>
<td>T07F10.12</td>
<td>UPSTREAM: 8233 bases</td>
<td></td>
</tr>
<tr>
<td>1286663</td>
<td>12866732</td>
<td>69</td>
<td>T07F10.10</td>
<td>ncRNA</td>
<td>T07F10.10</td>
<td>DOWNSTREAM: 8235 bases</td>
<td></td>
</tr>
<tr>
<td>1286663</td>
<td>12866732</td>
<td>69</td>
<td>T07F10.5</td>
<td>protein_coding</td>
<td>T07F10.5.2</td>
<td>DOWNSTREAM: 116 bases</td>
<td></td>
</tr>
<tr>
<td>1286663</td>
<td>12866732</td>
<td>69</td>
<td>T07F10.5</td>
<td>protein_coding</td>
<td>T07F10.5.1</td>
<td>DOWNSTREAM: 107 bases</td>
<td></td>
</tr>
<tr>
<td>12975514</td>
<td>12975570</td>
<td>56</td>
<td>R186.7</td>
<td>R186.7</td>
<td>protein_coding</td>
<td>R186.7</td>
<td>DOWNSTREAM: 2231 bases</td>
</tr>
<tr>
<td>12975514</td>
<td>12975570</td>
<td>56</td>
<td>R186.14</td>
<td>R186.14</td>
<td>ncRNA</td>
<td>R186.14</td>
<td>DOWNSTREAM: 8085 bases</td>
</tr>
<tr>
<td>12975514</td>
<td>12975570</td>
<td>56</td>
<td>R186.10</td>
<td>R186.10</td>
<td>ncRNA</td>
<td>R186.10</td>
<td>DOWNSTREAM: 1205 bases</td>
</tr>
<tr>
<td>12975514</td>
<td>12975570</td>
<td>56</td>
<td>R186.13</td>
<td>R186.13</td>
<td>ncRNA</td>
<td>R186.13</td>
<td>UPSTREAM: 69 bases</td>
</tr>
<tr>
<td>12975514</td>
<td>12975570</td>
<td>56</td>
<td>R186.2</td>
<td>srd-35</td>
<td>protein_coding</td>
<td>R186.2a</td>
<td>DOWNSTREAM: 9492 bases</td>
</tr>
<tr>
<td>12975514</td>
<td>12975570</td>
<td>56</td>
<td>R186.2</td>
<td>srd-35</td>
<td>protein_coding</td>
<td>R186.2b</td>
<td>DOWNSTREAM: 9492 bases</td>
</tr>
<tr>
<td>12975514</td>
<td>12975570</td>
<td>56</td>
<td>R186.3</td>
<td>R186.3</td>
<td>protein_coding</td>
<td>R186.3.1</td>
<td>UPSTREAM: 8304 bases</td>
</tr>
<tr>
<td>12975514</td>
<td>12975570</td>
<td>56</td>
<td>R186.3</td>
<td>R186.3</td>
<td>protein_coding</td>
<td>R186.3.2</td>
<td>UPSTREAM: 8302 bases</td>
</tr>
<tr>
<td>12975514</td>
<td>12975570</td>
<td>56</td>
<td>R186.8</td>
<td>R186.8</td>
<td>protein_coding</td>
<td>R186.8.2</td>
<td>UPSTREAM: 7625 bases</td>
</tr>
<tr>
<td>12975514</td>
<td>12975570</td>
<td>56</td>
<td>R186.8</td>
<td>R186.8</td>
<td>protein_coding</td>
<td>R186.8.1</td>
<td>UPSTREAM: 7532 bases</td>
</tr>
<tr>
<td>12975514</td>
<td>12975570</td>
<td>56</td>
<td>R186.4</td>
<td>lin-46</td>
<td>protein_coding</td>
<td>R186.4</td>
<td>DOWNSTREAM: 4782 bases</td>
</tr>
<tr>
<td>12975514</td>
<td>12975570</td>
<td>56</td>
<td>R186.5</td>
<td>shw-3</td>
<td>protein_coding</td>
<td>R186.5</td>
<td>INTRON</td>
</tr>
</tbody>
</table>

182