THE UNIVERSITY OF CHICAGO

THE NON-CANONICAL ROLE OF HOX GENES IN THE REGULATION AND MAINTENANCE OF *C. ELEGANS* MOTOR NEURON TERMINAL IDENTITY

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

COMMITTEE ON DEVELOPMENT, REGENERATION AND STEM CELL BIOLOGY

BY

WEIDONG FENG

CHICAGO, ILLINOIS

AUGUST 2022

Table of Contents

List of Figures

List of Tables

Acknowledgements

It's been almost seven years since I was first here in UChicago in the summer of 2015. It's also been the most dramatic seven years of my life. Looking back at the way I've been through, I feel complicated but first of all full of gratitude. I came as a naïve undergraduate student and I will leave carrying with me full of knowledge and wonderful memories. These treasures will be a lifelong blessing no matter what. My academic and personal growth are beyong measure. None of them will happen without being around so many wonderful people. I know that I certainly cannot express my appreciation enough with the following pale words. But please allow me to conclude this part my life journey with some nice words.

First of all, I want to thank my mentor, friend and colleague Dr. Paschalis Kratsios, who generously provide hands-on academic training and valuable lessons in my graduate school. I've been learning tremendously from him. With due respect and sincere gratitude, I wish all the best for him, as well for my fantastic labmates and for the future development of Kratsios Lab. I grew with the Lab and I hold solid belief that this lab is definitely going to thrive and continue beautiful work along its academic journey! Thank you, Paschalis, for everything. Words are just not enough at all and I am going to say proudly that I am really happy and satisfied will never regret joinging this lab.

Second, I would like to express my gratitude to Dr. Ellie Heckscher for her amazing mentorship, inspiration, and great help during the past years. It's so great to have you on my committee. And I thank my academic advisors Dr. Chip Ferguson for his valuable academic advice and comments which significantly improve the quality of my research. It's wonderful to have you on my committee. I also thank Dr. Marcelo Nobrega for his mentorship and great advice on my

committee. I thank Dr. Robert Carrillo for mentoring me during my rotation and supporting me always. I thank Dr. Urs Schmidt-Ott, Dr. Horne-Badovinac, Dr. Ed Munro for being on my Prelim Committee and Ciriculum Committee with great advice and help. I want to say big thank you particularly to Dr. Wei-jen Tang, Dr. Ilaria Rebay and Dr. Rick Fehon for their mentorship and great help during my summer program in 2015. I will not be able to be in Uchicgao without any of you! Another special thank you for Dr. Chenjian Li as my LIFE mentor!

Third, my life becomes so great with the companion of friends. I want to thank everyone current in or once in Kratsios Lab for the incredible atmosphere and environment. I want to thanks my friends in DRSB community! I thank my buddies for getting together through the years: Jiacheng Zhang, Qinpu He, Yinan Li, Xiaolin Huang, Can Dong, Hanyu Li, Siyu Chen, Ruijie Han, Guangpu Li, Wanhao Chi, Xiao Sun, Yue Liu, Rui He, Meizi Liu, Zhongshi Wang, Jiawei Liu, Will Yee, Jiajie Xu, Yupu Wang, Wenchao Liu, Anqi Yu and many others. I also want to thank my friends met at the Windmill Drama Club.

I also thank the entire DRSB community and Neuro community especially Sue Levison, Elena Rizzo and Laine-Nazaire for the great help. Thank you MGCB and the DevNeuro group as well. Special thank you to my best friend Dawei Wang for stepping into my life that means so much!

Last, I THANK my family for always on my side and staying together with me! Thank you Qian Liu my life partner, my soulmate, my exclusive life witness, and my wife! Our journey just begins with the best part. Thank you, my parents, grandparents, uncles, aunts and cousions. You are the cornerstone of my life.

Time goes and the journey continues. For everyone we crossed, are crossing, or will cross our trajectories, I wish it to be worthy and meaningful.

Abstract

Hox transcription factors play fundamental roles during early patterning, but they are also expressed continuously, from embryonic stages through adulthood, in the nervous system. However, the functional significance of their sustained expression remains unclear. In *C. elegans* motor neurons, we identified a non-canonical role of Hox genes in the regulation and maintenance of ventral nerve cord motor neuron terminal identity. First, we are among the first groups to establish the MN expression profile of Hox genes with single cell resolution. Second, we demonstrated the functions of Hox proteins in regulating neurotransmitter pathway genes and a broad range of motor neuron terminal identity genes. LIN-39 (Scr/Dfd/Hox4-5) is continuously required during post-embryonic life to maintain neurotransmitter identity, a core element of neuronal function. LIN-39 acts directly to co-regulate genes that define cholinergic identity (e.g., *unc-17/VAChT, cho-1/ChT*). We further show that LIN-39, MAB-5 (Antp/Hox6-8) and the transcription factor UNC-3 (Collier/Ebf) operate in a positive feedforward loop to ensure continuous and robust expression of cholinergic identity genes. Third, we identified a two-component design principle for homeostatic control of Hox gene expression in adult MNs (Hox transcriptional autoregulation is counterbalanced by negative UNC-3 feedback). Last, we described a Hox transcriptional switch prevented by UNC-3 to safeguard cholinergic motor neuron identity throughout life. These findings not only highlighted the noncanonical role but also established a terminal selector role for Hox proteins during post-embryonic life, critically broadening their functional repertoire from early patterning to the control of motor neuron terminal identity.

Chapter 1 Introduction and Background

1.1 Historical perspective on Hox genes

“The essential phenomenon is not that there has merely been a change, but that something has been changed into the likeliness of something else”, William Bateson in 1894, coined the term "homeosis" in his book, *Materials for the Study of Variation*. This was the first identification of homeotic transformation, after which Hox genes are named. But at that time, homeotic transformation (or “homeosis”) was basically descriptive and was not explained from a genetic point of view. In the next century, after the rediscovery of Mendel's work, evidence for a genetic basis of homeotic transformations was obtained by isolating homeotic mutants exemplified by the work of Calvin Bridges and Thomas Morgan. Following that, Hox genes complexes in Drosophila were identified and characterized by Thomas Kaufman and Edward Lewis, who together with Christiane Nüsslein-Volhard and Eric Wieschaus, were awarded the Nobel Prize in Physiology or Medicine in 1995 for their discoveries concerning the genetic control of early embryonic development. At the molecular level, the homeobox (see below) was discovered in 1983 by Ernst Hafen, Michael Levine, and William McGinnis in Walter Gehring's lab at the University of Basel, and Matthew P. Scott and Amy Weiner in Thomas Kaufman's lab at Indiana University in Bloomington. The knowledge and understandings on Hox genes grew dramatically following these seminal studies, making Hox one of the earliest characterized and famous gene families.

Mutations of Hox genes often lead to dramatic homeotic transformation phenotypes along the anterior-posterior axis (A-P axis), Hox genes constitute a subfamily of homeobox genes, which are characterized by the presence of a 180bp-long DNA sequence (homeobox) that encodes a

conserved 60-amino-acid-long DNA binding domain (homeodomain). Homeodomain exists in most eukaryotes and homeodomain proteins usually exert crucial developmental functions and are classified into multiple types such as LIM, POU, and TALE proteins. Hox proteins, a subfamily of homeodomain proteins, appeared firstly in the common ancestor of cnidarian and bilaterian and they function as transcription factors specifying the positional information in body plan formation (Carrol 1995; Ryan et al.,2007).

1.2 Functions of Hox genes during early stages of Development

The function of Hox proteins are well studied during early developmental stages (embryogenesis) and are highly conserved in bilateria (Rosa et al., 1999). They provide positional information essential for body plan formation by acting as decision makers rather than actually constructing body parts themselves. The target genes of Hox proteins promote cell division, cell adhesion, apoptosis, and cell migration and they are the executors fulfilling body plan formation and embryonic patterning (Krumlauf 1994; Pearson et al., 2005; Vachon et al., 1992; Capovilla and Botas 1998; Lohmann et al., 2002; Bromleigh and Freedman 2000). At the molecular level, Hox proteins recognize DNA sequences containing TAAT, similar to the rest of homeodomain proteins, but the flanking base pairs distinguish the binding behaviors of different homeodomain proteins (Hanes and Brent 1989, 1991). Hox proteins can bind as monomers or homodimers, however, to increase binding affinity and specificity, Hox proteins usually act in partnership as protein complexes with TALE family of cofactors such as Pbx and Meis proteins (Mann et al., 2009; Merabet and Galliot 2015).

In bilateria, Hox genes are often arranged in chromosomal clusters. The order of Hox genes within clusters corresponds with the temporal as well as the spatial order of their expression along the A-P axis during development (Pearson et al., 2005; Gaunt 2018). For example, the Hox gene at the 3' end of the cluster will usually express the earliest in the most anterior domain; while the more 5' Hox genes will express later and in more posterior domains. (**Figure 1.1**) This feature is termed temporal and spatial collinearity. In addition, 5′ Hox genes usually possess a dominant phenotype compared with those more 3′, which is termed posterior prevalence (or posterior dominance). Posterior prevalence and temporal/spatial collinearity are inter-connected reflecting the comprehensive regulatory mechanisms of Hox gene expression *(*Durston 2012; Limura and Pourquie 2007; Singh and Mishra 2008*)*.

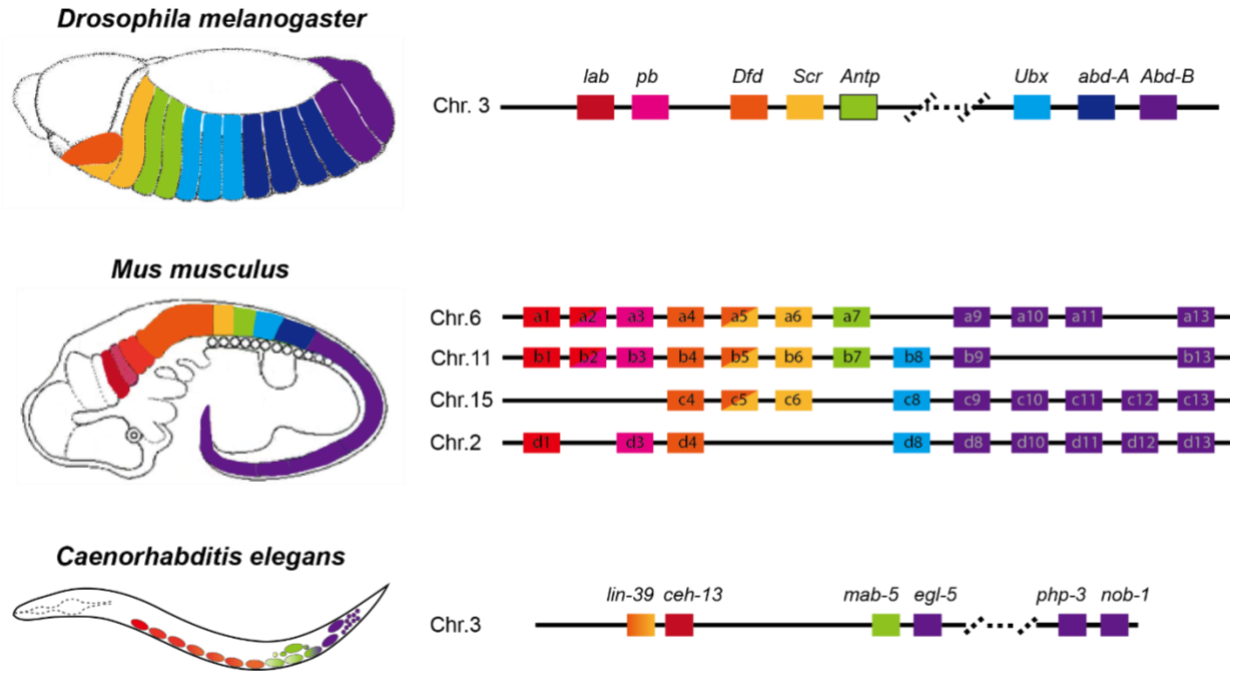

Figure 1.1 Genomic organization and colinear expression patterns of Hox genes in flies, mice and worms. Hox genes of the three organisms are color coded for their anteroposterior expression patterns, their genomic structures. and their possible phylogenic relationships. Partial overlap between *Hox gene expression is* not represented; Fly Hox gene abbreviations are: *lab, labial; pb, proboscipedia; Dfd, Deformed; Scr, Sex combs reduced; Antp, Antennapedia; Ubx, Ultrabithorax; abd-A, abdominal-A; Abd-B, Abdominal-B.* For mice, *Hoxa1* is shortened as *a1* indicated in boxes.

1.3 Regulation of Hox genes

Appropriate and tightly controlled regulation of Hox gene expression is crucial for body patterning. Fine-tuned regulatory mechanisms have been identified at multiple levels (transcriptional, post-transcriptional and even post-translational) in different organisms. In general, morphogenic fields regulate the proper initiation of Hox gene expression. For example, *Drosophila* maternal mRNA induces the expression of gap genes and pair-rule genes, whose products form a gradient and establish a morphogenic field. This morphogenic field initiates proper Hox gene expression during embryogenesis (Small et al., 1992). In vertebrates, Retinoic Acid (RA) and Fibroblastic Growth Factor (FGF) gradients form a morphogenic field and regulate differential Hox genes expression along A-P axis following 3'-5'/A-P temporal and spatial collinearity (Deuster 2008).

After the initiation of Hox gene expression, function domains of Hox proteins need to be fine-tuned and maintained during development. Multiple levels of regulation are involved: Positive and negative autoregulation has been reported for multiple Hox genes and potential Hox binding sites necessary for autoregulation have been identified *(*Delker et al., 2019; Packer et al., 1998; Crews and Pearson 2009*)*. The transcription of Hox genes can also be upregulated by trithorax group (TrxG) proteins or downregulated by the polycomb group (PcG) proteins, such as polycomb repressor complex 2 (PRC2) via histone methylation. Non-coding RNAs such as hox transcript antisense intergenic RNA (HOTAIR) might recruit PRC2 to Hox loci (Rinn et al., 2007). MicroRNAs located within Hox clusters have also been shown to inhibit more anterior Hox genes post-transcriptionally to fine-tune Hox protein levels (Lempradl and Ringrose 2008).

1.4 Functions of Hox genes in the nervous system

The roles of Hox genes during early development have been intensively investigated for decades, while much less is known about their role in later developmental stages. Accumulating evidence suggests that Hox genes are continuously expressed after embryogenesis and sustained to be expressed into adulthood *(*Allen et al., 2020; Coughlan et al., 2019; Hutlet et al., 2016; Kratsios et al., 2017; Lizen et al., 2017; Takahashi et al., 2004; Zheng et al., 2022*)*. Especially in the nervous system, Hox genes have enriched and extended expression, and affect multiple aspects of neuronal development, from the specification and survival of neural progenitors, to neuronal migration and axon guidance, as well as to the terminal differentiation of post-mitotic neurons (Di Bonito et al., 2013; Philippidou and Dasen, 2013; Parker and Krumlauf, 2020). It is intriguing to study if this conserved family of transcription factors also exerts important roles in later developmental and adult stages. To this end, my graduate thesis work focused on motor neurons (MNs) in *Caenorhabditis elegans* (*C. elegans*) ventral nerve cord (VNC).

1.5 Studying the function of Hox genes in *C. elegans* motor neurons

There are strong reasons for choosing *C. elegans* VNC MNs as our research model: First, *C. elegans* is a semi-transparent nematode about 1 mm in length with manageable culturing conditions as well as a fast life cycle of 3 days. Since 1963, Sydney Brenner proposed and performed research into the genetics of *C. elegans*, making it an extensively used model organism today. *C. elegans* was the first multicellular organism to have its whole genome sequenced, and was the first organism to have its connectome completed. In addition, the *C. elegans* community has generated a wealth of transgenic reporter strains for live imaging that are

wildly shared around the world. *C. elegans* is easily amenable to transgenesis and other genetic manipulations such as CRISPR/Cas9 gene editing, RNAi, and gene knock-in/out. Powerful large-scale forward and reverse genetic screens also accelerated genetic discoveries in *C. elegans*. Lastly, state-of-art techniques such as ChIP-seq, RNA-seq, ATAC-seq, and sm RNA FISH are well-established in *C. elegans,* making it a wonderful model organism to dive deeper into the investigation of developmental mechanisms.

C. elegans serves as a good model to study the role of Hox genes during later stages of MN development. First, there are only six Hox genes in C. elegans: one anterior Hox gene *ceh-13* (homolog of *Lab/Hox*1), two mid-body Hox genes *lin-39* (homolog of *Scr/Dfd/Hox4-5*) and *mab-5* (homolog of *Antp/Hox6-8*), and three posterior Hox genes *egl-5, nob-1* and *php-3* (all homologs of *AbdB/Hox9-13*). In comparison with the total number of 39 Hox genes in mammalian systems which are grouped into 4 tight clusters with greatly redundant functions, *C. elegans* Hox genes bear minimal redundancy. (**Figure 1.1**) In addition, only 4 out of 6 Hox genes are expressed in MNs (*ceh-13, lin-39, mab-5, and egl-5*) from the anterior to the posterior end of the VNC (**Figure 1.2-A**). Except for *ceh-13,* MNs are normally generated upon Hox gene mutations even in *lin-39, mab-5,* and *egl-5* triple mutants (Kenyon et al., 1997; Josephson et al., 2016*)*, unlike the case in *Drosophila* or mammalian systems where loss of Hox genes usually leads to cell death and early embryonic arrest (Alonso 2002; Bello et al., 2003; Arya et al., 2015; Lohmann et al., 2002*)*. This, together with the fact that VNC MNs are generated in two waves of development, during and after embryogenesis, makes *C. elegans* VNC MNs a great model to study the later role of Hox genes in post-mitotic neurons from development into adulthood.

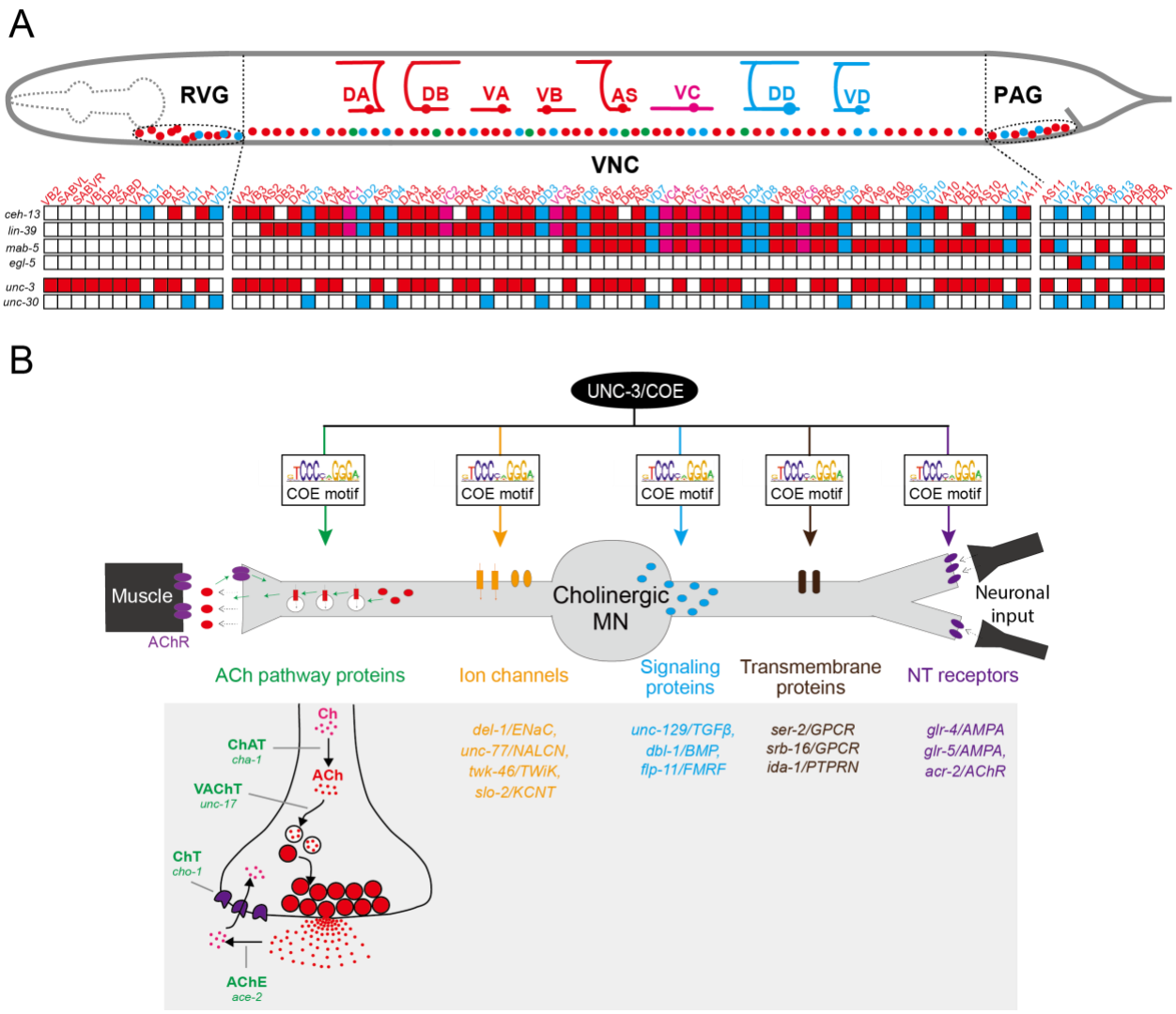

Figure 1.2 MN expression profile of *C. elegans* Hox genes, and schematic on the function of terminal selector UNC-3. (**A**) MN expression profile of *C. elegans Hox* gene *ceh-13, lin-39, mab-5, egl-5* and terminal selector gene *unc-3, unc-30* with single cell resolution. (From top row to bottom row) Each column represents individual MNs in the VNC with identity annotated above. MN subtypes are color coded and illustrated for their general morphology. Red (and pink which represents VC MNs only found in hermaphrodites) indicates cholinergic MNs while blue indicates GABAergic MNs. Empty box indicates no gene expression observed in this MN while different color-filled box indicates corresponding gene expression in specific MN subtypes, respectively. (**B**) Schematic demonstrating the terminal selector role of UNC-3 in cholinergic MNs. Cholinergic MNs receive neuronal input from right and synapse onto muscle targets on the left. Multiple groups of terminal identity genes color coded (in grey background) regulate MN functions. For example, genes involved in Acetylcholine (ACh) pathway *cha-1/ChAT* (Choline Acetyltransferase), *unc-17/VAChT* (Vesicular Acetylcholine Transporter), *ace-2/AChE* (Acetylcholinesterase), and *cho-1/ChT* (Choline Transporter) are illustrated on bottom left. UNC-3 binds to these genes directly through a conserved COE motif.

Based on morphology and connectivity, the VNC MNs in hermaphrodite *C. elegans* are categorized into seven subtypes (DA, DB, DD, VA, VB, AS, VC) *(*Von Stetina et al., 2006*)*. Among those, D-type MNs (DA, DB and DD) are born embryonically while the rest of them are generated at the first larval stage (L1) after embryogenesis. MNs can also be classified based on their neurotransmitter (NT) identity with DA, DB, VA, VB, AS and VC subtypes being cholinergic - they use Acetylcholine (ACh) as their NT. The DD and VD subtypes are GABAergic and use GABA. VC MNs are only found in hermaphrodites (They control egg-laying), but the rest of MNs subtypes are sex-shared (found in hermaphrodites and males) and control locomotion. Within each subtype, individual MNs intermingle along the VNC in an anteroposterior fashion between the retrovesicular ganglion (RVG) and the pre-anal ganglion (PAG). (**Figure 1.2-A**) For example, there are 9 DA MNs in total with 6 of them (DA2 to DA7) located within the VNC. DA2 is the most anterior one, and DA7 is the most posterior one in the VNC. Cell bodies of DA2-DA7 spread along the entire VNC intermingle with other MN subtypes. The linear distribution of MNs overlaps with the anteroposterior linear expression patterns of Hox genes in the VNC, which prompted us to study the role of Hox genes in VNC MN development.

1.6 The concept of terminal selectors

During the course of MN development, neuronal progenitors will initially be specified and then acquire distinct molecular, morphological and positional features. Then, MNs will gradually mature to obtain and later maintain their function-defining features *(*Catela and Kratsios 2019*)*. These features (e.g. NT synthesis enzymes and packaging proteins, neuropeptides, ion channels,

receptors, synapse proteins) need to be properly established and maintained throughout life to safeguard the unique **terminal identity** of individual MNs. In contrast to the well-studied early steps of MN development, the regulatory mechanisms that control MN terminal identity is less studied. Work in *C. elegans* and other systems suggest a common theme of co-regulatory logic of terminal identity by terminal selectors (and their combinations). Terminal selectors are crucial transcription factors (TFs) that co-regulate groups of effector genes defining identity features. Terminal selectors can function in combinations to initiate as well as maintain the expression of a large group of terminal identity genes *(*Hobert 2008, 2016; Hobert and Kratsios 2019*)*.

Two terminal selectors UNC-3/EBF and UNC-30/PITX are well studied regulating the terminal identity of VNC cholinergic and GABAergic MNs, respectively (Kratisos et al., 2012, 2015, 2017; Jin et al., 1994; Eastman et al., 1999; Cinar et al., 2005; Westmoreland et al., 2001*)*. In the VNC cholinergic MNs, UNC-3, the sole ortholog of the Collier/Olf/Ebf (COE) family of TFs, co-regulates nearly all terminal identity genes for all cholinergic MN subtypes (DA, DB, VA, VB, AS). (**Figure 1.2-B**) It functions through directly binding to conserved COE motifs in the cis-regulatory region of downstream terminal identity genes. These genes, such as *cho-1/ChT, unc-17/VAChT, ace-2/AChE* in the ACh pathway, *unc-129/TGFβ, acr-2,5,16/AChR*, and ion channels like *del-1/ENaC* and *slo-2/K-channel*, comprehensively shape the cholinergic identity of MNs. UNC-3 is necessary to establish and maintain in adulthood cholinergic MN terminal identities. Upon *unc-3* depletion, the expression of nearly all its downstream target genes is compromised and MNs lose their cholinergic identity. For GABAergic MNs, UNC-30 is the terminal selector to establish and maintain GABAergic terminal identities.

UNC-3 and UNC-30 are cell-type-specific TFs while Hox genes are region-specific TFs. They all continuously express in MNs and collaborate to regulate MN terminal identity genes. In this

dissertation, I will firstly review the emerging roles for Hox proteins in the last steps of neuronal development in worms, flies, and mice (**Chapter 2**) to provide some context for the new role of Hox genes I discovered in later stages of neuronal development. Then, focusing on *C. elegans* cholinergic MNs, I will demonstrate the maintenance of neurotransmitter identity by Hox proteins interacting with terminal selector UNC-3 through a homeostatic mechanism (**Chapter 3**). Lastly, I will discuss a case of MN terminal identity shift via a Hox transcriptional switch that is prevented by the terminal selector UNC-3 as means to safeguard cholinergic MN terminal identity (**Chapter 4**). Combining this evidence with unpublished data, I will discuss the role and the mechanisms of Hox genes in regulating MN terminal identities. My doctoral research provides the first evidence that Hox genes function as terminal selectors, offering new insights into how we view the function of these highly conserved TFs during later stages of neuronal development (**Chapter 5**).

1.7 References

Carroll, Sean B. "Homeotic genes and the evolution of arthropods and chordates." Nature 376.6540 (1995): 479-485.

Pearson, Joseph C., Derek Lemons, and William McGinnis. "Modulating Hox gene functions during animal body patterning." Nature Reviews Genetics 6.12 (2005): 893-904.

Ryan, J. F., et al. "Pre-Bilaterian Origins of the Hox Cluster and the Hox Code: Evidence from." (2007).

Krumlauf, Robb. "Hox genes in vertebrate development." Cell 78.2 (1994): 191-201.

Gaunt, Stephen J. "Hox cluster genes and collinearities throughout the tree of animal life." International Journal of Developmental Biology 62.11-12 (2018): 673-683.

De Rosa, Renaud, et al. "Hox genes in brachiopods and priapulids and protostome evolution." Nature 399.6738 (1999): 772-776.

Merabet, Samir, and Brigitte Galliot. "The TALE face of Hox proteins in animal evolution." Frontiers in genetics 6 (2015): 267.

Vachon, Gilles, et al. "Homeotic genes of the Bithorax complex repress limb development in the abdomen of the Drosophila embryo through the target gene Distal-less." Cell 71.3 (1992): 437-450.

Capovilla, Maria, and Juan Botas. "Functional dominance among Hox genes: repression dominates activation in the regulation of Dpp." Development 125.24 (1998): 4949-4957.

Lohmann, Ingrid, et al. "The Drosophila Hox gene deformed sculpts head morphology via direct regulation of the apoptosis activator reaper." *Cell* 110.4 (2002): 457-466.

Bromleigh, V. Carrie, and Leonard P. Freedman. "p21 is a transcriptional target of HOXA10 in differentiating myelomonocytic cells." *Genes & development* 14.20 (2000): 2581-2586.

Hanes, Steven D., and Roger Brent. "DNA specificity of the bicoid activator protein is determined by homeodomain recognition helix residue 9." *Cell* 57.7 (1989): 1275-1283.

Hanes, Steven D., and Roger Brent. "A genetic model for interaction of the homeodomain recognition helix with DNA." *Science* 251.4992 (1991): 426-430.

Mann, Richard S., Katherine M. Lelli, and Rohit Joshi. "Hox specificity: unique roles for cofactors and collaborators." *Current topics in developmental biology* 88 (2009): 63-101.

Small, Stephen, Adrienne Blair, and Michael Levine. "Regulation of even-skipped stripe 2 in the Drosophila embryo." *The EMBO journal* 11.11 (1992): 4047-4057.

Rinn, John L., et al. "Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs." *cell* 129.7 (2007): 1311-1323.

Lempradl, Adelheid, and Leonie Ringrose. "How does noncoding transcription regulate Hox genes?." *Bioessays* 30.2 (2008): 110-121.

Duester, Gregg. "Retinoic acid synthesis and signaling during early organogenesis." *Cell* 134.6 (2008): 921-931.

Durston, A. J. "Global posterior prevalence is unique to vertebrates: A dance to the music of time?." *Developmental Dynamics* 241.11 (2012): 1799-1807.

Iimura, Tadahiro, and Olivier Pourquié. "Hox genes in time and space during vertebrate body formation." *Development, growth & differentiation* 49.4 (2007): 265-275.

Singh, Narendra Pratap, and Rakesh K. Mishra. "A double-edged sword to force posterior dominance of Hox genes." *Bioessays* 30.11-12 (2008): 1058-1061.

Crews, Stephen T., and Joseph C. Pearson. "Transcriptional autoregulation in development." *Current Biology* 19.6 (2009): R241-R246.

Packer, Alan I., et al. "Expression of the murine Hoxa4 gene requires both autoregulation and a conserved retinoic acid response element." *Development* 125.11 (1998): 1991-1998.

Delker, Rebecca K., et al. "Low affinity binding sites in an activating CRM mediate negative autoregulation of the Drosophila Hox gene Ultrabithorax." *PLoS genetics* 15.10 (2019): e1008444.

Allen, A.M., et al., *A single-cell transcriptomic atlas of the adult Drosophila ventral nerve cord.* Elife, 2020. 9.

Coughlan, E., et al., *A Hox Code Defines Spinocerebellar Neuron Subtype Regionalization.* Cell Rep, 2019. 29(8): p. 2408-2421 e4.

Hutlet, B., et al., *Systematic expression analysis of Hox genes at adulthood reveals novel patterns in the central nervous system.* Brain Struct Funct, 2016. 221(3): p. 1223-43.

Kratsios, P., et al., *An intersectional gene regulatory strategy defines subclass diversity of C. elegans motor neurons.* Elife, 2017. 6.

Lizen, B., et al., *HOXA5 localization in postnatal and adult mouse brain is suggestive of regulatory roles in postmitotic neurons.* J Comp Neurol, 2017. 525(5): p. 1155-1175.

Takahashi, Y., et al., *Expression profiles of 39 HOX genes in normal human adult organs and anaplastic thyroid cancer cell lines by quantitative real-time RT-PCR system.* Exp Cell Res, 2004. 293(1): p. 144-53.

Zheng, C., H.M.T. Lee, and K. Pham, *Nervous system-wide analysis of Hox regulation of terminal neuronal fate specification in Caenorhabditis elegans.* PLoS Genet, 2022. 18(2): p. e1010092.

Di Bonito, M., Glover, J. C., and Studer, M. (2013). Hox genes and region-specific sensorimotor circuit formation in the hindbrain and spinal cord. Dev. Dyn. 242, 1348–1368. doi: 10.1002/dvdy.24055

Philippidou, P., and Dasen, J. S. (2013). Hox genes: choreographers in neural development, architects of circuit organization. Neuron 80, 12–34. doi: 10.1016/ j. neuron.2013.09.020

Parker, H. J., and Krumlauf, R. (2020). A Hox gene regulatory network for hindbrain segmentation. Curr. Top. Dev. Biol. 139, 169–203. doi: 10.1016/bs. ctdb.2020.03.001

Josephson, Matthew P., Adam M. Miltner, and Erik A. Lundquist. "Nonautonomous roles of MAB-5/Hox and the secreted basement membrane molecule SPON-1/F-Spondin in Caenorhabditis elegans neuronal migration." *Genetics* 203.4 (2016): 1747-1762.

Kenyon, C. J., et al. "The dance of the Hox genes: patterning the anteroposterior body axis of Caenorhabditis elegans." *Cold Spring Harbor Symposia on Quantitative Biology*. Vol. 62. Cold Spring Harbor Laboratory Press, 1997.

Alonso, Claudio R. "Hox proteins: sculpting body parts by activating localized cell death." *Current Biology* 12.22 (2002): R776-R778.

Bello, Bruno C., Frank Hirth, and Alex P. Gould. "A pulse of the Drosophila Hox protein Abdominal-A schedules the end of neural proliferation via neuroblast apoptosis." *Neuron* 37.2 (2003): 209-219.

Arya, R., et al. "Neural stem cell progeny regulate stem cell death in a Notch and Hox dependent manner." *Cell Death & Differentiation* 22.8 (2015): 1378-1387.

Lohmann, Ingrid, et al. "The Drosophila Hox gene deformed sculpts head morphology via direct regulation of the apoptosis activator reaper." *Cell* 110.4 (2002): 457-466.

Von Stetina, S.E., M. Treinin, and D.M. Miller, 3rd, The motor circuit. Int Rev Neurobiol, 869 2006. 69: p. 125-67.

Catela, Catarina, and Paschalis Kratsios. "Transcriptional mechanisms of motor neuron development in vertebrates and invertebrates." *Developmental biology* 475 (2021): 193-204.

Hobert, Oliver. "Regulatory logic of neuronal diversity: terminal selector genes and selector motifs." *Proceedings of the National Academy of Sciences* 105.51 (2008): 20067-20071.

Hobert, Oliver. "Terminal selectors of neuronal identity." *Current topics in developmental biology* 116 (2016): 455-475.

Hobert, Oliver, and Paschalis Kratsios. "Neuronal identity control by terminal selectors in worms, flies, and chordates." *Current Opinion in Neurobiology* 56 (2019): 97-105.

Kratsios, Paschalis, et al. "Coordinated regulation of cholinergic motor neuron traits through a conserved terminal selector gene." *Nature neuroscience* 15.2 (2012): 205-214.

Jin, Yishi, Roger Hoskins, and H. Robert Horvitz. "Control of type-D GABAergic neuron differentiation by C. elegans UNC-30 homeodomain protein." *Nature* 372.6508 (1994): 780-783.

Eastman, Catharine, H. Robert Horvitz, and Yishi Jin. "Coordinated Transcriptional Regulation of the unc-25Glutamic Acid Decarboxylase and the unc-47 GABA Vesicular Transporter by the Caenorhabditis elegans UNC-30 Homeodomain Protein." *Journal of Neuroscience* 19.15 (1999): 6225-6234.

Cinar, Hulusi, Sunduz Keles, and Yishi Jin. "Expression profiling of GABAergic motor neurons in Caenorhabditis elegans." *Current Biology* 15.4 (2005): 340-346.

Westmoreland, Joby J., et al. "Conserved Function of Caenorhabditis elegansUNC-30 and Mouse Pitx2 in Controlling GABAergic Neuron Differentiation." *Journal of Neuroscience* 21.17 (2001): 6810-6819.

Kratsios, Paschalis, et al. "Transcriptional coordination of synaptogenesis and neurotransmitter signaling." *Current Biology* 25.10 (2015): 1282-1295.

Kratsios, Paschalis, et al. "An intersectional gene regulatory strategy defines subclass diversity of C. elegans motor neurons." *Elife* 6 (2017): e25751.

Chapter 2 Emerging roles for Hox proteins in the last steps of neuronal development in worms, flies, and mice

Author contribution: **Figure 2.1** is generated by Paschalis Kratsios and Weidong Feng based on literature collected, summarized and organized by Weidong Feng. **Figure 2.2 and 2.3** are generated by Paschalis Kratsios and Yinan Li based on literature collected, summarized and organized by the same people.

2.1 Abstract

A remarkable diversity of cell types characterizes every animal nervous system. Previous studies have provided important insights into how neurons commit to a particular fate and establish circuits, but the mechanisms controlling later steps of neuronal development remain poorly understood. Hox proteins represent a conserved family of homeodomain transcription factors with well-established roles in anterior-posterior (A-P) body patterning and the early steps of nervous system development, including progenitor cell specification, neuronal migration, cell survival and circuit assembly. This review highlights recent studies in *C. elegans*, *Drosophila melanogaster* and mice that suggest non-canonical roles for Hox proteins in post-mitotic neurons during late stages of embryonic development and post-embryonic life. One key role is the transcriptional control of terminal identity features, such as expression of ion channels, neurotransmitter receptors and neuropeptides, which collectively ensure neuronal functionality. Altogether, these studies support the hypothesis that certain Hox proteins are continuously

required, from early development throughout post-embryonic life, to build and maintain a functional nervous system, thereby expanding their functions control early A-P patterning.

2.2 Introduction

Nervous system development is a multi-step process that generates a multitude of cell types. Dividing progenitor cells or neural stem cells will ultimately give rise to distinct types of neurons and glia. Newly born, post-mitotic neurons face a number of early challenges before participating into a functional neural circuit. They need to be molecularly specified, migrate to the right place, acquire distinct axo-dendritic morphology, and establish precise connections by identifying the right cellular partners. Studies in all major model organisms suggest that these early steps of nervous system development are often controlled by Hox proteins, a conserved family of homeodomain transcription factors critical for anterior-posterior (A-P) patterning and formation of the animal body plan. The early roles of Hox proteins during nervous system development have been summarized in excellent reviews (Parker and Krumlauf, 2020, Philippidou and Dasen, 2013). Here, we highlight recent studies in *C. elegans*, *Drosophila melanogaster*, and mice that uncovered new roles for Hox in the final steps of neuronal development. Emerging evidence suggests Hox genes control the expression of neuronal terminal identity features, such as the expression of neurotransmitter (NT) receptors, neuropeptides, ion channels, and NT receptors. Perhaps more strikingly, a number of Hox proteins are continuously expressed in invertebrate and vertebrate nervous systems and required to maintain terminal identity features and synapses of individual neuron types at late developmental stages and post-embryonic life. This review will discuss recent findings that point to an unexpected, continuous requirement for Hox in nervous system development and function.

2.3 Hox gene functions in the *C. elegans* nervous system

The *C. elegans* genome contains six Hox genes. The anterior Hox *ceh-13/Lab/Hox1* together with the mid-body Hox *(lin-39/Scr/Dfd/Hox3-5 and mab-5/Antp/Hox6-8)* and posterior Hox *(egl-5/AbdB/Hox9-13)* genes were identified thirty years ago (Clark et al., 1993, Costa et al., 1988, Van Auken et al., 2000), whereas two additional posterior Hox genes *(nob-1, php-3)* were discovered in 2000 (Van Auken et al., 2000). In total, *C. elegans* Hox genes are organized in 3 different sub-clusters located on Chromosome III (**Figure 2.1a**). Previous work revealed that *C. elegans* Hox genes control A-P patterning and the development of lateral epidermis and ventral ectoderm (Brunschwig et al., 1999, Chisholm, 1991, Clark et al., 1993, Costa et al., 1988, Cowing and Kenyon, 1992, Kenyon, 1986, Wang et al., 1993, Wittmann et al., 1997). During *C. elegans* nervous system development, Hox genes are utilized repeatedly at different stages, ranging from generation and specification of neural progenitors to the last steps of neuronal differentiation and maintenance of terminal identity features (**Table 2.1, 2.2**). Below, we review the role of Hox genes in *C. elegans* neurodevelopment, with an emphasis on recent studies uncovering Hox gene functions during later stages of development and postembryonic life.

2.3.1 Early roles for Hox genes in *C. elegans* nervous system development

2.3.1.1 Specification of neural progenitors

In *C. elegans*, distinct sets of ventral nerve cord motor neurons are generated in embryonic and postembryonic stages (Von Stetina et al., 2006). Twelve ventral ectodermal progenitor cells (termed P1-P12) generate twelve distinct P lineages, descendants of which eventually give rise to post-embryonic motor neurons. Hox genes act to set up A-P boundaries that differentiate P

lineage descendants in distinct zones (Clark et al., 1993, Kenyon, 1986, Salser et al., 1993, Wang et al., 1993).That is, the mid-body Hox genes *lin-39* and *mab-5* respectively regulate zones P3 to P8 and P7 to P11, whereas the posterior Hox gene *egl-5* regulates the P12 zone. Loss or aberrant expression of Hox genes will lead to homeotic transformation phenotypes. For example, loss of *mab-5* expression or ectopic *egl-5* expression in P11 will cause a P11 to P12 transformation (Jiang and Sternberg, 1998, Li et al., 2009). Besides the specification of neuronal progenitors (P cells), *egl-5* can also regulate the transdifferentiation of a rectal epithelial cell (Y) into the PDA motor neuron (Chisholm, 1991).

2.3.1.2 Control of cell survival

During early neuronal development, *C. elegans* Hox genes are also required for cell survival. The midbody Hox *lin-39* is crucial for the survival of precursor cells that will give rise to a class of postembryonic motor neurons. In the absence of *lin-39*, these precursors will undergo programmed cell death (Clark et al., 1993, Fixsen et al., 1985, Kalis et al., 2014). Similarly, the mid-body Hox gene *mab-5* promotes apoptosis within the P lineage of neuronal progenitors (P11.aaap, P12.aaap), thus preventing the generation of extra motor neurons (Liu et al., 2006). In other lineages, *mab-5* has been shown to inhibit as well as promote the generation of distinct neuroblasts (Salser and Kenyon, 1996).

2.3.1.3 Control of cell migration

The two Q neuroblasts (QL, QR) in *C. elegans* represent a powerful model to study cell migration. Their descendants migrate to well-defined positions along the A-P body axis, where they differentiate into sensory neurons and interneurons. The Hox gene *lin-39* is necessary for the anterior migration of the QR descendants, while *mab-5* dictates the posterior migration of QL descendants (Harris et al., 1996, Salser and Kenyon, 1992, Tamayo et al., 2013).

Mechanistically, *mig-13*, a single-pass transmembrane protein required for cell migration, has been identified as the Hox-dependent effector gene for Q neuroblast migration. LIN-39 establishes the anterior polarity of QR descendants by binding to the *mig-13* promoter and promoting *mig-13* expression, whereas MAB-5 inhibits the anterior polarity of QL descendants by associating with the *lin-39* promoter and down-regulating *lin-39* and *mig-13* expression (Harris et al., 1996, Salser and Kenyon, 1992, Sym et al., 1999, Wang et al., 2013). A similar role was reported for the anterior Hox gene *ceh-13.* Migration of the mechanosensory neurons AVM (descendant of QR) and PVM (descendant of QL) is significantly affected in animals lacking *ceh-13* activity (Tihanyi et al., 2010). Lastly, the posterior Hox gene *egl-5* controls migration of the hermaphrodite specific neurons (HSNs) that control egg laying. In *egl-5* mutants, HSNs do not migrate the full distance from their birthplace in the tail to the vulva (Chisholm, 1991).

2.3.2 Late roles for Hox genes in *C. elegans* nervous system development

Once post-mitotic neurons complete their migration, they extend axons and dendrites to make connections and establish circuits. The function of every neuronal circuit critically relies on the ability of its constituent neurons to communicate with each other via neurotransmitters and/or neuropeptides, as well as to display neuron type-specific electrophysiological signatures. These function-defining features are determined by the expression of neurotransmitter (NT) biosynthesis proteins, ion channels, neuropeptides, NT receptors, and cell adhesion molecules. Genes coding for such proteins have been termed “terminal identity” genes (Hobert, 2008, Hobert and Kratsios, 2019); they are expressed continuously – from late developmental stages through adulthood – to determine the final (mature) identity and function of each neuron type.

Recent studies on two different neuron types in *C. elegans*, namely the touch receptor neurons (TRNs) and nerve cord motor neurons (MNs), revealed a new role for Hox genes – the control of neuronal terminal identity (**Table 2.2**). We highlight these studies below.

2.3.2.1 Hox genes control terminal identity features of *C. elegans* TRNs

In *C. elegans*, there are six TRNs mediating sensory responses to light touch. TRNs are classified into four subtypes: (a) bilaterally symmetrical pairs of ALM and PLM neurons are located at the midbody and tail region, respectively and (b) single AVM and PVM neurons are located in the midbody (**Figure 2.1b**). ALM and PLM are born embryonically while AVM and PVM are generated post-embryonically. TRNs synapse onto and provide input to command interneurons (PVC, AVB, AVD, AVA), which stimulate downstream motor neurons, thus generating touch reflex responses. Early specification and differentiation of TRNs have been well investigated (Bounoutas and Chalfie, 2007), but how each TRN subtype acquires its unique terminal identity remains poorly understood.

At the behavioral level, animals lacking *egl-5* (posterior Hox) gene activity are touch-insensitive at the tail, suggesting defects in the development of the posteriorly located PLM neuron (Chisholm, 1991). Later studies demonstrated that *egl-5* is necessary for PLM terminal identity (Toker et al., 2003, Zheng et al., 2015a) (**Figure 2.1c**). In addition to *egl-5*, two other AbdB homologs in *C. elegans (nob-1, php-3)* control PLM development; *nob-1* is necessary for the generation of PLM precursors, whereas *php-3* together with *egl-5* diversifies PLM from its more anteriorly located counterpart, the ALM neuron. Lastly, *egl-5* controls PLM morphological characteristics, such as neurite length, by repressing anterior Hox genes (*lin-39, mab-5)* and TALE cofactors (Zheng et al., 2015a). The case of *egl-5* highlights a recurring theme of Hox

gene action across model systems – requirement for various facets of development of a specific neuron type.

In the more anteriorly located ALM neurons, the anterior Hox gene *ceh-13* regulates ALM terminal identity, as evidenced by reduced expression of a handful of terminal identity genes (*mec-4, mec-7, mec-17, mec-18*) in *ceh-13* mutant animals. Mechanistically, these studies proposed that CEH-13 and EGL-5 function as transcriptional guarantors by controlling the levels of expression of the terminal selector gene *mec-3*, which in turn is required for terminal identity of both ALM and PLM neurons (Zheng et al., 2015a, Zheng et al., 2015b). CEH-13 and EGL-5 increase the probability of *mec-3* transcriptional activation by the POU-homeodomain transcription factor UNC-86 via the same Hox/Pbx binding site in ALM and PLM neurons respectively. This molecular mechanism ensures robustness of TRN terminal differentiation.

It is interesting to note the multifaceted role of Hox genes in regulating different steps of neuronal development within the touch-reflex circuits in *C. elegans*: (a) Hox genes are involved in both early (e.g., neuronal migration, axon guidance) and late steps of TRN development (e.g., terminal differentiation). (b) All six *C. elegans* Hox genes affect TRN development in various ways: *ceh-13* regulates ALM terminal differentiation; *lin-39 and mab-5* regulates the migration of AVM/PVM precursor (from the Q neuroblasts); *egl-5* and *php-3* regulate PLM terminal differentiation; *nob-1* is necessary for the generation of PLM precursors. (c) Although it remains unclear how they control terminal identity genes (e.g., neurotransmitter biosynthesis proteins, ion channels, neuropeptides) in TRNs, two Hox proteins (CEH-13, EGL-5) appear to act directly as transcriptional guarantors of the *mec-3*, the terminal selector for all *C. elegans* TRNs. (d) Intriguingly, Hox genes control the development of neurons at different layers (sensory, interneuron, motor) of the touch-reflex circuit. Sensory TRN terminal differentiation requires

normal function of Hox genes, whereas the identity and function of command interneuron PVC (which receives sensory input from PLM) requires *egl-5* (Chisholm, 1991, Zheng et al., 2015a).

2.3.2.2 Hox genes control terminal identity features of ventral nerve cord motor neurons

Similar to the TRN studies described above, the availability of terminal identity markers for ventral nerve cord motor neurons (MNs) in *C. elegans* has critically advanced our mechanistic understanding of Hox gene function in the nervous system. Nine distinct classes of MNs are found in the nerve cord of *C. elegans* hermaphrodite animals. Based on neurotransmitter usage, they can be classified into two categories: cholinergic (SAB, DA, DB, VA, VB, AS, VC) and GABAergic (DD, VD) MNs (**Figure 2.1b**). The SAB, DA, DB, and DD classes are generated embryonically, whereas the VA, VB, VC, VD, and AS neurons are generated post-embryonically (Von Stetina et al., 2006). The terminal identity of most cholinergic MN classes in the nerve cord (SAB, DA, DB, VA, VB, AS) critically depends on the terminal selector UNC-3, member of the conserved family of Olf/Collier/Ebf(COE) family of TFs (Kratsios et al., 2015, Prasad et al., 2008, Prasad et al., 1998). Mechanistically, UNC-3 binds directly to the *cis*-regulatory region of terminal identity genes (e.g., Ach biosynthesis proteins, ion channels, neuropeptides) and activates their transcription. The homeodomain TF UNC-30 (PITX) acts in analogous manner in GABAergic (DD, VD) MNs (Eastman et al., 1999, Jin et al., 1994).

In the context of both cholinergic and GABAergic MNs, recent work demonstrated that Hox genes act as cofactors of terminal selectors (Feng et al., 2020, Kratsios et al., 2017). In cholinergic MNs, the mid-body Hox genes *lin-39* and *mab-5* collaborate with *unc-3* to activate expression of several terminal identity genes (*unc-129, del-1, acr-2, dbl-1, unc-77, slo-2)* (**Figure 2.1d**). Like UNC-3, chromatin immunoprecipitation experiments suggest that LIN-39 and MAB-5 act directly (Feng et al., 2020, Kratsios et al., 2017). In GABAergic MNs, *lin-39* and

mab-5 collaborate with *unc-30* to control terminal identity gene expression, as well. Apart from this UNC-3 co-factor role, *lin-39* is also the rate-limiting factor for ensuring cholinergic MN identity. In the absence of *unc-3*, LIN-39 no longer binds to the *cis*-regulatory region of cholinergic MN genes. Instead, it relocates and switches targets, resulting in ectopic activation of alternative identity genes (Feng et al., 2020). Hence, the terminal selector UNC-3 prevents a Hox transcriptional switch to safeguard cholinergic MN identity.

Are Hox genes required during adulthood to maintain terminal identity features and thereby ensure continuous functionality of individual neuron types? Inducible, protein depletion experiments using the auxin inducible degradation (AID) system demonstrated that the midbody Hox protein LIN-39 is required in adult life to maintain MN terminal identity features (Feng et al., 2020, Li et al., 2020). This finding is unexpected because Hox genes are mostly thought to act early during animal development. Additional work on Hox is needed in *C. elegans* and other models systems to rigorously test whether maintenance of neuronal identity is a key feature of Hox gene function in the nervous system.

The organization of cholinergic MNs into distinct subtypes along the A-P axis also offers an opportunity to dissect the molecular mechanisms underlying neuronal subtype identity. For example, the DA class of nine MNs can be subdivided into four subtypes based on cell boy position: DA1 is located at the anterior ganglion (retrovesicular ganglion [RVG]), DA2-7 are located at the VNC, and DA8-9 are found at the posterior ganglion (preanal ganglion [PAG]). In addition to their position, cholinergic MN subtypes do show distinct connectivity features and expression profiles of terminal identity genes (Kratsios et al., 2017). Hox genes control cholinergic MN subtype identity along the A-P axis of the *C. elegans* nervous system via an intersectional strategy that involves the terminal selector UNC-3 (Kratsios et al., 2017). For

example, UNC-3 is expressed in all 9 DA neurons, but collaborates with the mid-body Hox genes *lin-39* and *mab-5* in mid-body DA2-7 neurons to control their terminal identity (**Figure 2.1d**). Similarly, UNC-3 and the posterior Hox gene *egl-5* determine posterior MN (DA9) terminal identity (**Figure 2.1d**). In addition, *egl-5* also controls the appropriate synaptic wiring of DA9 neurons, illustrating that Hox proteins can coordinate connectivity and terminal identity features(Kratsios et al., 2017). Although the molecular mechanism of *egl-5* activity in posterior MNs is unknown, biochemical evidence suggest that LIN-39 – like UNC-3 - acts directly by binding on the *cis*-regulatory region of terminal identity genes. This direct mode of regulation further extends to intermediary TFs (*cfi-1/Arid3a, bnc-1/Bnc1/2*) responsible for MN subtype identity (Kerk et al., 2017, Li et al., 2020).

2.3.2.3 Posterior Hox gene *egl-5* controls the identity of serotonergic and dopaminergic neurons

In addition to its role on posterior MNs, the posterior Hox gene *egl-5* controls the identity of two other neuron types. The hermaphrodite specific neurons (HSNs) partially lose their ability to produce serotonin in *egl-5* mutants (Chisholm, 1991). Moreover, *egl-5* acts in tail sensory neurons of the *C. elegans* male. Upon *egl-5* genetic removal, these neurons do not adopt dopaminergic fate and cannot be induced to express dopamine (Lints and Emmons, 1999).
The role of the anterior Hox gene *ceh-13* during *C. elegans* neuronal terminal differentiation is largely elusive, partly due to the early larval lethality of *ceh-13* mutants (Brunschwig et al., 1999). A recent study suggested *ceh-13* controls terminal identity features of GABAergic motor neurons (DD1, DD2) located in the anterior ganglion, but the underlying mechanisms remain unknown (Aquino-Nunez et al., 2020).

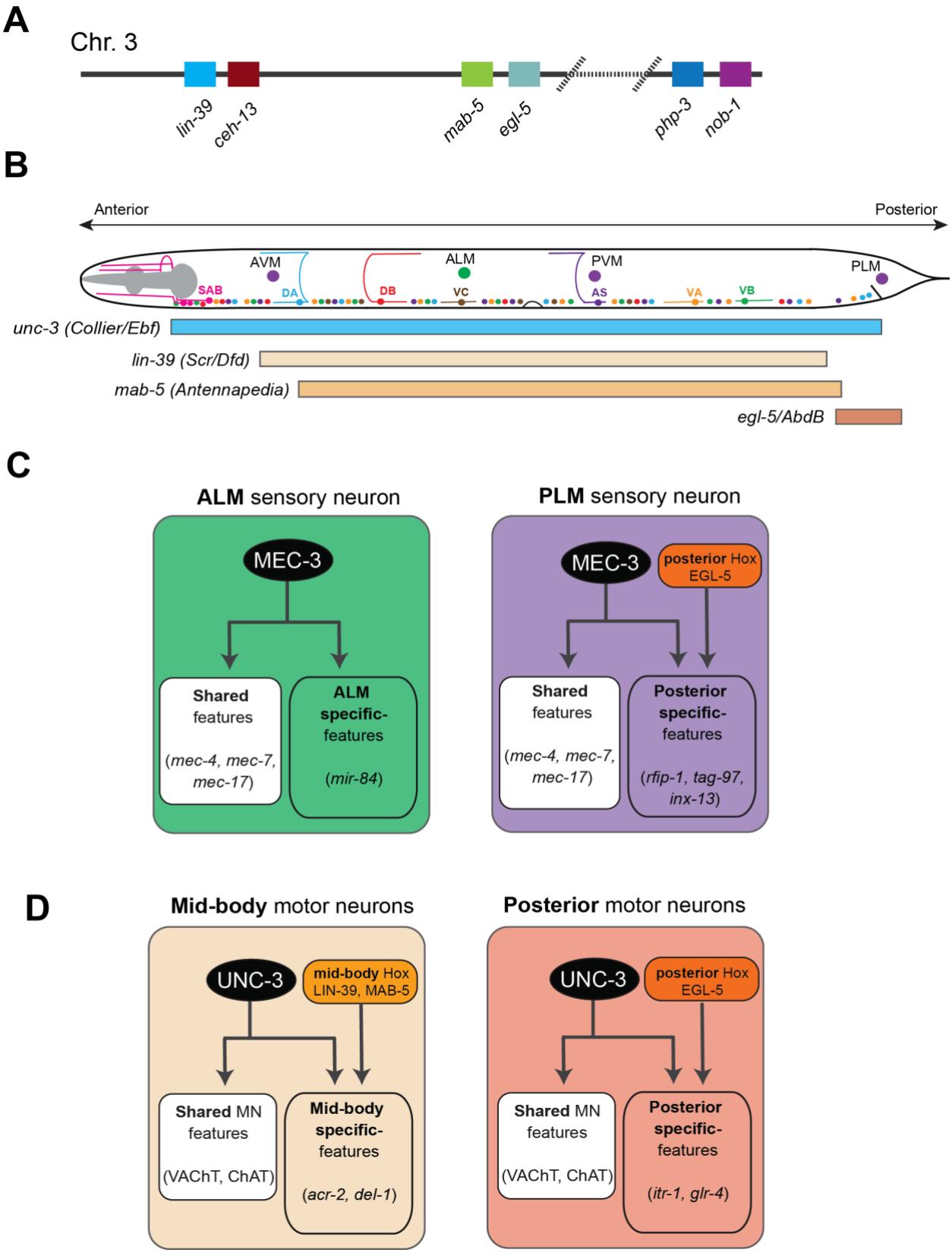

Figure 2.1 Hox gene functions in mechanosensory and motor neurons in *C. elegans*. (**a**)

Schematic of the C. elegans Hox gene cluster. (**b**) Schematic showing the cell body location of mechanosensory neurons (AVM, ALM, PVM, PLM) and cholinergic MNs in the ventral nerve cord (SAB, DA, DB, VA, VB, VC, AS). The GABAergic MNs are not shown. (**c**) The terminal selector MEC-3 controls ALM and PLM terminal identity. The activity of the posterior Hox gene *egl-5* diversifies PLM from ALM. Examples of terminal identity genes are shown in italics. (**d**) An intersectional strategy for the control of terminal identity of midbody (UNC-3, LIN-39, MAB-5) and posterior (UNC-3, EGL-5) MNs along the A-P axis of the *C. elegans* ventral nerve cord. See text for details.

2.4 Late Hox gene functions in the *Drosophila* nervous system

Hox genes were first discovered in the fruit fly *Drosophila melanogaster,* where genetic experiments during the 20th century identified mutants with dramatic phenotypes caused by homeotic transformations (e.g., legs instead of antennae, duplication of thoracic segments)(Nusslein-Volhard and Wieschaus, 1980). Subsequent studies showed that the principles of Hox gene function and their role in establishing the body plan along the A-P axis are conserved across species. All eight Hox genes in *Drosophila – labial (lab), proboscipedia (Thomas et al.), Deformed (Dfd), Sex combs reduced (Scr), Antennapedia (Seth R Taylor), Ultrabithorax (Ubx), abdominal-A (abd-A)* and *Abdominal-B (Abd-B)* – are located on chromosome 3 (**Figure 2.2a**). Based on their chromosomal location, *Drosophila* Hox genes are organized in two gene complexes. The Antennapedia complex or Antp-C (consisting of *lab, pb, Dfd, Scr,* and *Antp*) specifies the anterior body plan from the head to the anterior thorax, while the Bithorax complex or Bx-C (consisting of Ubx, abd-A, and Abd-B) specifies the segments in the posterior thorax and the abdomen (**Figure 2.2b)**. An important characteristic of Hox genes is "spatial collinearity", which describes the correspondence between the Hox gene order on the chromosome and their expression patterns along the A-P axis. This appears to be a highly conserved property of Hox genes and has been found in many other species (Gaunt, 2015).

During development, neurons in *Drosophila* arise from neuroblasts (NBs) located along the ventral nerve cord (VNC) (**Figure 2.2c**). These NBs possess the potency to generate any neuron type, but they give rise to unique types of neuronal progenies depending on their location along the A-P axis of the body. Such spatial pattern of distinct neuronal types correlates with the combinatorial expression pattern of Hox genes along the A-P axis of the *Drosophila* body. During neurogenesis, Hox gene activity guides NBs to exit the cell cycle and promotes (or blocks) apoptosis, eventually leading to a spatial map of unique neuron types (Estacio-Gomez and Diaz-Benjumea, 2014, Gummalla et al., 2014) (**Table 2.1**). In addition, Hox gene mutants also show axonal projection defects. Apart from their role in instructing the early events of neuronal development, such as neurogenesis and axon guidance (Estacio-Gomez and Diaz-Benjumea, 2014, Reichert and Bello, 2010), recent studies suggest Hox genes are also required to maintain the identity of terminal differentiated neurons.

Much of our current understanding of Hox gene function in the *Drosophila* nervous system derives from studies on the abdominal leucokinergic neurons (ABLKs), which express the neuropeptide *Leucokinin* (*Lk*) and are often used as a model system to study both embryonic and post-embryonic neurogenesis (Estacio-Gomez and Diaz-Benjumea, 2014). During embryonic neurogenesis, the NB5-5 progenitor gives rise to 7 pairs of embryonic ABLKs (eABLKs), one in each of the 7 abdominal segments (A1-7) of the VNC. *Lk* is not initially expressed in the eABLKs when they are born but becomes detectable at later developmental stages (first instar larva). Later, additional post-embryonic ABLKs (pABLKs) are generated (third instar larva), and express *Lk* in pupal stages. The cell type-specific expression of the terminal identity gene encoding *Lk* is critically dependent on Bx-C Hox gene activity (Estacio-Gomez et al., 2013).

Although *Lk* is a single Hox-dependent terminal identity gene, this study does suggest a later role for Hox in *Drosophila* neurons.

Hox genes are expressed in the neuroectoderm at early development, but then become silenced when NBs delaminate and are reactivated at later stages in specific neurons. The posterior abdominal Hox genes, Ubx and abd-a, are expressed in post-mitotic eABLKs in the first instar larvae, where they are redundantly required for the expression of *Lk*. Moreover, when both Ubx and abd-a are knocked down specifically from early second instar larvae, it results in loss of *Lk* expression in late third instar larvae (Estacio-Gomez et al., 2013), suggesting maintenance of *Lk* in eABLKs relies on continuous expression of Ubx and abd-a. On the other hand, the other posterior Hox gene Abd-b represses *Lk* expression in non-ABLK cells during both embryonic and larval neurogenesis. Similarly, Abd-b is continuously required to maintain the repression of *Lk*, as removing Abd-b from first instar larvae results in de-repression of *Lk* and increased number of ABLKs in third star larvae and adults. Although the underlying mechanisms remain unknown, these initial findings suggest that Hox genes can establish and maintain terminal identity features of post-mitotic neurons in *Drosophila*.

A second example of late Hox gene functions in *Drosophila* comes from head motor neurons (MNs) that innervate the mouth hook elevator (MHE) and depressor (MHD) muscles, which coordinate the elevation and depression of the mouth hook (MH). The anterior Hox gene *Dfd* is expressed in a subset of MNs that specifically innervate the MHE (Friedrich et al., 2016). These *Dfd*-expressing MNs play a critical role in controlling the MH-dependent motor behaviors, including hatching at the end of embryogenesis and feeding in larval stages. In *Dfd* mutants, while the number of these MNs remains unchanged, they fail to extend axonal projections to their muscle targets, resulting in failure to hatch. Intriguingly, removing *Dfd* after the

establishment of synaptic connections also results in impaired MH movements in larvae, suggesting *Dfd* is continuously required for the normal functions of these MNs (Friedrich et al., 2016). Genetically, *Dfd acts* upstream of a microtubule-organizing complex which is important for synapse stability even after their establishment. *Dfd* is continuously required to maintain the expression of Ankyrin2 extra large (Ank2-XL), which is known to be involved in determining the physical properties of synapses. Altogether, the above studies in *Drosophila* VNC and head motor neurons support the hypothesis that certain Hox genes are continuously required to establish and maintain neuronal identity and connectivity.

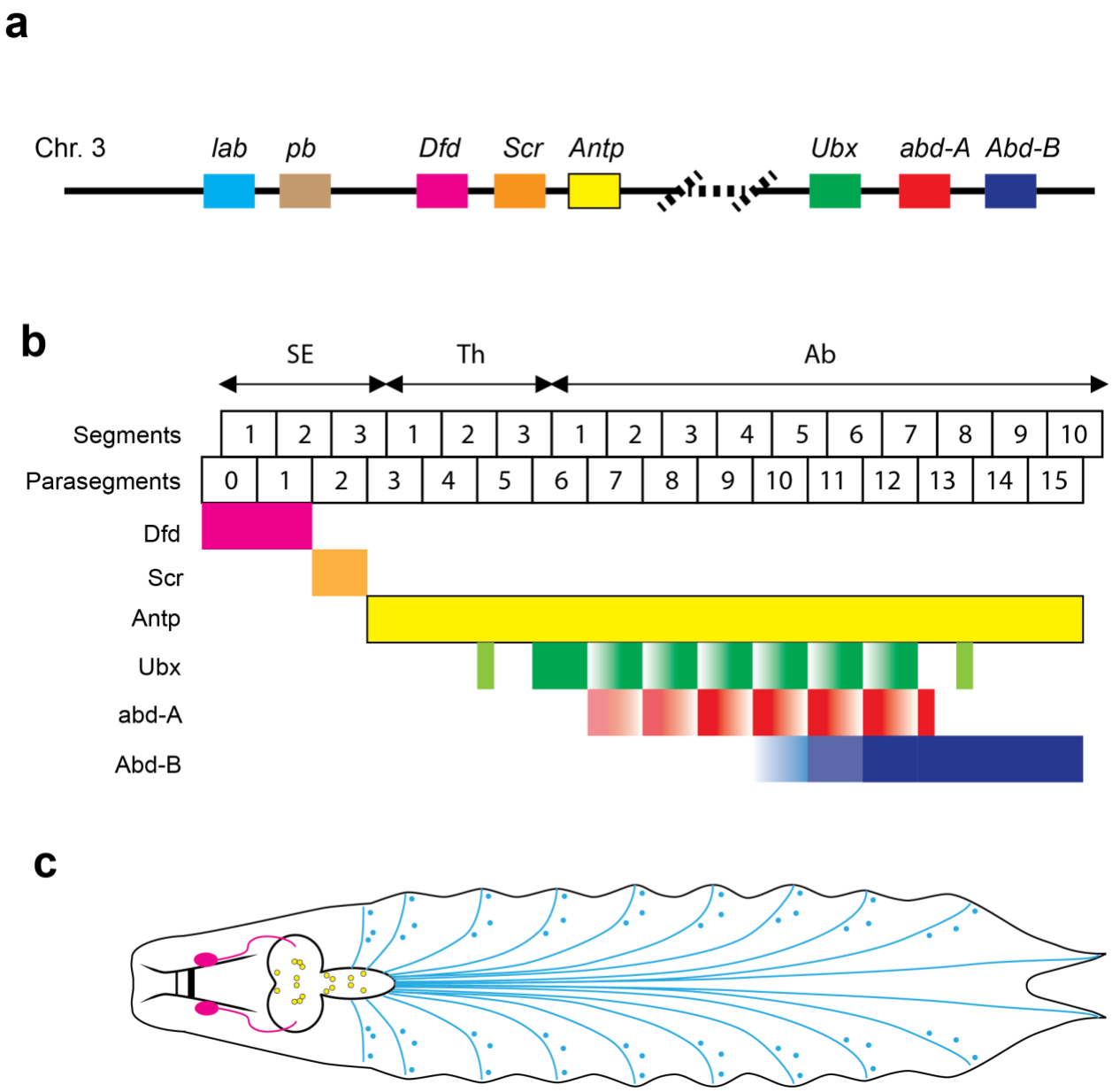

Figure 2.2 Hox gene expression in the *Drosophila* nerve cord**.** (**a**) Schematic of the Drosophila Hox gene cluster. (**b**) Six Hox genes are expressed in the Drosophila ventral nerve cord. Their expression pattern along the A-P axis is color-coded. SE: subsesophagus; Th: thorax; Ab: abdomen. Adapted from (Estacio-Gomez and Diaz-Benjumea, 2014). (**c**) Schematic of the nervous system in Drosophila larvae showing the brain and the peripheral nervous system in the ventral cord. Adapted from (Sokabe et al., 2016)

2.5 Late roles for Hox genes in the development of the mouse nervous system

During early vertebrate evolution, the single Hox gene cluster of vertebrate ancestors was duplicated twice, eventually giving rise to four clusters – *HoxA, HoxB, HoxC,* and *HoxD* in mammals (Soshnikova et al., 2013). In mice, these 4 clusters contain 39 Hox genes, which are further categorized into 13 paralog groups (PG) based on their relative position within the clusters and gene sequence (**Figure 2.3a**). The majority of these Hox genes are expressed in the mouse central nervous system (CNS) during development (Briscoe and Wilkinson, 2004, Krumlauf et al., 1993, Philippidou and Dasen, 2013).

A large body of work in the mouse hindbrain and spinal cord has uncovered critical roles for Hox genes in defining segment identity and establishing spatial gene expression patterns necessary for neuronal differentiation during early embryogenesis (Parker and Krumlauf, 2020, Philippidou and Dasen, 2013). Of note, these early Hox roles are conserved in zebrafish (**Table 2.1**). The expression of Hox genes in the hindbrain between embryonic day 7.5 (E7.5) to E9.5 appears strictly restricted within territories defined by rhombomere (transiently divided segments of the developing neural tube) boundaries (**Figure 2.3b**). Rhombomere boundaries create a series of anterior limits for Hox gene expression along the A-P axis. In the hindbrain, PG1-2 genes have more anterior boundaries compared to PG3-4 genes. On the other hand, PG5-13 genes are only expressed in the spinal cord, which is posterior to the hindbrain. The overall map of Hox gene expression along the A-P axis therefore displays collinearity with their corresponding location in the clusters (**Figure 2.3a-b**). The spatial pattern of Hox genes provides key information in cell fate determination and guiding neuronal differentiation.

Most studies in the mouse hindbrain and spinal cord focused on the early steps of neuronal development, thereby illuminating how Hox genes specify neural progenitor cells, or mediate

neuronal fate specification, and axonal path finding (Philippidou and Dasen, 2013) (**Table 2.1**). Meanwhile, it remains elusive whether mouse Hox genes control later steps in the developmental trajectory of mature, post-mitotic neurons.

Functional studies on Hox gene requirements in post-natal or adult animals are often challenging. They require conditional knockout models that would bypass the early developmental roles of Hox genes by removing gene activity in specific neuronal types at later time points. To generate such mouse models, it is crucial to establish a spatio-temporal expression map of Hox genes in the mouse nervous system. To this end, Hutlet et al. performed a systematic expression analysis of Hox genes in the adult mouse brain (Hutlet et al., 2016). This study found that the 24 Hox genes that are normally active during early development of the hindbrain continue to be expressed during adulthood. Neuroanatomical localization analysis revealed that these Hox genes are still expressed in adult post-mitotic neurons derived from rhombomeres, with visible anterior boundaries restricting individual Hox genes along the A-P axis. This indicates that the spatial collinearity rule is also maintained in adult hindbrain. Intriguingly, transcripts of Hox genes were identified in more anterior regions (forebrain) where they are not expressed during embryogenesis, suggesting Hox gene neo-expression in the adult CNS. More specifically, *Hoxb1, Hoxb3, Hoxb4, Hoxd3,* and *Hoxa5* transcripts were detected in both neocortex and the thalamus. Temporal analysis showed that their signal starts as early as the second postnatal week but becomes more robust only in the third postnatal week. In a separate study, Coughlan et al. have reported that the expression Hox PG9-11 genes is maintained and remains robust in spinocerebellar neurons until P7 (Coughlan et al., 2019), that is weeks after neuronal progenitor specification occurs. To summarize, the maintained Hox gene expression in the hindbrain and

spinal cord together with their neo-expression in the forebrain imply that Hox genes may play important roles in the post-natal CNS.

What are the biological functions of mouse Hox genes in post-mitotic neurons during late developmental and postnatal stages? While mouse studies that specifically eliminate *Hox* gene activity at postnatal stages are currently lacking, recent work suggests that mouse Hox genes, similar to their *C. elegans* counterparts, maintain the cell identity of terminally differentiated neurons. By inactivating Hox genes at successive developmental stages with conditional models, a recent study showed that *Hoxc8* is required for the maintenance of several terminal identity genes (*Nrg1, Mcam*, *Pappa*) in brachial MNs in the spinal cord (Catarina Catela, 2021). Interestingly, while these terminal differentiation genes require Hoxc8 for both initiation and maintenance of their expression, not all Hoxc8 target genes behave in the same way. In fact, the suite of Hoxc8 targets in brachial MNs is dynamic across different life stages. For example, the glycine receptor subunit alpha-2 (*Glra2*) appears significantly downregulated upon conditional knockout of *Hoxc8* at postnatal day 8 (p8), but is unaffected upon *Hoxc8* knockout at embryonic day 12 (e12). One possible explanation for this phenomenon is that *Glra2* is redundantly regulated by additional transcription factors at early stages, whose expression fades away later on and the regulation from *Hoxc8* becomes necessary for maintenance. Although the underlying mechanisms remain elusive, these findings suggest that Hox genes are continuously required in the mouse nervous system to control various aspects of neuronal development ranging from progenitor cell specification, cell survival, and neuronal terminal identity.

In mice, Hox genes are also important for neural circuit wiring. After their generation, neurons need to send out their axons and find the right partners to connect with. This process is called axon pathfinding and is regulated by the recognition of guidance cues through receptors

localized on the distal tips of axons (growth cones). Many studies have shown that correct expression of guidance cue receptors is often co-regulated by Hox genes (Di Bonito et al., 2013a, Geisen et al., 2008, Oury et al., 2006). For example, in the precerebellin anterior extramural migrating stream, Hox PG5 genes repress the repulsive Netrin receptor *Unc5b*, while Hox PG2 genes positively regulate it (Di Meglio et al., 2013). Moreover, *Hoxa2* is required for the expression of Slit receptor *Robo3* in commissural neurons in the hindbrain and *Robo2* in precerebellar poutine neurons (Di Bonito et al., 2013a). In the spinal cord, *Hox5* genes are required for proper connectivity of phrenic motor neurons to premotor interneurons and the diaphragm muscle (Philippidou et al., 2012, Vagnozzi et al., 2020). The phrenic motor neurons express a unique combination *Hox5*-dependent cell adhesion molecules of the Cadherin (Cdh) family (Vagnozzi et al., 2020), which is known to control neuronal connectivity across model systems. Importantly, early or late genetic removal of *Hox5* in mice affects diaphragm innervation, suggesting a continuous Hox requirement for establishment and maintenance of neuronal wiring (Philippidou et al., 2012). Lastly, two recent studies on *Hoxc8* provide strong evidence for Hox-mediated intrinsic of neuronal specificity. That is, *Hoxc8* removal specially in sensory neurons affects sensory-motor connectivity (Shin et al., 2020), whereas motor neuron-specific depletion of *Hoxc8* affects forelimb muscle innervation (Catela et al., 2016). Mechanistically, *Hoxc8* controls expression of axon molecules *Ret* and Gfrα to establish proper muscle innervation. Besides their role in axon guidance, many of the aforementioned axon guidance molecules are also required for synapse formation and plasticity. This leads to the possibility that Hox genes may also maintain synaptic plasticity at post-natal stages, an intriguing hypothesis that remains to be tested (Hutlet et al., 2016).

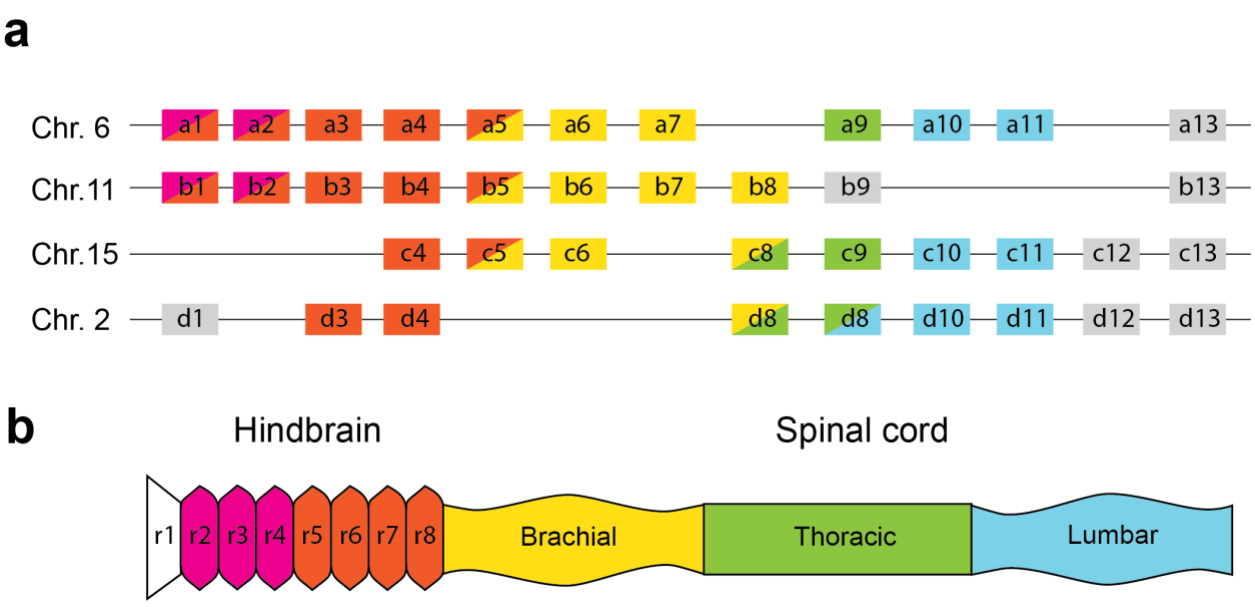

Figure 2.3 Hox gene expression in the mouse spinal cord**. (a)** The 39 Hox genes in mice are distributed in four clusters (a, b, c, d). (**b**) Region-specific Hox gene expression is shown along the rostrocaudal axis of the nervous system. Adapted from (Philippidou and Dasen, 2013).

2.6 Conclusions

In the nervous system, a large body of work has uncovered critical roles for Hox genes in the early steps of nervous system development, such as progenitor cell specification, neuronal migration, cell survival and circuit assembly. This review highlights recent studies that identified new functions for Hox genes in post-mitotic neurons during late developmental stages and postembryonic life, supporting the hypothesis that Hox genes control distinct aspects of neuronal development and function across different life stages. Hence, temporal inactivation of Hox gene activity is necessary to uncover the breadth of Hox gene functions in the nervous system. CRISPR/Cas9 genome editing coupled with powerful transcriptomic, biochemical, and behavioral approaches will help the realization of such an ambitious goal.

Table 2.1: A non-comprehensive list of Hox studies focused on early nervous system development.

Species	Gene	Description	Reference
C. elegans	*lin-39*	lin-39 is required for mid-body region-specific development (neuronal progenitor specification, vulva development) and Q neuroblast migration	(Clark et al., 1993)
C. elegans	*lin-39, mab-5*	lin-39 and mab-5 specify cell fates in posterior ventral nerve cord	(Salser et al., 1993)
C. elegans	*egl-5*	egl-5 specifies the P12 neuroectoblast fate via interactions with EGF, Wnt and other Hox genes	(Jiang and Sternberg, 1998)
C. elegans	*lin-39*	lin-39 promotes the survival of VC motor neuron precursor cells	(Fixsen et al., 1985)
C. elegans	*mab-5*	mab-5 promotes apoptosis, preventing the generation of extra motor neurons	(Liu et al., 2006)
C. elegans	*mab-5*	mab-5 closely controls the generation of neuroblasts within multiple lineages and fine-tunes the specification of lineages.	(Salser and Kenyon, 1996)
C. elegans	*mab-5*	*mab-5* dictates the posterior migration of QL descendants	(Salser and Kenyon, 1992) (Harris et al., 1996)

C. elegans	*ceh-13*	ceh-13 regulates the migration of Q neuroblast	(Tihanyi et al., 2010)
C. elegans	*egl-5*	egl-5 regulates the specification of PDA motor neuron and the migration of HSN	(Chisholm, 1991)
Drosophila	*lab*	*lab* is required for neuronal fate acquisition by postmitotic progenies of neuronal progenitors in the tritocerebrum.	(Hirth et al., 1998)
Drosophila	*abd-A*	*abd-A* specifies the time when apoptosis occurs and control the number of neurons produced from postembryonic neuroblast proliferation	(Bello et al., 2003)
Drosophila	*Ubx*	*Ubx* is required for the expression of the transcription factor *Dimm* in Va neurons	(Suska et al., 2011)
Drosophila	*Abd-B*	*Abd-B* prevents cell-autonomously the apoptosis of dMP2 and MP1 pioneer neurons	(Miguel-Aliaga and Thor, 2004)
Mouse	*Hoxd1*	*Hoxd1* is required for the specification of nociceptor axonal connectivity in the spinal cord	(Guo et al., 2011)
Mouse	*Hoxa2*	*Hoxa2* controls the formation of the sound localization circuit involving contralateral projections from the anterior VCN to the	(Di Bonito et al., 2013b)

		medial nucleus of the trapezoid body by regulating the Rig1 axon guidance receptor	
Mouse	*Hoxa5/Hoxc5*	*Hoxa5* and *Hoxc5* control the clustering and intramuscular branching of PMC motor neurons	(Philippidou et al., 2012)
Mouse	*Hoxc6*	Hoxc6 promotes motor neuron pool diversity within the LMC	(Lacombe et al., 2013)
Danio rerio	*hoxb1a*	hoxb1a regulates migration of the VIIth cranial nerve branchiomotor neurons	(McClintock et al., 2002)
Danio rerio	*hoxb1b*	*hoxb1b specifies hindbrain development via collaboration with cofactors Meis3 and Pbx4*	(Vlachakis et al., 2001)

Table 2.2: Hox gene studies focused on late steps of nervous system development.

Species	Gene	Description	Reference
C. elegans	*egl-5*	egl-5 regulates terminal differentiation of PLM	(Toker et al., 2003)
C. elegans	*egl-5*	EGL-5 is required for subtype-specific circuit formation by acting in both the sensory neuron and downstream interneuron to promote functional connectivity in touch receptor neurons.	(Zheng et al., 2015a)
C. elegans	*ceh-13*	CEH-13 functions cell non-autonomously to guide ALM migration and axonal outgrowth	(Zheng et al., 2015a)
C. elegans	*php-3*	PHP-3 makes PLM neurons morphologically distinct from ALM neurons independently with egl-5	(Zheng et al., 2015a)
C. elegans	*nob-1*	nob-1 is needed to generate the cells that become the PLM neurons	(Zheng et al., 2015b)
C. elegans	*ceh-13, egl-5*	CEH-13 and EGL-5 act as transcriptional guarantors to ensure reliable and robust mec-3 expression during terminal neuronal differentiation of touch receptor neurons	(Zheng et al., 2015b)
C. elegans	*lin-39, mab-5, egl-5*	Hox genes function as unc-3 co-factors to specify cholinergic motor neuron sub-class terminal identities	(Kratsios et al., 2017)

C. elegans	*lin-39, mab-5*	lin-39 and mab-5 regulates and maintains subtype specific terminal identities of both cholinergic and GABAergic motor neurons	(Feng et al., 2020)
C. elegans	*lin-39 mab-5*	Hox genes regulates cfi-1 to regulate ventral cord motor neuron terminal identity	(Li et al., 2020)
C. elegans	*egl-5*	egl-5 is crucial for HSN to adopt the serotonergic identity	(Chisholm, 1991)
C. elegans	*egl-5*	egl-5 regulates HSN terminal identity through regulating UNC-86	(Baum et al., 1999)
C. elegans	*egl-5*	egl-5 is required for the adoption of dopaminergic identity for ray cells through regulation of dbl-1	(Lints and Emmons, 1999)
C. elegans	*ceh-13*	*ceh-13* specifies the terminal identity of two GABAergic motor neurons DD1 and DD2	(Aquino-Nunez et al., 2020)
Drosophila	*Ubx/abd-a*	*Ubx* and *abd-a* are required to maintain the expression of the neuropeptide *Lk* in larval stages	(Estacio-Gomez et al., 2013)
Drosophila	*Dfd*	*Dfd* is continuously required to maintain the expression of Ankyrin2 extra large (Ank2-XL) and thus synaptic stability in head motor neurons (MNs) that innervate the mouth hood elevator (MHE) and depressor (MHD) muscles	(Friedrich et al., 2016)

Mouse	*Hoxc8*	*Hoxc8* is required for the maintenance of terminal identity genes *Nrg1*, *Mcam*, and *Pappa* in the spinal cord.	(Catarina Catela, 2021)
Mouse	*Hox5*	Late removal of *Hox5* genes depletes PMC motor neuron number and branches, suggesting it is continuously required for the survival of these neurons.	(Philippidou et al., 2012)

2.7 References

BOOBAKER, A. A. & BLAXTER, M. L. 2003. Hox Gene Loss during Dynamic Evolution of the Nematode Cluster. *Curr Biol,* 13**,** 37-40.

AQUINO-NUNEZ, W., MIELKO, Z. E., DUNN, T., SANTORELLA, E. M., HOSEA, C., LEITNER, L., MCCALLA, D., SIMMS, C., VEROLA, W. M., VIJAYKUMAR, S. & HUDSON, M. L. 2020. cnd-1/NeuroD1 Functions with the Homeobox Gene ceh-5/Vax2 and Hox Gene ceh-13/labial To Specify Aspects of RME and DD Neuron Fate in Caenorhabditis elegans. *G3 (Bethesda),* 10**,** 3071-3085.

BAUM, P. D., GUENTHER, C., FRANK, C. A., PHAM, B. V. & GARRIGA, G. 1999. The Caenorhabditis elegans gene ham-2 links Hox patterning to migration of the HSN motor neuron. *Genes Dev,* 13**,** 472-83.

BELLO, B. C., HIRTH, F. & GOULD, A. P. 2003. A pulse of the Drosophila Hox protein Abdominal-A schedules the end of neural proliferation via neuroblast apoptosis. *Neuron,* 37**,** 209-19.

BOUNOUTAS, A. & CHALFIE, M. 2007. Touch sensitivity in Caenorhabditis elegans. *Pflugers Arch,* 454**,** 691-702.

BRISCOE, J. & WILKINSON, D. G. 2004. Establishing neuronal circuitry: Hox genes make the connection. *Genes Dev,* 18**,** 1643-8.

BRUNSCHWIG, K., WITTMANN, C., SCHNABEL, R., BURGLIN, T. R., TOBLER, H. & MULLER, F. 1999. Anterior organization of the Caenorhabditis elegans embryo by the labial-like Hox gene ceh-13. *Development,* 126**,** 1537-46.

CATARINA CATELA, Y. W., KAILONG WEN, WEIDONG FENG, PASCHALIS KRATSIOS 2021. Control of spinal motor neuron terminal differentiation through sustained Hoxc8 gene activity. *bioRxiv*.

CATELA, C., SHIN, M. M., LEE, D. H., LIU, J. P. & DASEN, J. S. 2016. Hox Proteins Coordinate Motor Neuron Differentiation and Connectivity Programs through Ret/Gfralpha Genes. *Cell Rep,* 14**,** 1901-15.

CHISHOLM, A. 1991. Control of cell fate in the tail region of C. elegans by the gene egl-5. *Development,* 111**,** 921-32.

CLARK, S. G., CHISHOLM, A. D. & HORVITZ, H. R. 1993. Control of cell fates in the central body region of C. elegans by the homeobox gene lin-39. *Cell,* 74**,** 43-55.

COSTA, M., WEIR, M., COULSON, A., SULSTON, J. & KENYON, C. 1988. Posterior pattern formation in C. elegans involves position-specific expression of a gene containing a homeobox. *Cell,* 55**,** 747-56.

COUGHLAN, E., GARSIDE, V. C., WONG, S. F. L., LIANG, H., KRAUS, D., KARMAKAR, K., MAHESHWARI, U., RIJLI, F. M., BOURNE, J. & MCGLINN, E. 2019. A Hox Code Defines Spinocerebellar Neuron Subtype Regionalization. *Cell Rep,* 29**,** 2408-2421 e4.

COWING, D. W. & KENYON, C. 1992. Expression of the homeotic gene mab-5 during Caenorhabditis elegans embryogenesis. *Development,* 116**,** 481-90.

DI BONITO, M., GLOVER, J. C. & STUDER, M. 2013a. Hox genes and region-specific sensorimotor circuit formation in the hindbrain and spinal cord. *Dev Dyn,* 242**,** 1348-68.

DI BONITO, M., NARITA, Y., AVALLONE, B., SEQUINO, L., MANCUSO, M., ANDOLFI, G., FRANZE, A. M., PUELLES, L., RIJLI, F. M. & STUDER, M. 2013b. Assembly of the auditory circuitry by a Hox genetic network in the mouse brainstem. *PLoS Genet,* 9**,** e1003249.

DI MEGLIO, T., KRATOCHWIL, C. F., VILAIN, N., LOCHE, A., VITOBELLO, A., YONEHARA, K., HRYCAJ, S. M., ROSKA, B., PETERS, A. H., EICHMANN, A., WELLIK, D., DUCRET, S. & RIJLI, F. M. 2013. Ezh2 orchestrates topographic migration and connectivity of mouse precerebellar neurons. *Science,* 339**,** 204-7.

EASTMAN, C., HORVITZ, H. R. & JIN, Y. 1999. Coordinated transcriptional regulation of the unc-25 glutamic acid decarboxylase and the unc-47 GABA vesicular transporter by the Caenorhabditis elegans UNC-30 homeodomain protein. *J Neurosci,* 19**,** 6225-34.

ESTACIO-GOMEZ, A. & DIAZ-BENJUMEA, F. J. 2014. Roles of Hox genes in the patterning of the central nervous system of Drosophila. *Fly (Austin),* 8**,** 26-32.

ESTACIO-GOMEZ, A., MORIS-SANZ, M., SCHAFER, A. K., PEREA, D., HERRERO, P. & DIAZ-BENJUMEA, F. J. 2013. Bithorax-complex genes sculpt the pattern of leucokinergic neurons in the Drosophila central nervous system. *Development,* 140**,** 2139-48.

FENG, W., LI, Y., DAO, P., ABURAS, J., ISLAM, P., ELBAZ, B., KOLARZYK, A., BROWN, A. E. & KRATSIOS, P. 2020. A terminal selector prevents a Hox transcriptional switch to safeguard motor neuron identity throughout life. *Elife,* 9.

FIXSEN, W., STERNBERG, P., ELLIS, H. & HORVITZ, R. 1985. Genes that affect cell fates during the development of Caenorhabditis elegans. *Cold Spring Harb Symp Quant Biol,* 50**,** 99-104.

FRIEDRICH, J., SORGE, S., BUJUPI, F., EICHENLAUB, M. P., SCHULZ, N. G., WITTBRODT, J. & LOHMANN, I. 2016. Hox Function Is Required for the Development and Maintenance of the Drosophila Feeding Motor Unit. *Cell Rep,* 14**,** 850-60.

GAUNT, S. J. 2015. The significance of Hox gene collinearity. *Int J Dev Biol,* 59**,** 159-70.

GEISEN, M. J., DI MEGLIO, T., PASQUALETTI, M., DUCRET, S., BRUNET, J. F., CHEDOTAL, A. & RIJLI, F. M. 2008. Hox paralog group 2 genes control the migration of mouse pontine neurons through slit-robo signaling. *PLoS Biol,* 6**,** e142.

GUMMALLA, M., GALETTI, S., MAEDA, R. K. & KARCH, F. 2014. Hox gene regulation in the central nervous system of Drosophila. *Front Cell Neurosci,* 8**,** 96.

GUO, T., MANDAI, K., CONDIE, B. G., WICKRAMASINGHE, S. R., CAPECCHI, M. R. & GINTY, D. D. 2011. An evolving NGF-Hoxd1 signaling pathway mediates development of divergent neural circuits in vertebrates. *Nat Neurosci,* 14**,** 31-6.

HARRIS, J., HONIGBERG, L., ROBINSON, N. & KENYON, C. 1996. Neuronal cell migration in C. elegans: regulation of Hox gene expression and cell position. *Development,* 122**,** 3117-31.

HIRTH, F., HARTMANN, B. & REICHERT, H. 1998. Homeotic gene action in embryonic brain development of Drosophila. *Development,* 125**,** 1579-89.

HOBERT, O. 2008. Regulatory logic of neuronal diversity: terminal selector genes and selector motifs. *Proc Natl Acad Sci U S A,* 105**,** 20067-71.

HOBERT, O. & KRATSIOS, P. 2019. Neuronal identity control by terminal selectors in worms, flies, and chordates. *Curr Opin Neurobiol,* 56**,** 97-105.

HUTLET, B., THEYS, N., COSTE, C., AHN, M. T., DOSHISHTI-AGOLLI, K., LIZEN, B. & GOFFLOT, F. 2016. Systematic expression analysis of Hox genes at adulthood reveals novel patterns in the central nervous system. *Brain Struct Funct,* 221**,** 1223-43.

JIANG, L. I. & STERNBERG, P. W. 1998. Interactions of EGF, Wnt and HOM-C genes specify the P12 neuroectoblast fate in C. elegans. *Development,* 125**,** 2337-47.

JIN, Y., HOSKINS, R. & HORVITZ, H. R. 1994. Control of type-D GABAergic neuron differentiation by C. elegans UNC-30 homeodomain protein. *Nature,* 372**,** 780-3.

KALIS, A. K., KISSIOV, D. U., KOLENBRANDER, E. S., PALCHICK, Z., RAGHAVAN, S., TETREAULT, B. J., WILLIAMS, E., LOER, C. M. & WOLFF, J. R. 2014. Patterning of sexually dimorphic neurogenesis in the caenorhabditis elegans ventral cord by Hox and TALE homeodomain transcription factors. *Dev Dyn,* 243**,** 159-71.

KENYON, C. 1986. A gene involved in the development of the posterior body region of C. elegans. *Cell,* 46**,** 477-87.

KERK, S. Y., KRATSIOS, P., HART, M., MOURAO, R. & HOBERT, O. 2017. Diversification of C. elegans Motor Neuron Identity via Selective Effector Gene Repression. *Neuron,* 93**,** 80-98.

KRATSIOS, P., KERK, S. Y., CATELA, C., LIANG, J., VIDAL, B., BAYER, E. A., FENG, W., DE LA CRUZ, E. D., CROCI, L., CONSALEZ, G. G., MIZUMOTO, K. & HOBERT, O. 2017. An intersectional gene regulatory strategy defines subclass diversity of C. elegans motor neurons. *Elife,* 6.

KRATSIOS, P., PINAN-LUCARRE, B., KERK, S. Y., WEINREB, A., BESSEREAU, J. L. & HOBERT, O. 2015. Transcriptional coordination of synaptogenesis and neurotransmitter signaling. *Curr Biol,* 25**,** 1282-95.

KRUMLAUF, R., MARSHALL, H., STUDER, M., NONCHEV, S., SHAM, M. H. & LUMSDEN, A. 1993. Hox homeobox genes and regionalisation of the nervous system. *J Neurobiol,* 24**,** 1328-40.

LACOMBE, J., HANLEY, O., JUNG, H., PHILIPPIDOU, P., SURMELI, G., GRINSTEIN, J. & DASEN, J. S. 2013. Genetic and functional modularity of Hox activities in the specification of limb-innervating motor neurons. *PLoS Genet,* 9**,** e1003184.

LI, X., KULKARNI, R. P., HILL, R. J. & CHAMBERLIN, H. M. 2009. HOM-C genes, Wnt signaling and axial patterning in the C. elegans posterior ventral epidermis. *Dev Biol,* 332**,** 156-65.

LI, Y., OSUMA, A., CORREA, E., OKEBALAMA, M. A., DAO, P., GAYLORD, O., ABURAS, J., ISLAM, P., BROWN, A. E. & KRATSIOS, P. 2020. Establishment and maintenance of motor neuron identity via temporal modularity in terminal selector function. *Elife,* 9.

LINTS, R. & EMMONS, S. W. 1999. Patterning of dopaminergic neurotransmitter identity among Caenorhabditis elegans ray sensory neurons by a TGFbeta family signaling pathway and a Hox gene. *Development,* 126**,** 5819-31.

LIU, H., STRAUSS, T. J., POTTS, M. B. & CAMERON, S. 2006. Direct regulation of egl-1 and of programmed cell death by the Hox protein MAB-5 and by CEH-20, a C. elegans homolog of Pbx1. *Development,* 133**,** 641-50.

MCCLINTOCK, J. M., KHEIRBEK, M. A. & PRINCE, V. E. 2002. Knockdown of duplicated zebrafish hoxb1 genes reveals distinct roles in hindbrain patterning and a novel mechanism of duplicate gene retention. *Development,* 129**,** 2339-54.

MIGUEL-ALIAGA, I. & THOR, S. 2004. Segment-specific prevention of pioneer neuron apoptosis by cell-autonomous, postmitotic Hox gene activity. *Development,* 131**,** 6093-105.

NUSSLEIN-VOLHARD, C. & WIESCHAUS, E. 1980. Mutations affecting segment number and polarity in Drosophila. *Nature,* 287**,** 795-801.

OURY, F., MURAKAMI, Y., RENAUD, J. S., PASQUALETTI, M., CHARNAY, P., REN, S. Y. & RIJLI, F. M. 2006. Hoxa2- and rhombomere-dependent development of the mouse facial somatosensory map. *Science,* 313**,** 1408-13.

PARKER, H. J. & KRUMLAUF, R. 2020. A Hox gene regulatory network for hindbrain segmentation. *Curr Top Dev Biol,* 139**,** 169-203.

PHILIPPIDOU, P. & DASEN, J. S. 2013. Hox genes: choreographers in neural development, architects of circuit organization. *Neuron,* 80**,** 12-34.

PHILIPPIDOU, P., WALSH, C. M., AUBIN, J., JEANNOTTE, L. & DASEN, J. S. 2012. Sustained Hox5 gene activity is required for respiratory motor neuron development. *Nat Neurosci,* 15**,** 1636-44.

PRASAD, B., KARAKUZU, O., REED, R. R. & CAMERON, S. 2008. unc-3-dependent repression of specific motor neuron fates in Caenorhabditis elegans. *Dev Biol,* 323**,** 207-15.

PRASAD, B. C., YE, B., ZACKHARY, R., SCHRADER, K., SEYDOUX, G. & REED, R. R. 1998. unc-3, a gene required for axonal guidance in Caenorhabditis elegans, encodes a member of the O/E family of transcription factors. *Development,* 125**,** 1561-8.

REICHERT, H. & BELLO, B. 2010. Hox genes and brain development in Drosophila. *Adv Exp Med Biol,* 689**,** 145-53.

SALSER, S. J. & KENYON, C. 1992. Activation of a C. elegans Antennapedia homologue in migrating cells controls their direction of migration. *Nature,* 355**,** 255-8.

SALSER, S. J. & KENYON, C. 1996. A C. elegans Hox gene switches on, off, on and off again to regulate proliferation, differentiation and morphogenesis. *Development,* 122**,** 1651-61.

SALSER, S. J., LOER, C. M. & KENYON, C. 1993. Multiple HOM-C gene interactions specify cell fates in the nematode central nervous system. *Genes Dev,* 7**,** 1714-24.

SETH R TAYLOR, G. S., MOLLY REILLY, LORI GLENWINKEL, ABIGAIL POFF, REBECCA MCWHIRTER, CHUAN XU, ALEXIS WEINREB, MANASA BASAVARAJU, STEVEN J COOK, ALEC BARRETT, ALEXANDER ABRAMS, BERTA VIDAL, CYRIL

CROS, IBNUL RAFI, NENAD SESTAN, MARC HAMMARLUND, OLIVER HOBERT, DAVID M. MILLER III 2019. Expression profiling of the mature C. elegans nervous system by single-cell RNA-Sequencing. *bioRxiv*.

SHIN, M. M., CATELA, C. & DASEN, J. 2020. Intrinsic control of neuronal diversity and synaptic specificity in a proprioceptive circuit. *Elife,* 9.

SOKABE, T., CHEN, H. C., LUO, J. & MONTELL, C. 2016. A Switch in Thermal Preference in Drosophila Larvae Depends on Multiple Rhodopsins. *Cell Rep,* 17**,** 336-344.

SOSHNIKOVA, N., DEWAELE, R., JANVIER, P., KRUMLAUF, R. & DUBOULE, D. 2013. Duplications of hox gene clusters and the emergence of vertebrates. *Dev Biol,* 378**,** 194-9.

SUSKA, A., MIGUEL-ALIAGA, I. & THOR, S. 2011. Segment-specific generation of Drosophila Capability neuropeptide neurons by multi-faceted Hox cues. *Dev Biol,* 353**,** 72-80.

SYM, M., ROBINSON, N. & KENYON, C. 1999. MIG-13 positions migrating cells along the anteroposterior body axis of C. elegans. *Cell,* 98**,** 25-36.

TAMAYO, J. V., GUJAR, M., MACDONALD, S. J. & LUNDQUIST, E. A. 2013. Functional transcriptomic analysis of the role of MAB-5/Hox in Q neuroblast migration in Caenorhabditis elegans. *BMC Genomics,* 14**,** 304.

THOMAS, P. D., CAMPBELL, M. J., KEJARIWAL, A., MI, H., KARLAK, B., DAVERMAN, R., DIEMER, K., MURUGANUJAN, A. & NARECHANIA, A. 2003. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res,* 13**,** 2129-41.

TIHANYI, B., VELLAI, T., REGOS, A., ARI, E., MULLER, F. & TAKACS-VELLAI, K. 2010. The C. elegans Hox gene ceh-13 regulates cell migration and fusion in a non-colinear way. Implications for the early evolution of Hox clusters. *BMC Dev Biol,* 10**,** 78.

TOKER, A. S., TENG, Y., FERREIRA, H. B., EMMONS, S. W. & CHALFIE, M. 2003. The Caenorhabditis elegans spalt-like gene sem-4 restricts touch cell fate by repressing the selector Hox gene egl-5 and the effector gene mec-3. *Development,* 130**,** 3831-40.

VAGNOZZI, A. N., GARG, K., DEWITZ, C., MOORE, M. T., CREGG, J. M., JEANNOTTE, L., ZAMPIERI, N., LANDMESSER, L. T. & PHILIPPIDOU, P. 2020. Phrenic-specific transcriptional programs shape respiratory motor output. *Elife,* 9.

VAN AUKEN, K., WEAVER, D. C., EDGAR, L. G. & WOOD, W. B. 2000. Caenorhabditis elegans embryonic axial patterning requires two recently discovered posterior-group Hox genes. *Proc Natl Acad Sci U S A,* 97**,** 4499-503.

VLACHAKIS, N., CHOE, S. K. & SAGERSTROM, C. G. 2001. Meis3 synergizes with Pbx4 and Hoxb1b in promoting hindbrain fates in the zebrafish. *Development,* 128**,** 1299-312.

VON STETINA, S. E., TREININ, M. & MILLER, D. M., 3RD 2006. The motor circuit. *Int Rev Neurobiol,* 69**,** 125-67.

WANG, B. B., MULLER-IMMERGLUCK, M. M., AUSTIN, J., ROBINSON, N. T., CHISHOLM, A. & KENYON, C. 1993. A homeotic gene cluster patterns the anteroposterior body axis of C. elegans. *Cell,* 74**,** 29-42.

WANG, X., ZHOU, F., LV, S., YI, P., ZHU, Z., YANG, Y., FENG, G., LI, W. & OU, G. 2013. Transmembrane protein MIG-13 links the Wnt signaling and Hox genes to the cell polarity in neuronal migration. *Proc Natl Acad Sci U S A,* 110**,** 11175-80.

WITTMANN, C., BOSSINGER, O., GOLDSTEIN, B., FLEISCHMANN, M., KOHLER, R., BRUNSCHWIG, K., TOBLER, H. & MULLER, F. 1997. The expression of the C. elegans labial-like Hox gene ceh-13 during early embryogenesis relies on cell fate and on anteroposterior cell polarity. *Development,* 124**,** 4193-200.

ZHENG, C., DIAZ-CUADROS, M. & CHALFIE, M. 2015a. Hox Genes Promote Neuronal Subtype Diversification through Posterior Induction in Caenorhabditis elegans. *Neuron,* 88**,** 514-27.

ZHENG, C., JIN, F. Q. & CHALFIE, M. 2015b. Hox Proteins Act as Transcriptional Guarantors to Ensure Terminal Differentiation. *Cell Rep,* 13**,** 1343-1352.

Chapter 3 Maintenance of neurotransmitter identity by Hox proteins through a homeostatic mechanism

Author contribution: All data are generated by Weidong Feng except that in **Figure 3.5F-H**. Special acknowledgements to Minhkhoi Nguyen for his contribution to **Figure 3. 6G**. All figures are jointly generated by Weidong Feng and Paschalis Kratsios.

3.1 Summary

Hox transcription factors play fundamental roles during early patterning, but they are also expressed continuously, from embryonic stages through adulthood, in the nervous system. However, the functional significance of their sustained expression remains unclear. In *C. elegans* motor neurons (MNs), we find that LIN-39 (Scr/Dfd/Hox4-5) is continuously required during post-embryonic life to maintain neurotransmitter identity, a core element of neuronal function. LIN-39 acts directly to co-regulate genes that define cholinergic identity (e.g., *unc-17/VAChT, cho-1/ChT*). We further show that LIN-39, MAB-5 (Antp/Hox6-8) and the transcription factor UNC-3 (Collier/Ebf) operate in a positive feedforward loop to ensure continuous and robust expression of cholinergic identity genes. Finally, we identify a two-component design principle for homeostatic control of Hox gene expression in adult MNs: Hox transcriptional autoregulation is counterbalanced by negative UNC-3 feedback. These findings uncover a noncanonical role for Hox proteins during post-embryonic life, critically broadening their functional repertoire from early patterning to the control of neurotransmitter identity.

3.2 Introduction

Information flow in the nervous system critically relies on the ability of distinct neuron types to synthesize and package into synaptic vesicles specific chemical substances, known as neurotransmitters (NTs). Hence, a core functional feature of each neuron type is its NT identity, defined by the co-expression of genes encoding proteins necessary for the synthesis, packaging, and breakdown of a particular NT. Although rare instances of NT identity switching have been described in the nervous system [1, 2], it is generally accepted that individual neuron types acquire a specific NT identity during development and maintain it throughout life.

In the case of cholinergic neurons, the enzyme choline acetyltransferase (ChAT) synthesizes acetylcholine (ACh) from its precursor choline, the vesicular ACh transporter (VAChT) packages ACh into synaptic vesicles, and the enzyme acetylcholinesterase (AChE) breaks down ACh upon its release into choline. The choline transporter (ChT) reuptakes choline back into the cholinergic neuron (**Fig. 3.1A**) [3]. Co-expression of all these proteins throughout the life of a cholinergic neuron ensures the continuation of ACh biosynthesis, and thereby the communication of cholinergic neurons with their post-synaptic targets. Importantly, reduced expression of ACh pathway proteins has been associated with multiple neurological disorders [4-6]. Despite the widespread use of ACh in every nervous system [7, 8], how co-expression of ACh pathway proteins is controlled over time, from development through adulthood, is poorly understood. A handful of studies in *C. elegans* and mice identified LIM and POU homeodomain transcription factors as necessary during development for ACh pathway gene expression in various types of cholinergic neurons [9-12]. However, the mechanisms that maintain expression of ACh pathway genes during post-embryonic life are largely unknown.

Motor neurons (MNs) in the spinal cord of vertebrates and ventral nerve cord of many invertebrates use ACh to communicate with their muscle targets. The nematode *C. elegans* contains six different classes (types) of MNs within its ventral nerve cord that are cholinergic [13]. Five classes (DA, DB, VA, VB, AS) control locomotion and are sex-shared (found in *C. elegans* males and hermaphrodites), whereas one class (VC) controls egg-laying and is found only in hermaphrodites. Each MN class contains several members (DA = 9 neurons, DB = 7, VA = 12, VB = 11, AS = 11, VC = 6) that intermingle along the nerve cord. The cholinergic identity of these MNs is defined by the expression of *unc-17* (VAChT), *cha-1* (ChAT), *ace-2* (AChE), and *cho-1* (ChT) (**Fig. 3.1A**) [15, 16]. The rapid life cycle of *C. elegans* (~3 days from embryo to adult), the established methods to inactivate transcription factor activity in the adult, and the availability of reporter animals for all ACh pathway genes offer a unique opportunity to probe *in vivo,* and with single-cell resolution, the molecular mechanisms that maintain cholinergic identity during post-embryonic life.

Members of the HOX family of homeodomain transcription factors play fundamental roles during early development in anterior-posterior patterning and body plan formation [17, 18]. Seminal studies in worms, flies, zebrafish and mice have also established critical roles for Hox genes during the early development of the nervous system [19-23]. For example, Hox gene inactivation can affect the specification and/or survival of neuronal progenitors, as well as the differentiation, survival, migration, and/or connectivity of young post-mitotic neurons [19-23]. The focus on early development, however, combined with a lack of temporally controlled gene inactivation approaches that bypass early Hox pleiotropies (e.g., lethality, effects on neural progenitors) has resulted in an incomplete understanding of Hox gene functions in the nervous system. Emerging evidence in *C. elegans, Drosophila*, mice and humans shows that Hox genes

are expressed continuously, from development throughout adulthood, in certain neuron types [24-30]. However, the functional significance of sustained Hox gene expression in adult post-mitotic neurons remains unknown.

Individual neuron types acquire their NT identity during development and maintain it throughout life. Whether and how Hox proteins control NT identity is not known. Here, we show that the *C. elegans* Hox protein LIN-39 (Scr/Dfd/Hox4-5) is continuously required to maintain the expression of ACh pathway genes in cholinergic MNs, thereby securing their NT identity and function. In MNs that control locomotion, LIN-39 cooperates with another Hox protein MAB-5 (Antp/Hox6-8) and the transcription factor UNC-3 (Collier/Ebf) to directly activate the expression of ACh pathway genes. Moreover, LIN-39 and MAB-5 also regulate the expression levels of *unc-3*, thereby generating a positive feedforward loop that ensures robustness of ACh pathway gene expression throughout life. Importantly, we find that Hox gene expression is under homeostatic control and propose a two-component design principle that maintains optimal levels of Hox gene expression in MNs over time: Hox transcriptional autoregulation (component 1) is counterbalanced by negative UNC-3 feedback (component 2). Because Hox genes are expressed in the adult nervous system of invertebrate and vertebrate animals, their functional role on maintenance of NT identity may be deeply conserved.

3.3 Results

3.3.1 The Hox gene *lin-39 (Scr/Dfd/Hox4-5)* is required to maintain the cholinergic identity of motor neurons necessary for egg laying

We initially focused on the hermaphrodite-specific VC MNs, for which regulators of their cholinergic identity remain unknown. The Hox gene *lin-39* (*Scr/Dfd/Hox4-5*) is continuously expressed in VC neurons, from the time they are born (larval stage 1, L1) until adulthood (**Fig. 3.1B**) (**Fig. 3.8**)[31]. However, animals carrying strong loss-of-function alleles for *lin-39* lack VC neurons (*lin-39* controls VC survival) [32, 33], preventing us from testing its role in the regulation of VC cholinergic identity determinants *unc-17/VAChT* and *cho-1/ChT* (**Fig. 3.8**) [16]. To bypass this, we generated double mutant animals for *lin-39* and *ced-3* (VC cell death depends on *ced-3* caspase activity)[32, 33], and indeed observed that VC neurons are normally generated in these animals (**Fig. 3.1C**). Compared to controls, the number of VC neurons expressing *unc-17* and *cho-1* reporters was significantly reduced in *lin-39(n1760); ced-3(n1286)* double mutants in the adult (day 1) stage (**Fig. 3.1C-D**). Subsequent fluorescence intensity analysis with higher resolution revealed that expression of the cholinergic marker *unc-17::gfp* is significantly reduced, but not completely eliminated (**Fig. 3.1E**). For the *unc-17* and *cho-1* expression analysis, we used fosmid (~30kb-long genomic clones)-based reporters, which tend to faithfully recapitulate endogenous gene expression patterns [16]. Altogether, these findings suggest that, during larval development, *lin-39* controls the expression of VC cholinergic identity determinants (**Fig. 3.1F**).

The sustained expression of *lin-39* in VC MNs during adulthood raises the question of whether it is continuously required to maintain cholinergic identity. Because the *lin-39(n1760)* allele removes gene activity starting in early development, we used a previously validated *lin-*

39::mNG::3xFLAG::AID allele that codes for a LIN-39 protein fused to mNG::3xFLAG (serves as endogenous reporter) and auxin-inducible degron (AID) (enables LIN-39 depletion in a temporally-controlled manner) [34, 35] (**Fig. 3.1G**). We initiated auxin treatment at the last larval stage (L4) and maintained animals on auxin until the first day of adulthood (**Fig. 3.1H**). This led to efficient LIN-39 depletion in VC neurons and other nerve cord MNs that normally express *lin-39* (**Fig. 3.1H**). In day 1 adults, we witnessed a statistically significant decrease in the number of VC neurons expressing the *unc-17* reporter (**Fig. 3.1I**), indicating *lin-39* is required to maintain the expression of *unc-17/VAChT* at later life stages. Lastly, we examined two additional markers of VC terminal differentiation, *ida-1* (ortholog of human protein tyrosine phosphatase receptor type N [PTPRN])[36] and *glr-5* (ortholog of human glutamate ionotropic receptor GRIK1)[37]. Again, we found that *lin-39* is continuously required to maintain their expression during late larval and adult stages (**Fig. 3.8D-E**). In conclusion, *lin-39*, in addition to promoting VC survival during the L1 stage, is continuously required at later stages of life to control cholinergic identity (e.g., *unc-17*) and other features (*ida-1, glr-5*) of VC terminal differentiation (**Fig. 3.1F**).

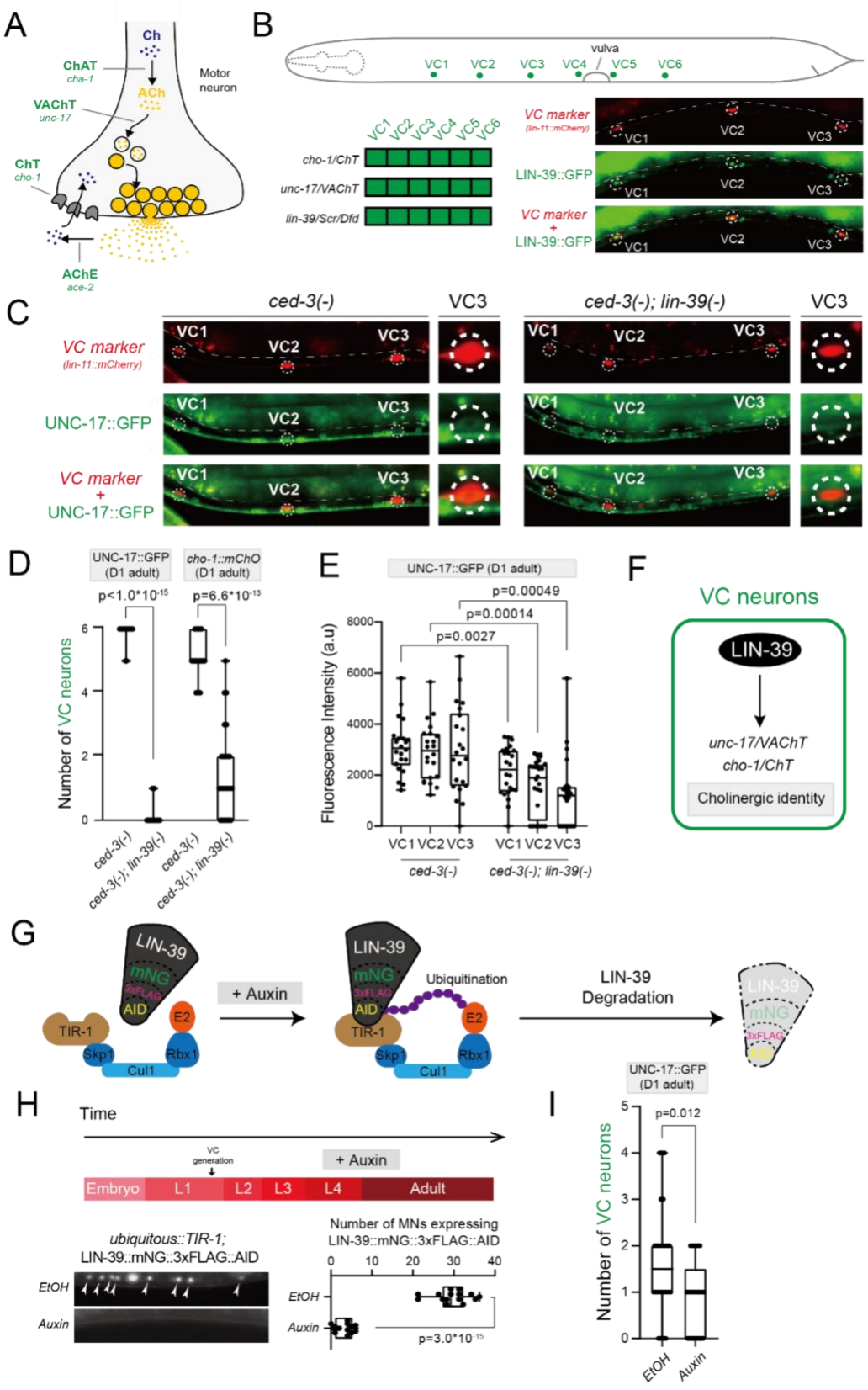

Figure 3.1 The Hox gene *lin-39 (Scr/Dfd/Hox4-5)* is required to maintain the cholinergic identity

of VC motor neurons. **A.** Schematic of the presynaptic bouton of a cholinergic motor neuron. Acetylcholine (Ach) pathway genes: *cha-1*/*ChAT*: Choline Acetyltransferase, *unc-17*/*VAChT*: Vesicular Acetylcholine Transporter, *cho-1*/*ChT*: Choline Transporter, *ace-2*/*AChE*: Acetylcholinesterase. **B.** Schematic of an adult *C. elegans* hermaphrodite highlighting the cell body position of the six VC motor neurons. Bottom left: Table summarizing co-expression of *cho-1, unc-17* and *lin-39* in all VC motor neurons. Bottom right: Representative image of a fosmid-based *lin-39* reporter *wgIs18* (LIN-39::GFP) in green channel and a VC specific marker (*lin-11::mCherry*) in red channel. Co-localization occurs in all six VC neurons, but only VC1-3 neurons are shown. **C.** Representative images of UNC-17::GFP co-expressed with *lin-11::mCherry* in *ced-3* (*n1286*) single mutant and *ced-3 (n1286); lin-39* (*n1760*) double mutant animals (day 1 adults). Inset shows expression in VC3. Dashed circles indicate cell body positions. Dashed white lines indicate the boundary of the intestine, which is auto-fluorescent in the green channel. **D.** Quantification of *unc-17* and *cho-1* reporter gene expression in *ced-3* (*n1286*) single mutant and *ced-3 (n1286); lin-39* (*n1760*) double mutant animals (day 1 adults). The fosmid-based reporter *otIs544* was used for *cho-1* (as *cho-1::SL2::mChOpti::H2B*). **E**. Quantification of fluorescence intensity of UNC-17::GFP in VC1-3 in *ced-3* (*n1286*) single and *ced-3 (n1286); lin-39* (*n1760*) double mutant animals. See Material and Methods for details (a.u, arbitrary units). **F**. Model of LIN-39 regulating cholinergic identity genes in VC MNs. **G**. Schematic of the auxin-inducible protein degradation system. Skp1, Cul1, Rbx1, and E2 are phylogenetically conserved components of the E3 ligase complex. Worms are expressing ubiquitously TIR1, a plant-specific substrate-recognizing subunit of the E3 ligase complex. In the presence of auxin, TIR1 binds to the AID fused to LIN-39, leading to ubiquitination and proteasomal degradation of LIN-39::mNG::3xFLAG::AID. **H.** Efficient depletion of LIN-39::mNG::3xFLAG::AID upon auxin treatment. Worms are grown on normal OP50 plates until L4 stages and then transferred to plates containing Auxin or EtOH (as control) and then scored 24 hours later at young adult stages. See Material and Methods for details. Representative images showing LIN-39::mNG::3xFLAG::AID expression in nerve cord motor neurons upon auxin or EtOH treatment with quantifications shown on the right. **I.** Quantification of the number of VC motor neurons expressing UNC-17::GFP upon auxin or EtOH treatment (Treatment initiated at L4 stage and images were scored at young adult stage). For quantification in **D**, **E, H** and **I**, box and whisker plots were used with presentation of all data points. Unpaired t-test with Welch's correction was performed and p-values were annotated. N>20.

3.3.2 The Hox gene *lin-39 (Scr/Dfd/Hox4-5)* controls the cholinergic identity of motor neurons necessary for locomotion

Prompted by our findings in MNs that control egg-laying, we next asked whether the control of NT identity by *lin-39* extends to MNs that control locomotion. Using the endogenous *lin-39::mNG::3xFLAG::AID* reporter allele, we found *lin-39* to be continuously expressed, from development to adulthood, in 28 of the 39 cholinergic MNs necessary for locomotion (**Fig. 3.2A**) (**Fig. 3.8B**) [31]. These neurons survive in animals carrying the strong loss-of-function *lin-39* allele *(n1760)* used in Figure 1, enabling assessment of putative *lin-39* effects on their cholinergic identity. Indeed, we found a significant decrease in the expression of *cho-1/ChT* in these MNs at larval stages (**Fig. 3.2B-C**). We obtained similar results for two additional cholinergic identity markers using an endogenous reporter for *unc-17 (unc-17::mKate)* and a fosmid-based reporter for *ace-2/AChE* (**Fig. 3.2D-E**), suggesting *lin-39* co-regulates the expression of several ACh pathway genes in MNs necessary for locomotion.

Due to its maintenance role in egg-laying MNs (**Fig. 3.1G-I**), we next asked whether *lin-39* is also required to maintain in the adult the cholinergic identity of MNs critical for locomotion. Again, we found that depletion of LIN-39 specifically during late larval and adult stages resulted in decreased expression of *unc-17/ChAT* (**Fig. 3.2F**), strongly suggesting a continuous LIN-39 requirement in these MNs.

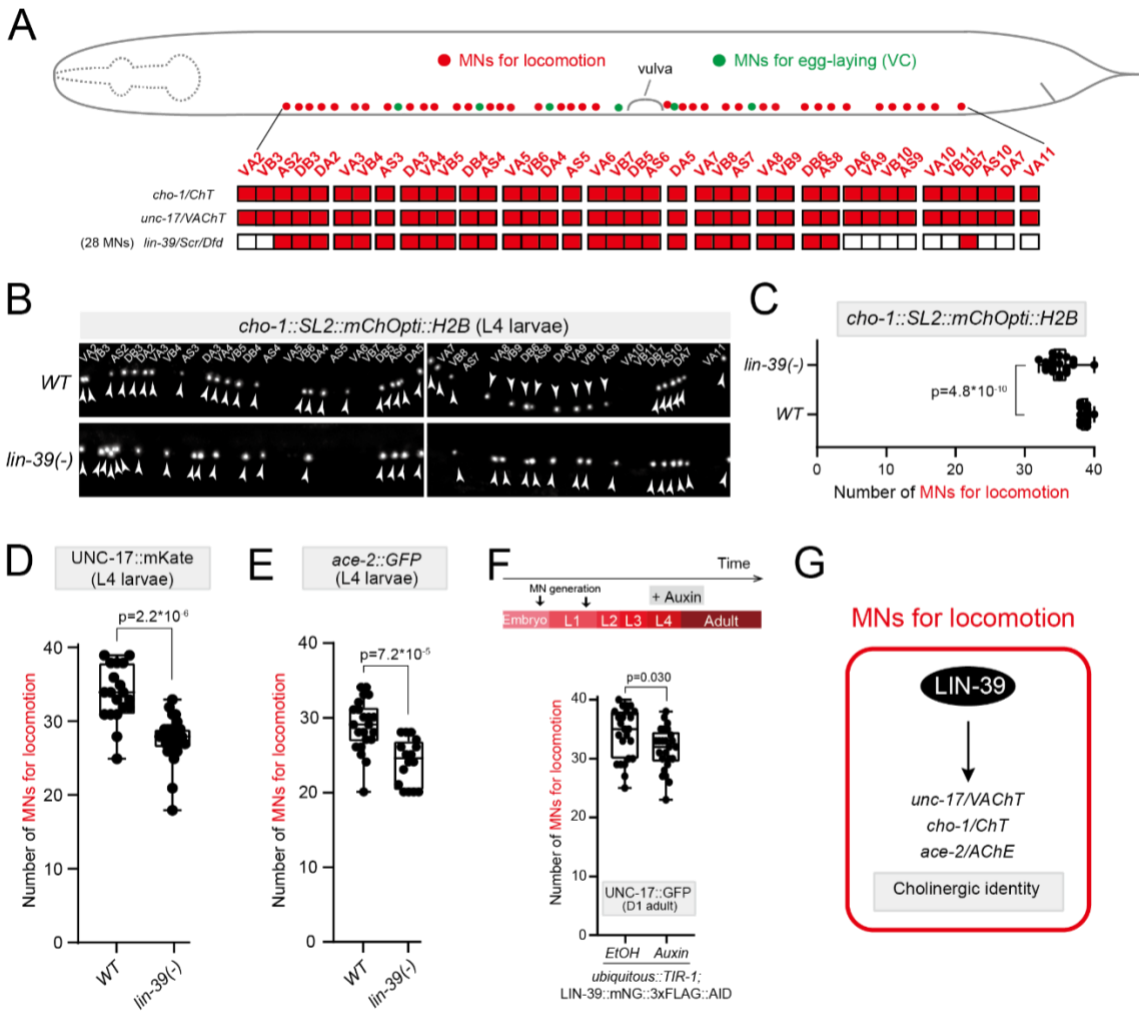

Figure 3.2 The Hox gene *lin-39 (Scr/Dfd/Hox4-5)* controls the cholinergic identity of motor neurons necessary for locomotion. **A.** Schematic of a *C. elegans* hermaphrodite showing the cell bodies of MNs that control locomotion (red) and egg-laying (green). Expression profile of *cho-1, unc-17 and lin-39* is summarized in the color-filled table. **B.** Representative images of a *cho-1* fosmid-based reporter in *WT* and *lin-39(n1760)* mutant animals (L4 stage). Motor neurons that control locomotion (e.g., DA, DB, VA, VB) express *cho-1* at L4. VC motor neurons mature later (day 1 adult) and do not express *cho-1* at L4. The identity of individual motor neurons is annotated. White arrowheads indicate MN nuclei because the *cho-1* reporter is localized to the nucleus. *mChOpti* signal is shown in white for better contrast. Left image: anterior to the vulva, Right image: posterior to the vulva. **C.** Quantification of the number of motor neurons expressing *cho-1* in panel **B**. **D-E**. Quantification of the number of sex-shared motor neurons expressing *unc-17* (endogenous reporter *ot907* [UNC-17::mKate]) and *ace-2* (fosmid-based reporter *otEx4432* [*ace-2::GFP]*) in WT and *lin-39(n1760)* mutant hermaphrodites at L4. **F.** Auxin treatment initiated at L4 and images were scored at young adult stage. Quantification of the number of sex-shared motor neurons expressing UNC-17::GFP upon auxin or EtOH treatment. For quantifications in **C-F**, box and whisker plots were used with presentation of all data points. Unpaired t-test with Welch's correction was performed and p-values were annotated. N>20. **G.** Schematic summarizing our findings in MNs that control locomotion

3.3.3 The Hox gene *lin-39* cooperates with *mab-5 (Antp, Hox6-8)* and *unc-3* (Collier/Ebf) to control the cholinergic identity of motor neurons

Compared to the strong effects observed in VC MNs (**Fig. 3.1D**), we observed modest effects on ACh pathway gene expression upon *lin-39* loss in MNs that control locomotion (**Fig. 3.2C-F**), suggesting additional factors partially compensate for *lin-39* in these neurons (**Fig. 3.2G**). To identify them, we followed a candidate approach. First, we reasoned that *mab-5* (*Antp, Hox6-8*), another Hox gene with continuous expression in a subset of *lin-39*-expressing MNs (**Fig. 3.3A**) (**Fig. 3.9**)[27], may work together with *lin-39* to control NT identity features. Indeed, we observed a decrease in the expression of the cholinergic marker *unc-17/VAChT* in double *lin-39(n1760); mab-5(e1239)* mutants compared to single *lin-39(n1760)* mutant animals (**Fig. 3.3D**). However, the residual expression of *unc-17/VAChT* in MNs of *lin-39; mab-5* double mutants suggests that, besides MAB-5, additional regulatory factors must cooperate with LIN-39. The transcription factor UNC-3, member of the Collier/Olf/Ebf (COE) family, is selectively expressed in MNs that control locomotion (**Fig. 3.3A**), and is the only factor previously known to control their cholinergic identity [15]. Hence, if *unc-3* compensates for the combined loss of *lin-39* and *mab-5*, we would predict stronger effects on ACh pathway gene expression in triple mutant *unc-3(n3435); lin-39(n1760); mab-5(e1239)* animals than in single *unc-3* or double *lin-39; mab-5* mutants. Indeed, this was the case for *cho-1/ChT* and *ace-2/AChE* expression (**Fig. 3.3B-C, E**). Because animals carrying the strong loss-of-function *unc-3* allele *(n3435)* [38] display a dramatic effect on *unc-17/VAChT* expression (UNC-3 is absolutely required for *unc-17/VAChT*) (**Fig. 3.3D**), we used a hypomorphic (weaker loss-of-function) *unc-3* allele *(ot837)*[34] to test the notion of cooperation. Indeed, we observed stronger effects on *unc-17/VAChT* expression in *unc-3 (ot837); lin-39(n1760); mab-5(e1239)* compared to *unc-3 (ot837)*

and double *lin-39; mab-5* mutants (**Fig. 3.3D**). We extended the analysis of the hypomorphic *unc-3* allele *(ot837)* to *cho-1/ChT,* and again observed stronger effects in triple mutant animals (**Fig. 3.3B-C, E**). Although the magnitude of the observed effects differs in Hox and *unc-3* mutants, this genetic analysis suggests that *lin-39*, *mab-5*, and *unc-3* synergize to control the expression of several ACh pathway genes in MNs necessary for *C. elegans* locomotion (**Fig. 3.3F**).

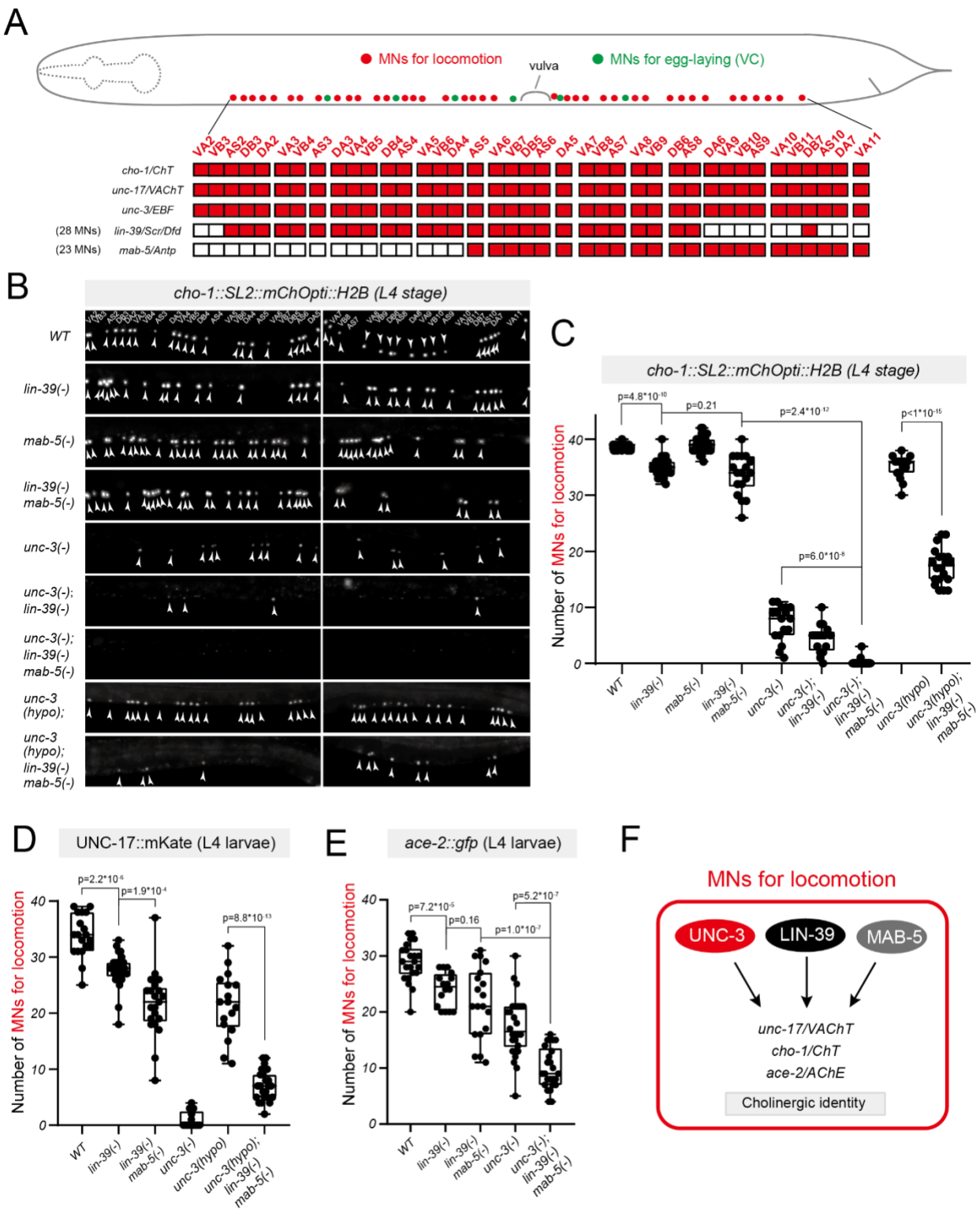

Figure 3.3 *lin-39* cooperates with *mab-5 (Antp, Hox6-8)* and *unc-3 (Collier/Ebf)* to control the cholinergic motor neuron identity. **A.** Schematic of a *C. elegans* hermaphrodite showing the cell bodies of MNs that control locomotion (red) and egg-laying (green). Expression profile of *cho-1*, *unc-17*, *unc-3*, *lin-39* and *mab-5* is illustrated in the color-filled table below. **B.** Representative

images of a *cho-1* fosmid-based reporter at L4 in the genetic background of *WT, lin-39(-), mab-5(-), lin-39(-) mab-5(-), unc-3(-), unc-3(-); lin-39(-), unc-3(-); lin-39(-) mab-5(-), unc-3(hypo), and unc-3(hypo); lin-39(-) mab-5(-).* Strong-loss-of-function (null) alleles of *lin-39(n1760), mab-5(e1239) and unc-3(n3435)* are annotated as *(-)* and the hypomorphic *unc-3* allele *ot837 [unc-3::mNG::AID]* as *(hypo*). The identity of individual motor neurons is annotated. White arrowheads indicate MN nuclei because the *cho-1* reporter is localized to the nucleus. *mChOpti* signal is shown in white for better contrast. Left image: anterior to the vulva, Right image: posterior to the vulva. **C**. Quantification of the number of motor neurons expressing *cho-1* in panel **B**. **D-E**. Quantification of the number of motor neurons (that control locomotion) expressing *unc-17* (endogenous reporter *ot907* [UNC-17::mKate]) and *ace-2* (fosmid-based reporter *otEx4432* [*ace-2::GFP*]) in multiple genetic backgrounds at L4 stage. For quantifications in **C-E**, box and whisker plots were used with presentation of all data points. Unpaired t-test with Welch's correction was performed and p-values were annotated. N>15. **F.** Schematic summarizing our findings in cholinergic motor neurons that control locomotion.

3.3.4 LIN-39 and UNC-3 act through distinct binding sites to activate ACh pathway gene expression in motor neurons

Interrogation of available ChIP-Seq datasets provided biochemical evidence of UNC-3 binding to the *cis*-regulatory regions of ACh pathway genes (*unc-17/VAChT, cho-1/ChT, ace-2/AChE)* (**Fig. 3.4A, D**) (**Fig. 3.10**) [39]. These bound regions contain cognate sites for UNC-3, and their previous mutational analysis strongly suggested UNC-3 acts directly to activate ACh pathway gene expression in cholinergic MNs[15]. Next, we examined available ChIP-Seq datasets for LIN-39 and MAB-5 [40], and witnessed coincident binding of LIN-39, MAB-5 and UNC-3 at *cis*-regulatory regions of *unc-17/VAChT, cho-1/ChT,* and *ace-2/AChE* (**Fig. 3.4A, D, Fig. 3.10**). We hypothesized that these co-bound regions (putative enhancers) are sufficient to drive reporter (*yfp*) gene expression in MNs. Indeed, a 280bp-long fragment of the *cho-1 cis*-regulatory region, as well as two fragments (1,000bp and 125bp) of the *unc-17* region, are sufficient to drive *yfp* expression in MNs (**Fig. 3.4A-B, D-F**). The *yfp* expression driven by these putative enhancer regions (*cho-1_280bp, unc-17_1,000bp, unc-17_125bp*) is *unc-3*- and Hox-dependent (**Fig. 3.4B-C, E-F**). Similar to our observations with an endogenous reporter for *unc-17* and a fosmid-

based reporter for *cho-1* (**Fig. 3.3C-D**), we corroborated the notion of synergy between UNC-3 and Hox by using the hypomorphic *unc-3* allele *(ot837)*. Stronger effects were observed in *unc-3 (ot837); lin-39; mab-5* triple mutants compared to *unc-3(ot837)* single or *lin-39; mab-5* double mutant animals (**Fig. 3.4C, E-F**). Altogether, our enhancer analysis in *unc-3* and Hox mutants identified specific *cis*-regulatory regions that require both UNC-3 and Hox (LIN-39, MAB-5) input to drive *unc-17/VAChT* and *cho-1/ChT* expression in cholinergic MNs.

Within the *cho-1_280bp* and *unc-17_125bp* enhancer regions, a single UNC-3 binding site (COE motif) is necessary for reporter gene expression in MNs (**Fig. 3.4A, D**) [15]. We bioinformatically searched for the presence of consensus LIN-39 binding sites (GATTGATG), which unlike MAB-5 sites, are well defined in *C. elegans* [40]. We found a single LIN-39 binding site 60bp upstream of the UNC-3 site in the *cho-1_280bp* region (**Fig. 3.4A**) and four LIN-39 sites flanking the UNC-3 binding site in the *unc-17_125bp* region (**Fig. 3.4D**). We tested the functional importance of the 4 LIN-39 sites by simultaneously deleting them in the context of transgenic reporter (*unc-17_125bp::yfp)* animals, and found a dramatic decrease in *yfp* expression in MNs (**Fig. 3.4D**). Altogether, this analysis combined with the ChIP-Seq results provide strong evidence that LIN-39 and UNC-3 bind directly to the *cis*-regulatory region of ACh pathway genes, and recognize distinct binding sites (**Fig. 3.4G**).

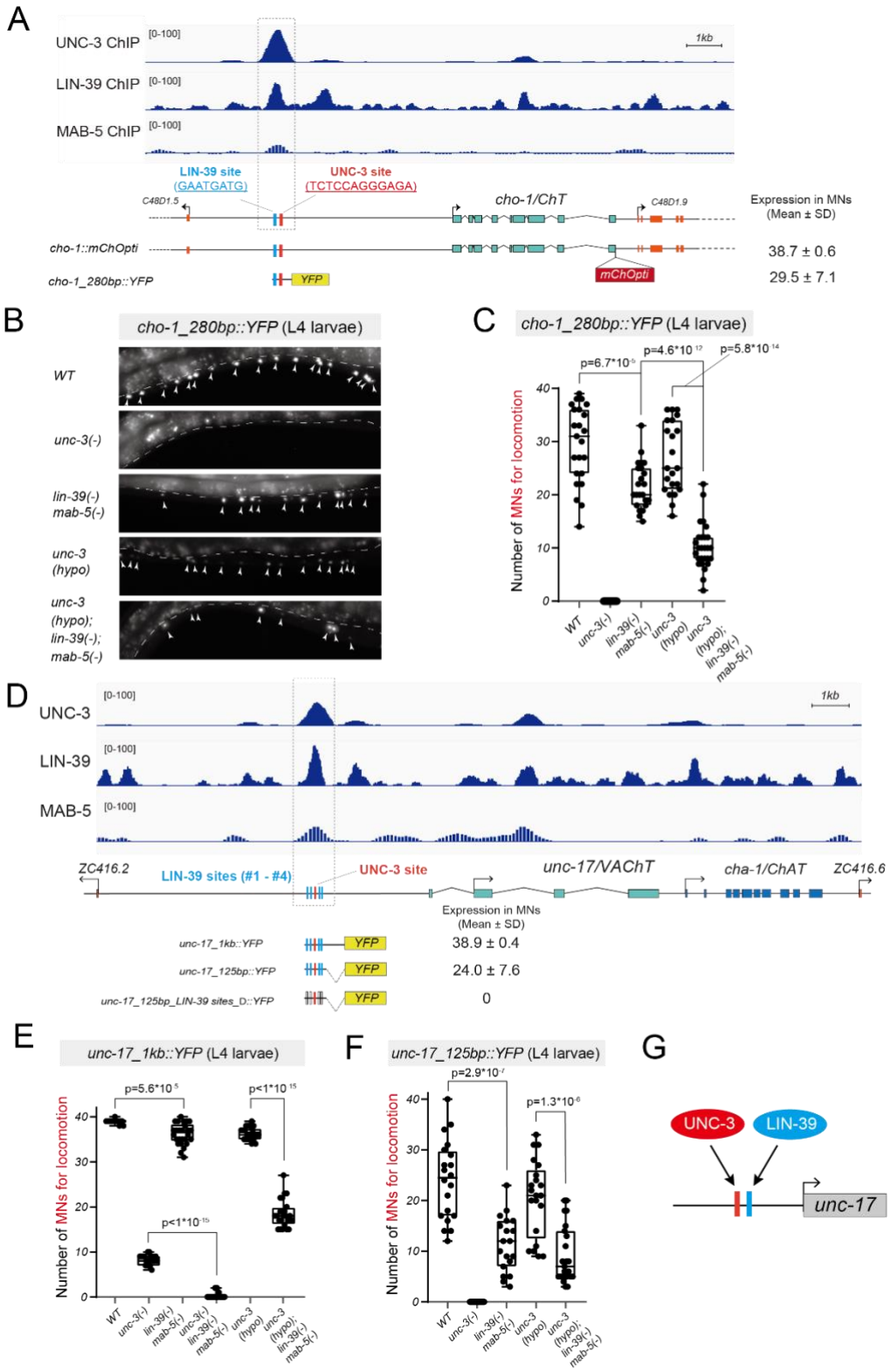

Figure 3.4 LIN-39 and UNC-3 act directly to activate expression of Ach pathway genes**. A.** Gene locus of *cho-1* with ChIP-seq tracks for UNC-3, LIN-39 and MAB-5. Molecular nature of

cho-1 reporter gene *cho-1::mChOpti* and *cho-1_280bp::YFP* is shown. Total number of MNs expressing each reporter is indicated in the format of Mean ± Standard Derivation (SD). Within a 280bp *cis*-regulatory region (grey dashed line), an UNC-3 site and a predicted LIN-39 site are highlighted in red and blue, respectively. **B.** Representative images of *cho-1_280bp::YFP* expression at L4 in *WT, unc-3(n3435), lin-39(n1760) mab-5(e1239), unc-3(hypo), and unc-3(hypo); lin-39(n1760); mab-5(e1239)* animals. Arrowheads indicate MN nuclei because *cho-1_280bp::YFP* has a nuclear localization signal (NLS). YFP signal is shown in white for better contrast. Dashed white lines indicate the boundary of the intestine, which is auto-fluorescent in the green channel. **C**. Quantification of the number of motor neurons expressing *cho-1 reporter* in panel **B**. **D.** Gene locus of *cho-1* with ChIP-seq tracks for UNC-3, LIN-39 and MAB-5. Expression in motor neurons is quantified for *unc-17* reporter gene *unc-17_1kb::YFP* and *unc-17_125bp::YFP* as well as its mutated version with LIN-39 site deletions. Within a 125bp *cis*-regulatory region framed by grey dashed line, an UNC-3 site and 4 predicted LIN-39 sites are highlighted in red and blue, respectively. **E-F**. Quantification of the number of motor neurons at L4 expressing the *unc-17_1kb::YFP* (**E**) and *unc-17_125bp::YFP* (**F**) reporters in *WT, unc-3(n3435), lin-39(n1760); mab-5(e1239), unc-3(n3435), lin-39(n1760); mab-5(e1239), unc-3(hypo), and unc-3(hypo); lin-39(n1760); mab-5(e1239)* animals. For all quantifications, box and whisker plots were used with presentation of all data points. Unpaired t-test with Welch's correction was performed and p-values were annotated. N>15. **G.** Schematic of *unc-17* regulation by UNC-3, LIN-39 and MAB-5. Red and blue rectangles indicate the binding sites of UNC-3 and LIN-39, respectively.

3.3.5 LIN-39 and MAB-5 control *unc-3* expression levels, thereby generating a positive feed-forward loop (FFL) for the control of cholinergic identity

Our findings suggest that Hox proteins LIN-39 and MAB-5, like UNC-3, directly control the expression of ACh pathway genes in MNs. However, LIN-39 and MAB-5 also bind extensively on the *cis*-regulatory region of *unc-3* (**Fig. 3.5A**), suggesting they directly regulate its expression as well. Indeed, expression of an endogenous *gfp* reporter for *unc-3* (UNC-3::GFP protein fusion) is significantly reduced in MNs of *lin-39* single mutants, and this effect is exacerbated in *lin-39; mab-5* double mutants (**Fig. 3.5B-C**). The number of MNs that express UNC-3::GFP is lower in *lin-39* and *lin-39; mab-5* animals (left graph in **Fig. 3.5C**). Compared to WT, we also observed a striking reduction in UNC-3::GFP fluorescence intensity in MNs of Hox mutant animals (34% reduction in *lin-39* and 43% in *lin-39; mab-5*) (middle graph in **Fig. 3.5C**), suggesting *lin-39* and *mab-5* are required for normal levels of UNC-3 expression.

The binding of LIN-39 and MAB-5 on the *unc-3* locus strongly suggests these Hox proteins control *unc-3* expression by acting at the level of transcription. Four lines of evidence support this possibility. First, CRISPR/Cas9-mediated mutation of a predicted LIN-39 binding site in the context of the endogenous *unc-3::gfp* reporter (UNC-3::GFP $^{\text{LIN-39 site MUT}}$) resulted in decreased expression in MNs (right graph in **Fig. 3.5C**). Second, we quantified endogenous *unc-3* mRNA levels with single cell resolution using RNA fluorescent *in situ* hybridization (RNA FISH)[41]. Compared to wild-type animals, we observed a consistent decrease of *unc-3* mRNA levels in cholinergic MNs of *lin-39; mab-5* mutants (**Fig. 3.5D**). Third, a small 223bp-long *cis*-regulatory element upstream of the first *unc-3* exon is sufficient to drive reporter gene (*mCherry*) expression in wild-type MNs (**Fig. 3.5A, E**). When we quantified numbers of MNs expressing this reporter, as well as its fluorescence intensity, we observed statistically significant differences

in Hox mutant animals (**Fig. 3.5F-G**). Like our observations with the UNC-3::GFP fusion (**Fig. 3.5B-C**), we observed a 28% reduction in fluorescence intensity of the *unc-3_223bp::mCherry* reporter in *lin-39* single mutants, which was exacerbated (34% reduction) in *lin-39; mab-5* double mutants (**Fig. 3.5G**). Lastly, we generated two fosmid-based reporters, in which we replaced two different parts of the *unc-3* locus (exon 3-11 [□Exon3-11], exon7-11 [□Exon7-11]) with *mCherry* while leaving the upstream and downstream *cis*-regulatory elements intact (**Fig. 3.5A**). In both cases, we found a significant reduction of *mCherry* expression in MNs of *lin-39; mab-5* double mutant animals (**Fig. 3.5H-I**).

Altogether, Hox proteins (LIN-39, MAB-5) act directly to activate the expression of ACh pathway genes in MNs (Hox → ACh genes, **Fig. 3.3-3.4**). Further, Hox proteins act at the level of transcription to control *unc-3*, which also controls ACh genes (Hox → *unc-3* → ACh genes), thereby generating a positive feed-forward loop (FFL) (**Fig. 3.5J**).

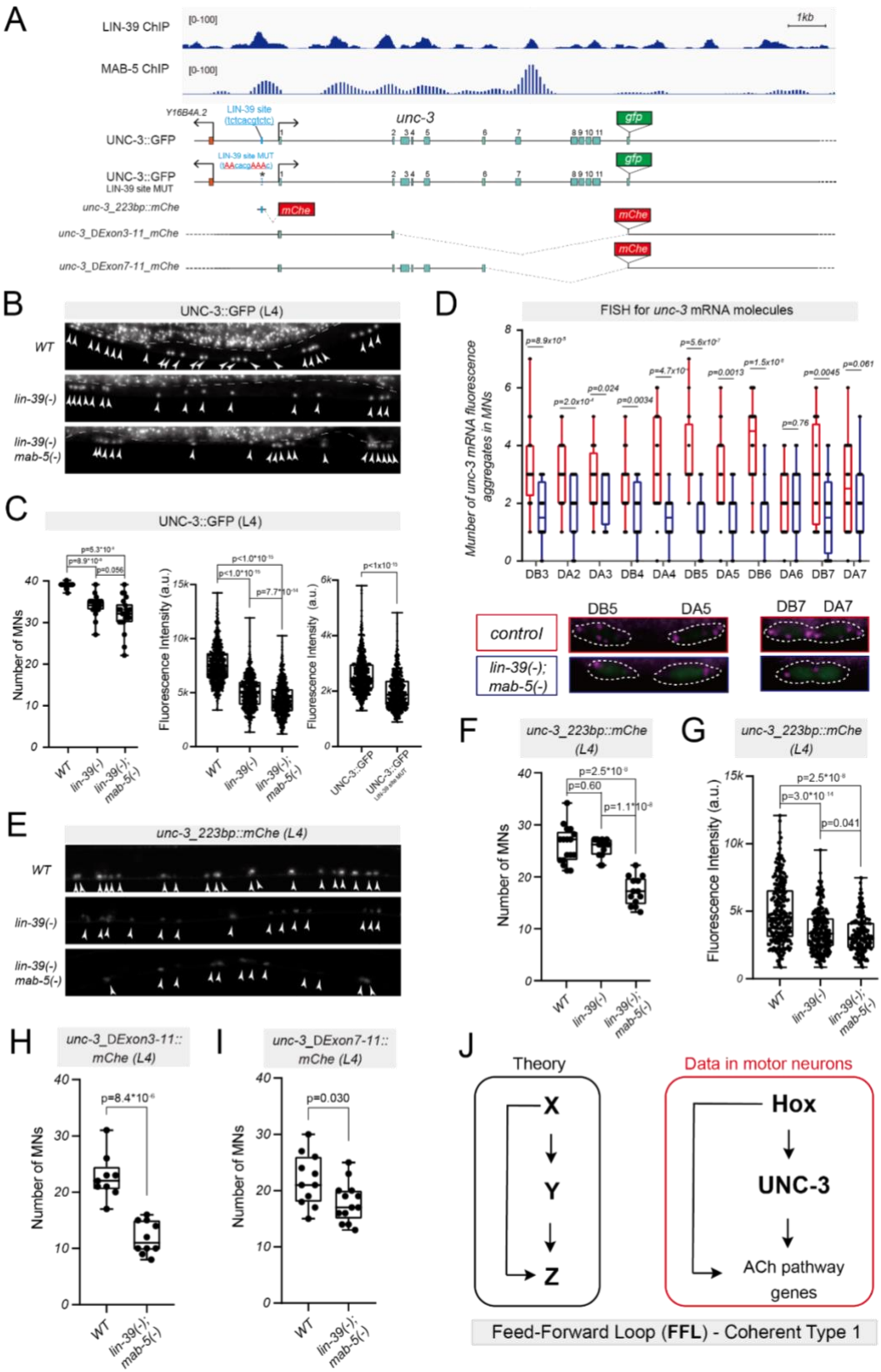

Figure 3.5 LIN-39 and MAB-5 control *unc-3* expression levels. **A.** Gene locus of *unc-3* with

ChIP-seq tracks for LIN-39 and MAB-5. LIN-39 and MAB-5. Molecular nature of endogenous *unc-3* reporters (UNC-3::GFP and UNC-3::GFP $^{\text{LIN-39 site MUT}}$) and transgenic *unc-3* reporters *(unc-3_223bp::mChe, unc-3_ΔExon3-11_mChe, unc-3_ΔExon7-11_mChe)* is shown. **B.** Representative images of L4 animals carrying the endogenous *unc-3::gfp* reporter in *WT, lin-39(n1760),* and *lin-39(n1760); mab-5(e1239)* backgrounds. N>20. **C.** Quantification of the number of MNs (left graph) or fluorescence intensity (middle graph) of UNC-3::GFP in *WT, lin-39(n1760), and lin-39(n1760); mab-5(e1239)* backgrounds. N>20. Right graph: Quantification of the number of MNs (at L4) expressing GFP in animals carrying the *UNC-3::G*FP and *UNC-3::GFP* $^{\textit{LIN-39 site MUT}}$ endogenous reporters. N=20. **D.** Quantification of the number of *unc-3* mRNA fluorescence aggregates in individual cholinergic MNs. Single-molecule mRNA FISH for *unc-3 (*using a Quasar 670 fluorescent probe) was performed in WT and *lin-39(n1760); mab-5(e1239)* animals at L1. For all quantifications, box and whisker plots were used with presentation of all data points. Unpaired t-test with Welch's correction was performed and p-values were annotated. N>20. Representative images of DB5, DA5, DB7, and DA7 are shown. WT data is colored red versus mutant in blue for quantification. Images are shown in merged channels. Magenta: signal from Quasar 670 probe for *unc-3* mRNA, Green: GFP reporter for cholinergic motor neurons (*unc-17::gfp*). **E.** Representative images of *unc-3_223bp::mChe* expression in L4 hermaphrodites in the genetic background of *WT, lin-39(n1760),* and *lin-39(n1760); mab-5(e1239)* **F-G.** Quantification of the number of MNs (**F**) or fluorescence intensity (**G**) of *unc-3_223bp::mChe* in panel **E**. For fluorescence intensity quantifications, single neurons expressing the mCherry reporters in WT and mutant backgrounds are evaluated for fluorescence intensity and plotted as individual data point in graph. N>600. In panel **G,** N>200. **H-I**. Quantification of the number of MNs expressing the *unc-3_ΔExon3-11_mChe* (**H**), and *unc-3_ΔExon7-11_mChe* (**I**) *in WT and lin-39(n1760); mab-5(e1239)* mutant animals. N=10. For all quantifications, box and whisker plots were used with presentation of all data points. Unpaired t-test with Welch's correction was performed and p-values were annotated. **J.** Schematic model of a coherent feed-forward loop (FFL) for the control of cholinergic identity by Hox and UNC-3.

3.3.6 Hox genes *lin-39* and *mab-5* maintain their expression in motor neurons through transcriptional autoregulation

How is Hox gene expression maintained in adult MNs? Available ChIP-Seq data show LIN-39 binding at its own locus (**Fig. 3.6A**), raising the possibility of *lin-39* maintaining its own expression in MNs through transcriptional autoregulation. To test this, we first determined which *cis*-regulatory elements of the *lin-39* locus are sufficient to drive reporter (*TagRFP*) expression in MNs. A 6.2kb element upstream of the locus fused to *TagRFP* only shows low levels of expression in a small number of nerve cord MNs (~6) during larval stages (**Fig. 3.6A**). This prompted us to study intronic elements. When intron 1 of *lin-39* was fused to *TagRFP*, we detected robust *TagRFP* expression in ~23 nerve cord MNs. These *lin-39* $^{intron\ 1}$*::TagRFP* animals display continuous expression in both larval and adult MNs (**Fig. 3.6A-B**), similar to animals carrying the endogenous *lin-39* reporter allele (*lin-39::mNG::3xFLAG::AID*), enabling us to test the idea of transcriptional autoregulation. Indeed, the number of *TagRFP* expressing MNs is significantly reduced in *lin-39* homozygous mutants carrying a null allele (*n1760*) (**Fig. 3.6B**), suggesting *lin-39* gene activity is necessary for its expression in MNs.

Because the *lin-39 (n1760)* allele removes gene activity starting in early embryo, the above findings do not address whether LIN-39 is continuously required to maintain its expression at later stages. To test this, we again used the AID system and efficiently depleted LIN-39 protein levels during late larval stages (**Fig. 3.6C**). Upon auxin administration, we found a significant reduction in the number of MNs expressing the *lin-39* $^{intron\ 1}$*::TagRFP* reporter (**Fig. 3.6D**), suggesting *lin-39* is continuously required to maintain its own expression.

These findings pinpoint the entirety of intron 1 as a putative *cis*-regulatory region used by LIN-39 to directly control its own expression. Next, we generated transgenic reporter animals that

split intron 1 in two fragments (832bp and 776bp). Animals carrying the 776bp fragment (*lin-39* ^{776bp}*::TagRFP)* revealed reporter gene expression in nerve cord MNs, but that was not the case for the 832 bp fragment (**Fig. 3.6A**). Within the 776bp fragment, we found eleven predicted LIN-39 binding sites (motifs) (**Fig. 3.6A**), and employed CRISPR/Cas9 gene editing to simultaneously mutate all of them in the context of the endogenous *lin-39::mNG::3xFLAG::AID* reporter allele. This led not only to a decrease in the number of MNs expressing the endogenous *lin-39::mNG::3xFLAG::AID* $^{sites\ 1\text{-}11\ MUT}$ reporter (**Fig. 3.6E**), but also in a reduction of its expression levels (**Fig. 3.6F**). These findings support a transcriptional autoregulation model where LIN-39 recognizes its cognate binding sites within intron 1 to regulate its expression in nerve cord MNs (**Fig. 3.6H**).

To investigate which of the 11 LIN-39 sites are functionally important, we mutated two sites with the highest bioinformatic prediction scores (sites #1 and #9 in **Fig. 3.6A**, see Materials and Methods). Mutation of either site in the context of *lin-39* ^{776bp}*::TagRFP* animals significantly decreased the number of *TagRFP* expressing MNs (**Fig. 3.6G**), indicating that these sites are necessary for autoregulation.

Lastly, we asked whether the Hox gene *mab-5*, like *lin-39*, controls its own expression in MNs. Supporting this possibility, ChIP-Seq data show extensive MAB-5 binding at its own locus, and expression of a *mab-5::gfp* reporter is significantly reduced in MNs of *mab-5* mutant animals (**Fig 3.11**). Altogether, these findings uncover a positive feedback mechanism (transcriptional autoregulation) required to maintain Hox gene expression in MNs (**Fig. 3.6I**).

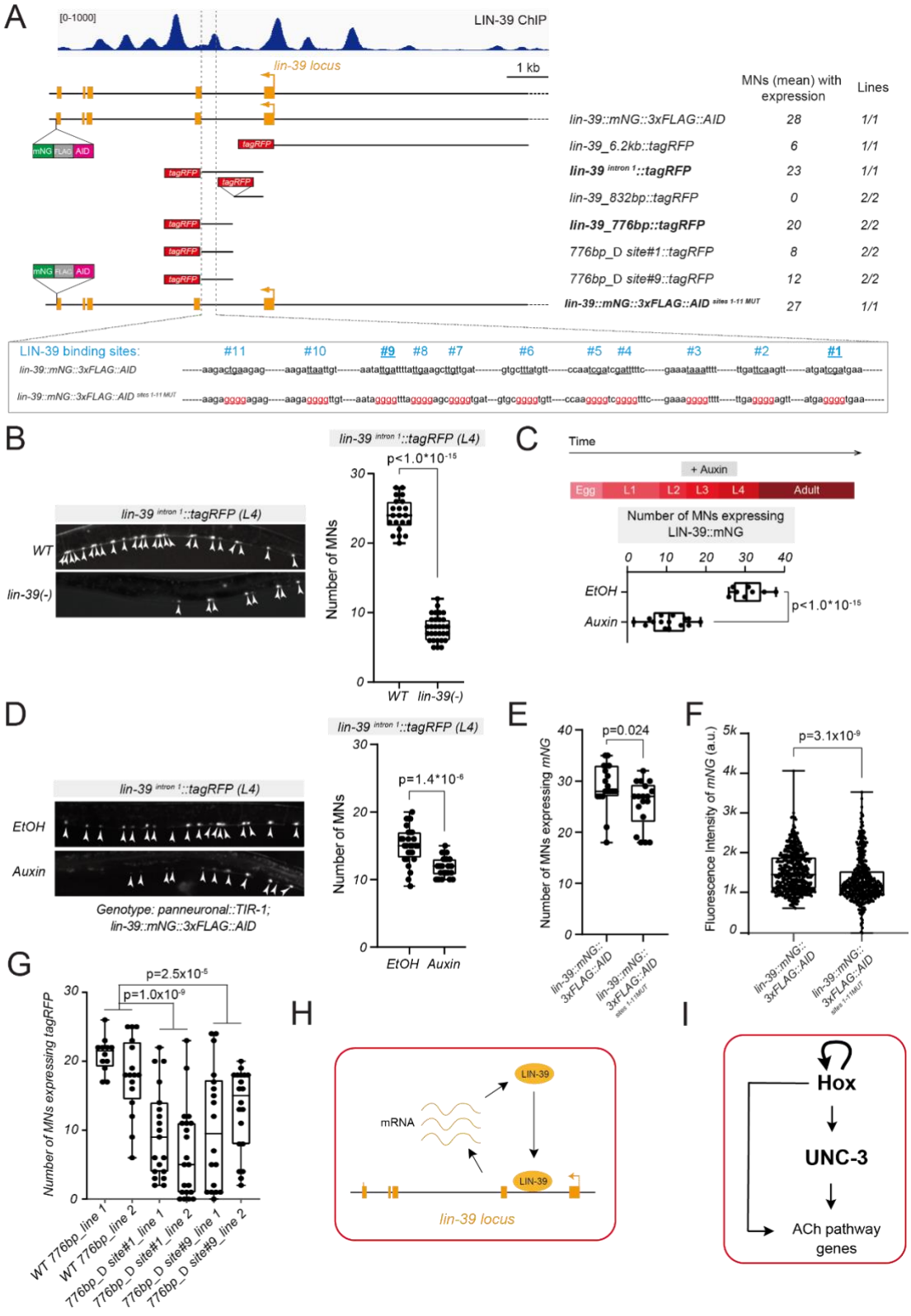

Figure 3.6 Transcriptional autoregulation of *lin-39* in motor neurons**. A.** Gene locus of *lin-39*

with ChIP-seq tracks for LIN-39. Molecular nature of endogenous *lin-39* reporters (*lin-39::mNG::3xFLAG::AID* and *lin-39::mNG::3xFLAG::AID*$^{sites1\text{-}11MUT}$*)* and transgenic *lin-39* reporters *(lin-39_6.2kb::tagRFP, lin-39*^{intron1}*::tagRFP, lin-39_832bp::tagRFP, lin-39_776bp::tagRFP, 776bp_□site#1::tagRFP, 776bp_□site#9::tagRFP*). The mean number of MNs that express each reporter is indicated on the right side. The precise location of the 11 predicted LIN-39 binding sites is shown, as well as their mutated version. **B**. Representative images and corresponding quantification of the number of MNs expressing *lin-39*^{intron1}*::tagRFP* in *WT* and *lin-39(n1760)* mutants at L4. **C**. Depletion of LIN-39::mNG::3xFLAG::AID upon auxin treatment. Worms expressing the *lin-39::mNG::3xFLAG::AID* together with TIR protein in all neurons (driven by pan-neuronal promoter *otti28*) were grown on normal OP50 plates until L2 stage, and then transferred to plates containing either auxin or EtOH (as control). Animals wre imaged 36 hours later at L4. See Material and Methods for more details. Quantification of LIN-39::mNG::3xFLAG::AID expression in motor neurons upon Auxin or EtOH treatment. **D**. Representative images and corresponding quantification of the number of MNs expressing *lin-39*^{intron1}*::tagRFP* treated with EtOH or auxin as described in panel **C**. **E-F**. Quantification (L4 stage) of the number of MNs expressing *mNG* (**E**) or the fluorescence intensity (**F**) of *mNG* in animals carrying the *lin-39::mNG::3xFLAG::AID* and *lin-39::mNG::3xFLAG::AID*$^{sites1\text{-}11MUT}$ alleles. In panel **F**, single neurons expressing *mNG* are evaluated for fluorescence intensity and plotted as individual data points. N>400. a.u., arbitrary units. **G**. Quantification of the number of MNs expressing the *lin-39_776bp::tagRFP* reporter (intact version and mutated version with LIN-39 site#1 or #9 mutated). Two independent transgenic lines for each version of the *lin-39_776bp::tagRFP* reporter were used for this analysis. For all quantifications, box and whisker plots were used with presentation of all data points. Unpaired t-test with Welch's correction was performed and p-values were annotated. N>15. **H.** Schematic model of *lin-39* transcriptional autoregulation. **I.** Schematic model showing the autoregulation of *lin-39* embedded at the top of a coherent feed-forward loop.

3.3.7 UNC-3 (Collier/Ebf) prevents high levels of Hox gene (*lin-39, mab-5*) expression

Low levels of LIN-39 lead to a failure to maintain ACh pathway gene expression (**Fig. 3.1-3.2**), whereas high levels of LIN-39 in cholinergic MNs result in locomotion defects [34]. Hence, we hypothesized that the levels of *lin-39* in cholinergic MNs must be tightly controlled to avoid detrimental effects on MN function (see Discussion). The findings described below suggest that tight control of Hox gene expression in MNs is achieved through a two-component mechanism, that is, Hox transcriptional autoregulation is counterbalanced by negative UNC-3 feedback. Because UNC-3 ChIP binding peaks are observed in *lin-39* and *mab-5* loci (**Fig. 3.7A, Fig. 3.12A**), we reasoned UNC-3 acts directly to prevent high levels of Hox gene expression in cholinergic MNs. Multiple lines of evidence support this idea. First, the expression levels of the endogenous *lin-39::mNG::3xFLAG::AID* reporter is increased in MNs of *unc-3* mutant animals (**Fig. 3.7B-C**). Second, the *lin-39* $^{intron\ 1}$*::TagRFP* reporter showed increased *TagRFP* expression in MNs of *unc-3* mutants (**Fig. 3.7D-E**). Lastly, RNA FISH revealed increased levels of *lin-39* mRNA molecules in individual MNs of *unc-3* mutants (**Fig. 3.7F**). Importantly, we extended this analysis to *mab-5*, and obtained similar results; *unc-3* gene activity is necessary to prevent high levels of *mab-5* expression in MNs (**Fig. 3.12B-C**).

Taken together, our findings uncover an intricate gene regulatory network for maintenance of cholinergic identity in *C. elegans* MNs. This network contains a positive FFL, where Hox proteins (LIN-39 and MAB-5) not only activate the expression of ACh pathway genes, but also control the expression of *unc-3*, a critical regulator of cholinergic identity in MNs (**Fig. 3.7G**). Moreover, the top level of this FFL, i.e., Hox, is under homeostatic control, ensuring that optimal levels of Hox gene expression are maintained in MNs. This is achieved through Hox transcriptional autoregulation counterbalanced by negative UNC-3 feedback (**Fig. 3.7G**).

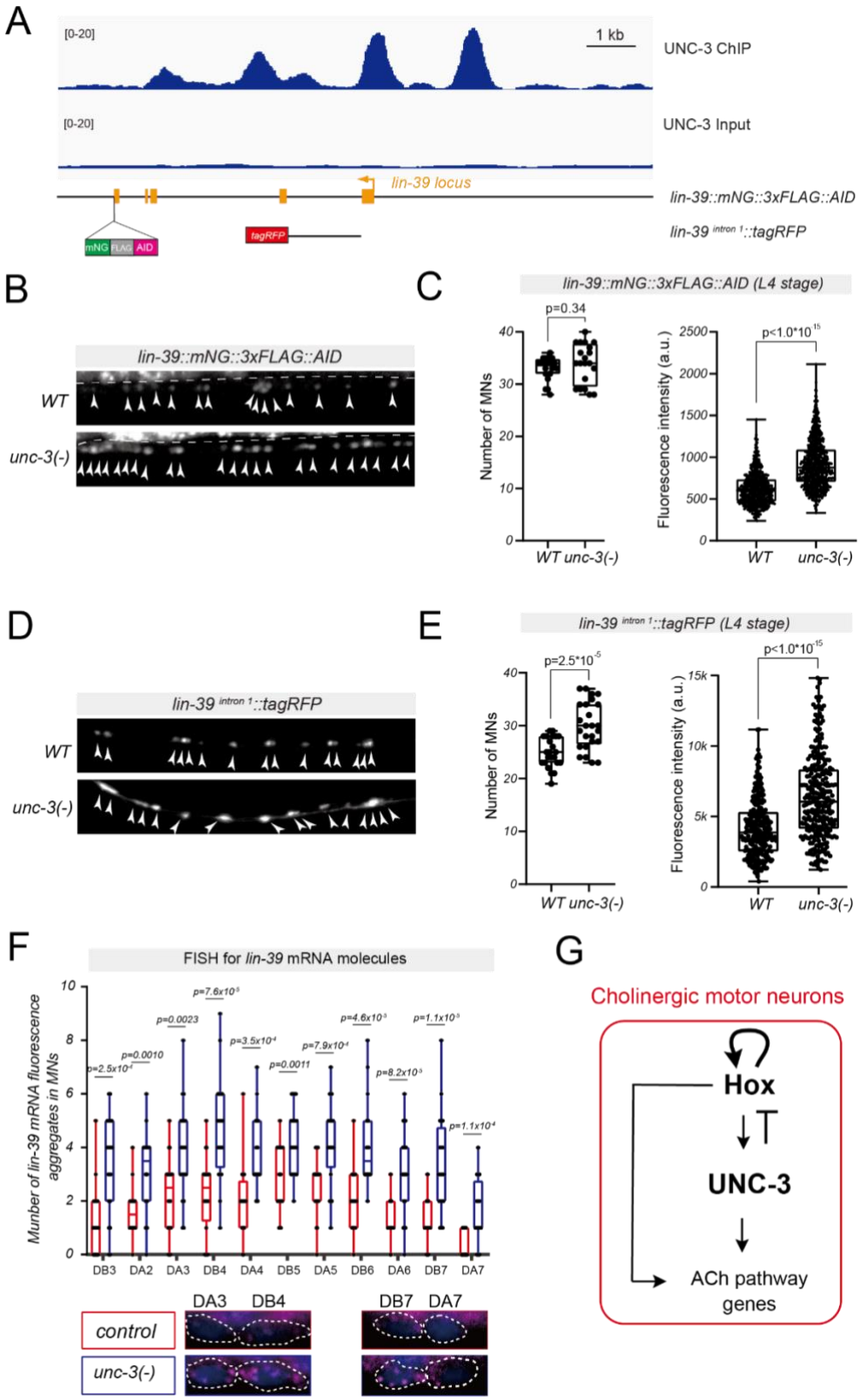

Figure 3.7 UNC-3 (Collier/Ebf) prevents high levels of *lin-39* expression. **A.** Gene locus of *lin-*

39 with ChIP-seq tracks for UNC-3. Molecular nature of *lin-39* reporter gene *lin-39::mNG::3xFLAG::AID* and *lin-39^{intron1}::tagRFP* is shown. **B-C**. Representative images (**B**) and quantifications (**C**) of the number of MNs expressing *lin-39::mNG::3xFLAG::AID* and the fluorescence intensity in *WT* and *unc-3(n3435)* animals at L4. **D-E**. Representative images (**D**) and quantifications (**E**) of the number and the fluorescence intensity of MNs expressing *lin-39^{intron1}::tagRFP* in *WT* versus *unc-3(n3435)* animals at L4. In panels **C** and **E**, single neurons expressing the indicated reporter gene are evaluated for fluorescence intensity and plotted as individual data point in graph. N>600. a.u., arbitrary unit. **F**. Quantification of the number of *lin-39* mRNA fluorescence aggregates in individual cholinergic MNs. Single-molecule mRNA FISH for *lin-39 (*using a Quasar 670 fluorescent probe) was performed in WT and *unc-3(n3435)* animals at L1. For all quantifications, box and whisker plots were used with presentation of all data points. Unpaired t-test with Welch's correction was performed and p-values were annotated. N>20. Representative images of DA3, DB4, DB7, and DA7 are shown. WT data is colored red versus mutant in blue for quantification. Images are shown in merged channels. Magenta: signal from Quasar 670 probe for lin-39 mRNA, Blue: DAPI for nuclei staining. **G.** Schematic model of the gene regulatory network for maintenance of cholinergic identity in *C. elegans* motor neurons

3.4 Discussion

Hox genes play fundamental roles in early patterning and body plan formation, but their functions in later stages of development and post-embryonic life remain largely unknown. In the context of the nervous system, Hox genes are known to control early events, such as specification, survival, and/or migration of progenitor cells and young post-mitotic neurons [19-23]. Here, we identify a new role for Hox proteins during later stages of nervous system development and post-embryonic life, critically extending their functional repertoire beyond early patterning. Using *C. elegans* nerve cord MNs as a model, we found that Hox gene activity is continuously required from development through adulthood for the control of NT identity, a core element of neuronal function. Moreover, Hox genes operate in a positive feedforward loop to safeguard the cholinergic identity of MNs: they not only control directly ACh pathway genes (e.g., *unc-17/VAChT, cho-1/ChT*), but also the expression of *unc-3* (*Collier/Ebf*), a master regulator of MN identity [15]. Lastly, we propose a homeostatic mechanism that maintains Hox gene expression in MNs at optimal levels, thereby ensuring robust expression of cholinergic identity determinants throughout life.

3.4.1 Hox genes *lin-39* (*Scr/Dfd, Hox4-5*) and *mab-5 (Antp, Hox6-8)* act as terminal selectors of cholinergic MN identity

Transcription factors that are continuously expressed in individual neuron types and act directly to co-regulate the expression of NT pathway genes have been termed “terminal selectors" [42-44]. Terminal selectors have been described to date for various neuron types in *C. elegans*, flies, simple chordates and mice, indicating the evolutionary conservation of terminal selector-based

mechanisms for the control of NT identity [45, 46]. The sustained expression of terminal selectors in specific neuron types suggests they are continuously required to maintain NT identity. However, a continuous requirement has been experimentally demonstrated only for a handful of terminal selectors to date [47].

Recent work in *C. elegans*, flies and mice indicates that certain Hox genes are continuously expressed, from development through adulthood, in distinct neuron types [24-30]. Functional studies in peptidergic neurons in *Drosophila*, as well as hindbrain and spinal neurons in mice, showed that Hox proteins control early facets of neuronal development (e.g., specification, survival, migration, connectivity) [19-23, 48], but whether they function as terminal selectors remains unknown - in part due to a lack of temporally controlled studies for Hox gene inactivation later in life. Through constitutive (genetic null alleles) and post-embryonic (inducible protein depletion) approaches, we found that the Hox gene *lin-39* is continuously required to maintain the cholinergic identity of *C. elegans* MNs. Biochemical (ChIP-Seq) and genetic evidence strongly suggests that LIN-39 and another Hox protein (MAB-5) act directly to co-regulate the expression of genes that encode integral components of the ACh biosynthetic pathway (*unc-17/VAChT, cho-1/ChT, ace-2/AChE*). Collectively, these findings reveal that Hox proteins can act as terminal selectors, uncovering a new role for these highly conserved transcription factors.

Because Hox genes are expressed in various neuron types in *C. elegans* [30], we surmise they may function as terminal selectors in other neurons as well. Supporting this notion, expression of a single marker of serotonergic identity (*tph-1*, tryptophan hydroxylase 1 [TPH1]) is affected in CP neurons of *lin-39* mutant animals [49, 50]. Similarly, expression of a dopaminergic identity marker (*cat-2*/tyrosine hydroxylase [TH]) is affected in tail sensory neurons of animals lacking

activity of the posterior Hox gene *egl-5* (*AbdB*) [51]. Future work is needed, however, to establish whether in those neurons LIN-39 and EGL-5 act as *bona fide* terminal selectors.

3.4.2 Hox and other subfamilies of homeodomain proteins act as terminal selectors

Homeodomain proteins are defined by a 60 amino acid motif (homeodomain) that directly contacts DNA [52]. Based on sequence similarities and/or the presence of other domains, several subfamilies of homeodomain proteins (e.g., HOX, LIM, POU, PRD) have been identified in every animal genome. In *C. elegans*, accumulating evidence suggests that members of LIM, POU, and PRD subfamilies can function as terminal selectors in specific neuron types [53] [31]. Here, we show that members of the HOX subfamily can act as terminal selectors in cholinergic MNs. A synthesis of our findings and previous studies on LIM, POU, and PRD proteins reveals an overarching theme: a single transcription factor family, the homeodomain proteins, is broadly used in the *C. elegans* nervous system to control neuronal identity by acting as terminal selectors. Supporting the evolutionary conservation of this theme, a number of LIM and POU homeodomain transcription factors can function as terminal selectors in the mouse nervous system [9, 10, 54-56].

3.4.3 Hox proteins cooperate with the terminal selector UNC-3 (Collier/Ebf) to control motor neuron cholinergic identity

Besides NT identity, terminal selectors are known to control additional neuron type-specific features of terminal differentiation, such as expression of ion channels, neuropeptides, NT receptors, and cell adhesion molecules [42, 44]. We find this to be the case for LIN-39. In sex-

specific MNs that control egg laying, LIN-39 is required for continuous expression of ACh pathway genes and additional terminal differentiation markers (e.g., *glr-5/GluR, ida-1/PTPRN*). Similarly, in MNs necessary for locomotion, LIN-39 controls ACh pathway genes (this study) and a handful of terminal differentiation markers (e.g., *del-1*/sodium channel SCNN1, *slo-2*/potassium-sodium activated channel KCNT) [27, 34]. But how can the same Hox protein (LIN-39) operate as a terminal selector in two different types of MNs?

Our findings support the idea of LIN-39 cooperating with distinct transcription factors in different MN types. In MNs that control locomotion, LIN-39 synergizes with MAB-5 and UNC-3, the latter is known to function as a terminal selector in these neurons [15]. Mechanistically, LIN-39 and UNC-3 control the same ACh pathway genes by recognizing distinct binding sites. However, whether LIN-39 and UNC-3 are being recruited to these sites in an additive or cooperative manner remains unresolved. In MNs that control egg-laying [14], UNC-3 is not present [15]. HLH-3 (bHLH protein of Achaete-Scute family) and LIN-11 (LIM homeodomain protein) are expressed in VC neurons [57, 58], constituting putative LIN-39 collaborators for the control of cholinergic identity in VC.

3.4.4 Hox and UNC-3 in a positive feed-forward loop (FFL): a form of redundancy engineering to ensure robust expression of cholinergic identity genes

Complex gene regulatory networks are composed of simple gene circuits called "network motifs"[59]. One of the most widely used motifs is the feed-forward loop (FFL), where transcription factor X activates a second transcription factor Y, and both activate their shared target gene Z (**Fig. 3.5J**). Our findings uncovered a FFL in cholinergic MNs, where a Hox protein (LIN-39) activates UNC-3, and both activate their shared targets (ACh pathway genes)

(**Fig. 3.5J**). FFLs have been described in transcriptional networks across species [59-61]. Based on systems biology classifications, there are 8 different types of FFLs – each characterized by the signs ("+" for activation, "–" for repression) of the transcriptional interactions within the motif. The FFL in MNs is coherent type 1 because all interactions are activating (**Fig. 3.5J**).

Why is there a need for this type of FFL in *C. elegans* MNs? Computational models and studies in bacteria indicate that a coherent type 1 FFL can ensure robust gene expression against perturbations through a process called "sign-sensitive delay" [61-63]. That is, a perturbation can decrease the levels of transcription factor X (LIN-39), but Z (ACh pathway genes) responds only at a delay once X levels decrease. The delay is due to the presence of Y (UNC-3). After X is decreased, it takes time for Y levels to decrease (depending on the degradation rate of Y) to a level insufficient to activate Z. Hence, the coherent type 1 FFL can be viewed as a form of redundancy engineering (or filter) to protect the shared target genes (Z) from fluctuations in the input (transcription factor X). Based on computational modeling and functional studies in bacteria [61-63], we propose that, in our system, LIN-39 (X) and UNC-3 (Y) operate in a coherent type 1 FFL to maintain robust expression levels of ACh pathway genes in MNs throughout life (**Fig. 3.5J**). Lastly, we note that coherent and incoherent FFLs can serve various functions in the developing nervous system. For example, they can act as developmental timers (by allowing a given transcription factor to control different genes at successive developmental stages) [64, 65], or as means to generate neuronal subtype diversity [66, 67].

The FFL we identified in *C. elegans* MNs forms the backbone of a transcriptional network, but it does not function in isolation. Rather, it is embedded within at least two additional loops: (a) a positive feedback loop that activates LIN-39 expression through transcriptional autoregulation,

and (b) negative feedback provided by UNC-3 to reduce LIN-39 expression (**Fig. 3.7G**). The significance of both is discussed below.

Consistent with our observations on Hox (LIN-39, MAB-5) and *unc-3*, two other *C. elegans* Hox proteins (CEH-13, EGL-5) control the transcription of another terminal selector (*mec-3*) in peripheral touch receptor neurons [68]. Although a FFL was not identified in that cellular context, the findings on *mec-3* and *unc-3* (this study) strongly suggest that Hox-mediated transcriptional control of terminal selectors may be a broadly applicable strategy to ensure robustness of gene expression in post-mitotic neurons.

3.4.5 A two-component design principle for homeostatic control of Hox gene expression in motor neurons

Mechanisms for the initiation of Hox gene expression are well-studied. It is known, for example, that during embryonic patterning, Polycomb Group (PcG) and trithorax Group (trxG) genes determine the levels and spatial expression of Hox genes [69-73]. It is also known that, during early nervous system development, morphogenetic gradients initially establish Hox gene expression, and this is further refined via Hox cross-regulatory interactions [23, 74]. However, the mechanisms that maintain Hox gene expression in post-mitotic neurons during later developmental and post-embryonic stages remain unknown.

Our constitutive (genetic null alleles) and temporally controlled (AID system) approaches combined with mutagenesis of Hox binding sites within the endogenous *lin-39* locus strongly suggest that LIN-39 maintains its own expression in MNs through transcriptional autoregulation (**Fig. 3.6H**). Subsequently, LIN-39 is required to maintain the expression of ACh pathway genes by operating at the top of a coherent FFL (**Fig. 3.6I**). However, the *lin-39* expression levels are

critical: low levels lead to a failure to maintain ACh pathway gene expression (**Fig. 3.1-3.2**), whereas high levels of LIN-39 in cholinergic MNs cause locomotion defects and "mixed neuronal identity", i.e., genes normally expressed in GABA neurons became ectopically expressed in cholinergic MNs [34]. We found that *lin-39* expression in MNs is under homeostatic control. The terminal selector UNC-3 prevents high levels of *lin-39* expression, thereby counterbalancing the positive effect of the transcriptional autoregulation of *lin-39*. From the perspective of Hox binding site affinity [75, 76], this mechanism may enable LIN-39 (when present at optimal levels) to bind at high-affinity sites in the *cis*-regulatory region of ACh pathway genes. The negative UNC-3 feedback prevents high levels of LIN-39, potentially avoiding LIN-39 binding to low-affinity sites in the *cis*-regulatory region of alternative (e.g., GABA) identity genes.

Altogether, we identified a two-component design principle (positive feedback of LIN-39 to itself, negative feedback from the terminal selector UNC-3) for the homeostatic control of Hox gene expression in motor neurons. These two components are embedded in a coherent FFL, which we propose to be necessary for maintenance of NT identity (**Fig. 3.7G**). Because Hox genes are continuously expressed in the adult fly, mouse, and human nervous systems, the gene regulatory architecture described here for the control of NT identity may constitute a broadly applicable principle.

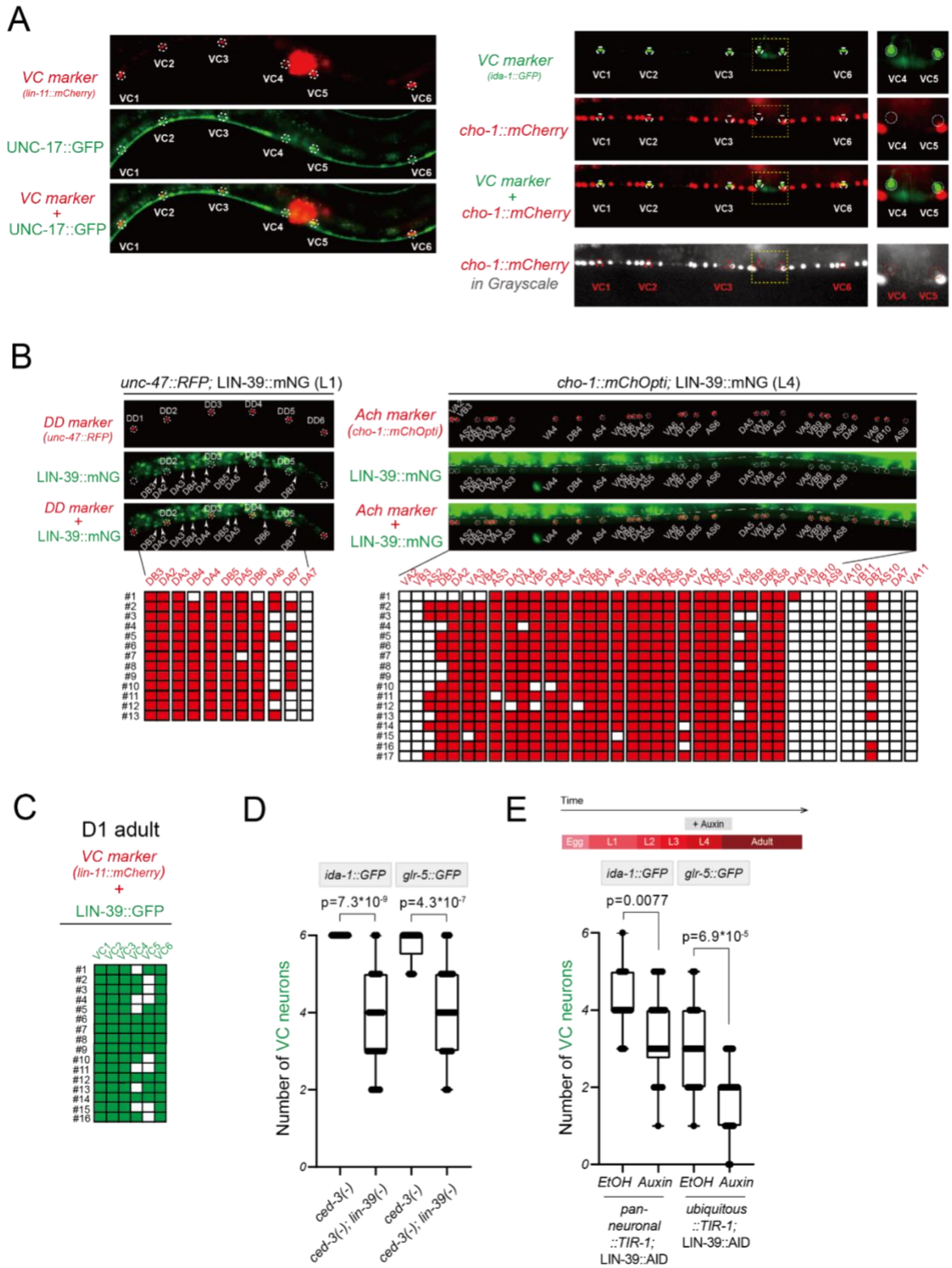

Figure 3.8 LIN-39 is necessary to maintain VC terminal identity features. **A**. Markers for *cho-1* and *unc-17* are expressed in all six VC MNs. Representative images showing *unc-17* (left) and *cho-1* (right) expression that colocalizes and known VC reporters in split and merged channels. Because *cho-1* expression is dim in VC4-5, a grayscale channel is shown **B-C**. LIN-39 is

continuously expressed in MNs that control locomotion and egg-laying. Representative images of endogenous *lin-39* reporter (*lin-39::mNG*). Markers of GABAergic (*unc-47::RFP*) and cholinergic (*cho-1::mChOpti*) motor neurons were use dot identify the *lin-39* expressing neurons with single-cell resolution. For panel **C** refer to **Figure 1B** for representative image. **D**. Quantification of *ida-1* and *glr-5* reporter gene expression in day 1 adult animals of *ced-3 (-)* and *ced-3(-); lin-39(-)* genotypes. **E.** LIN-39 is required to maintain *ida-1* and *glr-5* expression in adult VC neurons. Timeline for auxin application indicated on top with quantifications below. Box and whisker plots were used with presentation of all data points. Unpaired t-test with Welch's correction was performed and p-values were annotated. N>15.

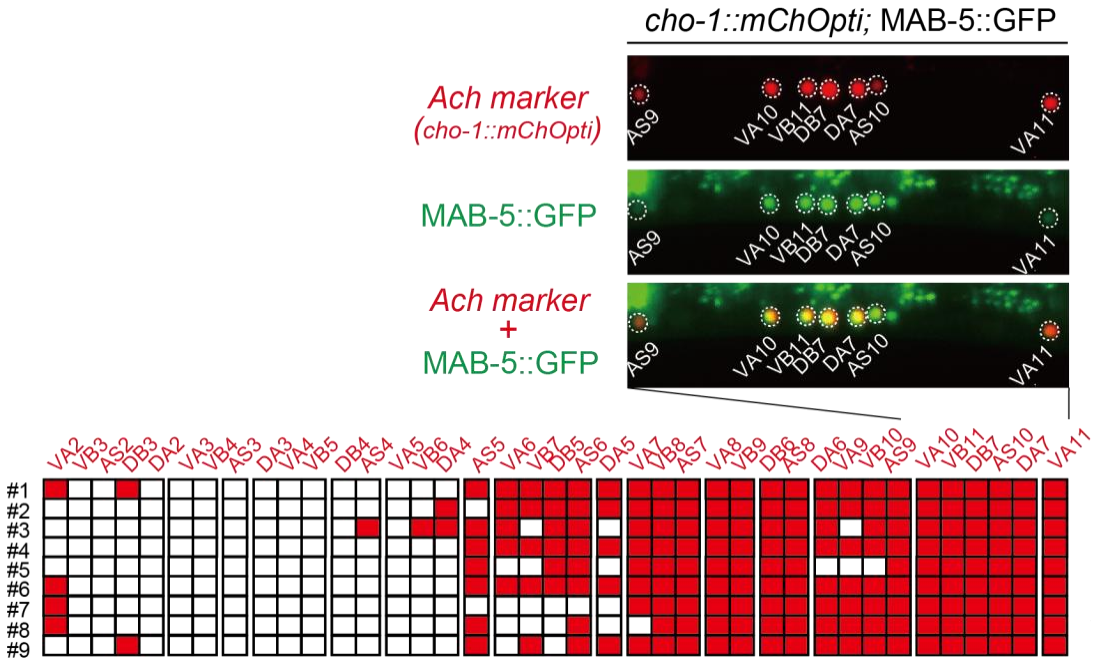

Figure 3.9 MAB-5 expression profile in cholinergic MNs. Representative co-localization images of MAB-5::GFP (fosmid reporter) with a cholinergic reporter that localizes to the nucleus (*cho-1::SL2::mChOpti::H2B*). Motor neuron nuclei are circled.

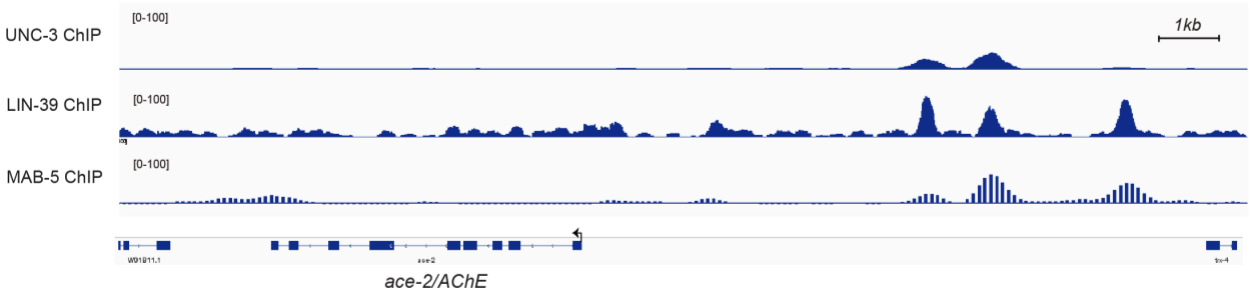

Figure 3.10 ChIP-seq binding peaks for UNC-3, LIN-39 and MAB-5 on *ace-2* locus. Schematic extracted from the IGV software showing the *ace-2* and ChIP-seq tracks for UNC-3, LIN-39 and MAB-5.

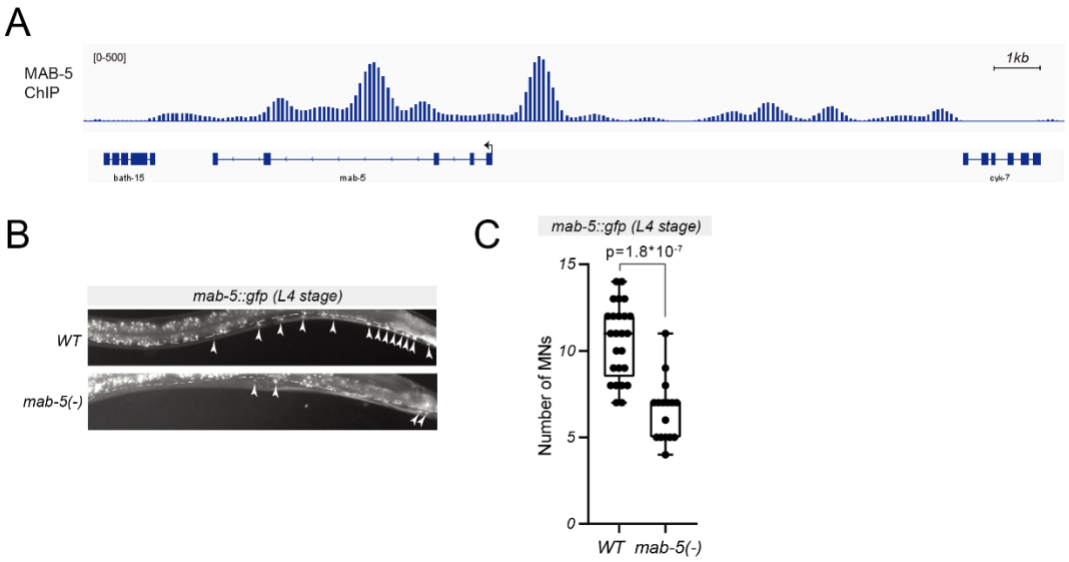

Figure 3.11 Transcriptional autoregulation of *mab-5* in motor neurons. **A**. Gene locus of *mab-5* with ChIP-seq tracks for MAB-5. **B**. Representative images of a *mab-5* transcriptional *gfp* reporter in posterior motor neurons of *WT* and *mab-5(e1239)* mutant animals at L4. **C**. Quantification of *mab-5::gfp* expression in panel **B**. Box and whisker plots were used with presentation of all data points. Unpaired t-test with Welch's correction was performed and p-values were annotated. N>15

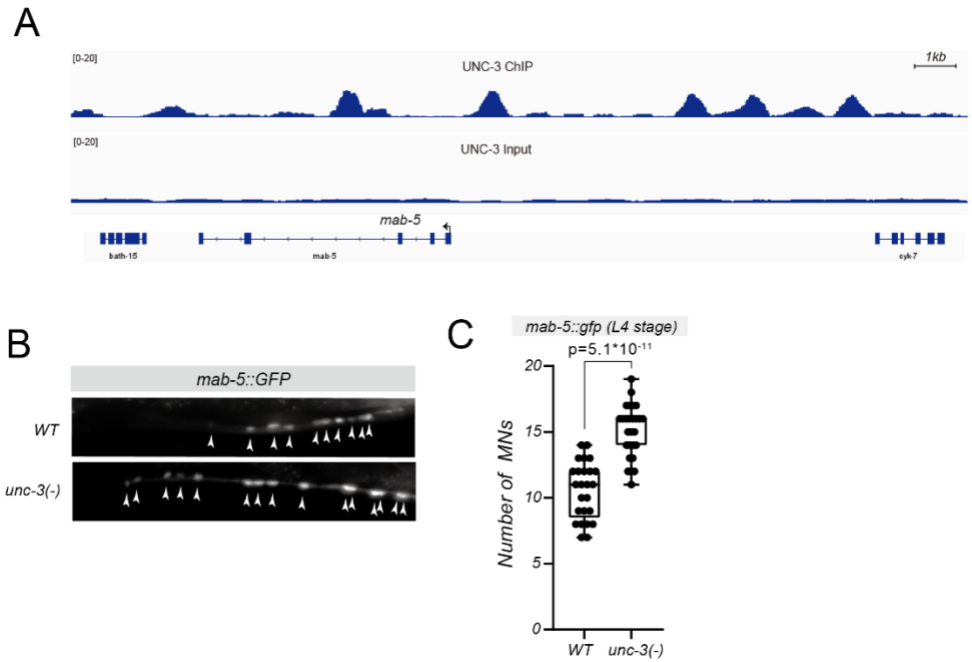

Figure 3.12 UNC-3 represses *mab-5* expression. **A**. Gene locus of *mab-5* with ChIP-seq tracks for UNC-3. **B**. Representative images of a *mab-5* transcriptional *gfp* reporter in posterior motor neurons of *WT* and *unc-3(n3435)* mutant animals at L4. **C**. Quantification of *mab-5::gfp* expression in panel **B**. Box and whisker plots were used with presentation of all data points. Unpaired t-test with Welch's correction was performed and p-values were annotated. N>15.

3.5 Material and Methods

C. elegans strains

Worms were grown at 15˚C, 20˚C or 25˚C on nematode growth media (NGM) plates seeded with bacteria (*E.coli* OP50) as food source [77]. Mutant alleles used in this study: *unc-3 (n3435) X, lin-39 (n1760) III, mab-5 (e1239) III, ced-3 (n1286) IV.* CRISPR-generated alleles: *unc-3 (ot837 [unc-3::mNG::3xFLAG::AID]) X, lin-39 (kas9 [lin-39::mNG::3xFLAG::AID]) III.*

All reporter strains used in this study are shown in **Table 3.1.**

Generation of transgenic reporter animals

Reporter gene fusions for *cis*-regulatory analysis were made using either PCR fusion [78] or Gibson Assembly Cloning Kit (NEB #5510S). Targeted DNA fragments were fused (ligated) to *tagrfp* coding sequence, which was followed by *unc-54 3' UTR.* Mutations or deletions LIN-39 binding sites were introduced via PCR mutagenesis. The product DNA fragments were either injected into young adult *pha-1(e2123)* hermaphrodites at 50ng/µl using *pha-1* (pBX plasmid) as co-injection marker (50 ng/µl) and further selected for survival, or injected into young adult N2 hermaphrodites at 50ng/µl (plus 50ng/µl pBX plasmid) using *myo-2::gfp* as co-injection marker (3 ng/µl) and further selected for GFP signal.

Targeted genome engineering

CRISPR/Cas9 genome editing was employed to generate: (a) PHX2477 strain carrying the *lin-39(syb2477 [lin-39_site1-11_MUT::mNG::3xFLAG::AID])* III allele, and (b) PHX5816 strain carrying the *unc-3(syb5816 [unc-3_LIN-39 site_MUT_GFP])* X allele. Gene editing was performed by SunyBiotech.

Temporally controlled protein degradation

AID-tagged proteins are conditionally degraded when exposed to auxin in the presence of TIR1 [35]. Animals carrying auxin-inducible alleles of *lin-39 (kas9 [lin-39::mNG::AID])* were crossed with *ieSi57* animals that express TIR1 ubiquitously or with otTi28 animals that express TIR-1 pan-neuronally. Auxin (indole-3-acetic acid [IAA]) was dissolved in ethanol (EtOH) to prepare 400 mM stock solutions which were stored at 4°C for up to one month. NGM agar plates were poured with auxin or ethanol added to a final concentration of 4 mM and allowed to dry overnight at room temperature. Plates were seeded with OP50 bacteria. To induce protein degradation, worms of the experimental strains were transferred onto auxin-coated plates and kept at 25°C. As a control, worms were transferred onto EtOH-coated plates instead. Auxin solutions, auxin-coated plates, and experimental plates were shielded from light.

Single Molecule RNA in situ Hybridization (sm RNA-FISH)

Egg preparation was performed before harvesting the synchronized L1 worms. Next, worms were fixed with 4% paraformaldehyde (PFA), washed twice with PBS, and permeabilized with 70% ethanol at 4°C overnight. Then, samples were hybridized with diluted mRNA probes (Stellaris; *unc-3*: unc-3 best Quasar 670 (SMF-1065-5); *lin-39*: lin-39 best Quasar 670 (SMF-1065-5)) in hybridization buffer (Stellaris#174; RNA FISH Hybridization Buffer (SMF-HB1-10)) following manufacturing protocols. Finally, samples were stained with DAPI, washed, and equilibrated with GLOX buffer. Worms were imaged at a fluorescence microscope (Zeiss, Axio Imager.Z2).

Microscopy

Worms were anesthetized using 100mM of sodium azide (NaN_3) and mounted on a 4% agarose pad on glass slides. Images were taken using an automated fluorescence microscope (Zeiss, Axio Imager.Z2). Acquisition of several z-stack images (each ~1 µm thick) was taken with Zeiss

Axiocam 503 mono using the ZEN software (Version 2.3.69.1000, Blue edition). Representative images are shown following max-projection of 1-8 µm Z-stacks using the maximum intensity projection type. Image reconstruction was performed using Image J software [79].

Motor neuron identification

Motor neurons were identified based on a combination of the following factors: (i) co-localization with fluorescent markers with known expression pattern, [ii] invariant cell body position along the ventral nerve cord, or relative to other MN subtypes, [iii] MN birth order, and (iv) number of MNs that belong to each subtype.

Bioinformatic analysis.

To predict the UNC-3 binding site (COE motif) in the *cis*-regulatory region of target genes, we used the MatInspector program from Genomatix [82]. The Position Weight Matrix (PWM) for the LIN-39 binding site is catalogued in the CIS-BP (Catalog of Inferred Sequence Binding Preferences database) [83]. To identify putative LIN-39 sites, we used FIMO (Find Individual Motif Occurrences)[84], which is one of the motif-based sequence analysis tools of MEME (Multiple Expectation maximization for Motif Elicitation) bioinformatics suite (http://meme-suite.org/). The p-value threshold for the analysis of LIN-39 target genes was set at $p<0.001$ for *unc-17* and *cho-1*, and at $p<0.01$ for *lin-39*.

Fluorescence Intensity (FI) Quantification

To quantify FI of individual MNs in the VNC, images of worms from different genetic backgrounds were taken with identical parameters through full-thickness z-stacks that cover the entire cell body. Image stacks were then processed and quantified for FI via FIJI. The focal plane in Z-stacks that has the brightest FI was selected for quantification to minimize background signals. Cell outline was manually selected, and FIJI was used to quantify the FI and area to get

the mean value for FI. After quantifying the FI of all MNs of interest, representative average image background FI was quantified and subtracted from the individual MN FI quantifications to get the mean FI in arbitrary units (a.u.).

Statistical analysis

For quantification, box and whisker plots were adopted to represent the quartiles in graph. The box includes data points from the first to the third quartile value with the horizontal line in box representing the mean value. Upper and lower limits indicate the max and min, respectively. Unpaired t-test with Welch's correction was performed and p-values were annotated. Visualization of data and p-value calculation were performed via GraphPad Prism Version 9.2.0 (283).

3.6 Acknowledgements

We thank the Caenorhabditis Genetics Center (CGC), which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440), for providing strains. We are grateful to Oliver Hobert and members of the Kratsios lab (Edgar Correa, Nidhi Sharma, Filipe Marques, Manasa Prahlad, Anthony Osuma) for comments on this manuscript. This work was funded by two NIH grants to P.K (R01 NS116365-01, R01 NS118078-01).

Table 3.1 List of *C. elegans* strains used in this study

Reporters and Transgenes	
OH11954	*lin-11::mCherry V + myo-2::GFP*
OP18	*unc-119(ed3) III; wgIs18 [lin-39::TY1::EGFP::3xFLAG + unc-119(+)]*
OH13083	*him-5(e1490) V; otIs576 [unc-17(fosmid)::GFP + lin-44::YFP]*
	him-5(e1490) otIs544 [cho-1(fosmid)::SL2::mCherry::H2B + pha-1(+)] V
CA1200	*ieSi57 [eft-3p::TIR1::mRuby::unc-54 3'UTR + Cbr-unc-119(+)] II; unc-119(ed3) III.*
OH14930	*otTi28 [unc-11prom8+ehs-1prom7+rgef-1prom2::TIR1::mTurquoise2::unc-54 3'UTR] X*
OH15568	*unc-17(ot907[unc-17::mKate2::3XFLAG]) IV*
OH9970	*otEx4432 [ace-2 fosmid_SL2:gfp, line 2.7D] + elt-2::DsRed*
OH11650	*pha-1(-); otEx5291 [cho-1_280bp_NLS_YFP+pha-1(+)]*
OH11684	*pha-1(-); otEx5308 [unc-17_125bp_NLS_YFP+pha-1(+)]*
OH11063	*otEx4992 [unc-3_fosmid_Exon2_mCherry+myo-2::GFP]*
OH11068	*otEx4995 [unc-3 fosmid_exon6_mCherry (myo-2::gfp)]*
OH11454	*otIs426 [Punc-17_1kb::YFP + myo-2::GFP]*
	otIs447 [unc-3p::mCherry + pha-1(+)] IV.
LX929	*vsIs48 [unc-17::GFP] X*
KRA257	*kasIs2 [lin-39_intron 1::tagRFP + pha-1 rescuing construct line 6.6] I*
	stIs10224 [lin-39p::HIS-24::mCherry + unc-119(+)]
BL5717	*inIs179 [pida-1::GFP] II; him-8(e1489) IV*

OH13105	*him-5(1490) otIs564 {otIs565 [unc-47(fosmid)::SL2::H2B::mChopti + pha-1(+)] }V*
	icIs270 [glr-5::GFP + lin-15(+)] X
OH3241	*juIs76 [unc-25p::GFP + lin-15(+)] II ; oyIs14 [sra-6::GFP + lin-15(+)] V*
OP27	*unc-119(ed3) III; wgIs27 [mab-5::TY1::EGFP::3xFLAG + unc-119(+)]*
CF453	*muIs16 [mab-5::GFP + dpy-20(+)] II; dpy-20(e1282) IV*
Mutants and CRISPR alleles	
PHX2477	*lin-39(syb2477 [lin-39_site1-11_MUT::mNG::AID]) III*
PHX5816	*unc-3(syb5816 [unc-3_LIN-39 site_MUT_GFP]) X*
MT3002	*ced-3(n1286) IV*
CB3531	*mab-5(e1239) III; him-5(e1490) V*
LE4023	*lin-39(n1760) mab-5(e1239)/hT2 III*
KRA45	*lin-39(n1760)/hT2 III, I*
MT10785	*unc-3(n3435) X*
KRA467	*lin-39(kas9[lin-39::mNG::AID]) III*
	unc-3(ot837[unc-3::mNeonGreen::AID]) X
	otEx5414 [Pace-2_180bp::unc-3 Pttx-3_CHE#4.1]

3.7 References

1. Dulcis, D., et al., *Neurotransmitter switching in the adult brain regulates behavior.* Science, 2013. 340(6131): p. 449-53.

2. Li, H.Q., et al., *Decoding Neurotransmitter Switching: The Road Forward.* J Neurosci, 2020. 40(21): p. 4078-4089.

3. Rand, J.B., *Acetylcholine.* WormBook, 2007: p. 1-21.

4. Casas, C., et al., *Early presymptomatic cholinergic dysfunction in a murine model of amyotrophic lateral sclerosis.* Brain Behav, 2013. 3(2): p. 145-58.

5. McKinley, J.W., et al., *Dopamine Deficiency Reduces Striatal Cholinergic Interneuron Function in Models of Parkinson's Disease.* Neuron, 2019. 103(6): p. 1056-1072 e6.

6. Zheng, L.F., et al., *Reduced expression of choline acetyltransferase in vagal motoneurons and gastric motor dysfunction in a 6-OHDA rat model of Parkinson's disease.* Brain Res, 2011. 1420: p. 59-67.

7. Gu, X. and X. Wang, *An overview of recent analysis and detection of acetylcholine.* Anal Biochem, 2021. 632: p. 114381.

8. Phillis, J.W., *Acetylcholine release from the central nervous system: a 50-year retrospective.* Crit Rev Neurobiol, 2005. 17(3-4): p. 161-217.

9. Cho, H.H., et al., *Isl1 directly controls a cholinergic neuronal identity in the developing forebrain and spinal cord by forming cell type-specific complexes.* PLoS Genet, 2014. 10(4): p. e1004280.

10. Lopes, R., et al., *Transcription factor LIM homeobox 7 (Lhx7) maintains subtype identity of cholinergic interneurons in the mammalian striatum.* Proc Natl Acad Sci U S A, 2012. 109(8): p. 3119-24.

11. Wenick, A.S. and O. Hobert, *Genomic cis-regulatory architecture and trans-acting regulators of a single interneuron-specific gene battery in C. elegans.* Dev Cell, 2004. 6(6): p. 757-70.

12. Zhang, F., et al., *The LIM and POU homeobox genes ttx-3 and unc-86 act as terminal selectors in distinct cholinergic and serotonergic neuron types.* Development, 2014. 141(2): p. 422-35.

13. Von Stetina, S.E., M. Treinin, and D.M. Miller, 3rd, *The motor circuit.* Int Rev Neurobiol, 2006. 69: p. 125-67.

14. Loots, G.G., et al., *Genomic deletion of a long-range bone enhancer misregulates sclerostin in Van Buchem disease.* Genome Res, 2005. 15(7): p. 928-35.

15. Kratsios, P., et al., *Coordinated regulation of cholinergic motor neuron traits through a conserved terminal selector gene.* Nat Neurosci, 2012. 15(2): p. 205-14.

16. Pereira, L., et al., *A cellular and regulatory map of the cholinergic nervous system of C. elegans.* Elife, 2015. 4.

17. Hughes, C.L. and T.C. Kaufman, *Hox genes and the evolution of the arthropod body plan.* Evol Dev, 2002. 4(6): p. 459-99.

18. Mallo, M., D.M. Wellik, and J. Deschamps, *Hox genes and regional patterning of the vertebrate body plan.* Dev Biol, 2010. 344(1): p. 7-15.

19. Di Bonito, M., J.C. Glover, and M. Studer, *Hox genes and region-specific sensorimotor circuit formation in the hindbrain and spinal cord.* Dev Dyn, 2013. 242(12): p. 1348-68.

20. Estacio-Gomez, A. and F.J. Diaz-Benjumea, *Roles of Hox genes in the patterning of the central nervous system of Drosophila.* Fly (Austin), 2014. 8(1): p. 26-32.

21. Joshi, R., R. Sipani, and A. Bakshi, *Roles of Drosophila Hox Genes in the Assembly of Neuromuscular Networks and Behavior.* Front Cell Dev Biol, 2021. 9: p. 786993.

22. Parker, H.J. and R. Krumlauf, *A Hox gene regulatory network for hindbrain segmentation.* Curr Top Dev Biol, 2020. 139: p. 169-203.

23. Philippidou, P. and J.S. Dasen, *Hox genes: choreographers in neural development, architects of circuit organization.* Neuron, 2013. 80(1): p. 12-34.

24. Allen, A.M., et al., *A single-cell transcriptomic atlas of the adult Drosophila ventral nerve cord.* Elife, 2020. 9.

25. Coughlan, E., et al., *A Hox Code Defines Spinocerebellar Neuron Subtype Regionalization.* Cell Rep, 2019. 29(8): p. 2408-2421 e4.

26. Hutlet, B., et al., *Systematic expression analysis of Hox genes at adulthood reveals novel patterns in the central nervous system.* Brain Struct Funct, 2016. 221(3): p. 1223-43.

27. Kratsios, P., et al., *An intersectional gene regulatory strategy defines subclass diversity of C. elegans motor neurons.* Elife, 2017. 6.

28. Lizen, B., et al., *HOXA5 localization in postnatal and adult mouse brain is suggestive of regulatory roles in postmitotic neurons.* J Comp Neurol, 2017. 525(5): p. 1155-1175.

29. Takahashi, Y., et al., *Expression profiles of 39 HOX genes in normal human adult organs and anaplastic thyroid cancer cell lines by quantitative real-time RT-PCR system.* Exp Cell Res, 2004. 293(1): p. 144-53.

30. Zheng, C., H.M.T. Lee, and K. Pham, *Nervous system-wide analysis of Hox regulation of terminal neuronal fate specification in Caenorhabditis elegans.* PLoS Genet, 2022. 18(2): p. e1010092.

31. Reilly, M.B., et al., *Unique homeobox codes delineate all the neuron classes of C. elegans.* Nature, 2020. 584(7822): p. 595-601.

32. Potts, M.B., D.P. Wang, and S. Cameron, *Trithorax, Hox, and TALE-class homeodomain proteins ensure cell survival through repression of the BH3-only gene egl-1.* Dev Biol, 2009. 329(2): p. 374-85.

33. Liu, H., et al., *Direct regulation of egl-1 and of programmed cell death by the Hox protein MAB-5 and by CEH-20, a C. elegans homolog of Pbx1.* Development, 2006. 133(4): p. 641-50.

34. Feng, W., et al., *A terminal selector prevents a Hox transcriptional switch to safeguard motor neuron identity throughout life.* Elife, 2020. 9.

35. Zhang, L., et al., *The auxin-inducible degradation (AID) system enables versatile conditional protein depletion in C. elegans.* Development, 2015. 142(24): p. 4374-84.

36. Hatzihristidis, T., et al., *A Drosophila-centric view of protein tyrosine phosphatases.* FEBS Lett, 2015. 589(9): p. 951-66.

37. Brockie, P.J., et al., *Differential expression of glutamate receptor subunits in the nervous system of Caenorhabditis elegans and their regulation by the homeodomain protein UNC-42.* J Neurosci, 2001. 21(5): p. 1510-22.

38. Prasad, B., et al., *unc-3-dependent repression of specific motor neuron fates in Caenorhabditis elegans.* Dev Biol, 2008. 323(2): p. 207-15.

39. Li, Y., et al., *Establishment and maintenance of motor neuron identity via temporal modularity in terminal selector function.* Elife, 2020. 9.

40. Boyle, A.P., et al., *Comparative analysis of regulatory information and circuits across distant species.* Nature, 2014. 512(7515): p. 453-6.

41. Ji, N. and A. van Oudenaarden, *Single molecule fluorescent in situ hybridization (smFISH) of C. elegans worms and embryos.* WormBook, 2012: p. 1-16.

42. Hobert, O., *Regulatory logic of neuronal diversity: terminal selector genes and selector motifs.* Proc Natl Acad Sci U S A, 2008. 105(51): p. 20067-71.

43. Hobert, O., *Regulation of terminal differentiation programs in the nervous system.* Annu Rev Cell Dev Biol, 2011. 27: p. 681-96.

44. Hobert, O., *Terminal Selectors of Neuronal Identity.* Curr Top Dev Biol, 2016. 116: p. 455-75.

45. Hobert, O., *A map of terminal regulators of neuronal identity in Caenorhabditis elegans.* Wiley Interdiscip Rev Dev Biol, 2016. 5(4): p. 474-98.

46. Hobert, O. and P. Kratsios, *Neuronal identity control by terminal selectors in worms, flies, and chordates.* Curr Opin Neurobiol, 2019. 56: p. 97-105.

47. Deneris, E.S. and O. Hobert, *Maintenance of postmitotic neuronal cell identity.* Nat Neurosci, 2014. 17(7): p. 899-907.

48. Dasen, J.S. and T.M. Jessell, *Hox networks and the origins of motor neuron diversity.* Curr Top Dev Biol, 2009. 88: p. 169-200.

49. Kalis, A.K., et al., *Patterning of sexually dimorphic neurogenesis in the caenorhabditis elegans ventral cord by Hox and TALE homeodomain transcription factors.* Dev Dyn, 2014. 243(1): p. 159-71.

50. Kalis, A.K., et al., *Hox proteins interact to pattern neuronal subtypes in Caenorhabditis elegans males.* Genetics, 2022. 220(4).

51. Lints, R. and S.W. Emmons, *Patterning of dopaminergic neurotransmitter identity among Caenorhabditis elegans ray sensory neurons by a TGFbeta family signaling pathway and a Hox gene.* Development, 1999. 126(24): p. 5819-31.

52. Burglin, T.R. and M. Affolter, *Homeodomain proteins: an update.* Chromosoma, 2016. 125(3): p. 497-521.

53. Hobert, O., *Homeobox genes and the specification of neuronal identity.* Nat Rev Neurosci, 2021. 22(10): p. 627-636.

54. Serrano-Saiz, E., et al., *BRN3-type POU Homeobox Genes Maintain the Identity of Mature Postmitotic Neurons in Nematodes and Mice.* Curr Biol, 2018. 28(17): p. 2813-2823 e2.

55. Song, N.N., et al., *Adult raphe-specific deletion of Lmx1b leads to central serotonin deficiency.* PLoS One, 2011. 6(1): p. e15998.

56. Zhao, Z.Q., et al., *Lmx1b is required for maintenance of central serotonergic neurons and mice lacking central serotonergic system exhibit normal locomotor activity.* J Neurosci, 2006. 26(49): p. 12781-8.

57. Perez, L.M. and A. Alfonso, *The Conserved ASCL1/MASH-1 Ortholog HLH-3 Specifies Sex-Specific Ventral Cord Motor Neuron Fate in Caenorhabditis elegans.* G3 (Bethesda), 2020. 10(11): p. 4201-4213.

58. Hobert, O., et al., *Control of neural development and function in a thermoregulatory network by the LIM homeobox gene lin-11.* J Neurosci, 1998. 18(6): p. 2084-96.

59. Alon, U., *Network motifs: theory and experimental approaches.* Nat Rev Genet, 2007. 8(6): p. 450-61.

60. Mangan, S. and U. Alon, *Structure and function of the feed-forward loop network motif.* Proc Natl Acad Sci U S A, 2003. 100(21): p. 11980-5.

61. Kalir, S., S. Mangan, and U. Alon, *A coherent feed-forward loop with a SUM input function prolongs flagella expression in Escherichia coli.* Mol Syst Biol, 2005. 1: p. 2005 0006.

62. Le, D.H. and Y.K. Kwon, *A coherent feedforward loop design principle to sustain robustness of biological networks.* Bioinformatics, 2013. 29(5): p. 630-7.

63. Mangan, S., A. Zaslaver, and U. Alon, *The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks.* J Mol Biol, 2003. 334(2): p. 197-204.

64. Baumgardt, M., et al., *Specification of neuronal identities by feedforward combinatorial coding.* PLoS Biol, 2007. 5(2): p. e37.

65. Gabilondo, H., et al., *Segmentally homologous neurons acquire two different terminal neuropeptidergic fates in the Drosophila nervous system.* PLoS One, 2018. 13(4): p. e0194281.

66. Hobert, O., *Development of left/right asymmetry in the Caenorhabditis elegans nervous system: from zygote to postmitotic neuron.* Genesis, 2014. 52(6): p. 528-43.

67. Johnston, R.J., Jr., *Lessons about terminal differentiation from the specification of color-detecting photoreceptors in the Drosophila retina.* Ann N Y Acad Sci, 2013. 1293: p. 33-44.

68. Zheng, C., F.Q. Jin, and M. Chalfie, *Hox Proteins Act as Transcriptional Guarantors to Ensure Terminal Differentiation.* Cell Rep, 2015. 13(7): p. 1343-1352.

69. Afzal, Z. and R. Krumlauf, *Transcriptional Regulation and Implications for Controlling Hox Gene Expression.* J Dev Biol, 2022. 10(1).

70. Deschamps, J., et al., *Initiation, establishment and maintenance of Hox gene expression patterns in the mouse.* Int J Dev Biol, 1999. 43(7): p. 635-50.

71. Gentile, C. and M. Kmita, *Polycomb Repressive Complexes in Hox Gene Regulation: Silencing and Beyond: The Functional Dynamics of Polycomb Repressive Complexes in Hox Gene Regulation.* Bioessays, 2020. 42(10): p. e1900249.

72. Kassis, J.A., J.A. Kennison, and J.W. Tamkun, *Polycomb and Trithorax Group Genes in Drosophila.* Genetics, 2017. 206(4): p. 1699-1725.

73. Mallo, M. and C.R. Alonso, *The regulation of Hox gene expression during animal development.* Development, 2013. 140(19): p. 3951-63.

74. Frank, D. and D. Sela-Donenfeld, *Hindbrain induction and patterning during early vertebrate development.* Cell Mol Life Sci, 2019. 76(5): p. 941-960.

75. Crocker, J., et al., *Low affinity binding site clusters confer hox specificity and regulatory robustness.* Cell, 2015. 160(1-2): p. 191-203.

76. Delker, R.K., et al., *Low affinity binding sites in an activating CRM mediate negative autoregulation of the Drosophila Hox gene Ultrabithorax.* PLoS Genet, 2019. 15(10): p. e1008444.

77. Brenner, S., *The genetics of Caenorhabditis elegans.* Genetics, 1974. 77(1): p. 71-94.

78. Hobert, O., *PCR fusion-based approach to create reporter gene constructs for expression analysis in transgenic C. elegans.* Biotechniques, 2002. 32(4): p. 728-30.

79. Schindelin, J., et al., *Fiji: an open-source platform for biological-image analysis.* Nat Methods, 2012. 9(7): p. 676-82.

80. Smemo, S., et al., *Obesity-associated variants within FTO form long-range functional connections with IRX3.* Nature, 2014. 507(7492): p. 371-5.

81. Seth R Taylor, G.S., Molly Reilly, Lori Glenwinkel, Abigail Poff, Rebecca McWhirter, Chuan Xu, Alexis Weinreb, Manasa Basavaraju, Steven J Cook, Alec Barrett, Alexander Abrams, Berta Vidal, Cyril Cros, Ibnul Rafi, Nenad Sestan, Marc Hammarlund, Oliver Hobert, David M. Miller III, *Expression profiling of the mature C. elegans nervous system by single-cell RNA-Sequencing.* bioRxiv, 2019.

82. Cartharius, K., et al., *MatInspector and beyond: promoter analysis based on transcription factor binding sites.* Bioinformatics, 2005. 21(13): p. 2933-42.

83. Weirauch, M.T., et al., *Determination and inference of eukaryotic transcription factor sequence specificity.* Cell, 2014. 158(6): p. 1431-1443.

84. Grant, C.E., T.L. Bailey, and W.S. Noble, *FIMO: scanning for occurrences of a given motif.* Bioinformatics, 2011. 27(7): p. 1017-8.

85. Faumont, S., et al., *An image-free opto-mechanical system for creating virtual environments and imaging neuronal activity in freely moving Caenorhabditis elegans.* PLoS One, 2011. 6(9): p. e24666.

Chapter 4 A terminal selector prevents a Hox transcriptional switch to safeguard motor neuron identity throughout life

Author contribution: All Figures are jointly generated by Paschalis Kratsios and Weidong Feng. Data in **Figure 4.1, 4.2, 4.3, 4.4, 4.5A-B, 4.6, 4.7, 4.8, 4.9, 4.11, 4.12, 4.13, 4.14, 4.15, 4.16A-B, 4.19,** and **Figure 4.20D** is generated by Weidong Feng. Data in **Figure 4.5C, 4.16C-D**, and **Figure 4.17** is generated by Yinan Li. Data in **Figure 4.10A-B** and **Figure 4.20A-C** is generated by André Brown group. Data in **Figure 4.18** is generated by Benayahu Elbaz.

4.1 Abstract

To become and remain functional, individual neuron types must select during development and maintain throughout life their distinct terminal identity features, such as expression of specific neurotransmitter receptors, ion channels and neuropeptides. Here, we report a molecular mechanism that enables cholinergic motor neurons (MNs) in the *C. elegans* ventral nerve cord to select and maintain their unique terminal identity. This mechanism relies on the dual function of the conserved terminal selector UNC-3 (Collier/Ebf). UNC-3 synergizes with LIN-39 (Scr/Dfd/Hox4-5) to directly co-activate multiple terminal identity traits specific to cholinergic MNs, but also antagonizes LIN-39's ability to activate terminal features of alternative neuronal identities. Loss of *unc-3* causes a switch in the transcriptional targets of LIN-39, thereby alternative, not cholinergic MN-specific, terminal features become activated and locomotion defects occur. The strategy of a terminal selector preventing a transcriptional switch may constitute a general principle for safeguarding neuronal identity throughout life.

4.2 Introduction

Every nervous system is equipped with distinct neuron types essential for different behaviors. Fundamental to nervous system function is the precise establishment and maintenance of neuron type-specific gene expression programs. Integral components of such programs are effector genes that encode proteins critical for neuronal function (e.g., neurotransmitter [NT] biosynthesis components, ion channels, NT receptors, neuropeptides) (Deneris and Hobert, 2014, Hobert, 2008, Hobert, 2011, Hobert, 2016). These effector genes, referred to as terminal identity genes herein, are expressed continuously, from development throughout life, in post-mitotic neurons in a combinatorial fashion (Hobert, 2008). Hence, it is the unique overlap of many effector gene products in a specific neuron type that determines its distinct terminal identity, and thereby function. However, the molecular mechanisms that select, in individual neuron types, which terminal identity genes should be expressed and which ones should be repressed are poorly defined. Understanding how neuron type-specific batteries of terminal identity genes are established during development and, perhaps most importantly, maintained throughout life represents one key step towards understanding how individual neuron types become and remain functional. Providing molecular insights into this fundamental problem may also have important biomedical implications, as defects in terminal identity gene expression are associated with a variety of neurodevelopmental and neurodegenerative disorders (Deneris and Hobert, 2014, Shibuya et al., 2011, Imbrici et al., 2013, Sgado et al., 2011).

Seminal genetic studies in multiple model systems revealed a widely employed molecular principle: neuron type-specific transcription factors (TFs) often coordinate the expression of "desired" terminal identity genes with the exclusion of "unwanted" terminal identity genes (Morey et al., 2008, Sagasti et al., 1999, Britanova et al., 2008, Cheng et al., 2004, Kala et al.,

2009, Lopes et al., 2012, Mears et al., 2001, Nakatani et al., 2007). These TFs exert a dual role: they are not only required to induce a specific set of terminal identity features critical for the function of a given neuron type, but also to simultaneously prevent expression of molecular features normally reserved for other neuron types. Consequently, neurons lacking these TFs fail to acquire their unique terminal identity, and concomitantly gain features indicative of alternative identities. For example, mouse striatal cholinergic interneurons lacking *Lhx7* lose their terminal identity and acquire molecular features indicative of GABAergic interneuron identity (Lopes et al., 2012). In midbrain neurons, removal of *Gata2* results in loss of GABAergic identity and simultaneous gain of terminal identity features specific to glutamatergic neurons (Kala et al., 2009). However, the molecular mechanisms underlying the dual function of most neuron type-specific TFs remain poorly defined. How can the same TF, within the same cell, promote a specific identity and simultaneously prevent molecular features of alternative neuronal identities? In principle, the same TF can simultaneously operate as direct activator of neuron type-specific terminal identity genes and direct repressor of alternative identity genes (Lodato et al., 2014, Wyler et al., 2016). Another possibility is indirect regulation. For example, a neuron type-specific TF can prevent adoption of alternative identity features by repressing expression of an intermediary TF that normally promotes such features (Cheng et al., 2004). Other mechanisms involving TF competition for cell type-specific enhancers or cell type-specific TF-TF interactions have also been described (see Discussion) (Andzelm et al., 2015, Gordon and Hobert, 2015, Rhee et al., 2016, Thaler et al., 2002). It remains unclear, however, whether these mechanisms of action of neuron type-specific TFs are broadly applicable in the nervous system. Although the aforementioned studies begin to explain how neurons select their terminal identity features during development (Morey et al., 2008, Sagasti et al., 1999, Britanova et al., 2008,

Cheng et al., 2004, Kala et al., 2009, Lopes et al., 2012, Mears et al., 2001, Nakatani et al., 2007), the function of neuron type-specific TFs is rarely assessed during post-embryonic stages. Hence, the molecular mechanisms that maintain neuronal terminal identity features, and thereby neuronal function, are largely unknown. Is the same neuron type-specific TF continuously required, from development through adulthood, to induce a specific set of terminal identity genes and simultaneously prevent "unwanted" features? Alternatively, a given neuron type could employ different mechanisms for selection (during development) and maintenance (through adulthood) of its function-defining terminal features. Addressing this fundamental problem has been challenging in the vertebrate nervous system, in part due to its inherent complexity and difficulty to track individual neuron types with single-cell resolution from embryo to adult.

To study how neurons select and maintain their terminal identity features, we use as a model the well-defined motor neuron (MN) subtypes of the *Caenorhabditis elegans* ventral nerve cord (equivalent to vertebrate spinal cord). Five cholinergic (DA, DB, VA, VB, AS) and two GABAergic (DD, VD) MN subtypes are located along the nerve cord and control locomotion (**Fig. 4.1A**) (Von Stetina et al., 2006, White et al., 1986). Because they are present in both *C. elegans* sexes (males and hermaphrodites), we will refer to them as "sex-shared" MNs. In addition, there are two subtypes of "sex-specific" cholinergic MNs: the hermaphrodite-specific VC neurons control egg laying (Portman, 2017, Schafer, 2005), and the male-specific CA neurons are required for mating (Schindelman et al., 2006) (**Fig. 4.1A**). In addition to distinct morphology and connectivity, each subtype can be molecularly defined by the combinatorial expression of known terminal identity genes, such as ion channels, NT receptors, and neuropeptides (**Fig. 4.1B**). An extensive collection of transgenic reporter *C. elegans* animals for MN subtype-specific terminal identity genes is available, thereby providing a unique opportunity

to investigate, at single-cell resolution, the effects of TF gene removal on developing and adult MNs.

UNC-3, the sole *C. elegans* ortholog of the Collier/Olf/Ebf (COE) family of TFs, is selectively expressed in all sex-shared cholinergic MNs of the nerve cord (**Fig. 4.1B**) (Kratsios et al., 2017, Kratsios et al., 2012, Pereira et al., 2015, Prasad et al., 2008, Prasad et al., 1998). Animals lacking *unc-3* display striking locomotion defects (Brenner, 1974). UNC-3 is known to directly activate a large battery of terminal identity genes expressed either in all sex-shared cholinergic MNs (e.g., the NT identity genes *unc-17/* VAChT and *cha-1/* ChAT), or in certain subtypes (e.g., ion channels, NT receptors, signaling molecules) (Kratsios et al., 2012) (**Fig. 4.1B-C**). Based on its ability to broadly co-regulate many distinct terminal identity features, *unc-3* has been classified as a terminal selector gene (Hobert, 2008). Besides its well-established function as activator of terminal identity genes in cholinergic MNs, whether and how UNC-3 can prevent expression of terminal features of alternative neuronal identities remains unclear.

Here, we describe a dual role for UNC-3 that enables sex-shared cholinergic MNs to select during development and maintain throughout life their terminal identity features. We find that UNC-3 is continuously required - from development through adulthood - not only to activate cholinergic MN identity genes, but also to prevent expression of terminal features normally reserved for other MN subtypes of the nerve cord, namely the sex-shared GABAergic VD neurons and sex-specific cholinergic MNs (CA, VC). These findings lend support to the notion that neuron type-specific TFs can promote a specific identity and simultaneously suppress features reserved for alternative, but functionally related, neuronal identities.

To uncover the molecular mechanism underlying the dual role of UNC-3, we conducted an unbiased genetic screen, which led to the identification of the Hox protein LIN-39

(Scr/Dfd/Hox4-5) as the intermediary factor necessary for expression of alternative neuronal identity features (e.g., VD, VC) in *unc-3*-depleted MNs. Unlike previously described cases of TFs that act indirectly to prevent alternative neuronal identities by repressing intermediary factors (discussed earlier), UNC-3 does not repress *lin-39* and both factors are co-expressed in cholinergic MNs. However, UNC-3 antagonizes the ability of LIN-39 to induce terminal features of alternative identities. Intriguingly, UNC-3 also synergizes with LIN-39 to co-activate multiple terminal identity features specific to cholinergic MNs. Consequently, loss of *unc-3* causes a switch in the transcriptional targets of LIN-39, thereby alternative, not cholinergic MN-specific, terminal identity features become activated and locomotion defects occur. Given that terminal selectors and Hox proteins are expressed in a multitude of neuron types across species (Deneris and Hobert, 2014, Hobert and Kratsios, 2019, Philippidou and Dasen, 2013, Estacio-Gomez and Diaz-Benjumea, 2014), the strategy of a terminal selector preventing a Hox transcriptional switch may constitute a general principle for safeguarding neuronal identity throughout life.

4.3 Results

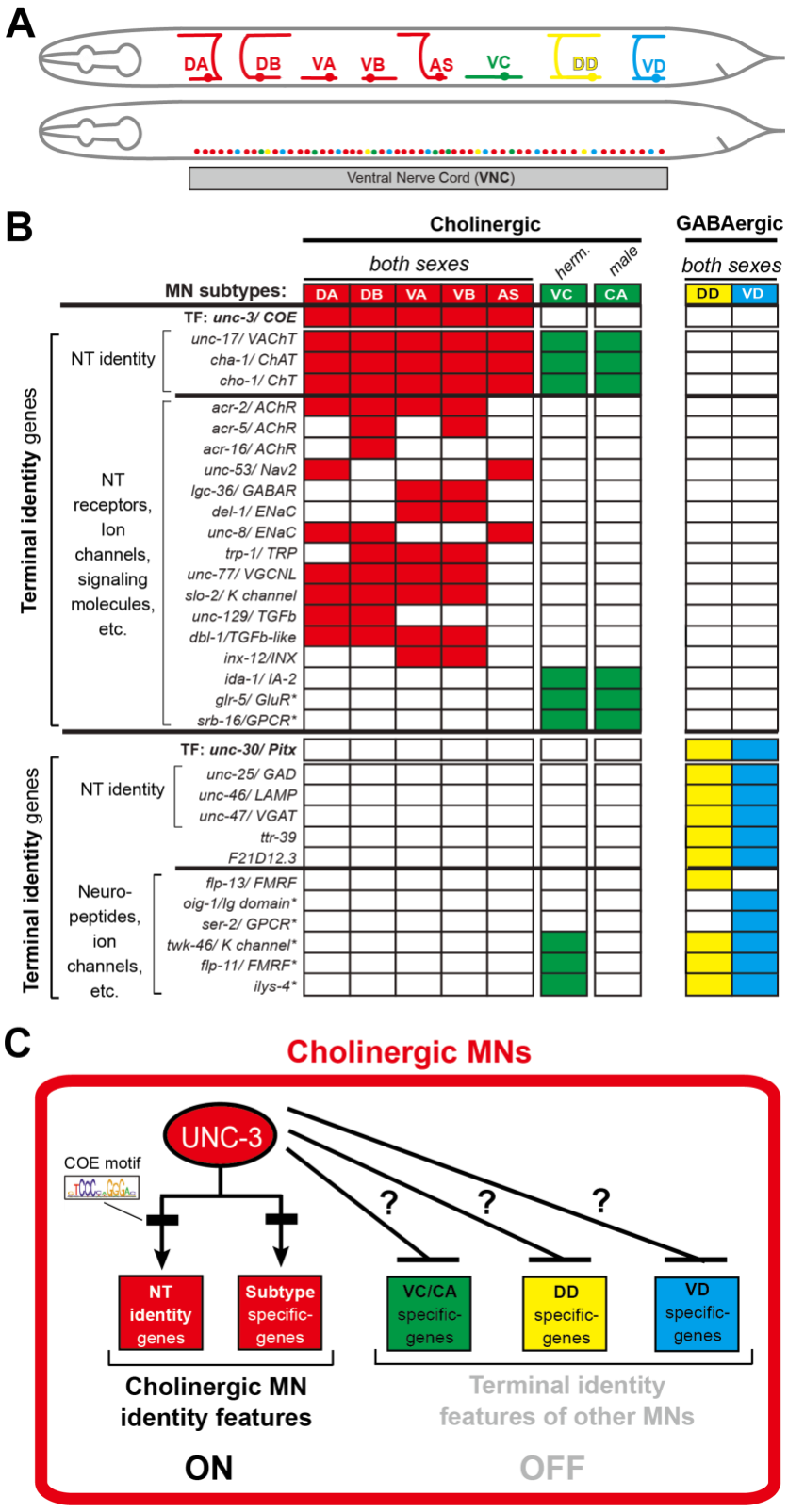

Figure 4.1 An extensive collection of terminal identity markers for distinct motor neuron subtypes of the *C. elegans* ventral nerve cord. **A:** Schematic showing distinct morphology for

each motor neuron subtype in the *C. elegans* hermaphrodite. Below, colored dots represent the invariant cell body position of all MNs of the ventral nerve cord (VNC). Red: 39 sex-shared cholinergic MNs (DA2-7 = 6 neurons, DB3-7 = 5, VA2-11 = 10, VB3-11 = 9, AS2-10 = 9); Green: 6 hermaphrodite-specific VC MNs; Yellow: 4 sex-shared GABAergic DD neurons (DD2-5 = 4); Blue: 9 sex-shared GABAergic VD neurons (VD3-11 = 9). With the exception of VC, all other subtypes have 1-3 extra neurons located at the flanking ganglia (retrovesicular and pre-anal) of the VNC (not shown). Individual neurons of each subtype intermingle along the VNC. **B:** Table summarizing expression of terminal identity markers for VNC MNs. The sex-shared GABAergic MNs (DD, VD) and the sex-specific MNs (VC, CA) do not express UNC-3. Conversely, the sex-shared cholinergic MNs (DA, DB, VA, VB, AS) and the sex-specific MNs (VC, CA) do not express UNC-30/Pitx. For the genes indicated with an asterisk (*), a detailed expression pattern is provided in **Figure 4.11**. Of note, the male-specific MNs of the CP subtype are also not shown. **C:** Schematic that summarizes the known function of UNC-3 (activator of cholinergic MN identity genes) and the question under investigation: does UNC-3 prevent expression of terminal identity features reserved for other MN subtypes?

4.3.1 UNC-3 has a dual role in distinct populations of ventral nerve cord (VNC) motor neurons

Neuron type-specific TFs often promote a specific identity and simultaneously suppress features reserved for other, functionally related neuronal types (Arlotta and Hobert, 2015). To test this notion for UNC-3, it was essential to identify a set of terminal identity markers for all *unc-3*-negative MN subtypes of the VNC, namely the GABAergic (VD, DD) and sex-specific (VC, CA) MNs (**Fig. 4.1B**). We undertook a candidate gene approach and examined the precise expression pattern of terminal identity genes (e.g., NT receptors, signaling proteins, ion channels, neuropeptides) reported to be expressed in *unc-3*-negative MNs (www.wormbase.org). In total, we carefully characterized at single-cell resolution the expression of 15 genes in wild-type animals of both *C. elegans* sexes at the fourth larval stage (L4) (see Methods and **Figure 4.11**). This analysis provided 9 terminal identity markers highly specific to *unc-3*-negative MNs that fall into four categories (**Fig. 4.1B**): (**a**) two VD-specific markers (*ser-2* / serotonin receptor [ortholog of HTR1D]*; oig-1 /* one Ig domain protein), (**b**) one DD-specific marker (*flp-13 /* FMRF-like neuropeptide*)*, (**c**) three markers for sex-specific (VC in hermaphrodites, CA in males) MNs (*glr-5* / glutamate receptor [ortholog of GRID/GRIK]; *srb-16* / serpentine GPCR receptor; *ida-1* / ortholog of protein tyrosine phosphatase PTPRN), and (**d**) three markers expressed in both GABAergic subtypes (DD, VD) and sex-specific MNs (*flp-11* / FRMR-like neuropeptide, *twk-46* / potassium channel [ortholog of KCNK1], *ilys-4 /* invertebrate-type lysozyme).

These 9 markers enabled us to test whether *unc-3*-depleted MNs gain expression of terminal features normally reserved for other MN subtypes. By using animals carrying a strong loss-of-function (null) allele for *unc-3 (n3435)* (Prasad et al., 2008), we first assessed any putative

effects on terminal markers for the sex-shared GABAergic MNs (DD, VD). Although the DD-specific marker *flp-13* is unaffected (**Figure 4.12A**), ectopic expression of the VD-specific markers (*ser-2, oig-1*) was observed in *unc-3*-depleted MNs (**Fig. 4.2A-B**). Interestingly, this ectopic expression was region-specific, observed in cholinergic MNs of the mid-body region of the VNC with 100% penetrance (**Fig. 4.2A-B**). Importantly, 12.1 ± 2.6 (mean ± STDV) out of the 39 *unc-3*-depleted MNs in the VNC were ectopically expressing these VD markers, suggesting that not all *unc-3*-depleted MNs acquire VD terminal identity features. Given that GABAergic and cholinergic MNs are generated in normal numbers in *unc-3* animals (Kratsios et al., 2012), the increase in the number of neurons expressing the VD markers cannot be attributed to early developmental defects affecting MN numbers. We next asked whether these ~12 MNs adopt additional VD terminal identity features, such as expression of genes involved in GABA biosynthesis (*unc-25/GAD* and *unc-47/VGAT)*, or selectively expressed in GABAergic MNs (*ttr-39, klp-4*). However, this does not appear to be the case, arguing against a complete cell fate switch (**Figure 4.12A**). We conclude that, in the absence of *unc-3*, cholinergic MNs not only lose their original terminal identity, but a third of them (~12 out of 39) in the mid-body VNC region also gain some terminal identity features normally reserved for the sex-shared VD neurons (**Fig. 4.1**). We will refer to these *unc-3*-depleted MNs as "VD-like" (**Fig. 4.2G**). We also uncovered the identity of these cells across multiple *unc-3* mutant animals and conclude that it is the same 12 neurons that become VD-like across animals (**Figure 4.12B**).

To test whether UNC-3 also prevents expression of terminal identity features of sex-specific cholinergic MNs, we examined three VC-specific terminal markers (*glr-5, srb-16, ida-1* in **Fig. 4.1B**) in hermaphrodite nematodes lacking *unc-3*. Again, we observed region-specific effects with 100% penetrance in the same cells across multiple animals (**Fig. 4.2C-D, Figure 4.12B)**.

All three markers were ectopically expressed in 10.5 ± 3.7 (mean ± STDV) of the 39 *unc-3*-depleted MNs located in the mid-body region of the VNC (**Fig. 4.2C-D**). These results are in agreement with a previous study reporting ectopic *ida-1* expression in *unc-3*-depleted MNs (Prasad et al., 2008). If these ~11 MNs fully adopt the VC terminal identity, then they should also express genes necessary for acetylcholine biosynthesis since VC neurons are cholinergic. However, this is not the case as expression of *unc-17*/VAChT and *cho-1*/ChT is dramatically affected in *unc-3*-depleted MNs (Kratsios et al., 2012). These data suggest that ~11 of the 39 *unc-3*-depleted MNs in the mid-body VNC region adopt some, but not all, VC terminal identity features. We will therefore refer to these *unc-3*-depleted MNs as "VC-like" (**Fig. 4.2G**).

Are the VD-like and VC-like neurons in *unc-3* hermaphrodites distinct populations? To test this, we generated *unc-3* hermaphrodites that carry a green fluorescent reporter for VC terminal identity (*ida-1::gfp*) and a red reporter for VD identity (*ser-2::rfp*). We found no overlap of the two reporters, indicating that the VD-like and VC-like neurons represent two distinct populations (**Fig. 4.2 E-F**). We further corroborated this result by taking advantage of the invariant lineage and cell body position of all MNs along the *C. elegans* nerve cord (**Figure 4.12B**). Of note, the VC-like population appears to be lineally related to VC neurons, whereas the VD-like population is not lineally related to VD neurons (**Figure 4.12B-C**). Lastly, terminal identity markers normally expressed in both VD and VC neurons (*flp-11, ilys-4, twk-46*) display an additive effect in *unc-3* mutants, as they are ectopically expressed in both VD-like and VC-like populations, further suggesting the presence of distinct *unc-3* MN populations (**Figure 4.12D-E**).

To summarize, there are 39 *unc-3*-expressing MNs along the wild-type nerve cord in hermaphrodites. While loss of *unc-3* uniformly leads to loss of cholinergic identity in all these MNs (Kratsios et al., 2012), one population (~12 MNs) acquires VD-like molecular features,

while a second population (~11 MNs) acquires VC-like molecular features, uncovering a dual role of UNC-3 in these populations (**Fig. 4.2G**). Of note, the remaining MNs (~16) in the VNC of *unc-3* mutants [39 - (12 VD-like + 11 VC-like) = 16] do not gain either VD or VC terminal identity features.

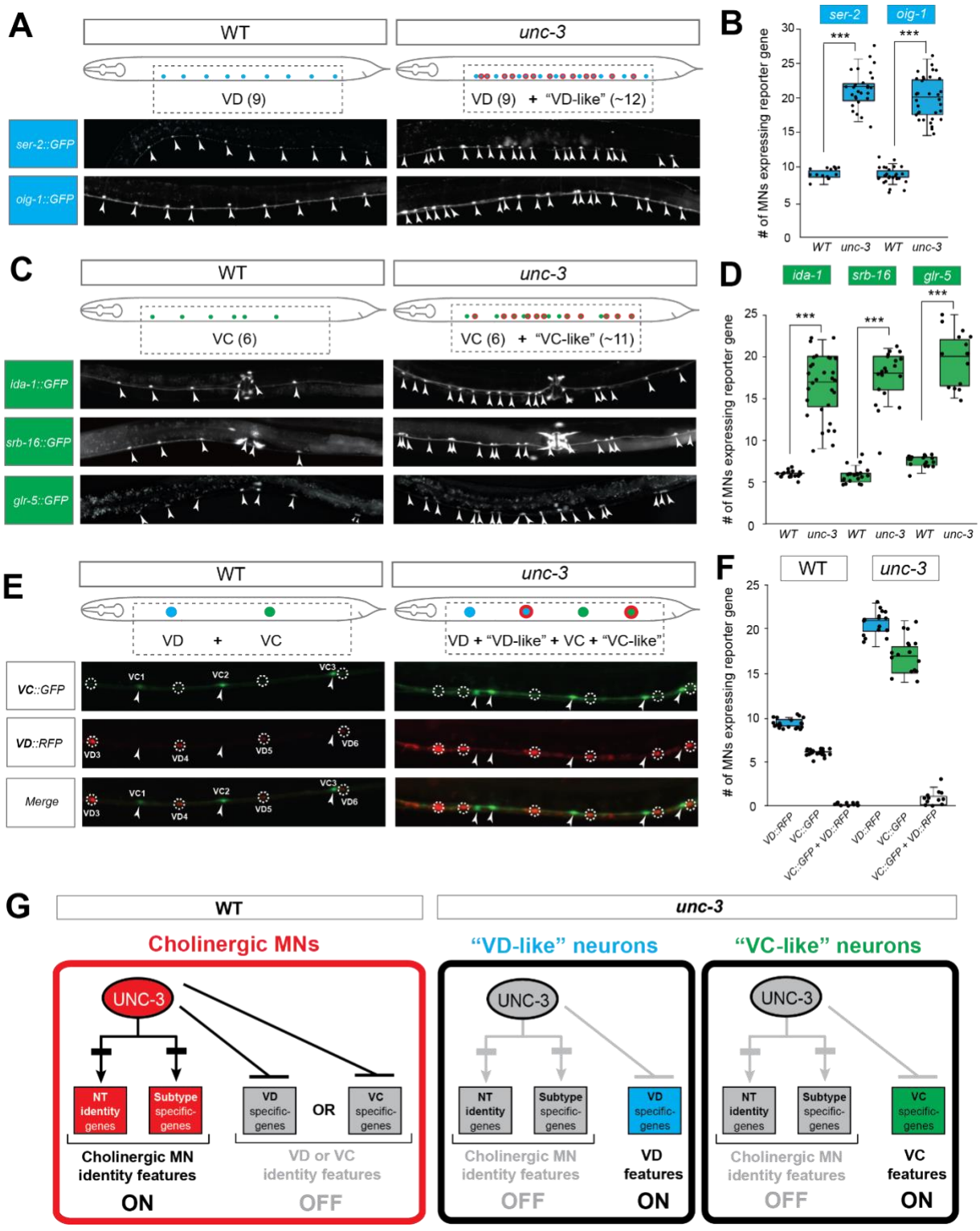

Figure 4.2: UNC-3 has a dual role in cholinergic ventral cord motor neurons**. A:** Terminal identity markers of VD neurons (*ser-2*, *oig-1*) are ectopically expressed in *unc-3*-depleted MNs. Representative images of larval stage 4 (L4) hermaphrodites are shown. Similar results were obtained in adult animals. Arrowheads point to MN cell bodies with *gfp* marker expression.

Green fluorescence signal is shown in white for better contrast. Dotted black box indicates imaged area. **B**: Quantification of VD markers (*ser-2, oig-1*) in WT and *unc-3 (n3435)* at L4. N > 15. *** $p < 0.001$. For details on box plots, see Materials and Methods. **C:** Terminal identity markers of VC neurons (*ida-1, srb-16, glr-5*) are ectopically expressed in *unc-3*-depleted MNs. Representative images of larval stage 4 (L4) hermaphrodites are shown. Similar results were obtained in adult animals. Arrowheads point to MN cell bodies with *gfp* marker expression. Green fluorescence signal is shown in white for better contrast. Dotted black box indicates imaged area. **D**: Quantification of VC markers (*ida-1, srb-16, glr-5*) in WT and *unc-3 (n3435)* at L4. Individual data points are dot-plotted. N > 15. *** $p < 0.001$. **E:** Distinct MNs acquire VC-like or VD-like terminal identity features in *unc-3 (n3435)* mutants. The VC marker in green (*ida-1::gfp*) and the VD marker in red (*ser-2::rfp*) do not co-localize in WT or *unc-3 (n3435)* mutants. Representative images are shown. Individual VC/VC-like and VD/VD-like neurons are circled (VC: dotted circles; VD: arrowheads) to highlight that an individual MN never expresses both markers. **F:** Quantification of data shown in E. N > 16. **G:** Schematic that summarizes the dual role of *unc-3*. Apart from activating cholinergic MN terminal identity genes, UNC-3 prevents expression of VD and VC terminal features in distinct cells ("VD-like" versus "VC-like").

4.3.2 The dual role of UNC-3 in cholinergic MNs extends to both *C. elegans* sexes

To test whether the dual function of UNC-3 applies to both sexes, we extended our analysis to *C. elegans* males. First, we showed that loss of *unc-3* in males resulted in loss of several cholinergic MN terminal identity features (**Figure 4.13**). Second, we observed ectopic expression of VD-specific terminal identity markers (*oig-1, ser-2*) in 11.9 ± 3.9 (mean ± STDV) out of the 39 *unc-3*-depleted MNs, indicating the presence of "VD-like" neurons in the male nerve cord (**Figure 4.13**). Lastly, we asked whether *unc-3* loss leads to ectopic expression of terminal identity markers (*ida-1, srb-16, glr-5*) for male-specific CA neurons. Indeed, we found this to be the case (**Figure 4.13**), suggesting the adoption of "CA-like" features by a population of *unc-3*-depleted MNs. Similar to hermaphrodites, these VD-like and CA-like cells were observed in the mid-body region of the male nerve cord with 100% penetrance (**Figure 4.13**).

Taken together, our findings uncover a dual role for UNC-3 in sex-shared cholinergic MNs. UNC-3 is not only required to activate cholinergic MN identity genes (Kratsios et al., 2012), but also to prevent expression of molecular features normally reserved for three other, functionally related neuronal subtypes of the nerve cord (VD, VC, CA). In both sexes, UNC-3 prevents expression of select terminal features of VD neurons in a specific population of cholinergic MNs. In a second population, UNC-3 prevents expression of terminal features normally reserved for sex-specific MNs, i.e., VC features in hermaphrodites and CA features in males. In the ensuing sections, we focus our analysis on *C. elegans* hermaphrodites to dissect the molecular mechanism underlying the dual role of UNC-3.

4.3.3 UNC-3 is continuously required to prevent expression of VD and VC terminal identity features

Neuron type-specific TFs that promote a specific identity and simultaneously prevent alternative features have been previously described (see Introduction). However, whether this dual role is required transiently (during development), or continuously (throughout life) remains unclear. The UNC-3 case provides an opportunity to distinguish between these two possibilities because ectopic expression of VC and VD features is observed at both larval and adult stages in *unc-3* null animals (**Fig. 4.13**). To this end, we employed the auxin-inducible degron (AID) system that enables depletion of UNC-3 in a temporally controlled manner (Zhang et al., 2015). This system requires tagging the UNC-3 protein with the AID degron fused to a fluorescent reporter gene (mNeonGreen, mNG). When UNC-3::mNG::AID and the plant-specific F-box protein TIR1 are co-expressed in MNs (by crossing animals carrying the *unc-3::mNG::AID* allele with *eft-3::TIR1* transgenic animals), application of the plant hormone auxin on these double transgenic animals induces degradation of UNC-3::mNG::AID (**Fig. 4.3A-C**). Auxin administration at the L4 stage (last larval stage before adulthood) on *unc-3::mNG::AID*; *eft-3::TIR1* animals resulted in a dramatic depletion of UNC-3 at day 1 adult animals (24 hours after auxin). UNC-3 depletion was accompanied by ectopic expression of VD and VC terminal identity features in nerve cord MNs, demonstrating a post-embryonic requirement for UNC-3 (**Fig. 4.3D-E**). Similar results were obtained when auxin was applied at different time points (**Fig. 4.3D**, legend). These findings suggest that UNC-3 is continuously required to prevent expression of VD and VC terminal identity features.

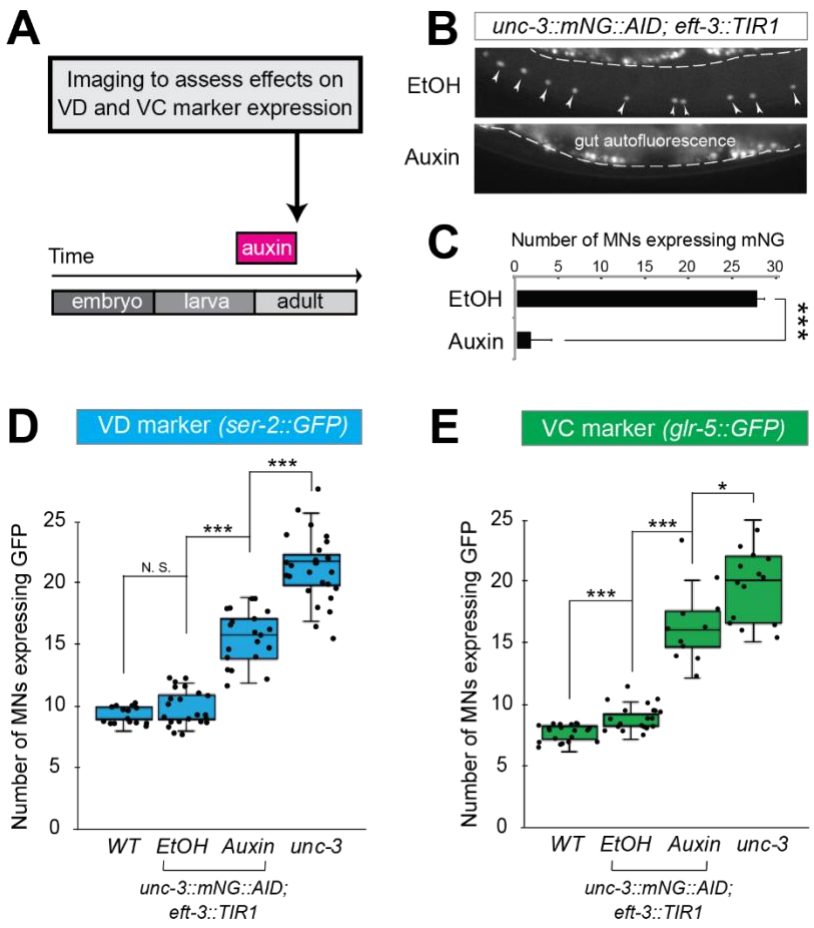

Figure 4.3 UNC-3 is continuously required to prevent expression of VD and VC terminal identity features. **A:** Schematic showing time window of auxin administration. **B:** Animals of the *unc-3::mNG::AID; eft-3::TIR1* genotype were either administered ethanol (EtOH) or auxin at the L4 stage. Twenty four hours later, expression of endogenous *unc-3* reporter (*unc-3::mNG::AID*) is severely reduced in the nuclei of VNC MNs (arrowheads) at the young adult stage (day 1). The same exact region was imaged in EtOH- and auxin-treated worms. mNG green fluorescent signal is shown in white for better contrast. White dotted line indicates the boundary of intestinal tissue (gut), which tends to be autofluorescent in the green channel. **C:** Quantification of number of MNs expressing the *unc-3::mNG::AID* reporter after EtOH (control) and auxin treatment. N > 12. *** $p < 0.001$. **D**: Auxin or ethanol (control) were administered at larval stage 3 (L3) on *unc-3::mNG::AID; eft-3::*TIR1 animals carrying the VD marker *ser-2::gfp*. Images were taken at the young adult stage (day 1.5). A significant increase in the number of MNs expressing the VD marker was evident in the auxin-treated animals compared to EtOH-treated controls. For comparison, quantification is provided for *ser-2::gfp* expressing MNs of wild-type animals and *unc-3(n3435)* mutants. Similar results were obtained when auxin was applied at L4 or day 1 adult animals. N >20. *** $p < 0.001$. **E**: Auxin or ethanol (control) were administered at larval stage 4 (L4) on *unc-3::mNG::AID; eft-3::*TIR1 animals carrying the VC marker *glr-5::gfp*. Images were taken at the young adult stage (day 2). A significant increase in the number of MNs expressing the VC marker was evident in the auxin-treated animals compared to EtOH-treated controls. N > 11. * $p < 0.05$; *** $p < 0.001$.

4.3.4 UNC-3 acts indirectly to prevent expression of VD and VC terminal identity genes

How does UNC-3 activate cholinergic MN identity genes and simultaneously prevent terminal features of alternative MN identities (e.g., VD, VC) (**Fig. 4.2G**)? Based on previous reports, the same TF, within the same neuron, can act as a direct activator for a set of genes and a direct repressor for another set of genes (Lodato et al., 2014, Wyler et al., 2016, Borromeo et al., 2014). While it is known that UNC-3 acts directly – through its cognate binding site (COE motif) – to activate expression of a large battery of cholinergic MN identity genes, we did not find any COE motifs in the *cis*-regulatory region of VD or VC terminal identity genes (**Table 4.2**). This contrasts the previously described function of UNC-3 as direct repressor (through the COE motif) of terminal identity genes in the chemosensory ASI neurons of *C. elegans* (Kim et al., 2005).

To test the possibility of indirect repression via an intermediary factor, we focused on VD neurons because, unlike VC neurons, a known activator of VD terminal features has been reported (Cinar et al., 2005, Eastman et al., 1999, Jin et al., 1994). In wild-type animals, the TF UNC-30, ortholog of human PITX1-3, is required to induce VD terminal identity genes. Since UNC-30 is not expressed in cholinergic MNs (Jin et al., 1994), we hypothesized that UNC-3 prevents expression of UNC-30/PITX, leading to inactivation of VD terminal identity genes. However, this is not the case because: (1) ectopic *unc-30* expression is not observed in *unc-3*-depleted MNs, and (2) the ectopic expression of the VD marker (*ser-2*) in *unc-3* mutants was not abolished in *unc-3; unc-30* double mutants (**Figure 4.14**). These observations suggest that UNC-3 may act indirectly to prevent expression of VD and VC terminal identity genes through as yet unknown intermediary factors.

4.3.5 The mid-body Hox protein LIN-39 (Scr/Dfd/Hox4-5) is the intermediary factor necessary for ectopic expression of VD and VC features in *unc-3* mutants

If the hypothesis of indirect repression is correct, mutation of the intermediary factor(s) in the *unc-3* mutant background would selectively eliminate ectopic expression of VD and/or VC terminal identity genes in *unc-3*-depleted MNs. To identify such factor(s), we embarked on an unbiased genetic screen. For the screen, we chose a transgenic *gfp* reporter strain for *flp-11*, an FMRF-like neuropeptide-encoding gene expressed in both VD and VC neurons (**Fig. 4.1B, Figure 4.11**), which is markedly affected by UNC-3 (**Fig. 4.4A-B, Figure 4.12D-E**). We mutagenized *unc-3 (n3435); flp-11::gfp* animals with ethyl methanesulfonate (EMS) and visually screened ~4,200 haploid genomes for mutants in which ectopic *flp-11::gfp* expression in *unc-3*-depleted MNs is suppressed. We isolated one mutant allele (*kas1*) (**Fig. 4.4A-B**). The phenotype was 100% penetrant as all *unc-3 (n3435); flp-11::gfp* animals carrying *kas1* in homozygosity consistently displayed a dramatic reduction in ectopic *flp-11* expression.

Gross morphological examination of *unc-3 (n3435); kas1*; *flp-11::gfp* hermaphrodites revealed that, unlike *unc-3 (n3435); flp-11::gfp* animals, the introduction of *kas1* is accompanied by a lack of the vulva organ (vulvaless phenotype). Upon a literature survey for TF mutants that are vulvaless, we stumbled across the mid-body Hox gene *lin-39* (ortholog of Dfd/Scr in flies and Hox4-5 in vertebrates) (Aboobaker and Blaxter, 2003, Clark et al., 1993), and hypothesized that the molecular lesion of *kas1* may lie in the *lin-39* locus. Indeed, Sanger sequencing uncovered a point mutation on the splice acceptor site (WT: A<u>G</u> > *kas1*: A<u>A</u>) in the second intron of *lin-39* (**Fig. 4.4C**). Similar to *unc-3 (n3435); kas1* animals, *unc-3 (n3435)* mutants carrying a previously published strong loss-of-function (premature STOP) allele of *lin-39 (n1760)* (Clark et al., 1993) displayed the same loss of ectopic *flp-11* expression (**Fig. 4.4A-C**), suggesting that

kas1 is a loss-of-function mutation of *lin-39*. The ectopic expression of *flp-11* in *unc-3(n3435); kas1* animals can be, at least partially, rescued by (1) selective expression of *lin-39* cDNA in cholinergic MNs, and (2) introduction of the *lin-39* wild-type locus in the context of a ~30kb genomic clone (fosmid) (**Figure 4.14**), corroborating that the *kas1* lesion in the *lin-39* locus is the phenotype-causing mutation.

Because *flp-11* is expressed in both VD and VC neurons, we next tested whether *lin-39* is required for ectopic expression of VD-specific (*ser-2, oig-1*) and VC-specific (*ida-1*) terminal identity genes in *unc-3*-depleted MNs. We found this to be the case by either generating *unc-3 (n3435); lin-39 (n1760)* double mutants (for VD markers) or by performing cholinergic MN-specific RNAi for *lin-39* in *unc-3 (n3435)* animals (for VC marker) (**Fig. 4.4D**). RNAi was necessary because VC neurons do not survive in *lin-39 (n1760)* animals (Potts et al., 2009), and the use of the *n1760* allele could confound our VC marker quantifications. Of note, all other nerve cord MN subtypes are normally generated in *lin-39 (n1760)* single and *unc-3 (n3435); lin-39 (n1760)* double mutants (Stefanakis et al., 2015), indicating that suppression of the *unc-3* phenotype, i.e., loss of ectopic VD gene expression in the double mutants is not due to MN elimination. Taken together, our genetic screen identified the mid-body Hox gene *lin-39* to be necessary for ectopic expression of both VD and VC terminal features in *unc-3*-depleted MNs (**Fig. 4.4F**).

Interestingly, this finding contradicts our initial hypothesis of UNC-3 repressing an intermediary TF in order to prevent expression of VD and VC features because *lin-39* is co-expressed with (not repressed by) *unc-3* in wild-type cholinergic MNs at the mid-body region of the VNC (**Fig. 4.4E**), as evident by our single-cell analysis of *unc-3* and *lin-39* reporters (**Figure 4.15**). Of note, 28 cholinergic MNs co-express *unc-3* and *lin-39*, which is in close agreement with the total

number of VD-like (12.1 ± 2.6) and VC-like (10.5 ± 3.7) cells observed in *unc-3* mutants (**Fig. 4.4E-F**). Moreover, we found that *lin-39* acts cell-autonomously as cholinergic MN-specific RNAi against *lin-39* in *unc-3 (n3435)* animals resulted in a significant reduction of ectopic terminal identity marker (*flp-11, ida-1)* expression (**Fig. 4.4B-D**). In the following Results sections, we describe the molecular mechanism through which UNC-3 and LIN-39/Hox select and maintain throughout life key terminal features of cholinergic MNs (**Fig. 4.4F**).

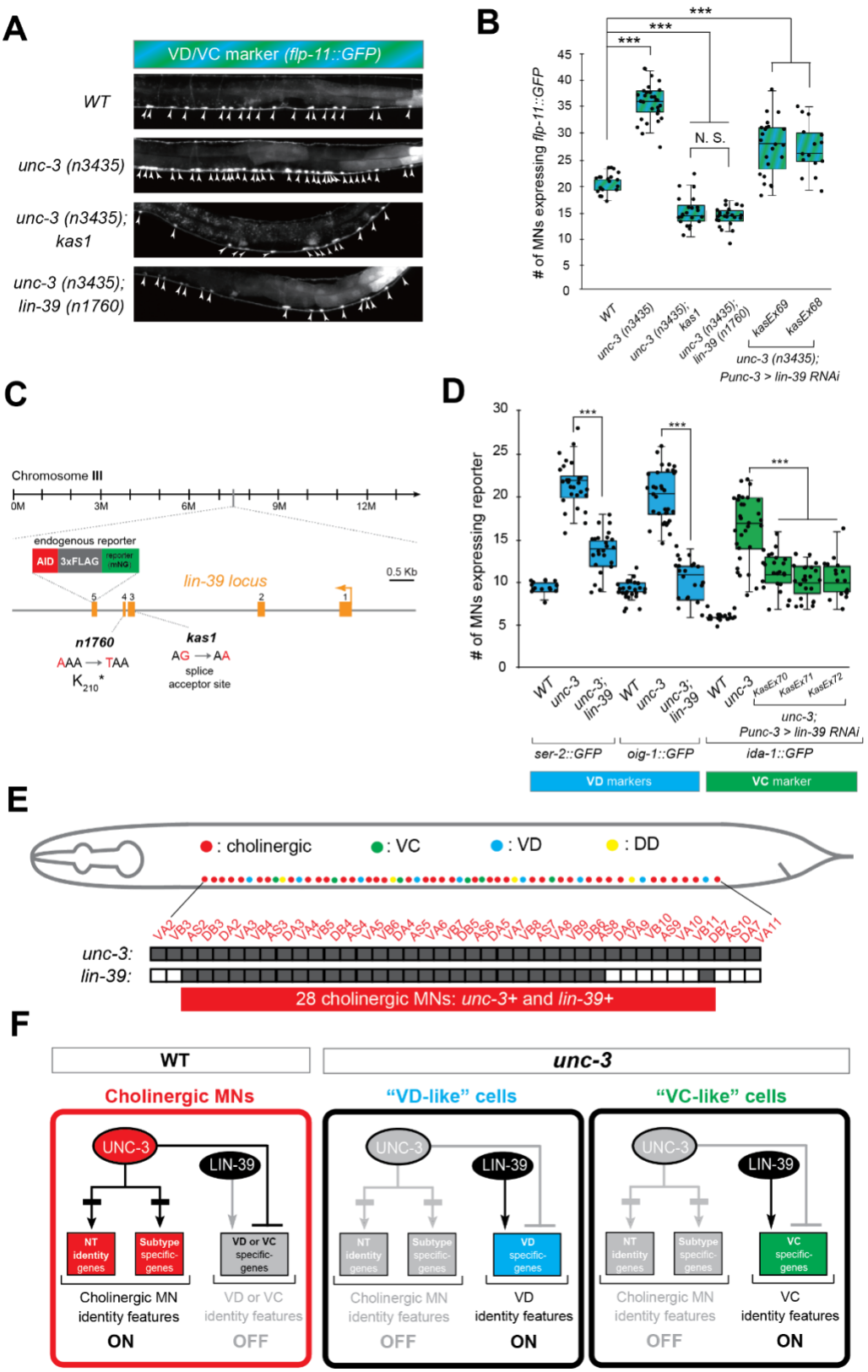

Figure 4.4 A genetic screen identifies the mid-body Hox protein LIN-39 (Scr/Dfd/Hox4-5) as necessary for ectopic expression of VD and VC terminal features. **A:** Representative images of

L4-stage WT, *unc-3(n3435)*, *unc-3(n3435); kas1*, and *unc-3(n3435); lin-39(n1760)* animals carrying *flp-11::gfp* (VD/VC marker). Arrowheads point to MN cell bodies with *gfp* marker expression. **B**: Quantification graph summarizing results from panel A. The two right-most bars show quantification of two independent transgenic lines driving *lin-39 RNAi* specifically in cholinergic MNs (*Punc-3 > lin-39 RNAi*) of *unc-3 (n3435)* mutants. N >15 . *** $p < 0.001$. N.S: not significant. **C:** Genetic locus of *lin-39*. Molecular lesions for *kas1* and *n1760* alleles are shown, as well as the *AID::3xFLAG::mNG* cassette inserted at the C-terminus (endogenous reporter). **D**: Quantification of two VD (*ser-2::gfp, oig-1::gfp*) and one VC (*ida-1::gfp*) markers in WT, *unc-3 (n3435)*, *unc-3(n3435); lin-39(n1760)* animals at L4. The three right-most bars show quantification of three independent transgenic lines driving *lin-39 RNAi* specifically in cholinergic MNs (*Punc-3 > lin-39 RNAi*) of *unc-3 (n3435)* mutants. N > 15. *** $p < 0.001$. **E**: Summary of *unc-3* and *lin-39* expression in cholinergic MNs. See **Figure 4.15** for raw data. **F**: Schematic that summarizes our findings. In the wild type (Faumont et al.) panel on the left, *lin-39* is normally expressed in cholinergic MNs but unable to induce expression of VD or VC genes. In the *unc-3* mutant, *lin-39* is now able to induce expression of alternative identity features (VD or VC) in distinct MN populations

4.3.6 UNC-3 prevents a switch in the transcriptional targets of LIN-39 in cholinergic motor neurons

What is the function of LIN-39 in wild-type cholinergic MNs of the VNC? Our previous findings suggested that LIN-39 and UNC-3, together with another mid-body Hox protein, MAB-5 (Antp/Hox6-8) (Salser et al., 1993), act synergistically to control expression of two cholinergic MN terminal identity genes (*unc-129,* ortholog of human BMP; *del-1* / Degenerin-like sodium channel [ortholog of human SCNN1G]) (Kratsios et al., 2017). To test the extent of this synergy, we examined in *lin-39* and *mab-5* null animals the expression of 4 additional cholinergic MN terminal identity genes known to be controlled by UNC-3 (*acr-2 / nicotinic acetylcholine receptor; dbl-1 / DPP/BMP-like; unc-77 / sodium channel [ortholog of human NALCN], slo-2 / potassium sodium-activated channel [ortholog of human KCNT1]*) (Kratsios et al., 2012). In all 4 cases, we found a statistically significant decrease in *lin-39* mutants, and this effect was exacerbated in *lin-39; mab-5* double mutants (**Fig. 4.5A-B**), indicating that the synergy of LIN-39 with MAB-5 (and UNC-3) extends to multiple terminal identity genes in cholinergic MNs (WT panel in **Fig. 4.5D**). The observed effects were 100% penetrant and consistent with the previously described region-specific expression pattern of *lin-39* and *mab-5* in VNC MNs (**Fig. 4.5A**)(Kratsios et al., 2017). Of note, while MAB-5 collaborates with LIN-39 to activate cholinergic MN identity genes (**Fig. 4.5B**), it does not affect the ectopic expression of VD or VC genes observed in *unc-3* mutants (**Figure 4.16**).

Since UNC-3 controls directly, via its cognate binding site, cholinergic MN terminal identity genes (Kratsios et al., 2012), we then asked whether this is the case for LIN-39. We analyzed available ChIP-Seq data for LIN-39 from the modENCODE project (Boyle et al., 2014) and found evidence for direct LIN-39 binding in the *cis*-regulatory of all 6 cholinergic MN terminal

identity genes (*unc-129, del-1, acr-2, dbl-1, unc-77, slo-2*) (**Fig. 4.5C, Figure 4.16**). Moreover, we identified multiple consensus LIN-39 binding sites (previously defined as GATTGATG) (Boyle et al., 2014) located within the LIN-39 ChIP-Seq peaks in the *cis*-regulatory region of the aforementioned genes (**Table 4.3**).

This analysis strongly suggests that LIN-39, similar to UNC-3, regulates directly the expression of multiple terminal identity genes in cholinergic MNs (**Fig. 4.5D**). However, in the absence of UNC-3, the function of LIN-39 in cholinergic MNs is modified. Instead of activating cholinergic MN identity genes, LIN-39 activates VD or VC terminal identity genes in *unc-3*-depleted MNs (**Fig. 4.4**). Taken together, our data suggest that UNC-3 antagonizes the ability of LIN-39 to activate alternative identity genes, thereby preventing a switch in the transcriptional targets of LIN-39 (model schematized in **Fig. 4.5D**). If this hypothesis is correct, one would expect decreased LIN-39 binding in the *cis*-regulatory region of cholinergic MN terminal identity genes in *unc-3* mutants. By performing ChIP-Seq for LIN-39 in *unc-3* mutant animals, we indeed observed decreased LIN-39 binding in the *cis*-regulatory region of the aforementioned genes (**Fig. 4.5C, Figure 4.16**). As a positive control, LIN-39 binding in *unc-3* mutant animals was observed in other loci, including the *lin-39* locus itself (**Figure 4.16**), consistent with the known role of LIN-39 in regulating its own expression (Niu et al., 2011). Similar results were obtained by ChIP-qPCR for LIN-39 targets in *unc-3* mutants animals (**Figure 4.16**). We conclude that, in the absence of UNC-3, LIN-39 is released from cholinergic MN terminal identity gene promoters, presumably leading to increased availability of LIN-39 and thereby activation of alternative identity genes.

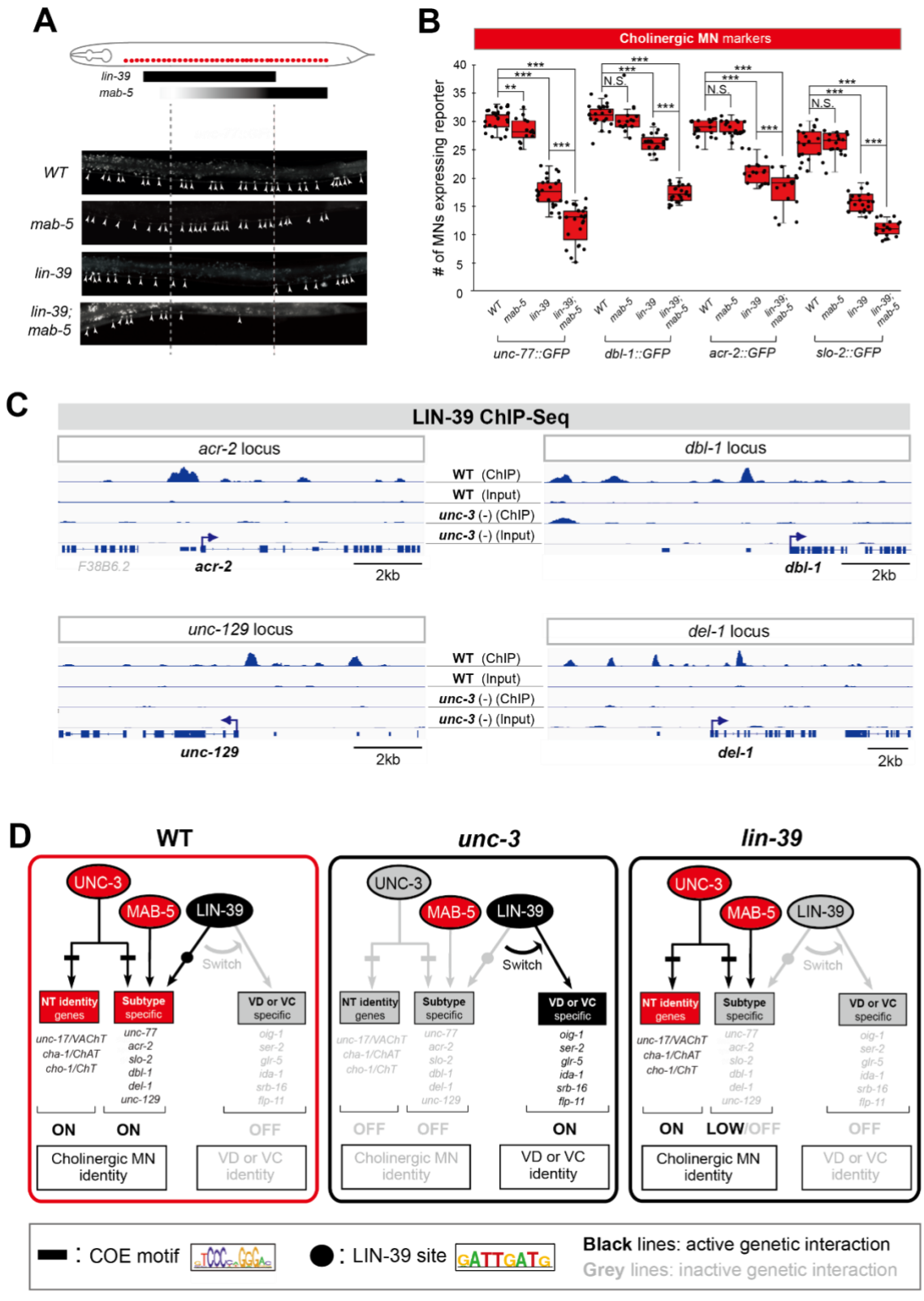

Figure 4.5 UNC-3 prevents a switch in the transcriptional targets of LIN-39 in cholinergic motor neurons. **A**: Schematic summarizing the expression pattern of *lin-39* and *mab-5* in VNC

cholinergic MNs. Below, representative images are shown of *unc-77::gfp* in *WT, lin-39 (n1760), mab-5 (1239)* and *lin-39 (n1760); mab-5 (1239)* animals at L4 stage. Arrowheads point to MN cell bodies with *gfp* marker expression. Green fluorescence signal is shown in white for better contrast. Dotted black box indicates imaged area. **B**: Quantification of cholinergic MN terminal identity markers (*unc-77, dbl-1, acr-2, slo-2*) in *WT, lin-39 (n1760), mab-5 (1239)* and *lin-39 (n1760); mab-5 (1239)* animals at L4. N > 15. ** $p < 0.01$; *** $p < 0.001$. **C:** ChIP-Seq tracks are shown for LIN-39 on four cholinergic MN terminal identity genes (*acr-2, unc-129, dbl-1, del-1*). The WT data come from the modENCODE project (Boyle et al., 2014), whereas the *unc-3 (-)* data were obtained by performing ChIP-Seq for LIN-39 on *unc-3 (n3435); lin-39 (kas9 [lin-39::mNG::3xFLAG::AID]* animals. **D**: Schematic showing the transcriptional switch in LIN-39 targets. In WT animals, UNC-3, MAB-5 and LIN-39 co-activate subtype-specific genes in cholinergic MNs (e.g., *unc-77, dbl-1, unc-129, acr-2*). In *unc-3* mutants, LIN-39 is no longer able to activate these genes, and instead switches to VD- or VC-specific terminal identity genes. Black font: gene expressed. Grey font: gene not expressed. Grey arrows indicate inactive genetic interactions. COE motif taken from (Kratsios et al., 2012) and LIN-39 site taken from (Weirauch et al., 2014) are represented with black rectangles and dots, respectively.

4.3.7 LIN-39 is continuously required to control expression of terminal identity genes in cholinergic MNs

The neuronal function of Hox proteins at post-developmental stages is largely unknown (Hutlet et al., 2016). The continuous expression of mid-body Hox *lin-39* in both developing and adult cholinergic MNs led us to investigate whether *lin-39* is required to maintain expression of terminal identity genes in these neurons. To test this idea, we employed clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9-based genome engineering and generated an auxin-inducible *lin-39* allele (*lin-39::mNG::3xFLAG::AID*) that also serves as an endogenous *lin-39* reporter (*mNG*). Animals carrying *lin-39::mNG::3xFLAG::*AID display no developmental phenotypes and show nuclear mNG expression in MNs located at the mid-body region of the VNC during development and adult stages (**Fig. 4.6A**), corroborating previous observations with a LIN-39 antibody (Maloof and Kenyon, 1998). Upon crossing the *lin-39::mNG::3xFLAG::AID* animals with the *eft-3::TIR1* line, we observed hypomorphic effects in the expression of two cholinergic MN identity genes (*acr-2*, *unc-77)* (**Figure 4.15C**). Although LIN-39 protein is present in the nuclei of cholinergic MNs of *lin-39::mNG::3xFLAG::AID; eft-3::TIR1* animals (**Figure 4.6A**), these effects are likely due to a mild reduction in LIN-39 levels triggered by TIR1. However, post-embryonic auxin administration on these animals resulted in efficient LIN-39 protein depletion and significantly enhanced these effects (**Fig. 4.6A-C**, **Figure 4.15C**). We therefore conclude that LIN-39 is continuously required to maintain terminal identity features in cholinergic MNs.

Next, we sought to determine whether LIN-39 is continuously required for the ectopic activation of VD and VC terminal features observed in *unc-3* null animals. Indeed, auxin administration at L4 stage on *unc-3(n3435); lin-39::mNG::3xFLAG::AID* animals carrying either a VD (*ser-2*),

VC (*glr-5*), or VD/VC (*flp-11*) marker resulted in a statistically significant suppression of the *unc-3* phenotype when compared to control (treated with ethanol) (**Fig. 4.6D-F**).

To sum up, our findings with the auxin-inducible (**Figure 4.15C**) and null *lin-39* alleles (**Fig. 4.4F, 4.5A-B**) indicate that, in the presence of UNC-3, LIN-39 is required to induce and maintain expression of cholinergic MN terminal identity genes (**Fig. 4.5D**). In the absence of UNC-3 (**Fig. 6**), LIN-39 is also continuously required - from development and possibly throughout life - for ectopic activation of VD and VC terminal identity genes (**Fig. 4.5D**).

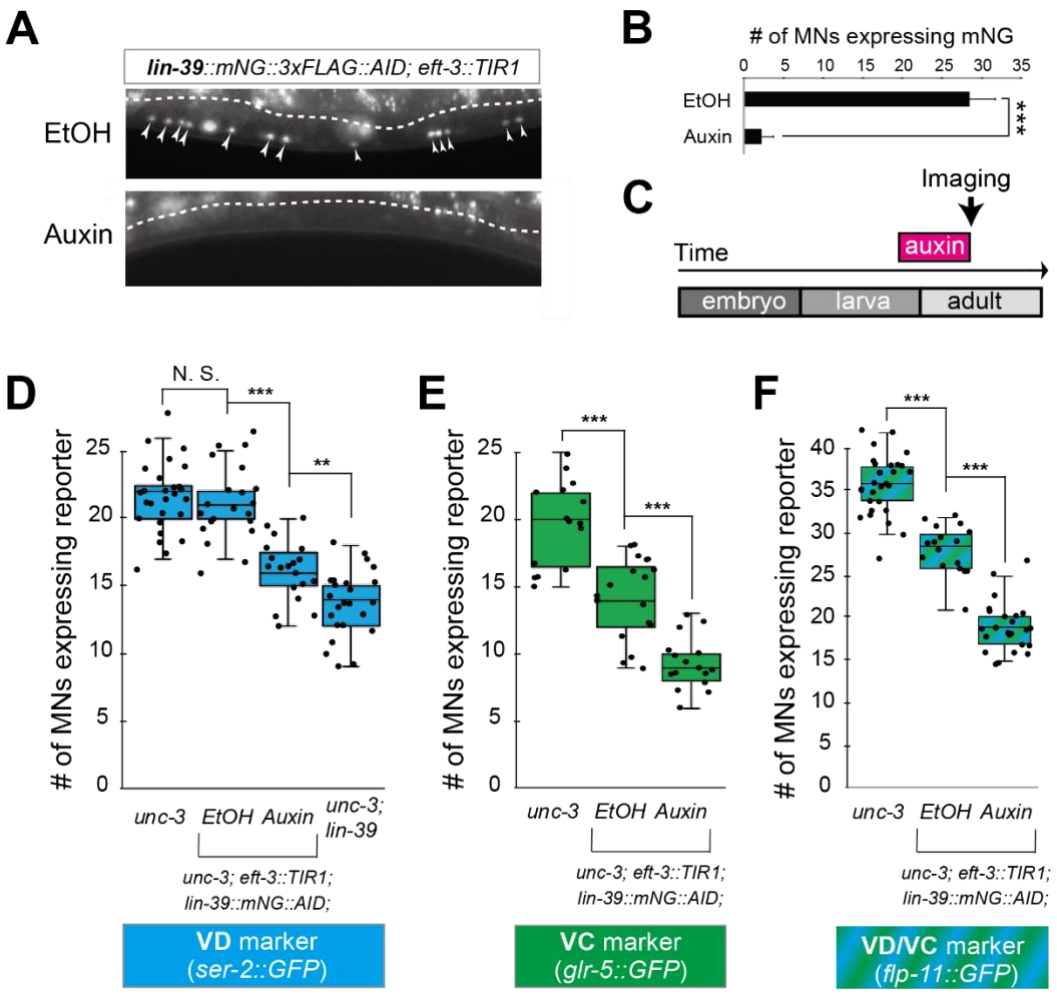

Figure 4.6 LIN-39 is continuously required to control expression of terminal identity genes. **A:** Animals of the *lin-39::mNG::3xFLAG::AID; eft-3::TIR1* genotype were either administered ethanol (EtOH) or auxin at the L3 stage. Twenty four hours later, expression of endogenous *lin-39* reporter (*lin-39::mNG::3xFLAG::AID*) is severely reduced in the nuclei of VNC MNs (arrowheads) at the young adult stage (day 1). mNG green fluorescent signal is shown in white for better contrast. White dotted line indicates the boundary of intestinal tissue (gut), which tends to be autofluorescent in the green channel. **B:** Quantification of number of MNs expressing the *lin-39::mNG::3xFLAG::AID* reporter after EtOH (control) and auxin treatment. N > 14. *** p < 0.001. **C:** Schematic showing time window of auxin administration. **D-F**: Auxin or ethanol (control) were administered at larval stage 4 (L4) on *unc-3 (n3435); lin-39::mNG::3xFLAG::AID; eft-3::TIR1* animals carrying either the VD marker *ser-2::gfp*, the VC marker *glr-5::gfp*, or the VD/VC marker *flp-11::gfp*. Images were taken at the young adult stage (day 1.6 for *ser-2*, day 1.8 for *glr-5* and day 2 for *flp-11*). A significant decrease in the number of MNs expressing the VD marker was evident in the auxin-treated animals compared to EtOH-treated controls. For comparison, quantification of marker expression is also provided in *unc-3 (n3435)* mutants. We note that hypomorphic effects in the ethanol treated group have been previously reported for other AID-tagged TFs in *C. elegans* (Kerk et al., 2017). Such effects appear to be target gene-specific, as they were observed for *glr-5* and *flp-11*, but not *ser-2* (**Fig. 4.6E-F**). N > 15. ** p < 0.01, *** p < 0.001, N. S: not significant.

4.3.8 LIN-39 is an activator of VD and VC terminal identity genes

The observation that *lin-39* is required for ectopic activation of both VD and VC terminal identity genes in *unc-3*-depleted MNs prompted us to examine the role of *lin-39* in VD and VC neurons of wild-type animals. Does LIN-39 control the same VD- and VC-specific terminal identity genes that become ectopically expressed in *unc-3* mutants?

To this end, we leveraged our endogenous *lin-39* reporter (*lin-39::mNG::3xFLAG::AID*) to assess expression in wild-type VD neurons at the mid-body region of the VNC, and found this to be the case (**Fig. 4.7A, Figure 4.15**). Next, we found that LIN-39 is required to induce expression of VD terminal identity genes (*ser-2, oig-1*) (**Fig. 4.7B**). To gain further mechanistic insights, we then asked whether *lin-39* acts together with UNC-30, the known activator of GABAergic MN identity genes (Eastman et al., 1999, Jin et al., 1994). Apart from confirming previous observations of UNC-30 controlling the VD-specific *oig-1* gene (Cinar et al., 2005, Howell et al., 2015), we also found that *ser-2* (**Fig. 4.7B**) and *flp-11* (**Figure 4.15E**) constitute novel UNC-30 targets in VD neurons. To test for synergistic effects, we focused on *ser-2* and *flp-11*, two VD-expressed terminal identity genes mildly affected in *lin-39* or *unc-30* single mutants. We generated *lin-39; unc-30* double mutants and observed stronger effects than either single mutant (**Fig. 4.7B, Figure 4.15E**). Such additive effects indicate that *lin-39* and *unc-30* act in parallel to activate VD terminal identity genes. Importantly, expression of other UNC-30 targets in GABAergic MNs, such as *flp-13* (DD-specific terminal identity marker) (Cinar et al., 2005, Shan et al., 2005, Yu et al., 2017) and genes expressed in both DD and VD neurons (*unc-25/GAD*, *unc-47/VGAT*), is unaffected in *lin-39* mutants (**Fig. 4.7B, Figure 4.15F**). Unlike UNC-30 that broadly controls multiple terminal features (NT identity and VD-specific terminal features) in VD neurons, we conclude that *lin-39* is selectively required for activation of VD-

specific terminal identity genes (**Fig. 4.7F**). To test for a maintenance role in VD neurons, we administered auxin at various post-developmental stages (L3, L4, day 1 adult) on animals carrying the *lin-39::mNG::3xFLAG::AID* allele. We found that LIN-39 is continuously required to maintain expression of the VD terminal identity gene *ser-2* (**Fig. 4.7C**).

The above genetic analysis indicates that LIN-39 and UNC-30/PITX activate expression of VD-specific genes (left panel in **Fig. 4.7F**). Similarly, LIN-39 and UNC-3 directly co-activate terminal identity genes in cholinergic MNs (left panel in **Fig. 4.5D**). Since the absence of UNC-3 leads to ectopic activation of VD-specific genes (**Fig. 4.5D**), we next considered the converse possibility: Does the absence of UNC-30/PITX lead to ectopic activation of cholinergic MN terminal identity genes in GABAergic VD neurons? However, this appears not to be the case as expression of 4 cholinergic MN markers *(acr-2, slo-1, unc-129, del-1),* normally co-activated by UNC-3 and LIN-39 (**Fig. 4.5A-B**), is unaffected in *unc-30* mutants (**Fig. 4.7D-F**).

Similar to its role in sex-shared VD neurons, does *lin-39* control expression of terminal identity genes in sex-specific VC neurons? We used the auxin-inducible *lin-39::mNG::3xFLAG::AID* allele to address this question because, unlike all other nerve cord MNs, the VC neurons do not survive in *lin-39 (n1760)* null animals (Potts et al., 2009). We applied auxin at a late larval stage (L3-L4) to knock-down LIN-39 and observed that VC neurons do not die, providing an opportunity to test for putative effects on VC terminal identity gene expression. Indeed, we found a statistically significant reduction in the number of VC neurons expressing *srb-16* (compare auxin and ethanol in **Figure 4.15D**).

Taken together, *lin-39* is required for expression of VD- and VC-specific terminal identity genes. In VD neurons, LIN-39 acts together with UNC-30/PITX to activate expression of VD-specific genes (left panel in **Fig. 4.7F**). Collectively, these findings on VD and VC neurons together with

observations on cholinergic MNs (**Fig. 4.5D**) show that, in different MN subtypes, the mid-body Hox gene *lin-39* controls expression of distinct terminal identity genes, likely due to collaboration with distinct TFs (i.e., UNC-3 and MAB-5 in cholinergic MNs versus UNC-30 in VD neurons [compare **Fig. 4.5D** and **Fig. 4.7F**]).

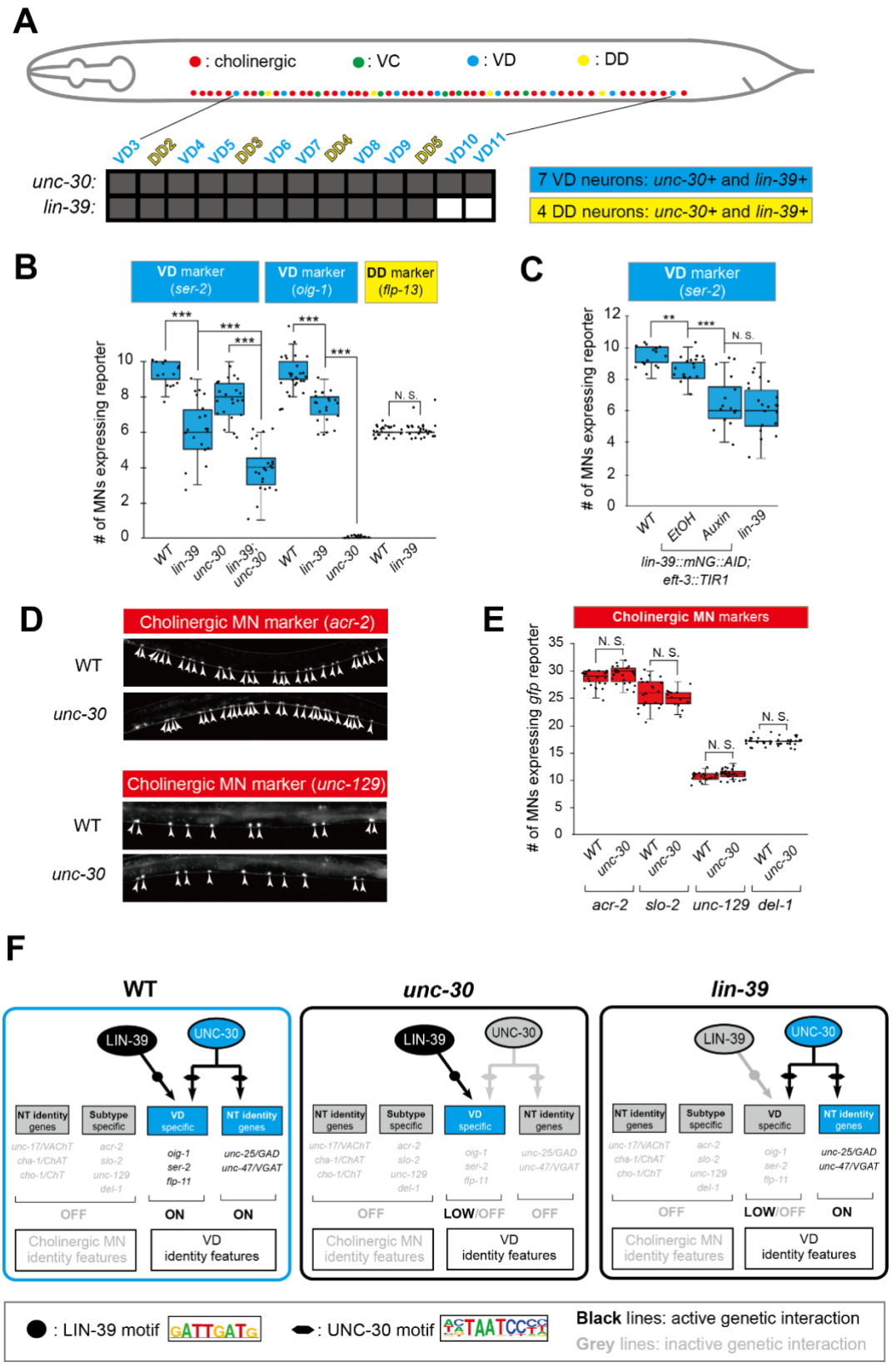

Figure 4.7 LIN-39 is an activator of VD terminal identity genes. **A:** Schematic summarizing

unc-30 and *lin-39* expression in VD and DD neurons populating the VNC. In addition, 4 VD and 2 DD neurons are located in ganglia flanking the VNC (not shown because they were excluded from our analysis). Raw data on *lin-39* expression described in **Figure 4.15 B:** Quantification of two VD (*ser-2::gfp, oig-1::gfp*) and one DD (*flp-13::gfp*) markers in WT and *lin-39 (n1760)* animals at L4. Both VD markers were also tested in *unc-30 (e191)* mutants. Double *lin-39 (n1760); unc-30 (e191)* mutants showed a more severe reduction in expression of the VD marker *ser-2::gfp* compared to each single mutant. N > 15. *** $p < 0.001$. N. S: not significant. **C:** Auxin or ethanol (control) were administered at larval stage 3 (L3) on *lin-39::mNG::3xFLAG::AID; eft-3::TIR1* animals carrying the VD marker *ser-2::gfp*. Images were taken at the young adult stage (day 1.5). A significant decrease in the number of MNs expressing the VD marker was evident in the auxin-treated animals compared to EtOH-treated controls. Similar results were obtained when auxin administration occurred at L4 or day 1 adult animals. For comparison, quantification of marker expression is also provided in WT and *lin-39 (n1760)* animals. N > 15. ** $p < 0.01$, *** $p < 0.001$. N. S: not significant. **D**: Several terminal identity markers of cholinergic neurons (*acr-2, slo-2, unc-129, del-1*) are not ectopically expressed in *unc-30*-depleted GABAergic MNs. A strong loss-of-function allele *e191* for *unc-30* was used (Brenner, 1974, Eastman et al., 1999). Arrowheads point to MN cell bodies with *gfp* marker expression. Green fluorescence signal is shown in white for better contrast. **E**: Quantification of data presented in panel D. N. S: not significant. **F**: Schematic summarizing the function of LIN-39 and UNC-30 in GABAergic VD neurons. LIN-39 site is taken from (Weirauch et al., 2014). UNC-30 site is taken from (Yu et al., 2017).

4.3.9 LIN-39 acts through distinct *cis*-regulatory elements to control *oig-1* expression in VD and VD-like motor neurons

Does LIN-39 act directly or indirectly to activate VD and VC terminal identity genes? Analysis of available ChIP-Seq data (modENCODE project) indicates direct regulation of these genes by LIN-39 (**Fig. 4.8A, Figure 4.17**). However, the low resolution of ChIP-Seq data does not allow the identification of the exact DNA sequence recognized by LIN-39. Therefore, we interrogated the *cis*-regulatory region of two VD terminal identity genes *oig-1* and *ser-2* for the presence of the consensus LIN-39 binding site GATTGATG (Boyle et al., 2014) and found several copies located within the boundaries of the LIN-39 ChIP-Seq peaks in *oig-1* and *ser-2* (**Table 4.3**). To test the functionality of these putative LIN-39 binding sites, we honed in on *oig-1* and performed a systematic *cis*-regulatory analysis in the context of transgenic reporter animals. A previous study identified a minimal 125bp *cis*-regulatory element (contained within the LIN-39 peak boundaries) upstream of *oig-1* as sufficient to drive reporter gene expression in VD neurons (Howell et al., 2015) (**Fig. 4.8A**). We independently confirmed this observation, and further found that the 125bp element contains a single LIN-39 site. Mutation of this site in the context of transgenic *oig-1* reporter animals (*oig-1* $^{125bp\ LIN\text{-}39\ site\ MUT}$*::tagRFP*) leads to a significant reduction of tagRFP expression in VD neurons (**Fig. 4.8A-B**), phenocoping the effect observed in *lin-39 (n1760)* null mutants (**Fig. 4.7B**). We conclude that, in wild-type animals, LIN-39 acts directly, by recognizing its cognate site, to activate expression of the VD-specific gene *oig-1*. Interestingly, a functional binding site for UNC-30/PITX also exists in this 125bp element (Howell et al., 2015, Yu et al., 2017), and is spaced 11 base pairs apart from the LIN-39 site (**Fig. 4.7D**), indicating that LIN-39 and UNC-30 control *oig-1* by recognizing distinct and in close proximity *cis*-regulatory motifs. Moreover, available UNC-30 ChIP-Seq data further

support this possibility as UNC-30 and LIN-39 ChIP-Seq peaks largely overlap at this 125bp element (**Fig. 4.8A**). Lastly, deletion of the region where LIN-39 and UNC-30 peaks overlap in the context of a 2.6kb *oig-1* reporter (*oig-1* $^{2.6kb\ LIN\text{-}39\ peak\ \#3\ DEL}$) abolish reporter expression in VD neurons (**Fig. 4.8A-B**).

We next asked whether LIN-39 acts through the same or distinct *cis*-regulatory elements to drive *oig-1* expression in VD versus VD-like neurons of *unc-3* mutants. While *oig-1* reporters in the context of a large (32.5 kb) genomic clone (fosmid) or a 2.6 kb intergenic region do show expression in both VD and VD-like neurons of *unc-3* mutants, reporter animals carrying 1 kb of *cis*-regulatory sequence (that contains the 125bp element) immediately upstream of ATG showed expression only in VD neurons (**Fig. 4.8A**). Conversely, a distal 1.6 kb element displayed expression in VD-like cells, but no expression in VD neurons of either WT or *unc-3* animals, suggesting the VD and VD-like elements are physically separated on the genome. Within the 1.6kb element, there is a LIN-39 binding peak (peak #4) based on available ChIP-Seq data on WT animals. Deletion of this peak in the context of a 2.6kb *oig-1* reporter (*oig-1* $^{2.6kb\ LIN\text{-}39\ peak\ \#4\ DEL}$) resulted in loss of expression in VD-like cells of *unc-3* mutants, whereas reporter expression was maintained in VD neurons (**Fig. 4.8A-B**). This analysis strongly suggests that LIN-39 acts through distinct *cis*-regulatory elements to activate *oig-1* expression in VD versus VD-like cells.

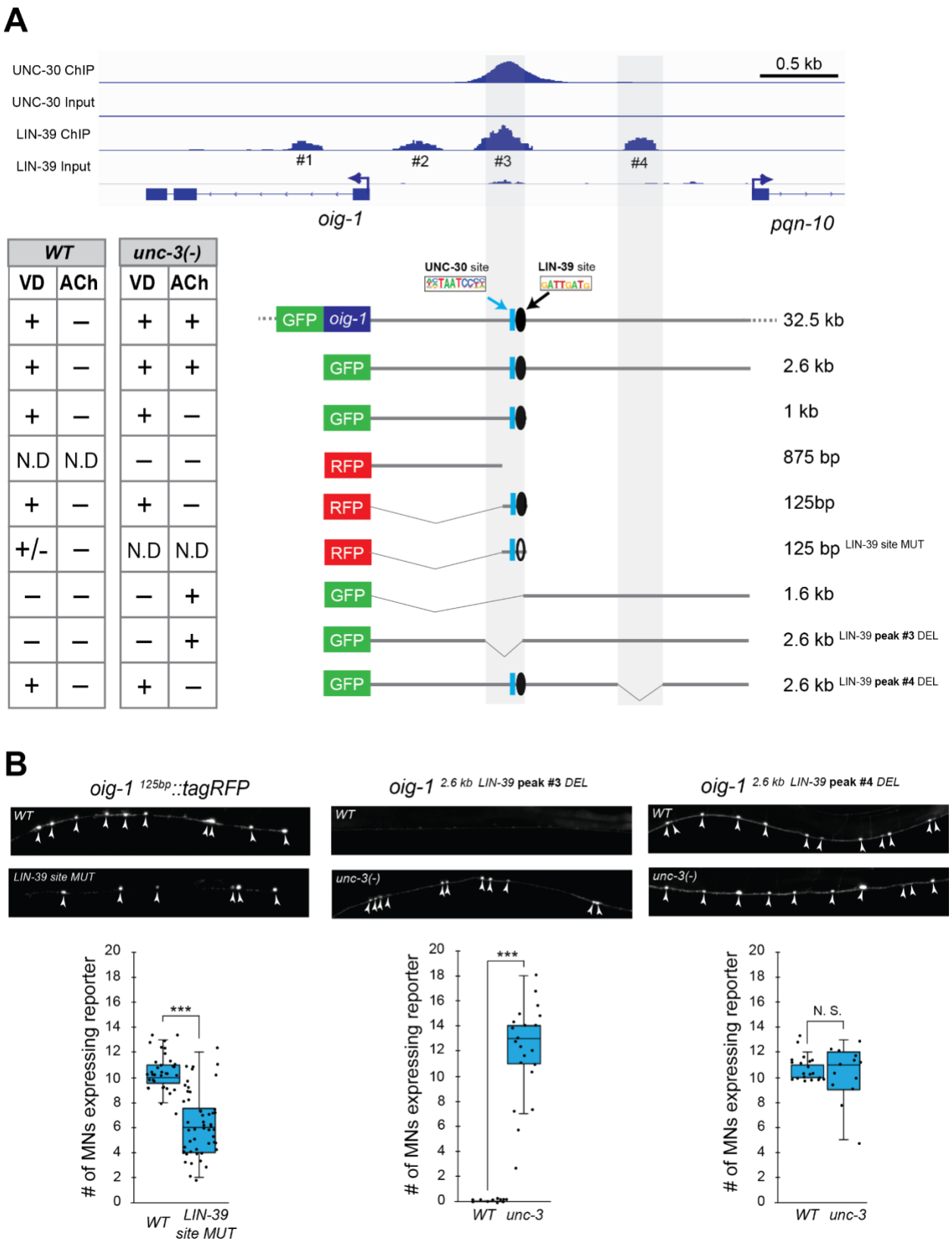

Figure 4.8 LIN-39 acts through distinct *cis*-regulatory elements to activate *oig-1* expression in

VD and VD-like neurons. **A**: ChIP-Seq tracks are shown for UNC-30 (the top two) and LIN-39 (the bottom two) on VD gene oig-1 locus. The UNC-30 data were obtained from (Yu et al., 2017). The LIN-39 data come from the modENCODE project (Boyle et al., 2014). Four LIN-39 peaks are annotated with peak#3 largely overlapping with UNC-30 peak. The results of *cis*-regulatory analysis in both WT and *unc-3* mutants are shown in the lower panel (aligned to the ChIP-seq tracks). Expression patterns of at least two transgenic lines were analyzed for each construct. "+" indicates consistent and bright expression in ventral nerve cord (VNC) MNs (either VD or cholinergic). "+/−" indicates consistent and bright expression in noticeable less number of VNC MNs. "−" indicates no or extremely dim expression in VNC MNs. "N.D.": Not determined. In the schematic of the transgenes, a known UNC-30 site is shown as a blue box and a bioinformatically predicted LIN-39 site is represented as a black circle (filled circle indicates the presence of the site while unfilled one indicates deletion of the site). MUT indicates deletion of the LIN-39 site and DEL indicates deletion of the respective LIN-39 peak region. **B**: Images (top part) and quantifications (bottom part) of selected constructs in the *cis*-regulatory analysis shown in **A**. Animals carrying the *oig-1*^{125bp}*::tagRFP* (left panel) with the LIN-39 site deleted show reduced tagRFP reporter expression in VD neurons; animals carrying the *oig-1*$^{2.6kb\ LIN\text{-}39\ peak\ \#3\ DEL}$ (middle panel) ectopically express the reporter in cholinergic MNs of *unc-3* mutants, but not in WT animals; animals carrying the *oig-1*$^{2.6kb\ LIN\text{-}39\ peak\ \#4\ DEL}$ (right panel) do not show ectopic reporter expression in *unc-3*-depleted MNs, but do show VD expression in both wild-type and *unc-3* mutants. N > 12. *** $p < 0.001$. N. S: not significant.

4.3.10 The LIN-39-mediated transcriptional switch depends on UNC-3 and LIN-39 levels

How does the absence of UNC-3 lead to ectopic and *lin-39*-dependent activation of VD terminal identity genes in cholinergic MNs (**Fig. 4.9D**)? In principle, UNC-3 and LIN-39 could physically interact in order to co-activate expression of cholinergic MN terminal identity genes. In the absence of *unc-3*, this interaction would be disrupted and LIN-39 becomes available, in cholinergic MNs, to assume its VD function, i.e., to activate VD-specific terminal identity genes (**Fig. 4.9D**). Although our co-immunoprecipitation (co-IP) experiments on UNC-3 and LIN-39 in a heterologous system (HEK cells) did not provide evidence for physical interaction (**Figure 4.18**), the heterologous context of this experiment still leaves open the possibility that, in cholinergic MNs *in vivo*, UNC-3 directly (or indirectly) recruits LIN-39 on terminal identity gene promoters. This scenario is supported by the observed decrease of LIN-39 binding on cholinergic MN gene loci in *unc-3* mutants (**Fig. 4.5C**). Lastly, the gene dosage experiments presented below firmly suggest there is a close stoichiometric relationship between UNC-3 and LIN-39, reminiscent of LIM homeodomain TF stoichiometries described in vertebrate MNs (Song et al., 2009).

Because the decrease of LIN-39 binding is accompanied by ectopic activation of VD terminal identity genes in *unc-3*-depleted MNs, we hypothesized that LIN-39 is the rate-limiting factor present in limited amount in cholinergic MNs. That is, in the presence of UNC-3, LIN-39 activates cholinergic MN identity genes, but in its absence LIN-39 becomes available to activate alternative identity (e.g., VD) genes. Quantification of the endogenous expression levels of both proteins indeed showed lower levels of LIN-39 expression compared to UNC-3 (**Fig. 4.9A**). Supporting the aforementioned hypothesis, we found a gene dosage relationship between *unc-3* and *lin-39*. Loss of one *unc-3* copy (*unc-3 (n3435)/+*) caused slight ectopic expression of VD

genes (*ser-2* in **Fig. 4.9B** and *flp-11* in **Figure 4.19A**), but that ectopic expression is decreased by loss of one *lin-39* copy (*unc-3 (n3435)/+; lin-39 (n1760)/+*) (**Fig. 4.9B, Figure 4.19A**). Accordingly, loss of one *lin-39* copy in *unc-3* null animals *(unc-3 (n3435); lin-39 (n1760)/+)* also reduced, but did not eliminate, ectopic expression of VD genes (**Fig. 4.9A, Figure 4.19A**). Moreover, knock-down of *unc-3* with RNAi specifically in cholinergic MNs also led to ectopic expression of VD genes (**Fig. 4.9C, Figure 4.19B-C**), whereas ectopic expression of UNC-3 in VD neurons resulted in repression of VD gene expression, presumably by recruiting LIN-39 away from VD promoters (**Fig. 4.9C, Figure 4.19B-C**). Lastly, we asked whether LIN-39 is sufficient to induce expression of VD terminal identity genes in cholinergic MNs. Indeed, we found this to be the case (**Fig. 4.9C, Figure 4.19D-E**). In conclusion, we propose that the LIN-39-mediated transcriptional switch observed in *unc-3* mutants critically depends on UNC-3 and LIN-39 levels, with the latter being the rate-limiting factor (**Fig. 4.9D**).

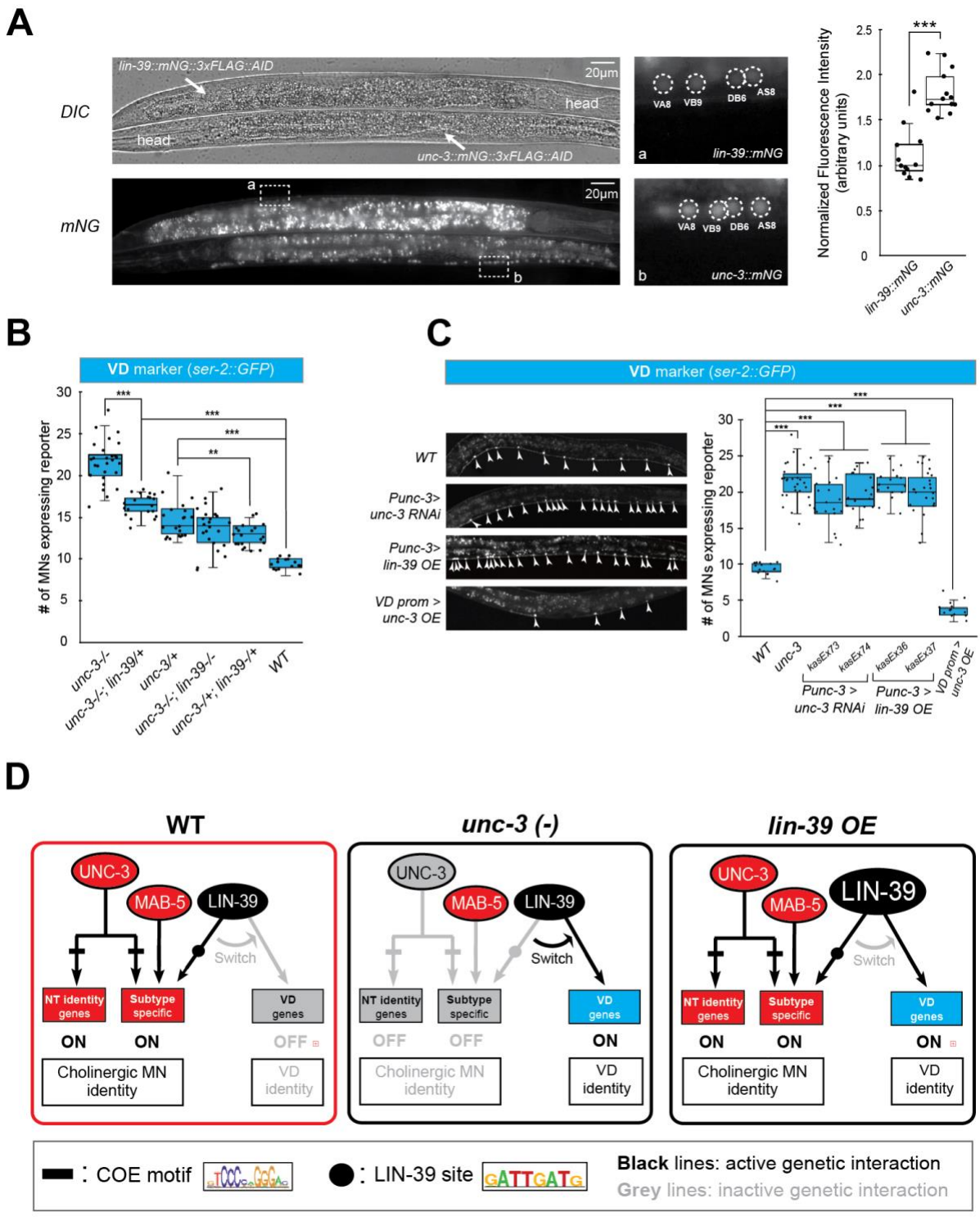

Figure 4.9 Gene dosage experiments suggest that LIN-39 is the rate-limiting factor. **A:** The endogenous expression levels of UNC-3 are higher than LIN-39. The DIC and mNG channels of two worms next to each other on the same slide with the genotype of *lin-39::mNG::3xFLAG::AID* (top, anterior right, ventral up) and *unc-3::mNG::3xFLAG::AID* (bottom, anterior left, ventral down) respectively are shown on the left. VNC regions indicated

by dashed frame (a, b) are zoomed in at the middle panel with dashed circles around MN nuclei. The identities of these MNs are shown (e.g., VA8, VB9). The same cholinergic MNs have stronger expression levels of endogenous *unc-3::mNG* than *lin-39::mNG*. Quantification of the fluorescence intensities is shown on the right panel. For details on the quantification, see Materials and Methods. N = 12. *** $p < 0.001$. **B:** Quantification of the VD marker (*ser-2::gfp*) in *unc-3 (n3435), unc-3 (n3435); lin-39 (n1760)/+, unc-3 (n3435)/+, unc-3 (n3435); lin-39 (n1760), unc-3 (n3435)/+; lin-39 (n1760)/+*, and WT animals at L4. N > 15. ** $p < 0.01$, *** $p < 0.001$. **C:** Representative images of the VD marker (*ser-2::gfp*) expression on the left in L4 stage transgenic animals that either down-regulate *unc-3* in cholinergic MNs (*Punc-3 > unc-3 RNAi*), over-express *lin-39* in cholinergic MNs (*Punc-3 > lin-39 OE*), or over-express *unc-3* in VD neurons (VD prom [*unc-47* prom] > *unc-3 OE*). Arrowheads point to MN cell bodies with *gfp* marker expression. Green fluorescence signal is shown in white for better contrast. Quantification is provided on the right. Two independent transgenic lines were used for *Punc-3 > unc-3 RNAi* and *Punc-3 > lin-39 OE.* N > 13. *** $p < 0.001$. **D:** Schematic summarizing the gene dosage experiments.

4.3.11 Ectopic expression of VD terminal identity genes in cholinergic motor neurons is associated with locomotion defects

The dual role of UNC-3 revealed by our molecular analysis (**Fig. 4.2**) led us to posit that the severe locomotion defects observed in *unc-3* animals may represent a composite phenotype (Brenner, 1974, Yemini et al., 2013). In other words, these defects are not only due to loss of expression of cholinergic MN terminal identity determinants (e.g., *unc-17*/VAChT, *cha-1*/ChAT, *del-1*/Degenerin-like sodium channel, *acr-2*/acetylcholine receptor [ortholog of CHRNE]), but also due to the ectopic expression of VD and VC terminal features (e. g., *ser-2*/serotonin receptor [ortholog of HTR1D], *flp-11*/FRMR-like neuropeptide, *glr-5*/Glutamate receptor [ortholog of GRID], *srb-16*/GPCR) in *unc-3*-depleted MNs. To genetically separate these distinct molecular events, we generated *unc-3 (n3435); lin-39 (n1760)* double mutants, which do display loss of cholinergic MN terminal identity genes, but the ectopic expression of VD and VC terminal features is suppressed (**Fig. 4.4F**). We predicted that if ectopic expression of VD and VC genes contributes to locomotion defects, then *unc-3 (n3435)* mutants would display more severe locomotion defects than *unc-3 (n3435); lin-39 (n1760)* double mutants. To test this, we performed high-resolution behavioral analysis of freely moving adult (day 1) *C. elegans* animals using automated multi-worm tracking technology (Javer et al., 2018b, Yemini et al., 2013). This analysis can quantitate multiple features related to *C. elegans* locomotion (*e.g.,* speed, crawling amplitude, curvature, pause, forward and backward locomotion) and, most importantly, each feature can be localized to a specific part of the nematode's body (e. g., head, mid-body, tail). Since *unc-3* and *lin-39* expression uniquely overlaps in mid-body nerve cord MNs that innervate mid-body muscles, we hypothesized that loss of *unc-3* and/or *lin-39* genes would have effects on mid-body posture and motion, and thereby focused our analysis on mid-body curvature features.

Of the 49 mid-body features examined, 29 were significantly different in *unc-3* single mutants when compared to wild-type (N2 strain) animals (see **Table 4.4** for all 49 features). Intriguingly, 12 of these 29 features (41.37%) were significantly suppressed in *unc-3; lin-39* double mutants (**Fig. 4.10A, Figure 4.20A, Table 4.4**), suggesting that suppression of these behavioral defects could be attributed to suppression of the ectopically expressed VD and VC terminal identity genes in these double mutants. We found no evidence for suppression of the remaining 17 features in *unc-3; lin-39* double mutants, likely due to the fact that UNC-3 also controls other terminal identity genes, such as NT pathway genes (**Fig. 4.9D**), independently of LIN-39.

Next, we asked whether ectopic expression of VD terminal identity genes in otherwise wild-type animals can lead to locomotion defects. To test this, we took advantage of our transgenic animals that selectively over-express LIN-39 in cholinergic MNs (*Pcho-1 > LIN-39*, *Punc-3 > LIN-39*) (**Fig. 4.9C, Figure 4.20D**). First, we confirmed that in these animals, expression of cholinergic MN terminal identity genes is unaffected (**Fig. 4.20D**). Second, we found that LIN-39 overexpression led to ectopic activation of VD, but not VC, terminal identity genes in cholinergic MNs (**Figure 4.20D**), providing an opportunity to specifically assess the consequences of ectopic VD gene expression on animal locomotion. We found that 9 of the 29 (31.03%) mid-body features affected in *unc-3 (n3435)* animals were also altered in animals over-expressing *lin-39* in cholinergic MNs (**Fig. 4.10B, Figure 4.20B-C, Table 4.4**).

In conclusion, our behavioral analysis is in agreement with our molecular findings. At the molecular level, we found that *lin-39*/Hox is necessary for the ectopic expression of VD terminal identity genes in *unc-3* mutants. At the behavioral level, this *lin-39*-dependent, ectopic expression of terminal identity genes is accompanied by locomotion defects.

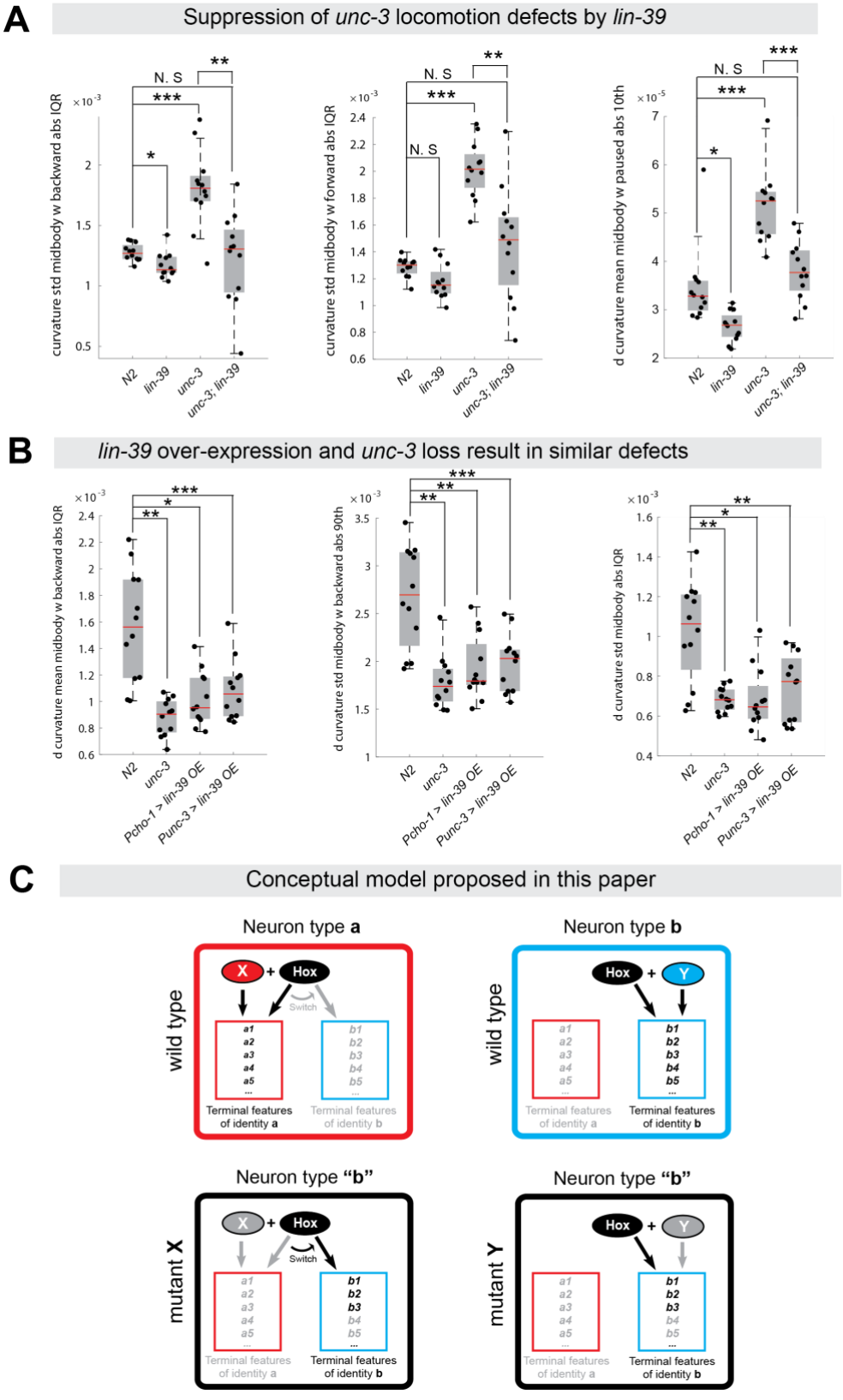

Figure 4.10 Ectopic expression of VD terminal identity genes in cholinergic motor neurons is

associated with locomotion defects. **A**: Examples of three mid-body locomotion features that are significantly affected in *unc-3 (n3435)* animals, but markedly improved in *unc-3 (n3435); lin-39 (n1760)* double mutant animals. Each black dot represents a single adult animal. The unit for the first two graphs is 1/microns. The unit for the graph on the right is 1/(microns*seconds). N = 12. Additional mid-body features affected in *unc-3 (n3435)* animals, but improved in *unc-3 (n3435); lin-39 (n1760)* mutants are provided in **Figure 4.20** and **Table 4.4**. * $p < 0.01$, ** $p < 0.001$, *** $p < 0.0001$. **B**: Examples of three mid-body locomotion features affected in *unc-3 (n3435)* mutants and animals over-expressing *lin-39* in cholinergic MNs. Each black dot represents a single adult animal. The unit for the Y axis is 1/(microns*seconds). N = 12. Additional mid-body features affected in *unc-3 (n3435)* and *lin-39* over-expressing animals are provided in **Figure 4.20** and **Table 4.4**. * $p < 0.01$, ** $p < 0.001$, *** $p < 0.0001$. **C:** Conceptual model summarizing the findings of this paper. Grey font: not expressed gene. Black font: expressed gene. Grey arrow: inactive genetic interaction. Black arrow: active genetic interaction.

4.4 Discussion

During development, individual neuron types must select their unique terminal identity features, such as expression of NT receptors, ion channels and neuropeptides. Continuous expression of these features - from development through adulthood - is essential for safeguarding neuronal terminal identity, and thereby ensuring neuronal function (Deneris and Hobert, 2014, Hobert, 2011, Hobert, 2016). Here, we provide critical insights into the mechanisms underlying selection and maintenance of neuron type-specific terminal identity features by using the well-defined MN populations of the *C. elegans* nerve cord as a model. First, we report that, in cholinergic MNs, the terminal selector-type TF UNC-3 has a dual role; UNC-3 is not only required to promote cholinergic MN identity features (Kratsios et al., 2012), but also to prevent expression of multiple terminal features normally reserved for three other ventral cord neuron types (VD, VC, CA). Second, we provide evidence that cholinergic MNs can secure their terminal identity throughout life by continuously relying on UNC-3's dual function. Third, we propose an unusual mechanism underlying this dual function, as we find UNC-3 necessary to prevent a switch in the transcriptional targets of the mid-body Hox protein LIN-39 (Scr/Dfd/Hox4-5) (**Fig. 4.9D**). Lastly, our findings shed light upon the poorly explored, post-embryonic role of Hox proteins in the nervous system by uncovering that LIN-39 is continuously required to maintain expression of multiple terminal identity genes in MNs.

4.4.1 UNC-3 determines the function of the rate-limiting factor LIN-39/Hox in cholinergic motor neurons

Numerous cases of neuron type-specific TFs with a dual role have been previously described in both vertebrate and invertebrate models systems (Britanova et al., 2008, Cheng et al., 2004, Kala et al., 2009, Lopes et al., 2012, Mears et al., 2001, Morey et al., 2008, Nakatani et al., 2007, Sagasti et al., 1999). Although the underlying mechanisms often remain unclear, recent studies proposed two modes of action. First, such TFs can act directly to activate "desired" terminal identity features and repress (also directly) alternative identity features (Lodato et al., 2014, Wyler et al., 2016). Second, neuron type-specific TFs can act indirectly by controlling intermediary factors. For example, in the mouse spinal cord, a complex of three TFs (Isl1, Lhx3, NLI) specifies MN identity by recognizing specific DNA elements in the *cis*-regulatory region of MN-specific genes and the homeodomain TF Hb9, and activates their expression. Hb9 functions as a transcriptional repressor of alternative (V2a interneuron) identity genes, thereby consolidating MN identity (Lee et al., 2008, Song et al., 2009, Thaler et al., 2002). An analogous mechanism operates in V2a interneurons and involves Chx10, a homeodomain protein that represses alternative neuronal identity programs (Clovis et al., 2016, Lee et al., 2008). Hence, mouse mutants for *Hb9* or *Chx10* result in ectopic expression of alternative identity genes (Arber et al., 1999, Clovis et al., 2016, Thaler et al., 1999). Several TFs (*unc-4*/Uncx, *mab-9*/Tbx20, *unc-55*/COUP, *bnc-1*/BNC) with repressor activity are known to control aspects of cholinergic MN development in *C. elegans* (Kerk et al., 2017, Pflugrad et al., 1997, Pocock et al., 2008, Von Stetina et al., 2007, Winnier et al., 1999). However, their genetic removal did not result in ectopic expression of alternative (VD, VC) identity features in cholinergic MNs (data not shown). Although we cannot exclude the involvement of yet-to-be identified transcriptional

repressors acting downstream of UNC-3, our genetic and biochemical analyses led us to propose the following mechanism underlying UNC-3's dual role.

In cholinergic MNs, *unc-3* and *lin-39* are co-expressed, albeit the latter in lower levels (**Fig. 4.9A**), suggesting that LIN-39/Hox is a rate-limiting factor whose function is determined by UNC-3. In wild-type animals, UNC-3 and LIN-39 occupy *cis*-regulatory elements of cholinergic MN terminal identity genes, resulting in their activation (**Fig. 4.5C**) (Kratsios et al., 2012). In the absence of UNC-3, LIN-39 is released from these elements and becomes available to activate alternative identity genes, such as VD-specific terminal identity genes. Several lines of evidence support this conclusion. First, ChIP-seq data show that LIN-39 binding is decreased in cholinergic MN gene loci (**Fig. 4.5C**). Second, our gene dosage experiments show that either lowering *unc-3* levels or increasing *lin-39* levels in cholinergic MNs results in ectopic activation of VD identity genes (**Fig. 4.9B-C**). Lastly, we performed an extensive *cis*-regulatory analysis of one VD-specific gene (*oig-1*) and identified the element through which LIN-39 acts to induce *oig-1* expression in VD-like cells of *unc-3* mutants (**Fig. 4.8**). Together, these data suggest that the role of UNC-3 in cholinergic MNs is not simply to activate gene expression with LIN-39, but also to "recruit" LIN-39 away from promoters of alternative identity genes, thereby antagonizing its ability to activate those genes. Supporting this scenario, ectopic UNC-3 expression in VD neurons results in decreased expression of VD-specific genes (**Fig. 4.9C**). Given that the mouse ortholog of *unc-3*, Ebf2, is co-expressed with Hox genes in cholinergic MNs of the spinal cord (Catela et al., 2019, Kratsios et al., 2017), the molecular mechanism described here may be conserved across species. Interestingly, a seminal study recently described a conceptually similar mechanism in the mouse retina, where CRX recruits MEF2D to retina-specific enhancers, resulting in selective activation of photoreceptor genes (Andzelm et al., 2015).

4.4.2 Insights into how neurons maintain their terminal identity features throughout life

Is there a need for mechanisms that continuously prevent expression of alternative identity features in a post-mitotic neuron? Or, do such mechanisms become superfluous once neurons have restricted their developmental potential by committing to a specific terminal identity? This fundamental question is poorly explored, in part due to the fact that most neuron type-specific TFs have been studied during embryonic stages. For example, it is not known whether CRX is continuously required to activate retina-specific enhancers and simultaneously prevent expression of alternative identity genes (Andzelm et al., 2015). Our temporally controlled protein depletion experiments uncovered a continuous requirement for the dual role of UNC-3. Post-embryonic depletion of UNC-3 not only results in failure to maintain cholinergic MN terminal features (Kratsios et al., 2012), but is also accompanied by ectopic expression of alternative identity features (e.g., VD, VC). These findings reveal a simple and economical mechanism that can enable individual neuron types to select and maintain their distinct terminal identity features. That is, the same TF is continuously required - from development throughout life - to not only activate neuron type-specific identity genes, but also prevent expression of alternative identity features.

4.4.3 Maintenance of terminal identity features: A new function of Hox proteins in the nervous system

Across model systems, a large body of work on motor neurons and other neuron types has established that, during early development, Hox proteins are required for neuronal diversity, cell survival, axonal path finding and circuit assembly (Baek et al., 2013, Catela et al., 2016, Estacio-

Gomez and Diaz-Benjumea, 2014, Estacio-Gomez et al., 2013, Karlsson et al., 2010, Mendelsohn et al., 2017, Miguel-Aliaga and Thor, 2004, Moris-Sanz et al., 2015, Philippidou and Dasen, 2013). However, the function and downstream targets of Hox proteins during post-embryonic stages are largely unknown. Our contributions towards this knowledge gap are twofold. First, we found that the mid-body Hox protein LIN-39 is continuously required, from development through adulthood, to control expression of MN terminal identity genes, thereby revealing a novel role for Hox proteins in maintaining neuronal identity. Second, we uncovered multiple terminal identity genes as downstream targets of LIN-39 in different MN subtypes (cholinergic MNs: *acr-2, dbl-1, unc-77, slo-2;* VD neurons: *oig-1, ser-2, flp-11;* VC neurons: *srb-16).* Since continuous expression of these genes is essential for MN function, these findings may provide a molecular explanation for the uncoordinated locomotion defects observed in *lin-39* mutants (**Figure 4.20**). Given the maintained expression of Hox genes in the adult nervous system of flies, mice and humans (Baek et al., 2013, Takahashi et al., 2004, Hutlet et al., 2016), our findings may be broadly transferable.

4.4.4 Impact on the concept of terminal selector genes

TFs able to broadly activate many distinct terminal identity features of a specific neuron type (e.g., NT biosynthesis components, NT receptors, ion channels, neuropeptides) have been termed "terminal selectors" (Hobert, 2008). Several dozens of terminal selectors have been described thus far in multiple model systems including worms, flies and mice (Hobert, 2011, Hobert, 2016, Hobert and Kratsios, 2019). However, it is unclear whether terminal selectors are also required to prevent expression of alternative identity features. Our findings suggest this to be the case by revealing a dual role for UNC-3, the terminal selector of cholinergic MN identity in *C. elegans*.

In the future, it will be interesting to see whether other terminal selectors also exert a dual role in order to safeguard neuronal terminal identity. Supporting this possibility, Pet-1, the terminal selector of mouse serotonergic neurons has been recently shown to repress several terminal identity genes (Wyler et al., 2016).

4.4.5 Limitations and lessons learned about the control of neuronal terminal identity

The examination of multiple MN terminal identity markers at single-cell resolution enabled us to make an interesting observation. Although all *unc-3*-depleted nerve cord MNs uniformly lose their cholinergic identity, one subpopulation acquires VD terminal features ("VD-like" neurons) and another subpopulation acquires VC terminal features ("VC-like" neurons). This intriguing observation may be analogous to findings described in the mammalian neocortex, where genetic removal of the TF *Satb2* leads to loss of pyramidal neuron identity (UL1 subtype), and concomitant gain of molecular features specific to two other pyramidal neuron subtypes (DL, UL2) (Britanova et al., 2008). Together, the cases of UNC-3 and Satb2 support the notion that neuron type-specific TFs often suppress features of functionally related neuronal subtypes (Arlotta and Hobert, 2015).

Although our study employs an extensive repertoire of terminal identity markers for distinct MN subtypes, the extent of alternative identity features (e.g., VD, VC) being ectopically expressed in *unc-3*-depleted MNs remains unknown. Future unbiased transcriptional profiling of *unc-3*-depleted MNs could help address this issue. In addition, the strong axonal defects in MNs of *unc-3* mutants preclude any further attempts to assess whether the observed VD-like and VC-like cells, as defined by molecular markers, also acquire morphological features of VD and VC neurons, respectively (Prasad et al., 1998). However, the VD-like neurons of *unc-3* mutants do

not acquire GABAergic identity like wild-type VD neurons (**Figure 4.12A**), arguing against a complete cell fate transformation.

4.4.6 Evolutionary implications of this study

Our findings highlight the employment of economical solutions to evolve novel cell types in the nervous system. The same Hox protein (LIN-39) collaborates with distinct terminal selectors in different MNs, and this collaboration determines the specificity of LIN-39/Hox function. In GABAergic (VD) neurons, LIN-39 works together with UNC-30/PITX to control expression of VD terminal identity genes, whereas in cholinergic MNs LIN-39 synergizes with UNC-3 to control cholinergic MN identity genes (**Fig. 4.7I**). We speculate that the *unc-3* mutant "state" may constitute the "ground state". That is, the "VD-like" neurons, for example, in *unc-3* mutants that express LIN-39 may represent an ancient cell type that was altered to become a new cell type through the recruitment of distinct terminal selectors (conceptual model in **Fig. 4.10C**). Hence, the amount of genetic information required for evolution of new cell types is kept to minimum. The recruitment of UNC-30/PITX enabled "VD-like" cells to fully adopt GABAergic VD neuron terminal identity, as evident by the ability of UNC-30/PITX to control expression of GABA synthesis proteins (Eastman et al., 1999, Jin et al., 1994) (**Fig. 4.7F**). Similarly, recruitment of UNC-3 enabled "VD-like" cells to become cholinergic MNs. In this "new" cholinergic cell type, UNC-3 exerts a dual role: it antagonizes the ability of LIN-39 to activate VD-specific genes, and also synergizes with LIN-39 to co-activate cholinergic MN terminal identity genes (**Fig. 4.9D**). We hope the strategy described here of a terminal selector preventing a Hox transcriptional switch may provide a conceptual framework for future studies on terminal identity and evolution of neuronal cell types.

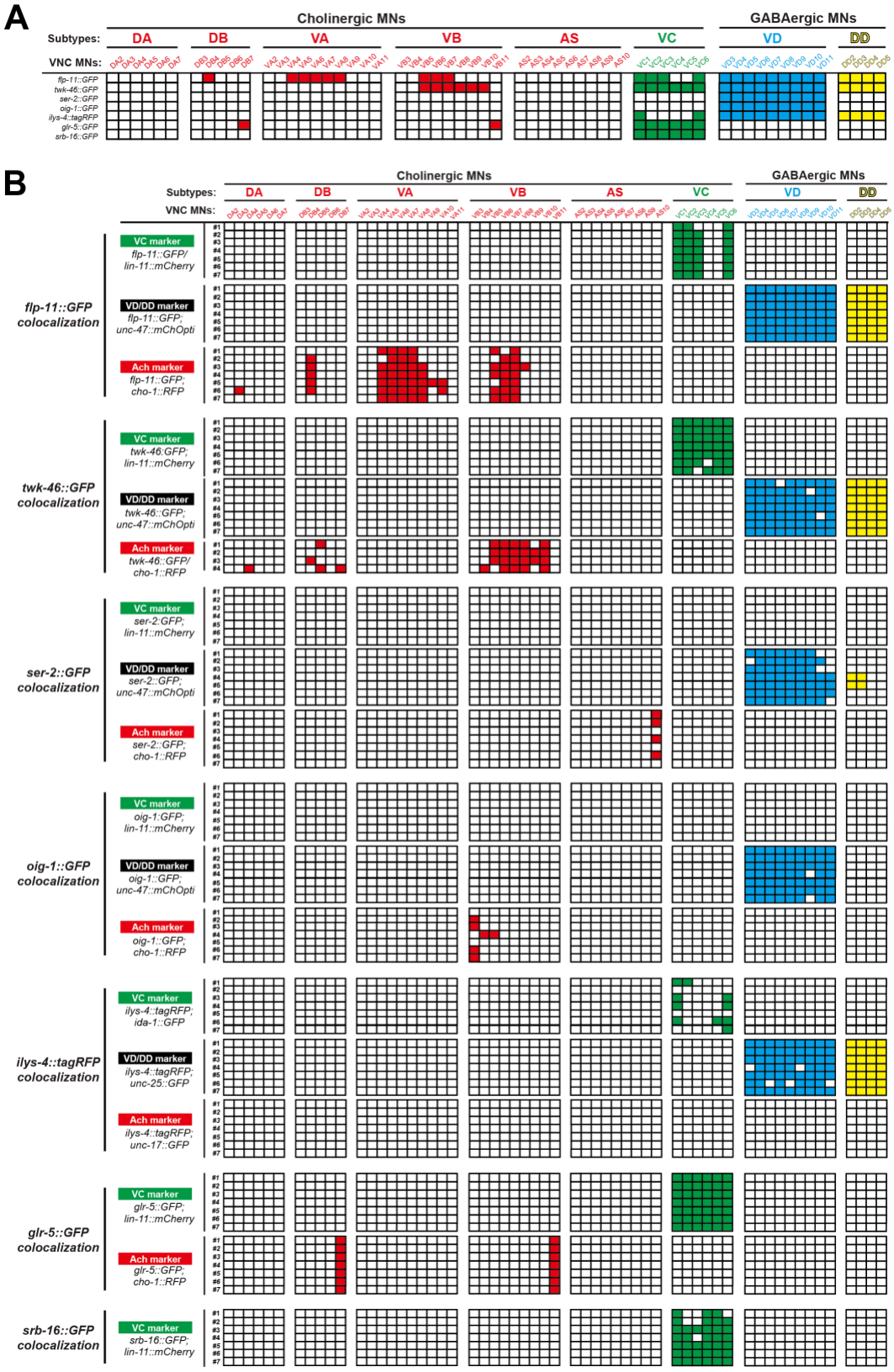

Figure 4.11 Detailed characterization of the expression pattern of VC and VD terminal identity

markers. **A**: An expression map of VC and VD terminal identity genes with single-cell resolution at larval stage 4 (L4). Each column represents an individual motor neuron (MN) in the ventral nerve cord (VNC) of a hermaphrodite worm. MN subtypes are color-coded. This table represents a summary of the co-localization analysis described in panel B. Apart from the 7 genes listed, we also examined 5 additional terminal identity genes (*bra-1, dhc-1, rgs-4, vhp-1, vps-25*) but found no evidence for expression in VD or VC neurons. **B**: Co-localization analysis of VD and VC terminal identity genes. Each column represents an individual MN in the VNC of a hermaphrodite worm. Each row represents a randomly-selected worm co-expressing the respective VD or VC marker and one of the known identity markers: *unc-47::mChOpti or unc-25::gfp* for GABAergic (VD and DD) MN identity, *cho-1::rfp or unc-17::gfp* for cholinergic MN identity and *lin-11::mCherry or ida-1::gfp* for VC MN identity. A color-filled lattice indicates co-expression of the known identity marker and the selected marker in the individual MN.

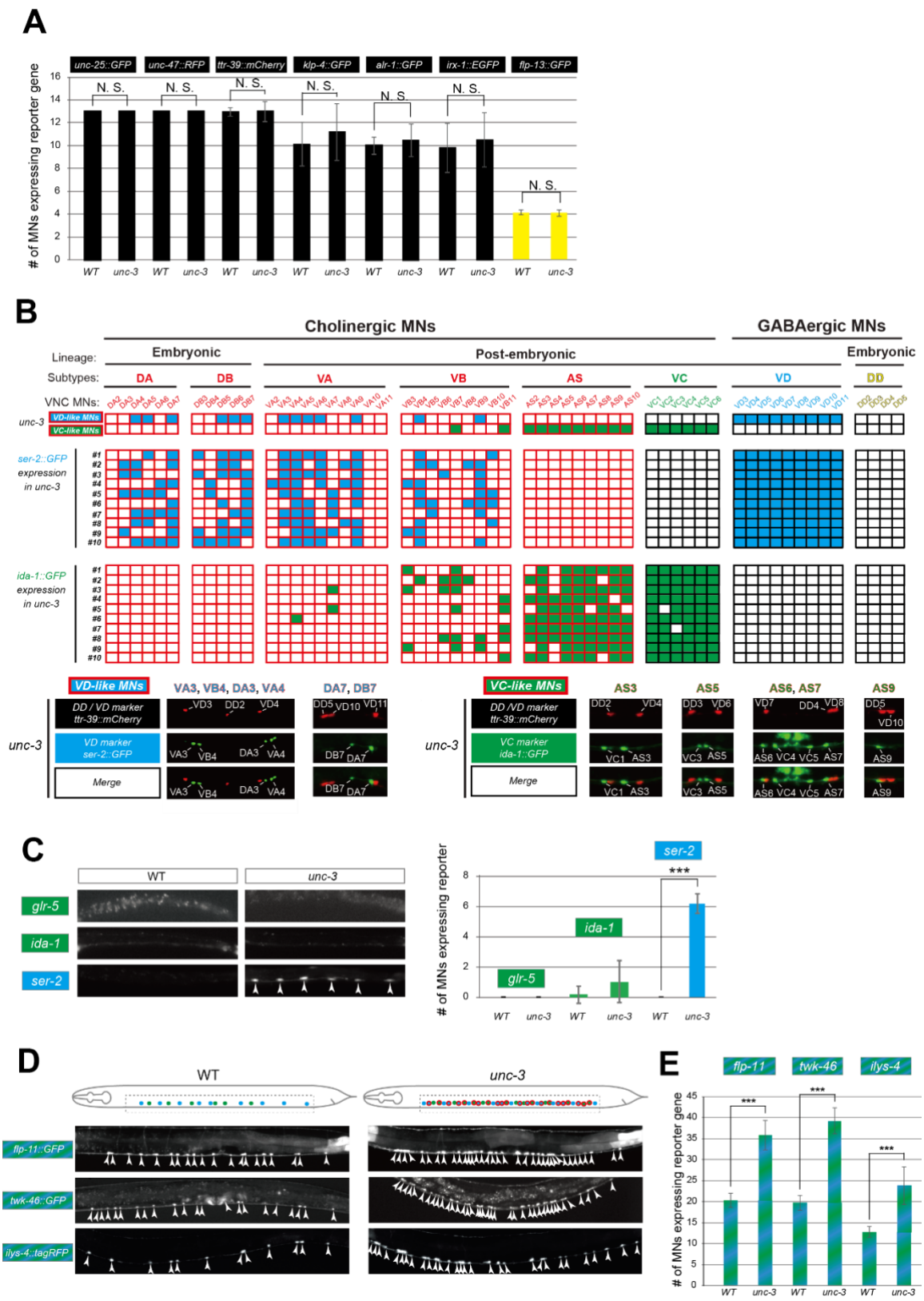

Figure 4.12 UNC-3 selectively prevents expression of VD and VC terminal identity features in distinct cholinergic MNs. **A:** Quantification of six VD/DD (*unc-25::gfp, unc-47::rfp, ttr-*

39::mCherry, klp-4::gfp, alr-1::gfp and irx-1::egfp) and one DD (*flp-13::gfp*) markers in WT and *unc-3 (n3435)* animals at L4. No statistically significant effects were observed in *unc-3* mutants. N > 13. N. S: not significant. **B**: A table showing that VD and VC identity genes are ectopically expressed in distinct cholinergic MNs upon *unc-3* depletion. Top two rows: A summary map of VD and VC gene expression with single-cell resolution in *unc-3* mutants, which is based on the co-localization analysis shown below. Each column represents an individual MN in the VNC of an *unc-3*-depleted hermaphrodite worm. Ten randomly selected worms that co-express VD/DD marker *ttr-39::mCherry* and VD marker *ser-2::gfp* or VC marker *ida-1::gfp* are analyzed. The expression of *ttr-39::mCherry* is not affected by *unc-3* depletion, which serves as a positional landmark for reference. This landmark together with the invariant MN cell body position along the VNC enable the identification of distinct *unc-3*-depleted MNs that lose cholinergic features and concomitantly gain either "VD-like" or "VC-like" features. Below this table, examples of the co-localization analysis are provided. Representative magnified images are shown. The identity of *unc-3*-depleted MNs that acquire VC or VD terminal features is shown. Note that VC-like terminal features are mainly acquired by MNs of the AS cholinergic subtype, whereas VD-like terminal features are acquired by select members of four cholinergic MN subtypes (DA, DB, VA, VB). **C**: The DA and DB cholinergic MNs are born embryonically and therefore present in the VNC at L1. Corroborating our results from panel B, only the VD terminal identity marker (*ser-2*) shows ectopic expression in DA and DB neurons of unc-3 mutants at L1. This is not the case for the VC terminal identity markers (*glr-5, ida-1*). Representative images are shown on the left. Arrowheads point to MN cell bodies with ectopic gfp marker expression. Green fluorescence signal is shown in white for better contrast. Quantification is on the right. N > 22. *** $p < 0.001$. **D:** Terminal identity markers of VD/VC MNs (*flp-11, twk-46 and ilys-4*) are ectopically expressed in *unc-3*-depleted cholinergic MNs. Representative images of larval stage 4 (L4) hermaphrodites are shown. Similar results were obtained in adult animals. Arrowheads point to MN cell bodies with marker expression. Fluorescence signal is shown in white for better contrast. Dotted black box indicates imaged area. **E**: Quantification of data shown in panel D.. N > 18. *** $p < 0.001$.

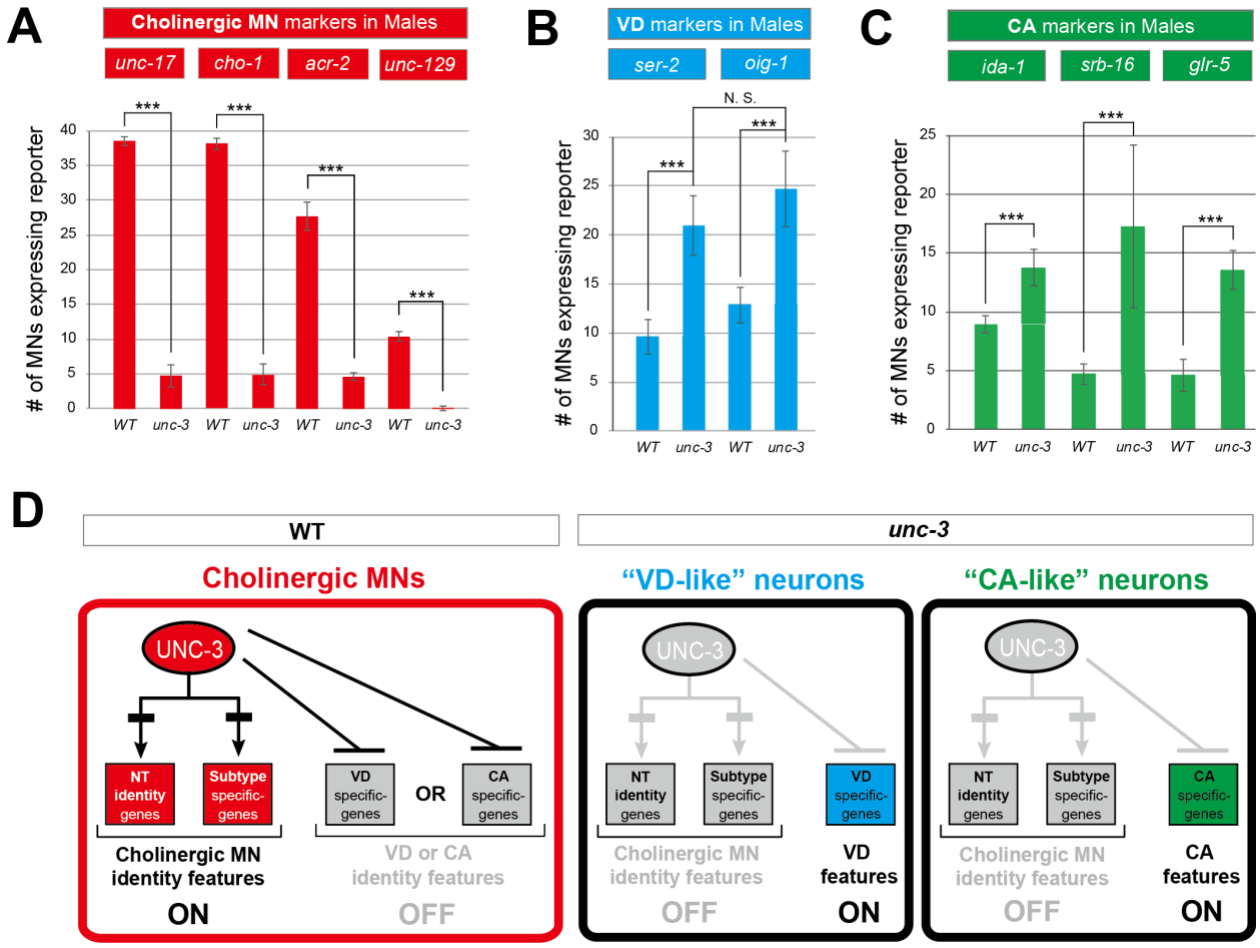

Figure 4.13 The dual role of UNC-3 in cholinergic MNs extends to both *C. elegans* sexes. **A:** Quantification of four sex-shared cholinergic MN markers (*unc-17, cho-1, acr-2, unc-129*) expression in WT and *unc-3 (n3435)* male animals at L4. N > 13. *** p < 0.001. **B**: Quantification of VD marker (*ser-2 and oig-1*) expression in L4 stage WT and *unc-3 (n3435)* male animals. No significant difference was found between the number of MNs ectopically expressing each marker. N > 11. *** p < 0.001. N. S: not significant. **C:** Quantification of the male-specific CA markers (*ida-1, srb-16 and glr-5*) in young adult stage WT or *unc-3 (n3435)* male animals. N > 14. *** p < 0.001. **D:** Schematic summarizing the dual role of UNC-3 in *C. elegans* males.

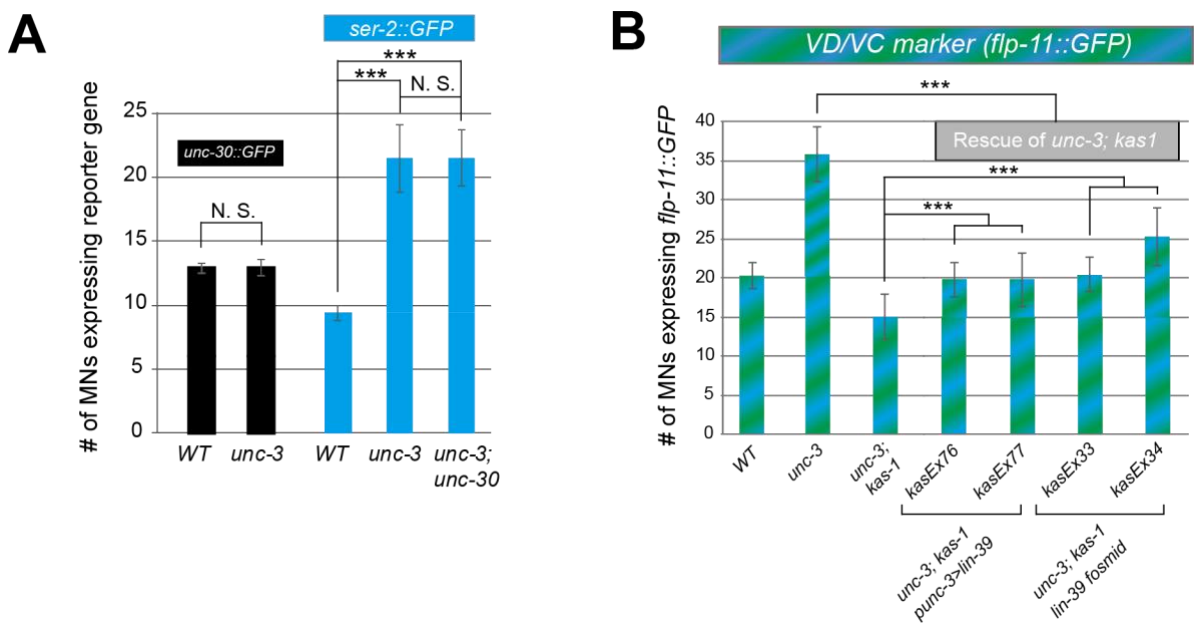

Figure 4.14 Ectopic expression of VD terminal identity markers in *unc-3* mutants requires LIN-39 but not UNC-30. **A**: Quantification of *unc-30::gfp* and VD marker *ser-2::gfp* in WT and *unc-3 (n3435)* animals at L4, with *ser-2::gfp* further examined in *unc-3 (n3435); unc-30 (e191)* double mutants. Expression of *ser-2::gfp* is equally affected in *unc-3 (n3435)* single and *unc-3 (n3435); unc-30 (e191)* double mutants. N > 15. *** $p < 0.001$. N. S: not significant. **B**: Quantification of the VD/VC marker *flp-11::gfp* expression in L4 stage WT, *unc-3 (n3435)* mutant, *unc-3 (n3435); kas1*, and transgenic animals that rescue *unc-3 (n3435); kas1* background with expression of *lin-39* in cholinergic MNs (*Punc-3 > lin-39*) or expression of GFP-tagged *lin-39* fosmid. Two independent transgenic lines were used for *Punc-3 > lin-39 OE* and *lin-39* fosmid. Of note, the partial rescue observed with the *Punc-3 > lin-39 OE* lines likely arises because the fragment of *unc-3* promoter used to drive *lin*-39 is not expressed in all 39 cholinergic MNs. N > 16. *** $p < 0.001$.

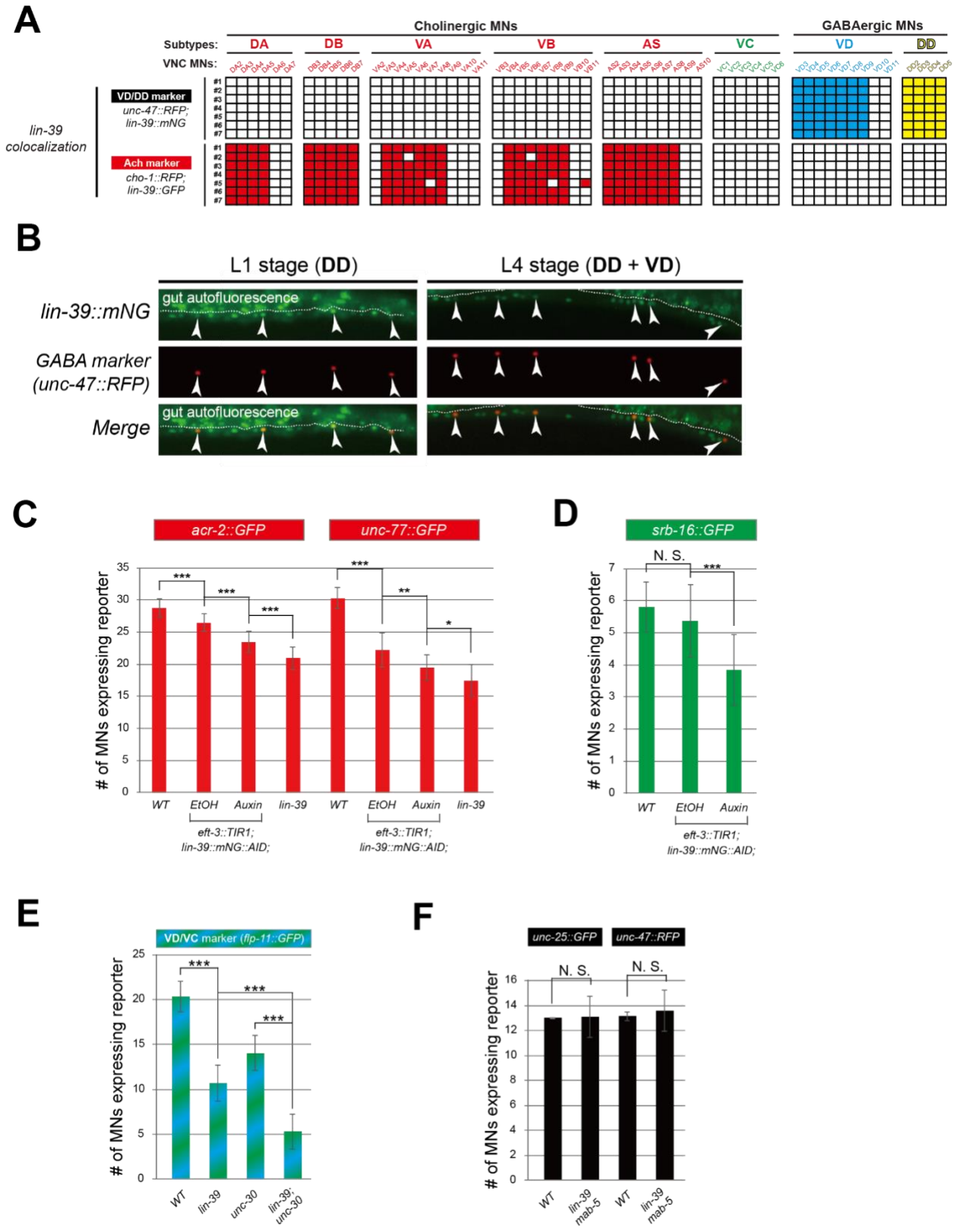

Figure 4.15 LIN-39 is continuously required to activate distinct terminal identity genes in sex-

shared and sex-specific cholinergic MNs. **A:** Co-localization analysis to determine the *lin-39* expression pattern with single-cell resolution. Each column represents an individual MN in the VNC of a hermaphrodite worm. Top seven rows represent seven randomly-selected worms carrying a GABA marker (*unc-47::RFP)* and the *lin-39::mNG::3xFLAG::A*ID endogenous reporter allele. Bottom seven rows represent seven randomly-selected worms carrying a cholinergic marker (*cho-1::RFP* and the *lin-39* fosmid-based reporter *wgIs18 [lin-39^{fosmid}::GFP]*. A color-filled lattice indicates co-expression of the known identity marker and *lin-39* in the individual MN. **B:** Representative images of L1- and L4-stage animals co-expressing *unc-47::rfp* (labels only DD at L1, labels both DD and VD at L4) and endogenous *lin-39* marker (*lin-39::mNG*). Arrowheads point to cell bodies of DD and VD MNs co-expressing both markers. White dotted line indicates the boundary of intestinal tissue (gut), which tends to be autofluorescent in the green channel. **C:** Auxin or ethanol (control) were administered at larval stage 4 (L4) on *lin-39::mNG::3xFLAG::AID; eft-3::TIR1* animals carrying the sex-shared cholinergic MN markers *acr-2::gfp* and *unc-77::gfp*. Images were taken at the young adult stage (day 1 for *acr-2::gfp* and day 2.4 for *unc-77::gfp*) and the number of MNs expressing these markers was quantified. A statistically significant decrease is evident in the auxin-treated animals compared to EtOH-treated controls. For comparison, quantification of marker expression is also provided in WT and *lin-39 (n1760)* animals. N > 12. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. **D:** Auxin or ethanol (control) were administered at larval stage 3 (L3) on *lin-39::mNG::3xFLAG::AID; eft-3::TIR1* animals carrying the VC marker *srb-16::gfp.* Quantification was performed at the young adult stage (day 1.5). A significant decrease in the number of MNs expressing the VC marker was evident in the auxin-treated animals compared to EtOH-treated controls. For comparison, quantification of marker expression is also provided in WT animals. N > 15. ** $p < 0.01$, *** $p < 0.001$. N. S: not significant. **E:** Quantification of VD/VC marker *flp-11::gfp* in WT, *lin-39 (n1760), unc-30 (e191)* and *lin-39 (n1760); unc-30 (e191)* double mutant animals at L4 stage. Double mutants showed a more severe reduction in *flp-11::gfp* expression compared to each single mutant. N > 20. *** $p < 0.001$. **F:** Expression of GABA biosynthesis markers (*unc-25, unc-47*) is not affected in *lin-39 (n1760); mab-5 (e1239)* double mutants at the L4 stage. N > 15; N. S: not significant.

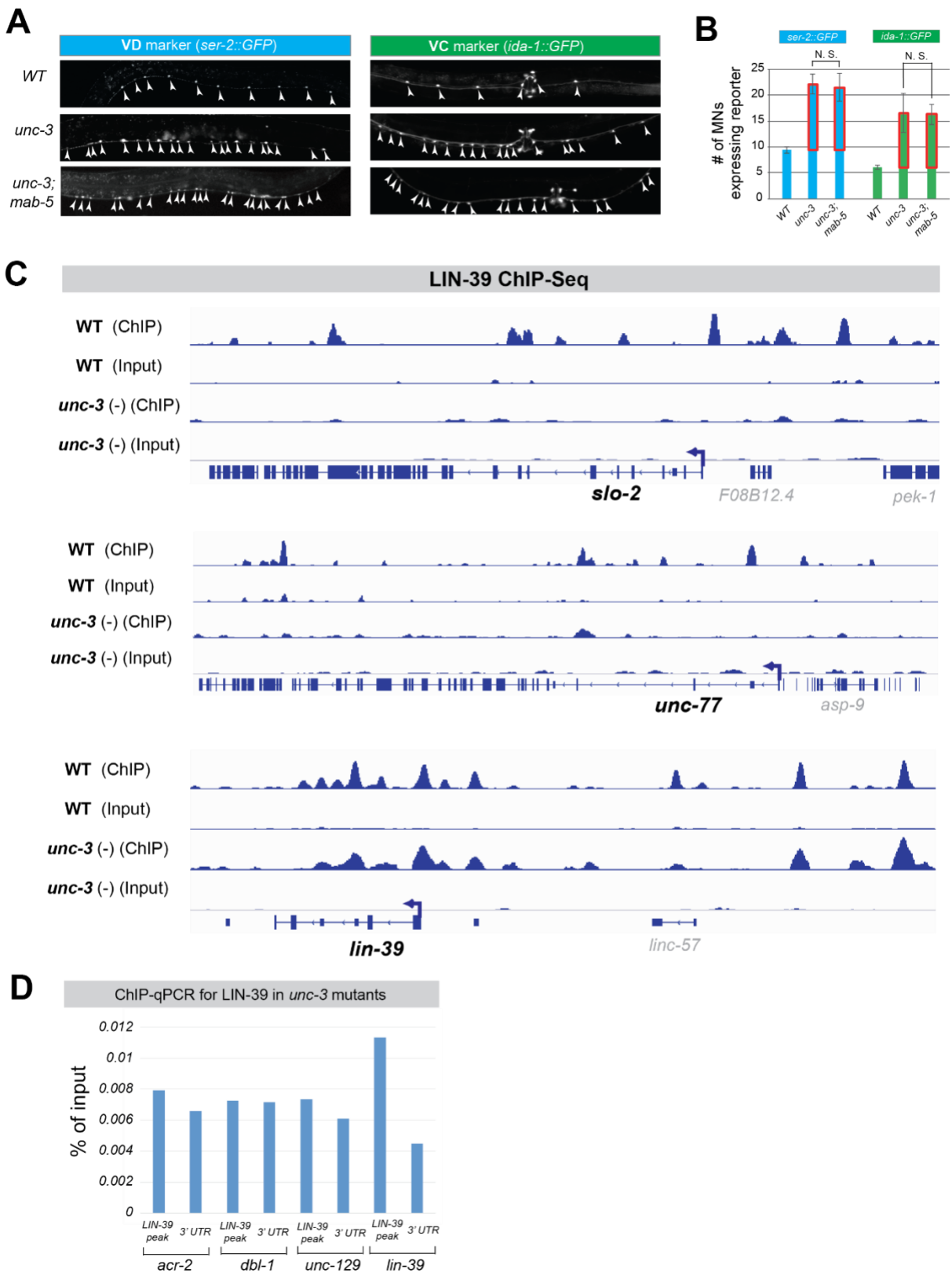

Figure 4.16 MAB-5 is not required for ectopic VD or VC marker expression and LIN-39

binding on cholinergic MN genes is affected in *unc-3* mutants. **A:** The ectopic expression of VD (*ser-2*) or VC (*ida-1*) terminal identity markers in *unc-3* mutants is not affected by loss of *mab-5*. Strong loss-of-function (null) alleles for *unc-3 (n3435)* and *mab-5 (e1239)* were used. Representative images of L4 stage animals are shown. Arrowheads point to MN cell bodies with *gfp* marker expression. Green fluorescence signal is shown in white for better contrast. **B:** Quantification of data shown in panel D. N > 15. N. S: not significant. **C:** ChIP-Seq tracks are shown for LIN-39 on two cholinergic MN terminal identity genes (*slo-2, unc-77*) and *lin-39* locus (positive control). The WT data come from the modENCODE project (Boyle et al., 2014), whereas the *unc-3 (-)* data were obtained by performing ChIP-Seq for LIN-39 on *unc-3 (n3435); lin-39 (kas9 [lin-39::mNG::3xFLAG::AID]* animals. **D:** ChIP-qPCR was performed on *unc-3 (n3435); wgIs18 (lin-39* ^{fosmid}*::GFP)* animals to examine LIN-39 binding at four target genes (*acr-2, dbl-1, unc-129, lin-39*).

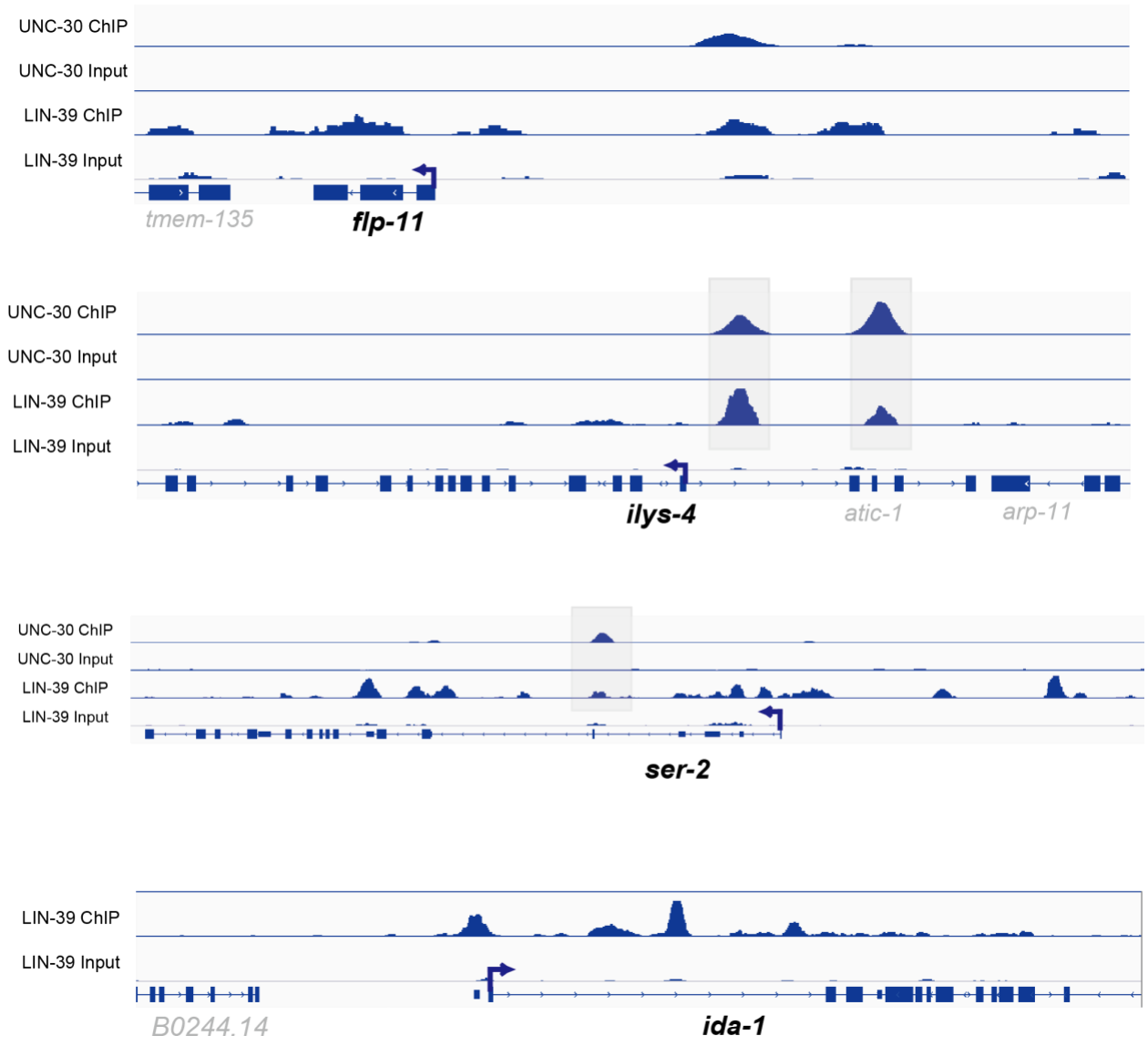

Figure 4.17 LIN-39 binds directly to the *cis*-regulatory region of VD and VC terminal identity genes. ChIP-Seq tracks for UNC-30 and LIN-39 are shown for the VD-expressed genes *flp-11*, *ilys-4*, and *ser-2*. ChIP-Seq tracks for LIN-39 are shown for the VC-specific terminal identity gene *ida-1*. ChIP-Seq data for LIN-39 were generated through the modENCODE project (Boyle et al., 2014) and ChIP-Seq data for UNC-30 were previously generated (Yu et al., 2017). ChIP-seq data were visualized using Integrative Genomics Viewer (Siponen et al.) (Thorvaldsdottir et al., 2013) and snapshots of ChIP-Seq tracks were generated.

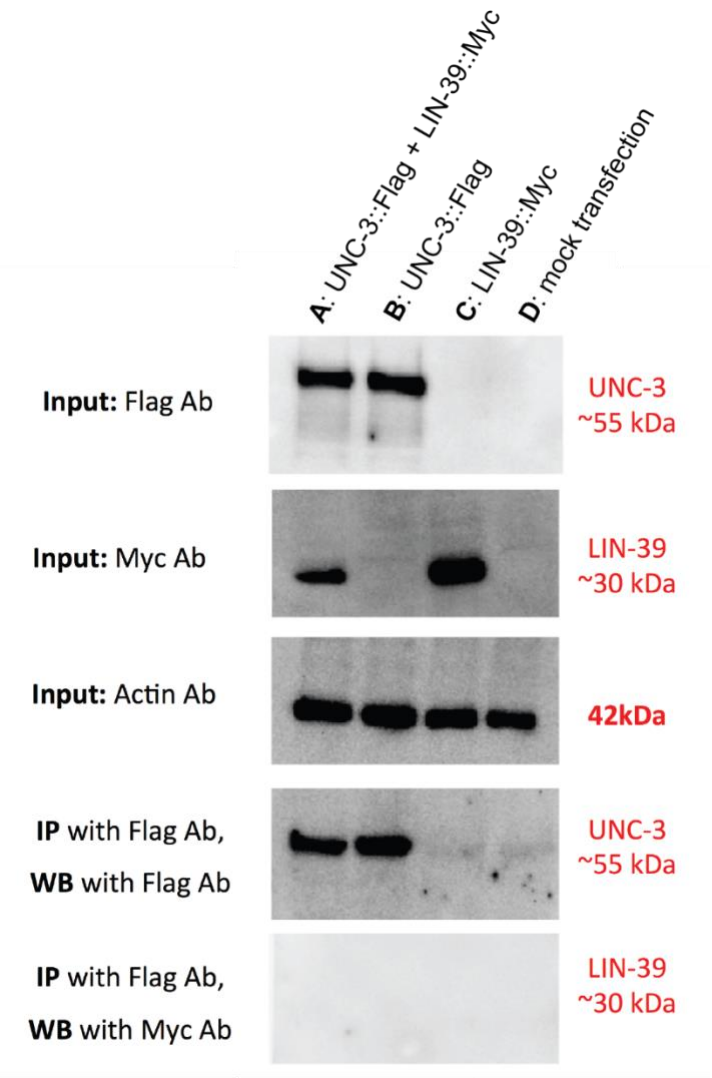

Figure 4.18 UNC-3 does not physically interact with LIN-39 in a heterologous system. (A) HEK293 cells were transfected with two mammalian expression plasmids encoding Flag-tagged UNC-3 and a plasmid encoding Myc-tagged LIN-39, or (B) with a single plasmid encoding the Flag-tagged UNC-3, or (C) with a single plasmid encoding the Myc-tagged LIN-39, or (D) were mock transfected. 10 ug of the protein lysate were subjected to western blot analysis using antibodies against Flag, Myc and Actin (upper three panels). 100 ug of these cell lysates were used for immunoprecipitation of the Flag-tagged UNC-3, using anti-Flag antibody coated beads (lower two panels).

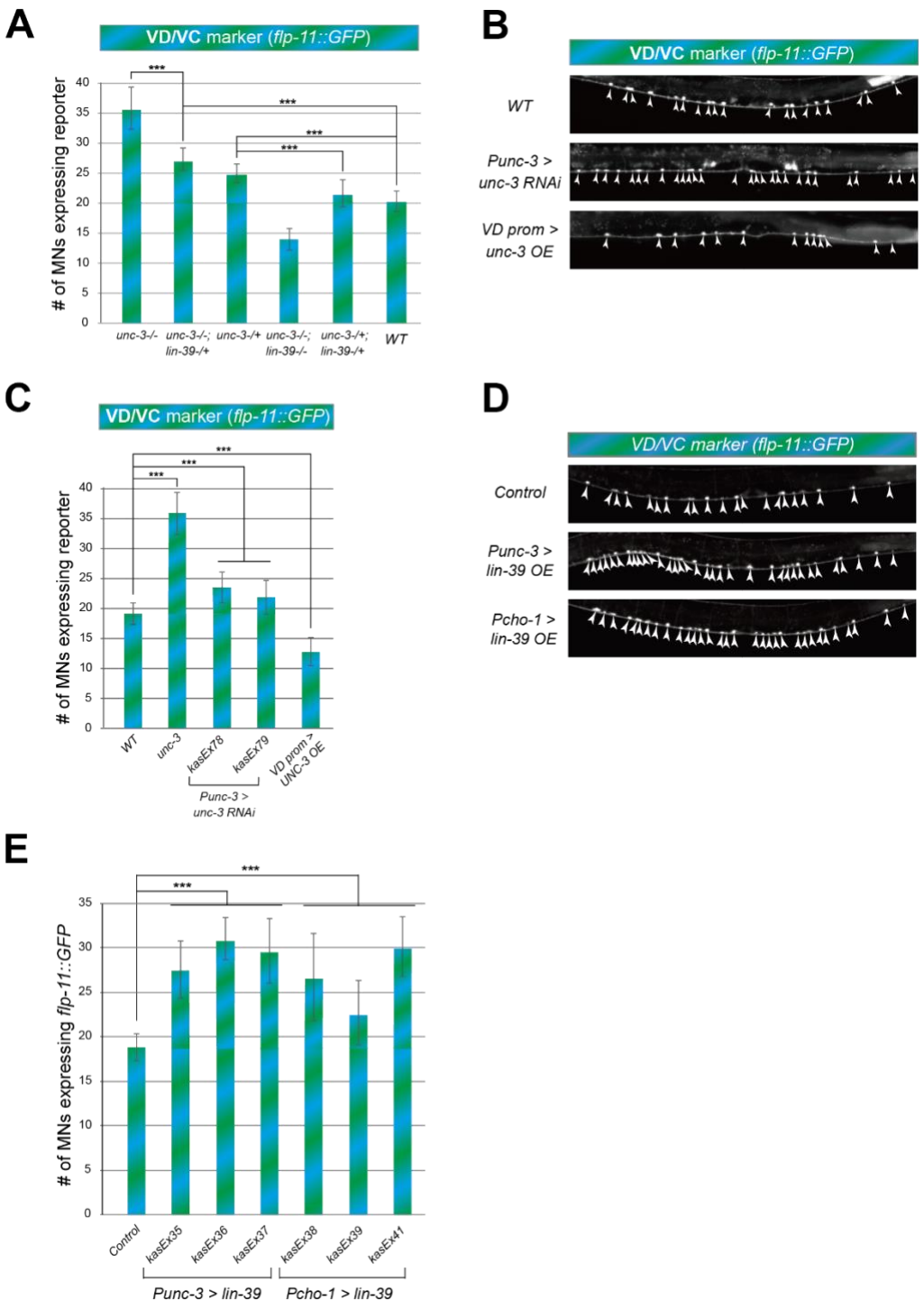

Figure 4.19 UNC-3 and LIN-39 levels are crucial for ectopic expression of VD/VC terminal identity marker *flp-11*. **A:** Quantification of the VD/VC marker *flp-11::gfp* in *unc-3 (n3435), unc-3 (n3435); lin-39 (n1760)/+, unc-3 (n3435)/+, unc-3 (n3435); lin-39 (n1760), unc-3 (n3435)/+; lin-39 (n1760)/+*, and WT animals at L4. N > 18. *** p < 0.001. **B**: Representative images of the VD/VC marker *flp-11::gfp* expression in L4 stage WT or transgenic animals that

either down-regulate *unc-3* in cholinergic MNs (*Punc-3 > unc-3 RNAi*) or over-express *unc-3* in GABAergic (including VD) neurons (*Punc-47 > unc-3 OE*). Arrowheads point to MN cell bodies with *gfp* expression. Green fluorescence signal is shown in white for better contrast. **C:** Quantification of data shown in panel B. Two independent transgenic lines were used for *Punc-3 > unc-3 RNAi*. N > 17. *** $p < 0.001$. Quantification of *flp-11::gfp* is also shown in WT and *unc-3 (n3435)* null animals for comparison. **D**: Representative images of the VD/VC marker *flp-11::gfp* expression in L4 stage WT or transgenic animals that over-express *lin-39* in cholinergic MNs driven by two different cholinergic MN promoters(*Punc-3 > lin-39 OE, Pcho-1 > lin-39 OE*). Arrowheads point to MN cell bodies with marker expression. Green fluorescence signal is shown in white for better contrast. **E:** Quantification of data shown in panel D. Three independent transgenic lines were used for *Punc-3 > lin-39 OE* and *Pcho-1 > lin-39 OE*. N =19. *** $p < 0.001$.

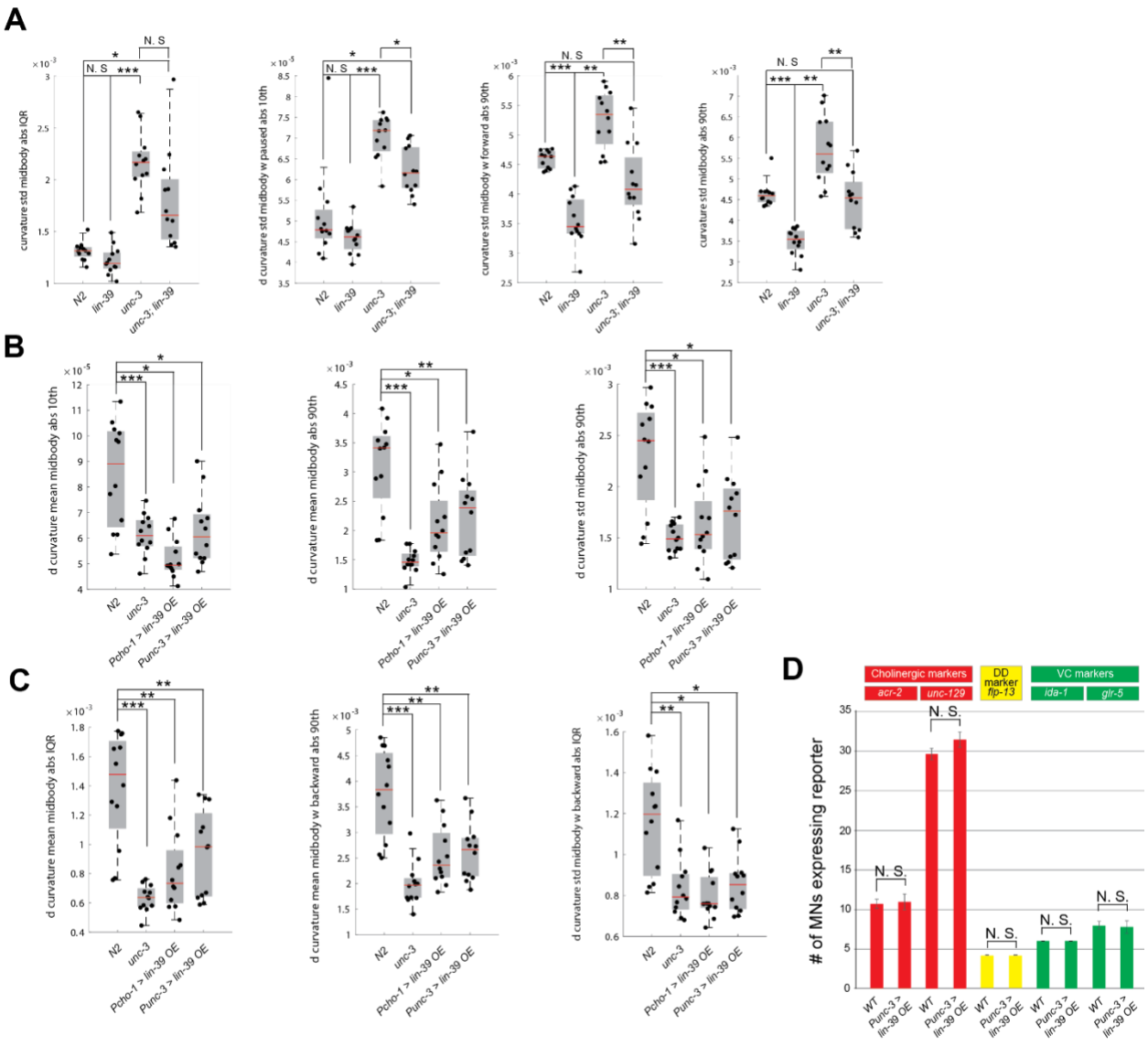

Figure 4.20 Automated worm tracking analysis on *unc-3* and *unc-3; lin-39* mutants. **A**: Additional mid-body locomotion features that are significantly affected in *unc-3 (n3435)* animals, but markedly improved in *unc-3 (n3435); lin-39 (n1760)* double mutant animals. Each black dot represents a single adult animal. The unit for the first two graphs is 1/microns. The unit for the graph on the right is 1/(microns*seconds). N = 12. For a comprehensive list of mid-body features see **Table S3**. * p < 0.01, ** p < 0.001, *** p < 0.0001. **B-C**: Additional mid-body locomotion features affected in *unc-3 (n3435)* mutants and animals over-expressing *lin-39* in cholinergic MNs. Each black dot represents a single adult animal. The unit for the y axis is 1/(microns*seconds). N = 12. For a comprehensive list of mid-body features see **Table 4.4**. * p < 0.01, ** p < 0.001, *** p < 0.0001. **D:** Over-expression of LIN-39 in cholinergic MNs (*Punc-3 > LIN-39 OE*) does not affect expression of terminal identity genes specific to cholinergic (*acr-2, unc-129*), GABAergic DD (*flp-13*) or VC (*ida-1, glr-5*) motor neurons. L4 stage animals were imaged and quantified. N > 15.

4.5 Acknowledgements

We thank the Caenorhabditis Genetics Center (CGC), which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440), for providing strains. We thank Anthony Osuma, Melanie Le Gouez, and Minhkhoi Nguyen for generating *lin-39* and *oig-1* reporter strains. We are grateful to Oliver Hobert, Elizabeth Heckscher, Robert Carillo, Catarina Catela, and Daniele Canzio for comments on this manuscript. This work was funded by an NINDS grant (R00NS084988) and a Whitehall Foundation grant to P.K.

4.6 Material and Methods

C. elegans strains

Worms were grown at 15˚C, 20˚C or 25˚C on nematode growth media (NGM) plates seeded with bacteria (*E.coli* OP50) as food source (Brenner, 1974).

Forward genetic screen

EMS mutagenesis was performed on *unc-3 (n3435); ynIs40 [flp-11::GFP]* animals using standard procedures (Kutscher and Shaham, 2014). Mutagenized L4 animals were visually screened at a dissecting fluorescence microscope for changes in *flp-11::GFP* expression in VNC MNs. One mutant (*kas1*) was retrieved.

Generation of transgenic reporter animals

Reporter gene fusions for *cis*-regulatory analysis of terminal identity genes were made using either PCR fusion (Hobert, 2002) or Gibson Assembly Cloning Kit (NEB #5510S). Targeted DNA fragments were fused (ligated) to *tagrfp* coding sequence, which was followed by *unc-54 3' UTR.* The TOPO XL PCR cloning kit was used to introduce the PCR fusion fragments into the

pCR-XL-TOPO vector (Invitrogen). Mutations on LIN-39 motifs were introduced via mutagenesis PCR. The product DNA fragments were either injected into young adult *pha-1(e2123)* hermaphrodites at 50ng/µl using *pha-1* (pBX plasmid) as co-injection marker (50 ng/µl) and further selected for survival, or injected into young adult N2 hermaphrodites at 50ng/µl (plus 50ng/µl pBX plasmid) using *myo-2::gfp* as co-injection marker (3 ng/µl) and further selected for GFP signal.

The fosmid clone WRM0616aE11 (genomic region: III:7519128..7554793) (Source BioScience) that contains the entire *lin-39* locus was linearized by restriction enzyme digestion, mixed with sonicated bacterial genomic DNA (12 ng/µl) and injected into young adult N2 hermaphrodites at 15 ng/µl using *myo-2::gfp* as co-injection marker (3 ng/µl).

Generation of transgenic animals for RNAi or over-expression

The cDNA (for over-expression) or the exon-rich genomic region (for RNAi) of *unc-3* and *lin-39* were amplified by PCR and then ligated to cholinergic (*cho-1, unc-3*) or GABAergic (*unc-47*) MN promoters using Gibson Assembly Cloning Kit (NEB #5510S). For *unc-3* RNAi, we targeted exons 2-5 with the following primers: FRW: GTCTGTAAAAGATGAGAACCAGCGG, RVS: CTGTCAATAATAACTGGATCGCTGG. For *lin-39* RNAi, we targeted exons 3-5 with the following primers: FRW: gtggtcaaactccgaacttaaagtg, RVS: gaaggggcgagaaatgtgtgataac. For over-expression constructs, DNA products were purified using a PCR purification protocol (QIAGEN), and then injected into young adult WT hermaphrodites at 50 ng/µl together with 50 ng/µl pBS plasmid (filler DNA) and 3 ng/µl of *myo-2::gfp* (co-injection marker). For RNAi constructs, complementary sense and anti-sense exon-rich genomic regions of *unc-3* and *lin-39* were PCR purified and injected into young adult WT or *unc-3 (n3435)* hermaphrodites each at 100 ng/µl with *myo-*

2::gfp as co-injection marker (3 ng/µl) following previously established procedures (Esposito et al., 2007).

Targeted genome engineering

To generated the *lin-39 (kas9 [lin-39::mNG::AID])* allele, CRISPR/Cas9 genome editing was employed to introduce the *mNG::3xFLAG::AID* cassette into the *lin-39* gene locus before the stop codon. Micro-injection, selection and strain establishment were performed as previously described (Dickinson et al., 2015).

Temporally-controlled protein degradation

In the presence of TIR1, AID-tagged proteins are conditionally degraded when exposed to auxin in the presence of TIR1 (Zhang et al., 2015). Animals carrying auxin-inducible alleles of *lin-39 (kas9 [lin-39::mNG::AID])* or *unc-3 (ot837 [unc-3::mNG::AID])* (Kerk et al., 2017) were crossed with ieSi57 [*eft-3prom::tir1*] animals that express TIR1 ubiquitously. Auxin (indole-3-acetic acid [IAA]) was dissolved in ethanol (EtOH) to prepare 400 mM stock solutions which were stored at 4°C for up to one month. NGM agar plates with fully grown OP50 bacteria were coated with auxin solution to a final concentration of 4 mM, and allowed to dry overnight at room temperature. To induce protein degradation, worms of the experimental strains were transferred onto auxin-coated plates and kept at 20°C. As control, worms were transferred onto EtOH-coated plates instead. Auxin solutions, auxin-coated plates, and experimental plates were shielded from light.

Microscopy

Worms were anesthetized using 100mM of sodium azide (NaN_3) and mounted on a 4% agarose pad on glass slides. Images were taken using an automated fluorescence microscope (Zeiss, Axio Imager.Z2). Acquisition of several z-stack images (each ~1 µm thick) was taken with Zeiss

Axiocam 503 mono using the ZEN software (Version 2.3.69.1000, Blue edition). Representative images are shown following max-projection of 1-8 µm Z-stacks using the maximum intensity projection type. Image reconstruction was performed using Image J software (Schindelin et al., 2012).

Chromatin Immunoprecipitation (ChIP)

ChIP assay was performed as previously described (Yu et al., 2017, Zhong et al., 2010) with the following modifications. Synchronized *unc-3 (n3435); lin-39 (kas9 [lin-39::mNG::3xFLAG::AID]* worms at L1 stage were cultured on 10 cm plates seeded with OP50 at 20°C overnight. Early L3 worms were cross-linked and resuspended in FA buffer supplemented with protease inhibitors (150 mM NaCl, 10 µl 0.1 M PMSF, 100 µl 10% SDS, 500 µl 20% N-Lavroyl sarsosine sodium, 2 tablets of cOmplete ULTRA Protease Inhibitor Cocktail [Roche Cat.# 05892970001] in 10ml FA buffer). The sample was then sonicated using a Covaris S220 at the following settings: 200 W Peak Incident Power, 20% Duty Factor, 200 Cycles per Burst for 60 seconds. Samples were transferred to centrifuge tubes and spun at the highest speed for 15 min. The supernatant was transferred to a new tube, and 5% of the material was saved as input and stored at -20°C. Twenty (20) µl of equilibrated anti-FLAG M2 magnetic beads (Sigma-Aldrich M8823) were added to the remainder. The *lin-39 (kas9 [lin-39::mNG::3xFLAG::AID])* CRIPSR-generated allele was used in order to precipitate the immunocomplex comprising the endogenous LIN-39 protein and the bound DNA. The immunocomplex was incubated and rotated overnight at 4°C. On the next day, the beads were washed at 4°C twice with 150 mM NaCl FA buffer (5 min each), once with 1M NaCl FA buffer (5 min). The beads were transferred to a new centrifuge tube and washed twice with 500 mM NaCl FA buffer (10 min each), once with TEL buffer (0.25 M LiCl, 1% NP-40, 1% sodium deoxycholate, 1mM EDTA, 10 mM Tris-

HCl, pH 8.0) for 10 min, twice with TE buffer (5 min each). The immunocomplex was then eluted in 200 µl elution buffer (1% SDS in TE with 250 mM NaCl) by incubating at 65°C for 20 min. The saved input samples were thawed and treated with the ChIP samples as follows. One (1) µl of 20 mg/ml proteinase K was added to each sample and the samples were incubated at 55°C for 2 hours and then at 65°C overnight (12-20 hours) to reverse cross-link. The immunoprecipitated DNA was purified with Ampure XP beads (A63881) according to manufacturer's instructions. Library preparation and Illumina sequencing was performed at the Genomics Core facility of the University of Chicago.

Real-time quantitative PCR (qPCR) analysis of ChIP DNA.

ChIP was performed on *unc-3 (n3435); wgIs18 (lin-39* ^{fosmid}*::GFP)* animals as described above. qPCR analysis of ChIP DNA was performed to probe enrichment of predicted LIN-39 binding sites at four target genes (*acr-2, dbl-1, unc-129, lin-39*). Three biological replicates were included. The primers used are provided in 5'-3' orientation: *acr-2* LIN-39 site (FRW: acattcgcaccaacaaagcg; RVS: aaaggacggacccaacagac), *acr-2* 3' UTR (FRW: tttcagcgccacatgtgtttg; RVS: attgcctagtgattctgagtagagg), *dbl-1* LIN-39 site (FRW: gcacaatccctcgggatcaa; RVS: TAAGTTTTGCGCTGCTGCTG), *dbl-1* 3' UTR (FRW: atacccgcttctatgtcgcc; RVS: ccgtgacacattgcaccaaa), *unc-129* LIN-39 site (FRW: attcgtgtctcgcagggaac; RVS: atagaggaaccggcaaaggtg), *unc-129 3' UTR* (FRW: ttctgtctgtacatcttccctacc; RVS: tttgccaagaaacaaagagagcag), lin-39 LIN-39 site (FRW: gacgtctccctctttctcctc; RVS: tccgctttctgagactccac), *lin-39* 3'UTR (FRW: gttcaagaaaaatattgtgcgttcc; RVS: catttttcgctcgaactgatgga). The amplification was conducted in a QuantStudio 3 using the Power SYBR Green PCR Master Mix (ThermoFisher Cat.# 4367659),

with the following program: Step 1: 95°C for 10 min; Step 2: 95°C for 15 s; Step 3: 60°C for 1 min. Repeat steps 2-3 for 40 times.

Motor neuron identification

Motor neuron (MN) subtypes were identified based on combinations of the following factors: (a) co-localization with fluorescent markers with known expression pattern, (b) invariant cell body position along the ventral nerve cord, or relative to other MN subtypes, (c) MN birth order, and (d) number of MNs that belong to each subtype.

Bioinformatic analysis.

To predict the UNC-3 binding site (COE motif) in the *cis*-regulatory region of *unc-129, del-1, acr-2, unc-77* and *slo-2*, we used the MatInspector program from Genomatix (Cartharius et al., 2005). The Position Weight Matrix (PWM) for the LIN-39 binding site is catalogued in the CIS-BP (Catalog of Inferred Sequence Binding Preferences database) (Weirauch et al., 2014). To identify putative LIN-39 sites on the *cis*-regulatory regions of *unc-129, del-1, acr-2, unc-77, slo-2, oig-1*, and *ser-2,* we used FIMO (Find Individual Motif Occurrences)(Grant et al., 2011), which is one of the motif-based sequence analysis tools of the MEME (Multiple Expectation maximization for Motif Elicitation) bioinformatics suite (http://meme-suite.org/). To predict the binding site for the transcription factor UNC-30, we performed FIMO analysis using the UNC-30 binding motif (WNTAATCHH) described in (Cinar et al., 2005). The p-value threshold for the analysis was set at $p < 0.005$.

Automated worm tracking

Worms were maintained as mixed stage populations by chunking on NGM plates with *E. coli* OP50 as the food source. The day before tracking, 30-40 L4 larvae were transferred to a seeded NGM plate and incubated at 20ºC for approximately 24 hours. Five adults are picked from the

incubated plates to each of the imaging plates (see below) and allowed to habituate for 30 minutes before recording for 15 minutes. Imaging plates are 35 mm plates with 3.5 mL of low-peptone (0.013% Difco Bacto) NGM agar (2% Bio/Agar, BioGene) to limit bacteria growth. Plates are stored at 4ºC for at least two days before use. Imaging plates are seeded with 50µl of a 1:10 dilution of OP50 in M9 the day before tracking and left to dry overnight with the lid on at room temperature.

Behavioral feature extraction and analysis.

All videos were analyzed using Tierpsy Tracker (Javer et al., 2018a) to extract each worm's position and posture over time. These postural data were then converted into a set of behavioral features as previously described (Javer et al., 2018b). From the total set of features, we only considered 48 that are related to midbody posture and motion, as well as the midbody width (see **Table S3** for feature descriptions and their average values for each strain). For each strain comparison, we performed unpaired two-sample t-tests independently for each feature. The false discovery rate was controlled at 5% across all strain and feature comparisons using the Benjamini Yekutieli procedure (Kim and van de Wiel, 2008). The p-value threshold to control the false discovery rate at 0.05 is 0.0032.

Cloning, western blot, and immunoprecipitation.

UNC-3, and LIN-39 cDNAs were cloned into the mammalian expression vectors pcDNA 3.1(+)-C-Flag plasmid and the pcDNA 3.1(+)-N-Myc plasmid by GeneScript, to generate C-terminus Flag-tagged UNC-3 and N-terminus Myc-tagged LIN-39. The constructs were verified by sequencing at the sequencing core facility of University of Chicago. The tagged proteins were expressed in HEK293 cells. Protein expression was detected by standard western blot. Expression of Myc tagged LIN-39 was detected using anti-Myc (Abcam, #ab9106), expression

of Flag-tagged UNC-3 in the total cell lysate was detected using mouse anti-Flag (Sigma, #F3165), expression of Flag-tagged UNC-3 in the IP was detected using rabbit anti-Flag (Sigma, #SAB4301135). Immunoprecipitation of Flag-tagged UNC-3 was performed using Flag antibody coated beads (Sigma, #A2220). For the IP, the "Clean-Blot IP Detection Reagent" (Thermo Fisher, #21230) was used as secondary antibody.

Quantification of Fluorescence Intensity

Images of worms carrying the *lin-39::mNG::3xFLAG::AID* or *unc-3::mNG:: 3xFLAG::AID* alleles were taken on the same slide with the same camera settings at the same development stage. Acquisition of four z-stack images (each 0.53 µm thick) covering the middle portion of targeted MN cell bodies was taken with Zeiss Axiocam 503 mono using the ZEN software (Version 2.3.69.1000, Blue edition). Image reconstruction was performed using Image J software following average-projection the Z-stacks using the average intensity projection protocol. The chosen cells for quantification of mNG fluorescence intensity for both genotypes are the same 10 cholinergic MNs: AS2, DB3, DA2, VA3, VB4, AS3, DA3, VA4, VB5 and DB4. Targeted cell areas were manually selected with minimum background as region of interest (ROI) and the total fluorescence Intensity was measured, calculated, and then represented by Image J as Integrated Density – IntDen (ROI). Background was additionally selected and IntDen (Background) was calculated. The net fluorescence intensity increase is represented as NeIncr = IntDen(ROI) / Area(ROI) – IntDen (Background) / Area(Background). 12 NetIncrs of both genotypes were calculated and data were normalized by dividing the Median of NetIncr (*lin-39::mNG::3xFLAG::AID*) for better contrast and presented as arbitrary units (a.u).

Statistical analysis.

For data quantification, graphs show values expressed as mean ± standard deviation (STDV). The statistical analyses were performed using the unpaired *t*-test (two-tailed). Calculations were performed using the Evan's Awesome A/B Tools online software (https://www.evanmiller.org/ab-testing/t-test.html). Differences with $p < 0.05$ were considered significant. Quantifications are provided in the form of box-and-whisker plots (Tukey boxplot) with individual data point dot-plotted. In all boxplots, middle horizontal line represents the median value (equals to Q2). The box illustrates the interquartile range (IQR), i.e., from Q1 to Q3. The upper limit indicates either the maximum value if maximum < Q3 + 1.5*IQR, or the value that is not higher than Q3 + 1.5*IQR. Similarly, the lower limit indicates either the minimum value or the value that is not lower than Q1 - 1.5*IQR.

Table 4.1 Key Resources Table

Key Resources Table				
Reagent type (species) or resource	**Designation**	**Source or reference**	**Identifiers**	**Additional information**
Gene *(Caenorhabditis elegans)*	*unc-3*	Wormbase	WBGene00006743	
Gene *(Caenorhabditis elegans)*	*unc-30*	Wormbase	WBGene00006766	
Gene *(Caenorhabditis elegans)*	*lin-39*	Wormbase	WBGene00003024	
Gene *(Caenorhabditis elegans)*	*mab-5*	Wormbase	WBGene00003102	
Strain (*Caenorhabditis elegans*)	*unc-3 (n3435)*	Bob Horvitz (MIT,	MT10785	Null Allele: deletion

		Cambridge MA)		
Strain (*Caenorhabditis elegans*)	*unc-30 (e191)*	Caenorhabditis Genetics Center	CB845	Allele: substitution
Strain (*Caenorhabditis elegans*)	*lin-39(n1760) / dpy-17(e164) unc-32(e189) III.*	Caenorhabditis Genetics Center	MT4009	Null Allele: substitution
Strain (*Caenorhabditis elegans*)	*mab-5 (n1239) III; him-5 (e1490) V*	Caenorhabditis Genetics Center	CB3531	Allele: substitution
Strain (*Caenorhabditis elegans*)	*him-8 (e1489) IV*	Caenorhabditis Genetics Center	CB1489	Allele: substitution
Strain (*Caenorhabditis elegans*)	*ieSi57 II; unc-3 (ot837 [unc-3::mNG::AID])*	Caenorhabditis Genetics Center	OH13988	CRISPR-generated allele
Strain (*Caenorhabditis elegans*)	*lin-39 (kas9 [lin-39::mNG::AID])*	This paper	KRA110	See Material and Method, Section

				Targeted genome editing
Strain (*Caenorhabditis elegans*)	*ieSi57 [eft-3prom::tir1]*	Caenorhabditis Genetics Center	CA1200	Genotype: *ieSi57 II; unc-119(ed3) III.*
Strain (*Caenorhabditis elegans*)	*ser-2::gfp*	Caenorhabditis Genetics Center	OH2246	Genotype: *otIs107 I*
Strain (*Caenorhabditis elegans*)	*oig-1::gfp*	Caenorhabditis Genetics Center	OH3955	Genotype: *pha-1(e2123) III; otEx193*
Strain (*Caenorhabditis elegans*)	*ida-1::gfp*	Caenorhabditis Genetics Center	BL5717	Genotype: *inIs179 II; him-8(e1489) IV*
Strain (*Caenorhabditis elegans*)	*glr-5::gfp*	Aixa Alfonso (University of Illinois, Chicago IL)	AL270	Genotype: *icIs270 X*

Strain (*Caenorhabditis elegans*)	*srb-16::gfp*	Caenorhabditis Genetics Center	BC14820	Genotype:*dpy-5(e907) I; sEx14820*
Strain (*Caenorhabditis elegans*)	*flp-11::gfp*	Caenorhabditis Genetics Center	NY2040	Genotype: *ynIs40 V*
Strain (*Caenorhabditis elegans*)	*twk-46::gfp*	Caenorhabditis Genetics Center	BC13337	Genotype: *dpy-5(e907) I; sIs12928 V*
Strain (*Caenorhabditis elegans*)	*ilys-4::tagrfp*	This paper	KRA22	Genotype: *pha-1(e2123) III; kasEx22*
Strain (*Caenorhabditis elegans*)	*flp-13::gfp*	Caenorhabditis Genetics Center	NY2037	Genotype: *ynIs37 III*
Strain (*Caenorhabditis elegans*)	*lin-11::mCherry*	Oliver Hobert (Columbia University, New York NY)	OH11954	Genotype: *lin-11::mCherry + myo-2::GFP V*

Strain (*Caenorhabditis elegans*)	*klp-4::gfp*	Caenorhabditis Genetics Center	BC11799	Genotype: *dpy-5(e907) I; sEx11799*
Strain (*Caenorhabditis elegans*)	*alr-1::egfp*	Caenorhabditis Genetics Center	OP200	Genotype: *unc-119(ed3) III; wgIs200 X*
Strain (*Caenorhabditis elegans*)	*irx-1::egfp*	Caenorhabditis Genetics Center	OP536	Genotype: *unc-119(tm4063) III; wgIs536 I*
Strain (*Caenorhabditis elegans*)	*del-1::gfp*	Caenorhabditis Genetics Center	NC138	
Strain (*Caenorhabditis elegans*)	*acr-2::gfp*	Caenorhabditis Genetics Center	CZ631	Genotype: *juIs14 IV*
Strain (*Caenorhabditis elegans*)	*unc-129::gfp*	Caenorhabditis Genetics Center	*evIs82b*	Genotype: *evIs82b IV*

Strain (*Caenorhabditis elegans*)	*dbl-1::gfp*	Caenorhabditis Genetics Center	BW1935	Genotype: *unc-119(ed3) III; ctIs43 him-5(e1490) V*
Strain (*Caenorhabditis elegans*)	*nca-1::gfp*	Caenorhabditis Genetics Center	BC15028	Genotype: *dpy-5(e907) I; sEx15028*
Strain (*Caenorhabditis elegans*)	*slo-2::gfp*	Caenorhabditis Genetics Center	BC10749	Genotype: *dpy-5(e907) I; sEx10749*
Strain (*Caenorhabditis elegans*)	*ttr-39::mCherry*	Caenorhabditis Genetics Center	CZ8332	Genotype: *juIs223 IV*
Strain (*Caenorhabditis elegans*)	*cho-1::rfp*	Caenorhabditis Genetics Center	OH13646	Genotype: *pha-1(e2123) III; him-5(e1490) otIs544 V*
Strain (*Caenorhabditis elegans*)	*unc-17::gfp*	Caenorhabditis Genetics Center	LX929	Genotype: *vsIs48 X*

Strain (*Caenorhabditis elegans*)	*unc-25::gfp*	Caenorhabditis Genetics Center	CZ13799	Genotype: *juIs76 II*
Strain (*Caenorhabditis elegans*)	*unc-47::mChOpti*	Caenorhabditis Genetics Center	OH13105	Genotype: *him-5(e1490) otIs564 V*
Strain (*Caenorhabditis elegans*)	*unc-30::gfp*	Caenorhabditis Genetics Center	OP395	Genotype: *unc-119(tm4063) III; wgIs395*
Strain (*Caenorhabditis elegans*)	*ser-2::rfp*	Mark Alkema (University of Massachusetts, Worcester MA)	AL270	Genotype: *zfIs8 IV*
Strain (*Caenorhabditis elegans*)	*oig-1(fosmid)::GFP*	Caenorhabditis Genetics Center	OH11809	Genotype: *otIs450*
Strain (*Caenorhabditis elegans*)	*lin-39::gfp*	Caenorhabditis Genetics Center	OP18	Genotype: *unc-119(ed3) III; wgIs18*

Genetic reagent (*Caenorhabditis elegans*)	*Poig-1_1kb::gfp*	Oliver Hobert (Columbia University, New York NY)	otEx5993 otEx5994 otEx5995	See Table S4
Genetic reagent (*Caenorhabditis elegans*)	*Poig-1_1.6kb::gfp*	This paper	kasEx147 kasEx148	See Material and Method and Table S4
Genetic reagent (*Caenorhabditis elegans*)	*Poig-1_2.6kb_LIN-39 site #3 DEL::gfp*	This paper	kasEx149 kasEx150	See Material and Method and Table S4
Genetic reagent (*Caenorhabditis elegans*)	*Poig-1_2.6kb_LIN-39 site #4 DEL::gfp*	This paper	kasEx151 kasEx152	See Material and Method and Table S4
Genetic reagent (*Caenorhabditis elegans*)	*Poig-1_125bp_::tagrfp*	This paper	kasEx80 kasEx81 kasEx82	See Material and Method and Table S4
Genetic reagent (*Caenorhabditis elegans*)	*Poig-1_ LIN-39 site mut 125bp_::tagrfp*	This paper	kasEx91 kasEx92 kasEx93	See Material and Method and Table S4

Genetic reagent (*Caenorhabditis elegans*)	*Punc-3_558bp>lin-39 RNAi + myo-2::gfp*	This paper	kasEx68 kasEx69 kasEx70 kasEx71 kasEx72	See Material and Method
Genetic reagent (*Caenorhabditis elegans*)	*Punc-3_558bp>unc-3 RNAi + myo-2::gfp*	This paper	kasEx73 kasEx74 kasEx78 kasEx79	See Material and Method
Genetic reagent (*Caenorhabditis elegans*)	*Punc-3_558bp>lin-39 cDNA OE + myo-2::gfp*	This paper	kasEx35 kasEx36 kasEx37 kasEx76 kasEx77	See Material and Method
Genetic reagent (*Caenorhabditis elegans*)	*Punc-47>unc-3 cDNA + myo-2::gfp*	This paper	kasEx75	See Material and Method
Genetic reagent (*Caenorhabditis elegans*)	*Pcho-1_280bp>lin-39 cDNA OE + myo-2::gfp*	This paper	kasEx38 kasEx39 kasEx41	See Material and Method

Genetic reagent (*Caenorhabditis elegans*)	*lin-39 fosmid WRM0616aE11 + myo-2::gfp*	This paper	kasEx33 kasEx34	See Material and Method
Antibody	anti-Myc (Rabbit polyclonal)	Abcam	#ab9106	
Antibody	anti-Flag (Mouse monoclonal)	Sigma	#F3165	
Antibody	anti-Flag (Rabbit polyclonal)	Sigma,	#SAB4301135	
Antibody	Clean-Blot IP Detection Reagent (Mouse monoclonal)	Thermo Fisher	#21230	
Antibody	Flag antibody coated beads (Mouse monoclonal)	Sigma,	#A2220	
Antibody	anti-FLAG M2 magnetic beads	Sigma-Aldric	M8823	

	(Mouse monoclonal)			
Recombinant DNA reagent	pcDNA 3.1(+)-C-Flag (Plasmid)	Genscript	pcDNA 3.1(+)	C-terminus Flag-tagged UNC-3
Recombinant DNA reagent	pcDNA 3.1(+)-N-Myc (Plasmid)	Genscript	pcDNA 3.1(+)	N-terminus Myc-tagged LIN-39
Recombinant DNA reagent	Fosmid clone WRM0616aE11	Source BioScience	WRM0616aE11	*lin-39::GFP* fosmid clone
Commercial assay or kit	Gibson Assembly Cloning Kit	NEB	#5510S	
Commercial assay or kit	QIAquick PCR Purification Kit	QIAGEN	#28104	
Commercial assay or kit	Ampure XP beads	Beckman Coulter Life Sciences	A63881	
Commercial assay or kit	TOPO XL-2 Complete PCR Cloning Kit	Thermo Fisher	K8050	

Chemical compound, drug	Auxin (indole-3-acetic acid)	Alfa Aesar	#10196875	
Software, algorithm	ZEN	ZEISS	Version 2.3.69.1000, Blue edition	
Software, algorithm	Image J	Image J	Version 1.52i	
Software, algorithm	RStudio	RStudio	Version 1.2.5001	
Software, algorithm	Adobe Photoshop CS6	Adobe	Version 13.0 x64	
Software, algorithm	Adobe Illustrator CS6	Adobe	Version 16.0.0 x64	

Table 4.2 UNC-3 binding sites (COE motifs) are not found in the *cis*-regulatory region of VD- and VC-expressed terminal identity genes.

Terminal identity gene	**Expression**	**Effect in *unc-3 (-)***	**COE motif**
unc-17	Cholinergic MNs	Loss of expression in MNs	Yes
cho-1	Cholinergic MNs		Yes
acr-2	Cholinergic MNs		Yes
del-1	Cholinergic MNs		Yes
unc-129	Cholinergic MNs		Yes
nca-1	Cholinergic MNs		Yes
slo-2	Cholinergic MNs		Yes
ser-2	VD	Ectopic expression in MNs	No
oig-1	VD		No
flp-11	DD/VD/VC		No
twk-46	DD/VD/VC		No
ilys-4	DD/VD/VC		No
glr-5	VC		No
ida-1	VC		No
srb-16	VC		No

Table 4.3 Novel LIN-39/Hox targets in cholinergic and GABAergic (VD) motor neurons.

LIN-39 targets in cholinergic MNs				LIN-39 targets in GABAergic VD MNs				
	COE motif (relative to ATG)	**LIN-39** sites (relative to ATG)	**P value**			**UNC-30** sites (relative to ATG)	**LIN-39** sites (relative to ATG)	**P value**
unc-129	*COE1 (-262-240)* *COE2 (-346-324)* *COE3 (-458-436)*	*LIN-39 sites:* *#1 (-444-437)* *#2 (-421-414)* *#3 (-376-369)*	*#1 p=0.000119* *#2 p=0.00474* *#3 p=0.0014*	*oig-1**	*Peak#4*	*N. D.*	*LIN-39 sites:* *#1 (-1957-1950)* *#2 (-1868-1861)* *#3 (-1804-1797)*	*#1 p=0.00162* *#2 p=0.00144* *#3 p=0.000645*
					Peak#3	*UNC-30 site:* *(-955-951)*	*#4 (-974-967)* *#5 (-843-836)* *#6 (-801-794)*	*#4 p=0.00377* *#5 p=0.00474* *#6 p=0.00441*
del-1	*COE1 (-194-172)* *COE2 (-2374-2352)*	*LIN-39 sites:* *#1 (-276-269)* *#2 (-323-316)* *#3 (-389-382)* *#4 (-487-480)* *#5 (-2445-2438)* *#6 (-2609-2602)*	*#1 p=0.00259* *#2 p=0.00386* *#3 p=0.00187* *#4 p=0.00402* *#5 p=0.0000396* *#6 p=0.00483*	*ser-2**		*UNC-30 site:* *#1 (-253-246)*	*LIN-39 site:* *#1 (-311-304)* *#2 (-254-247)* *#3 (-198-191)* *#4 (-183-176)* *#5 (-139-132)* *#6 (+121-128)* *#7 (+141-148)*	*#1 p=0.00368* *#2 p=0.00274* *#3 p=0.00472* *#4 p=0.00181* *#5 p=0.00144* *#6 p=0.00461* *#7 p=0.000491*
*acr-2**	*COE2 (-483-461)* *COE1 (-147-126)*	*LIN-39 sites:* *#1 (-791-784)* *#2 (-759-752)* *#3 (-646-639)* *#4 (-580-564)* *#5 (-486-479)* *#6 (-482-475)* *#7 (-468-461)* *#8 (-464-457)*	*#1 p=0.000864* *#2 p=0.000574* *#3 p=0.0016* *#4 p=0.0047* *#5 p=0.00205* *#6 p=0.00135* *#7 p=0.0045* *#8 p=0.00325*					
*dbl-1**	*COE1 (-1433-1411)* *COE2 (-1231-1210)*	*LIN-39 sites:* *#1 (-1475-1468)* *#2 (-1460-1453)* *#3 (-1217-1210)* *#4 (-1209-1202)* *#5 (-1087-1080)* *#6 (-1078-1071)* *#7 (-1035-1-28)*	*#1 p=0.00474* *#2 p=0.00441* *#3 p=0.000947* *#4 p=0.00199* *#5 p=0.000513* *#6 p=0.0027* *#7 p=0.000574*					

Note

1. Asterisk (*) highlights novel LIN-39 targets; N. D: Not Determined.
2. The selected cis-regulatory regions are LIN-39 ChIP-seq peaks from modENCODE project which fall within the DNA sequence used for our reporter constructs (except for del-1).
3. All UNC-3 binding sites (COE motifs 23bp) and LIN-39 binding sites have either been previously described in Kratsios et al. 2012 or been predicted by FIMO search ($p < 0.005$).
4. UNC-30 binding site on ser-2 locus was predicted by FIMO search. UNC-30 site on oig-1 was experimentally validated in Howell et al., 2015.

Table 4.4 Comprehensive list of mid-body features of worm tracking

Feature Name	Detailed description of feature and unit of measurement	N2	lin-39 (n1760)		unc-3 (n3435)		unc-3(n3435);lin-39(n1760)		pha-1;lin-39 OE#9 (+myo-2::GFP)		punc-3;lin-39 OE#2(+myo-2::GFP)	
		mean values	mean values	p-value compared	mean values	p-value compared t	mean values	p-value compared to	mean values	p-value compare	mean values	p-value compared to N2
width_midbody_10th	10th percentile of midbody width	75.0234915415	84.0014581409	0.0000541582	75.7438627879	0.6494456802	74.1518300374	0.6571079053	76.0198186239	0.5693865227	70.4414596558	0.1051941051
curvature_mean_midbody_abs_10th	10th percentile of the absolute value of midbody mean curvature	0.0002777676	0.0002327734	0.0768794467	0.0004309297	0.0000002730	0.0004839073	0.0000006603	0.0002584885	0.1643420590	0.0002785886	0.9447651673
curvature_mean_midbody_abs_90th	90th percentile of the absolute value of midbody mean curvature	0.0034673289	0.0030230876	0.0083845538	0.0053191672	0.0000000000	0.0056124115	0.0000000003	0.0033422481	0.2675993582	0.0034500406	0.8567483491
curvature_std_midbody_abs_10th	10th percentile of the absolute value of the standard deviation of midbody curvature	0.0021135609	0.0011625531	0.0000000000	0.0016177845	0.0001301998	0.0011763271	0.0000000000	0.0019590682	0.0483568078	0.0021511699	0.6400130041
curvature_std_midbody_abs_90th	90th percentile of the absolute value of the standard deviation of midbody curvature	0.0046334698	0.0034972655	0.0000000085	0.0057134557	0.0002669070	0.0044927783	0.5197099486	0.0044166663	0.0430227399	0.0045141600	0.4266389034
d_curvature_mean_midbody_abs_10th	10th percentile of the absolute value of the time derivative of the mean midbody curvature	0.0000849470	0.0000382200	0.0000002414	0.0000612700	0.0011629970	0.0000399035	0.0000002794	0.0000519888	0.0000312533	0.0000630176	0.0053265570
d_curvature_mean_midbody_abs_90th	90th percentile of the absolute value of the time derivative of the mean midbody curvature	0.0031035888	0.0013205179	0.0000021215	0.0014777993	0.0000004485	0.0008164117	0.0000000010	0.0021197299	0.0029460833	0.0022443292	0.0096708231
d_curvature_std_midbody_abs_10th	10th percentile of the absolute value of the time derivative of the standard deviation of the midbody curvature	0.0000960781	0.0000593124	0.0000053025	0.0000797277	0.0096514558	0.0000638274	0.0000132937	0.0000673399	0.0000611033	0.0000765326	0.0070644192
d_curvature_std_midbody_abs_90th	90th percentile of the absolute value of the time derivative of the standard deviation of the midbody curvature	0.0022994295	0.0012433658	0.0000138812	0.0014991598	0.0000428674	0.0010417098	0.0000000546	0.0016378089	0.0023017096	0.0016899912	0.0046789031
curvature_mean_midbody_abs_IQR	Inter-quartile range of the absolute value of midbody mean curvature	0.0017277871	0.0015032336	0.0045787916	0.0026425603	0.0000000000	0.0029855980	0.0000065773	0.0016942977	0.5527287773	0.0017033175	0.6620404400
curvature_std_midbody_abs_IQR	Inter-quartile range of the absolute value of the standard deviation of midbody curvature	0.0013103093	0.0012196298	0.0618287955	0.0021628844	0.0000000012	0.0018009138	0.0017644253	0.0013053204	0.8992526454	0.0012463767	0.0607536392
d_curvature_mean_midbody_abs_IQR	Inter-quartile range of the absolute value of the time derivative of the mean midbody curvature	0.0013825508	0.0005308285	0.0000011202	0.0006315870	0.0000010665	0.0003646393	0.0000000058	0.0008168329	0.0004272971	0.0009399762	0.0044146378
d_curvature_std_midbody_abs_IQR	Inter-quartile range of the absolute value of the time derivative of the standard deviation of the midbody curvature	0.0010235785	0.0005427537	0.0000101053	0.0006822453	0.0001520221	0.0004947625	0.0000004060	0.0006810817	0.0006037740	0.0007376303	0.0036028350
curvature_mean_midbody_w_backward_abs_10th	10th percentile of the absolute value of the midbody mean curvature while the worm is moving backwards	0.0002999378	0.0002971000	0.9355872435	0.0003589157	0.3589007157	0.0003410464	0.4390529191	0.0002817723	0.3366939842	0.0003099948	0.6673932494
curvature_mean_midbody_w_backward_abs_90th	90th percentile of the absolute value of the midbody mean curvature while the worm is moving backwards	0.0035582572	0.0033383697	0.2101658902	0.0042925112	0.0486529721	0.0040628662	0.0299233497	0.0034029360	0.2793559979	0.0033472649	0.0573803242
curvature_std_midbody_w_backward_abs_10th	10th percentile of the absolute value of the standard deviation of midbody curvature while the worm is moving backwards	0.0022459962	0.0013374187	0.0000000004	0.0015006697	0.0000051457	0.0013253095	0.0000000891	0.0021043842	0.0739884609	0.0022738890	0.7463735814
curvature_std_midbody_w_backward_abs_90th	90th percentile of the absolute value of the standard deviation of midbody curvature while the worm is moving backwards	0.0045807281	0.0035819632	0.0000000041	0.0049183895	0.2282985921	0.0037708153	0.0005156780	0.0044625608	0.1382139922	0.0045658811	0.9031339795
d_curvature_mean_midbody_w_backward_abs_10th	10th percentile of the absolute value of the time derivative of the mean midbody curvature while the worm is moving backwards	0.0000976440	0.0000801408	0.0268216124	0.0000876158	0.1991041113	0.0000806002	0.0383193743	0.0000761800	0.0009465612	0.0000794515	0.0053852502
d_curvature_mean_midbody_w_backward_abs_90th	90th percentile of the absolute value of the time derivative of the mean midbody curvature while the worm is moving backwards	0.0037574183	0.0022774685	0.0002639400	0.0019876470	0.0000018304	0.0017908597	0.0000258625	0.0025441181	0.0005373482	0.0026104904	0.0007964295
d_curvature_std_midbody_w_backward_abs_10th	10th percentile of the absolute value of the time derivative of the standard deviation of the midbody curvature while the worm is moving backwards	0.0001025507	0.0000922019	0.1047891334	0.0001006677	0.7756317796	0.0000869873	0.0546936587	0.0000821016	0.0002899332	0.0000862369	0.0034877705
d_curvature_std_midbody_w_backward_abs_90th	90th percentile of the absolute value of the time derivative of the standard deviation of the midbody curvature while the worm is moving backwards	0.0026798610	0.0017535136	0.0001833321	0.0017800531	0.0000331780	0.0014308085	0.0000016107	0.0019294388	0.0004423787	0.0019788707	0.0006729491
curvature_mean_midbody_w_backward_abs_IQR	Inter-quartile range of the absolute value of midbody mean curvature while the worm is moving backwards	0.0018129091	0.0016286181	0.0959289715	0.0021112025	0.1097861057	0.0024260217	0.0043159821	0.0017752878	0.7190108237	0.0016615056	0.0690643683
curvature_std_midbody_w_backward_abs_IQR	Inter-quartile range of the absolute value of the standard deviation of midbody curvature while the worm is moving backwards	0.0012775135	0.0011698254	0.0074125217	0.0018062320	0.0000132221	0.0012311231	0.6768625847	0.0013048159	0.4153599969	0.0012632331	0.5997724968
d_curvature_mean_midbody_w_backward_abs_IQR	Inter-quartile range of the absolute value of the time derivative of the mean midbody curvature while the worm is moving backwards	0.0015662006	0.0009690312	0.0003946130	0.0008784648	0.0000209110	0.0007114153	0.0000021108	0.0010134521	0.0004688463	0.0010848996	0.0020447479
d_curvature_std_midbody_w_backward_abs_IQR	Inter-quartile range of the absolute value of the time derivative of the standard deviation of the midbody curvature while the worm is moving backwards	0.0011570715	0.0007905192	0.0006321441	0.0008355111	0.0009132618	0.0006760672	0.0000100635	0.0007974736	0.0001730004	0.0008542122	0.0014244824
curvature_mean_midbody_w_forward_abs_10th	10th percentile of the absolute value of the midbody mean curvature while the worm is moving forwards	0.0002778324	0.0002765450	0.9447894715	0.0004281933	0.0000434282	0.0004249867	0.0004782919	0.0002731396	0.6817742007	0.0002675620	0.3066211832
curvature_mean_midbody_w_forward_abs_90th	90th percentile of the absolute value of the midbody mean curvature while the worm is moving forwards	0.0034204693	0.0030798557	0.0231579707	0.0047215342	0.0000000003	0.0046727294	0.0000254648	0.0033974867	0.8227015754	0.0034174786	0.9751073851
curvature_std_midbody_w_forward_abs_10th	10th percentile of the absolute value of the standard deviation of midbody curvature while the worm is moving forwards	0.0021413441	0.0013285234	0.0000000107	0.0015754924	0.0000018870	0.0013653509	0.0000000034	0.0020495333	0.1688893534	0.0021227607	0.8147249611
curvature_std_midbody_w_forward_abs_90th	90th percentile of the absolute value of the standard deviation of midbody curvature while the worm is moving forwards	0.0045888340	0.0035490889	0.0000000349	0.0052595603	0.0001694619	0.0041836474	0.0413500333	0.0044909744	0.1003649176	0.0045085451	0.5403072941
d_curvature_mean_midbody_w_forward_abs_10th	10th percentile of the absolute value of the time derivative of the mean midbody curvature while the worm is moving forwards	0.0001150786	0.0000968866	0.1140507608	0.0001064716	0.3279159476	0.0000906372	0.0406650618	0.0000959174	0.0458824449	0.0000885674	0.0056509386
d_curvature_mean_midbody_w_forward_abs_90th	90th percentile of the absolute value of the time derivative of the mean midbody curvature while the worm is moving forwards	0.0033083285	0.0019425167	0.0002715226	0.0024375517	0.0032282332	0.0018092533	0.0000301379	0.0028812906	0.2386578903	0.0026073606	0.0639356682
d_curvature_std_midbody_w_forward_abs_10th	10th percentile of the absolute value of the time derivative of the standard deviation of the midbody curvature while the worm is moving forwards	0.0001153462	0.0001010945	0.1155251073	0.0001157007	0.9592284601	0.0000962295	0.0145811982	0.0000991145	0.0401790137	0.0000949928	0.0099871183
d_curvature_std_midbody_w_forward_abs_90th	90th percentile of the absolute value of the time derivative of the standard deviation of the midbody curvature while the worm is moving forwards	0.0023944387	0.0017399644	0.0056813757	0.0020302421	0.0374865079	0.0014546201	0.0000127969	0.0021062296	0.2069178038	0.0019066550	0.0367409340
curvature_mean_midbody_w_forward_IQR	Inter-quartile range of the absolute value of midbody mean curvature while the worm is moving forwards	0.0017011287	0.0015185416	0.0245280017	0.0024034993	0.0000009329	0.0023950084	0.0007653730	0.0017009846	0.9980711412	0.0016869970	0.7590757450
curvature_std_midbody_w_forward_abs_IQR	Inter-quartile range of the absolute value of the standard deviation of midbody curvature while the worm is moving backwards	0.0012844236	0.0011786164	0.0212442527	0.0020136720	0.0000000001	0.0014562931	0.1767221245	0.0012853976	0.9781777930	0.0012171814	0.0305291303
d_curvature_mean_midbody_w_forward_abs_IQR	Inter-quartile range of the absolute value of the time derivative of the mean midbody curvature while the worm is moving forwards	0.0015558771	0.0009431710	0.0005068267	0.0010872310	0.0007165725	0.0008383038	0.0000174240	0.0013036747	0.1277862419	0.0011738786	0.0151407626
d_curvature_std_midbody_w_forward_abs_IQR	Inter-quartile range of the absolute value of the time derivative of the standard deviation of the midbody curvature while the worm is moving forwards	0.0011144976	0.0008327796	0.0078481175	0.0009615403	0.0514345770	0.0007050433	0.0000194592	0.0009712876	0.1788487262	0.0008642725	0.0106701807
curvature_mean_midbody_w_paused_abs_10th	10th percentile of the absolute value of the midbody mean curvature while the worm is paused	0.0002866092	0.0002472546	0.4496687520	0.0004546605	0.0002519762	0.0005014652	0.0000376338	0.0002386124	0.1567351466	0.0002896316	0.9283735389
curvature_mean_midbody_w_paused_abs_90th	90th percentile of the absolute value of the midbody mean curvature while the worm is paused	0.0036729479	0.0029507615	0.0362770484	0.0055575930	0.0000047746	0.0056809866	0.0000039126	0.0032135919	0.1291540719	0.0036368668	0.9052448042
curvature_std_midbody_w_paused_abs_10th	10th percentile of the absolute value of the standard deviation of midbody curvature while the worm is paused	0.0018722855	0.0010756615	0.0000000000	0.0016549170	0.1021898881	0.0011788228	0.0000000028	0.0018027356	0.4614978826	0.0020349324	0.1432216746
curvature_std_midbody_w_paused_abs_90th	90th percentile of the absolute value of the standard deviation of midbody curvature while the worm is moving forwards	0.0047380475	0.0033401756	0.0001423871	0.0058934023	0.0064937956	0.0045217988	0.5465904388	0.0043072988	0.1623896692	0.0044628394	0.4046081022
d_curvature_mean_midbody_w_paused_abs_10th	10th percentile of the absolute value of the time derivative of the mean midbody curvature while the worm is paused	0.0000346095	0.0000265448	0.0041624792	0.0000513314	0.0000282086	0.0000381568	0.2392972968	0.0000296470	0.0553153527	0.0000312372	0.1778372947
d_curvature_mean_midbody_w_paused_abs_90th	90th percentile of the absolute value of the time derivative of the mean midbody curvature while the worm is paused	0.0007263018	0.0005139740	0.1130784630	0.0010372204	0.0302987422	0.0007348483	0.9488855056	0.0005509137	0.1856665796	0.0005994649	0.3349197333
d_curvature_std_midbody_w_paused_abs_10th	10th percentile of the absolute value of the time derivative of the standard deviation of the midbody curvature while the worm is paused	0.0000512209	0.0000458702	0.1387083172	0.0000702844	0.0000294553	0.0000621547	0.0072441140	0.0000448552	0.0815161504	0.0000463300	0.1684714898
d_curvature_std_midbody_w_paused_abs_90th	90th percentile of the absolute value of the time derivative of the standard deviation of the midbody curvature while the worm is moving forwards	0.0011638812	0.0007382971	0.3258006241	0.0012306293	0.8767875830	0.0010045150	0.7107970357	0.0007075911	0.2926826312	0.0007343568	0.3211237370
curvature_mean_midbody_w_paused_abs_IQR	Inter-quartile range of the absolute value of midbody mean curvature while the worm is paused	0.0017429772	0.0014518945	0.0628281829	0.0027748237	0.0000002374	0.0030490838	0.0000231235	0.0016484939	0.4950460459	0.0017879797	0.7754367136
curvature_std_midbody_w_paused_abs_IQR	Inter-quartile range of the absolute value of the standard deviation of midbody curvature while the worm is paused	0.0015763014	0.0011782266	0.1043163292	0.0022310444	0.0162214628	0.0018144581	0.3863122214	0.0012973439	0.2469649504	0.0012625205	0.1955406899
d_curvature_mean_midbody_w_paused_abs_IQR	Inter-quartile range of the absolute value of the time derivative of the mean midbody curvature while the worm is paused	0.0003332315	0.0002287682	0.0966883102	0.0004682438	0.0424707921	0.0003356207	0.9695080994	0.0002479617	0.1701224586	0.0002699929	0.3041898537
d_curvature_std_midbody_w_paused_abs_IQR	Inter-quartile range of the absolute value of the time derivative of the standard deviation of the midbody curvature while the worm is moving forwards	0.0005355164	0.0003505853	0.3223514717	0.0005678328	0.8616503027	0.0004791289	0.7610099664	0.0003378634	0.2906752708	0.0003484370	0.3165440759

4.7 References

ABOOBAKER, A. A. & BLAXTER, M. L. 2003. Hox Gene Loss during Dynamic Evolution of the Nematode Cluster. *Curr Biol,* 13**,** 37-40.

ANDZELM, M. M., CHERRY, T. J., HARMIN, D. A., BOEKE, A. C., LEE, C., HEMBERG, M., PAWLYK, B., MALIK, A. N., FLAVELL, S. W., SANDBERG, M. A., RAVIOLA, E. & GREENBERG, M. E. 2015. MEF2D drives photoreceptor development through a genome-wide competition for tissue-specific enhancers. *Neuron,* 86**,** 247-63.

ARBER, S., HAN, B., MENDELSOHN, M., SMITH, M., JESSELL, T. M. & SOCKANATHAN, S. 1999. Requirement for the homeobox gene Hb9 in the consolidation of motor neuron identity. *Neuron,* 23**,** 659-74.

ARLOTTA, P. & HOBERT, O. 2015. Homeotic Transformations of Neuronal Cell Identities. *Trends Neurosci,* 38**,** 751-62.

BAEK, M., ENRIQUEZ, J. & MANN, R. S. 2013. Dual role for Hox genes and Hox co-factors in conferring leg motoneuron survival and identity in Drosophila. *Development,* 140**,** 2027-38.

BORROMEO, M. D., MEREDITH, D. M., CASTRO, D. S., CHANG, J. C., TUNG, K. C., GUILLEMOT, F. & JOHNSON, J. E. 2014. A transcription factor network specifying inhibitory versus excitatory neurons in the dorsal spinal cord. *Development,* 141**,** 2803-12.

BOYLE, A. P., ARAYA, C. L., BRDLIK, C., CAYTING, P., CHENG, C., CHENG, Y., GARDNER, K., HILLIER, L. W., JANETTE, J., JIANG, L., KASPER, D., KAWLI, T., KHERADPOUR, P., KUNDAJE, A., LI, J. J., MA, L., NIU, W., REHM, E. J., ROZOWSKY, J., SLATTERY, M., SPOKONY, R., TERRELL, R., VAFEADOS, D., WANG, D., WEISDEPP, P., WU, Y. C., XIE, D., YAN, K. K., FEINGOLD, E. A., GOOD, P. J., PAZIN, M. J., HUANG, H., BICKEL, P. J., BRENNER, S. E., REINKE, V., WATERSTON, R. H., GERSTEIN, M.,

WHITE, K. P., KELLIS, M. & SNYDER, M. 2014. Comparative analysis of regulatory information and circuits across distant species. *Nature,* 512**,** 453-6.

BRENNER, S. 1974. The genetics of Caenorhabditis elegans. *Genetics,* 77**,** 71-94.

BRITANOVA, O., DE JUAN ROMERO, C., CHEUNG, A., KWAN, K. Y., SCHWARK, M., GYORGY, A., VOGEL, T., AKOPOV, S., MITKOVSKI, M., AGOSTON, D., SESTAN, N., MOLNAR, Z. & TARABYKIN, V. 2008. Satb2 is a postmitotic determinant for upper-layer neuron specification in the neocortex. *Neuron,* 57**,** 378-92.

CARTHARIUS, K., FRECH, K., GROTE, K., KLOCKE, B., HALTMEIER, M., KLINGENHOFF, A., FRISCH, M., BAYERLEIN, M. & WERNER, T. 2005. MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics,* 21**,** 2933-42.

CATELA, C., CORREA, E., WEN, K., ABURAS, J., CROCI, L., CONSALEZ, G. G. & KRATSIOS, P. 2019. An ancient role for collier/Olf/Ebf (COE)-type transcription factors in axial motor neuron development. *Neural Dev,* 14**,** 2.

CATELA, C., SHIN, M. M., LEE, D. H., LIU, J. P. & DASEN, J. S. 2016. Hox Proteins Coordinate Motor Neuron Differentiation and Connectivity Programs through Ret/Gfralpha Genes. *Cell Rep,* 14**,** 1901-15.

CHENG, L., ARATA, A., MIZUGUCHI, R., QIAN, Y., KARUNARATNE, A., GRAY, P. A., ARATA, S., SHIRASAWA, S., BOUCHARD, M., LUO, P., CHEN, C. L., BUSSLINGER, M., GOULDING, M., ONIMARU, H. & MA, Q. 2004. Tlx3 and Tlx1 are post-mitotic selector genes determining glutamatergic over GABAergic cell fates. *Nat Neurosci,* 7**,** 510-7.

CINAR, H., KELES, S. & JIN, Y. 2005. Expression profiling of GABAergic motor neurons in Caenorhabditis elegans. *Curr Biol,* 15**,** 340-6.

CLARK, S. G., CHISHOLM, A. D. & HORVITZ, H. R. 1993. Control of cell fates in the central body region of C. elegans by the homeobox gene lin-39. *Cell,* 74**,** 43-55.

CLOVIS, Y. M., SEO, S. Y., KWON, J. S., RHEE, J. C., YEO, S., LEE, J. W., LEE, S. & LEE, S. K. 2016. Chx10 Consolidates V2a Interneuron Identity through Two Distinct Gene Repression Modes. *Cell Rep,* 16**,** 1642-1652.

DENERIS, E. S. & HOBERT, O. 2014. Maintenance of postmitotic neuronal cell identity. *Nat Neurosci,* 17**,** 899-907.

DICKINSON, D. J., PANI, A. M., HEPPERT, J. K., HIGGINS, C. D. & GOLDSTEIN, B. 2015. Streamlined Genome Engineering with a Self-Excising Drug Selection Cassette. *Genetics,* 200**,** 1035-49.

EASTMAN, C., HORVITZ, H. R. & JIN, Y. 1999. Coordinated transcriptional regulation of the unc-25 glutamic acid decarboxylase and the unc-47 GABA vesicular transporter by the Caenorhabditis elegans UNC-30 homeodomain protein. *J Neurosci,* 19**,** 6225-34.

ESPOSITO, G., DI SCHIAVI, E., BERGAMASCO, C. & BAZZICALUPO, P. 2007. Efficient and cell specific knock-down of gene function in targeted C. elegans neurons. *Gene,* 395**,** 170-6.

ESTACIO-GOMEZ, A. & DIAZ-BENJUMEA, F. J. 2014. Roles of Hox genes in the patterning of the central nervous system of Drosophila. *Fly (Austin),* 8**,** 26-32.

ESTACIO-GOMEZ, A., MORIS-SANZ, M., SCHAFER, A. K., PEREA, D., HERRERO, P. & DIAZ-BENJUMEA, F. J. 2013. Bithorax-complex genes sculpt the pattern of leucokinergic neurons in the Drosophila central nervous system. *Development,* 140**,** 2139-48.

FAUMONT, S., RONDEAU, G., THIELE, T. R., LAWTON, K. J., MCCORMICK, K. E., SOTTILE, M., GRIESBECK, O., HECKSCHER, E. S., ROBERTS, W. M., DOE, C. Q. & LOCKERY, S. R. 2011. An image-free opto-mechanical system for creating virtual

environments and imaging neuronal activity in freely moving Caenorhabditis elegans. *PLoS One,* 6**,** e24666.

GORDON, P. M. & HOBERT, O. 2015. A competition mechanism for a homeotic neuron identity transformation in C. elegans. *Dev Cell,* 34**,** 206-19.

GRANT, C. E., BAILEY, T. L. & NOBLE, W. S. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics,* 27**,** 1017-8.

HOBERT, O. 2002. PCR fusion-based approach to create reporter gene constructs for expression analysis in transgenic C. elegans. *Biotechniques,* 32**,** 728-30.

HOBERT, O. 2008. Regulatory logic of neuronal diversity: terminal selector genes and selector motifs. *Proc Natl Acad Sci U S A,* 105**,** 20067-71.

HOBERT, O. 2011. Regulation of terminal differentiation programs in the nervous system. *Annu Rev Cell Dev Biol,* 27**,** 681-96.

HOBERT, O. 2016. Terminal Selectors of Neuronal Identity. *Curr Top Dev Biol,* 116**,** 455-75.

HOBERT, O. & KRATSIOS, P. 2019. Neuronal identity control by terminal selectors in worms, flies, and chordates. *Curr Opin Neurobiol,* 56**,** 97-105.

HOWELL, K., WHITE, J. G. & HOBERT, O. 2015. Spatiotemporal control of a novel synaptic organizer molecule. *Nature,* 523**,** 83-7.

HUTLET, B., THEYS, N., COSTE, C., AHN, M. T., DOSHISHTI-AGOLLI, K., LIZEN, B. & GOFFLOT, F. 2016. Systematic expression analysis of Hox genes at adulthood reveals novel patterns in the central nervous system. *Brain Struct Funct,* 221**,** 1223-43.

IMBRICI, P., CAMERINO, D. C. & TRICARICO, D. 2013. Major channels involved in neuropsychiatric disorders and therapeutic perspectives. *Front Genet,* 4**,** 76.

JAVER, A., CURRIE, M., LEE, C. W., HOKANSON, J., LI, K., MARTINEAU, C. N., YEMINI, E., GRUNDY, L. J., LI, C., CH'NG, Q., SCHAFER, W. R., NOLLEN, E. A. A., KERR, R. & BROWN, A. E. X. 2018a. An open-source platform for analyzing and sharing worm-behavior data. *Nat Methods,* 15**,** 645-646.

JAVER, A., RIPOLL-SANCHEZ, L. & BROWN, A. E. X. 2018b. Powerful and interpretable behavioural features for quantitative phenotyping of Caenorhabditis elegans. *Philos Trans R Soc Lond B Biol Sci,* 373.

JIN, Y., HOSKINS, R. & HORVITZ, H. R. 1994. Control of type-D GABAergic neuron differentiation by C. elegans UNC-30 homeodomain protein. *Nature,* 372**,** 780-3.

KALA, K., HAUGAS, M., LILLEVALI, K., GUIMERA, J., WURST, W., SALMINEN, M. & PARTANEN, J. 2009. Gata2 is a tissue-specific post-mitotic selector gene for midbrain GABAergic neurons. *Development,* 136**,** 253-62.

KARLSSON, D., BAUMGARDT, M. & THOR, S. 2010. Segment-specific neuronal subtype specification by the integration of anteroposterior and temporal cues. *PLoS Biol,* 8**,** e1000368.

KERK, S. Y., KRATSIOS, P., HART, M., MOURAO, R. & HOBERT, O. 2017. Diversification of C. elegans Motor Neuron Identity via Selective Effector Gene Repression. *Neuron,* 93**,** 80-98.

KIM, K., COLOSIMO, M. E., YEUNG, H. & SENGUPTA, P. 2005. The UNC-3 Olf/EBF protein represses alternate neuronal programs to specify chemosensory neuron identity. *Dev Biol,* 286**,** 136-48.

KIM, K. I. & VAN DE WIEL, M. A. 2008. Effects of dependence in high-dimensional multiple testing problems. *BMC Bioinformatics,* 9**,** 114.

KRATSIOS, P., KERK, S. Y., CATELA, C., LIANG, J., VIDAL, B., BAYER, E. A., FENG, W., DE LA CRUZ, E. D., CROCI, L., CONSALEZ, G. G., MIZUMOTO, K. & HOBERT, O.

2017. An intersectional gene regulatory strategy defines subclass diversity of C. elegans motor neurons. *Elife,* 6.

KRATSIOS, P., STOLFI, A., LEVINE, M. & HOBERT, O. 2012. Coordinated regulation of cholinergic motor neuron traits through a conserved terminal selector gene. *Nat Neurosci,* 15**,** 205-14.

KUTSCHER, L. M. & SHAHAM, S. 2014. Forward and reverse mutagenesis in C. elegans. *WormBook***,** 1-26.

LEE, S., LEE, B., JOSHI, K., PFAFF, S. L., LEE, J. W. & LEE, S. K. 2008. A regulatory network to segregate the identity of neuronal subtypes. *Dev Cell,* 14**,** 877-89.

LODATO, S., MOLYNEAUX, B. J., ZUCCARO, E., GOFF, L. A., CHEN, H. H., YUAN, W., MELESKI, A., TAKAHASHI, E., MAHONY, S., RINN, J. L., GIFFORD, D. K. & ARLOTTA, P. 2014. Gene co-regulation by Fezf2 selects neurotransmitter identity and connectivity of corticospinal neurons. *Nat Neurosci,* 17**,** 1046-54.

LOPES, R., VERHEY VAN WIJK, N., NEVES, G. & PACHNIS, V. 2012. Transcription factor LIM homeobox 7 (Lhx7) maintains subtype identity of cholinergic interneurons in the mammalian striatum. *Proc Natl Acad Sci U S A,* 109**,** 3119-24.

MALOOF, J. N. & KENYON, C. 1998. The Hox gene lin-39 is required during C. elegans vulval induction to select the outcome of Ras signaling. *Development,* 125**,** 181-90.

MEARS, A. J., KONDO, M., SWAIN, P. K., TAKADA, Y., BUSH, R. A., SAUNDERS, T. L., SIEVING, P. A. & SWAROOP, A. 2001. Nrl is required for rod photoreceptor development. *Nat Genet,* 29**,** 447-52.

MENDELSOHN, A. I., DASEN, J. S. & JESSELL, T. M. 2017. Divergent Hox Coding and Evasion of Retinoid Signaling Specifies Motor Neurons Innervating Digit Muscles. *Neuron,* 93**,** 792-805 e4.

MIGUEL-ALIAGA, I. & THOR, S. 2004. Segment-specific prevention of pioneer neuron apoptosis by cell-autonomous, postmitotic Hox gene activity. *Development,* 131**,** 6093-105.

MOREY, M., YEE, S. K., HERMAN, T., NERN, A., BLANCO, E. & ZIPURSKY, S. L. 2008. Coordinate control of synaptic-layer specificity and rhodopsins in photoreceptor neurons. *Nature,* 456**,** 795-9.

MORIS-SANZ, M., ESTACIO-GOMEZ, A., SANCHEZ-HERRERO, E. & DIAZ-BENJUMEA, F. J. 2015. The study of the Bithorax-complex genes in patterning CCAP neurons reveals a temporal control of neuronal differentiation by Abd-B. *Biol Open,* 4**,** 1132-42.

NAKATANI, T., MINAKI, Y., KUMAI, M. & ONO, Y. 2007. Helt determines GABAergic over glutamatergic neuronal fate by repressing Ngn genes in the developing mesencephalon. *Development,* 134**,** 2783-93.

NIU, W., LU, Z. J., ZHONG, M., SAROV, M., MURRAY, J. I., BRDLIK, C. M., JANETTE, J., CHEN, C., ALVES, P., PRESTON, E., SLIGHTHAM, C., JIANG, L., HYMAN, A. A., KIM, S. K., WATERSTON, R. H., GERSTEIN, M., SNYDER, M. & REINKE, V. 2011. Diverse transcription factor binding features revealed by genome-wide ChIP-seq in C. elegans. *Genome Res,* 21**,** 245-54.

PEREIRA, L., KRATSIOS, P., SERRANO-SAIZ, E., SHEFTEL, H., MAYO, A. E., HALL, D. H., WHITE, J. G., LEBOEUF, B., GARCIA, L. R., ALON, U. & HOBERT, O. 2015. A cellular and regulatory map of the cholinergic nervous system of C. elegans. *Elife,* 4.

PFLUGRAD, A., MEIR, J. Y., BARNES, T. M. & MILLER, D. M., 3RD 1997. The Groucho-like transcription factor UNC-37 functions with the neural specificity gene unc-4 to govern motor neuron identity in C. elegans. *Development,* 124**,** 1699-709.

PHILIPPIDOU, P. & DASEN, J. S. 2013. Hox genes: choreographers in neural development, architects of circuit organization. *Neuron,* 80**,** 12-34.

POCOCK, R., MIONE, M., HUSSAIN, S., MAXWELL, S., PONTECORVI, M., ASLAM, S., GERRELLI, D., SOWDEN, J. C. & WOOLLARD, A. 2008. Neuronal function of Tbx20 conserved from nematodes to vertebrates. *Dev Biol,* 317**,** 671-85.

PORTMAN, D. S. 2017. Sexual modulation of sex-shared neurons and circuits in Caenorhabditis elegans. *J Neurosci Res,* 95**,** 527-538.

POTTS, M. B., WANG, D. P. & CAMERON, S. 2009. Trithorax, Hox, and TALE-class homeodomain proteins ensure cell survival through repression of the BH3-only gene egl-1. *Dev Biol,* 329**,** 374-85.

PRASAD, B., KARAKUZU, O., REED, R. R. & CAMERON, S. 2008. unc-3-dependent repression of specific motor neuron fates in Caenorhabditis elegans. *Dev Biol,* 323**,** 207-15.

PRASAD, B. C., YE, B., ZACKHARY, R., SCHRADER, K., SEYDOUX, G. & REED, R. R. 1998. unc-3, a gene required for axonal guidance in Caenorhabditis elegans, encodes a member of the O/E family of transcription factors. *Development,* 125**,** 1561-8.

RHEE, H. S., CLOSSER, M., GUO, Y., BASHKIROVA, E. V., TAN, G. C., GIFFORD, D. K. & WICHTERLE, H. 2016. Expression of Terminal Effector Genes in Mammalian Neurons Is Maintained by a Dynamic Relay of Transient Enhancers. *Neuron,* 92**,** 1252-1265.

SAGASTI, A., HOBERT, O., TROEMEL, E. R., RUVKUN, G. & BARGMANN, C. I. 1999. Alternative olfactory neuron fates are specified by the LIM homeobox gene lim-4. *Genes Dev,* 13**,** 1794-806.

SALSER, S. J., LOER, C. M. & KENYON, C. 1993. Multiple HOM-C gene interactions specify cell fates in the nematode central nervous system. *Genes Dev,* 7**,** 1714-24.

SCHAFER, W. R. 2005. Egg-laying. *WormBook***,** 1-7.

SCHINDELIN, J., ARGANDA-CARRERAS, I., FRISE, E., KAYNIG, V., LONGAIR, M., PIETZSCH, T., PREIBISCH, S., RUEDEN, C., SAALFELD, S., SCHMID, B., TINEVEZ, J. Y., WHITE, D. J., HARTENSTEIN, V., ELICEIRI, K., TOMANCAK, P. & CARDONA, A. 2012. Fiji: an open-source platform for biological-image analysis. *Nat Methods,* 9**,** 676-82.

SCHINDELMAN, G., WHITTAKER, A. J., THUM, J. Y., GHARIB, S. & STERNBERG, P. W. 2006. Initiation of male sperm-transfer behavior in Caenorhabditis elegans requires input from the ventral nerve cord. *BMC Biol,* 4**,** 26.

SGADO, P., DUNLEAVY, M., GENOVESI, S., PROVENZANO, G. & BOZZI, Y. 2011. The role of GABAergic system in neurodevelopmental disorders: a focus on autism and epilepsy. *Int J Physiol Pathophysiol Pharmacol,* 3**,** 223-35.

SHAN, G., KIM, K., LI, C. & WALTHALL, W. W. 2005. Convergent genetic programs regulate similarities and differences between related motor neuron classes in Caenorhabditis elegans. *Dev Biol,* 280**,** 494-503.

SHIBUYA, K., MISAWA, S., ARAI, K., NAKATA, M., KANAI, K., YOSHIYAMA, Y., ITO, K., ISOSE, S., NOTO, Y., NASU, S., SEKIGUCHI, Y., FUJIMAKI, Y., OHMORI, S., KITAMURA, H., SATO, Y. & KUWABARA, S. 2011. Markedly reduced axonal potassium

channel expression in human sporadic amyotrophic lateral sclerosis: an immunohistochemical study. *Exp Neurol,* 232**,** 149-53.

SIPONEN, M. I., WISNIEWSKA, M., LEHTIO, L., JOHANSSON, I., SVENSSON, L., RASZEWSKI, G., NILSSON, L., SIGVARDSSON, M. & BERGLUND, H. 2010. Structural determination of functional domains in early B-cell factor (EBF) family of transcription factors reveals similarities to Rel DNA-binding proteins and a novel dimerization motif. *J Biol Chem,* 285**,** 25875-9.

SONG, M. R., SUN, Y., BRYSON, A., GILL, G. N., EVANS, S. M. & PFAFF, S. L. 2009. Islet-to-LMO stoichiometries control the function of transcription complexes that specify motor neuron and V2a interneuron identity. *Development,* 136**,** 2923-32.

STEFANAKIS, N., CARRERA, I. & HOBERT, O. 2015. Regulatory Logic of Pan-Neuronal Gene Expression in C. elegans. *Neuron,* 87**,** 733-50.

TAKAHASHI, Y., HAMADA, J., MURAKAWA, K., TAKADA, M., TADA, M., NOGAMI, I., HAYASHI, N., NAKAMORI, S., MONDEN, M., MIYAMOTO, M., KATOH, H. & MORIUCHI, T. 2004. Expression profiles of 39 HOX genes in normal human adult organs and anaplastic thyroid cancer cell lines by quantitative real-time RT-PCR system. *Exp Cell Res,* 293**,** 144-53.

THALER, J., HARRISON, K., SHARMA, K., LETTIERI, K., KEHRL, J. & PFAFF, S. L. 1999. Active suppression of interneuron programs within developing motor neurons revealed by analysis of homeodomain factor HB9. *Neuron,* 23**,** 675-87.

THALER, J. P., LEE, S. K., JURATA, L. W., GILL, G. N. & PFAFF, S. L. 2002. LIM factor Lhx3 contributes to the specification of motor neuron and interneuron identity through cell-type-specific protein-protein interactions. *Cell,* 110**,** 237-49.

THORVALDSDOTTIR, H., ROBINSON, J. T. & MESIROV, J. P. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform,* 14**,** 178-92.

VON STETINA, S. E., FOX, R. M., WATKINS, K. L., STARICH, T. A., SHAW, J. E. & MILLER, D. M., 3RD 2007. UNC-4 represses CEH-12/HB9 to specify synaptic inputs to VA motor neurons in C. elegans. *Genes Dev,* 21**,** 332-46.

VON STETINA, S. E., TREININ, M. & MILLER, D. M., 3RD 2006. The motor circuit. *Int Rev Neurobiol,* 69**,** 125-67.

WEIRAUCH, M. T., YANG, A., ALBU, M., COTE, A. G., MONTENEGRO-MONTERO, A., DREWE, P., NAJAFABADI, H. S., LAMBERT, S. A., MANN, I., COOK, K., ZHENG, H., GOITY, A., VAN BAKEL, H., LOZANO, J. C., GALLI, M., LEWSEY, M. G., HUANG, E., MUKHERJEE, T., CHEN, X., REECE-HOYES, J. S., GOVINDARAJAN, S., SHAULSKY, G., WALHOUT, A. J. M., BOUGET, F. Y., RATSCH, G., LARRONDO, L. F., ECKER, J. R. & HUGHES, T. R. 2014. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell,* 158**,** 1431-1443.

WHITE, J. G., SOUTHGATE, E., THOMSON, J. N. & BRENNER, S. 1986. The structure of the nervous system of the nematode Caenorhabditis elegans. *Philos Trans R Soc Lond B Biol Sci,* 314**,** 1-340.

WINNIER, A. R., MEIR, J. Y., ROSS, J. M., TAVERNARAKIS, N., DRISCOLL, M., ISHIHARA, T., KATSURA, I. & MILLER, D. M., 3RD 1999. UNC-4/UNC-37-dependent repression of motor neuron-specific genes controls synaptic choice in Caenorhabditis elegans. *Genes Dev,* 13**,** 2774-86.

WYLER, S. C., SPENCER, W. C., GREEN, N. H., ROOD, B. D., CRAWFORD, L., CRAIGE, C., GRESCH, P., MCMAHON, D. G., BECK, S. G. & DENERIS, E. 2016. Pet-1 Switches Transcriptional Targets Postnatally to Regulate Maturation of Serotonin Neuron Excitability. *J Neurosci,* 36**,** 1758-74.

YEMINI, E., JUCIKAS, T., GRUNDY, L. J., BROWN, A. E. & SCHAFER, W. R. 2013. A database of Caenorhabditis elegans behavioral phenotypes. *Nat Methods,* 10**,** 877-9.

YU, B., WANG, X., WEI, S., FU, T., DZAKAH, E. E., WAQAS, A., WALTHALL, W. W. & SHAN, G. 2017. Convergent Transcriptional Programs Regulate cAMP Levels in C. elegans GABAergic Motor Neurons. *Dev Cell,* 43**,** 212-226 e7.

ZHANG, L., WARD, J. D., CHENG, Z. & DERNBURG, A. F. 2015. The auxin-inducible degradation (AID) system enables versatile conditional protein depletion in C. elegans. *Development,* 142**,** 4374-84.

ZHONG, M., NIU, W., LU, Z. J., SAROV, M., MURRAY, J. I., JANETTE, J., RAHA, D., SHEAFFER, K. L., LAM, H. Y., PRESTON, E., SLIGHTHAM, C., HILLIER, L. W., BROCK, T., AGARWAL, A., AUERBACH, R., HYMAN, A. A., GERSTEIN, M., MANGO, S. E., KIM, S. K., WATERSTON, R. H., REINKE, V. & SNYDER, M. 2010. Genome-wide identification of binding sites defines distinct functions for Caenorhabditis elegans PHA-4/FOXA in development and environmental response. *PLoS Genet,* 6**,** e1000848.

Chapter 5 Discussion and Future Directions

5.1 Summary

In summary, this dissertation focused on the non-canonical role of Hox genes in the regulation and maintenance of *C. elegans* motor neuron terminal identity. MNs in the VNC acquire their functional terminal identity during the last steps of neuronal development and maintain their identity throughout life. My findings on Hox genes surprisingly revealed non-canonical roles during these processes. First, Hox genes were intensively studied during embryogenesis and early steps of neuronal development while here we demonstrated the functions of Hox proteins in regulating NT pathway genes and a broad range of MN terminal identity genes. Second, Hox genes are continuously expressed throughout life in MNs and we are among the first groups to establish the MN expression profile of Hox genes with single cell resolution. Third, the sustained expression of Hox genes corresponds to their functions in maintaining MN terminal identities. Fourth, Hox genes are region-specific TFs exerting conserved roles in the regulation of MN terminal identity in a region-specific manner collaborating with distinct cell-type-specific TFs (terminal selectors) in different cell types. Lastly, Hox proteins and terminal selectors are involved in a regulatory network with mutual positive/negative regulations and they together, engage in a coherent feedforward loop to safeguard MN terminal identity. These findings bring new insights into the function of Hox gene studies and advance our understanding of the establishment and maintenance of neuronal identities.

5.2 Discussion

5.2.1 A non-canonical role for Hox proteins in the last steps of MN development

Hox genes are well known to confer segmental identities to anteroposterior body patterning. Downstream targets of Hox proteins are involved in fundamental processes during early embryogenesis, such as cell migration and survival. Whether Hox genes continue to be expressed and carry out important functions during later stages of development is far less studied. Here, we found a non-canonical role for Hox genes in *C. elegans* VNC MNs reflects. We established a MN expression profile of all *C. elegans* Hox genes with single cell resolution and demonstrated their functional significance during the last steps of MN development (**Figure 1.1 and A1**). The continuous expression of Hox genes safeguards MN terminal identity features, such as NT synthesis and packaging, neuropeptides, ion channels, and receptor proteins etc.

5.2.2 MN terminal identity genes identified as novel targets of Hox proteins

Identifying and characterizing the downstream target genes of Hox proteins has always been a major challenge in the field. Through this thesis work, we have identified and characterized a number of new Hox targets, broadening the spectrum of genes regulated by Hox proteins. The summary of identified ceh-13 targets is *mig-13/CD320/LDLRAD2, unc-53/NAV1, unc-129/GDF10,* and *acr-2/AChR.* The summary of identified lin-39 targets is irx-1/TALE, oig-1/OneIG, unc-53/NAV1, cho-1/ChT, unc-17/VAChT, ace-2/AChE, del-1/ENaC, unc-77/NALCN, twk-46/TWiK, slo-2/KCNT, unc-129/TGFβ, dbl-1/BMP, flp-11/FMRF, ser-2/GPCR, srb-16/GPCR, ida-1/PTPRN, glr-4/AMPA, glr-5/AMPA, acr-2/AChR.

5.2.3 Hox proteins regulate downstream targets in MNs in a region-specific manner.

Through studying the regulation of these novel Hox targets, we found a conserved theme of spatial regulation in line with the spatial collinearity of Hox genes. These Hox downstream targets are expressed in a cell-type-specific manner while they are under the regulation of Hox proteins in a region-specific manner. Hox proteins, on the other hand, regulate a broad range of downstream targets regardless of the cell type. Hence, Hox genes function as spatial indicators for a particular terminal identity gene expressed in a specific MN subtype: More anterior Hox proteins regulate downstream gene expression in the anterior MNs of this subtype, while posterior Hox genes regulate target expression in the posterior MNs of the same subtype.

Examples could be found in *Kratsios et al., 2017* where distinct DA and VA MNs localized along the VNC are regulated via different Hox genes along the A-P axis. Mid-body expression of terminal identity genes is regulated by mid-body Hox genes *lin-39* and *mab-5,* while posterior expression of the same target is under the regulation of posterior Hox gene *egl-5*. The same logic of spatial regulation of Hox genes is also characterized in **Chapter 4** in agreement with what is published for Hox genes during early developmental stages. The conserved and well-controlled regulation of downstream targets by Hox proteins also indicates mechanisms that fine-tune and maintain distinct Hox domains during both early and late stages of development.

5.2.4 Mutual regulation of Hox genes at the transcriptional level

As discussed in **Chapter 1**, the regulation of Hox gene expression domains has been characterized in different levels (transcriptionally, post-transcriptionally and post-translationally). In MN development, only a few pieces of evidence suggest the mutual

regulation of Hox proteins. Through unpublished data (**Figure D1**), we demonstrated that the more posterior Hox genes *lin-39* and *egl-5* represses a transcription reporter of the more anterior Hox gene *che-13,* while the posterior Hox gene *egl-5* represses a transcription reporter of *lin-39* as well. Hence, it seems that the posterior prevalence mechanism of Hox proteins applies to the later stages of MN development, involving the transcriptional regulation (repression) of anterior Hox genes via posterior Hox proteins.

5.2.5 Hox proteins collaborate with terminal selectors to regulate terminal identity features of distinct MN subtypes.

The cross-regulation of Hox proteins ensures their exact functional domains, but how can the same Hox protein co-regulate terminal identity genes expressed in distinct MN subtypes? The first hypothesis we tested is collaboration with cell-type-specific TFs (usually terminal selectors). The intersectional co-regulatory mechanism of Hox proteins and UNC-3 cooperatively regulating cholinergic MN terminal identity genes among distinct MN subclasses has been enlightening, leading us to investigate if similar molecular mechanisms are adopted for the regulation of terminal identity of other MN subtypes.

For example, UNC-3 and UNC-30 are terminal selectors for the VNC cholinergic MNs and GABAergic MNs, respectively, co-regulating a large spectrum of terminal identity genes. One terminal identity gene *mig-13* is expressed in both cholinergic and GABAergic MNs and is a target of UNC-3 and UNC-30 (**Figure F1**), respectively. *mig-13* encodes an evolutionarily conserved transmembrane protein, cell-autonomously regulating the asymmetric distribution of actin cytoskeleton in the leading migratory edge. Previous research has characterized *mig-13*

under the regulation of the anterior Hox gene *ceh-13* and my unpublished data confirmed this finding. In *ceh-13* mutants, *mig-13* expression is completely compromised in both cholinergic and GABAergic MNs both at L1 and L4 stages. Post-embryonic MNs are normally generated in *ceh-13* mutants (unpublished data) but they no longer express the *mig-13* GFP reporter.

This data prompts us to hypothesize that the regulation of *mig-13* relies on the collaboration of Hox proteins with distinct terminal selectors (UNC-3 in cholinergic MNs and UNC-30 in GABAergic MNs) in different MN subtypes. Besides *ceh-13*, mid-body Hox genes *lin-39* and *mab-5* are of particular interest to test this hypothesis due to their overlapping expression with *unc-3* or *unc-30* in *mig-13*-expressing MNs.

5.2.6 Mutual regulation between Hox proteins and terminal selectors

Not only do terminal selectors collaborate with Hox proteins to regulate terminal identity genes, but there is mutual regulation between them. For example, in the VNC cholinergic MNs, mid-body Hox proteins LIN-39 and MAB-5 activate the transcription of *unc-3*, while UNC-3 in turn, represses the transcription of *lin-39* and *mab-5* (**Chapter** 3). These interactions bear great functional significance in the regulation of cholinergic MN terminal identity. Hox genes *lin-39* and *mab-5* on the one hand, undergo transcriptional autoregulation to maintain their expression in later stages of development. On the other hand, LIN-39 and MAB-5 activate *unc-3* transcription. Both will positively regulate cholinergic identity gene expression such as *cho-1/ChT* and *unc-17/VAChT* in the ACh pathway, forming a type I coherent feedforward loop. Theoretically, Hox protein levels can accumulate, if for example the turnover rate of Hox

proteins is low. The negative feedback from UNC-3 provides a balancing mechanism to control the levels of Hox proteins to further safeguard the developmental outcome of cholinergic MNs.

5.2.7 Hox genes function at the top of a transcriptional network for MN terminal identity.

Hox genes exert a later role in the last steps of MN developments by directly regulating a broad range of terminal identity genes for different MN subtypes. Hox genes can maintain their expression in post-mitotic MNs through positive autoregulation. Their precise levels of expression, on the other hand are balanced through the interplay of Hox positive autoregulation and negative UNC-3 feedback.

Controlled levels of Hox gene expression are tightly linked with MN terminal identity. Gene dosage experiments that affect Hox levels led to "mixed" MN identity. For example, in the balanced WT situation for cholinergic MNs, endogenous UNC-3 levels are significantly higher than LIN-39. As a result, repression of *lin-39* by UNC-3 dominates, and counterbalance the positive autoregulation of *lin-39*. This leads to a controlled lower level of LIN-39. However, upon *unc-3* overexpression, the balance is broken with complete tilt toward extremely low levels of LIN-39. Consequently, cholinergic MNs lose their identity: some VA and VB MNs start to express terminal identity reporter of DA and DB MNs such as *unc-129::GFP*. This is later found to involve the UNC-3 co-repressor BNC-1. In WT situation, BNC-1 is activated by LIN-39 in VA and VB MNs to antagonize the activation of *unc-129::GFP* by UNC-3. Upon *unc-3* overexpression, the balanced is disrupted and BNC-1 levels dramatically decrease, leading to a failure to repress *unc-129::GFP* expression in VA and VB MNs. As a result, *unc-129::GFP* is de-repressed in VA and VB MNs, which means the MNs now adopted a mixed identity by

partially obtaining DA and DB identity. Therefore, cholinergic MNs intra-identity is mis-regulated (unpublished data in **Figure G1**).

On the other hand, upon *unc-3* depletion or *lin-39* overexpression, the UNC-3-LIN-39 balance is tilted in the other direction with more LIN-39 and less or no UNC-3. Then, MNs lose their cholinergic identity and ectopically express terminal identity genes indicative of alternative identities such as VD-like and VC-like identity (**Chapter 4**). This way, cholinergic MNs lose their identity to obtain alternative identity features.

Another reason to place Hox proteins instead of terminal selectors like UNC-3 at the top of a transcriptional network of MN terminal identity is the biased nature of UNC-3-LIN-39 balance. UNC-3 is endogenously higher than LIN-39 and further increase of UNC-3 levels will be amplified leading to tilted balance. On the other hand, LIN-39 levels are always in check through simultaneous positive autoregulation and negative UNC-3 feedback, with fluctuations in LIN-39 levels manageable for rebalance. Ongoing research in Kratsios Lab is aiming at resolving the negative autoregulation of UNC-3 to make the balance more robust for perturbations or fluctuations.

5.2.8 Hox proteins function as terminal selectors.

Upon investigating the functions of Hox proteins in C. elegans MNs, we conclude that they are terminal selectors as well. Hox proteins are region-specific terminal selectors that dictate MN terminal identities in collaboration with cell-type-specific TFs in distinct MN subtypes. Based on our identification of a broad range of Hox downstream targets, we conclude Hox proteins are able to regulate multiple levels of terminal differentiation. First, they are able to regulate cell-

type-specific TFs that serve as terminal selector. Second, they are able to regulate co-repressors of terminal selectors. Third, they are able to activate terminal identity genes of different subtypes in a region-specific manner. Lastly, they are able to mediate terminal identity switch.

A recent article in *Nature* proposed a homeodomain TF code for all C. elegans neurons with the concept that a different combinatorial expression of Homeodomain TFs in individual neuron types dictates distinct neuronal cell fate. In VNC MNs, it might be possible to establish a similar homeodomain code based on combinatorial Hox gene expression and function. Unpublished single-cell RNA-seq data suggests a clear clustering of MN subtypes based on differential expression combinations of Hox genes. In line with these ideas and data, I wish to speculate that the expression profile of Hox genes is of crucial importance for adoption of MN terminal identity.

5.2.9 *C. elegans* Hox genes function later during development

It is interesting to note that the later role of Hox genes is mainly identified and characterized in *C. elegans*. Growing evidence in *Drosophila* and mammalian systems also support the conclusion of this thesis. In *C. elegans*, genes that cause large-scale homeotic transformation phenotypes during embryogenesis are not Hox genes. Hox genes, in *C. elegans*, behave more like region-specific selector genes that function during later development stages in various tissue. This will partially explain the survival of most Hox mutants in *C. elegans*, unlike the Hox mutant phenotypes in flies and mice. One interesting potential explanation for this is rooted in the lineage. *C. elegans animals* are not technically segmented animals and cell fate specifications happens along with the A-P polarity establishment during cell divisions within lineages. Hox

genes are later initiated in a non-lineage manner. The different periods of time when Hox genes are initiated and function and when global A-P identity formation are somehow separated. More and more evidence suggests that only after animal hatching, during larval stages, Hox genes behave more like homeotic selector genes and resemble the common ancestor of Hox genes. It's speculated that during evolution, the animal body plan became more sophisticated and neuronal control of appendages and other tissues became more demanding. Hox genes perhaps evolved to take this task and initiate theie expression earlier and earlier and become more and more indispensable undergoing stricter regulation and function with redundancy.

5.2.10 Hox genes and cancer.

The novel role of Hox proteins in later stages of neuronal development is actually not entirely brand new. It seems that the Hox genes are repetitively used multiple times during the course of development to achieve similar roles in different developmental contexts. After the establishment of primary, A-P axis, animal bodies need to generate secondary A-P axis especially in segments/appendages and especially in the nervous system. Hox genes, this time during later stages of development, can also deposit positional information to instruct patterning on finer and finer levels of development (e.g. MN subtype differentiation). A later role for Hox proteins has also been reported in carcinogenesis, another level of reuse during development, albeit abnormal. Hox gene expression is often altered in cancer and Hox genes can function as both oncogenes as well as tumor suppressor genes. Growing evidence links deregulation of Hox genes to crucial processes of oncogenesis, tumor proliferation, invasion, angiogenesis and metastasis. Aberrations of Hox gene expression have already been reported in multiple types of malignancy

such as breast cancer, lung cancer, prostate cancer and leukemia etc. To this end, understanding the role of Hox genes in later stages of development may reveal novel insights to fully grasp the spectrum of Hox functions, possibly providing therapeutic possibilities.

5.3 Limitations and Future Directions

5.3.1 Potential role of other Hox genes in MN development

The model raised above was mainly based on our investigation of mid-body Hox gene *lin-39* and *mab-5*. Whether other *C. elegans* Hox genes exert similar roles in VNC MNs is largely unknown and can be a great next step to study. There are 4 Hox genes with expression in the VNC MNs: *ceh-13, lin-39, mab-5* and *egl-5*. One of them, the anterior Hox gene *ceh-13* is of particular interest. *ceh-13* is the least studied Hox gene, partially due to the embryonic/early larval lethality phenotypes. We know little on the function of *ceh-13* during later stages of MN development. Unpublished data suggests that *ceh-13* regulates 4 terminal identity genes (*mig-13, unc-53/NAV1, unc-129/TGFβ, acr-2/AChR*) in embryonic MNs (**Figure B1**). Making use of Auxin-induced-degradation system, we were able to bypass the lethality to conditionally remove CEH-13 protein during later stages of development upon Auxin treatment and study its function. This way, our preliminary data indicates that *ceh-13* also undergoes transcriptional autoregulation to maintain its expression in later stages of larval development.

For the posterior Hox gene *egl-5,* previous studies have identified terminal identity genes as downstream targets such as *itr-1/ITPR1, glr-4/AMPA, mig-13,* and *flp-18/FMRF*. We have also established a Auxin inducible allele of *egl-5(syb2361) [egl-5::mNG::AID]* and this allele will facilitate further studies in characterization the potential autoregulatory mechanisms as well as

testing if *egl-5* is required for maintenance of posterior MN terminal identity features throughout life.

5.3.2 A discussion on the temporal resolution of our studies

A limitation of this thesis work resides on the lack of adequate temporal resolution when compared to the spatial single-cell resolution we achieved. We have illustrated, in single cell resolution, the MN expression profile of C. elegans Hox genes *ceh-13, lin-39, mab-5* and *egl-5*, but we did not precisely establish when exactly do those Hox genes initiate to express and function in MNs and whether Hox gene expression changes overtime. This temporal resolution bears great significance as well. Our model in general indicates regulatory mechanisms but doesn't specify if the mechanisms can be applied throughout life. There are at least three stages when the involved mechanisms can differ: Hox gene initiation in MNs, Hox gene regulation during MN terminal differentiation, and Hox gene maintenance throughout life of MNs. Most of our analysis focused on the last larval stage (L4) and adult stages and we are able to therefore make a strong case for the role of Hox genes in the last steps of MN development and maintenance of MN terminal identity. However, the same mechanisms might not operate during the initiation of Hox gene expression or even during the establishment of MN terminal identities. Our preliminary evidence suggests that *lin-39* and *ceh-13* initiate their expression in 2-3 fold larval stages in embryonic MNs. We don't show with solid evidence whether Hox genes are expressed in their progenitors, nor do we demonstrate the expression and function of Hox genes in the generation of post-embryonic MNs at late L1 stage. Our genetic analysis utilized null alleles of Hox genes which eliminate gene function from early on globally. We cannot

distinguish mechanisms involved between initiation versus maintenance of MN identity. More precisely, we cannot conclude with full confidence that besides the maintenance role of Hox genes on MN terminal identity, whether Hox are required to also initiae MN terminal identity.

The above limitations also point to new future directions that dive deeper in temporal resolution to firstly establish Hox gene MN expression profile temporally and then characterize their functions during different stages, possibly upon conditional knock-out/down (through the AID protein depletion system, for example). This way, we can have a complete understanding of the roles of Hox genes not only during the last steps of MN development, but also during the establishment of MN terminal identity.

5.3.3 Further steps to achieve deeper mechanistic understandings

Mechanistically, we lack thorough understandings on what happens specifically at the molecular level. Hox genes are notoriously known for their low binding affinity to DNA sequences. Bioinformatic analysis can provide us with potential Hox binding sites but we need more evidence to reveal molecular details. We performed ChIP-seq of LIN-39 in order to systematically identify downstream targets, but we only got suboptimal quality data. We plan to unbiasedly identify targets of the other Hox genes as well in VNC MNs. For that, we need to optimize protocols and perform ChIP-seq for CEH-13, MAB-5 and EGL-5.

Hox genes can bind to DNA as monomers, homodimers, as well as heterodimers, heterotrimers, or protein complexes with TALE family of co-factors. PBX and MEIS family of proteins function as Hox co-factors across species. Their orthologs are found in *C. elegans* (CEH-20/PBX, UNC-62/MEIS). Our unpublished data suggests CEH-20 is necessary for regulating

terminal identity genes under the regulation of Hox proteins such as *glr-4/AMPA, acr-2/AChR, unc-129/TGFβ, ida-1/PTPRN, ser-2/GPCR, flp-11/FMRF, glr-5/AMPA* and *ilys-4/LYSozyme*. Additionally, CEH-20 is also required for the autoregulation of *ceh-13, lin-39* and *mab-5*. Whether UNC-62 exert similar roles is unknown. What are the Hox genes that CEH-20 collaborates with to regulate MN terminal identities is also unknown.

In addition to PBX and MEIS co-factors, Hox proteins also collaborate with terminal selectors. In **Chapter 3**, we characterized the interactions between LIN-39 and UNC-3 in cholinergic MNs. Unpublished data (**Figure E1**) also points to the possibility of similar mechanisms in GABAergic MNs that involves co-regulations and even interactions between Hox proteins (possibly CEH-13 and LIN-39) with UNC-30. However, we don't know the molecular details on how Hox proteins and terminal selectors co-regulate MN terminal identity genes. Available ChIP-seq data suggests Hox proteins and terminal selectors have binding peaks in close proximity. We knew from a heterologous in vitro cell culture system that UNC-3 and LIN-39 do not physically interact, but this possibility is not firmly excluded. We do not know if one protein might indirectly recruit the other one or if the process involves epigenetic modifications. Further biochemical and genomic investigations on Hox functions in MNs can focus on the molecular details on the *cis*, *trans*, or epigenetic interactions.

5.3.4 Limitations on quantifications and other technical challenges

Through this dissertation, quantifications are done manually or semi-manually in a non-blind manner. This would certainly bring limitations. Moreover, the quantifications might not fully reflect what really occurs endogenously. For example, endogenous reporter genes can express at

an extremely low level and below certain detection thresholds at the current fluorescence microscopy conditions. Sometimes, we arbitrarily set up a threshold for our quantifications of the number of MNs expressing reporter genes in WT animals, and when expression levels are below this threshold, we quantified that as zero/no expression. When we quantified the fluorescence intensity instead to improve our resolution, we are also limited to the high noise/signal ratio and background subtraction. To bypass this, we developed a quantification method that involves manual selection of MNs with circling the outline of MN cell bodies and quantifying the fluorescence intensity through the FIJI imaging software. These quantification limitations can be overcome with automatic algorithms in the future. And the resolution of imaging can be improved through confocal microscopy.

APPENDIX A Expression profile of *C. elegans* Hox genes in ventral nerve cord motor neurons with single cell resolution

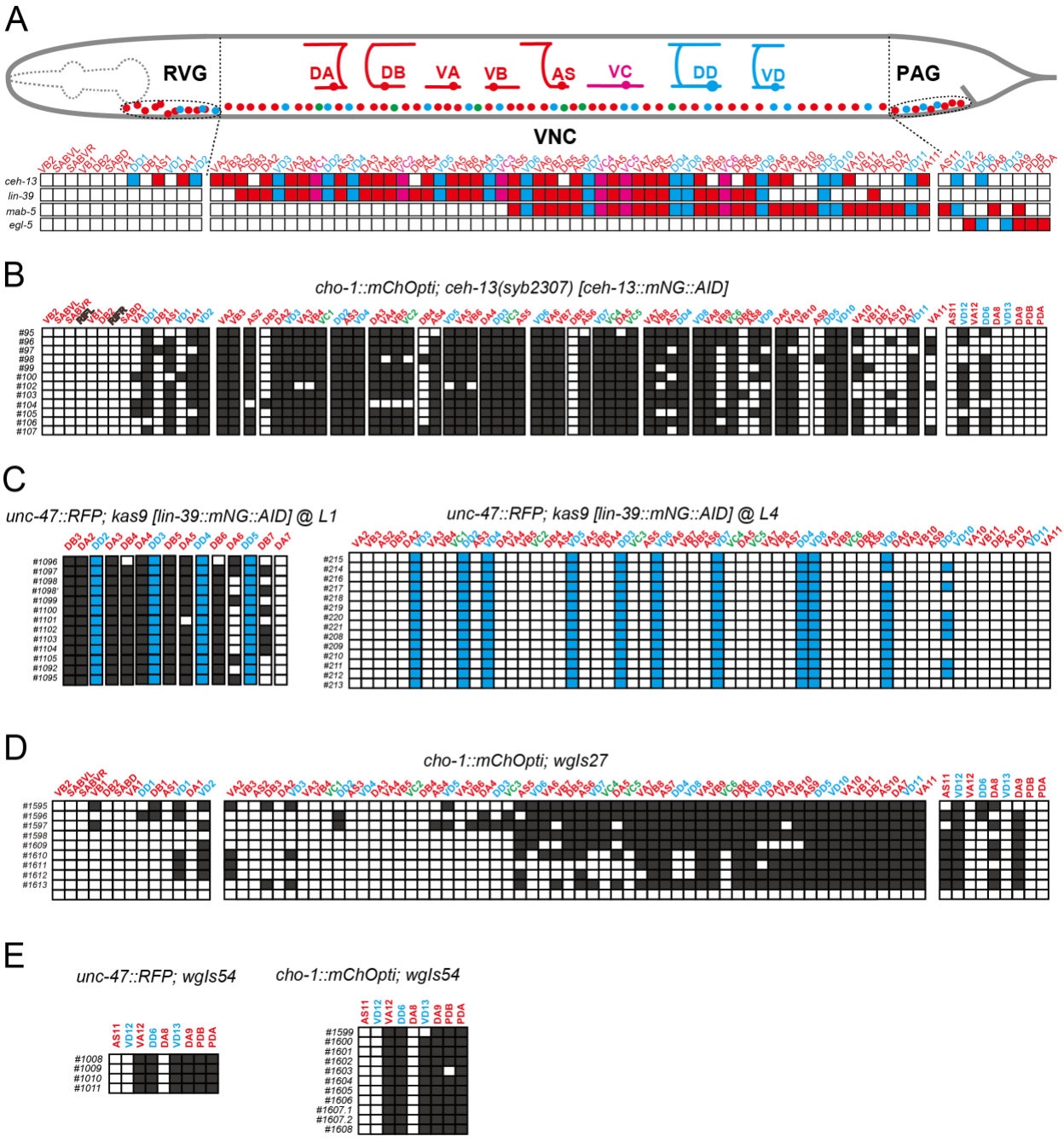

Figure A1 Expression profile of Hox genes in VNC MNs with single cell resolution. **(A)** Summary of *ceh-13, lin-39, mab-5, and egl-5* expression profile. **(B-E)** Colocalization analysis for those genes respectively

APPENDIX B A preliminary investigation of ceh-13/Lab in autoregulation and regulation of motor neuron terminal identity

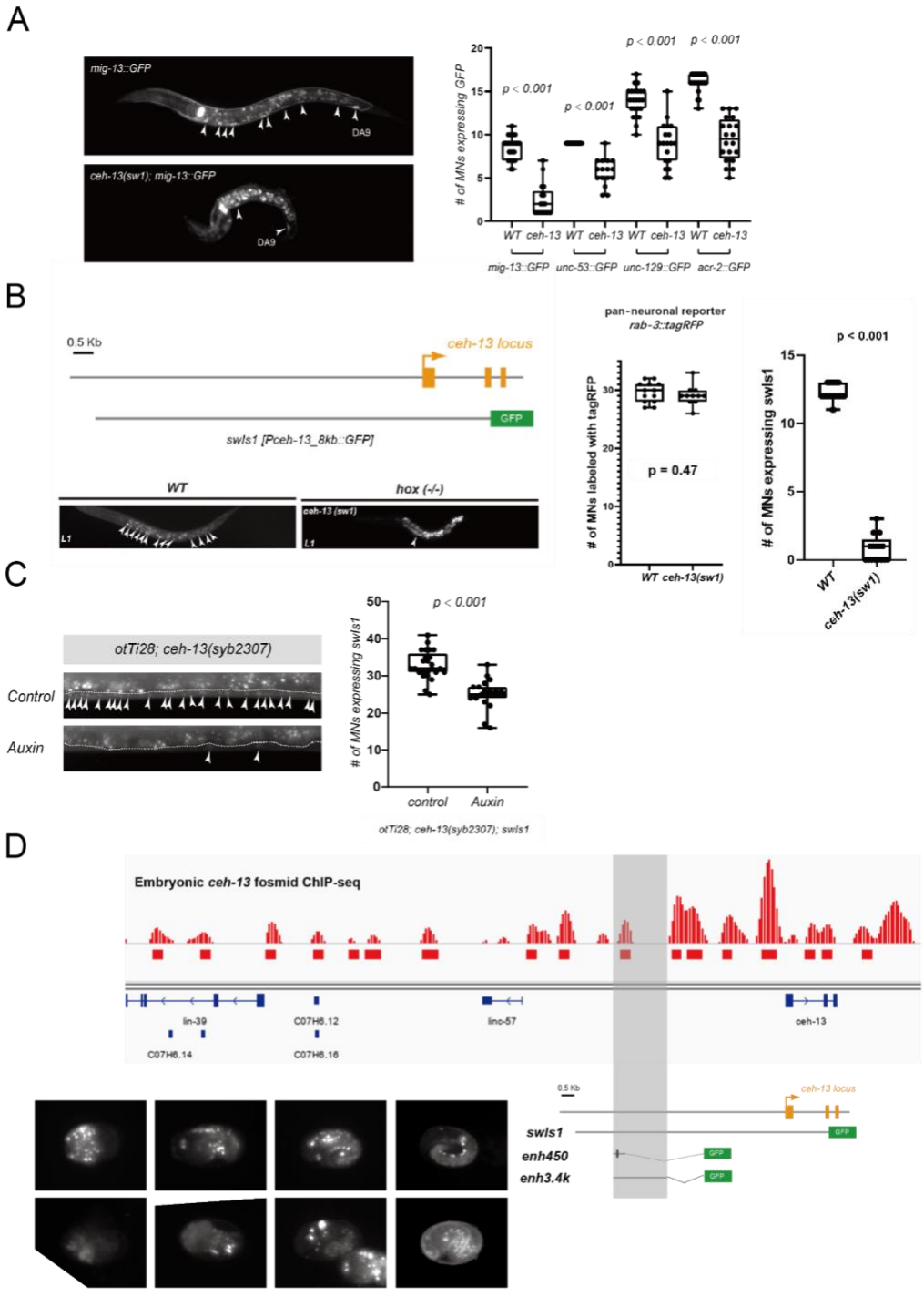

Figure B1 *ceh-13* undergoes autoregulation and regulates MN terminal identity. (**A**) Representative images of *mig-13* reporter gene *muIs42 [mig-13::GFP]* expression in WT versus

ceh-13(sw1) genetic null mutant background (Left side). *ceh-13* mutant worms were imaged only if they were arrested for growth at L1 stage. One the right side, quantifications of the number of MNs expressing reporter genes of *mig-13, unc-53, unc-129 and acr-2*. All cases show significant decrease of expression upon *ceh-13* mutants indicating *ceh-13* is required for reporter gene expression. (**B**) On the left side is the genetic locus of *ceh-13* overlay with a reporter gene of *ceh-13 swIs1 [Pceh-13_8kb::GFP]* and its expression in WT versus *ceh-13(sw1)* background. Representative images are on the left bottom and quantifications are on the rightest side. In the middle is a quantification of pan neuronal reporter *rab-3::tagRFP* expression in WT versus ceh-13 null background, indicating the total number of MNs does not decrease upon *ceh-13* null but MN numbers expressing *swIs1* decreased, which suggests for the potential autoregulation of *ceh-13*. (**C)** One the left is the representative images of *ceh-13(syb2307) [ceh-13::mNG::AID]* expression in control situation versus Auxin-treatment situation. Auxin will induce CEH-13::mNG::AID protein complex degradation. The control was done with no Auxin treatment but with ethanol treatment which was used to dissolve Auxin. On the right is the quantification of the number of MNs expressing swIs1 in worms expressing *ceh-13::mNG::AID* with pan neuronal TIR-1 protein which functions as Auxin receptor upon ethanol versus Auxin treatment. The significant decrease indicates that CEH-13 is required to maintain *ceh-13* reporter gene *swIs1* expression in MNs. Treatment was administrated at L1 or L2 stages and worms were scored after 2 days' treatment at room temperature when they reached L3 or L4 stages. (**D**) Preliminary work to identify potential *ceh-13* autoregulation site in a 450bp enhancer region. On the top is embryonic CEH-13 ChIP-seq track overlay with *ceh-13* locus with corresponding genetic natures of several reporter genes including *enh450* that was based on a 450bp enhancer region of *ceh-13*. This reporter was published to drive expression during embryogenesis. We found in details that from bean stage to comma stage to two-fold and three-fold stage embryos (representative images illustrated this on the top row from left to right). *enh450* is able to drive expression but no longer during development. Within *enh450* there is a predicted CEH-13 binding site highlighted with black filled circle on schematic. Upon mutation of this site, embryonic expression is great compromised (representative images illustrated this on the left bottom). This data demonstrated the importance of this site for autoregulation but was not sufficient to interpret if autoregulation is required for initiation or maintenance of *ceh-13* expression during embryogenesis. Nor can we interpret the mechanisms for maintaining *ceh-13* expression after three-fold stage and later on at larval stages. Other cis-regulatory regions remain to be identified in the future.

APPENDIX C A preliminary investigation of mab-5 in autoregulation and regulation of VC motor neuron terminal identity

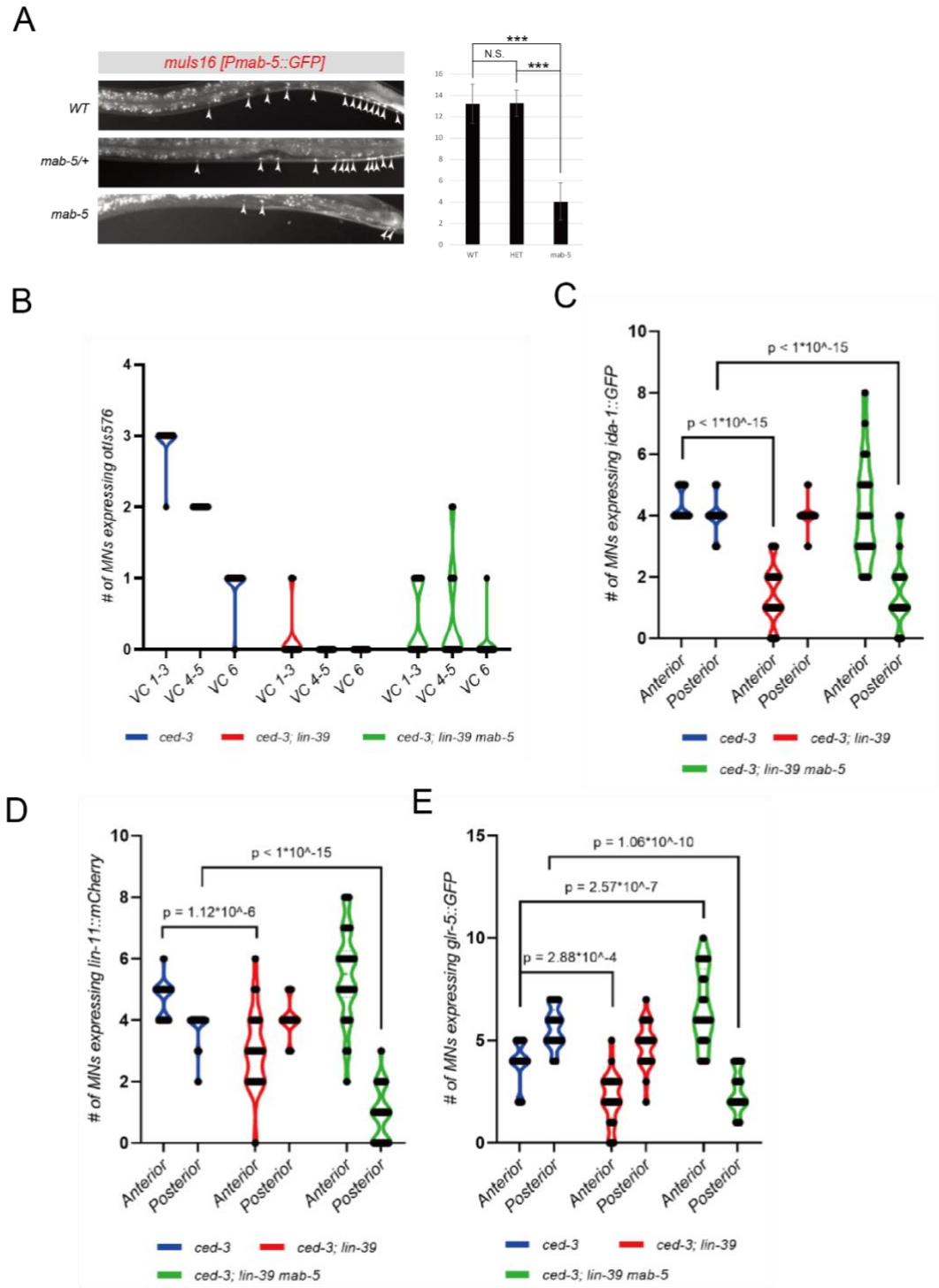

Figure C1 *mab-5* undergoes autoregulation and regulates VC MN terminal identity. (**A**) represen

tative images (left) and quantifications (right) of *mab-5* reporter gene *muIs16* expression in WT, *mab-5(ne1239)* heterozygous and homozygous background. *e1239* is a genetic *null* allele of *mab-5*. This data indicates that *mab-5* undergoes autoregulation in VNC MNs and the autoregulation is sensitive to *mab-5* levels. (**B-E**) Quantifications of VC MN terminal identity gene reporter (*otIs576 [unc-17_fosmid_GFP], ida-1::GFP; lin-11::mCherry, glr-5::GFP*) expression in *ced-3, ced-3; lin-39,* and *ced-3; lin-39 mab-5* mutant background (data shown as violin plot). In *lin-39* mutants, VC progenitors undergo apoptosis which relies on the function of CED-3 caspase. In *ced-3* mutants, VC MNs in *lin-39* animals will survive. Comparing between the number of MNs expressing VC reporters in *ced-3; lin-39* versus *ced-3; lin-39 mab-5* triple mutants in VC4-6 where *mab-5* is expressed, we wish to study the function of *mab-5* in regulating VC terminal identity. We cannot be conclusive at this time without further evidence. But it's possible that mab-5 functions as a repressor for VC identity genes. A special future direction is to study the role of *mab-5* in MN migration along the VNC as in *mab-5* mutants, MNs tend to migrate towards anterior side.

APPENDIX D A transcriptional cross-regulatory network of Hox genes in motor neurons

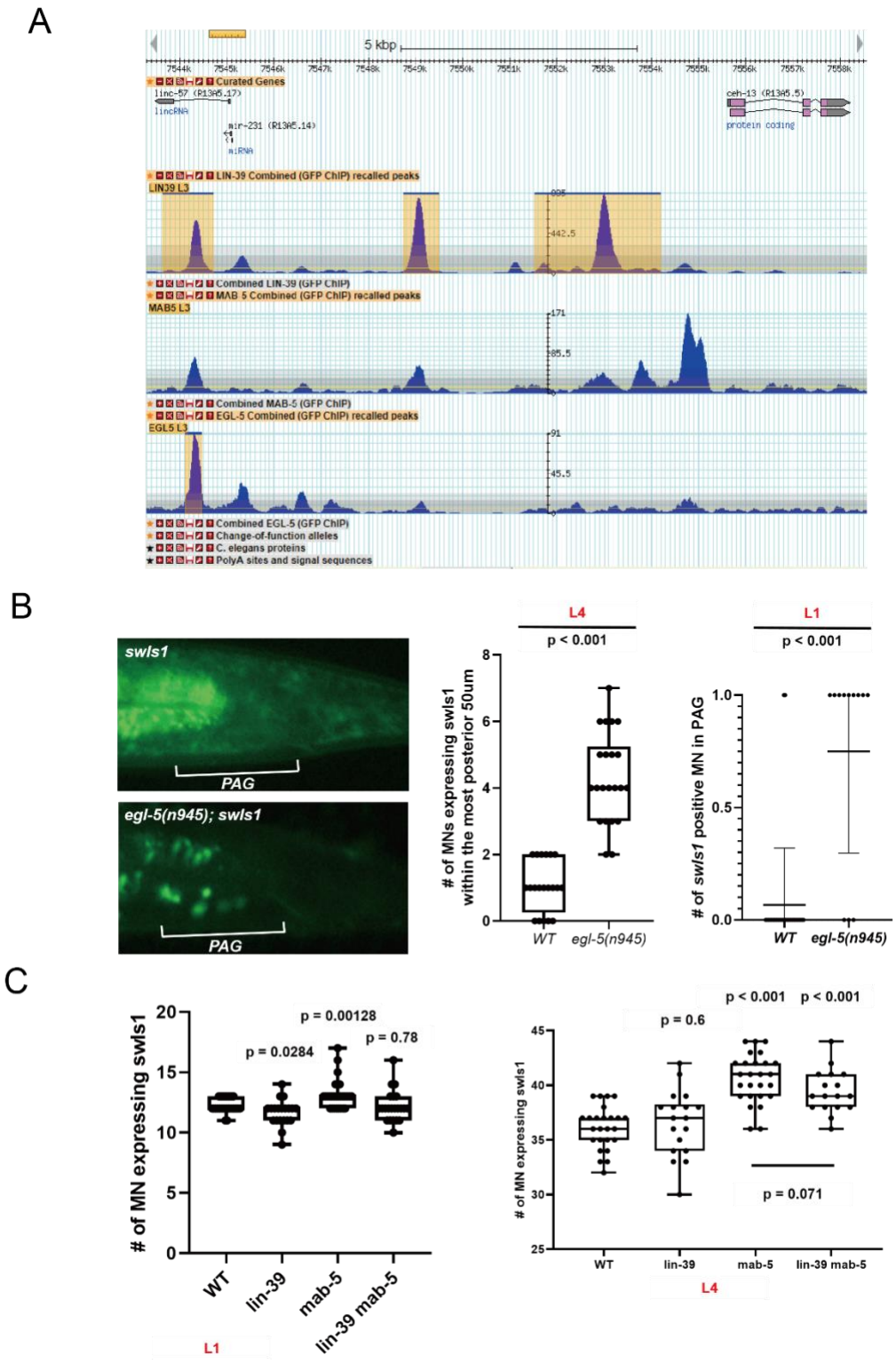

Figure D1 *ceh-13* is transcriptionally repressed by posterior Hox proteins.

(A) ChIP-seq tracks of LIN-39, MAB-5, and EGL-5 from modENCODE project overlay with *ceh-13* locus, highlighting enriched binding peaks of posterior Hox proteins onto *ceh-13* region. This is a strong evidence suggesting the crossregulation of Hox genes at the transcriptional level. (**B**) Representative images and quantifications of *ceh-13* reporter *swIs1* in PAG in WT versus genetic null mutant of *egl-5(n945)* background at L4 and L1 stages. EGL-5 is necessary for repressing *ceh-13* expression in PAG from L1 to L4 stages. (**C**) Quantifications of *swIs1* expression in *WT, lin-39, mab-5, and lin-39 mab-5* double mutant background. Phenotypes are only observed in *mab-5 and lin-39 mab-5* double mutants, indicating that mab-5 rather than *lin-39* represses *ceh-13* expression at both L1 and L4 stages.

APPENDIX E Potential role of Hox genes in regulating GABAergic MN identity

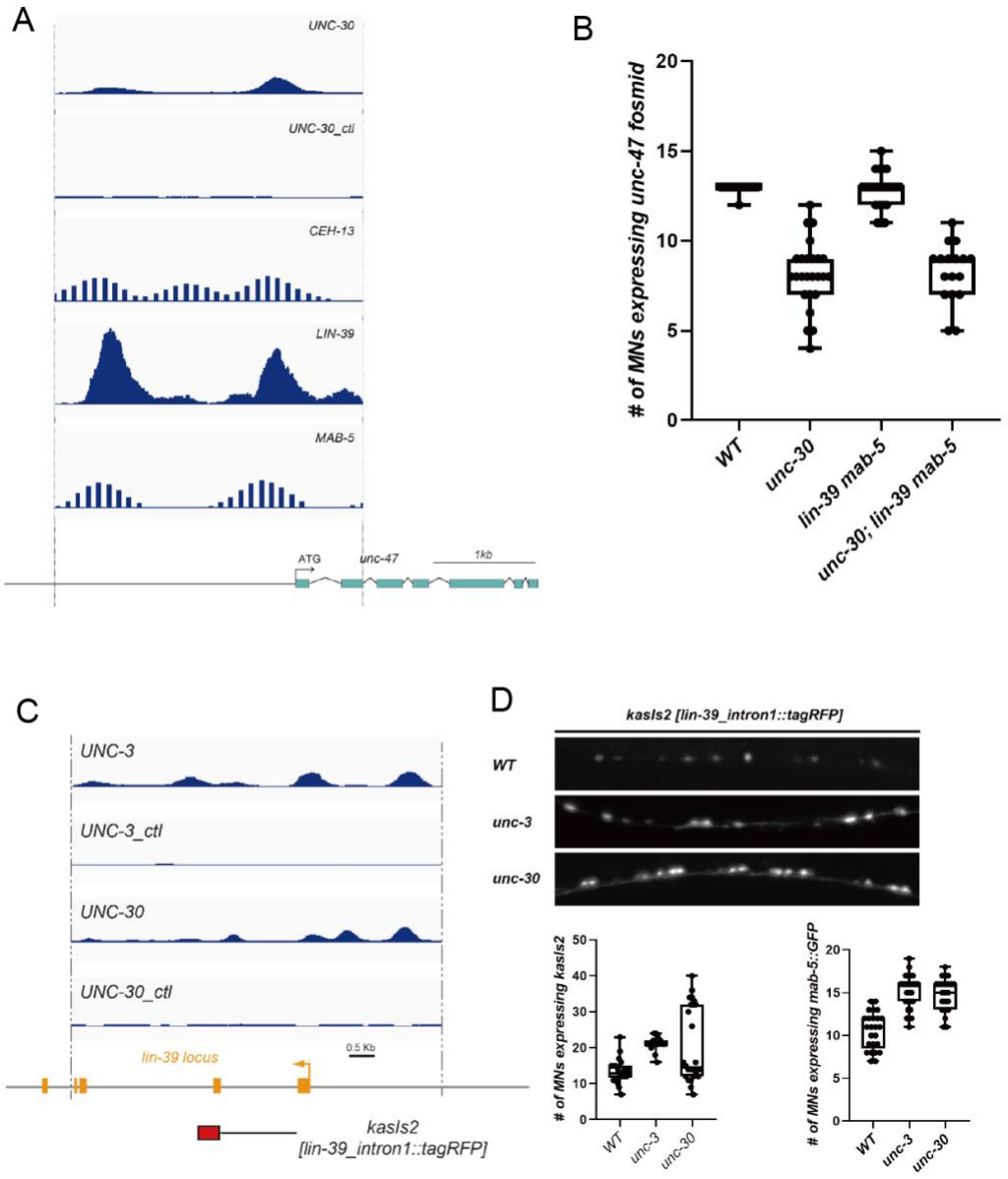

Figure E1 Potential role of Hox genes in GABAergic MNs. (**A**) ChIP-seq tracks of UNC-30, CEH-13, LIN-39, and MAB-5 from modENCODE project overlay with *unc-47/VGAT* locus. Of note, LIN-39 binding peaks are highly enriched, suggesting a potential role in regulating *unc-47 expression*. (**B**) Quantificaiton of the number of MNs expressing *unc-47_fosmid_RFP* in WT, *unc-30, lin-39 mab-5, and unc-30; lin-39 mab-5* triple mutant background. Similar to the observation in cholinergic MNs, *unc-30* mutant is not sufficient to completely abolish *unc-47* expression and *lin-39 mab-5* mutants lead to very mild effect. In triple mutant, however, the phenotype is not enhanced, different from cholinergic case. It's possible that the Hox gene regulating GABAergic identity gene expression is not *lin-39 or mab-5* but instead, *ceh-13* that shows enriched expression in GABAergic MNs. (**C-D**) Evidence suggesting both UNC-3 and UNC-30 might repress lin-39 and mab-5 expression.

APPENDIX F *mig-13*, a case study highlighting the intersection of regulation of terminal identities by different Hox genes and terminal selectors

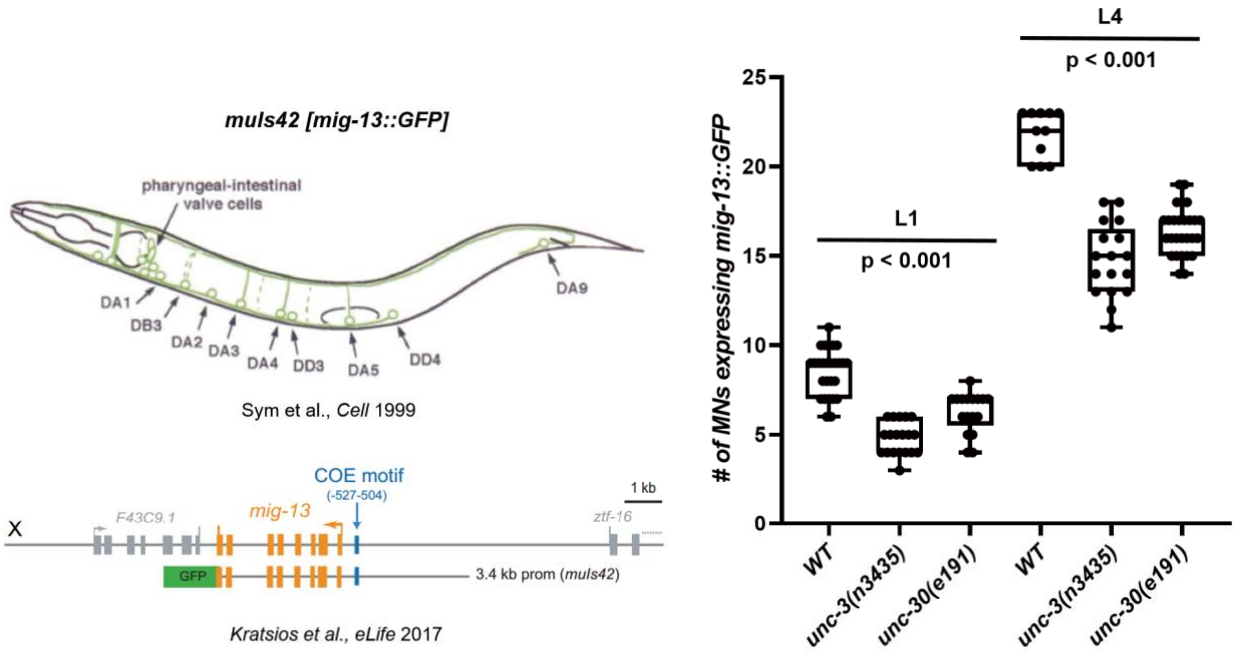

Figure F1 *mig-13* is regulated by both UNC-3 and UNC-30. Reporter gene of *mig-13 muIs42 [mig-13::GFP]* is expressed in both cholinergic and GABAergic MNs. (Figure is modified based on Sym et al., *Cell* 1999 and Kratsios et al., *eLife* 2017). On the right we demonstrated the regulation of *mig-13* expression by both *unc-3 and unc-30*. Since we have demonstrated that *mig-13* is a downstream target of *ceh-13*, it remains an intriguing question to test if *mig-13* is also regulated by *lin-39 or mab-5*. It's an interesting hypothesis that terminal identify genes like *mig-13* undergoes intersectional regulation by the collaboration of terminal selectors and Hox proteins where terminal selectors provide cell type specificity while Hox proteins provide spatial specificity. They intersect to regulate the same terminal identity gene expression in different MN subtypes within different positions on A-P axis.

APPENDIX G Levels of Hox proteins and terminal selector fine-tune terminal identity.

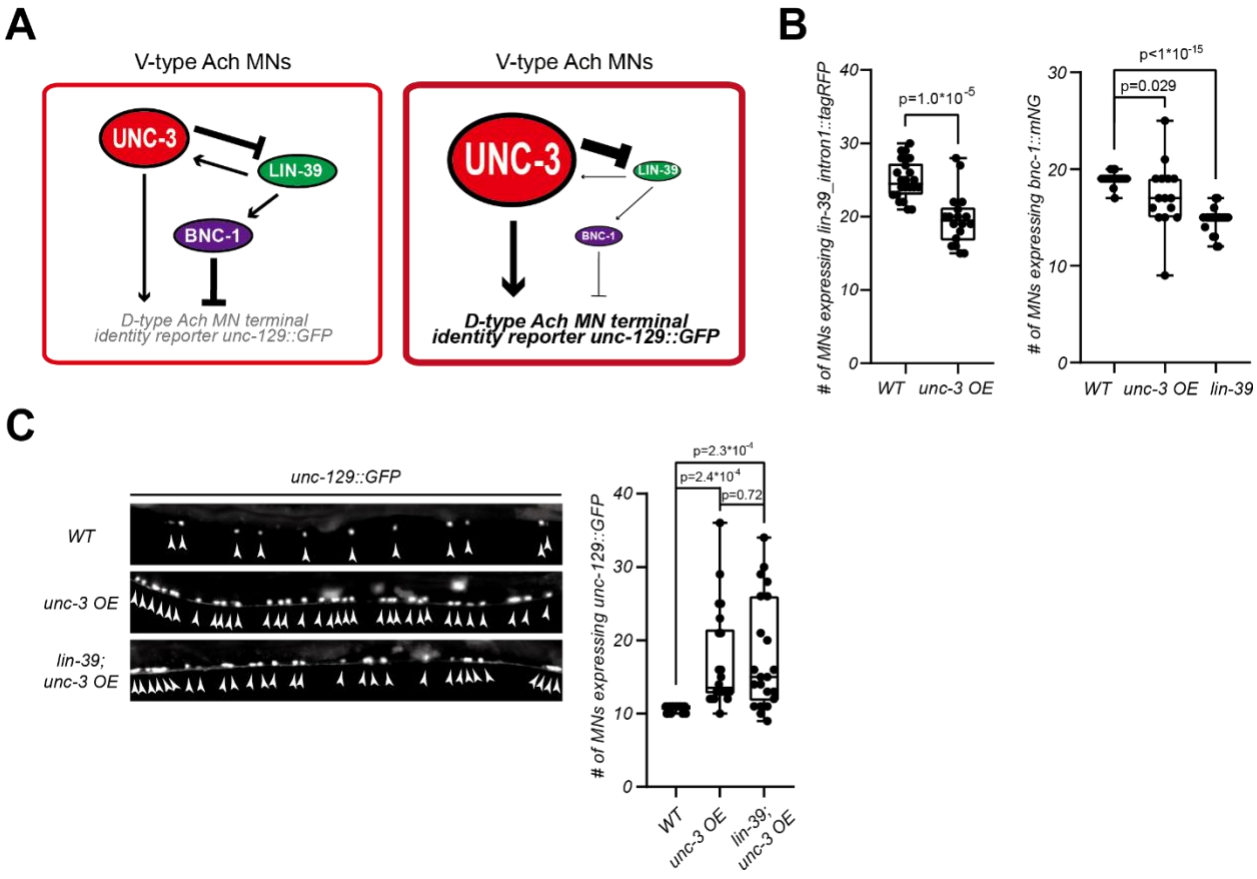

Figure G1 LIN-39 and UNC-3 levels are crucial for cholinergic MN identity. (**A**) Schematic demonstrating the molecular mechanisms of V-type MN identity regulation by the levels of UNC-3 and LIN-39. In WT scenario, the levels of LIN-39 are well controlled and downstream target BNC-1 serves as co-repressor antagonizing the functions of UNC-3 in activating D-type cholinergic identity gene *unc-129* expression. Upon *unc-3* overexpression, levels of LIN-39 permanently decrease, along with BNC-1, then UNC-3 is able to ectopically activate *unc-129* in V-type MNs to disrupt cholinergic identity. **(B)** Genetic evidence demonstrating the regulation of *lin-39 by unc-3* and the regulation of *bnc-1 by lin-39* in panel **A**. (**C**) Representative and quantifications as genetic evidence for the above arguments with analysis of *unc-129::GFP* expression.