THE UNIVERSITY OF CHICAGO

MOLECULAR MECHANISMS OF NEURONAL IDENTITY MAINTENANCE IN THE NEMATODE *CAENORHABDITIS ELEGANS*

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 NEUROBIOLOGY

BY
YINAN LI

CHICAGO, ILLINOIS
AUGUST 2022

TABLE OF CONTENTS

LIST OF FIGURES

ABBREVIATIONS

ARID	AT-rich Interaction Domain
COE	Collier/Olf/Ebf
MN	Motor Neuron
PAG	Preanal Ganglion
RVG	Retrovesicular Ganglion
TF	Transcription Factor
VNC	Ventral Nerve Cord

ACKNOWLEDGMENTS

It has been a wonderful journey learning, training, and doing research in Neuroscience. I appreciate this valuable experience for always keeping me intellectually engaged, pushing me to ask critical questions and think rigorously, and more importantly preparing me for greater challenges in my future endeavors. The last six years have been nothing short of successes and failures, and joy and frustration, yet one of the most important lessons I have learned is that perseverance is the key to achieving my goals. I hope I will never forget this lesson, as it will benefit me even for decades after graduate school.

I would like to thank my parents not only for loving and caring about me, but also for their guidance. Whenever I am in doubts of myself or don't know what to do, they are the beacon that shines the way ahead of me. As I grow up, we have made many right decisions *together*, and I truly appreciate their listening. I am so grateful for being their son, and I wish nothing but happiness and good health for them. I love my family more than anything.

The sweat from the last six years finally culminates into this work. However, it is *not* my own sweat. Nothing described here could have been achieved without my advisor Paschalis Kratsios. I would like to thank him for introducing me to *C. elegans* - amazing and adorable creatures that help advance science in ways I could not have imagined. I also appreciate the hands-on training from Paschalis and his everlasting curiosity in science that never ceases to motivate me during my pursuit in Neuroscience. I am also grateful to members on my committee for their unwavering support over the years - Edwin Ferguson and Alex Ruthenburg for giving critical opinions and motivating me to think deeper in my projects; Xiaoxi Zhuang for reminding me of the importance of jumping out of my project and thinking about the big picture; Xiaochang Zhang for valuable advice on key techniques employed in my projects as well as career advice. I feel privileged standing on the shoulders of these giants.

Last but not least, I would like to thank my co-workers in the Kratsios lab, especially Anthony Osuma, Jayson Smith, Filipe Marques, Edgar Correa, Hsin-Chiao Huang, Muna Okebalama, Pauline Dao, Olivia Gaylord, and Jihad Aburas who directly helped me in the work described here. It has been a great honor working with these talented people. I will miss the days we spent together.

ABSTRACT

The nervous system is composed of a variety of neuron types. Each distinct neuron type can be distinguished by their morphological features, physical and electrical properties, connectivity patterns, and perhaps more importantly, the expression of neuron type-specific gene batteries. Due to this diversity in neuronal identities, different neuron types may form unique and functioning neural circuits which control a wide range of activities we carry out every day. It is therefore intriguing to understand how neuronal diversity is achieved from a developmental biology perspective. Over the history of developmental neurobiology, a number of studies have been dedicated to elucidating the early steps of neuronal identity acquirement, such as subtype specification, axon pathfinding. However, the final step of neuron development which is the adoption of the stable terminal state as a fully differentiated post-mitotic neuron has been poorly investigated. Recent work on the motor neurons of the nematode Caenorhabditis elegans highlights the role of a special class of transcription factors termed terminal selectors in defining the functional properties of a terminally differentiated neurons. Mounting genetic analyses indicate that terminal selectors directly control the expression of genes tightly related to the functions of fully differentiated neurons (e.g. neurotransmitter synthesis genes, ion channels, neuropeptides etc.), which are termed terminal identity genes. Nevertheless, given that a post-mitotic neuron must remain functioning throughout the life of the animal, it is still unclear how the maintenance of terminal identity gene expression over such a long period of time is accomplished. Are there other gene regulators besides the terminal selectors involved in this long-term process? What are the mechanistical details of terminal identity gene expression maintenance in the later stages of an animal's life? Addressing these questions will not only help us get a comprehensive model of how neuronal identity remains stable, but also provide clues to the etiology of late-onset neurodegenerative diseases where neurons fail to maintain their identity and eventually die. Therefore, the central goal of this thesis is to investigate how the continuous expression of terminal identity

genes is achieved by focusing on the motor neurons of *C. elegans*.

To investigate the maintenance of terminal identity gene expression, I focus on one example which is the glutamate receptor subunit gene *glr-4*. First, I characterized the endogenous expression pattern of *glr-4*, showing it is only expressed in specific but not all cholinergic MNs. I found that a novel AT-rich interaction domain (ARID) family transcription factor CFI-1 acts as a repressor and prevents the cholinergic terminal selector UNC-3 and the mid-body Hox proteins LIN-39 and MAB-5 from activating *glr-4* in all cholinergic MNs. Using a combination of inducible protein degradation system, genetic analysis, and promoter bashing analysis, I showed that CFI-1 is continuously required to maintain the repression of *glr-4* and it directly represses *glr-4* by binding to the *glr-4* promoter via a conserved binding motif for the ARID protein family. Furthermore, I performed Chromatin Immunoprecipitation analysis followed by sequencing (ChIP-seq) for CFI-1 and revealed its genome-wide binding pattern. The genome-wide binding profile of CFI-1 displays significant overlap with that of the terminal selector UNC-3, suggesting that CFI-1 plays a critical role in co-regulating terminal identity genes with UNC-3 in late developmental stages thereby maintaining neuronal identity in adulthood.

Second, I explored the upstream regulatory mechanisms of the expression of *cfi-1*, which appears to be crucial from the results mentioned above. With promoter bashing analysis, I found that although *cfi-1* is broadly expressed in the nervous system and muscle cells in the head, a distal enhancer of *cfi-1* is sufficient to drive its expression in the MNs. Careful examination of this enhancer showed that *cfi-1* expression in the MNs is regulated by UNC-3 and LIN-39 and MAB-5. Interestingly, the Hox proteins are only required for the initiation of *cfi-1* expression during larval stages but are dispensable in adult worms. On the other hand, UNC-3 is only important for the maintenance of *cfi-1* in adulthood but is not

necessary to drive the initiation of *cfi-1* early on. This finding provides novel perspectives to the functions of the terminal selector UNC-3 (apart from its known role in directly regulating terminal identity genes), as it may indirectly contribute to the maintenance of neuronal identity by maintaining the expression of other transcription factors which then act on the terminal identity genes.

Overall, these results corroborate previous findings on the properties of terminal selectors and substantially expand our knowledge on neuronal identity maintenance by identifying a key gene regulator of the process and reporting its regulatory mechanisms. Results from this thesis will therefore make a leap towards a comprehensive understanding of how the nervous system remains stable throughout our life.

CHAPTER 1

INTRODUCTION

In this chapter, **section 1.4** is partially reprinted from Feng, Li et al., an review article published in *Frontiers in Neuroscience* in which I am a primary author. The work is included with permission from all authors.

Relevant Publication

Feng, W.*, Li, Y.* and Kratsios, P., 2021. **Emerging Roles for Hox Proteins in the Last Steps of Neuronal Development in Worms, Flies, and Mice**. *Frontiers in Neuroscience*, 15. (* Equal contribution)

1.1 The problem of neuronal identity

The definition of neuronal identity is a fundamental problem and historically the root of neuroscience. It dates to the classic work of Ramón y Cajal that culminated in what would become known as the "neuron doctrine", where he classified anatomically distinct processing units now known as neurons (Finger, 2001). Advancements in electrophysiology further elevated our understanding of the physical properties of neurons, and enabled neuronal identity classification based on their electrophysiological features and connectivity patterns (Hamill et al., 1981; Neher and Sakmann, 1976; Sakmann and Neher, 1984; Verkhratsky and Parpura, 2014). Over the last decade or two, the advent of RNA sequencing technique added an additional layer to the definition of neuronal identity, which is their molecular profiles.

The modern research on neuronal identity puts emphasis on the expression and underlying genetic regulation of neuron type-specific gene batteries. These gene batteries define the phenotypic features of diverse neuronal identities and provide basis for their functions

(Hobert et al., 2010). Mounting molecular profiling studies have significantly enriched our understanding of the diversity of neuron types (Eberwine et al., 1992; Esumi et al., 2008; Tang et al., 2009; Tietjen et al., 2003). Naturally, the problem of neuronal identity then becomes a question of how these neuron type-specific gene batteries are regulated. In this thesis, the discussion of neuronal identity focuses on the genetic regulation of the function-defining gene batteries of a neuron.

1.2 Establishment versus maintainance of neuronal terminal identity

Terminally differentiated neurons are post-mitotic, thus they no longer undergo cell division and need to remain stable and functioning throughout the life of the animal. Therefore, a comprehensive understanding of neuronal identity cannot be achieved without studying the regulation of the neuron type-specific gene batteries over the entire life of a neuron. The long-term stability in the functions of post-mitotic neurons is based upon stability in the expression of terminal identity genes (Deneris and Hobert, 2014). These terminal identity genes encode neuropeptides, neurotransmitter (NT) synthesis related genes, NT receptors, ion channels, enzymes etc. and are continuously expressed throughout the life of the neuron. While research in the past has provided insights in how initiation of terminal identity gene is established during early development, much is unknown about how maintenance of terminal identity gene expression is achieved over the full lifespan of an animal, which can last decades in many species. Characterizing the maintenance mechanisms of terminal differentiation gene expression will not only fill in the knowledge gap of neuronal development, but provide clues to what could go wrong in scenarios of late on-set neurodegenerative diseases. In this thesis, I will hone in on one terminal identity gene as an example, the glutamate receptor subunit *glr-4*, and investigate the mechanisms underlying its expression maintenance.

Terminal selectors are a special group of transcription factors and they have been found so far in worms, flies, chordates, and mice (Deneris and Hobert, 2014; Hobert and Kratsios, 2019; Konstantinides et al., 2018). They have two important properties - a) they are continuously expressed in neurons from development to adulthood; b) they control the terminally differentiate state of a neuron with the ability to regulate a large pool of terminal identity genes in a given neuron type . For instance, in the cholinergic motor neurons of the nematode *C. elegans*, the phylogenetically conserved Collier/Olf/Ebf (COE)-type transcription factor (TF) UNC-3/Ebf acts as a terminal selector. Genetic analysis revealed that UNC-3 directly controls the transcription of all known terminal identity genes in the cholinergic MNs, including *unc-17/VAChT, cha-1/ChAT, ace-2/AChE*, and *cho-1/ChT* etc. (Kratsios et al., 2011). While previous work suggested that UNC-3 is required both for the initiation and maintenance of the cholinergic terminal identity genes, it is unclear whether it shows this unified behavior towards all the target genes it regulates as a TF. In Chapter 3 of this thesis, I will explore a novel property of UNC-3, which is it is only required for the expression maintenance but not initiation of a transcription factor CFI-1. This behavior is different from its regulation of terminal identity genes and may be extended to additional transcription factor targets. Our results strongly suggest that terminal selectors may adopt different mechanisms when regulating the establishment and maintenance of neuronal identity.

1.3 *C. elegans* ventral cord motor neurons as a model to study neuronal terminal identity

To study the problem of neuronal identity maintenance, this work adopts the motor neurons (MNs) of the nematode *Caenorhabditis elegans* as the research system. This system is an excellent platform for studying neuronal identity due to 4 reasons.

First, there are 8 unique types of MNs (SAB, DA, DB, VA, VB, AS, DD, and VD) whose cell body locates in the ventral cord of the animal (Figure 1.1). These distinct neuron types can be distinguished by their i) morphology and connectivity patterns (e.g. sending axons and innervating the dorsal or ventral body wall muscles); ii) functions (e.g. controlling forward or backward locomotion); and iii) molecular profiles (e.g. SAB, DA, DB, VA, VB, and AS neurons are cholinergic and express a repertoire of genes related to the NT acetylcholine, whereas DD and VD neurons are GABAergic and express genes related to GABA). A previous study suggests that the unique molecular profiles of these MN types are achieved by a group of spatial-specifically expressed transcription factors (TFs) (Kerk et al., 2017). This rich gene regulatory network presents promising candidates for studying the maintenance aspect of terminal identity gene expression.

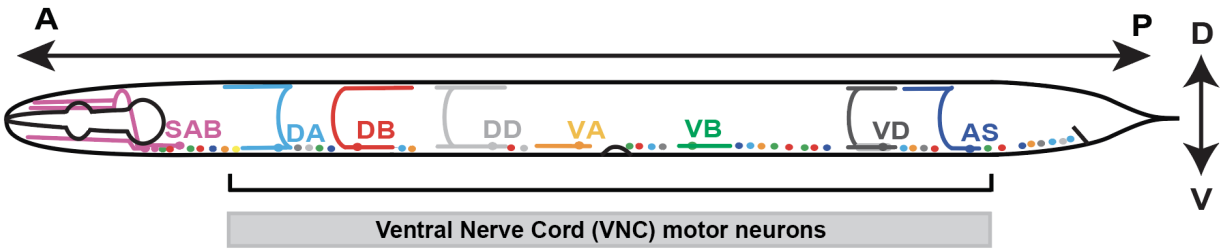


Figure 1.1: **Types of MNs in the *C. elegans* ventral cord.**
Schematic illustrating the MN system in the *C. elegans* ventral nerve cord. There are 8 types of MNs in *C. elegans*, six of which (SAB, DA, DB, VA, VB, and AS) are cholinergic while the remaining two (DD and VD) are GABAergic.

Second, the anatomical mapping of the 302 neurons in the *C. elegans* hermaphrodites is well established (White et al., 1986). Taking advantage of the remarkable wealth of neuron type-specific fluorescent reporters, it is convenient to analyze gene expression at single cell level.

Third, despite a much shorter lifespan comparing to humans, the development of *C. elegans* is well characterized and is composed of embryonic, larval (L1 – L4), and adult stages. Three of the 8 MN types are born during embryogenesis, while the rest five are born post-embryonically and become mature at the end of the L1 larval stage. Nevertheless, all 8 MN types remain functioning once they are born and throughout life, requiring active maintenance of their unique terminal identity gene batteries.

Last but not the least, *C. elegans* is an excellent model organism to perform genetic studies due to the ease of applying power genetic tools like the CRISPR/Cas9 technique (Ghanta and Mello, 2020; Jinek et al., 2012). It is also convenient to perform worm tracking for behavioral analysis. Thus, they present an ideal model to study how neuronal identity is maintained throughout life.

1.4 Hox genes control terminal identity features of ventral nerve cord MNs

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 1.2). 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-3*0 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 model 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 1.2). Similarly, UNC-3 and the posterior Hox gene *egl-5* determine posterior MN (DA9) terminal identity (Figure 1.2). 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).

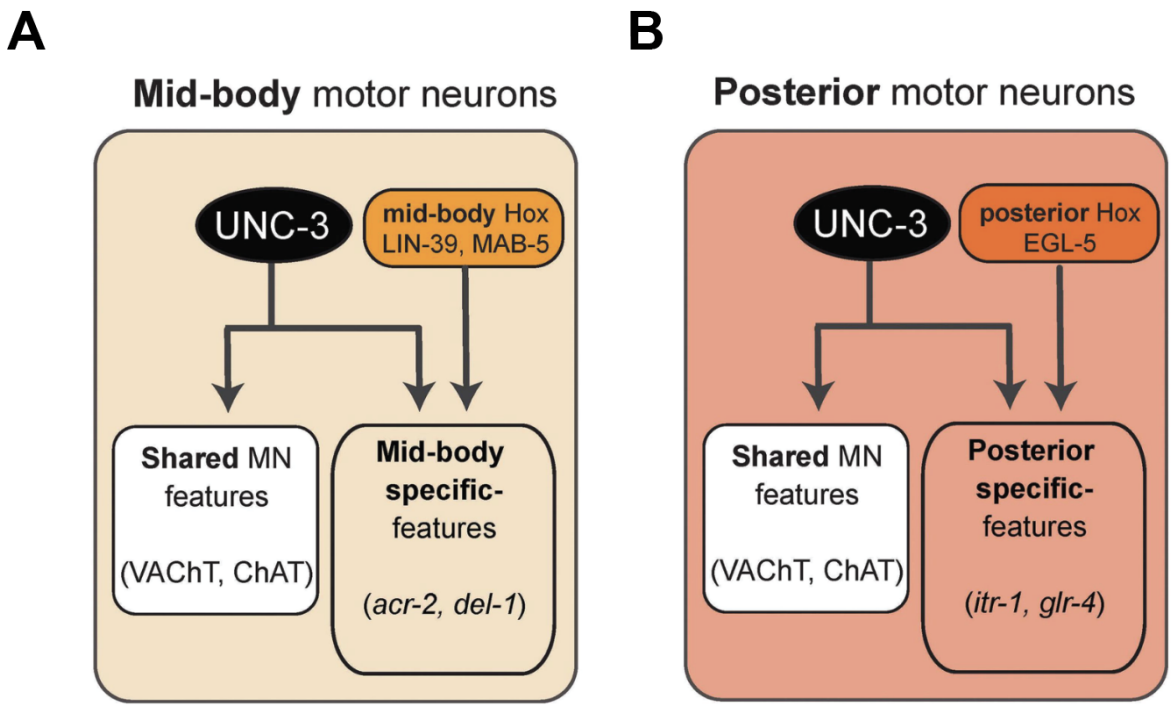


Figure 1.2: **Hox gene functions in the MNs of *C. elegans*.**
A-B, An intersectional strategy for the control of terminal identity of midbody (**A**) and posterior (**B**) MNs along the A-P axis of the *C. elegans* ventral nerve cord.

1.5 The ARID family transcription factors and their role in regulating neuronal identity

The ARID (AT-rich interaction domain) proteins are a large protein family that can be found in all eukaryotes and plays critical role in development and tissue-specific gene expression (Patsialou et al., 2005). This protein family is recognized by a highly conserved ARID domain, which is a helix-turn-helix motif-based DNA binding domain. Previous work on ARID proteins in humans, fruit flies, and yeast revealed that despite their shared ARID domain, not all ARID proteins have the same DNA-binding behavior. While the human ARID3a and *Drosophila Dri (dead ringer)* proteins bind to a consensus AT-rich motif, the human ARID1a protein which is a member of the SWI/SNF chromatin remodeling complex does not show preference for specific DNA sequences (Dallas et al., 2000; Gregory et al., 1996; Herrscher et al., 1995; Whitson et al., 1999).

The gene *cfi-1* encodes the only ARID protein in *C. elegans* (Shaham and Bargmann, 2002). As a homolog of ARID3 subfamily of ARID proteins in humans, previous work suggested that CFI-1 also shows DNA sequence preference and binds to a consensus AT-rich motif (Shaham and Bargmann, 2002). Genetic analyses suggest that CFI-1 acts as a regulator of cell identity in specific neuron types (Kerk et al., 2017; Shaham and Bargmann, 2002). A detailed review of these genetic studies will be covered in Chapter 3. It appears that CFI-1 may function as either a gene activator or repressor depending on the neuron types. Nevertheless, it is unknown whether CFI-1, potentially playing a key role in determining neuronal identity, is also involved in maintaining their terminal identity beyond early development and neurogenesis.

1.6 Aim of this study

Motor neurons (MNs) transmit information from the central nervous system directly to muscles, thus are essential for controlling movements. MNs are highly diverse in terms of their morphology, connectivity patterns and molecular profiles. Degeneration of MN subtypes has been implicated in various severe diseases, including spinal muscular atrophy (SMA) and amyotrophic lateral sclerosis (ALS). Despite much effort in elucidating the mechanisms of early MN development, the molecular mechanisms underlying the maintenance of MN diversity remain poorly understood.

To investigate the mechanisms of MN identity maintenance, this study focuses on one terminal identity gene *glr-4*, a glutamate receptor subunit in the nematode *C. elegans*, which is expressed specifically in certain but not all MNs of the VNC. The expression of *glr-4* in these MNs begins during larval stages and lasts throughout life. Previous work suggests that the conserved Collier/Olf/Ebf (COE)-type transcription factor (TF) UNC-3/Ebf is required to activate *glr-4* (Kratsios et al., 2011). Given that UNC-3 is expressed in all cholinergic MNs, an intriguing question then arises - How UNC-3 only activates *glr-4* selectively in a subset of these MNs but not in all of them? A genetic analysis suggests that the *C. elegans* ARID protein CFI-1 acts as a repressor and prevents UNC-3 from activating *glr-4* in the rest of the MNs. However, it is unknown whether CFI-1 is continuously required to maintain the repression of *glr-4*.

In this thesis, Chapter 2 will investigate the role of CFI-1 in maintaining the neuron -type specific expression pattern of *glr-4* in adult worms. Chapter 3 will dive deeper into the upstream regulation of *cfi-1*, which turns out to be key player in maintaining neuronal identity. Together, our results uncover a sophisticated gene regulatory network involving UNC-3, CFI-1, and Hox proteins that is dedicated in establishing and maintaining neuronal

identity diversity. The findings presented in this thesis may provide general mechanisms that apply to other species, hence contributing to a more comprehensive understanding of how the stability of post-mitotic neurons is achieved.

1.7 References

Dallas, P.B., Pacchione, S., Wilsker, D., Bowrin, V., Kobayashi, R., and Moran, E. (2000). The human SWI-SNF complex protein p270 is an ARID family member with non-sequence-specific DNA binding activity. Mol. Cell. Biol. 20, 3137–3146.

Deneris, E.S., and Hobert, O. (2014). Maintenance of postmitotic neuronal cell identity. Nat. Neurosci. 17, 899–907.

Eberwine, J., Yeh, H., Miyashiro, K., Cao, Y., Nair, S., Finnell, R., Zettel, M., and Coleman, P. (1992). Analysis of gene expression in single live neurons. Proc. Natl. Acad. Sci. USA 89, 3010–3014.

Esumi, S., Wu, S.-X., Yanagawa, Y., Obata, K., Sugimoto, Y., and Tamamaki, N. (2008). Method for single-cell microarray analysis and application to gene-expression profiling of GABAergic neuron progenitors. Neurosci. Res. 60, 439–451.

Feng, W., Li, Y., Dao, P., Aburas, J., Islam, P., Elbaz, B., Kolarzyk, A., Brown, A.E., and Kratsios, P. (2020). A terminal selector prevents a Hox transcriptional switch to safeguard motor neuron identity throughout life. Elife 9.

Finger, S. (2001). Origins of Neuroscience: A History of Explorations Into Brain Function (Oxford University Press).

Ghanta, K.S., and Mello, C.C. (2020). Melting dsDNA Donor Molecules Greatly Improves Precision Genome Editing in *Caenorhabditis elegans*. Genetics 216, 643–650.

Gregory, S.L., Kortschak, R.D., Kalionis, B., and Saint, R. (1996). Characterization of the dead ringer gene identifies a novel, highly conserved family of sequence-specific DNA-binding proteins. Mol. Cell. Biol. 16, 792–799.

Hamill, O.P., Marty, A., Neher, E., Sakmann, B., and Sigworth, F.J. (1981). Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches. Pflugers Arch. 391, 85–100.

Herrscher, R.F., Kaplan, M.H., Lelsz, D.L., Das, C., Scheuermann, R., and Tucker, P.W.

(1995). The immunoglobulin heavy-chain matrix-associating regions are bound by Bright: a B cell-specific trans-activator that describes a new DNA-binding protein family. Genes Dev. 9, 3067–3082.

Hobert, O., and Kratsios, P. (2019). Neuronal identity control by terminal selectors in worms, flies, and chordates. Curr. Opin. Neurobiol. 56, 97–105.

Hobert, O., Carrera, I., and Stefanakis, N. (2010). The molecular and gene regulatory signature of a neuron. Trends Neurosci. 33, 435–445.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816–821.

Kerk, S.Y., Kratsios, P., Hart, M., Mourao, R., and Hobert, O. (2017). Diversification of *C. elegans* motor neuron identity via selective effector gene repression. Neuron 93, 80–98.

Konstantinides, N., Kapuralin, K., Fadil, C., Barboza, L., Satija, R., and Desplan, C. (2018). Phenotypic convergence: distinct transcription factors regulate common terminal features. Cell 174, 622–635.e13.

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

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

Li, Y., Osuma, A., Correa, E., Okebalama, M.A., Dao, P., Gaylord, O., Aburas, J., Islam, P., Brown, A.E., and Kratsios, P. (2020). Establishment and maintenance of motor neuron identity via temporal modularity in terminal selector function. Elife 9.

Neher, E., and Sakmann, B. (1976). Single-channel currents recorded from membrane of denervated frog muscle fibres. Nature 260, 799–802.

Patsialou, A., Wilsker, D., and Moran, E. (2005). DNA-binding properties of ARID family proteins. Nucleic Acids Res. 33, 66–80.

Sakmann, B., and Neher, E. (1984). Patch clamp techniques for studying ionic channels in excitable membranes. Annu. Rev. Physiol. 46, 455–472.

Shaham, S., and Bargmann, C.I. (2002). Control of neuronal subtype identity by the *C. elegans* ARID protein CFI-1. Genes Dev. 16, 972–983.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. Nat. Methods 6, 377–382.

Tietjen, I., Rihel, J.M., Cao, Y., Koentges, G., Zakhary, L., and Dulac, C. (2003). Single-cell transcriptional analysis of neuronal progenitors. Neuron 38, 161–175.

Verkhratsky, A., and Parpura, V. (2014). History of electrophysiology and the patch clamp. Methods Mol. Biol. 1183, 1–19.

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

Whitson, R.H., Huang, T., and Itakura, K. (1999). The novel Mrf-2 DNA-binding domain recognizes a five-base core sequence through major and minor-groove contacts. Biochem. Biophys. Res. Commun. 258, 326–331.

CHAPTER 2

CONTROL OF NEURONAL TERMINAL DIFFERENTIATION THROUGH CELL CONTEXT-DEPENDENT CFI-1/ARID3 FUNCTIONS

This Chapter is a full reprint of Li et al., *bioRxiv*, in which I am the primary author. The work is included with permission from all authors. To be noted, the manuscript here has not yet been formally peer-reviewed and may undergo revision.

Relevant Publication

Yinan Li, Jayson J. Smith, Filipe Marques, Anthony Osuma, Hsin-Chiao Huang, Paschalis Kratsios. (2022) **Control of neuronal terminal differentiation through cell context-dependent CFI-1/ARID3 functions.** *bioRxiv*; doi: `https://doi.org/10.1101/2022.07.04.498728`

2.1 Abstract

ARID3 transcription factors are expressed in the nervous system, but their functions and mechanisms of action are largely unknown. Here, we generated *in vivo* a genome-wide binding map for CFI-1, the sole *C. elegans* ARID3 ortholog. We identified 6,396 protein-coding genes as putative direct targets of CFI-1, most of which (77%) are expressed in post-mitotic neurons and encode terminal differentiation markers (e.g., neurotransmitter receptors, ion channels, neuropeptides). To gain mechanistic insights, we focused on two neuron types. In sensory neurons (IL2 class), CFI-1 exerts a dual role: it acts directly to activate, and indirectly to repress, distinct terminal differentiation genes. In motor neurons, however, CFI-1 acts directly as a repressor, continuously antagonizing three transcriptional activators (UNC-3/Ebf, LIN-39/Hox4-5, MAB-5/Hox6-8). By focusing on a glutamate receptor gene

(*glr-4/GRIK1*), we found CFI-1 exerts its repressive activity through proximal binding to the *glr-4* locus. Further, the core DNA binding domain of CFI-1 is partially required for *glr-4* repression in motor neurons. Altogether, this study uncovers cell context-dependent mechanisms through which a single ARID3 protein controls the terminal differentiation of distinct neuron types.

2.2 Introduction

Members of the ARID family of proteins are found in plants, yeast, fungi, and invertebrate and vertebrate animals (Kortschak et al., 2000, Patsialou et al., 2005, Wilsker et al., 2002, Wilsker et al., 2005). ARID family proteins are expressed either ubiquitously or in a tissue-specific fashion and control various biological processes, such as cell proliferation, differentiation, and embryonic patterning (Wilsker et al., 2002, Wilsker et al., 2005). Additionally, mutations in ARID family proteins are associated with cancer and several neurodevelopmental disorders (Shang et al., 2015, Bramswig et al., 2017, Kosho et al., 2014, Miyake et al., 2014, Smith et al., 2016, Lin et al., 2014).

Humans possess fifteen ARID family proteins, divided into seven subfamilies (ARID1-5, JARID1-2) based on the degree of sequence similarity (Wilsker et al., 2005). The AT-Rich Interaction Domain (ARID), after which the family is named, was first identified in ARID3 proteins. These bind DNA in a sequence-specific manner, prefer AT-rich sequences, and are known to function as transcription factors (Wilsker et al., 2005). The ARID5 subfamily also encodes transcription factors (Patsialou et al., 2005, Wilsker et al., 2002), but the remaining five subfamilies (ARID1-2, ARID4, JARID1-2) encode proteins that bind DNA in a non-sequence-specific manner (Patsialou et al., 2005). For example, ARID1A, ARID1B, and ARID2 constitute subunits of the SWI/SNF (BAF/PBAF) chromatin-remodeling complex

that can move and/or eject nucleosomes. Although the precise functions of all ARID proteins are not known, accumulating evidence suggests they can act either as positive or negative regulators of gene transcription, or as components of chromatin-remodeling complexes (Kortschak et al., 2000, Patsialou et al., 2005, Wilsker et al., 2002, Wilsker et al., 2005).

Mouse *Bright (Arid3a)* and *Drosophila dead ringer (retained)* are the founding members of the ARID family, and belong to the ARID3 subfamily, which is specific to metazoans. Single orthologs exist in *C. elegans* (CFI-1) and *Drosophila* (*dead ringer*), whereas mammals contain three orthologs (ARID3A-C) (Fig. 1A) (Kortschak et al., 2000, Wilsker et al., 2005, Gregory et al., 1996, Shaham and Bargmann, 2002). A defining feature of ARID3 proteins is the extended ARID (eARID) domain, a ∼40 residue-long region next to the core ARID domain (Wilsker et al., 2002, Wilsker et al., 2005) (Kortschak et al., 2000). Structural studies showed both ARID and eARID domains contact DNA (Iwahara and Clubb, 1999, Iwahara et al., 2002).

ARID3 proteins have several early developmental roles, as identified by genetic studies. Mice lacking *Bright/Arid3a* display early embryonic lethality due to defects in hematopoiesis (Webb et al., 2011). *Bright/Arid3a* is best studied in B cell lineages, where it acts as an activator and increases immunoglobulin transcription(Herrscher et al., 1995, Ratliff et al., 2014, Webb et al., 2011, Webb et al., 1998). However, *Bright/Arid3a* is also critical for embryonic stem cell differentiation (An et al., 2010, Popowski et al., 2014, Rhee et al., 2014). In this context, it can act either as an activator or repressor of gene expression (Rhee et al., 2014). Similar to mice lacking *Bright/Arid3a*, null mutants for dead ringer in *Drosophila* display early lethality (Shandala et al., 1999, Shandala et al., 2002). Dead ringer is essential for anterior-posterior patterning and muscle development in the fly embryo, and can act either as an activator or repressor of gene transcription (Hader et al., 2000, Shandala et al., 1999,

Valentine et al., 1998). Lastly, Arid3a and Arid3b have been associated with tumorigenesis by acting as direct inducers of cell cycle regulators (Lestari et al., 2012, Saadat et al., 2021).

Genetic studies in *Drosophila* and *C. elegans* have also identified late developmental roles of ARID3 proteins in the nervous system. In *Drosophila* larvae, *dead ringer* is expressed in distinct neuron types and controls axonal pathfinding (Ditch et al., 2005, Shandala et al., 2003, Sibbons, 2004), though its downstream targets in neurons - and thus, whether dead ringer behaves as an activator or repressor – remain unknown. In *C. elegans*, *cfi-1* is selectively expressed in head muscle and several neuron types: the IL2 sensory neurons, the AVD, PVC, and LUA interneurons, and various classes of motor neurons (Shaham and Bargmann, 2002, Li et al., 2020). Because *C. elegans* animals lacking *cfi-1* are viable (Shaham and Bargmann, 2002), these mutant strains provided a glance into the potential functions of ARID3 proteins during post-embryonic life. Candidate approaches that examined a handful of effector genes encoding neurotransmitter (NT) biosynthesis proteins and receptors suggested that CFI-1 acts as an activator of gene expression both in sensory neurons (IL2) and interneurons (AVD, PVC) (Shaham and Bargmann, 2002, Zhang et al., 2014, Ahn et al., 2022). However, in ventral nerve cord motor neurons, CFI-1 is thought to act as a repressor of the glutamate receptor gene *glr-4/GRIK1* (Kerk et al., 2017). The molecular mechanisms underlying the differential activities of CFI-1 (activator versus repressor) in these distinct neuron types remain unknown. Elucidating such mechanisms in *C. elegans* may provide clues as to how CFI-1 orthologs in other species control cell differentiation. Lastly, unbiased approaches to identify the in vivo targets of CFI-1 (and any other ARID3 protein) in the nervous system are currently lacking, preventing a comprehensive understanding of the neuronal functions controlled by ARID3 transcription factors.

Here, we performed chromatin immunoprecipitation for CFI-1 followed by sequencing

(ChIP-Seq). By generating an in vivo binding map on the *C. elegans* genome, we identified 6,396 protein-coding genes as putative direct targets of CFI-1, the majority of which (77%) are expressed in post-mitotic neurons. Gene ontology analysis suggests CFI-1 is primarily involved in the process of neuronal terminal differentiation. To gain mechanistic insight into how CFI-1 controls the terminal differentiation of different neuron types, we focused on head sensory neurons (IL2 class) and nerve cord motor neurons (DA, DB, VA, and VB classes). In sensory IL2 neurons, CFI-1 exerts a dual role: it acts directly to activate and indirectly to repress distinct terminal differentiation genes (e.g., NT receptors, ion channels). In nerve cord motor neurons, however, CFI-1 acts directly to repress expression of the glutamate receptor gene *glr-4/GRIK1*. CRISPR/Cas9-mediated mutagenesis of endogenous CFI-1 binding sites suggests proximal binding to the *glr-4* locus is necessary for repression, advancing our understanding of ARID3-mediated gene repression. Importantly, the core DNA binding domain of CFI-1 is partially required for *glr-4* repression in motor neurons. Altogether, this study offers mechanistic insights into cell context-dependent functions of CFI-1 (ARID3), a critical regulator of the terminal differentiation program of distinct neuron types.

2.3 Materials and Methods

2.3.1 C. elegans strain culture

Worms were grown at 20℃ or 25℃ on nematode growth media (NGM) plates supplied with *E. coli* OP50 as food source (Brenner, 1974).

2.3.2 Generation of transgenic animals carrying transcriptional fusion reporters and overexpression or rescue constructs

Reporter gene fusions for *cis*-regulatory analyses of *glr-4* were made with PCR fusion (Hobert, 2002). Genomic regions were amplified and fused to the coding sequence of tagrfp followed by the *unc-54* 3' UTR. PCR fusion DNA fragments were injected into young adult *pha-1(e2123)* hermaphrodites at 50 ng/µl together with *pha-1* (pBX plasmid) as co-injection marker (50 ng/µl). To generate animals with *cfi-1* overexpression in the SAB neurons, the unc-4 promoter was fused to the cDNA sequence of *cfi-1* followed by the *unc-54* 3' UTR. The fluorescent co-injection marker *myo-2::gfp* was used (2 ng/µl) and the PCR fusion DNA fragments were injected into young adult animals carrying the *glr-4::tagrfp* reporter at 50 ng/µl. To generate transgenic animals carrying different versions of the *cfi-1* cDNA rescue constructs (WT, ΔARID, ΔeARID, ΔHTH), the *cfi-1* enhancer driving expression in motor neurons was fused to the corresponding version of *cfi-1* cDNA followed by the *unc-54* 3' UTR. The fluorescent co-injection marker *myo-2::gfp* was used (2 ng/µl) and the PCR fusion DNA fragments were injected into young adults of *cfi-1(-)* mutants carrying the *glr-4::tagrfp* reporter at 50 ng/µl.

2.3.3 Targeted genome editing

The endogenous *glr-4* reporter allele *syb3680 [2×NLS::mScarlet::glr-4]* was generated by SunyBiotech via CRISPR/Cas9 genome editing by inserting the *2×NLS::mScarlet* cassette immediately after the ATG of *glr-4*. Moreover, the endogenous *glr-4* reporter allele *syb5348 [2×NLS::mScarlet::SL2::glr-4*$^{11\ CFI\text{-}1\ sites\ MUT}$*]* that carries nucleotide substitutions in eleven CFI-1 binding sites was also generated by SunyBiotech. The endogenous *glr-4* reporter allele *kas29 [2×NLS::mScarlet::SL2::glr-4*$^{COE1\ MUT}$*]* that carries nucleotide substitutions in a single UNC-3 binding site (COE1 motif) was generated in the Kratsios lab by using homology dependent repair and inserting a synthesized DNA fragment that carries the desired mutations.

2.3.4 Microscopy

Imaging slides were prepared by anesthetizing worms with sodium azide (NaN_3, 100 mM) and mounting them on a 4% agarose pad on glass slides. Images were taken with an automated fluorescence microscope (Zeiss, Axio Imager Z2). Images containing several z stacks (0.50 µm intervals between stacks) were 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 2-5 µm Z-stacks using the maximum intensity projection type. Image reconstruction was performed with Image J (Schindelin et al., 2012).

2.3.5 Motor neuron subtype identification

Motor neuron subtypes were identified based on combinations of the following factors: [1] co-localization with or exclusion from additional reporter transgene with known expression patterns; [2] Invariant position of neuronal cell bodies along the ventral nerve cord, [3] Birth

order of specific motor neuron subtypes (e.g., during embryonic or post-embryonic stages); [4] Total cell numbers in each motor neuron subtype.

2.3.6 Bioinformatic prediction of binding motifs

Information of the CFI-1 binding motif is curated in the Catalog of Inferred Sequence Binding Preferences database (`http://cisbp.ccbr.utoronto.ca`). To predict and identify CFI-1 binding motifs in the *glr-4* promoter, we utilized tools provided by MEME (Multiple Expectation maximization for Motif Elicitation) bioinformatics suite (`http://meme-suite.org/`), and performed FIMO (Find Individual Motif Occurrences) motif scanning analysis.

2.3.7 Chromatin Immunoprecipitation (ChIP)

ChIP assay was performed as previously described, with the following modifications (Yu et al. 2017; Zhong et al. 2010). Synchronized L1 *cfi-1(syb1778[3xFLAG::cfi-1])* worms and N2 worms were cultured on 10 cm plates seeded with OP50 at 20℃ overnight. Late L2 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-Lauroyl sarsosine sodium, 2 tablets of cOmplete ULTRA Protease Inhibitor Cocktail [Roche Cat.# 05892970001] in 10ml FA buffer). For each IP experiment, 200 µl worm pellet was collected. 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 1 min. 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℃. The remainder was incubated with FLAG antibody at 4℃ overnight. Wild-type (N2) worms do not carry the GFP tag and serve as negative control. The *cfi-1(syb1778[3xFLAG::cfi-1])* CRIPSR gener-

ated allele was used in order to immunoprecipitate the endogenous CFI-1 protein. On the next day, 20 ul Dynabeads Protein G (1004D) was added to the immunocomplex which was then incubated for 2 hr at 4℃. The beads then 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℃ 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℃ for 2 hours then 65℃ overnight (12-20 hours) to reverse cross-link. The immonuprecipitated DNA was purified with Ampure XP beads (A63881) according to manufacturer's instructions.

2.3.8 ChIP-Seq data analysis

Unique reads were mapped to the *C. elegans* genome (ce10) with bowtie2 (Langmead and Salzberg 2012). Peak calling was then performed with MACS2 (minimum q-value cutoff for peak detection: 0.005) (Zhang et al. 2008). For visualization purposes, the sequencing depth was normalized to 1x genome coverage using bamCoverage provided by deepTools (Ramírez et al. 2016) and peak signals were shown in Integrated Genome Viewer (Siponen et al.). Heatmap of peak coverage in regard to CFI-1 enrichment center was generated with NGSplot (Shen et al. 2014). The average profile of peaks binding to TSS region was generated with ChIPseeker (Yu et al. 2015).

2.3.9 Enrichment of CFI-1 targets in IL2 expressed genes

The top 1,000 highest expressed genes in IL2 (transcripts per million, tpm) were mined from available single-cell RNA-sequencing data (CeNGEN). This dataset was computationally compared to a dataset of CFI-1 ChIP-Seq targets using the 'semi_join' function in R (package Dplyr 1.0.7). This generated a new data frame containing genes in the scRNA-seq dataset that are also putatively bound by CFI-1. Similarly, the 'set_diff' function (Dplyr 1.0.7) was used to generate a new data frame containing genes that are expressed in IL2 based on scRNA-seq but which are not found in the CFI-1 ChIP-seq dataset. Gene list analysis (PANTHER 17.0) was performed on both data frames to functionally classify all genes based on protein class ontology.

2.3.10 Temporally controlled protein degradation

Temporally controlled protein degradation was achieved with the auxin-inducible degradation system (Zhang et al., 2015). TIR1 expression was driven by the pan-neuronal promoter in the transgene *otTi28[unc-11prom8+ehs-1prom7+rgef-1prom2::TIR1::mTurquoise2::unc-54 3'UTR]*. To induce degradation of CFI-1 proteins, we used the allele *kas16[cfi-1::mNG::AID]*. Worms at the L4 stage were grown at 20℃ on NGM plates coated with 4 nM auxin (indole-3-acetic acid [IAA] dissolved in ethanol) or ethanol (negative control) for 2 days before testing (see figure legends for exact time in specific experiments). All plates were shielded from light.

2.3.11 Single molecule RNA fluorescent in situ hybridization (smRNA FISH)

Synchronized L1 worms were collected from the plates and washed with M9 buffer 3 times. Worms were incubated in the fixation buffer (3.7% formaldehyde in 1x PBS) for 45 minutes at room temperature. Worms were then washed twice with 1x PBS, resuspended in 70% ethanol and left at 4℃ for two nights. After removing the ethanol, worms were incubated in the wash buffer (10% formamide in 2x SSC buffer) for 5 minutes and the wash buffer was removed afterwards. A *glr-4* probe was designed using the Stellaris Probe Designer website (Biosearch Technologies). The probe was mixed in hybridization buffer (0.1 g/ml dextran sulfate [Sigma D8906-50G], 1 mg/ml Escherichia coli tRNA [ROCHE 10109541001], 2 mM vanadyl ribonucleotide complex [New England Biolabs S1402s], 0.2 mg/ml RNase-free BSA [Ambion AM2618], 10% formamide) and added to the worms. The hybridization buffer was removed and worms were washed twice in wash buffer (DAPI was added during the second wash and incubated for 30 minutes in the dark for nuclear counterstaining). Worms were washed once in 2x SSC, incubated in GLOX buffer (0.4% glucose, 0.1 M Tris-HCl, 2x SSC) for 2 minutes for equilibration, and the resuspended in GLOX buffer with glucose oxidase and catalase added. The samples were then examined under the fluorescent microscope.

2.3.12 Real-time PCR assay for glr-4 expression level analysis

Synchronized L4 stage wildtype and *cfi-1(-)* worms were collected, and mRNA was extracted. cDNA library was prepared using the Superscript first strand cDNA synthesis kit (Invitrogen #11904-018). RT-PCR TaqMan assays for the genes *glr-4* (assay ID: Ce02435302_g1) and *pmp-3* (Ce02485188_m1) were performed, and the expression level of *glr-4* was determined in each genotype after normalizing to the expression of the housekeeping gene *pmp-3*.

2.3.13 Harsh touch behavioral assay

Harsh touch was delivered with a platinum wire pick as previously described (Li et al 2001, Marques et al 2019). The stimulus was applied from above the animals by pressing down with the edge of the pick on the tail of non-moving adult animals. Each animal was tested only once, and we scored as normal response worms that moved forward. Results are presented as fractions of animals that were responding normally.

2.3.14 Statistical analysis

For data quantification, graphs show values expressed as mean ± standard error mean (SEM) of animals. The statistical analyses were performed using the unpaired t-test (two-tailed). Calculations were performed using the GraphPad QuickCalcs online software (`http://www.graphpad.com/quickcalcs/`). Differences with $p < 0.05$ were considered significant.

Data Availability

The accession number for the CFI-1 ChIP-Seq data is GEO: GSE205628.

Author contributions

Y. L., Conceptualization, Data curation, Investigation, Visualization, Methodology, Writing— review and editing; J.J.S. Formal analysis, Validation, Investigation, Writing—review and editing; F.M., A.O., H.C.H., Formal analysis, Validation, Investigation; P. K., Conceptualization, Supervision, Investigation, Funding acquisition, Project administration, Writing—original draft, review and editing.

Competing Interests

The authors declare no competing interests.

2.4 Results

2.4.1 A map of CFI-1/ARID3 binding on the C. elegans genome

To identify CFI-1 binding events, we first generated an endogenous reporter strain through in-frame insertion of the flag epitope sequence (*3×flag*) immediately after the *cfi-1* start codon (Figure. 2.1B). Immunostaining against FLAG on adult *3×flag::cfi-1* animals showed nuclear expression in head muscle cells, as well as in neurons of the head, ventral nerve cord, and tail regions (Figure. 2.1C-E), indicating this reporter allele faithfully recapitulates the known expression pattern of *cfi-1* (Kerk et al., 2017, Shaham and Bargmann, 2002). Unlike *cfi-1* null animals (Shaham and Bargmann, 2002), homozygous *3×flag::cfi-1* animals do not display any defects in posterior touch response (Figure. 2.2). This suggests that insertion of the *3×flag* sequence does not alter *cfi-1* gene function. We therefore conducted ChIP-Seq using a FLAG antibody on homozygous *3×flag::cfi-1* animals at the third larval stage (L3), as all *cfi-1*-expressing cells are generated by this stage.

Our ChIP-Seq experiment revealed strong enrichment of CFI-1 binding in the *C. elegans* genome, identifying 14,806 unique binding peaks (q-value cutoff: 0.05) (Figure. 2.1F-G). The CFI-1 peaks are predominantly located between 0 and 3kb upstream of transcription start sites (Figure. 2.1H), suggesting CFI-1 acts at promoter and enhancer regions to regulate gene expression. Altogether, ChIP-Seq for CFI-1 generated the first *in vivo* binding map of an endogenously tagged ARID3 protein, offering an opportunity to comprehensively identify the biological processes controlled by CFI-1.

2.4.2 The majority of CFI-1/ARID3 target genes encode neuronal terminal differentiation markers

Subsequent bioinformatic analysis of the 14,806 CFI-1 binding peaks revealed 6,396 protein-coding genes as putative CFI-1 targets (see Materials and Methods). Because the majority of *cfi-1*-expressing cells are neurons (Figure. 2.1C-E) (Kerk et al., 2017, Shaham and Bargmann, 2002), we reasoned that a significant portion of the 6,396 protein-coding genes may be expressed in the nervous system. To test this, we used available single-cell expression profiles (CeNGEN project: `www.cengen.org`) for all known *cfi-1*-expressing neurons (IL2, URA, AVD, PVC, LUA, DA, DB, VA, VB, DD, VD), and indeed found that 77.1% of the global CFI-1 targets (4,931 out of 6,396) are expressed in these neurons (Figure. 2.1J). To gain insights into the biological functions of CFI-1, we conducted gene ontology (GO) analysis with PANTHER (Mi et al., 2013). Strikingly, the majority of CFI-1 target genes (∼70%) encodes proteins essential for neuronal terminal differentiation (e.g., NT receptors, transporters, ion channels, transmembrane receptors, cell adhesion molecules) (Figure. 2.1I-K). The second largest category (23% of CFI-1 targets) contains transcription factors, chromatin factors, as well as proteins involved in DNA/RNA metabolism (Figure. 2.1I-K), suggesting CFI-1 can affect gene expression indirectly through these factors. Altogether, the downstream targets identified via our unbiased approach suggest that CFI-1 plays a prominent role in neuronal terminal differentiation.

2.4.3 CFI-1/ARID3 acts directly to activate terminal differentiation genes in IL2 sensory neurons

Although *cfi-1* is expressed in several neuron types, a handful of CFI-1 target genes have only been identified in head sensory neurons of the IL2 class (Figure. 2.3A-B) (Shaham

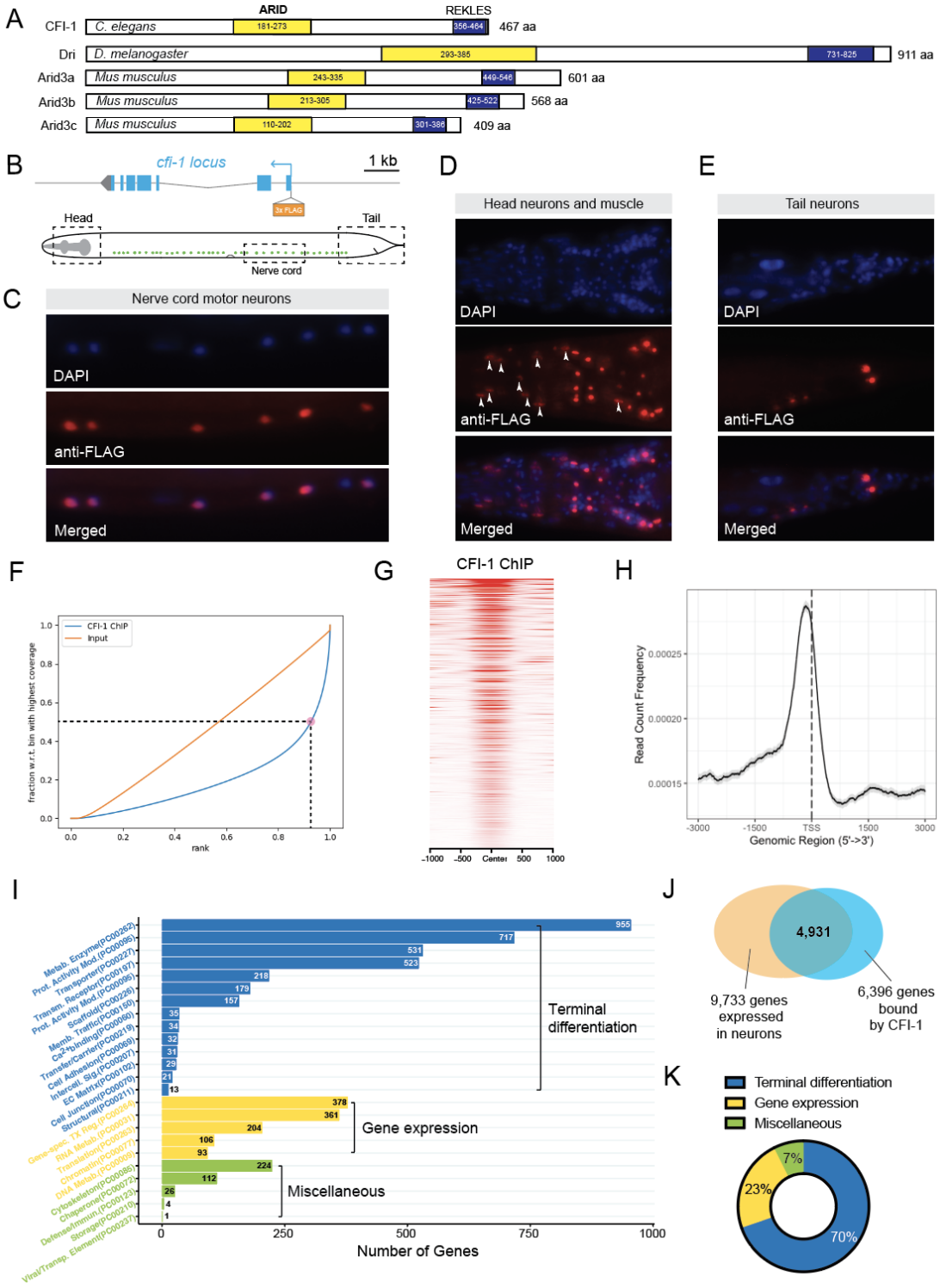


Figure 2.1: **Mapping genome-wide CFI-1 binding with ChIP-seq.**
A, Schematics of the coding sequences of CFI-1 and its *Drosophila* homolog *Dead ringer (Dri)* and Mouse homologs Arid3a-c. The ARID (yellow) and REKLES (blue) domains are highlighted.

Figure 2.1, continued.
B, Diagram of the *3×flag::cfi-1* allele. The endogenous CFI-1 proteins are tagged with 3xFLAG which are inserted immediately after the ATG. (Bottom) Schematic of *C. elegans* with dashed line boxes highlighting regions where immunostaining images are shown in **C-E**.
C-E, The expression of 3xFLAG::CFI-1 fusion protein is confirmed by immunostaining (DAPI, blue; anti-FLAG, red) in the ventral cord motor neurons (**C**), muscles (indicated by arrowheads) and neurons in the head (**D**), and neurons in the tail (**E**)
F, Fingerprint plot indicating localized, strong enrichment of CFI-1 binding events in the genome. Specifically, when counting the reads contained in 86% of all genomic bins, only 50% of the maximum number of reads are reached, which indicates 14% of the genome contain half of total reads.
G, Heatmap of CFI-1 ChIP-seq signal around 1.0 kb of the center of the binding peak.
H, Summary plot of CFI-1 ChIP-seq signal with a 95% confidence interval (grey area) around 3.0 kb of the transcription start site (Kadkhodaei et al.). The average signal peak is detected at ∼140 bp upstream of the TSS.
I, Graph summarizing protein class ontology analysis of global CFI-1 target genes identified by ChIP-seq. A total number of 7,995 genes are analyzed, 4,984 of which hit known protein class terms.
J, Pie chart summarizing three main categories of genes bound by CFI-1.
K, Venn diagram showing that 77.1% of the protein-coding genes (4,931 out of 6,396) that are bound by CFI-1 are expressed in the nervous system based on RNA-Seq data from the CenGEN project.

and Bargmann, 2002, Zhang et al., 2014). In these neurons, genetic experiments suggested CFI-1 influences gene expression both positively and negatively (Shaham and Bargmann, 2002, Zhang et al., 2014). CFI-1 activates various terminal differentiation genes (e.g., *cho-1/ChT*, *unc-17/VAChT*, *gcy-19* [receptor-type guanylate cyclase], *klp-6* [kinesin-like protein], *unc-5* [netrin receptor]), and represses expression of two ion channel-encoding genes (*pkd-2/Polycystin-2 like 1 [PKD2L1]* and *lov-1/Polycystin-1 like 3 [PKD1L3]*) associated with polycystic kidney disease (Zheng et al., 2018b). It remained unknown, however, whether CFI-1 acts directly or indirectly to control these genes. Leveraging our ChIP-Seq dataset, we found that CFI-1 binds directly to all known terminal differentiation genes (e.g., *cho-1/ChT*, *unc-17/VAChT*) that require *cfi-1* gene activity for their activation in IL2 neurons (Figure. 2.3C). However, we did not detect any binding in the *cis*-regulatory regions of

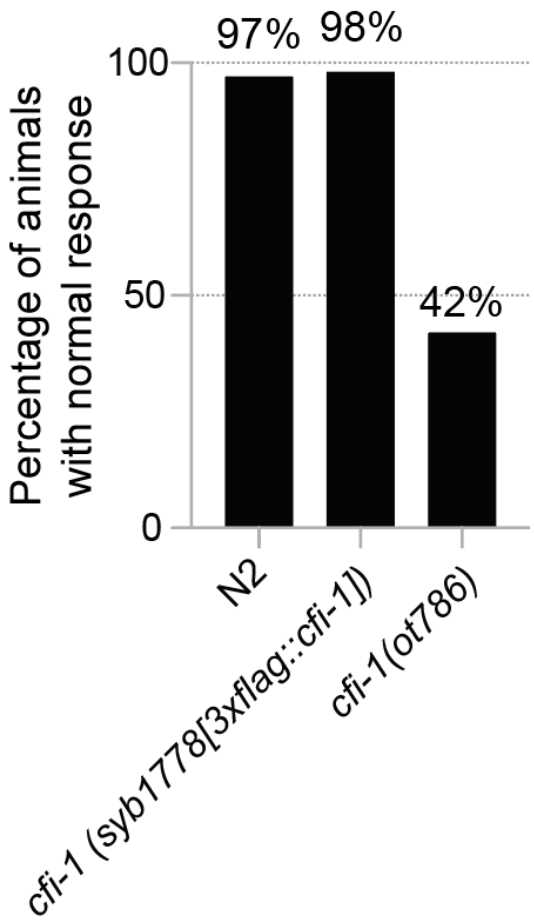


Figure 2.2: **Harsh touch assay on *C. elegans* animals carrying the *3×flag::cfi-1* allele used for ChIP-Seq.**
Summary of the percentage of animals that show normal response to harsh touch stimulation for each genotype. Normal response is determined by forward locomotion upon harsh touch of the tail. See Materials and Methods for details. The null *ot786* allele of *cfi-1* is used as a positive control. N = 60.

genes repressed by *cfi-1* (*pkd-2*, *lov-1*) (Fig. 2D). Altogether, biochemical evidence (ChIP-Seq) combined with genetic studies (Shaham and Bargmann, 2002, Zhang et al., 2014) strongly suggest that in IL2 sensory neurons, CFI-1 exerts a dual role: it directly activates a set of terminal differentiation genes, but indirectly (via intermediary factors) represses the expression of ion channel-encoding genes (*pkd-2/Polycystin-2* and *lov-1/Polycystin-1 like*) (Figure. 2.3B).

We next undertook an unbiased approach to investigate whether genes expressed in mature IL2 neurons are also bound by CFI-1. To this end, we used available single-cell expression profiles (CeNGEN project: `www.cengen.org`) and selected the most highly expressed genes (top 1,000) in IL2 neurons. Strikingly, most of these genes (∼70%) are bound by CFI-1 (Figure. 2.3E). Among them, GO analysis revealed an overrepresentation of terminal differentiation genes (Figure. 2.3E). This analysis provides biochemical evidence to support the idea that CFI-1 directly activates scores of terminal differentiation genes in IL2 neurons.

2.4.4 CFI-1/ARID3 acts directly to repress endogenous glr-4/GRIK1 expression in cholinergic motor neurons

Our findings in IL2 sensory neurons suggest CFI-1 is a direct activator and indirect repressor of gene expression (Figure. 2.3B-D). Next, we interrogated the function of CFI-1 in cholinergic nerve cord motor neurons that control locomotion (Figure. 2.4A). Using an endogenous reporter allele (Li et al., 2020), we found that *cfi-1* is selectively expressed in 29 cholinergic motor neurons (of the DA, DB, VA, VB classes) located in the mid-body region (Figure. 2.4A, Figure. 2.5). Next, we asked whether in these neurons CFI-1 functions as an activator of gene expression. To this end, we examined for *cfi-1* dependency five available motor neuron-specific terminal differentiation markers (*twk-40*, *twk-43* [TWiK potassium channels]; *ncs-2* [neuronal calcium sensor]; *npr-29* [neuropeptide]; *dbl-1* [Bmp-like]) (Li and Kratsios, 2021). These showed no difference in expression in motor neurons of *cfi-1(-)* mutants (Figure. 2.6). Consistently, a previous study also found that three other terminal differentiation markers (*acr-5* [acetylcholine receptor], *del-1* [SCNN1 sodium channel], *inx-12* [gap junction protein]) are not affected in motor neurons of *cfi-1(-)* mutants (Kerk et al., 2017). Interestingly, our ChIP-seq data indicate that these eight genes are all bound

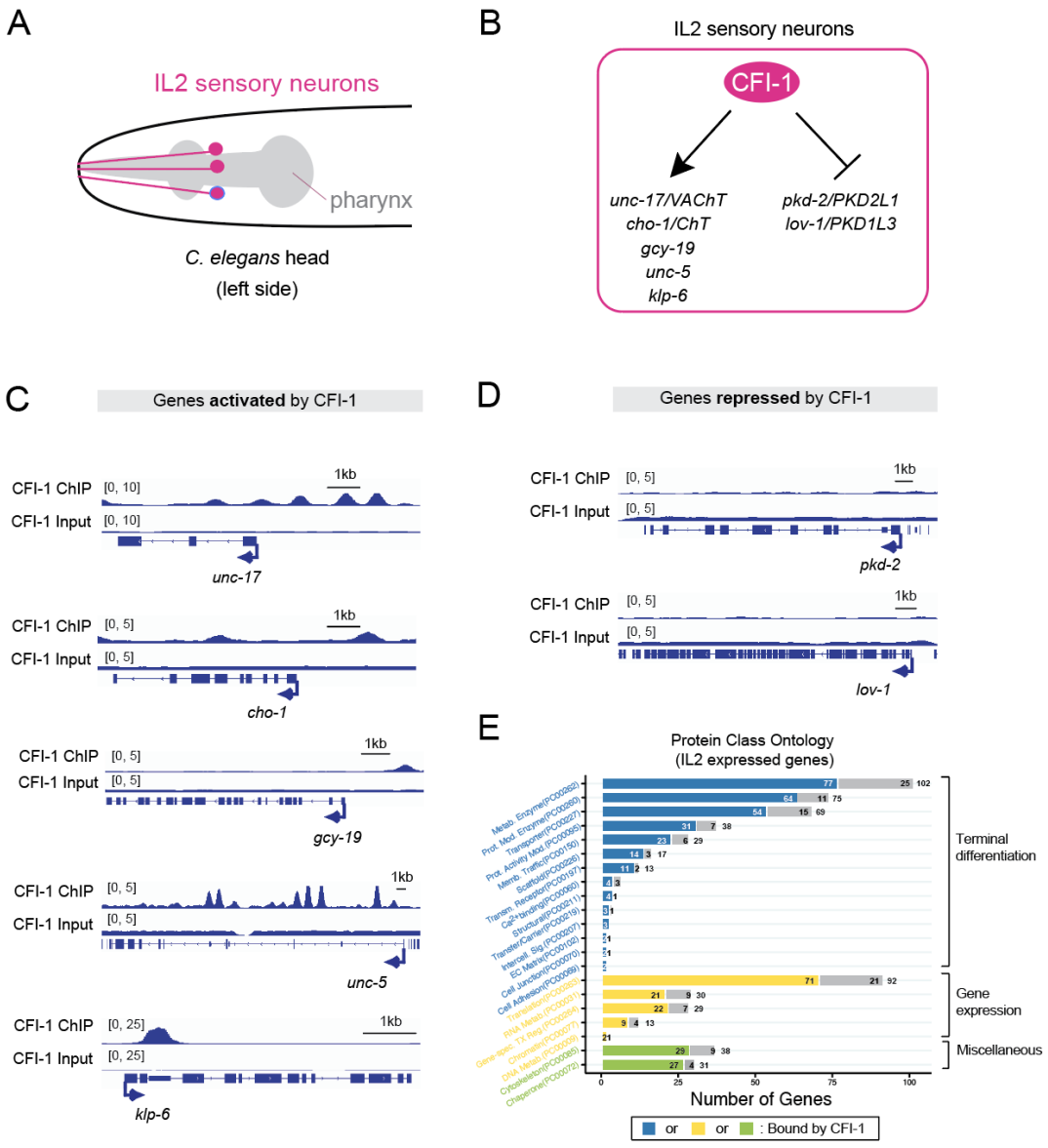


Figure 2.3: **CFI-1 directly activates terminal differentiation genes in the IL2 sensory neurons.**
A, Diagram showing the location of the IL2 sensory neurons.
B, Model summarizing known targets of CFI-1 in the IL2 sensory neurons.
C, CFI-1 binding signals at the loci of genes activated by CFI-1 in the IL2 neurons.
D, CFI-1 binding signals at the loci of genes repressed by CFI-1 in the IL2 neurons.
E, Protein class ontology analysis of genes expressed in IL2 sensory neurons binned by CFI-1 binding based on ChIP-seq data. Bar colors correspond to colors in **Figure. 2.1I**: Blue, genes involved in terminal differentiation; yellow, genes involved in regulating gene expression; green, miscellaneous genes.

by CFI-1 (Figure. 2.6), raising the possibility of CFI-1 operating redundantly with other transcription factors to activate expression of terminal differentiation genes in motor neurons.

The terminal differentiation gene *glr-4* (ortholog of human GRIK1 [glutamate inotropic receptor kainite type subunit 1]) is the only known CFI-1 target in motor neurons, where it is negatively regulated by CFI-1 (Kerk et al., 2017), providing an opportunity to obtain mechanistic insights into how ARID3 proteins mediate gene repression in the nervous system. We therefore carried out an in-depth investigation focused on *glr-4*, as detailed below.

Because previous studies employed transgenic reporters (Brockie et al., 2001, Kerk et al., 2017), the endogenous expression pattern of *glr-4* in *C. elegans* neurons remained unclear. We therefore generated an endogenous *glr-4* reporter allele by inserting the *2xNLS::mScarlet::SL2* cassette immediately after the start codon (see Materials and Methods) and established the *glr-4* expression pattern in motor neurons with single-cell resolution (Figure. 2.4B). Consistent with previous studies (Brockie et al., 2001, Kerk et al., 2017), we observed high levels of *glr-4* (mScarlet) expression in head neurons, as well as in SAB motor neurons that innervate head muscle (Figure. 2.4A, C). This endogenous reporter also revealed new sites of expression. In the nerve cord of WT animals, we detected low levels of *glr-4* (mScarlet) expression in 14 of the 29 *cfi-1*-expressing motor neurons, as well as in AS motor neurons, which do not express *cfi-1* (Figure. 2.4A-B,D, Figure. 2.5). Further, the observed levels of *glr-4* expression in head (SAB) and ventral cord motor neurons were independently confirmed by available scRNA-Seq data (CeNGEN project: `www.cengen.org`) (Figure. 2.4E).

Next, we tested whether endogenous *glr-4* expression in motor neurons depends on *cfi-1* gene activity. Indeed, expression of the *glr-4* (mScarlet) reporter allele is increased in motor neurons of *cfi-1* loss-of-function mutants, as we observed a higher number of cells expressing

glr-4, and at higher levels, compared to WT motor neurons (Figure. 2.4D). No effects on *glr-4* were observed in SAB neurons, as they do not express *cfi-1* (arrowheads in Figure. 2.4C). Moreover, we performed single-molecule mRNA fluorescent *in situ* hybridization (smRNA FISH) in WT and *cfi-1* mutant animals. Loss of *cfi-1* led to increased levels of *glr-4* mRNA in nerve cord motor neurons (Fig. 3F). Lastly, these results were corroborated by RT-PCR in WT and *cfi-1* mutant animals (Figure. 2.4G). Altogether, we conclude CFI-1 limits the endogenous expression of *glr-4/GRIK1* in nerve cord motor neurons. In WT animals, we can either detect low or no *glr-4* expression in motor neurons, whereas loss of *cfi-1* results in robust *glr-4* expression in these cells (Figure. 2.4H). Because ChIP-Seq revealed extensive CFI-1 binding immediately upstream of the *glr-4* locus (Figure. 2.4B), we propose CFI-1 acts as a direct repressor of *glr-4/GRIK1*.

2.4.5 CFI-1/ARID3 is required to maintain glr-4 repression in nerve cord motor neurons

The cfi-1-expressing motor neurons (DA, DB, VA, VB) are generated in two waves (DA/DB are born during embryogenesis; VA and VB at larval stage 1 [L1]) (Figure. 2.7A-B). Although all *cfi-1*-expressing motor neurons have been generated by L2, we found no *glr-4* (mScarlet) expression in WT animals at this stage. However, we observed a progressive increase in the number of WT motor neurons expressing low levels of *glr-4* (mScarlet) at subsequent stages (L3, L4, Day 2 [D2] adult), indicating a correlation between *glr-4* expression and motor neuron maturation (Figure. 2.7A).

To test when *cfi-1* gene activity is required for repression, we monitored endogenous *glr-4* (mScarlet) expression in motor neurons of WT and *cfi-1* mutant animals at larval (L2, L3, L4) and adult (day 2) stages. Compared to controls, we identified a statistically significant

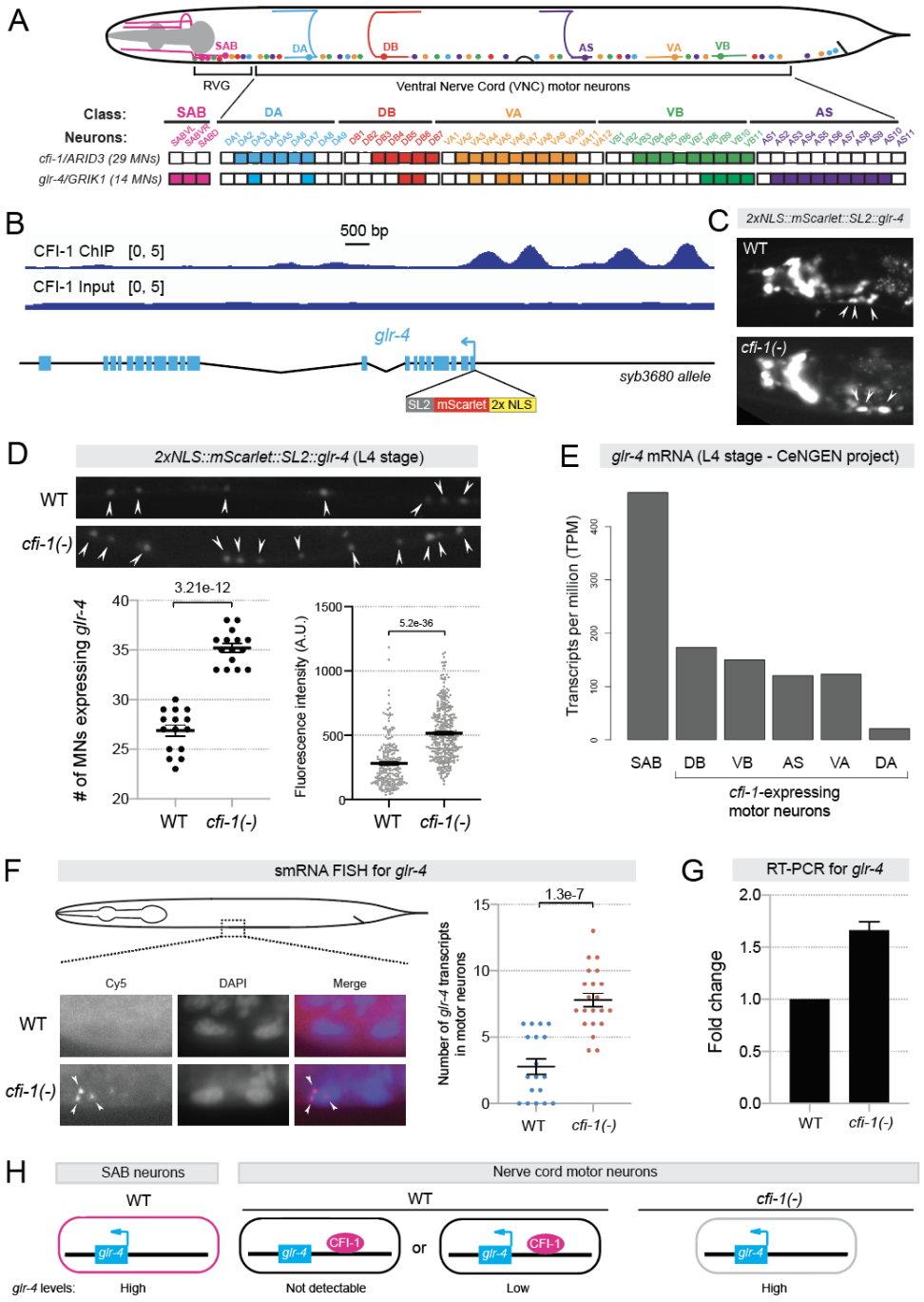


Figure 2.4: **CFI-1/Arid3a represses *glr-4/GRIK1* in the ventral cord motor neurons**

A, Summary of the endogenous expression patterns of *cfi-1* and *glr-4* in the five cholinergic nerve cord motor neuron subtypes and the SAB neurons in the retrovesicular ganglion (RVG). The expression patterns are determined by colocalization with neuron subtype-specific reporters. Filled boxes represent positive expression, while empty boxes indicate no detectable expression.

Figure 2.4, continued.
B, CFI-1 binding signal at the *glr-4* locus. (Bottom) Design of an endogenously tagged, nuclear localized fluorescent reporter allele of *glr-4 (2×NLS::mScarlet::SL2::glr-4)*. The reporter cassette is inserted immediately downstream of the ATG of *glr-4* with CRISPR/Cas9.
C, Representative fluorescent micrographs showing that the expression of 2xNLS::mScarlet::SL2::GLR-4 is unaffected in SAB neurons (arrowheads) in *cfi-1(-)* mutants. Bright white signal to the left of the SAB neurons indicates *glr-4* expression in head neurons.
D, *2×NLS::mScarlet::SL2::glr-4* expression in the ventral cord motor neurons (indicated by arrowheads in the representative images) in WT and *cfi-1(-)* animals at L4 stage. Ectopic expression of *glr-4* was detected in *cfi-1(-)* mutants. (Bottom) Each dot in the quantification graph of total number of motor neurons (left) represents an individual animal. Each dot in the fluorescence intensity quantification graph (right) represents an individual motor neuron that shows *glr-4* expression. For simplicity, only motor neurons in the anterior VNC are included in the fluorescence intensity quantification. p-values are indicated in the graphs. $N \geq 15$.
E, RNA-seq data from the CeNGEN project showing expression of *glr-4* transcripts in SAB, DB, VB, AS, VA, and DA motor neurons.
F, Single molecule fluorescent *in situ* hybridization (smFISH) shows mild ectopic expression of *glr-4* transcripts in the motor neurons of L1 *cfi-1(-)* animals. Left: represented images showing ectopic *glr-4* mRNA molecules (Cy5, red) in the nucleus (DAPI, blue) of a motor neuron in *cfi-1(-)* mutants. Right: quantification of the number of *glr-4* transcripts detected in the anterior ventral nerve cord in WT and *cfi-1(-)* animals. p-value is indicated in the graph. $N \geq 18$.
G, RT-PCR from whole worm lysates showing upregulation of *glr-4* transcripts in *cfi-1(-)* mutants.
H, Models summarizing the repressive regulation of *glr-4* expression by CFI-1 in the SAB neurons versus ventral nerve cord motor neurons.

increase in the number of *glr-4* (mScarlet)-expressing motor neurons at L3, L4 and adult (day 2) stages (Figure. 2.7A). Next, we used a conditional *cfi-1* allele (*mNG::3×FLAG::AID::cfi-1*) that enables temporally controlled CFI-1 protein depletion upon administration of the plant hormone auxin (Li et al., 2020, Zhang et al., 2015). Depletion of CFI-1 during the first 2 days of adulthood led to a significant increase in the number of motor neurons expressing *glr-4* (mScarlet) (Figure. 2.7sB-C), suggesting *cfi-1* is continuously required to maintain *glr-4/GRIK1* repression in the adult (see Discussion).

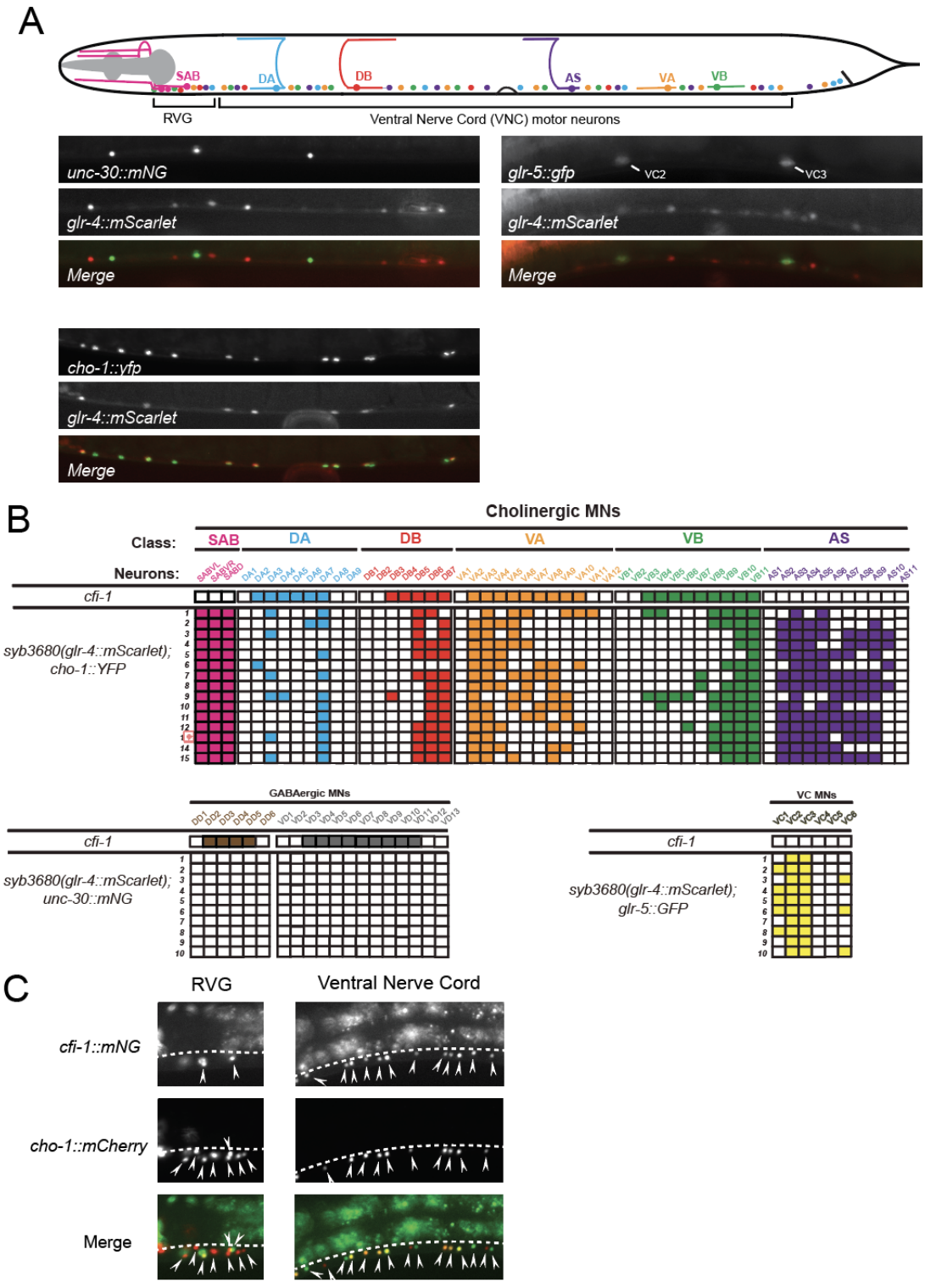


Figure 2.5: **Expression of *glr-4* in nerve cord motor neurons.**

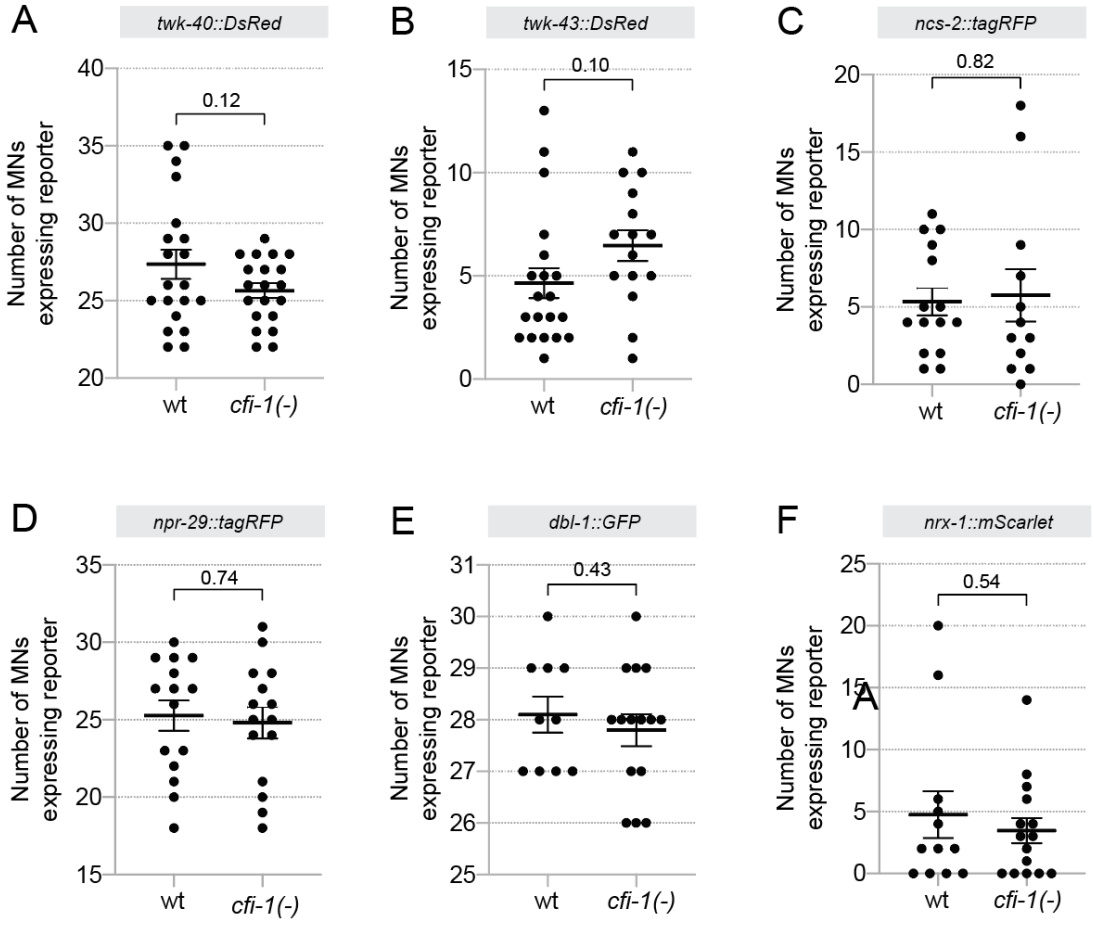
Figure 2.5, continued.
A, Worm schematic with motor neuron subtypes. Representative images showing colocalization analysis of v expression and motor neuron subtype-specific reporters *unc-30::mNG* (marker for GABAergic motor neurons), *glr-5::gfp* (marker for VC neurons), and *cho-1::yfp* (marker for cholinergic motor neurons).
B, Summary *glr-4* expression in VNC MNs. Filled boxes indicate expression, while empty boxes indicate no detectable expression. Each row represents an individual worm scored at larval stage 4 (L4) for colocalization of *glr-4* and motor neuron subtype-specific reporters.
C, Representative images showing the colocalization analysis for *cfi-1* expression in RVG and ventral cord motor neurons. The *cho-1::mCherry* reporter is expressed in cholinergic motor neurons.

2.4.6 CFI-1/ARID3 is sufficient to repress glr-4 expression

To test whether *cfi-1* is sufficient to repress *glr-4*, we ectopically expressed *cfi-1* in the SAB neurons. Using an SAB-specific-promoter (*unc-4*) to drive *cfi-1*, we observed a significant decrease in the number of SAB neurons expressing a *glr-4* reporter gene (Figure. 2.7D-E), indicating *cfi-1* is not only necessary (Figure. 2.4), but also sufficient to repress *glr-4* expression.

2.4.7 CFI-1 binding sites proximal to glr-4 are necessary for repression in motor neurons

CFI-1 binds to both proximal and distal *cis*-regulatory elements upstream of *glr-4* (Figure. 2.8A). To precisely identify the elements through which CFI-1 mediates repression, we conducted *cis*-regulatory analysis in the context of transgenic reporter animals. When tagRFP was driven by distal regulatory elements (2.23kb or 938bp), we did not observe differences in the number of tagRFP expressing motor neurons between WT and *cfi-1* mutants (Figure. 2.8B). However, we found an increase in the number of tagRFP-expressing motor neurons in *cfi-1* mutants when tagRFP was driven by a 3.7kb element (Figure. 2.8B), suggesting this element contains sequences necessary for CFI-1 repression. A translational reporter (GLR-



G

Gene	CFI-1 binding
twk-40	*YES*
twk-43	*YES*
ncs-2	*YES*
npr-29	*YES*
dbl-1	*YES*
nrx-1	*YES*
acr-5	*YES*
del-1	*YES*
inx-12	*YES*

Figure 2.6: **The expression of motor neuron terminal differentiation genes is not affected in *cfi-1* mutant animals.**
A-F, Quantification of the number of MNs expressing fluorescent reporters for terminal identity genes in WT and cfi-1 (ot786) mutant animals. Stage: L4. P-value is indicated in the graph. N = 10.
G, All genes are bound by CFI-1 based on ChIP-seq data.

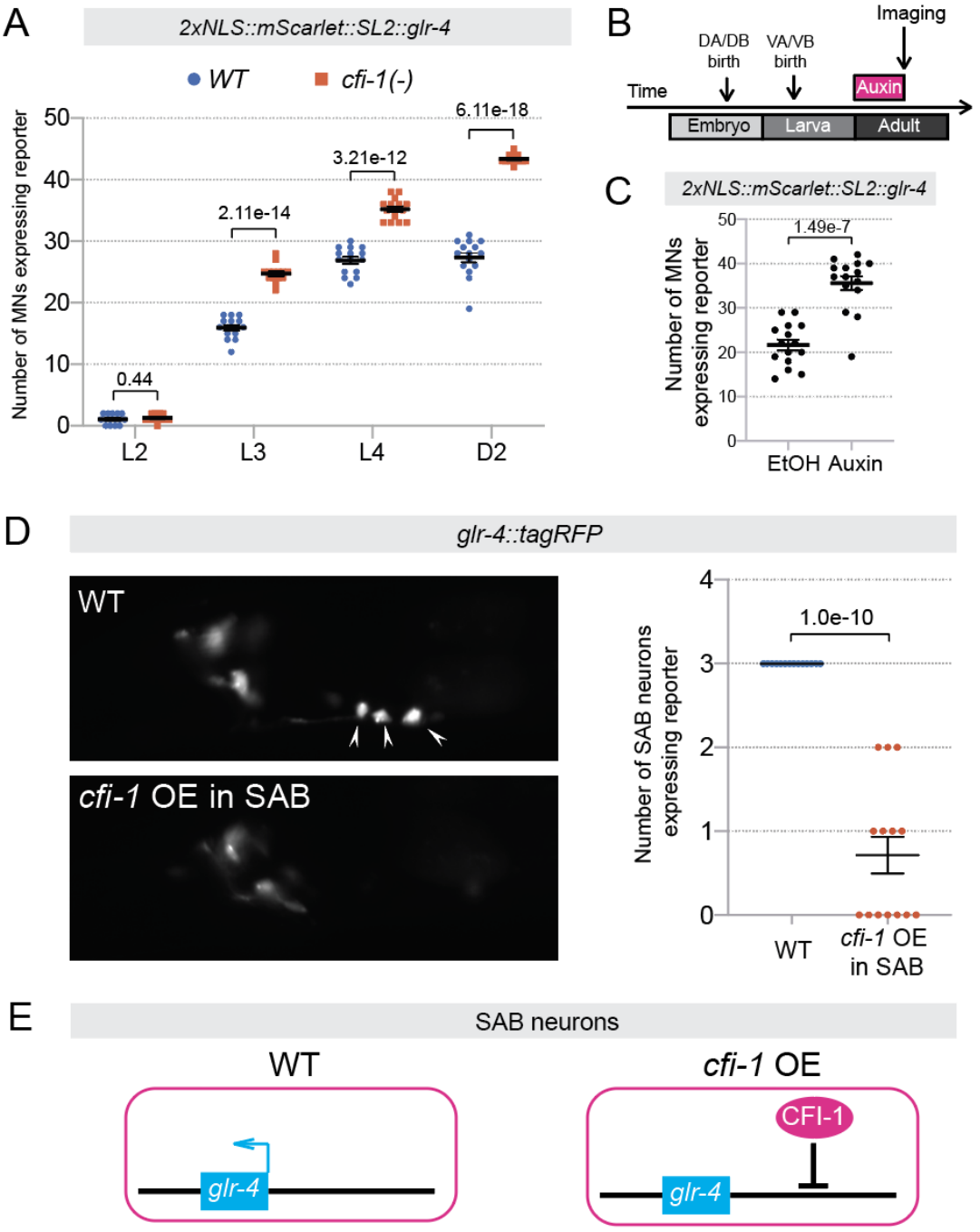


Figure 2.7: **CFI-1 is sufficient to repress *glr-4* and continuously required to maintain this repression in ventral cord motor neurons.**
A, Quantification of the number of MNs expressing *glr-4* in WT and *cfi-1(-)* animals at four developmental stages – L2, L3, L4, and day 2 adults. p-values are indicated in the graph. N $\geq$ 12.

Figure 2.7, continued.
B, Diagram showing the timeline for the administration of auxin to induce degradation of CFI-1 protein in *kas16[mNG::AID::cfi-1]*; *otTi28[unc-11prom8+ehs-1prom7+rgef-1prom2::TIR1::mTurquoise2::unc-54 3'UTR]* animals. *otTi28* drives expression of TIR1 specifically in neurons.
C, Quantification graph comparing expression of the endogenous glr-4 reporter between the ethanol group (control) and auxin group on *kas16[mNG::AID::cfi-1]*; *otTi28[unc-11prom8+ehs-1prom7+rgef-1prom2::TIR1::mTurquoise2::unc-54 3'UTR]* animals. Ectopic expression of *2×NLS::mScarlet::SL2::glr-4* in motor neurons was detected upon CFI-1 protein knock-down. p-values are indicated in the graph. N = 15.
D, CFI-1 is sufficient to repress *glr-4* expression. Left: representative images showing loss of *glr-4::tagRFP* expression in the SAB neurons (arrowheads) upon overexpression (OE) of *cfi-1*. Right: quantification of the number of SAB neurons expressing *glr-4::tagRFP* in wildtype and *cfi-1* overexpression animals. p-value is indicated in the graph. N $\geq$ 14.
E, Model summarizing the sufficiency of CFI-1 to repress *glr-4* in SAB neurons.

4::GFP) driven by a 4.9kb element yielded similar results (Figure. 2.8B). Interestingly, a shorter tagRFP reporter (3.14kb) that specifically lacks the most proximal CFI-1 binding peak did not show an increase in the number of tagRFP expressing motor neurons in *cfi-1* mutants (Figure. 2.8B), suggesting proximal CFI-1 binding to the *glr-4* locus is needed for repression.

We next sought to determine whether proximal CFI-1 binding sites are required for *glr-4* repression. The CFI-1 site (NNATHDNN) has been previously determined *in vitro* through protein binding microarrays (Figure. 2.8A) (Weirauch et al., 2014). Within the most proximal region of *glr-4*, we identified eleven predicted CFI-1 binding sites (Figure. 2.8A). To test their functionality, we introduced nucleotide substitutions to all eleven sites in the context of the endogenous *glr-4* (mScarlet) reporter through CRISPR/Cas9 genome editing (Figure. 2.8C). This manipulation nearly phenocopied the *cfi-1* null mutant phenotype, as it led to a dramatic increase in the number of mScarlet-expressing motor neurons (Figure. 2.8C-D). We conclude that CFI-1 binding sites located in the proximal region of *glr-4* are necessary for its repression in motor neurons.

2.4.8 The transcription factor UNC-3 (Collier/Ebf) and two Hox proteins (LIN-39, MAB-5) activate basal levels of glr-4/GRIK1 expression in motor neurons

Because *glr-4* is expressed at basal levels in cholinergic motor neurons of WT animals (Figure. 2.4), we reasoned this occurs due to CFI-1 antagonizing the function of *glr-4* activators in these neurons. The transcription factor UNC-3 (Collier/Ebf) and the Hox proteins LIN-39 (Scr/Dfd/Hox4-5) and MAB-5 (Antp/Hox6-8) are known to act as transcriptional activators in cholinergic motor neurons (Feng et al., 2020, Kerk et al., 2017) (Figure. 2.9A), leading us to hypothesize that they can also activate *glr-4* expression. Indeed, we found that expression of the endogenous *glr-4* (mScarlet) reporter is reduced in unc-3 loss-of-function mutant animals (Figure. 2.9B). Importantly, the decrease of *glr-4* expression was significantly exacerbated in *unc-3; lin-39; mab-5* triple mutants compared to *unc-3* single mutants, indicating that these three factors cooperate to activate *glr-4* in cholinergic motor neurons (Figure. 2.9B). Interestingly, the Hox requirement is only revealed in the absence of *unc-3* gene activity, as *glr-4* expression appears normal in *lin-39; mab-5* double mutants.

2.4.9 The most proximal UNC-3 biding site is necessary for glr-4 expression in motor neurons

Interrogation of available ChIP-Seq data for UNC-3, LIN-39 and MAB-5 showed overlapping binding upstream of *glr-4* (Figure. 2.9C), suggesting a direct mode of activation by these factors. To functionally test this notion, we focused on UNC-3 because its binding site (termed COE motif) is well-defined in the *C. elegans* genome (Kratsios et al., 2012, Li et al.,

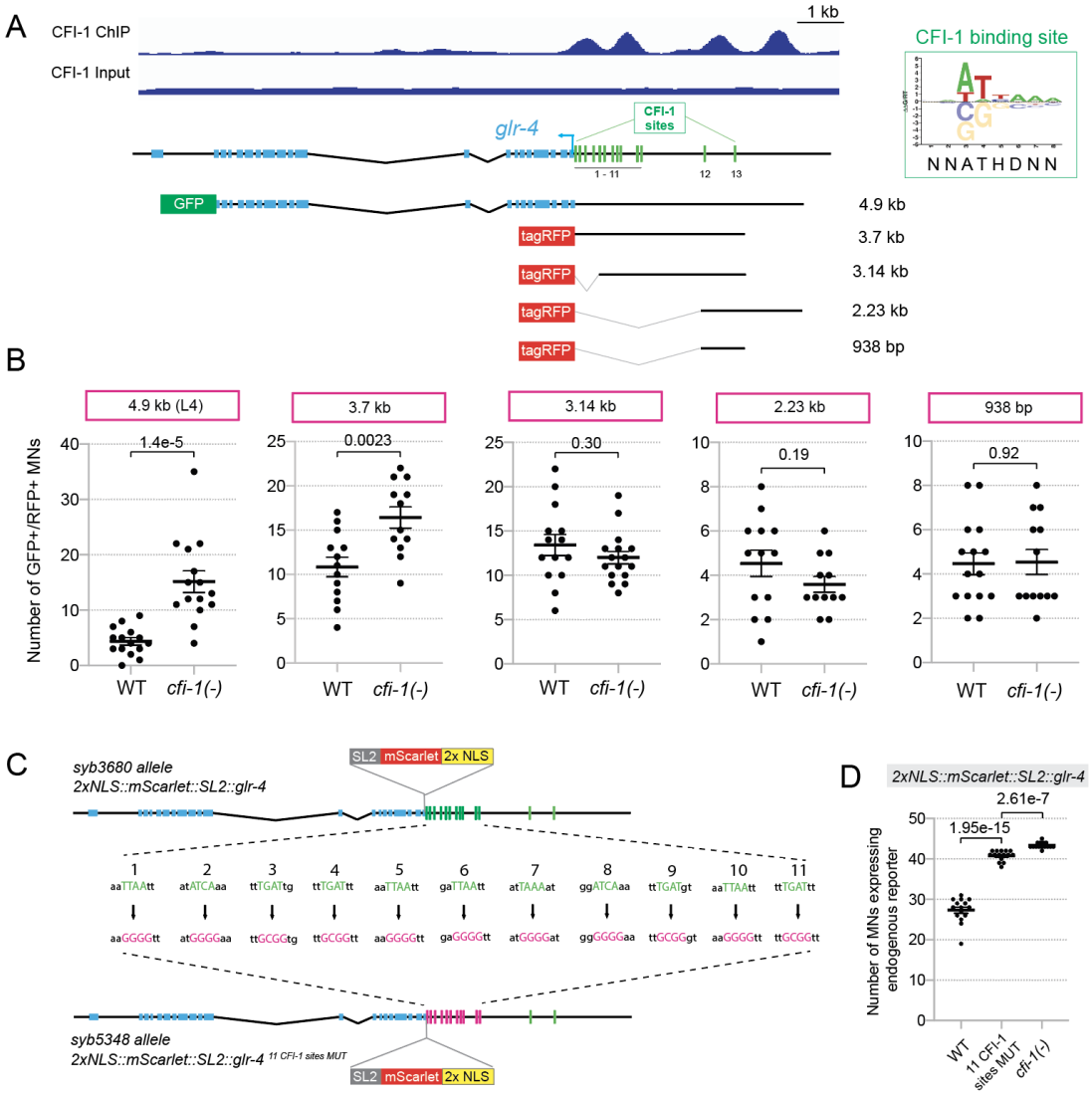


Figure 2.8: **CFI-1 directly represses *glr-4* by binding to its promoter via a conserved binding motif.**
A, CFI-1 ChIP-seq tracks at the *glr-4* locus. (Bottom) Schematics of a series of *glr-4* transgenic reporter lines which contain different parts of the *cis*-regulatory regions upstream of the gene. CFI-1 recognizes a highly conserved binding motif shared by the ARID family (right). Using bioinformatic analysis, 13 of these binding motifs were identified in the *cis*-regulatory region upstream of *glr-4*, which overlap with the CFI-1 binding peaks.

Figure 2.8, continued.
B, Quantification of the number of motor neurons expressing the reporters of *glr-4* shown in panel A in WT and *cfi-1(-)* animals in day 2 adults. Only the reporters that contain the proximal *glr-4* promoter regions show ectopic expression in *cfi-1(-)* mutants. p-values are indicated in the graph. All reporters were analyzed in the adult (D2), except the 4.9kb reporter which was analyzed at the L4 stage. N $\geq$ 13.
C, Schematic depicting the design of the *2×NLS::mScarlet::SL2::glr-4*$^{11\ CFI\text{-}1\ sites\ MUT}$ allele. Point mutations were introduced to the 11 CFI-1 binding motifs that fall in the proximal CFI-1 binding peaks at the *glr-4* promoter.
D, Quantification of the number of motor neurons expressing WT *2×NLS::mScarlet::SL2::glr-4* reporter and the *2×NLS::mScarlet::SL2::glr-4*$^{11\ CFI\text{-}1\ sites\ MUT}$ allele at day 2 adult stage. Ectopic expression of the reporter was observed with the point mutation allele, which shows a total number of *glr-4*-expressing motor neurons comparable to what was observed in *cfi-1(-)* null mutants.

2020). Through a bioinformatic search (Materials and Methods), we found two COE motifs, one proximal and one distal to the *glr-4* locus (Figure. 2.9C). Using CRISPR/Cas9 genome editing, we introduced nucleotide substitutions to the proximal COE motif (COE1) in the context of the endogenous *glr-4* (mScarlet) reporter (Figure. 2.9C). Animals carrying this *glr-4* (mScarlet) COE1 MUT reporter allele showed a significant reduction in the number of mScarlet-expressing motor neurons, reminiscent of the effect seen in *unc-3 (-)* null mutants (Fig. 6D). These data indicate that, in WT animals, the most proximal COE motif is necessary for basal *glr-4* expression in motor neurons.

2.4.10 CFI-1 antagonizes the ability of UNC-3 to activate glr-4 expression in motor neurons

Because UNC-3 activates basal *glr-4* expression in motor neurons of WT animals, we wondered whether it also controls the increased levels of *glr-4* expression observed in *cfi-1* mutants. We found this to be the case through double mutant analysis. The number of mScarlet-expressing cells in *cfi-1; unc-3* mutants is dramatically decreased compared to *cfi-*

1 single mutants (Figure. 2.9E). These data indicate that CFI-1 antagonizes the ability of UNC-3 to activate *glr-4* (mScarlet) expression in motor neurons.

Because the proximal UNC-3 binding site (COE motif) is required to activate *glr-4* expression in WT motor neurons (Figure. 2.9D), this site may also be necessary for the increased expression of *glr-4* in motor neurons of *cfi-1* mutants. Indeed, we observed a significant reduction in the number of mScarlet-expressing cells in *cfi-1* mutants carrying the *glr-4* (mScarlet) COE1 MUT reporters compared to the intact version of the *glr-4* (mScarlet) reporter (Figure. 2.9E). Lastly, we hypothesized that the extensive binding of CFI-1 (four binding peaks identified by ChIP-seq) immediately upstream of the *glr-4* locus may limit the ability of UNC-3 to access the locus, resulting in basal levels of *glr-4* expression in WT motor neurons. To test this, we performed ChIP-Seq for UNC-3 in WT and *cfi-1* null mutant animals (at the L3 stage). We found that UNC-3 binding on the *glr-4* locus remains largely unaltered upon *cfi-1* loss (Fig. 6C), suggesting UNC-3 can access the locus independently of the presence of CFI-1.

2.4.11 The core ARID domain of CFI-1 is partially required for glr-4 repression

To gain molecular insights into ARID3-mediated gene repression, we deleted portions of the CFI-1 DNA-binding domain and then conducted rescue assays to assess *glr-4* expression. ARID3 proteins are defined by the eARID domain, a ∼40 residue-long highly conserved domain immediately following the core ARID domain (Figure. 2.10A). Because structural studies on *Dead ringer* showed that eARID contacts DNA (Iwahara and Clubb, 1999, Iwahara et al., 2002, Patsialou et al., 2005), we tested whether the CFI-1 eARID domain is required for *glr-4* repression. Transgenic expression of either WT CFI-1 or CFI-1 lacking

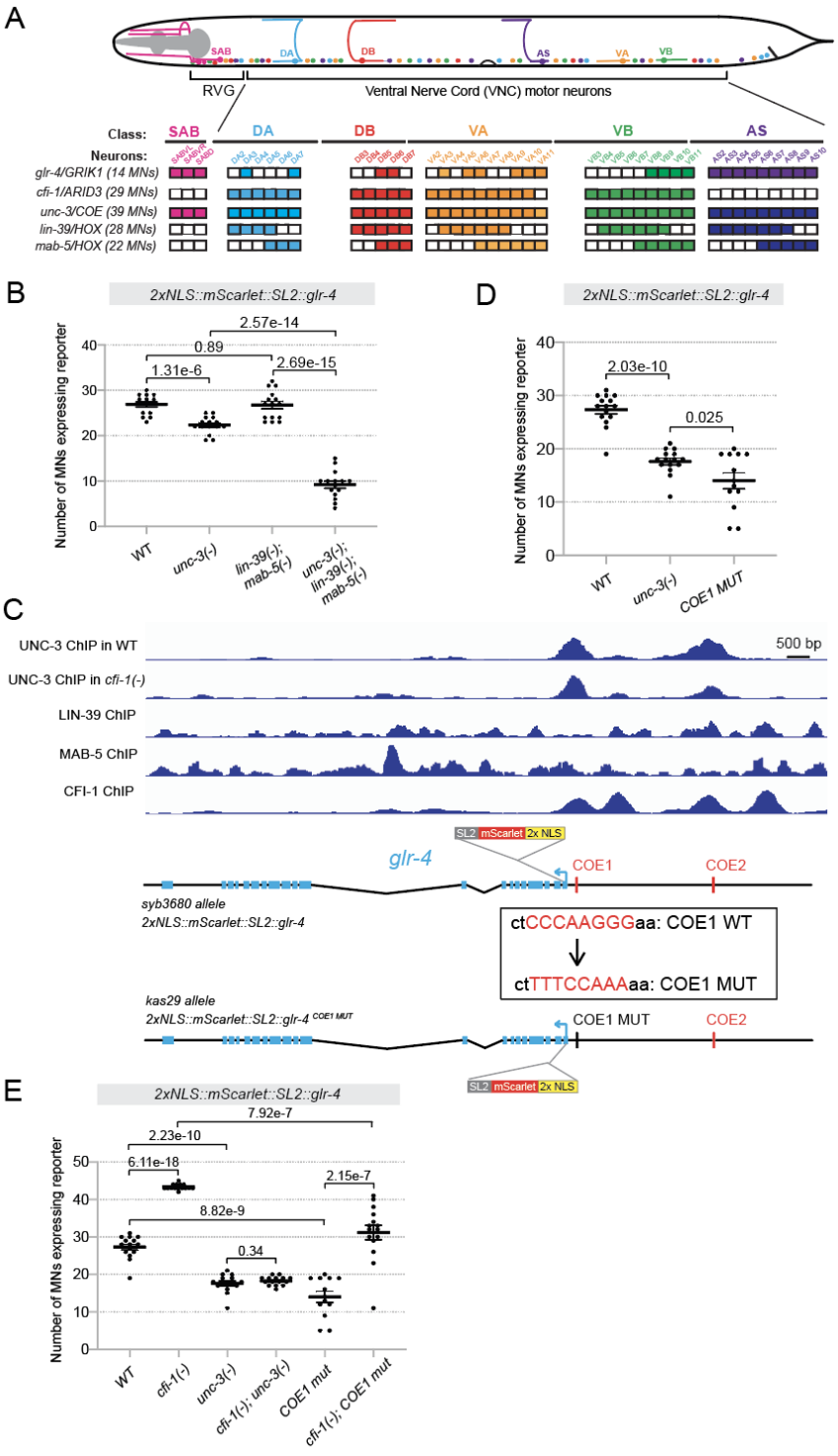


Figure 2.9: **CFI-1 represses *glr-4* by counteracting activation from the cholinergic terminal selector UNC-3 and Hox proteins LIN-39 and MAB-5.**

Figure 2.9, continued.
A, Summary of the endogenous expression patterns of *cfi-1*/Arid3, *glr-4*/GRIK1, *unc-3*/COE, *lin-39*/HOX, and *mab-5*/HOX in the five cholinergic motor neuron subtypes and in the SAB neurons of the retrovesicular ganglion (RVG). Filled boxes represent positive expression, while empty boxes indicate no detectable expression.
B, Graph summarizing quantification results of the expression of the *2×NLS::mScarlet::SL2::glr-4* reporter in WT animals, *cfi-1(-)* mutants, *lin-39(-); mab-5(-)* double mutants, and *unc-3(-); lin-39(-); mab-5(-)* triple mutants at the L4 stage. p-values are indicated in the graph. N = 15.
C, ChIP-seq binding peaks for UNC-3 (in WT animals and *cfi-1(-)* mutants), CFI-1, LIN-39, and MAB-5 at the *glr-4* locus. (Bottom) Schematic showing the details in the design of *2×NLS::mScarlet::SL2::glr-4*$^{COE1\ MUT}$. Point mutations were introduced to the proximal COE motif (UNC-3 binding site) to the *glr-4* gene.
D, Graph summarizing expression of the *2×NLS::mScarlet::SL2::glr-4* reporter in WT animals and *unc-3(-)* mutants, and the expression of the *2×NLS::mScarlet::SL2::glr-4*$^{COE1\ MUT}$ allele. Quantification was conducted at the day 2 adult stage. P-values are indicated in the graph. N $\geq$ 13.
E, Graph summarizing expression of the *2×NLS::mScarlet::SL2::glr-4* reporter in WT animals, *cfi-1(-)* mutants, *unc-3(-)* mutants, and *cfi-1(-); unc-3(-)* double mutants and the expression of the *2×NLS::mScarlet::SL2::glr-4*$^{COE1\ MUT}$ allele in WT animals and *cfi-1(-)* mutants in day 2 adults. p-values are indicated in the graph. N $\geq$ 13.

the eARID domain (ΔeARID) in motor neurons of *cfi-1* null mutant animals led to complete rescue. That is, *glr-4* expression was no longer observed in motor neurons when either WT or ΔeARID CFI-1 were provided (p = 0.37) (Figure. 2.10B), indicating eARID is dispensable for CFI-1-mediated gene repression. Next, we mutated the helix-turn-helix (HTH) domain within the core ARID region, as the HTH domain of *Dead ringer* contacts the major groove of DNA (Iwahara and Clubb, 1999, Iwahara et al., 2002, Patsialou et al., 2005). Again, transgenic expression of CFI-1 lacking the HTH domain (ΔHTH) in motor neurons of *cfi-1* mutants led to significant repression of *glr-4* expression -the effect is comparable to WT CFI-1 (p = 0.05) (Fig. 7B). However, transgenic expression of CFI-1 lacking the entire core ARID (ΔARID) domain (including the HTH domain) in motor neurons of *cfi-1* mutants led to partial rescue, i.e., ΔARID CFI-1 did not completely repress *glr-4* expression compared to WT CFI-1 (p = 0.0011) (Figure. 2.10B). Altogether, our analysis suggests that the eARID

and HTH domains of CFI-1 are dispensable, but the core ARID domain is partially required for *glr-4* repression in motor neurons.

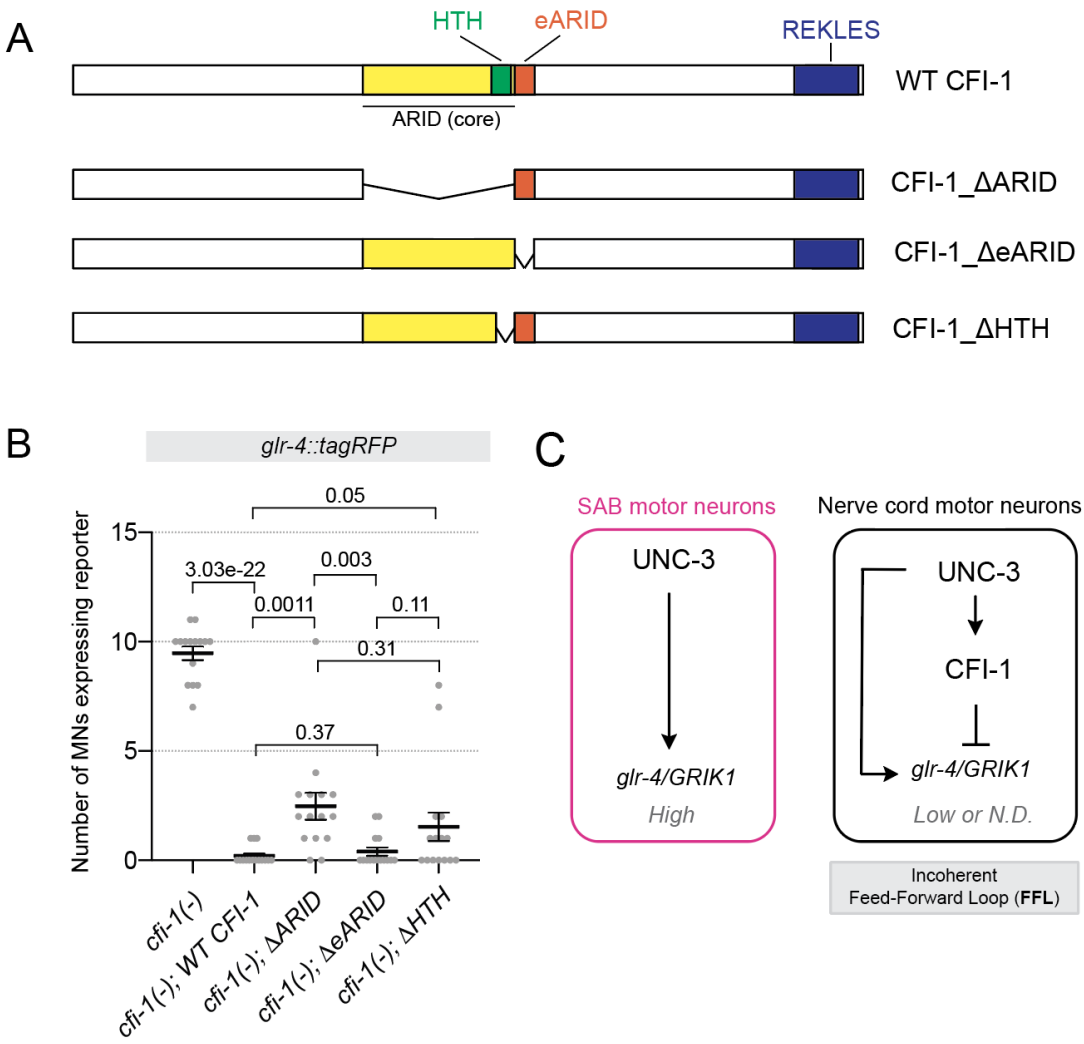


Figure 2.10: **Protein motif analysis of CFI-1.**
A, Schematics of the four CFI-1 cDNA constructs tested for rescue effects including the WT cDNA, cDNA with deletion of the ARID domain (ΔARID), cDNA with deletion of the eARID domain (ΔeARID), and cDNA with deletion of the helix-turn-helix structure (ΔHTH).
B, Quantification results of the number of motor neurons showing expression of *glr-4::tagrfp* in *cfi-1(ot786)* mutant animals expressing the rescue constructs.
C, Schematic model summarizing our findings in SAB and nerve cord motor neurons.

2.5 Discussion

ARID3 proteins are the founding members of the ARID family, but their functions in the nervous system are poorly understood. Here, we performed ChIP-Seq for CFI-1, the sole ARID3 ortholog in *C. elegans*, and established its genome-wide binding map *in vivo*. We found that most CFI-1 target genes (77%) are expressed in post-mitotic neurons and encode proteins critical for terminal differentiation and neuronal function (e.g., neurotransmitter receptors, ion channels, neuropeptides). Further, our study offers mechanistic insights into how a single ARID3 protein controls the terminal differentiation program of distinct neuron types, uncovering cell context-depending CFI-1 functions.

2.5.1 ARID3 proteins bind to both proximal and distal cis-regulatory regions

in vitro assays like PBM (protein binding microarrays) and SELEX (systematic evolution of ligands by exponential enrichment) have defined the binding site of CFI-1, Dead ringer and Arid3a (Mathelier et al., 2014, Nitta et al., 2015, Weirauch et al., 2014). To assess their mechanism of action, however, unbiased assays to monitor genome-wide binding of ARID3 proteins remain necessary. ChIP-Seq in *C. elegans* for endogenously tagged CFI-1 revealed binding events predominantly located between 0 and 3kb upstream of transcription start sites (Figure. 2.1H). Because the size of intergenic regions for most *C. elegans* genes is less than 3kb (Dupuy et al., 2004, Nelson et al., 2004), our *in vivo* binding map strongly suggests that CFI-1 can act both at proximal (e.g., promoters) and distal (e.g., enhancers) regions to regulate gene expression. Consistently, a previous *in vitro* study found that overexpressed Arid3a in mouse embryonic stem cells also binds to both proximal and distal *cis*-regulatory elements (Rhee et al., 2014).

2.5.2 The genome-wide binding map suggests a prominent role for CFI-1 in neuronal terminal differentiation

Our current knowledge of ARID3 functions in the nervous system remains rudimentary. In vertebrate nervous systems, the role of ARID3 proteins is completely unknown. In *Drosophila*, *dead ringer* has been implicated in the control of axonal pathfinding (Ditch et al., 2005, Shandala et al., 2003, Sibbons, 2004). In *C. elegans*, a handful of effector genes have been identified as CFI-1 targets in distinct neuron types (Glenwinkel et al., 2021, Kerk et al., 2017, Shaham and Bargmann, 2002, Zhang et al., 2014). Hence, a comprehensive understanding of ARID3-mediated biological processes in the nervous system is lacking.

To address this knowledge gap, we globally identified 6,396 protein-coding genes as putative direct targets of CFI-1, most of which (77%) are expressed in post-mitotic *C. elegans* neurons. GO analysis suggested a prominent role for CFI-1 in neuronal terminal differentiation, as ∼70% of its target genes encode NT receptors, transporters, ion channels, transmembrane receptors, cell adhesion molecules, etc. Moreover, 23% of CFI-1 targets encode transcription factors, chromatin factors, and proteins involved in DNA/RNA metabolism, suggesting CFI-1 can affect gene expression indirectly through these factors. Altogether, our ChIP-Seq dataset illuminates the biological processes under the control of an ARID3 protein in *C. elegans*.

2.5.3 CFI-1 acts as a terminal selector in IL2 sensory neurons

Transcription factors that bind directly to the *cis*-regulatory region of terminal differentiation genes (e.g., NT biosynthesis components, NT receptors, ion channels, neuropeptides,

membrane proteins) and activate their expression have been termed "terminal selectors" (Hobert, 2008, Hobert, 2016b, Hobert and Kratsios, 2019). Terminal selectors are continuously required in individual neuron types to initiate and maintain expression of terminal differentiation genes, thereby safeguarding neuronal functionality throughout life. To date, terminal selectors have been described in *C. elegans, Drosophila*, simple chordates and mice (Hobert and Kratsios, 2019), suggesting an evolutionarily conserved role for these critical regulators of neuronal differentiation. For most terminal selectors, however, biochemical evidence for direct binding to their target genes is currently lacking. For example, CFI-1 is a candidate terminal selector in IL2 sensory neurons because five terminal differentiation genes (*cho-1/ChT, unc-17/VAChT, gcy-19* [receptor-type guanylate cyclase], *klp-6* [kinesin-like protein], *unc-5* [netrin receptor]) fail to be properly expressed in *cfi-1* mutants (Zhang et al., 2014). Here, we provide biochemical evidence that CFI-1 binds directly to all these genes, indicating CFI-1 acts as a direct activator. Further, our analysis of the top 1,000 highly expressed genes in IL2 neurons revealed that the majority of CFI-1 binding (∼70%) occurs at terminal differentiation genes, consistent with an *in silico* prediction study for CFI-1 binding (Glenwinkel et al., 2021). Altogether, a synthesis of our findings with the aforementioned studies indicate that CFI-1 functions as a bona fide terminal selector and directly activates IL2 terminal differentiation genes. This mode of action is reminiscent of mouse Arid3a, which is known to also act as a direct activator of gene expression in cells outside the nervous system (Herrscher et al., 1995, Ratliff et al., 2014, Rhee et al., 2014).

Accumulating evidence suggests that terminal selectors act in combination with other transcription factors to determine the differentiation of individual neuron types (Glenwinkel et al., 2021, Lloret-Fernandez et al., 2018). Supporting this notion, genetic and in silico prediction studies found that CFI-1 collaborates with the POU homeodomain transcription factor UNC-86 to activate IL2-specific terminal differentiation genes (Glenwinkel et al., 2021,

Zhang et al., 2014). Similarly, it was proposed that CFI-1 collaborates with two different homeodomain proteins, UNC-42/Prop-1 like and CEH-14/LIM, to respectively control the terminal differentiation of AVD and PVC interneurons (Berghoff et al., 2021, Glenwinkel et al., 2021), albeit the underlying mechanisms remain unclear.

2.5.4 Insights into ARID3-mediated gene repression

In IL2 sensory neurons, CFI-1 exerts a dual role (Figure. 2.3). It functions as a direct activator of IL2-specific terminal differentiation genes and indirect repressor of ion channel-encoding genes (*pkd-2, lov-1*), which are normally expressed in CEM neurons (cells responsible for pheromone detection) (Chasnov et al., 2007). Hence, it promotes IL2 terminal differentiation and inhibits an alternative neuronal identity (CEM). However, in nerve cord motor neurons, CFI-1 functions as a direct repressor of the glutamate receptor-encoding gene *glr-4/GRIK1*. Altogether, these observations indicate that a single ARID3 protein can control the terminal differentiation program of distinct neuron types through mechanisms that depend on cell context, possibly due to CFI-1 participating in distinct neuron type-specific regulatory complexes that function either as dedicated activators or repressors.

Through an in-depth regulatory analysis of *glr-4/GRIK1*, our findings advance our understanding of how ARID3 proteins mediate gene repression in several aspects. First, *cis*-regulatory analysis combined with mutagenesis of endogenous CFI-1 binding sites strongly suggests that proximal CFI-1 binding to the *glr-4* locus is required for repression (Figure. 2.8). Second, CFI-1 antagonizes three conserved transcription factors (UNC-3/Ebf, LIN-39/Hox4-5, MAB-5/Hox6-8) that directly activate basal levels of *glr-4/GRIK1* expression in nerve cord motor neurons (Figure. 2.9). Third, ChIP-Seq for UNC-3 in WT and *cfi-1(-)* mutants did not reveal any changes in the UNC-3 binding pattern to *glr-4*, arguing against a

model in which *glr-4* is repressed because the repressor (CFI-1) competes with the activator (UNC-3) for binding to the same *cis*-regulatory elements. Lastly, we found that deletion of highly conserved protein domains (eARID, HTH) predicted to bind DNA (based on Dead ringer structural studies) did not affect the ability of CFI-1 to repress *glr-4* expression (Figure. 2.10). This suggests that either eARID and HTH act redundantly, or other CFI-1 domains are responsible for *glr-4* repression. Supporting the latter, deletion of the entire core ARID domain (including the HTH domain) of CFI-1 led to a partial failure to repress *glr-4* in motor neurons.

One unresolved question is how exactly CFI-1 represses gene expression in *C. elegans* motor neurons. Because proximal binding is required for *glr-4* repression, it is possible that CFI-1 interferes with the function of the basal transcription complex (Baumann et al., 2010). Another possibility stems from patterning studies in the *Drosophila* embryo, where Dead ringer binds to Groucho, a well-characterized transcriptional corepressor (Hader et al., 2000, Valentine et al., 1998). In mouse stem cells, Arid3a directly represses pluripotency genes by recruiting histone deacetylases (HDACs) (Rhee et al., 2014). Future studies are needed to determine whether CFI-1 acts through any of these repressive mechanisms in motor neurons.

s

2.5.5 CFI-1 continuously antagonizes the activator function of terminal selectors

The *glr-4* gene receives positive regulatory input from three conserved transcription factors (UNC-3/Ebf, LIN-39/Hox4-5, MAB-5/Hox6-8) and negative input from CFI-1. UNC-3 and the Hox proteins LIN-39 and MAB-5 are known terminal selectors in *C. elegans* nerve cord motor neurons (Feng et al., 2020, Kratsios et al., 2017, Kratsios et al., 2012). As such, they

are continuously required from embryonic to adult stages to activate expression of multiple terminal differentiation genes (Feng et al., 2020, Li et al., 2020). On the other hand, our constitutive (null alleles) and conditional (temporally controlled protein depletion) approaches also revealed a continuous requirement for CFI-1 in motor neurons (Figure. 2.7). This argues against a transient role where, for example, CFI-1 establishes a repressive chromatin environment during early development and its activity becomes unnecessary at later stages of life. Instead, we favor a model where CFI-1 is required continuously to prevent high levels of *glr-4* expression driven by terminal selectors (UNC-3, Hox) (Figure. 2.10C). Such model is also supported by the continuous requirement of two other repressor proteins (BNC-1/Bnc1, MAB-9/Tbx20) in *C. elegans* motor neurons (Kerk et al., 2017). Antagonism between repressor proteins and terminal selectors has also been reported in *C. elegans* touch receptor neurons with EGL-44/Tead3 and EGL-46/Insm2 repressing terminal selector target genes (Zheng et al., 2018a). Further, studies in mice indicated that the activity of two terminal selectors, *Nurr1* and *Crx*, which respectively control dopamine neuron and photoreceptor identities, is counteracted by repressor proteins (Otx2, Nr2e3)(Di Salvio et al., 2010, Peng et al., 2005). Additional work is needed, however, to determine whether all these repressor proteins (e.g., EGL-44, Otx2) act directly and are continuously required.

2.5.6 Evolutionary implications

The terminal selector UNC-3 is required to maintain *cfi-1* expression in nerve cord motor neurons (Li et al., 2020). Hence, the repressor protein (CFI-1) and *glr-4* are both targets of UNC-3, thereby generating an incoherent feedforward loop (FFL) (Figure. 2.10C). In SAB motor neurons, however, CFI-1 is not expressed, and UNC-3 is able to drive high levels of *glr-4* expression (Figure. 2.10C)(Kratsios et al., 2015). From an evolutionary perspective, incoherent FFLs have been proposed to diversify a ground state into various substates

(Hobert, 2016a). One can envision an ancestral state where an UNC-3 ortholog is present in a relatively homogeneous population of motor neurons, but the recruitment of a repressor (CFI-1) enabled their diversification. Hence, the UNC-3 → CFI-1 –I *glr-4* incoherent FFL may help distinguish, at the molecular level, nerve cord motor neurons that control locomotion from SAB motor neurons that control head movement. In agreement with this idea, incoherent FFLs are known to diversify gustatory neurons in *C. elegans* and photoreceptor cells in *Drosophila* (Etchberger et al., 2009, Johnston, 2013).

2.6 References

AHN, S., YANG, H., SON, S., LEE, H. S., PARK, D., YIM, H., CHOI, H. J., SWOBODA, P. & LEE, J. 2022. The C. elegans regulatory factor X (RFX) DAF-19M module: A shift from general ciliogenesis to cell-specific ciliary and behavioral specialization. Cell Rep, 39, 110661.

AN, G., MINER, C. A., NIXON, J. C., KINCADE, P. W., BRYANT, J., TUCKER, P. W. & WEBB, C. F. 2010. Loss of Bright/ARID3a function promotes developmental plasticity. Stem Cells, 28, 1560-7.

BAUMANN, M., PONTILLER, J. & ERNST, W. 2010. Structure and basal transcription complex of RNA polymerase II core promoters in the mammalian genome: an overview. Mol Biotechnol, 45, 241-7.

BERGHOFF, E. G., GLENWINKEL, L., BHATTACHARYA, A., SUN, H., VAROL, E., MOHAMMADI, N., ANTONE, A., FENG, Y., NGUYEN, K., COOK, S. J., WOOD, J. F., MASOUDI, N., CROS, C. C., RAMADAN, Y. H., FERKEY, D. M., HALL, D. H. & HOBERT, O. 2021. The Prop1-like homeobox gene unc-42 specifies the identity of synaptically connected neurons. Elife, 10.

BRAMSWIG, N. C., CALUSERIU, O., LUDECKE, H. J., BOLDUC, F. V., NOEL, N. C., WIELAND, T., SUROWY, H. M., CHRISTEN, H. J., ENGELS, H., STROM, T. M. & WIECZOREK, D. 2017. Heterozygosity for ARID2 loss-of-function mutations in individuals with a Coffin-Siris syndrome-like phenotype. Hum Genet, 136, 297-305.

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

BROCKIE, P. J., MADSEN, D. M., ZHENG, Y., MELLEM, J. & MARICQ, A. V. 2001. Differential expression of glutamate receptor subunits in the nervous system of Caenorhabditis elegans and their regulation by the homeodomain protein UNC-42. J Neurosci, 21, 1510-22.

CHASNOV, J. R., SO, W. K., CHAN, C. M. & CHOW, K. L. 2007. The species, sex, and

stage specificity of a Caenorhabditis sex pheromone. Proc Natl Acad Sci U S A, 104, 6730-5.

DI SALVIO, M., DI GIOVANNANTONIO, L. G., ACAMPORA, D., PROSPERI, R., OMODEI, D., PRAKASH, N., WURST, W. & SIMEONE, A. 2010. Otx2 controls neuron subtype identity in ventral tegmental area and antagonizes vulnerability to MPTP. Nat Neurosci, 13, 1481-8.

DITCH, L. M., SHIRANGI, T., PITMAN, J. L., LATHAM, K. L., FINLEY, K. D., EDEEN, P. T., TAYLOR, B. J. & MCKEOWN, M. 2005. Drosophila retained/dead ringer is necessary for neuronal pathfinding, female receptivity and repression of fruitless independent male courtship behaviors. Development, 132, 155-64.

DOITSIDOU, M., FLAMES, N., TOPALIDOU, I., ABE, N., FELTON, T., REMESAL, L., POPOVITCHENKO, T., MANN, R., CHALFIE, M. & HOBERT, O. 2013. A combinatorial regulatory signature controls terminal differentiation of the dopaminergic nervous system in C. elegans. Genes Dev, 27, 1391-405.

DUPUY, D., LI, Q. R., DEPLANCKE, B., BOXEM, M., HAO, T., LAMESCH, P., SEQUERRA, R., BOSAK, S., DOUCETTE-STAMM, L., HOPE, I. A., HILL, D. E., WALHOUT, A. J. & VIDAL, M. 2004. A first version of the Caenorhabditis elegans Promoterome. Genome Res, 14, 2169-75.

ETCHBERGER, J. F., FLOWERS, E. B., POOLE, R. J., BASHLLARI, E. & HOBERT, O. 2009. Cis-regulatory mechanisms of left/right asymmetric neuron-subtype specification in C. elegans. Development, 136, 147-60.

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.

GLENWINKEL, L., TAYLOR, S. R., LANGEBECK-JENSEN, K., PEREIRA, L., REILLY, M. B., BASAVARAJU, M., RAFI, I., YEMINI, E., POCOCK, R., SESTAN, N., HAMMARLUND, M., MILLER, D. M., 3RD & HOBERT, O. 2021. In silico analysis of the transcrip-

tional regulatory logic of neuronal identity specification throughout the C. elegans nervous system. Elife, 10.

GREGORY, S. L., KORTSCHAK, R. D., KALIONIS, B. & SAINT, R. 1996. Characterization of the dead ringer gene identifies a novel, highly conserved family of sequence-specific DNA-binding proteins. Mol Cell Biol, 16, 792-9.

HADER, T., WAINWRIGHT, D., SHANDALA, T., SAINT, R., TAUBERT, H., BRONNER, G. & JACKLE, H. 2000. Receptor tyrosine kinase signaling regulates different modes of Groucho-dependent control of Dorsal. Curr Biol, 10, 51-4.

HERRSCHER, R. F., KAPLAN, M. H., LELSZ, D. L., DAS, C., SCHEUERMANN, R. & TUCKER, P. W. 1995. The immunoglobulin heavy-chain matrix-associating regions are bound by Bright: a B cell-specific trans-activator that describes a new DNA-binding protein family. Genes Dev, 9, 3067-82.

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. 2016a. A map of terminal regulators of neuronal identity in Caenorhabditis elegans. Wiley Interdiscip Rev Dev Biol, 5, 474-98.

HOBERT, O. 2016b. 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.

IWAHARA, J. & CLUBB, R. T. 1999. Solution structure of the DNA binding domain from Dead ringer, a sequence-specific AT-rich interaction domain (ARID). EMBO J, 18, 6084-94.

IWAHARA, J., IWAHARA, M., DAUGHDRILL, G. W., FORD, J. & CLUBB, R. T. 2002. The structure of the Dead ringer-DNA complex reveals how AT-rich interaction domains

(ARIDs) recognize DNA. EMBO J, 21, 1197-209.

JOHNSTON, R. J., JR. 2013. Lessons about terminal differentiation from the specification of color-detecting photoreceptors in the Drosophila retina. Ann N Y Acad Sci, 1293, 33-44.

KADKHODAEI, B., ITO, T., JOODMARDI, E., MATTSSON, B., ROUILLARD, C., CARTA, M., MURAMATSU, S., SUMI-ICHINOSE, C., NOMURA, T., METZGER, D., CHAMBON, P., LINDQVIST, E., LARSSON, N. G., OLSON, L., BJORKLUND, A., ICHINOSE, H. & PERLMANN, T. 2009. Nurr1 is required for maintenance of maturing and adult midbrain dopamine neurons. J Neurosci, 29, 15923-32.

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.

KORTSCHAK, R. D., TUCKER, P. W. & SAINT, R. 2000. ARID proteins come in from the desert. Trends Biochem Sci, 25, 294-9.

KOSHO, T., MIYAKE, N. & CAREY, J. C. 2014. Coffin-Siris syndrome and related disorders involving components of the BAF (mSWI/SNF) complex: historical review and recent advances using next generation sequencing. Am J Med Genet C Semin Med Genet, 166C, 241-51.

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.

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.

LESTARI, W., ICHWAN, S. J., OTSU, M., YAMADA, S., ISEKI, S., SHIMIZU, S. & IKEDA, M. A. 2012. Cooperation between ARID3A and p53 in the transcriptional activation of p21WAF1 in response to DNA damage. Biochem Biophys Res Commun, 417, 710-6.

LI, Y. & KRATSIOS, P. 2021. Transgenic reporter analysis of ChIP-Seq-defined enhancers identifies novel target genes for the terminal selector UNC-3/Collier/Ebf. MicroPubl Biol, 2021.

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.

LIN, C., SONG, W., BI, X., ZHAO, J., HUANG, Z., LI, Z., ZHOU, J., CAI, J. & ZHAO, H. 2014. Recent advances in the ARID family: focusing on roles in human cancer. Onco Targets Ther, 7, 315-24.

LLORET-FERNANDEZ, C., MAICAS, M., MORA-MARTINEZ, C., ARTACHO, A., JIMENO-MARTIN, A., CHIRIVELLA, L., WEINBERG, P. & FLAMES, N. 2018. A transcription factor collective defines the HSN serotonergic neuron regulatory landscape. Elife, 7.

MARQUES, F., SARO, G., LIA, A. S., POOLE, R. J., FALQUET, L. & GLAUSER, D. A. 2019. Identification of avoidance genes through neural pathway-specific forward optogenetics. PLoS Genet, 15, e1008509.

MATHELIER, A., ZHAO, X., ZHANG, A. W., PARCY, F., WORSLEY-HUNT, R., ARENILLAS, D. J., BUCHMAN, S., CHEN, C. Y., CHOU, A., IENASESCU, H., LIM, J., SHYR, C., TAN, G., ZHOU, M., LENHARD, B., SANDELIN, A. & WASSERMAN, W. W. 2014. JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. Nucleic Acids Res, 42, D142-7.

MI, H., MURUGANUJAN, A., CASAGRANDE, J. T. & THOMAS, P. D. 2013. Large-scale

gene function analysis with the PANTHER classification system. Nat Protoc, 8, 1551-66.

MIYAKE, N., TSURUSAKI, Y. & MATSUMOTO, N. 2014. Numerous BAF complex genes are mutated in Coffin-Siris syndrome. Am J Med Genet C Semin Med Genet, 166C, 257-61.

NELSON, C. E., HERSH, B. M. & CARROLL, S. B. 2004. The regulatory content of intergenic DNA shapes genome architecture. Genome Biol, 5, R25.

NITTA, K. R., JOLMA, A., YIN, Y., MORGUNOVA, E., KIVIOJA, T., AKHTAR, J., HENS, K., TOIVONEN, J., DEPLANCKE, B., FURLONG, E. E. & TAIPALE, J. 2015. Conservation of transcription factor binding specificities across 600 million years of bilateria evolution. Elife, 4.

PATSIALOU, A., WILSKER, D. & MORAN, E. 2005. DNA-binding properties of ARID family proteins. Nucleic Acids Res, 33, 66-80.

PENG, G. H., AHMAD, O., AHMAD, F., LIU, J. & CHEN, S. 2005. The photoreceptor-specific nuclear receptor Nr2e3 interacts with Crx and exerts opposing effects on the transcription of rod versus cone genes. Hum Mol Genet, 14, 747-64.

POPOWSKI, M., TEMPLETON, T. D., LEE, B. K., RHEE, C., LI, H., MINER, C., DEKKER, J. D., ORLANSKI, S., BERGMAN, Y., IYER, V. R., WEBB, C. F. & TUCKER, H. 2014. Bright/Arid3A acts as a barrier to somatic cell reprogramming through direct regulation of Oct4, Sox2, and Nanog. Stem Cell Reports, 2, 26-35.

RATLIFF, M. L., TEMPLETON, T. D., WARD, J. M. & WEBB, C. F. 2014. The Bright Side of Hematopoiesis: Regulatory Roles of ARID3a/Bright in Human and Mouse Hematopoiesis. Front Immunol, 5, 113.

RHEE, C., LEE, B. K., BECK, S., ANJUM, A., COOK, K. R., POPOWSKI, M., TUCKER, H. O. & KIM, J. 2014. Arid3a is essential to execution of the first cell fate decision via direct embryonic and extraembryonic transcriptional regulation. Genes Dev, 28, 2219-32.

SAADAT, K., LESTARI, W., PRATAMA, E., MA, T., ISEKI, S., TATSUMI, M. & IKEDA, M. A. 2021. Distinct and overlapping roles of ARID3A and ARID3B in regulating E2Fdependent

transcription via direct binding to E2F target genes. Int J Oncol, 58.

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.

SHAHAM, S. & BARGMANN, C. I. 2002. Control of neuronal subtype identity by the C. elegans ARID protein CFI-1. Genes Dev, 16, 972-83.

SHANDALA, T., KORTSCHAK, R. D., GREGORY, S. & SAINT, R. 1999. The Drosophila dead ringer gene is required for early embryonic patterning through regulation of argos and buttonhead expression. Development, 126, 4341-9.

SHANDALA, T., KORTSCHAK, R. D. & SAINT, R. 2002. The Drosophila retained/dead ringer gene and ARID gene family function during development. Int J Dev Biol, 46, 423-30.

SHANDALA, T., TAKIZAWA, K. & SAINT, R. 2003. The dead ringer/retained transcriptional regulatory gene is required for positioning of the longitudinal glia in the Drosophila embryonic CNS. Development, 130, 1505-13.

SHANG, L., CHO, M. T., RETTERER, K., FOLK, L., HUMBERSON, J., ROHENA, L., SIDHU, A., SALIGANAN, S., IGLESIAS, A., VITAZKA, P., JUUSOLA, J., O'DONNELL-LURIA, A. H., SHEN, Y. & CHUNG, W. K. 2015. Mutations in ARID2 are associated with intellectual disabilities. Neurogenetics, 16, 307-14.

SIBBONS, J. 2004. Identifying the roles of dead ringer in the Drosophila eye Ph.D., The University of Adelaide.

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.

SMITH, J. A., HOLDEN, K. R., FRIEZ, M. J., JONES, J. R. & LYONS, M. J. 2016. A novel familial autosomal dominant mutation in ARID1B causing neurodevelopmental delays, short stature, and dysmorphic features. Am J Med Genet A, 170, 3313-3318.

VALENTINE, S. A., CHEN, G., SHANDALA, T., FERNANDEZ, J., MISCHE, S., SAINT, R. & COUREY, A. J. 1998. Dorsal-mediated repression requires the formation of a multiprotein repression complex at the ventral silencer. Mol Cell Biol, 18, 6584-94.

WEBB, C. F., BRYANT, J., POPOWSKI, M., ALLRED, L., KIM, D., HARRISS, J., SCHMIDT, C., MINER, C. A., ROSE, K., CHENG, H. L., GRIFFIN, C. & TUCKER, P. W. 2011. The ARID family transcription factor bright is required for both hematopoietic stem cell and B lineage development. Mol Cell Biol, 31, 1041-53.

WEBB, C. F., SMITH, E. A., MEDINA, K. L., BUCHANAN, K. L., SMITHSON, G. & DOU, S. 1998. Expression of bright at two distinct stages of B lymphocyte development. J Immunol, 160, 4747-54.

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.

WILSKER, D., PATSIALOU, A., DALLAS, P. B. & MORAN, E. 2002. ARID proteins: a diverse family of DNA binding proteins implicated in the control of cell growth, differentiation, and development. Cell Growth Differ, 13, 95-106.

WILSKER, D., PROBST, L., WAIN, H. M., MALTAIS, L., TUCKER, P. W. & MORAN, E. 2005. Nomenclature of the ARID family of DNA-binding proteins. Genomics, 86, 242-51.

ZHANG, F., BHATTACHARYA, A., NELSON, J. C., ABE, N., GORDON, P., LLORET-FERNANDEZ, C., MAICAS, M., FLAMES, N., MANN, R. S., COLON-RAMOS, D. A. & HOBERT, O. 2014. The LIM and POU homeobox genes ttx-3 and unc-86 act as terminal selectors in distinct cholinergic and serotonergic neuron types. Development, 141, 422-35.

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.

ZHENG, C., JIN, F. Q., TRIPPE, B. L., WU, J. & CHALFIE, M. 2018a. Inhibition of cell fate repressors secures the differentiation of the touch receptor neurons of Caenorhabditis elegans. Development, 145.

ZHENG, W., YANG, X., HU, R., CAI, R., HOFMANN, L., WANG, Z., HU, Q., LIU, X., BULKLEY, D., YU, Y., TANG, J., FLOCKERZI, V., CAO, Y., CAO, E. & CHEN, X. Z. 2018b. Hydrophobic pore gates regulate ion permeation in polycystic kidney disease 2 and 2L1 channels. Nat Commun, 9, 2302.

CHAPTER 3

ESTABLISHMENT AND MAINTENANCE OF MOTOR NEURON IDENTITY VIA TEMPORAL MODULARITY IN TERMINAL SELECTOR FUNCTION

This Chapter is a full reprint of Li et al., *Elife*, in which I am the primary author. The work is included with permission from all authors.

Relevant Publication

Li, Y., Osuma, A., Correa, E., Okebalama, M.A., Dao, P., Gaylord, O., Aburas, J., Islam, P., Brown, A.E. and Kratsios, P., 2020. "**Establishment and maintenance of motor neuron identity via temporal modularity in terminal selector function.**" *Elife*, 9, p.e59464.

For tables and supplementary files, please refer to `https://doi.org/10.7554/eLife.59464`

3.1 Abstract

Terminal selectors are transcription factors (TFs) that establish during development and maintain throughout life post-mitotic neuronal identity. We previously showed that UNC-3/Ebf, the terminal selector of *C. elegans* cholinergic motor neurons (MNs), acts indirectly to prevent alternative neuronal identities (Feng et al., 2020). Here, we globally identify the direct targets of UNC-3. Unexpectedly, we find that the suite of UNC-3 targets in MNs is modified across different life stages, revealing "temporal modularity" in terminal selector function. In all larval and adult stages examined, UNC-3 is required for continuous expression of various protein classes (e.g., receptors, transporters) critical for MN function. However, only in late larvae and adults, UNC-3 is required to maintain expression of MN-specific TFs. Minimal disruption of UNC-3's temporal modularity via genome engineering affects locomotion. Another *C. elegans* terminal selector (UNC-30/Pitx) also exhibits tem-

poral modularity, supporting the potential generality of this mechanism for the control of neuronal identity.

3.2 Introduction

Nervous system development is a multi-step process that culminates in the generation of distinct neuron types necessary for animal behavior. Seminal studies in many model systems have begun to elucidate the molecular mechanisms that control the early steps of neuronal development, such as specification of neural progenitors and generation of post-mitotic neurons (Catela and Kratsios, 2019, Doe, 2017, Greig et al., 2013, Jessell, 2000, Lodato and Arlotta, 2015, Perry et al., 2017). However, our understanding of the final steps of neuronal development is very rudimentary. Once neurons become post-mitotic, how do they acquire their unique functional features, such as neurotransmitter synthesis, electrical activity and signaling properties? And, perhaps more importantly, how do neurons maintain such features throughout post-embryonic life?

Terminal selectors represent one class of transcription factors (TFs) with continuous expression – from development through adulthood – in specific neuron types (Hobert, 2008, Hobert, 2011, Hobert, 2016b, Garcia-Bellido, 1975). A defining feature of terminal selectors is the ability to activate the expression of effector genes, whose protein products determine the terminally differentiated state, and thereby function, of a given neuron type. Such effector genes, herein referred to as "terminal identity genes", are expressed continuously in specific neuron types and endow them with distinct functional and phenotypic properties. Examples include neurotransmitter (NT) biosynthesis components, NT receptors, ion channels, neuropeptides, and adhesion molecules (Hobert, 2008, Hobert, 2011, Hobert, 2016b). Numerous studies support the idea that terminal selectors establish during development and

maintain throughout life neuronal identity (and function) by activating expression of terminal identity genes (Deneris and Hobert, 2014, Hobert, 2008, Hobert, 2016b, Hobert and Kratsios, 2019). Multiple cases of terminal selectors have been described thus far in worms, flies, chordates and mice (Deneris and Hobert, 2014, Hobert and Kratsios, 2019, Konstantinides et al., 2018), suggesting high conservation of this type of regulators. Importantly, human mutations in terminal selectors and their effector target genes have been linked to either developmental or degenerative conditions of the nervous system (Deneris and Hobert, 2014).

However, the molecular mechanisms through which terminal selectors establish and maintain neuronal identity are poorly understood, in part due to two major challenges. First, the majority of studies follow a candidate approach focused on a select subset of terminal identity genes (Flames and Hobert, 2009, Hobert, 2016a, Lopes et al., 2012). Hence, the extent of terminal identity features and breadth of biological processes controlled by terminal selectors remain largely unknown. Addressing this knowledge gap requires unbiased methods for the identification of terminal selector target genes, but such approaches have only been applied to a limited number of terminal selectors to date (Corbo et al., 2010, Housset et al., 2013, Kadkhodaei et al., 2013, Wyler et al., 2016, Yu et al., 2017). Second, the continuous expression of terminal selectors represents an additional challenge because it is not known whether their function - in a particular neuron type - remains the same, or changes at different life stages. This is partly due to the difficulty of tracking individual neurons in the complex vertebrate nervous system throughout embryonic and post-natal life. Hence, longitudinal studies in simple model organisms are needed to determine whether terminal selectors control an identical suite of target genes across different stages (e.g., embryo, adult), or whether the suite of targets can change over time. Addressing these two challenges may extend our knowledge of how terminal selectors control neuronal identity, as well as advance

our understanding of how cellular identity is established and maintained.

This study focuses on UNC-3, the sole *C. elegans* ortholog of the Collier/Olf/Ebf (COE) family of TFs (Dubois and Vincent, 2001). UNC-3 acts as a terminal selector in cholinergic motor neurons (MNs) of the *C. elegans* ventral nerve cord (Kratsios et al., 2012). Importantly, mutations in EBF3, a human ortholog of UNC-3, cause a neurodevelopmental syndrome characterized by motor developmental delay (Blackburn et al., 2017, Chao et al., 2017, Harms et al., 2017, Sleven et al., 2017). A previous study proposed that UNC-3 controls cholinergic MN identity in *C. elegans* by activating directly the expression of various terminal identity genes (e.g., acetylcholine biosynthesis components, ion channels, NT receptors, neuropeptides), which were identified via a candidate approach (Kratsios et al., 2012). More recently, it was demonstrated that UNC-3 can also act indirectly to prevent the adoption of alternative neuronal identities (Feng et al., 2020). Lastly, animals lacking *unc-3* gene activity display severe locomotion defects (Brenner, 1974, Feng et al., 2020), suggesting UNC-3 may broadly control gene expression in cholinergic MNs. However, an unbiased identification of UNC-3 targets, as well as a longitudinal analysis of *unc-3* mutants are currently lacking.

Here, we employ chromatin immunoprecipitation followed by DNA sequencing (ChIP-Seq) and report the identification of ∼3,500 protein-coding genes as putative direct targets of UNC-3. Protein class ontology analysis suggests that UNC-3, besides terminal identity genes, also controls additional biological processes, such as neuronal metabolism and downstream gene regulatory networks comprised of numerous TFs and nucleic acid-binding proteins. These findings help obtain a comprehensive understanding of terminal selector function.

Through a longitudinal analysis of *unc-3* mutants at embryonic, larval and adult stages, we identified two groups of target genes with distinct temporal requirements for UNC-3 in cholinergic MNs. One group encodes multiple classes of proteins (e.g., receptors, secreted molecules, TFs) that require UNC-3 for both embryonic initiation and post-embryonic maintenance of their expression. Contrasting this stable mode of regulation over time, a second group of targets consists exclusively of TFs (*cfi-1/Arid3a, bnc-1/BNC1-2, mab-9/Tbx20, ceh-44/CUX1-2, nhr-40/nuclear hormone receptor*) that do not require UNC-3 for initiation, but depend on UNC-3 activity for maintenance. Hence, the suite of UNC-3 targets in cholinergic MNs is modified across different life stages, a phenomenon we term "temporal modularity" in terminal selector function. To provide mechanistic insights, we focused on the second group of targets and identified a molecular mechanism for their *unc-3*-independent initiation that relies on Hox proteins. Importantly, preventing UNC-3's ability to selectively maintain expression of a single TF (*cfi-1/Arid3a*) from the second target group led to locomotion defects, indicating minimal disruption of temporal modularity affects animal behavior. Lastly, we provide evidence for temporal modularity in the function of UNC-30/PITX, the terminal selector of GABAergic MN identity in *C. elegans* (Eastman et al., 1999, Jin et al., 1994). Because terminal selectors have been identified in both invertebrate and vertebrate nervous systems (Deneris and Hobert, 2014, Hobert and Kratsios, 2019), we hypothesize that temporal modularity in their function may be a general mechanism for the establishment and maintenance of neuronal identity.

3.3 Materials and Methods

3.3.1 C. elegans strain culture

Worms were grown at 20℃ or 25℃ on nematode growth media (NGM) plates supplied with *E. coli* OP50 as food source (Brenner, 1974). A list of all *C. elegans* strains used is provided in Key Resources Table, which is available as supplementary data at `https://doi.org/10.7554/eLife.59464`.

3.3.2 Generation of transgenic animals carrying transcriptional fusion reporters

Reporter gene fusions for *cis*-regulatory analyses and validation of newly identified UNC-3 target genes were made with PCR fusion (Hobert, 2002). Genomic regions were amplified and fused to the coding sequence of tagrfp followed by the *unc-54* 3' UTR. To mutate the LIN-39 binding motif, the reporter fusion was first introduced into the pCR-XL-TOPO vector using the TOPO XL PCR cloning kit (Invitrogen). Then, mutagenesis PCR was performed, and single clones containing plasmids that carry the desired mutation were isolated. PCR fusion DNA fragments were injected into young adult *pha-1(e2123)* hermaphrodites at 50 ng/µl together with *pha-1* (pBX plasmid) as co-injection marker (50 ng/µl).

3.3.3 Chromatin Immunoprecipitation (ChIP)

ChIP assay was performed as previously described (Yu et al. 2017; Zhong et al. 2010) with the following modifications. Synchronized L1 *unc-3 (ot839[unc-3::gfp])* worms and N2 worms were cultured on 10 cm plates seeded with OP50 at 20°C overnight. Late L2 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-Lauroyl sarsosine sodium, 2 tablets of cOmplete ULTRA Protease Inhibitor Cocktail [Roche Cat.# 05892970001] in 10ml FA buffer). For each IP experiment, 200 µl worm pellet was collected. 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 1 min. 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. The remainder was incubated with 2 µl GFP antibody (Abcam Cat.# ab290) at 4°C overnight. Wild-type (N2) worms do not carry the GFP tag and serve as negative control. The *unc-3 (ot839[unc-3::gfp])* CRIPSR generated allele was used in order to immunoprecipitate the endogenous UNC-3 protein. On the next day, 20 µl Dynabeads Protein G (1004D) was added to the immunocomplex, which was then incubated for 2 hr at 4°C. The beads then were washed at 4°C twice with 150 mM NaCl FA buffer (5 min each), and 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, and 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 then 65°C overnight (12-20 hours) to reverse cross-link. The immonuprecipitated DNA was purified with Ampure XP beads (A63881) according to manufacturer's instructions.

3.3.4 ChIP-sequencing data analysis

Unique reads were mapped to the *C. elegans* genome (ce10) with bowtie2 (Langmead and Salzberg, 2012). Peak calling was then performed with MACS2 (minimum q-value cutoff for peak detection: 0.005). For visualization purposes, the sequencing depth was normalized to 1x genome coverage using bamCoverage provided by deepTools (Ramirez et al., 2016), and peak signals were shown in IGV. Heatmap of peak coverage in regard to UNC-3 enrichment center was generated with NGSplot (Shen et al., 2014). The average profile of peaks binding to TSS region was generated with ChIPseeker (Yu et al., 2015). To study the distribution of peaks genome-wide, the peaks were annotated using annotatePeaks.pl provided by Homer (Heinz et al., 2010), and each peak was assigned to a gene with the nearest TSS. For de novo motif discovery, sequences containing 100bp around the centers of each peak (from -50bp to +50bp) were extracted and supplied to findMotifsGenome.pl provided by Homer.

3.3.5 Protein Class Ontology analysis using PANTHER

Protein Class Ontology analysis was performed on 1,478 UNC-3-bound genes out of the 3,502 protein-coding genes. The number of genes is significantly lower than the number of peaks because Panther analysis only considers genes with known protein class terms.

3.3.6 Targeted genome editing

The *cfi-1* endogenous reporter strain *kas16 [mNG::AID::cfi-1]* was generated by employing CRISPR/Cas9 genome editing, inserting the *mNG::3xFLAG::AID* cassette immediately after the ATG of *cfi-1*. The *cfi-1* enhancer knock-out allele *mNG::AID::cfi-1* $^{\Delta\ enhancer\ (769\ bp)}$ was generated by using two guide RNAs flanking the *cfi-1* enhancer to guide excision of the genomic region, which was then followed by homology dependent repair (HDR) to create

a 769 bp deletion (-11,329 bp to -12,097 bp). The UNC-3 binding motif mutation allele *mNG::AID::cfi-1*$^{8\ COE\ motifs\ mut}$ was generated by creating nucleotide substitutions in the repair template, which carries homology arms complementary to the *cfi-1* enhancer region and is then introduced into the genome through HDR.

3.3.7 Microscopy

Imaging slides were prepared by anesthetizing worms with sodium azide (NaN_3, 100 mM) and mounting them on a 4% agarose pad on glass slides. Images were taken with an automated fluorescence microscope (Zeiss, Axio Imager Z2). Images containing several z stacks (0.50 µm intervals between stacks) were 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 2-5 µm Z-stacks using the maximum intensity projection type. Image reconstruction was performed with Image J (Schindelin et al., 2012).

3.3.8 Motor neuron subtype identification

The identification of specific MN subtypes expressing a given UNC-3 target gene was assessed based on the following: (a) co-localization with fluorescent reporters that label specific MN subtypes; (b) Invariant position of neuronal cell bodies along the ventral nerve cord; (c) Birth order of specific motor neuron subtypes (e.g. during embryogenesis or post-embryogenesis); (d) Total cell numbers in each motor neuron subtype.

3.3.9 Bioinformatic prediction of binding motifs

Information of the LIN-39 binding motif is curated in the Catalog of Inferred Sequence Binding Preferences database (`http://cisbp.ccbr.utoronto.ca`). To predict and identify LIN-39 binding motifs and UNC-3 binding motifs (identified in this paper) in the *cfi-1* enhancer (-11,391 bp to -12,146 bp), we utilized tools provided by MEME (Multiple Expectation maximization for Motif Elicitation) bioinformatics suite (`http://meme-suite.org/`), and performed FIMO (Find Individual Motif Occurrences) motif scanning analysis.

3.3.10 Temporally controlled protein degradation

Temporally controlled protein degradation was achieved with the auxin-inducible degradation system (Zhang et al., 2015). TIR1 expression was driven by the ubiquitously active *eft-3* promoter in the transgene *ieSi57 [eft-3prom::tir1]*, or a transgene that drives TIR1 selectively in neurons (*otTi28*). To induce degradation of proteins (LIN-39, CFI-1, UNC-3), the following alleles were used: *lin-39 (kas9 [lin-39::mNG::AID])*, *cfi-1 (kas16 [mNG::AID::cfi-1])*, and *unc-3 (ot837 [unc-3::mNG::AID])*. Worms at specific developmental stages (see figure legends for details) were grown at 20℃ on NGM plates coated with 4 nM auxin (indole-3-acetic acid [IAA] dissolved in ethanol) or ethanol (negative control) for 1 day or 4 days before tested (see figure legends for more details). All plates were shielded from light.

3.3.11 Worm tracking

Worms were maintained as mixed stage populations by chunking on NGM plates with *E. coli* OP50 as the food source. Worms were bleached and the eggs were allowed to hatch in M9 buffer to arrest as L1 larvae. L1 larvae were refed on OP50 and allowed to grow to day 2 or adulthood. On the day of tracking, five worms were 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 4℃ for at least two days before use. Imaging plates are seeded with 50 µl 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.

3.3.12 Behavioral feature extraction and analysis

All videos were analyzed using Tierpsy Tracker to extract each worm's position and posture over time (Javer et al., 2018a). These postural data were then converted into a set of behavioral features selected from a large set of features as previously described (Javer et al., 2018b). 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 (Benjamini, 2001).

3.3.13 Statistical analysis

For data quantification, graphs show values expressed as mean ± standard error mean (SEM) of animals. The statistical analyses were performed using the unpaired t-test (two-tailed). Calculations were performed using the GraphPad QuickCalcs online software (`http://www.graphpad.com/quickcalcs/`). Differences with $p < 0.05$ were considered significant.

Data accessibility

The accession number for the UNC-3 ChIP-seq data is GEO: GSE143165.

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 the lab of Oliver Hobert for providing the OH14930 strain and Kaiyuan Tang for generating reporter gene constructs. We are grateful to Edwin Ferguson, Oliver Hobert, Daniele Canzio, Catarina Catela, and Weidong Feng for comments on this manuscript. This work was supported by a training grant [University of Chicago Initiative for Maximizing Student Development (IMSD), 2R25GM109439-06] to A.O, a training grant (T32 GM007183) to E. C., a Whitehall Foundation grant to P.K., grants from National Institute of Neurological Disorders and Stroke (NINDS) of the NIH (Award Numbers R00NS084988 and R21NS108505) to P.K., and a Medical Research Council grant MC-A658-5TY30 to A.E.X.B.

Author contributions

Y.L. and P.K. conceived the concept, designed the experiments and wrote the paper. A.O. and E.C. conducted validation of targets identified by ChIP-seq that is described in Figure 3.5. M.A.O performed imaging for part of the data in Figure 3.4B. P.D. and O.G. generated reporter constructs # 2 - # 12 described in Figure 3.7A. P.I. and A.E.B led behavioral analyses described in Figure 3.12. J.A. performed microinjections.

Conflict of interest statement

The authors declare no competing interests.

3.4 Results

3.4.1 Identifying the global targets of UNC-3 via ChIP-Seq

To identify in an unbiased manner putative UNC-3 target genes, we employed chromatin immunoprecipitation followed by DNA sequencing (ChIP-Seq). We used a reporter strain with in-frame GFP sequences inserted immediately upstream of the stop codon of the endogenous *unc-3* gene (Figure 3.1A). Expression of UNC-3::GFP fusion protein was observed in the nucleus of 53 cholinergic MNs (SAB subtype = 3 neurons, DA = 9, DB = 7, VA = 12, VB = 11, AS = 11) and 19 other neurons known to express *unc-3* (Figure 3.1A), indicating that this reporter faithfully recapitulates the endogenous *unc-3* expression pattern (Pereira et al., 2015, Prasad et al., 1998). Insertion of GFP does not detectably alter the function of UNC-3 since expression of known UNC-3 targets (*cho-1/ChT, unc-17/VAChT, ace-2/AChE*) is unaffected in *unc-3::gfp* animals (Figure 3.1B). Unlike *unc-3* null mutants, *unc-3::gfp* animals do not display locomotion defects. We therefore performed ChIP-seq on *unc-3::gfp* animals at larval stage 2 (L2) because all *unc-3* expressing neurons have been generated by this stage.

Our ChIP-seq dataset revealed strong enrichment of UNC-3 binding in the genome by identifying a total of 6,892 unique binding peaks (q-value cutoff: 0.05) (Figures 3.1C and 3.2). The majority of UNC-3 binding peaks (91.95%) are predominantly located between 0 and 3 kb upstream of a transcription start site, whereas only a small fraction is found in introns (2.05%) or downstream of a gene (0.99%), suggesting UNC-3 chiefly acts at promoters and enhancers to regulate gene expression (Figures 3.1D and 3.1F). Through *de novo* motif discovery analysis (see Materials and Methods), we identified a 12 bp pseudo-palindromic sequence overrepresented in the UNC-3 binding peaks (Figure 3.1E), which highly resembles the binding site of UNC-3 vertebrate orthologs (Treiber et al., 2010a, Treiber et al., 2010b, Wang and Reed, 1993, Wang et al., 1997). To test the quality of our ChIP-Seq results, we

sought to determine whether UNC-3 binding peaks are present in the *cis*-regulatory region of all previously described UNC-3 targets in cholinergic MNs, because these neurons constitute the majority of *unc-3*-expressing cells (53 out of 72 cells). Previous studies identified 10 terminal identity genes as putative direct UNC-3 targets and 43 terminal identity genes whose expression is affected by genetic removal of *unc-3* (Kratsios et al., 2015, Kratsios et al., 2012). In the current study, we found UNC-3 binding peaks in 9 out of 10 (90%) direct UNC-3 targets and 38 of the 43 (88.37%) downstream targets of UNC-3, indicating high quality in the ChIP-Seq results (Figures 3.3A and 3.2, Supplementary File 1). Moreover, ChIP-Seq for UNC-3 appears highly sensitive, as it identified peaks in *unc-3*-dependent genes expressed in a limited number of neurons (e.g., glr-4/GluR is expressed in 4 out of 72 *unc-3*+ neurons) (Figure 3.3A). In conclusion, our ChIP-Seq experiment generated a comprehensive map of UNC-3 binding in the *C. elegans* genome and provided biochemical evidence to the hypothesis that UNC-3 binds to the *cis*-regulatory region of multiple terminal identity genes, consolidating UNC-3's function as a terminal selector of cholinergic MN identity.

Our bioinformatic analysis of the UNC-3 binding peaks revealed 3,502 protein-coding genes as putative UNC-3 targets (see Materials and Methods). To classify these new targets, we performed gene ontology (GO) analysis focused on protein classes using PANTHER (Mi et al., 2013). This analysis revealed three broad categories (Figure 3.3B, Supplementary File 2). First, there is a preponderance (42.18% of the total number of UNC-3 targets classified by PANTHER) of terminal identity genes (e.g., 114 transporter proteins, 111 receptors and trans-membrane proteins, 37 signaling molecules, 11 cell adhesion molecules), suggesting that UNC-3 broadly affects multiple features of neuronal terminal identity. The second overrepresented category (24.07% of UNC-3 targets) contains a large number of proteins involved in the control of gene expression, such as 239 nucleic acid binding proteins (16.77%) and 104 TFs (7.3%), highlighting the possibility of an extensive network of gene regulatory

factors downstream of UNC-3. The third category (24.14%) consists of genes coding for various types of enzymes (e.g., hydrolases, ligases, oxidoreductases), suggesting a new role for UNC-3 in neuronal metabolic pathways. Together, this analysis unravels the breadth of biological processes potentially under the direct control of UNC-3. The ChIP-Seq experiment was performed on endogenously tagged UNC-3, which is expressed in 53 cholinergic MNs of the nerve cord and 19 other neurons located in the *C. elegans* head and tail (Pereira et al., 2015). Since MNs are the majority (53 cells) of *unc-3*-expressing cells (72 in total), a significant portion of the UNC-3-bound genes may be expressed in MNs. To test this, we used available single-cell RNA-Seq data (CeNGEN project: `www.cengen.org`) that identified 576 transcripts specifically enriched in cholinergic MNs (Seth R Taylor, 2019). We found that 52.95% of these MN-expressed genes are bound by UNC-3 and fall in the aforementioned gene categories (Supplementary File 3), thereby constituting putative UNC-3 targets in cholinergic MNs.

3.4.2 cis-regulatory analysis reveals novel TFs as direct UNC-3 targets in motor neurons

Our ChIP-Seq results provide an opportunity to reveal new roles for UNC-3, beyond the direct control of terminal identity genes. To this end, we focused on the 104 TFs identified by ChIP-Seq as putative UNC-3 targets (Figure 3.3B-C). To functionally test whether the UNC-3 bound DNA elements upstream of these TF-encoding genes carry information critical for gene expression, we carried out a *cis*-regulatory analysis. We isolated and fused to RFP elements located upstream of 16 randomly selected TFs of different families (e.g., homeobox, nuclear hormone receptors, Zn finger) (Figure 3.3C, Table 1). We generated transgenic reporter animals and found that 10 of these TF reporters were sufficient to drive RFP expression in ventral cord MNs (Table 1). Next, we identified 9 TFs (*nhr-1, nhr-40,*

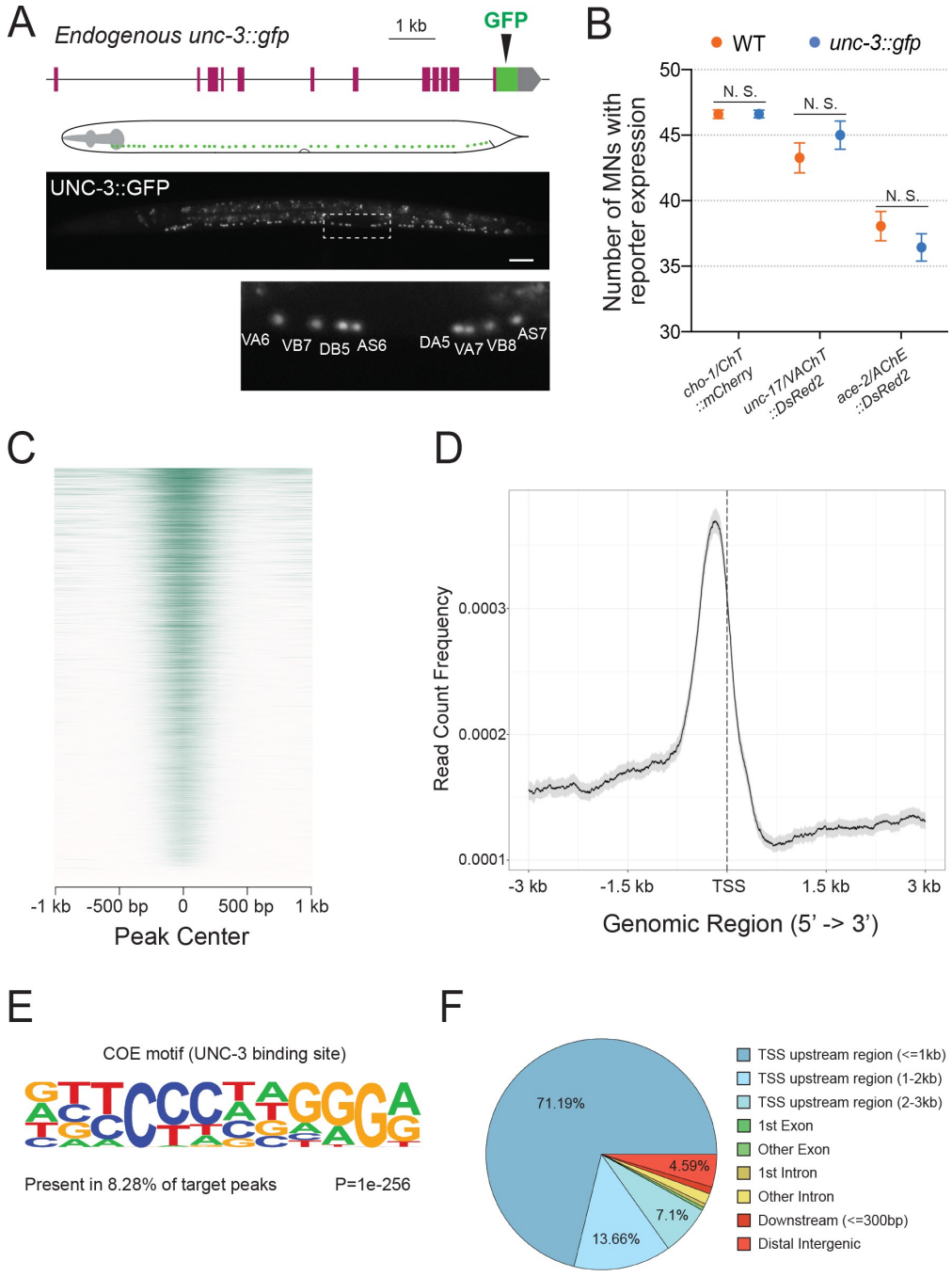


Figure 3.1: **Mapping UNC-3 binding genome-wide with ChIP-Seq.**
A, Diagram illustrating the endogenous reporter allele of UNC-3. GFP is inserted immediately upstream of *unc-3*'s stop codon. Below, a representative image at L2 stage showing expression of UNC-3::GFP fusion protein in cholinergic MN nuclei. Region highlighted in dashed box is enlarged. Scale bar, 20 µm.
B, Quantification of terminal identity gene markers that report expression of known UNC-3 targets (*cho-1/ChT, unc-17/VAChT, ace-2/AChE*) in WT and *ot839 [unc-3::gfp]* animals at the L4 stage (N = 15). N. S.: not significant.

Figure 3.1, continued.
C, Heatmap of UNC-3 ChIP-Seq signal around 1.0 kb of the center of the binding peak.
D, Summary plot of UNC-3 ChIP-Seq signal with a 95% confidence interval (grey area) around 3.0 kb of the TSS. The average signal peak is detected at ∼200 bp upstream of the TSS.
E, *de novo* motif discovery analysis of 6,892 UNC-3 binding peaks identifies a 12 bp-long UNC-3 binding motif.
F, Pie chart summarizes genomic distribution of UNC-3 ChIP-Seq signal

mab-9, ztf-26, ceh-44, zfh-2, cfi-1, bnc-1, nhr-19) with expression in cholinergic MNs. Some of those are also expressed in the *unc-3*-negative GABAergic MNs of the nerve cord (Table 1). Interestingly, one reporter (*nhr-49*) is exclusively expressed in GABAergic MNs. Expression of five TFs (*nhr-1, nhr-19, nhr-49, zfh-2, ztf-26*) in ventral cord MNs has not been previously described. The remaining 5 TFs (*cfi-1/Arid3a, bnc-1/Bnc1/2, mab-9/Tbx20, nhr-40, ceh-44/Cux1*) are known to be expressed in subsets of *unc-3*-positive MNs (Kerk et al., 2017, Pocock et al., 2008, Brozova et al., 2006), but our analysis revealed *cis*-regulatory elements sufficient for their MN expression. Next, we tested for *unc-3* dependency at the L4 stage, and found that 8 of the 9 TF reporters (*nhr-1, nhr-40, zfh-2, ztf-26, ceh-44, cfi-1, bnc-1, mab-9*) with expression in cholinergic MNs are positively regulated by *unc-3*; their expression significantly decreased in *unc-3* mutants (Table 1, Figure 3.5). On the other hand, nhr-49, which is normally expressed in GABAergic MNs, is negatively regulated by *unc-3*; ectopic expression was observed in cholinergic MNs of *unc-3* mutants (Table 1, Figure 3.5). Together, this cis-regulatory analysis revealed novel TFs as direct UNC-3 targets in cholinergic MNs. Besides its known role as an activator of terminal identity genes (Kratsios et al., 2015, Kratsios et al., 2012), these findings suggest UNC-3 can also act directly to either activate or repress expression of multiple TF-encoding genes. Alternatively, UNC-3 binding could facilitate recruitment of other factors that function either as activators or repressors. Altogether, these data uncover an extensive gene regulatory network downstream of UNC-3.

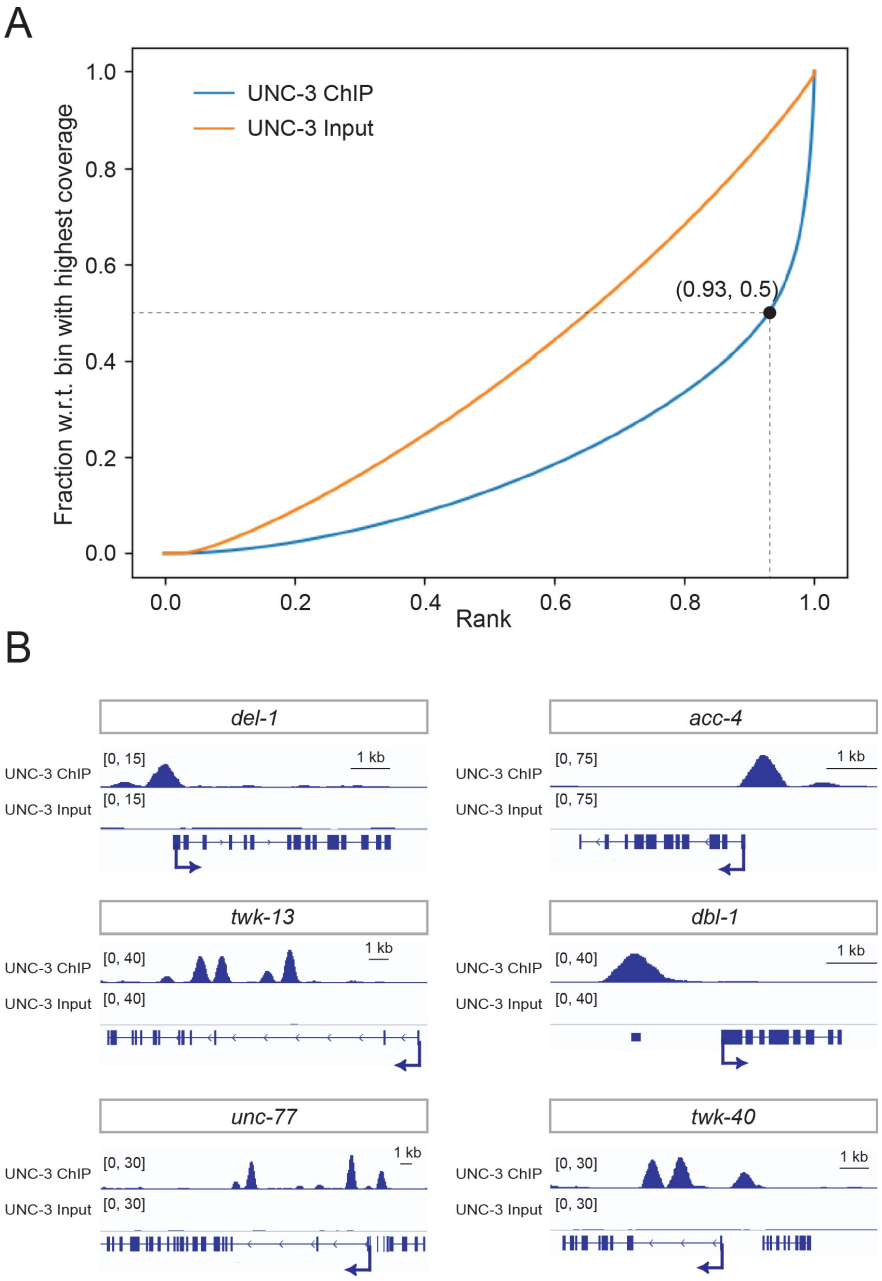


Figure 3.2: **UNC-3 ChIP-Seq results yield genome-wide enrichment of UNC-3.** **A**, Fingerprint plot indicating localized, strong enrichment of UNC-3 in the genome. Specifically, when counting the reads contained in 93% of all genomic bins (data point 0.93, 0.5 on UNC-3 ChIP curve), only 50% of the maximum number of reads are reached, which indicates 7% of the genome contains half of total sequencing reads from the ChIP sample. **B**, Snapshots of UNC-3 ChIP-Seq and input (negative control) signals at the *cis*-regulatory regions of known UNC-3 targets (*del-1, acc-4, twk-13, dbl-1, unc-77, twk-40*).

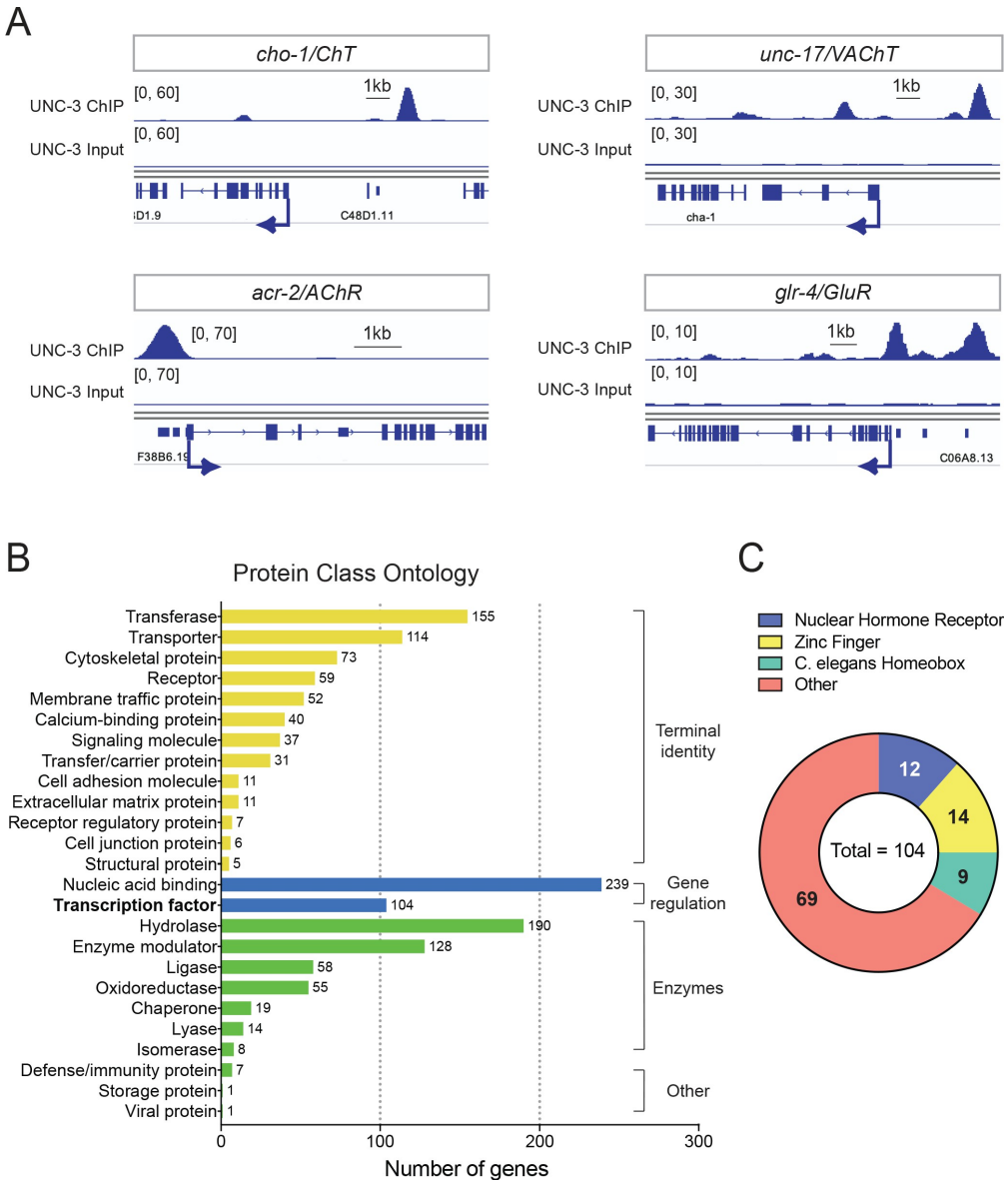


Figure 3.3: **Global analyses of UNC-3 ChIP-Seq data.**
A, Snapshots of UNC-3 ChIP-Seq and input (negative control) signals at the cis-regulatory regions of known UNC-3 targets (*cho-1/ChT, unc-17/VAChT, acr-2/AChR, glr-4/GluR*).
B, Graph summarizing protein class ontology analysis of putative target genes of UNC-3 identified by ChIP-Seq. Out of the 3,502 protein-coding UNC-3 targets, 1,425 encode for proteins with known protein class terms and these were the ones considered by PANTHER. This analysis classifies UNC-3 targets into 3 broad categories: terminal identity genes, gene expression regulators, and enzymes.
C, Pie chart breaking down TF families that show UNC-3 binding.

3.4.3 Temporal modularity of UNC-3 function in cholinergic motor neurons

Terminal selectors are continuously expressed, from development through adulthood, in specific neuron types. However, it remains unclear whether – in the same neuron type – a terminal selector controls an identical suite of targets across different life stages, or the suite of targets can change over time. The case of UNC-3 offers an opportunity to address this issue because its direct targets (terminal identity genes and newly identified TFs [Table 1]) are continuously expressed in cholinergic MNs. However, *unc-3* dependency of terminal identity genes was mostly tested at a single developmental stage, the last larval stage (L4) (Kratsios et al., 2012). We therefore performed a longitudinal analysis to determine whether target gene dependency on *unc-3* remains stable or changes at different life stages.

First, we tested 4 terminal identity genes (*acr-2/AChR, unc-129/TGFbeta, glr-4/GluR, unc-17/VAChT*) at larval (L2, L4) and adult (day 1) stages and found that their MN expression, at every stage, critically depends on UNC-3 (Figure 3.4A). Because this analysis relied on animals lacking *unc-3* gene activity since early development (a null allele was used), whether UNC-3 is continuously required to maintain expression of these genes remained unclear. We therefore employed the auxin-inducible degradation (AID) system to deplete the endogenous UNC-3 protein in cholinergic MNs at late larval and young adult stages (Zhang et al., 2015). Compared to controls, we found a significant decrease in the number of adult MNs expressing *acr-2/AChR, glr-4/GluR*, and *unc-17/VAChT* in animals treated with auxin (Figure 3.6). These findings indicate UNC-3 is required to initiate and maintain the expression of terminal identity genes, consolidating its role as a terminal selector of cholinergic MN fate.

However, a different picture emerged after testing the 8 TF reporters that are positively

regulated by UNC-3 (Table 1). Similar to terminal identity genes, the expression of 3 TFs (*zfh-2/Zfhx3, ztf-26, nhr-1*) critically depends on UNC-3 at every stage (L2, L4, adult) (Figure 3.5), suggesting UNC-3 controls initiation and maintenance of their expression. In striking contrast, the early expression (L2 stage) of five TFs (*cfi-1, bnc-1, mab-9, nhr-40, ceh-44*) does not require UNC-3 (Figure 3.4B). However, maintenance of their expression during late larval and/or adult stages does depend on UNC-3 (Figure 3.4B). Hence, this longitudinal analysis revealed two groups of targets with distinct requirements for UNC-3 at different life stages. One group consists of terminal identity genes (*acr-2/AChR, unc-129/TGFbeta, glr-4/GluR, unc-17/VAChT*) and TFs (*zfh-2/Zfhx3, ztf-26, nhr-1*) that require UNC-3 for both initiation and maintenance of expression ("initiation and maintenance" module, Figures 3.4A and 3.4C). The second group consists exclusively of TFs (*cfi-1/Arid3a, bnc-1/Bnc1, mab-9/Tbx20, ceh-44/Cux1, nhr-40*) that depend on UNC-3 for maintenance, but not initiation ("maintenance-only" module, Figures 3.4B-C).

Collectively, these findings suggest that, in cholinergic MNs, the suite of UNC-3 target genes can partially change at different life stages. In all larval and adult stages examined, UNC-3 is required for the continuous expression of one set of genes ("initiation and maintenance" module). However, only in late larvae and adults, UNC-3 is required to maintain expression of another set of genes ("maintenance-only" module). To describe this phenomenon, we use the term "temporal modularity in UNC-3 function" given that the function of a TF in a specific cell type and at a particular life stage is determined by the suite of targets it controls in that cell type and at that stage (Figure 3.4C). In the following sections, we hone in on a single target (*cfi-1/Arid3a*) from the "maintenance-only" module, aiming to dissect the molecular mechanisms underlying the temporal modularity of UNC-3 function in cholinergic MNs.

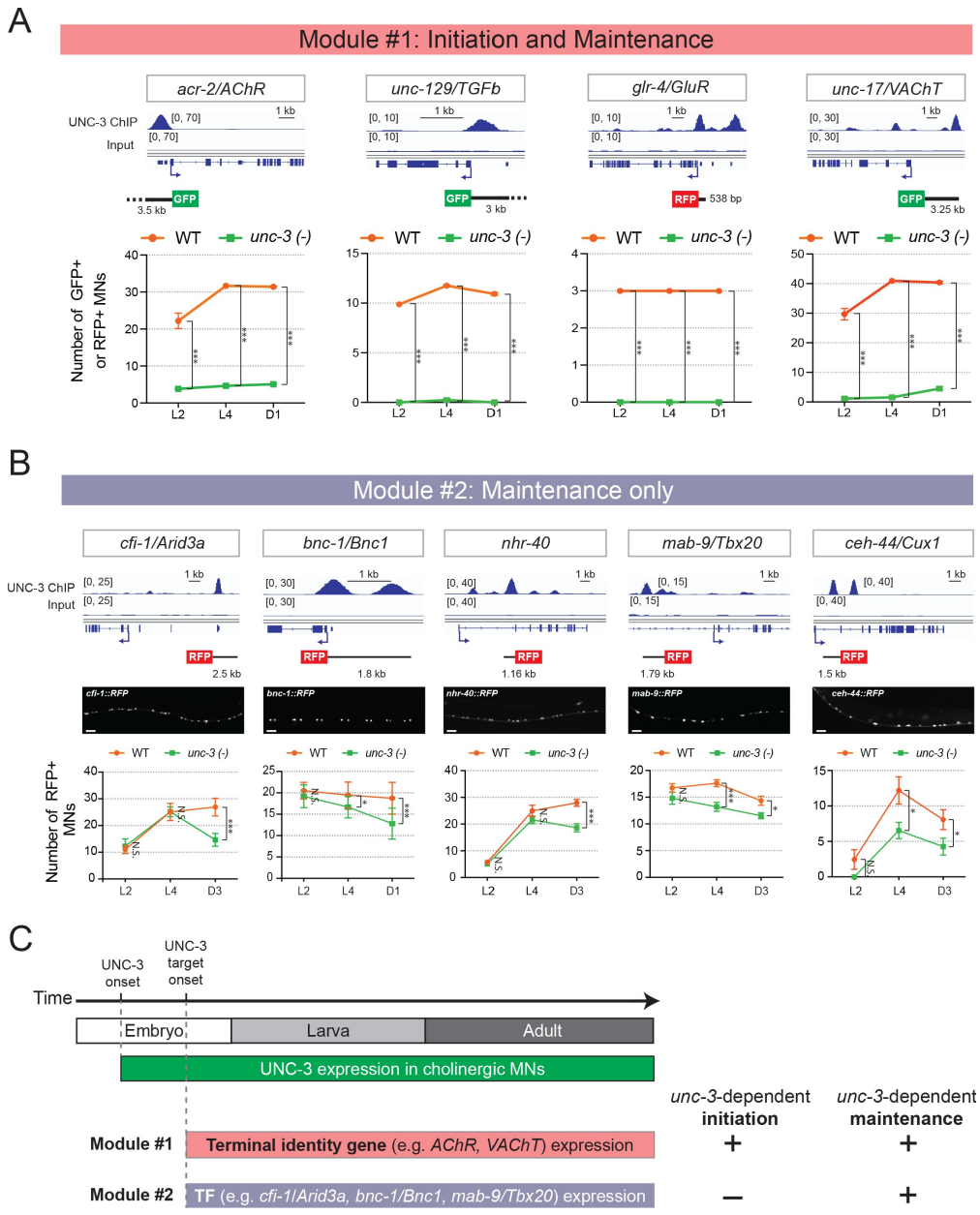


Figure 3.4: **Terminal identity genes and transcription factors display distinct temporal requirements for UNC-3.**
A, Top: snapshots of UNC-3 ChIP-Seq and input (negative control) signals at the cis-regulatory regions of 4 cholinergic terminal identity genes (*acr-2/AChR, unc-129/TFGb, glr-4/GluR*, and *unc-17/VAChT*). The length of DNA elements included in each reporter is shown. Bottom: quantification of terminal identity gene reporters in WT and *unc-3 (n3435)* animals at 3 different developmental stages – L2, L4, and day 1 adults ($N \geq 12$). UNC-3 is required for both initiation and maintenance of all 4 terminal identity genes. *** $p < 0.001$.

Figure 3.4, continued.
B, Top: snapshots of UNC-3 ChIP-Seq and input (negative control) signals at the cis-regulatory regions of 5 transcription factors (*cfi-1/Arid3a, bnc-1/Bnc1, nhr-40, mab-9/Tbx20, and ceh-44/Cux1*). The length of DNA elements included in each reporter is shown. Middle: representative images of WT L4 animals showing expression of the transgenic reporters in MNs. Scale bar, 20 µm. Bottom: quantification of transcription factor reporters in WT and *unc-3 (n3435)* animals at 3 different developmental stages – L2, L4, and young adults (day 1 or day 3) ($N \geq 12$). UNC-3 is required for maintenance, but not initiation of the 5 TFs. N.S.: not significant, * $p < 0.05$, *** $p < 0.001$.
C, Schematic summarizing the phenomenon of temporal modularity in UNC-3 function. The first module consists of terminal identity genes and TFs, which require UNC-3 for both initiation and maintenance of gene expression. The second module consists exclusively of TFs that require UNC-3 only for maintenance.

3.4.4 A distal enhancer is necessary for initiation and maintenance of cfi-1/Arid3a expression in MNs

Our *cis*-regulatory analysis suggests that maintenance, but not initiation, of *cfi-1* expression depends on UNC-3 (Figure 3.4B). We therefore hypothesized that the sole UNC-3 binding peak on the *cfi-1* locus (located ∼12 kb upstream) demarcates an enhancer element selectively required for maintenance (Figures 3.4B and 3.7A). If this were to be the case, then it would be logical to assume that a separate *cis*-regulatory element would control *cfi-1* initiation in MNs. To test this assumption, we conducted an unbiased *cis*-regulatory analysis *in vivo* by generating a series of 12 transgenic reporter (GFP or RFP) animals, with each reporter carrying small and contiguous DNA fragments spanning a ∼15 kb region (Figure 3.7A). Surprisingly, this analysis did not reveal a separate initiation element. Instead, it showed that the same 2.5 kb distal element (reporter #7) that drives RFP expression in subsets of *unc-3*-expressing cholinergic MNs (DA, DB, VA, VB subtypes) at larval and adult stages (Figure 3.4B), is also sufficient for embryonic (3-fold stage) MN expression (Figures 3.7A - C). In addition, this 2.5 kb element also showed expression at all these stages in *unc-3*-negative neurons of the nerve cord, namely the GABAergic (DD, VD subtypes) MNs (Figures 3.7A and 3.7C), which will be discussed later in Results. We conclude that this

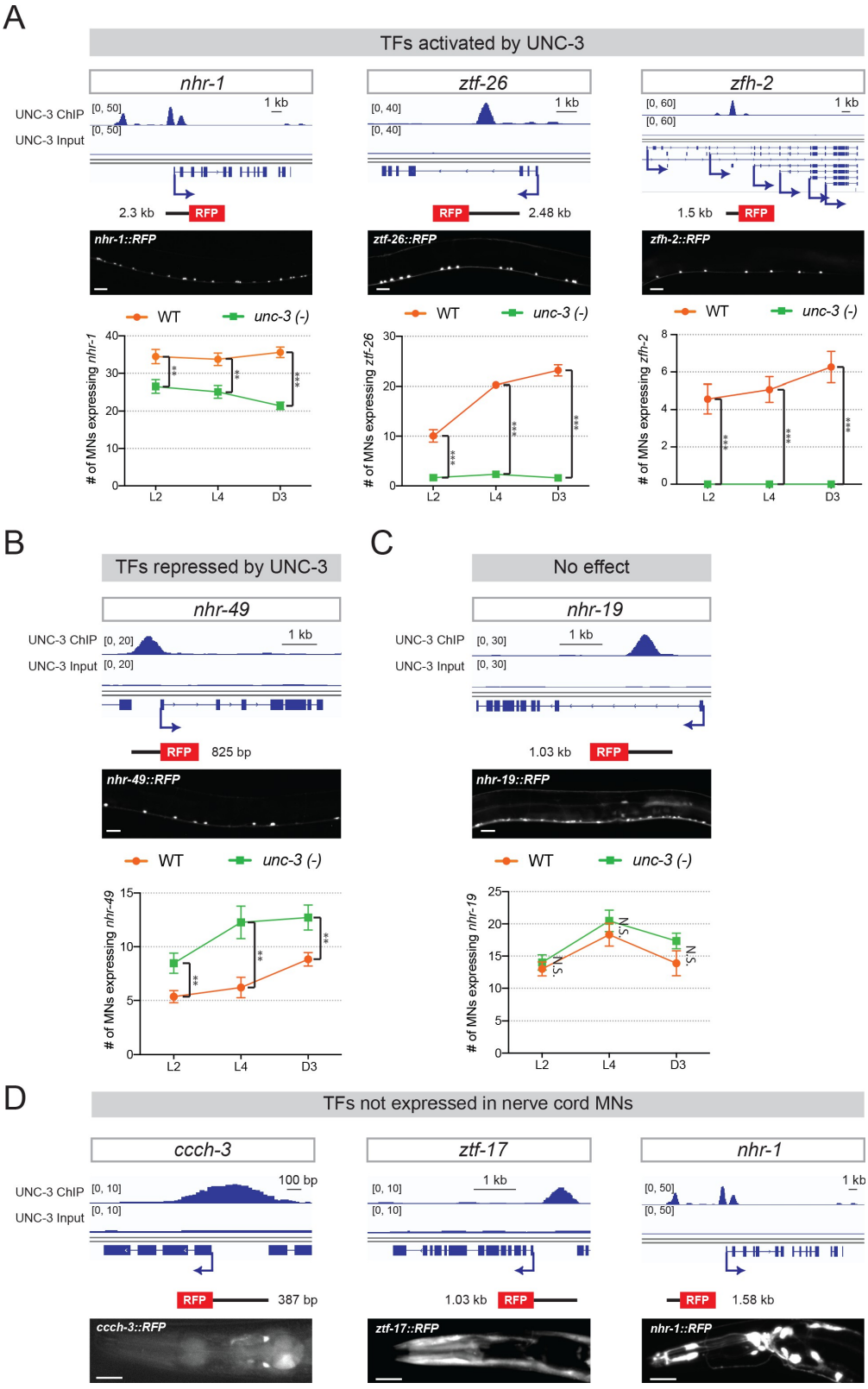


Figure 3.5: **UNC-3 directly controls the expression of several TF reporters in MNs.**

Figure 3.5, continued.
A-C, Top: snapshots of UNC-3 ChIP-Seq and input (negative control) signal at the *cis*-regulatory regions of 5 TF-encoding gens (*nhr-1, ztf-26, zfh-2, nhr-49, nhr-19*). Transgenic RFP reporters contain the *cis*-regulatory regions bound by UNC-3, as well as flanking sequences. Middle: representative images of WT L4 animals showing expression of the transgenic reporters in nerve cord MNs. Scale bar, 20 µm. Bottom: quantification of TF reporters in WT and *unc-3 (n3435)* animals at 3 different developmental stages – L2, L4, and day 3 adult. *nhr-1, ztf-26*, and *zfh-2* are activated by UNC-3 (**A**), while *nhr-49* is repressed by UNC-3 (**B**) and *nhr-19* does not appear to be controlled by UNC-3 (**C**). N $\geq$ 15. N.S.: not significant, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
D, Three TF reporters (*ccch-3, ztf-17, nhr-1*) are not expressed in nerve cord MNs, but show expression in head neurons. Top: snapshots of UNC-3 ChIP-seq and input (negative control) signals at the *cis*-regulatory regions of *ccch-3, ztf-17*, and *nhr-1*. Transgenic RFP reporters contain the *cis*-regulatory regions bound by UNC-3. Bottom: representative images of WT L4 animals showing expression of the transgenic reporters in some unidentified neurons of the head. It is known that UNC-3 is expressed in some head neurons.

enhancer element is sufficient for initiation and maintenance of *cfi-1* reporter expression in nerve cord MNs.

To test the necessity of this element, we first generated via CRISPR/Cas9 an endogenous mNeonGreen (mNG) reporter allele for *cfi-1*, which also carries an auxin-inducible degron (AID) tag (*mNG::AID::cfi-1*), enabling inducible depletion of CFI-1 (depletion experiments are described later in Results). Animals carrying the *mNG::AID::cfi-1* AID allele do not show any developmental phenotypes, suggesting that the *mNG::AID* tag does not detectably alter *cfi-1* activity. This reporter showed expression in subsets of *unc-3*-expressing MNs (DA, DB, VA, VB subtypes), GABAergic nerve cord MNs (DD, VD subtypes), tail and head neurons, as well as head muscle (Figures 3.7A-C), a pattern consistent with previous studies describing *cfi-1* expression (Shaham and Bargmann, 2002, Kerk et al., 2017). We determined the onset of the endogenous *cfi-1* reporter (*mNG::AID::cfi-1*) in MNs to be at the 3-fold embryonic stage, coinciding with the onset of transgenic reporters (#7 and #8) containing the distal enhancer (Figure 3.7B). Next, we employed CRISPR/Cas9 genome editing and deleted 769 bp that constitute the core of the UNC-3 binding peak (located ∼12 kb upstream) in the

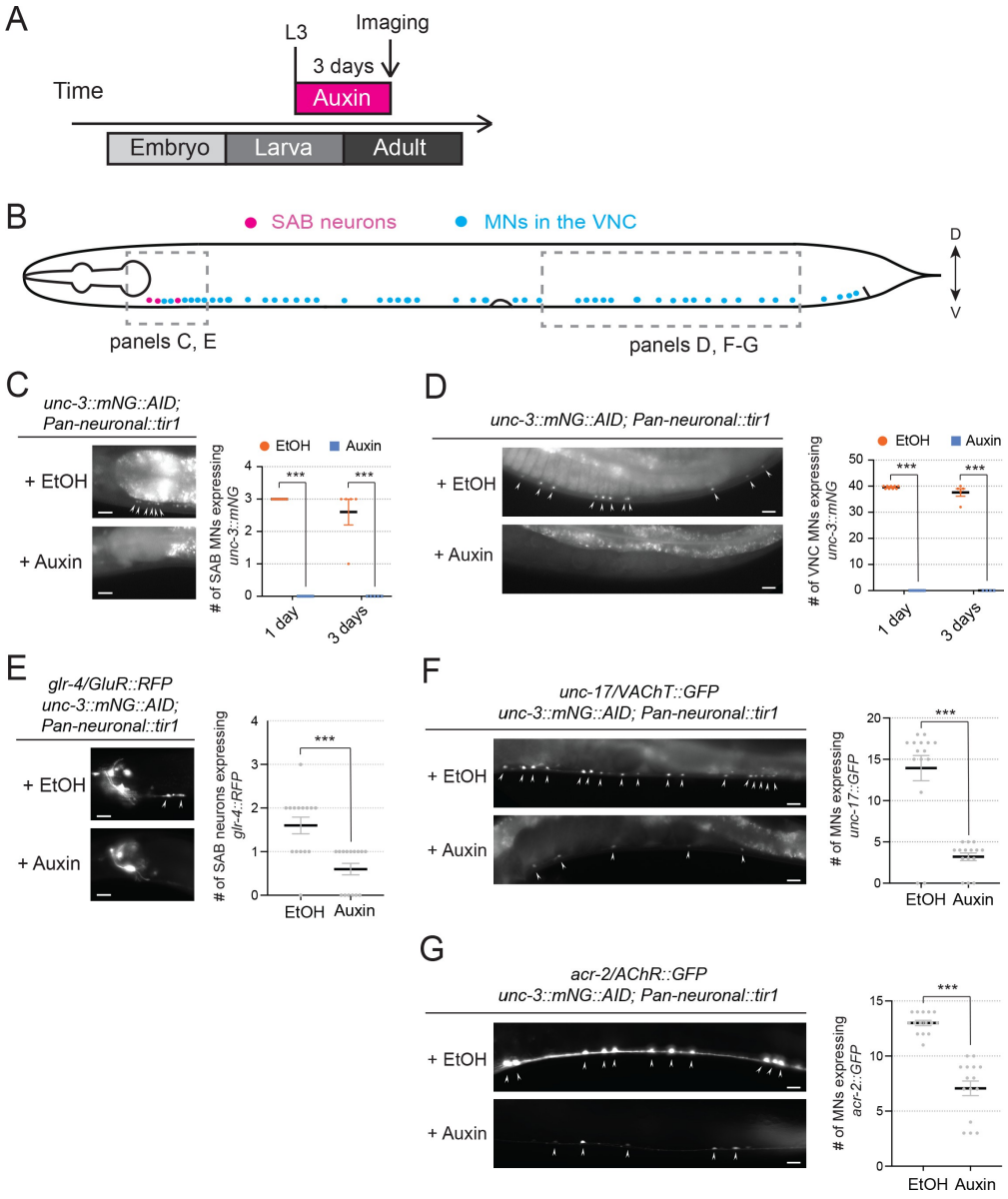


Figure 3.6: **UNC-3 is required to maintain the expression of *glr-4/GluR*, *unc-17/VAChT*, and *acr-2/AChR* in cholinergic motor neurons.**
A, Diagram illustrating the design of UNC-3 knock-down experiments. Auxin (or EtOH, as control) was applied to L3 worms carrying *unc-3::mNG::AID* and a pan-neuronal *tir-1* allele (with reporters of target genes to be tested). Examination of reporter expression was performed following 3 days of continuous auxin/EtOH administration.

Figure 3.6, continued.
B, Diagram illustrating the worm body with highlights of the locations of the SAB neurons (in the anterior) and the MNs along the VNC. Two boxed regions focus on the SAB neurons and the MNs in the posterior half of the worm body, which are shown in the representative images in the next panels.
C-D, Demonstration of UNC-3 knock-down efficiency. Left: representative images showing expression of *unc-3::mNG::AID* after 3 days of auxin or EtOH administration in the SAB neurons (**C**) or MNs in the posterior VNC (**D**); Right: Quantification of the number of SAB neurons (**C**) or MNs in the posterior VNC (**D**) showing expression of *unc-3::mNG::AID* following auxin/EtOH treatment. To get a preliminary idea of the efficiency, 9 worms from the auxin group and 10 worms from the EtOH group were checked one day after starting the treatment. To demonstrate the knock-down is persistent, 5 worms from each group were checked 3 days after starting the treatment. *** $p < 0.001$.
E, *glr-4/GluR, unc-17/VAChT*, and *acr-2/AChR* require UNC-3 for both initiation and maintenance of their expression. Left: representative images showing the expression of *glr-4::RFP* (**E**), *unc-17::GFP* (**F**), and *acr-2::GFP* (**G**) following 3 days or auxin/EtOH treatment. Downregulation of these reporters is apparent in terms of both the number of cells showing positive expression and the fluorescent intensity of the positive cells; Right: quantification of the number of SAB neurons expressing *glr-4::RFP* (**E**), and the number of MNs in the posterior VNC expressing *unc-17::GFP* (**F**) and *acr-2::GFP* from worms shown on the left (N = 15). *** $p < 0.001$.

context of the endogenous *cfi-1* reporter (*mNG::AID::cfi-1* $^{\Delta\ enhancer\ (769\ bp)}$). We found that *mNG::AID::cfi-1* expression is selectively eliminated in cholinergic (DA, DB, VA, VB) and GABAergic (DD, VD) nerve cord MNs at all life stages examined (3-fold embryo, L4, Day 1 adult) (Figures 3.7A-B, quantification of cholinergic MNs shown in Figure 3.7D).

We conclude that a distal 2.5kb enhancer (located ∼12 kb upstream of *cfi-1*) is sufficient for *cfi-1* expression in nerve cord MNs. Genome editing suggests that a 769 bp sequence within this 2.5kb enhancer is required for both initiation and maintenance of *cfi-1* in nerve cord MNs (Figure 3.7D). In the ensuing sections, we test the hypothesis that this enhancer integrates UNC-3 input for *cfi-1* maintenance in cholinergic MNs, as well as UNC-3-independent input for *cfi-1* initiation in these neurons.

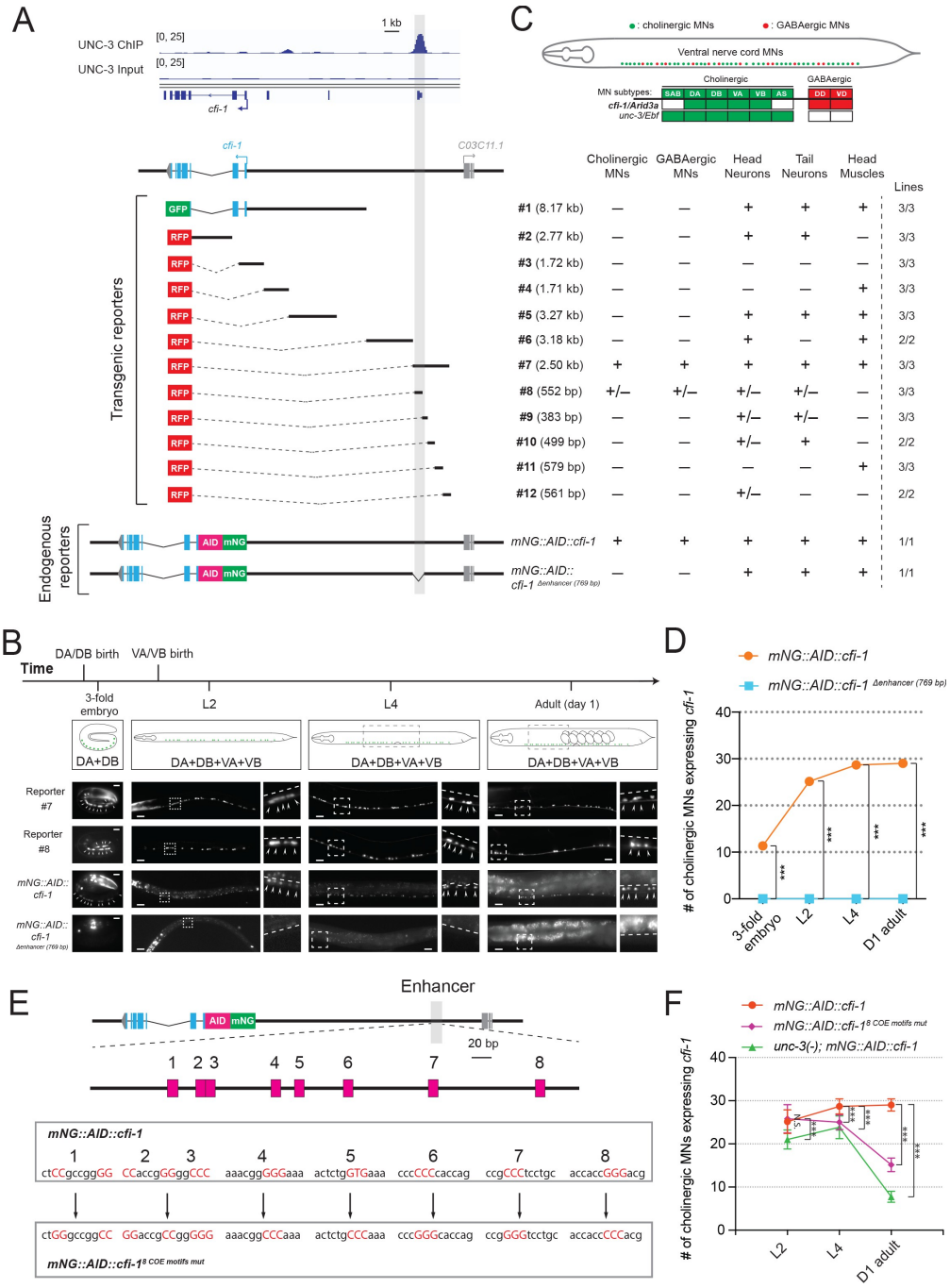


Figure 3.7: **UNC-3 acts through a distal enhancer to maintain *cfi-1* expression in cholinergic motor neurons.**

Figure 3.7, continued.
A, Top: Snapshots of UNC-3 ChIP-Seq and input (negative control) signals at the *cis*-regulatory region of *cfi-1*. The grey bar highlights an UNC-3 binding peak located ∼12 kb upstream of the TSS of *cfi-1* (-11,391 bp to -12,146 bp). Bottom: schematic showing the strategy of constructing *cfi-1* reporters. Twelve transcriptional fusion reporters (#1 [-1 bp to -8,170 bp], #2 [993 bp to 3,764 bp], #3 [547 bp to -1,173 bp], #4 [-1,164 bp to -2,875 bp], #5 [-2,865 bp to -6,141 bp], #6 [-8,162 bp to -11,346 bp], #7 [-11,329 bp to -13,824 bp], #8 [-11,329 bp to -11,881 bp], #9 [-11,851 bp to -12,234 bp], #10 [-12,223 bp to -12,722 bp], #11 [-12,705 bp to -13,284 bp], and #12 [-13,263 bp to -13,824 bp]) carry cis-regulatory regions fused to fluorescent reporters (GFP or RFP). The endogenous reporter alleles (*mNG::AID::cfi-1* and *mNG::AID::cfi-1* $^{\Delta\ enhancer\ (769\ bp)}$) have an in-frame fluorescent protein mNeonGreen (mNG) insertion immediately after the ATG of *cfi-1*. The enhancer KO allele *mNG::AID::cfi-1* $^{\Delta\ enhancer\ (769\ bp)}$ carries a 769 bp deletion (-11,329 bp to -12,097 bp). Table on the right summarizes the expression pattern of each reporter allele at L4 stage. N ≥ 12. +: reporter expressed, –: reporter not expressed, +/–: reporter partially expressed in the respective neurons. Number of independent transgenic lines tested for each reporter is shown on the right.
B, Representative images showing the expression of reporter #7, reporter #8, *mNG::AID::cfi-1*, and *mNG::AID::cfi-1* $^{\Delta\ enhancer\ (769\ bp)}$) at specific life stages. Areas highlighted in dashed boxes are enlarged and presented on the right side of each picture. The onset of *cfi-1* expression occurs at the 3-fold embryonic stage. mNG+ MNs are annotated with arrowheads. Scale bars, 5 µm (3-fold embryos); 20 µm (larvae and adults).
C, Schematic summarizing the expression pattern of *cfi-1* and *unc-3* in nerve cord MNs.
D, Quantification of the number of cholinergic MNs expressing endogenous *cfi-1 (mNG::AID::cfi-1)* in WT and animals carrying the enhancer deletion (*mNG::AID::cfi-1* $^{\Delta\ enhancer\ (769\ bp)}$). Deletion of the enhancer element located ∼12 kb upstream of the TSS of *cfi-1* completely abolishes *cfi-1* expression in MNs at all tested stages. A red fluorescent marker (*ttr-39::mCherry*) for GABAergic MNs was used to exclude these neurons from the quantification. Cholinergic MNs expressing *cfi-1* were positive for mNG and negative for mCherry.
E, Bioinformatic analysis predicted 8 UNC-3 binding sites (COE motifs, shown as pink boxes) in the *cfi-1* enhancer region, which displays UNC-3 binding (-11,391 bp to -12,146 bp). Using CRISPR/Cas9, these 8 motifs were mutated by substituting duplets or triplets of nucleotides as shown below, thereby generating the strain *cfi-1 (syb1856 [mNG::AID::cfi-1*$^{8\ COE\ motifs\ mut}$*])*.
F, Quantification of the number of cholinergic MNs expressing the endogenous *cfi-1* reporter (*mNG::AID::cfi-1*) in WT and *unc-3 (n3435)* animals, as well as in animals with mutated COE motifs (*mNG::AID::cfi-1*$^{8\ COE\ motifs\ mut}$) at L2, L4, and day 1 adult stages (N ≥ 12). N.S.: not significant, *** $p < 0.001$.

3.4.5 UNC-3 maintains cfi-1 expression in cholinergic MNs via direct activation of the distal enhancer

The binding of UNC-3 to the distal enhancer strongly suggests UNC-3 acts directly to maintain *cfi-1* expression in cholinergic MNs (Figure 3.7A). However, the UNC-3 peak is spread across several hundred base pairs due to the inherently low ChIP-Seq resolution. Hence, the precise DNA sequences recognized by UNC-3 remained unknown. Through bioinformatic analysis (see Materials and Methods), we identified 8 putative UNC-3 binding sites (COE motifs) within the 769 bp distal enhancer (Figure 3.7E). Using CRISPR/Cas9 technology, we simultaneously mutated all 8 motifs in the context of the endogenous *cfi-1* reporter allele (*mNG::AID::cfi-1*$^{8\ COE\ motifs\ mut}$) by substituting nucleotides known to be critical for DNA binding of UNC-3 orthologs (Treiber et al., 2010a, Wang et al., 1993) (Figure 3.7E). During the L2 stage, expression of mNG in MNs is not affected in *mNG::AID::cfi-1*$^{8\ COE\ motifs\ mut}$ animals, indicating early *cfi-1* expression occurs normally (Figure 3.7F). Intriguingly, mNG expression is significantly down-regulated in cholinergic MNs at later larval (L4) and adult (day 1) stages, resembling the phenotype of *unc-3* null mutants (Figure 3.7F). We conclude that UNC-3 binds to the distal enhancer and directly acts through one or more of these 8 COE motifs to maintain *cfi-1* expression in cholinergic MNs.

Previous studies in the nervous system have shown that a TF can maintain its own expression via transcriptional activation either by itself (positive auto-regulation), or in partnership with other TFs (Leyva-Diaz and Hobert, 2019, Scott et al., 2005, Xue et al., 1992). We found though that *cfi-1* does not auto-regulate and UNC-3 binding at the distal enhancer occurs normally in *cfi-1* null mutants (Figure 3.8), excluding a potential involvement of CFI-1 in its own maintenance.

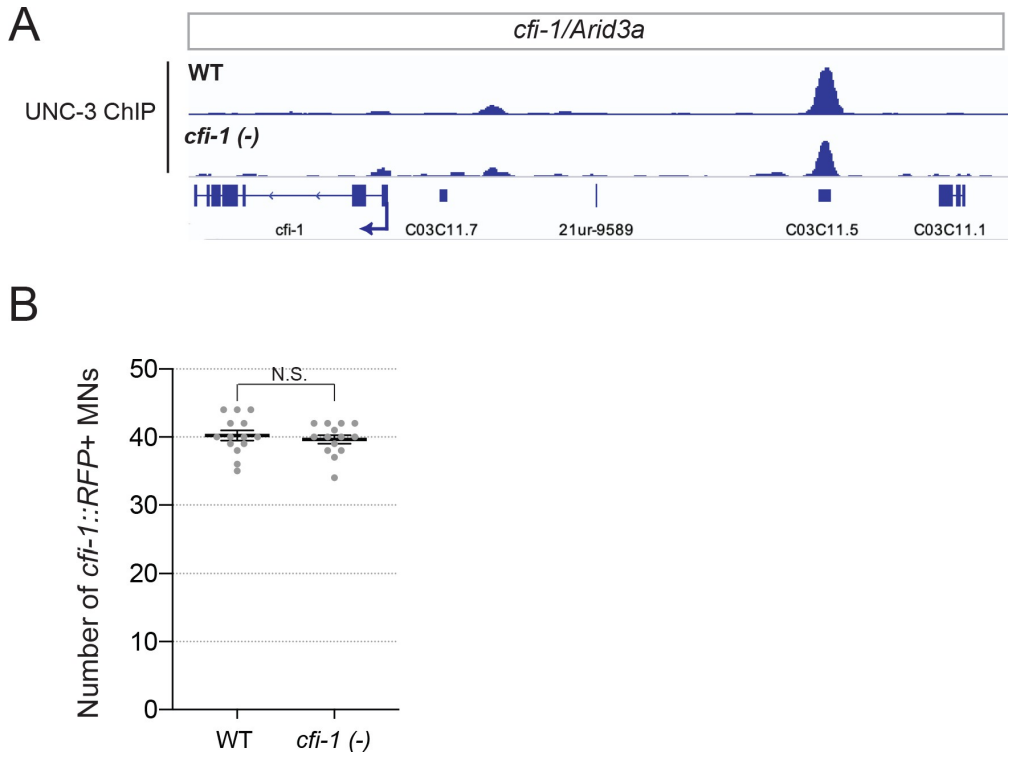


Figure 3.8: **CFI-1 does not auto-regulate.**
A, Snapshots of UNC-3 ChIP-seq signal at the *cfi-1* locus in WT and *cfi-1 (ot786)* animals. UNC-3 binds to the *cfi-1* enhancer normally in *cfi-1 (ot786)* mutants.
B, Quantification of the number of MNs expressing the transcriptional fusion reporter #7($cfi\text{-}1^{2.5kb}$*::RFP*) in WT worms and *cfi-1 (ot786)* mutants (N = 14). No significant difference was detected between WT and *cfi-1 (ot786)*, suggesting that the expression of reporter #7 ($cfi\text{-}1^{2.5kb}$*::RFP*) is not altered upon genetic removal of *cfi-1*. N.S.: not significant.

3.4.6 LIN-39 (Scr/Dfd/Hox4-5) and MAB-5 (Antp/Hox6-8) control cfi-1 expression in cholinergic MNs through the same distal enhancer

If UNC-3 exerts a maintenance role, what are the factors that initiate *cfi-1* expression in MNs? Previous work showed that two Hox proteins, LIN-39 (Scr/Dfd/Hox4-5) and MAB-5 (Antp/Hox6-8), control *cfi-1* expression in MNs (Kratsios et al., 2017). However, analysis was performed at the last larval stage (L4) and transgenic *cfi-1* reporter animals were used. Hence, it is unclear whether LIN-39 and MAB-5 are required for initiation of endogenous *cfi-1*.

Because *lin-39; mab-5* double null mutants are viable (Liu and Fire, 2000), we performed a longitudinal analysis and found that expression of the endogenous *mNG::AID::cfi-1* reporter in cholinergic MNs is severely affected at embryonic (3-fold), larval (L2, L4) and adult (D1) stages (Figure 3.9A-C). Since onset of *cfi-1* expression occurs at the 3-fold embryonic stage (Figure 3.7B), these results suggest LIN-39 and MAB-5 are required for *cfi-1* initiation. Conversely, initiation of *cfi-1* expression is not affected in a null mutant of *unc-3* (Figure 3.9B-C). Moreover, available ChIP-Seq data from modENCODE (Boyle et al., 2014) indicate that LIN-39 and MAB-5 bind to the *cfi-1* distal enhancer (Figure 3.9A). Expression of a 2.5 kb transgenic *cfi-1* reporter (reporter #7) that carries the distal enhancer is significantly affected in lin-39; mab-5 double mutants at early larval (L2) stages (Figure 3.9A, D). These results strongly suggest that LIN-39 and MAB-5 activate *cfi-1* expression directly. Although the DNA sequence of the MAB-5 binding site is not known, mutation of a single, bioinformatically predicted LIN-39 binding site (wild-type: aaTTGAtg > mutated: aaGGGGtg) within the enhancer led to a decrease in reporter gene expression at L2 (Figure 3.9A, E). This decrease was weaker compared to *lin-39; mab-5* double mutants (Figure 3.9C), likely due to compensation by MAB-5. Indeed, LIN-39 and MAB-5 appear to act synergistically because endogenous *cfi-1* expression (*mNG::AID::cfi-1*) is mildly affected in

lin-39 single mutants, but severely affected in *lin-39; mab-5* double mutants (Figure 3.9F). We conclude that the Hox proteins LIN-39 and MAB-5 are necessary for *cfi-1* initiation in cholinergic MNs (left panel, Figure 3.9G), and act through the same distal enhancer utilized by UNC-3 to maintain *cfi-1* (right panel, Figure 3.9G).

Because LIN-39 and MAB-5 are continuously expressed – from embryo through adulthood – in cholinergic MNs (Feng et al., 2020), it is likely that these Hox proteins, like UNC-3, are also required for maintenance. To test this, we used the auxin-inducible degradation (AID) system and depleted the endogenous LIN-39 protein at the last larval stage (L4) by using a previously described *lin-39::mNG::AID* allele (Feng et al., 2020, Zhang et al., 2015). However, expression of *cfi-1* was unaffected in the adult (Figure 3.10). This negative result could be attributed to low and undetectable levels of LIN-39 and/or functional compensation by MAB-5. We therefore used *lin-39; mab-5* double (null) mutants and crossed them to *unc-3* null animals. Of note, Hox genes (*lin-39, mab-5*) and *unc-3* do not cross-regulate their expression (Kratsios et al., 2017). If Hox proteins, similar to UNC-3, are required for *cfi-1* maintenance in cholinergic MNs, stronger effects should be present in *unc-3; lin-39; mab-5* triple mutants compared to *unc-3* single mutants. Indeed, we found this to be the case in day 1 adult animals (graph on the right, Figure 3.9C). Supporting a maintenance role for Hox, mutation of the LIN-39 binding site within the *cfi-1* enhancer (reporter #8) led to a sustained decrease in reporter gene expression from L2 to adult stages (Figure 3.9E).

Together, our findings suggest the Hox proteins LIN-39 and MAB-5 control initiation and maintenance of *cfi-1* in cholinergic MNs via the same distal *cis*-regulatory region (enhancer) utilized by UNC-3 to maintain *cfi-1* (Figure 3.9G). However, this region bears distinct UNC-3 and LIN-39 binding sites.

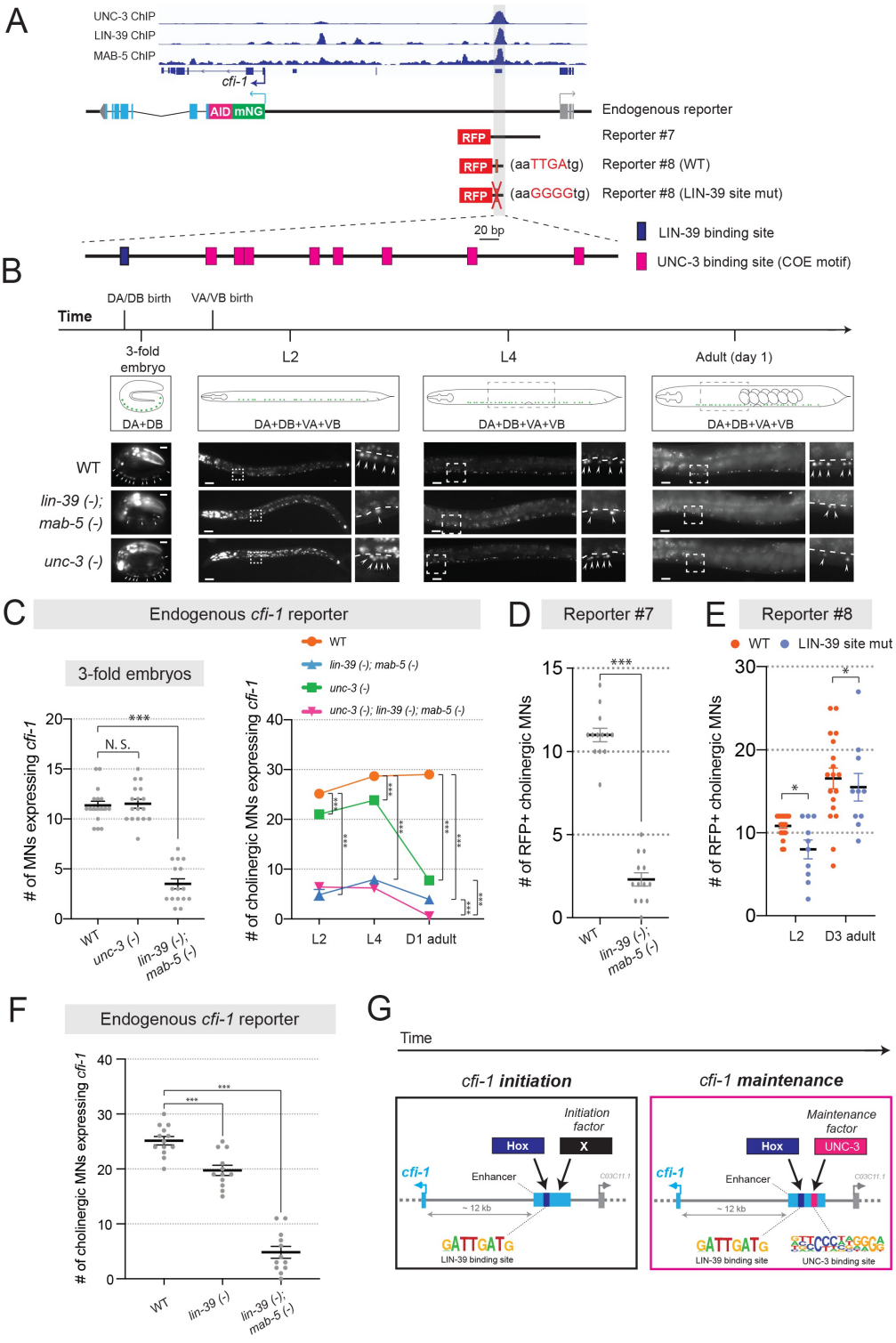


Figure 3.9: **UNC-3 and Hox control *cfi-1* expression in cholinergic MNs.**

Figure 3.9, continued.
A, A snapshot of UNC-3 (L2 stage), LIN-39 (L3 stage), and MAB-5 (L2 stage) ChIP-Seq signals at the *cfi-1* locus. UNC-3, LIN-39, and MAB-5 bind to the same *cfi-1* enhancer (highlighted in gray). Below: Schematics illustrating the reporters used in the rest of the figure.
B, Representative images showing the expression of the *mNG::AID::cfi-1* in WT, *unc-3 (n3435)*, and *lin-39 (n1760); mab-5 (e1239)* animals during 3-fold embryonic, L2, L4, and day 1 adult stages. *cfi-1* is expressed in 4 cholinergic MN subtypes (DA, DB, VA, and VB). DA and DB are born embryonically, while VA and VB are born post-embryonically. Areas highlighted in dashed boxes are enlarged and presented on the right side of each picture. mNG+ MNs are annotated with arrowheads. Scale bars, 5 µm (3-fold embryos); 20 µm (larvae and adults).
C, Quantification of the number of cholinergic MNs expressing the endogenous *cfi-1* reporter (*mNG::AID::cfi-1*) in WT animals, *unc-3 (n3435)* mutants, *lin-39 (n1760); mab-5 (e1239)* double mutants, and *unc-3 (n3435); lin-39 (n1760); mab-5 (e1239)* triple mutants during 3-fold embryonic, L2, L4, and day 1 adult stages ($N \geq 12$). N.S.: not significant, *** $p < 0.001$. A red fluorescent marker (ttr-39::mCherry) for GABAergic MNs was used to exclude these neurons from the quantification. Cholinergic MNs expressing *cfi-1* were positive for mNG and negative for mCherry.
D, Quantification of the expression of transgenic *cfi-1* reporter #7 in WT and *lin-39 (n1760); mab-5 (e1239)* animals at L2 stage ($N = 13$). Reporter expression is strongly affected in *lin-39 (-); mab-5 (-)* double mutants. *** p ¡ 0.001.
E, Quantification of the WT transgenic reporter #8 and the same reporter with the LIN-39 binding site mutated (point mutations) at larval (L2) and adult (D3) stages ($N \geq 13$). * $p < 0.05$.
F, Quantification of the expression of the *mNG::AID::cfi-1* allele in WT animals, *lin-39 (n1760)* single mutants, and *lin-39 (n1760); mab-5 (e1239)* double mutants at the L2 stage ($N \geq 12$). While the number of cholinergic MNs with *cfi-1* expression is mildly decreased in *lin-39* single mutants, more severe effects are observed in double mutants. *** $p < 0.001$.
G, Schematic summarizing the mechanisms underlying initiation and maintenance of *cfi-1* expression in cholinergic MNs.

3.4.7 cfi-1/Arid3a is required post-embryonically to maintain MN subtype identity

Expression of *cfi-1* in cholinergic MNs (DA, DB, VA, VB) is maintained throughout life by Hox and UNC-3 (Figure 3.9G). But why is it important to ensure continuous *cfi-1* expression? Although its function in VA and VB remains unknown, CFI-1 is required during early development to establish the identity of DA and DB subtypes by acting as a tran-

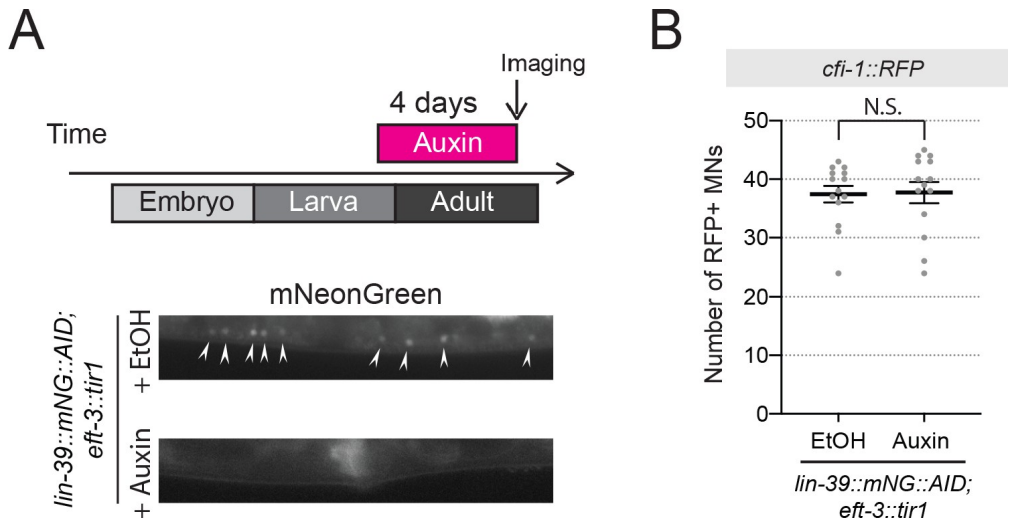


Figure 3.10: **Auxin-inducible depletion of LIN-39 at larval stage 4 (L4) does not affect *cfi-1* expression in nerve cord MNs.**
A, Representative images showing expression of LIN-39::mNG::AID after treatment with ethanol and auxin (negative control). LIN-39::mNG::AID is degraded and mNG fluorescent signal becomes undetectable in MNs, when worms are imaged after 4 days of auxin treatment. Arrowheads indicate MN nuclei in the nerve cord.
B, Quantification of the number of MNs expressing *glr-4::RFP* in *mNG::AID::cfi-1* worms, worms carrying the *cfi-1* enhancer deletion (*mNG::AID::cfi-1* $^{\Delta\ enhancer\ (769\ bp)}$), worms with mutated COE motifs (*mNG::AID::cfi-1*$^{8\ COE\ motifs\ mut}$), as well as in *cfi-1* null mutants at L2, L4, and day 2 adult stages (N = 15). N.S.: not significant, *** $p < 0.001$. The remaining expression of *cfi-1* in $\sim$ 15 MNs of *mNG::AID::cfi-1*$^{8\ COE\ motifs\ mut}$ animals (Figure 3.7F) is likely the reason for not observing *glr-4/GluR* ectopic expression in MNs of these animals.

scriptional repressor (Kerk et al., 2017). In *cfi-1* null animals, glutamate receptor subunit 4 (*glr-4/GluR*), a terminal identity gene normally activated by UNC-3 in another MN subtype (SAB), becomes ectopically expressed in DA and DB neurons (Figure 3.11). In animals lacking *cfi-1* expression specifically in MNs at all stages (*mNG::AID::cfi-1* $^{\Delta\ enhancer\ (769\ bp)}$, Figure 3.7A-B), we also observed ectopic *glr-4* expression in these neurons (Figure 3.11). Hence, early global removal of *cfi-1*, or MN-specific loss of *cfi-1* both lead to DA and DB neurons adopting a mixed identity. However, whether CFI-1 is required post-embryonically to continuously prevent DA and DB from obtaining a mixed identity is not known.

To enable CFI-1 protein depletion selectively at post-embryonic stages, we used the *mNG::AID::cfi-1* reporter allele (Figure 3.7A), which also serves as a conditional allele as it carries the auxin-inducible degron (AID) (Zhang et al., 2015). Auxin administration at the first larval (L1) stage resulted in efficient depletion of *mNG::AID::cfi-1* expression, which was undetectable 2 days later (Figure 3.11). At the L3 stage (2 days upon continuous auxin treatment), we observed ectopic expression of *glr-4* in DA and DB neurons. These results suggest that CFI-1 is required post-embryonically to prevent DA and DB neurons from adopting mixed identity, underscoring the critical role of UNC-3 and Hox in maintaining *cfi-1* expression (Figure 3.9G).

3.4.8 Minimal disruption of temporal modularity in UNC-3 function leads to locomotion defects

Animals carrying *unc-3* null alleles display severe locomotion defects (Feng et al., 2020), likely due to combined defects in the expression of UNC-3 targets from both "initiation and maintenance" and "maintenance-only" modules (Figure 3.4C). Genes from the "initiation and maintenance" module include (among others) terminal identity genes coding for ACh biosynthesis components (Figure 3.4A). Hence, it is conceivable that loss of *unc-3* can lead to defects in ACh biosynthesis, likely contributing to locomotion defects. However, it is unclear whether, in cholinergic MNs, the maintained expression of any of the UNC-3 targets from the "maintenance-only" module is critical for locomotion. To test this, we focused on *cfi-1/Arid3a* and used the CRISPR-engineered allele (*mNG::AID::cfi-1*$^{8\ COE\ motifs\ mut}$) that selectively affects maintenance, but not initiation, of *cfi-1* expression in cholinergic MNs (Figure 3.7E-F). This allele minimally disrupts temporal modularity in UNC-3 function because UNC-3 can still control all its targets from both modules (Figure 3.4C), except one (*cfi-1*). As controls, we used animals carrying: (a) the endogenous *cfi-1* reporter allele

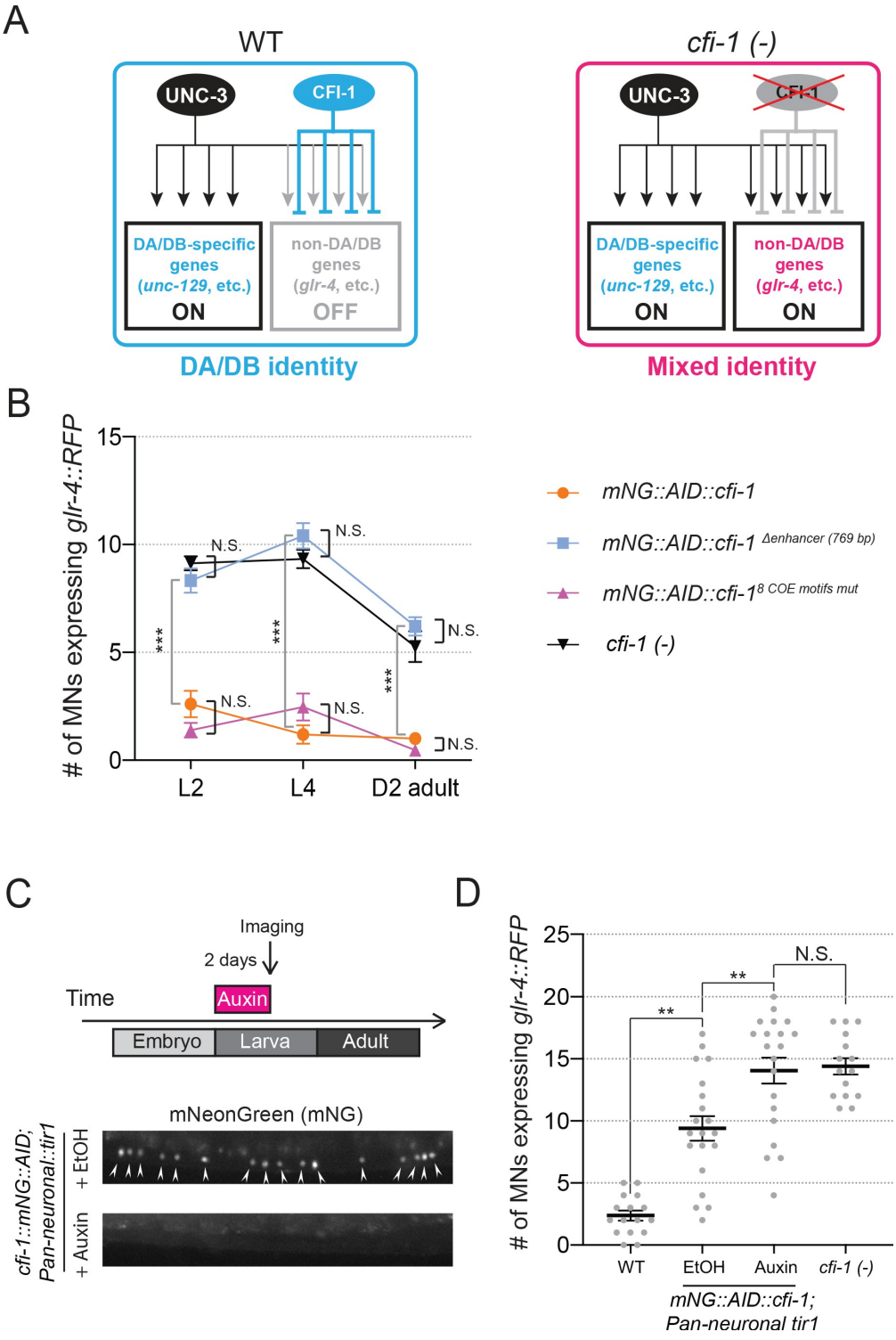


Figure 3.11: **CFI-1 is required post-embryonically to maintain DA and DB neuronal identities.**

Figure 3.11, continued.
A, CFI-1 regulates DA and DB MN identity by repressing the glutamate receptor subunit *glr-4/GluR* and possibly other genes. In *cfi-1 (-)* mutants, *glr-4/GluR* becomes ectopically activated by UNC-3 in DA and DB MNs, and these neurons adopt a mixed molecular identity.
B, Representative images showing expression of mNG::AID::CFI-1 after treatment with ethanol and auxin. Upon two days of continuous auxin treatment (onset of treatment at L1), mNG::AID::CFI-1 is degraded and mNG expression in MNs (arrowheads) becomes undetectable.
C, Quantification of MNs expressing *glr-4::RFP* was performed on L4 worms (2 days after the onset of auxin treatment). Some mild hypomorphic effects in the expression of *glr-4::RFP* was observed in the EtOH group (negative control), potentially due to mild reduction in CFI-1 levels triggered by TIR1 even in the absence of auxin. A significant increase in the number of MNs expressing *glr-4::RFP* was evident in auxin-treated worms in comparison to control animals. WT and *cfi-1 (-)* data are also provided for comparison. $N \geq 20$. N.S.: not significant, ** $p < 0.01$.

(*mNG::AID::cfi-1*), (b) a putative null *cfi-1* allele (ot786) (Kerk et al., 2017), in which *cfi-1* activity is affected in MNs and other neuron types of the motor circuit (Pereira et al., 2015, Shaham and Bargmann, 2002), and (c) a deletion of the distal enhancer (*mNG::AID::cfi-1* Δ *enhancer (769 bp)*), in which both initiation and maintenance of *cfi-1* are abrogated in nerve cord MNs (Figure 3.12A). We performed a high-resolution behavioral analysis of freely moving adult (day 2) animals of the above genotypes using automated multi-worm tracking technology (Javer et al., 2018b, Yemini et al., 2013). We found several features related to *C. elegans* locomotion (e.g., body curvature, velocity) severely affected in *cfi-1 (ot786)* putative null animals (Figure 3.12B-G). Compared to *cfi-1* null mutants, animals carrying the *mNG::AID::cfi-1*$^{8\ COE\ motifs\ mut}$ allele (selective disruption of *cfi-1* maintenance) display milder, but statistically significant locomotion defects in the adult (Figure 3.12B-G). As expected, these defects were also present in animals carrying the *mNG::AID::cfi-1* Δ *enhancer (769 bp)* allele, in which both initiation and maintenance of *cfi-1* expression is affected. In summary, we specifically disrupted in cholinergic MNs the maintained expression of a single UNC-3 target (*cfi-1* from the "maintenance-only" module) by using the *mNG::AID::cfi-1*$^{8\ COE\ motifs\ mut}$ allele and observed locomotion defects. This analysis sug-

gests that minimal disruption of temporal modularity in UNC-3 function can affect animal behavior.

3.4.9 Hox proteins and UNC-3 control bnc-1/BNC expression in cholinergic motor neurons

Our *cis*-regulatory analysis suggested that five TFs (*cfi-1/Arid3a, bnc-1/BNC1-2, mab-9/Tbx20, ceh-44/CUX1-2, nhr-40/nuclear hormone receptor*) require UNC-3 selectively for maintenance (Figure 3.4B). An in-depth analysis of *cfi-1/Arid3a* revealed that Hox proteins (LIN-39, MAB-5) and UNC-3 ensure the continuous expression of *cfi-1* in subsets of *unc-3*-positive MNs (Figures 3.4, 3.7, and 3.9). We next asked whether a similar mechanism applies to the regulation of *bnc-1/BNC*, which is also expressed in a subset of *unc-3*-positive MNs (VA, VB) and prevents them from adopting a mixed identity (Figure 3.13) (Kerk et al., 2017). Using an endogenous reporter allele (*bnc-1::mNG::AID*), we found that LIN-39 and MAB-5 are required for bnc-1 expression at all stages examined (L2, L4, day 1 adult; Figure 3.13), suggesting a role for Hox in *bnc-1* initiation and maintenance. Next, we found that UNC-3 is absolutely required for *bnc-1* maintenance in the adult (day 1), albeit weaker effects were also observed at L2 and L4 (Figure 3.13). Similar to *cfi-1*, these findings strongly suggest that endogenous *bnc-1* expression depends on Hox and UNC-3.

3.4.10 Temporal modularity of UNC-30/PITX function in GABAergic motor neurons

Is temporal modularity observed in the function of other terminal selectors? To address this, we focused on UNC-30/PITX, the terminal selector of GABAergic MN (DD, VD) identity in

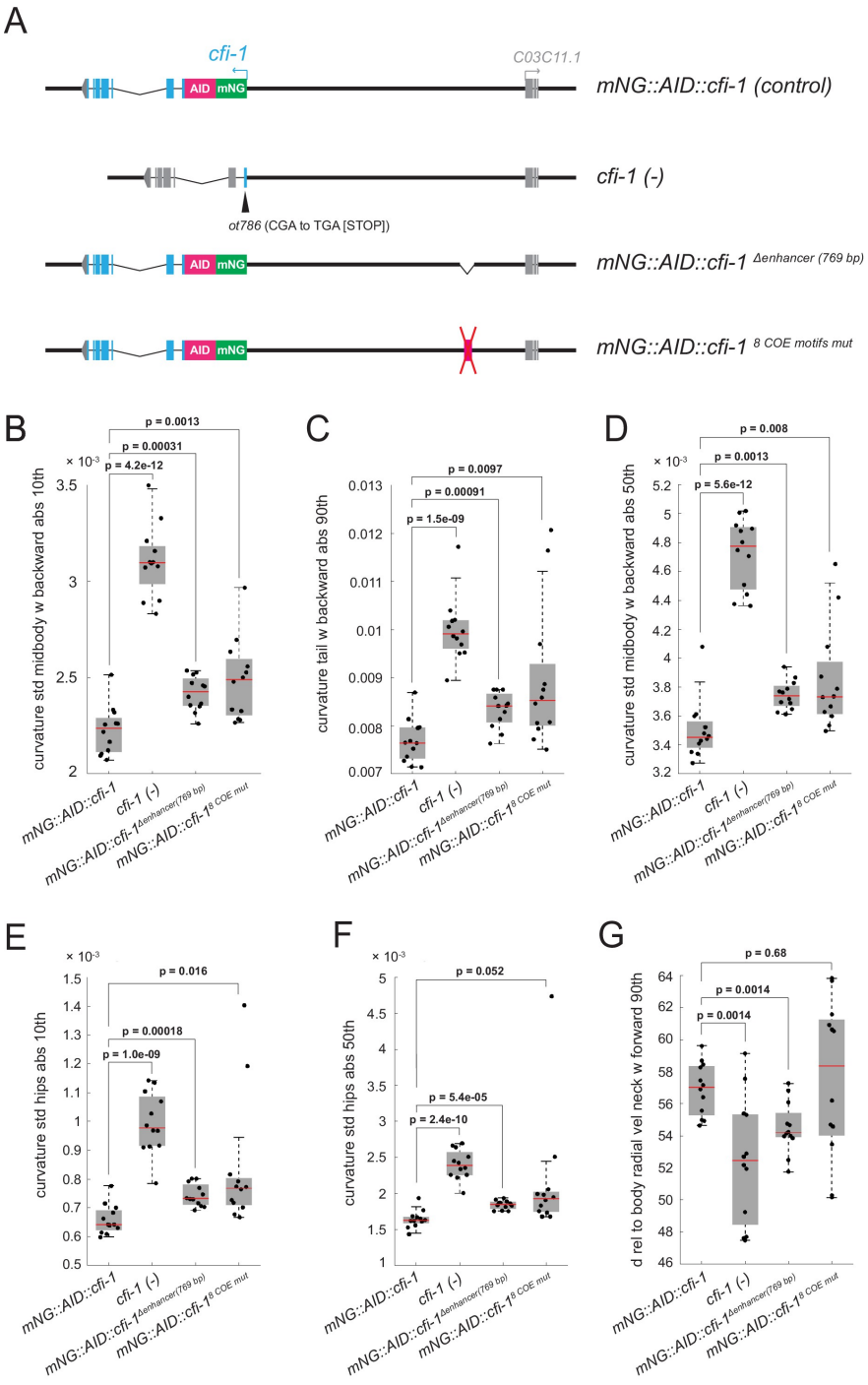


Figure 3.12: **Disruption of temporal modularity in UNC-3 function leads to locomotion defects.**

Figure 3.12, continued.
A, Schematics illustrating four *cfi-1* alleles tested for behavioral analysis.
B-G, Examples of six locomotion features significantly disrupted in animals carrying a putative null (*ot786*) allele for *cfi-1*. Locomotion analysis was performed on day 2 adult worms. Animals lacking *cfi-1* expression (initiation and maintenance) specifically in MNs (*mNG::AID::cfi-1* $^{\Delta\ enhancer\ (769\ bp)}$) and animals unable to maintain cfi-1 expression in cholinergic MNs (*mNG::AID::cfi-1*$^{8\ COE\ motifs\ mut}$) display locomotion defects when compared to control *mNG::AID::cfi-1* animals. As expected, these defects are milder when compared to animals carrying the *cfi-1 (ot786)* allele. Panels B-F show locomotion features related to body curvature, whereas panel G shows radial velocity of the neck. A detailed description of each locomotion feature is provided below:
B, curvature_std_midbody_w_backward_abs_10th: 10th percentile of the absolute value of the standard deviation of the curvature of the midbody, while worm is moving backwards.
C, curvature_tail_w_backward_abs_90th: 90th percentile of the absolute value of the curvature of the tail, while worm is moving backwards.
D, curvature_std_midbody_w_backward_50th: 50th percentile of the standard deviation of the curvature of the midbody, while worm is moving backwards.
E, curvature_std_hips_abs_10th: 10th percentile of the absolute value of the standard deviation of the curvature of the hips.
F, curvature_std_hips_abs_50th: 50th percentile of the absolute value of the standard deviation of the curvature of the hips
G, d_rel_to_body_radial_vel_neck_w_forward_90th: 90th percentile of the derivative of radial velocity of the neck relative to the centroid of the midbody points, while worm is moving forwards.

the *C. elegans* nerve cord (Figure 3.14A) (Jin et al., 1994). UNC-30/PITX is known to directly activate the expression of several terminal identity genes (e.g., *unc-25/GAD* [glutamic acid decarboxylase], *unc-47/VAGT* [vesicular GABA transporter]) (Eastman et al., 1999), but a longitudinal analysis of target gene expression in *unc-30* null animals is lacking. Using reporter strains and methodologies similar to those used for UNC-3, we found that terminal identity gene (*unc-25/GAD, unc-47/VAGT*) expression is affected in GABAergic MNs of *unc-30* mutants at all stages examined (3-fold embryo, L2, L4, adult [day 1]) (Figures 3.14B and 3.14D), suggesting a requirement for initiation and maintenance. As mentioned earlier (Figure 3.7C), *cfi-1* is also expressed in GABAergic MNs (Figure 3.14A). Since UNC-3 is required for *cfi-1* maintenance in cholinergic MNs, we asked whether UNC-30/PITX plays a similar role to maintain *cfi-1* expression in GABAergic MNs. We found no effect at the

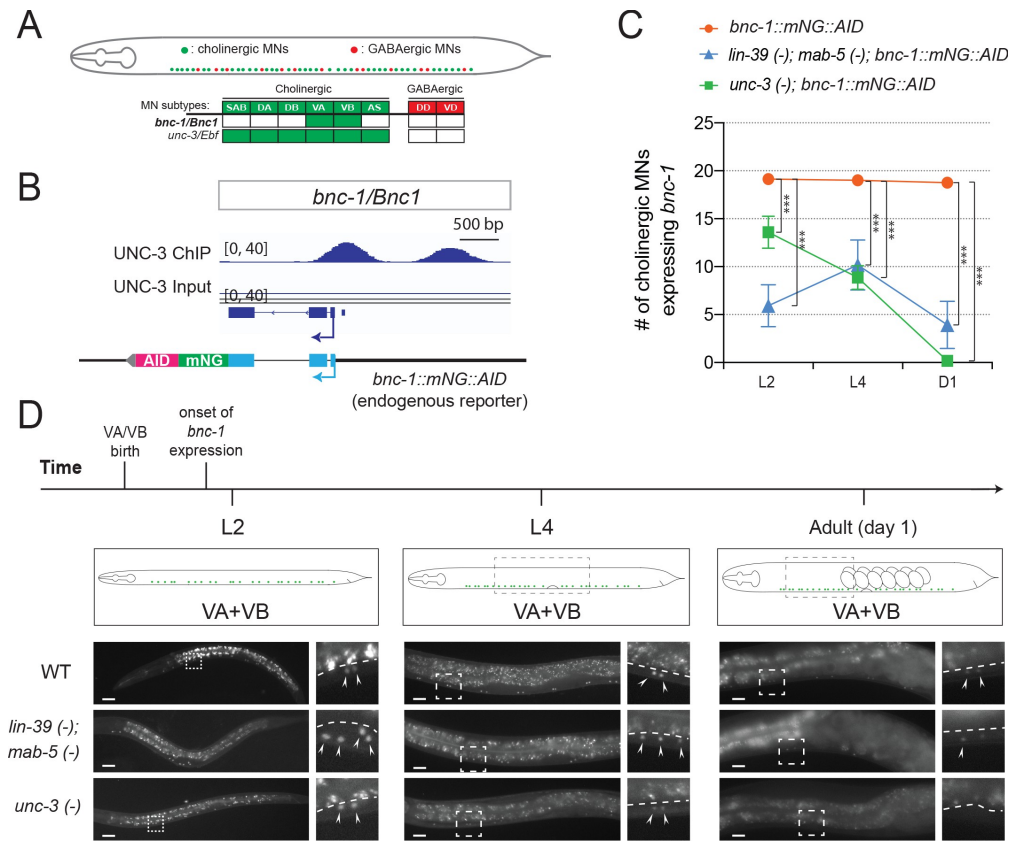


Figure 3.13: **Hox proteins and UNC-3 control *bnc-1* expression in VA and VB neurons.**

A, Schematic summarizing the expression of *bnc-1/Bnc1* and *unc-3/Ebf* in MN subtypes of the *C. elegans* ventral nerve cord.

B, A snapshot of UNC-3 ChIP-seq and input (negative control) signals at the cis-regulatory region of bnc-1.

C, Quantification of expression of the endogenous reporter *bnc-1::mNG::AID* in WT, *unc-3 (n3435)*, and *lin-39 (n1760); mab-5 (e1239)* animals during L2, L4, and day 1 adult stages ($N \geq 12$). *** $p < 0.001$.

D, Representative images of the endogenous reporter *bnc-1::mNG::AID* in WT, *unc-3 (e151)*, and *lin-39 (n1760); mab-5 (e1239)* animals during L2, L4, and day 1 adult stages. bnc-1 is expressed in two cholinergic MN subtypes (VA and VB). Areas highlighted in dashed boxes are enlarged and presented on the right side of each picture. Arrowheads point to nuclei of VA and VB neurons that express the reporter. Above the white dashed line lies the intestine, which is autofluorescent in the green channel. Scale bar, 20 µm.

embryonic stage (3-fold), but progressively stronger effects at larval (L2, L4) and adult (day 1) stages (Figure 3.14C), indicating that, in GABAergic MNs, UNC-30/PITX is selectively required for the maintenance, but not initiation of *cfi-1* expression. Taken together, our findings indicate that, like UNC-3 in cholinergic MNs (Figure 3.4C), the function of UNC-30 in GABAergic MNs is organized into two modules (module #1: initiation and maintenance; module #2: maintenance-only) (Figures 3.14D-E), suggesting temporal modularity may be a shared feature among terminal selector type-TFs.

To gain mechanistic insights, we analyzed available UNC-30 ChIP-Seq data at the L2 stage (Yu et al., 2017). UNC-30 binds to the cis-regulatory region of terminal identity genes (unc-25/GAD, unc-47/VAGT) (Figure 3.14B), confirming previous observations (Eastman et al., 1999). UNC-30 also binds to the same distal enhancer of *cfi-1* in GABAergic MNs, as UNC-3 does in cholinergic MNs (Figure 3.14C). However, the UNC-30 binding sites are distinct from the UNC-3 sites in this enhancer (Figure 3.14C). CRISPR-mediated deletion of this enhancer abolished *cfi-1* expression in both cholinergic and GABAergic MNs (Figures 3.7A-B). This finding suggests that maintenance of *cfi-1* expression in two different neuron types relies on the same enhancer receiving UNC-30/PITX input in GABAergic MNs and UNC-3/EBF input in cholinergic MNs. Interestingly, these results provide an example of "enhancer pleiotropy" (Sabaris et al., 2019), in which the same cis-regulatory element is used to control gene expression in different neuron types.

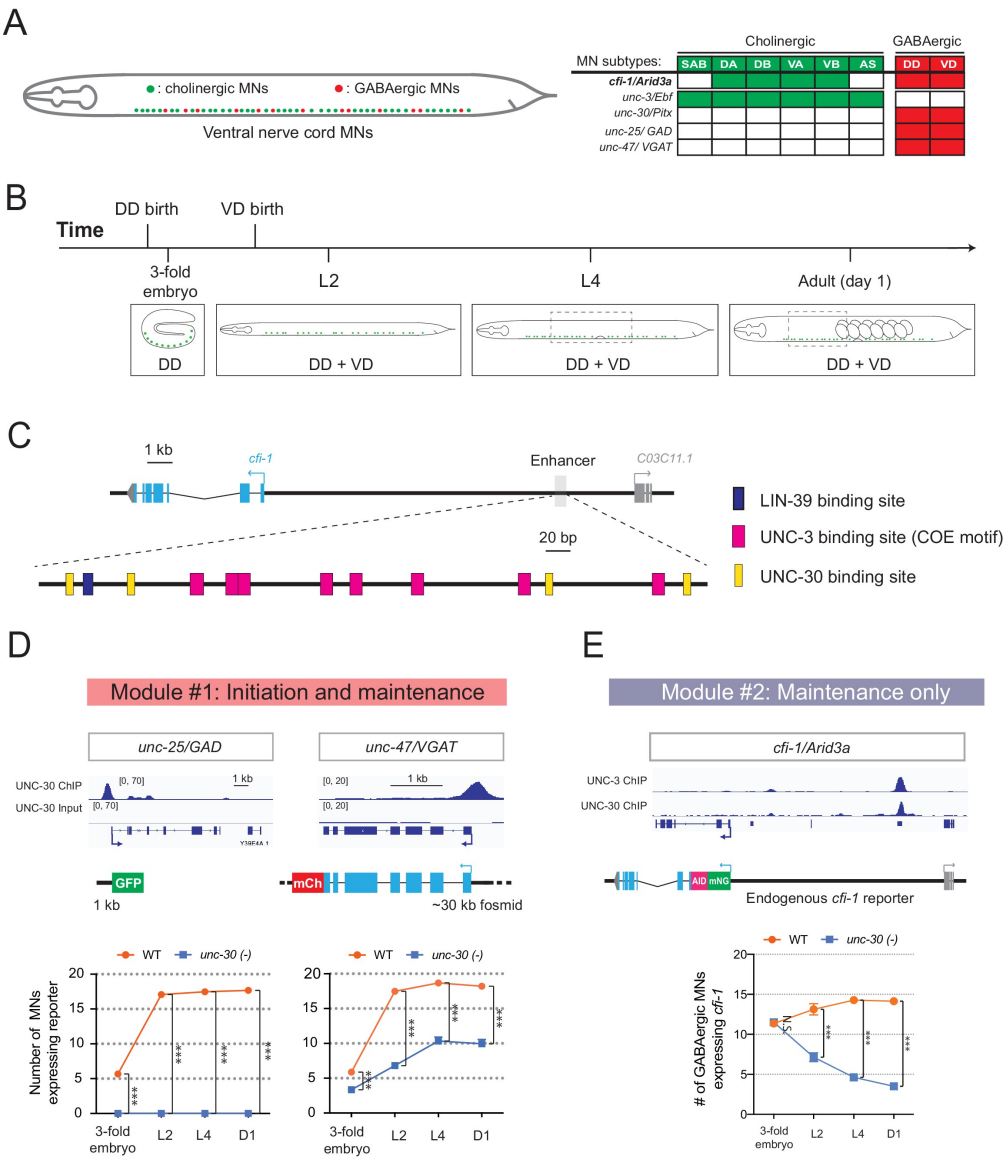


Figure 3.14: **Temporal modularity in UNC-30/Pitx function in GABAergic MNs.** **A**, Schematic summarizing the expression of *cfi-1/Arid3a, unc-3/Ebf, unc-30/Pitx, unc-25/GAD*, and *unc-47/VGAT* in MN subtypes of the *C. elegans* ventral nerve cord.
B, Schematic showing time of birth and cell body position of GABAergic nerve cord MNs. DD neurons are born embryonically. VD neurons are born post-embryonically.
C, Bioinformatic analyses predict 4 UNC-30 binding sites (yellow boxes) in the *cfi-1* enhancer. The location of UNC-3 and LIN-39 binding sites are also shown.

Figure 3.14, continued.
D, Top: snapshots of UNC-30 ChIP-Seq and input (negative control) signals at the *cis*-regulatory regions of 2 GABAergic terminal identity genes (*unc-25/GAD, unc-47/VGAT*). Bottom: quantification of the expression of transgenic reporters in WT and *unc-30 (e191)* animals at 4 different developmental stages – 3-fold embryo, L2, L4, and day 1 adults (N = 15). UNC-30 is required for both initiation and maintenance of *unc-25/GAD* and *unc-47/VGAT*. *** $p < 0.001$.
E, Top: a snapshot of UNC-3 ChIP-Seq and UNC-30 ChIP-Seq signals at the *cfi-1* locus. Bottom: quantification of the number of MNs expressing the endogenous reporter *mNG::AID::cfi-1* in WT and *unc-30 (e191)* animals. All *cfi-1*-expressing MNs in the ventral cord (cholinergic and GABAergic MNs) were counted in 3-fold embryos due to a lack of a specific marker that labels GABAergic MNs in embryos. Expression of *cfi-1* specifically in GABA neurons was quantified at L2, L4, and day 1 adult stages ($N \geq 12$). At those stages, cholinergic MNs were identified based on a fluorescent marker (*cho-1::mChOpti*), which are ruled out during scoring. GABAergic MNs were scored positive for *cfi-1* expression when the *mNG::AID::cfi-1* (green) expression co-localized with *ttr-39::mCherry* (red). N.S.: not significant, *** $p < 0.001$.

3.5 Discussion

Terminal selectors are continuously expressed TFs that determine the identity and function of individual neuron types (Hobert, 2008, Hobert, 2016b). However, the breadth of biological processes controlled by terminal selectors remains unclear. Moreover, whether terminal selectors control an identical suite of target genes across different life stages, or the suite of targets can change over time is largely unexplored. Filling such knowledge gaps can help us understand how cellular identity is established during development and maintained throughout life, a fundamental goal in the field of developmental biology. In this study, we focus on UNC-3/Ebf, the terminal selector of cholinergic MNs in the *C. elegans* nerve cord. Through ChIP-Seq, we identify in an unbiased manner the direct targets of UNC-3, uncovering the breadth of biological processes potentially controlled by this terminal selector. Unexpectedly, we find two groups of target genes with distinct temporal requirements for UNC-3 in cholinergic MNs. One group encodes different classes of proteins (e.g., receptors, transporters) that require UNC-3 activation at all life stages examined. However, a second group exclusively encodes MN-specific TFs that selectively require UNC-3 for maintenance of their expression

at late larval and adult stages. Hence, the suite of UNC-3 targets in cholinergic MNs is partially modified across different life stages, a phenomenon we term "temporal modularity" in terminal selector function. Minimal disruption of this modularity by selectively removing the ability of UNC-3 to maintain expression of a single target gene (the TF *cfi-1/Arid3a* from the second group) led to locomotion defects, highlighting the necessity of temporal modularity for animal behavior.

3.5.1 Temporal modularity in terminal selector function may represent a general principle for neuronal subtype diversity

Why is there a need for temporal modularity in the function of continuously expressed TFs, such as terminal selectors? The case of UNC-3 suggests that temporal modularity is necessary for generating and maintaining neuronal subtype diversity. UNC-3 is continuously expressed in six MN subtypes (SAB, DA, DB, VA, VB, AS) of the *C. elegans* ventral nerve cord (Figure 3.7C) (Pereira et al., 2015). However, most of UNC-3 targets from either module (modules # 1 and #2, Figure 3.4C) are expressed in some, but not all, of these subtypes. By ensuring the continuous expression of genes from module #1 ("initiation and maintenance" module, Figure 3.4C), such as terminal identity genes [(Kratsios et al., 2012) and this study], UNC-3 consolidates the unique identity of each MN subtype. However, target genes from module #2 ("maintenance-only" module, Figure 3.4C) escape initiation by UNC-3, but do require UNC-3 for maintenance. Interestingly, all module #2 targets code for TFs (*cfi-1/Arid3a, bnc-1/BNC1-2, mab-9/Tbx20, ceh-44/Cux1, nhr-40*), specifically expressed in subsets of the six subtypes (SAB, DA, DB, VA, VB, AS) ((Kerk et al., 2017) and this study). Three of these factors (*cfi-1/Arid3a, bnc-1/Bnc1, mab-9/Tbx20*) are thought to act as transcriptional repressors to prevent the adoption of mixed MN identity (Kerk et al., 2017). For example, loss of *cfi-1* during early development results in "mixed" identity of DA and DB subtypes,

as these cells, in addition to their normal molecular signature, also acquire expression of genes (e.g., *glr-4/GluR*) normally expressed in other MNs (Figure 3.11). Through protein depletion experiments at post-embryonic stages, we show here that CFI-1 is continuously required to prevent DA and DB neurons from acquiring a mixed identity. Hence, UNC-3 indirectly maintains the unique identity of individual MN subtypes (DA, DB) by selectively maintaining expression of the TF *cfi-1* (from module #2) during late larval and adult stages. In addition, UNC-3 acts directly to control DA and DB identity by ensuring the continuous expression of *unc-129/TGFb*, a DA- and DB-specific terminal identity gene (from module #1) (Figures 3.4A and 3.11). Hence, temporal modularity in UNC-3 function can be envisioned as a "double safety" mechanism for the generation and maintenance of MN diversity.

The mechanism of temporal modularity may represent a general paradigm of how terminal selectors establish and maintain subtype identity within a class of functionally related neurons. Indeed, the same terminal selector is often continuously expressed in multiple subtypes of a given neuronal class (Hobert and Kratsios, 2019). In mice, for example, the terminal selectors *Nurr1* and *Pet1* are expressed in several subtypes of dopaminergic and serotonergic neurons, respectively (Kadkhodaei et al., 2009, Okaty et al., 2015). However, future studies are needed to determine whether temporal modularity in the function of these terminal selectors is necessary for establishing and maintaining neuronal subtype identity.

3.5.2 Temporal modularity offers key insights into how terminal selectors control neuronal identity over time

The prevailing hypothesis of how terminal selectors establish and maintain neuronal identity is that they bind constantly, from development through adulthood, on the *cis*-regulatory region of terminal identity genes, and thereby continuously activate their expression (Deneris

and Hobert, 2014, Hobert, 2008, Hobert and Kratsios, 2019). For most terminal selectors, however, biochemical evidence for binding and longitudinal analysis of terminal selector mutant animals are currently lacking. Our analysis on module #1 genes (e.g., terminal identity genes) supports the aforementioned hypothesis; ChIP-Seq for UNC-3 demonstrated binding on these genes (Figures 3.4A and 3.4C). Moreover, our longitudinal genetic analysis combined with the UNC-3 protein depletion experiments in adult MNs demonstrate that UNC-3 is continuously required to maintain terminal identity gene expression.

Interestingly, the analysis on module #2 genes provides new mechanistic insights into how terminal selectors control neuronal identity over time. Instead of constantly activating the same set of genes, as predicted by the above hypothesis, terminal selectors can also modify the suite of their target genes at different life stages (temporal modularity), suggesting their function can be more dynamic than previously thought. This is based on the finding that UNC-3 is selectively required for maintenance, not initiation, of five conserved TF-encoding genes from module #2 (*cfi-1/Arid3a, bnc-1/BNC1-2, mab-9/Tbx20, ceh-44/Cux1, nhr-40*) (Figures 3.4B-C). Hence, some targets (module #1) require constant UNC-3 input from development through adulthood, whereas others (module #2) require UNC-3 only for maintenance.

By honing in on one TF (*cfi-1/Arid3a*) downstream of UNC-3, we identify a mechanism that potentially enables initiation and maintenance of module #2 genes (Figure 3.9G). That is, the Hox proteins LIN-39 and MAB-5 control initiation of *cfi-1/Arid3a* in cholinergic MNs independently of UNC-3, but *cfi-1* maintenance in these same neurons depends on both Hox and UNC-3 (Figure 3.9G). Mechanistically, we propose that Hox-dependent initiation and Hox/UNC-3-dependent maintenance of *cfi-1* are "funneled" through the same distal enhancer, which bears both Hox and UNC-3 binding sites. This suggests embryonic

initiation and post-embryonic maintenance of expression of a particular gene, in a specific cell type, can be achieved by distinct TF combinations acting upon the same *cis*-regulatory region (enhancer) (Figure 3.9G). This somewhat surprising mechanism differs from previous fly and mouse studies reporting distinct and physically separated *cis*-regulatory regions necessary for either initiation or maintenance of cell type-specific gene expression (Ellmeier et al., 2002, Johnson et al., 2011, Manzanares et al., 2001, Pfeffer et al., 2002, Rhee et al., 2016).

In summary, our findings critically extend the mechanisms underlying UNC-3 function. Previous work demonstrated that, in cholinergic MNs, UNC-3 not only activates expression of terminal identity genes (Kratsios et al., 2012), but also prevents alternative neuronal identities (Feng et al., 2020). Here, we report that the suite of UNC-3 targets in these neurons can be partially modified at different life stages, offering key insights into how terminal selectors control neuronal identity over time.

3.5.3 Hox proteins collaborate with stage-specific TFs to establish and maintain MN terminal identity

During early neuronal development, Hox proteins are essential for cell survival, neuronal diversity, and circuit assembly (Baek et al., 2013, Catela et al., 2016, Estacio-Gomez and Diaz-Benjumea, 2014, Karlsson et al., 2010, Miguel-Aliaga and Thor, 2004, Moris-Sanz et al., 2015, Philippidou and Dasen, 2013). However, their post-embryonic neuronal functions remain elusive. Moreover, Hox proteins are often continuously expressed in multiple cell types of a given body region (Baek et al., 2013, Hutlet et al., 2016, Takahashi et al., 2004), raising the question of how do they achieve cell type-specificity in their function. Our findings on mid-body Hox proteins LIN-39 and MAB-5 begin to address this question.

LIN-39 and MAB-5 are continuously expressed in multiple cell types located at the *C. elegans* mid-body region (Clark et al., 1993, Cowing and Kenyon, 1992, Feng et al., 2020, Maloof and Kenyon, 1998). We find that LIN-39 and MAB-5 exert a cell type-specific function; they are required to initiate (in embryo) and maintain (post-embryonically) *cfi-1* expression in specific subsets of cholinergic MNs (DA, DB, VA, VB; Figure 3.7C). Such specificity likely arises through collaboration with distinct TFs responsible for either initiation or maintenance of *cfi-1* (Figure 3.9G). Supporting this scenario, Hox (LIN-39, MAB-5) and UNC-3 are co-expressed in DA, DB, VA, and VB neurons (Feng et al., 2020). Moreover, UNC-3 is selectively required for *cfi-1* maintenance (not initiation) in these neurons (Figure 3.9G). We surmise that other, yet-to-be-identified factors collaborate with Hox proteins at early stages to initiate *cfi-1* expression specifically in DA, DB, VA, and VB neurons. Such initiation-specific factors likely act through the same distal enhancer because its deletion completely abolished *cfi-1* expression in these neurons at early (and late) stages (Figures 3.7D and 3.9G). In summary, we propose that Hox proteins collaborate with distinct TFs over time, that is, initiation-specific factors and the terminal selector UNC-3, to ensure continuous expression of *cfi-1/Arid3a* in specific subtypes of cholinergic MNs. This mechanism may extend to the regulation of terminal identity genes, as we previously showed that LIN-39 and UNC-3 are required to maintain expression of *acr-2*/ human CHRNA1 (acetylcholine receptor subunit) and *unc-77*/ human NALCN (sodium channel) in cholinergic MNs (Feng et al., 2020). Altogether, these findings offer mechanistic insights into the recently proposed hypothesis that Hox proteins in *C. elegans* collaborate with terminal selectors to establish and maintain neuronal terminal identity (Kratsios et al., 2017, Zheng et al., 2015).

3.5.4 Limitations of this study

By conducting a longitudinal analysis for 14 UNC-3 target genes, we identified two groups with distinct temporal requirements (modules #1 and #2 in Figure 3.4C). It is likely though that UNC-3 temporally controls other targets through additional modules. For example, animals lacking *unc-3* display severe axon guidance defects in cholinergic MNs (Prasad et al., 1998), but the underlying mechanisms remain unknown. Since axon guidance molecules are often expressed in a transient fashion during early neuronal development, we speculate that UNC-3 may transiently activate expression of such molecules, a possibility that would add another temporal module in UNC-3 function. Another likely scenario is an "initiation-only" module where UNC-3 is responsible only for onset, but not maintenance, of a yet-to-be-identified set of targets.

In addition, the breadth of biological processes controlled by terminal selectors remains largely unknown. Our ChIP-Seq analysis potentially implicates UNC-3 in a range of biological processes. First, this dataset significantly extends previous reports on the role of UNC-3 in neuronal terminal identity (Kim et al., 2005, Kratsios et al., 2012) by identifying hundreds of terminal identity genes (42.18% of total ChIP-Seq hits) as putative UNC-3 targets. Second, the ChIP-Seq dataset suggests new roles for UNC-3 in neuronal metabolic pathways (24.14% of UNC-3 target genes code for enzymes) and gene regulatory networks (24.07% of UNC-3 targets are TFs and nucleic acid-binding proteins). However, future RNA-Sequencing studies in *unc-3*-depleted MNs are necessary to correlate gene expression changes with UNC-3 ChIP-Seq data, and thereby identify bona fide targets and biological processes under direct UNC-3 control. Such analysis may also uncover any indirect effects on gene expression in *unc-3*-depleted MNs, perhaps arising due to a partial cell fate transformation previously reported for these neurons (Feng et al., 2020).

3.5.5 Temporal modularity may be a shared feature among continuously expressed TFs

This study suggests that the suite of targets of two *C. elegans* terminal selectors (UNC-3/Ebf and UNC-30/Pitx) can be modified over time, providing evidence for temporal modularity in their function. Given that terminal selectors, as well as other neuron type-specific TFs with continuous expression, have been described in both invertebrate and vertebrate species (Deneris and Hobert, 2014, Hobert and Kratsios, 2019, Mayer et al., 2018, Mi et al., 2018), it will be interesting to determine the potential generality of the temporal mechanism described here. Supporting this possibility, the terminal selector of serotonergic neurons in mice (Pet-1) activates expression of serotonin biosynthesis proteins during development, but appears to switch transcriptional targets at later life stages (Wyler et al., 2016). Outside the nervous system, cell type-specific TFs with continuous expression have been described in worms, flies and mice (Pikkarainen et al., 2004, Soler et al., 2012, Wiesenfahrt et al., 2016, Zhou et al., 2017). Future studies will determine whether the principle of temporal modularity is widely employed for the control of cell type identity.

3.6 References

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.

BENJAMINI, Y. Y., D. 2001. The control of the false discovery rate in multiple testing under dependency. The Annals of Statistics, 29, 1165-1188.

BLACKBURN, P. R., BARNETT, S. S., ZIMMERMANN, M. T., COUSIN, M. A., KAI-WAR, C., PINTO, E. V. F., NIU, Z., FERBER, M. J., URRUTIA, R. A., SELCEN, D., KLEE, E. W. & PICHURIN, P. N. 2017. Novel de novo variant in EBF3 is likely to impact DNA binding in a patient with a neurodevelopmental disorder and expanded phenotypes: patient report, in silico functional assessment, and review of published cases. Cold Spring Harb Mol Case Stud, 3, a001743.

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., RO-ZOWSKY, 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.

BROZOVA, E., SIMECKOVA, K., KOSTROUCH, Z., RALL, J. E. & KOSTROUCHOVA, M. 2006. NHR-40, a Caenorhabditis elegans supplementary nuclear receptor, regulates embryonic and early larval development. Mech Dev, 123, 689-701.

CATELA, C. & KRATSIOS, P. 2019. Transcriptional mechanisms of motor neuron development in vertebrates and invertebrates. Dev Biol.

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.

CHAO, H. T., DAVIDS, M., BURKE, E., PAPPAS, J. G., ROSENFELD, J. A., MCCARTY, A. J., DAVIS, T., WOLFE, L., TORO, C., TIFFT, C., XIA, F., STONG, N., JOHNSON, T. K., WARR, C. G., UNDIAGNOSED DISEASES, N., YAMAMOTO, S., ADAMS, D. R., MARKELLO, T. C., GAHL, W. A., BELLEN, H. J., WANGLER, M. F. & MALICDAN, M. C. 2017. A Syndromic Neurodevelopmental Disorder Caused by De Novo Variants in EBF3. Am J Hum Genet, 100, 128-137.

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.

CORBO, J. C., LAWRENCE, K. A., KARLSTETTER, M., MYERS, C. A., ABDELAZIZ, M., DIRKES, W., WEIGELT, K., SEIFERT, M., BENES, V., FRITSCHE, L. G., WEBER, B. H. & LANGMANN, T. 2010. CRX ChIP-seq reveals the cis-regulatory architecture of mouse photoreceptors. Genome Res, 20, 1512-25.

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

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

DOE, C. Q. 2017. Temporal Patterning in the Drosophila CNS. Annu Rev Cell Dev Biol, 33, 219-240.

DUBOIS, L. & VINCENT, A. 2001. The COE–Collier/Olf1/EBF–transcription factors: structural conservation and diversity of developmental functions. Mech Dev, 108, 3-12.

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.

ELLMEIER, W., SUNSHINE, M. J., MASCHEK, R. & LITTMAN, D. R. 2002. Combined deletion of CD8 locus cis-regulatory elements affects initiation but not maintenance of CD8 expression. Immunity, 16, 623-34.

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.

FEDERICI, T. & BOULIS, N. M. 2006. Gene-based treatment of motor neuron diseases. Muscle Nerve, 33, 302-23.

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.

FLAMES, N. & HOBERT, O. 2009. Gene regulatory logic of dopamine neuron differentiation. Nature, 458, 885-9.

GARCIA-BELLIDO, A. 1975. Genetic control of wing disc development in Drosophila. Ciba Found Symp, 0, 161-82.

GREIG, L. C., WOODWORTH, M. B., GALAZO, M. J., PADMANABHAN, H. & MACKLIS, J. D. 2013. Molecular logic of neocortical projection neuron specification, development and diversity. Nat Rev Neurosci, 14, 755-69.

HARMS, F. L., GIRISHA, K. M., HARDIGAN, A. A., KORTUM, F., SHUKLA, A., ALAWI, M., DALAL, A., BRADY, L., TARNOPOLSKY, M., BIRD, L. M., CEULEMANS, S., BEBIN, M., BOWLING, K. M., HIATT, S. M., LOSE, E. J., PRIMIANO, M., CHUNG, W. K., JUUSOLA, J., AKDEMIR, Z. C., BAINBRIDGE, M., CHARNG, W. L., DRUMMONDBORG, M., ELDOMERY, M. K., EL-HATTAB, A. W., SALEH, M. A., BEZIEAU, S., COGNE, B., ISIDOR, B., KURY, S., LUPSKI, J. R., MYERS, R. M., COOPER, G. M. & KUTSCHE, K. 2017. Mutations in EBF3 Disturb Transcriptional Profiles and Cause Intellectual Disability, Ataxia, and Facial Dysmorphism. Am J Hum Genet, 100, 117-127.

HEINZ, S., BENNER, C., SPANN, N., BERTOLINO, E., LIN, Y. C., LASLO, P., CHENG,

J. X., MURRE, C., SINGH, H. & GLASS, C. K. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell, 38, 576-89.

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. 2016a. A map of terminal regulators of neuronal identity in Caenorhabditis elegans. Wiley Interdiscip Rev Dev Biol, 5, 474-98.

HOBERT, O. 2016b. 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.

HOUSSET, M., SAMUEL, A., ETTAICHE, M., BEMELMANS, A., BEBY, F., BILLON, N. & LAMONERIE, T. 2013. Loss of Otx2 in the adult retina disrupts retinal pigment epithelium function, causing photoreceptor degeneration. J Neurosci, 33, 9890-904.

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.

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.

JESSELL, T. M. 2000. Neuronal specification in the spinal cord: inductive signals and transcriptional codes. Nat Rev Genet, 1, 20-9.

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.

JOHNSON, S. A., HARMON, K. J., SMILEY, S. G., STILL, F. M. & KAVALER, J. 2011. Discrete regulatory regions control early and late expression of D-Pax2 during external sensory organ development. Dev Dyn, 240, 1769-78.

KADKHODAEI, B., ALVARSSON, A., SCHINTU, N., RAMSKOLD, D., VOLAKAKIS, N., JOODMARDI, E., YOSHITAKE, T., KEHR, J., DECRESSAC, M., BJORKLUND, A., SANDBERG, R., SVENNINGSSON, P. & PERLMANN, T. 2013. Transcription factor Nurr1 maintains fiber integrity and nuclear-encoded mitochondrial gene expression in dopamine neurons. Proc Natl Acad Sci U S A, 110, 2360-5.

KADKHODAEI, B., ITO, T., JOODMARDI, E., MATTSSON, B., ROUILLARD, C., CARTA, M., MURAMATSU, S., SUMI-ICHINOSE, C., NOMURA, T., METZGER, D., CHAMBON, P., LINDQVIST, E., LARSSON, N. G., OLSON, L., BJORKLUND, A., ICHINOSE, H. & PERLMANN, T. 2009. Nurr1 is required for maintenance of maturing and adult midbrain dopamine neurons. J Neurosci, 29, 15923-32.

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.

KONSTANTINIDES, N., KAPURALIN, K., FADIL, C., BARBOZA, L., SATIJA, R. & DESPLAN, C. 2018. Phenotypic Convergence: Distinct Transcription Factors Regulate Common Terminal Features. Cell, 174, 622-635 e13.

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.

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.

LANGMEAD, B. & SALZBERG, S. L. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods, 9, 357-9.

LEYVA-DIAZ, E. & HOBERT, O. 2019. Transcription factor autoregulation is required for acquisition and maintenance of neuronal identity. Development, 146.

LIU, J. & FIRE, A. 2000. Overlapping roles of two Hox genes and the exd ortholog ceh-20 in diversification of the C. elegans postembryonic mesoderm. Development, 127, 5179-90.

LODATO, S. & ARLOTTA, P. 2015. Generating neuronal diversity in the mammalian cerebral cortex. Annu Rev Cell Dev Biol, 31, 699-720.

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.
MANZANARES, M., BEL-VIALAR, S., ARIZA-MCNAUGHTON, L., FERRETTI, E., MARSHALL, H., MACONOCHIE, M. M., BLASI, F. & KRUMLAUF, R. 2001. Independent regulation of initiation and maintenance phases of Hoxa3 expression in the vertebrate hindbrain involve auto- and cross-regulatory mechanisms. Development, 128, 3595-607.
MAYER, C., HAFEMEISTER, C., BANDLER, R. C., MACHOLD, R., BATISTA BRITO, R., JAGLIN, X., ALLAWAY, K., BUTLER, A., FISHELL, G. & SATIJA, R. 2018. Developmental diversification of cortical inhibitory interneurons. Nature, 555, 457-462.
MI, D., LI, Z., LIM, L., LI, M., MOISSIDIS, M., YANG, Y., GAO, T., HU, T. X., PRATT, T., PRICE, D. J., SESTAN, N. & MARIN, O. 2018. Early emergence of cortical interneuron diversity in the mouse embryo. Science, 360, 81-85.
MI, H., MURUGANUJAN, A., CASAGRANDE, J. T. & THOMAS, P. D. 2013. Large-scale gene function analysis with the PANTHER classification system. Nat Protoc, 8, 1551-66.
MIGUEL-ALIAGA, I. & THOR, S. 2004. Segment-specific prevention of pioneer neuron apoptosis by cell-autonomous, postmitotic Hox gene activity. Development, 131, 6093-105.
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.
OKATY, B. W., FRERET, M. E., ROOD, B. D., BRUST, R. D., HENNESSY, M. L., DEBAIROS, D., KIM, J. C., COOK, M. N. & DYMECKI, S. M. 2015. Multi-Scale Molecular Deconstruction of the Serotonin Neuron System. Neuron, 88, 774-91.
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.
PERRY, M., KONSTANTINIDES, N., PINTO-TEIXEIRA, F. & DESPLAN, C. 2017. Gen-

eration and Evolution of Neural Cell Types and Circuits: Insights from the Drosophila Visual System. Annu Rev Genet, 51, 501-527.

PFEFFER, P. L., PAYER, B., REIM, G., DI MAGLIANO, M. P. & BUSSLINGER, M. 2002. The activation and maintenance of Pax2 expression at the mid-hindbrain boundary is controlled by separate enhancers. Development, 129, 307-18.

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

PIKKARAINEN, S., TOKOLA, H., KERKELA, R. & RUSKOAHO, H. 2004. GATA transcription factors in the developing and adult heart. Cardiovasc Res, 63, 196-207.

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.

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.

RAMIREZ, F., RYAN, D. P., GRUNING, B., BHARDWAJ, V., KILPERT, F., RICHTER, A. S., HEYNE, S., DUNDAR, F. & MANKE, T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res, 44, W160-5.

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.

SABARIS, G., LAIKER, I., PREGER-BEN NOON, E. & FRANKEL, N. 2019. Actors with Multiple Roles: Pleiotropic Enhancers and the Paradigm of Enhancer Modularity. Trends Genet, 35, 423-433.

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.

SCOTT, M. M., KRUEGER, K. C. & DENERIS, E. S. 2005. A differentially autoregulated Pet-1 enhancer region is a critical target of the transcriptional cascade that governs serotonin neuron development. J Neurosci, 25, 2628-36.

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.

SHAHAM, S. & BARGMANN, C. I. 2002. Control of neuronal subtype identity by the C. elegans ARID protein CFI-1. Genes Dev, 16, 972-83.

SHEN, L., SHAO, N., LIU, X. & NESTLER, E. 2014. ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. BMC Genomics, 15, 284.

SLEVEN, H., WELSH, S. J., YU, J., CHURCHILL, M. E., WRIGHT, C. F., HENDERSON, A., HORVATH, R., RANKIN, J., VOGT, J., MAGEE, A., MCCONNELL, V., GREEN, A., KING, M. D., COX, H., ARMSTRONG, L., LEHMAN, A., NELSON, T. N., DECIPHERING DEVELOPMENTAL DISORDERS, S., STUDY, C., WILLIAMS, J., CLOUSTON, P., HAGMAN, J. & NEMETH, A. H. 2017. De Novo Mutations in EBF3 Cause a Neurodevelopmental Syndrome. Am J Hum Genet, 100, 138-150.

SOLER, C., HAN, J. & TAYLOR, M. V. 2012. The conserved transcription factor Mef2 has multiple roles in adult Drosophila musculature formation. Development, 139, 1270-5.

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.

TREIBER, N., TREIBER, T., ZOCHER, G. & GROSSCHEDL, R. 2010a. Structure of an Ebf1:DNA complex reveals unusual DNA recognition and structural homology with Rel proteins. Genes Dev, 24, 2270-5.

TREIBER, T., MANDEL, E. M., POTT, S., GYORY, I., FIRNER, S., LIU, E. T. & GROSSCHEDL, R. 2010b. Early B cell factor 1 regulates B cell gene networks by activation, repression, and transcription- independent poising of chromatin. Immunity, 32, 714-25.

WANG, M. M. & REED, R. R. 1993. Molecular cloning of the olfactory neuronal transcription factor Olf-1 by genetic selection in yeast. Nature, 364, 121-6.

WANG, M. M., TSAI, R. Y., SCHRADER, K. A. & REED, R. R. 1993. Genes encoding components of the olfactory signal transduction cascade contain a DNA binding site that may direct neuronal expression. Mol Cell Biol, 13, 5805-13.

WANG, S. S., TSAI, R. Y. & REED, R. R. 1997. The characterization of the Olf-1/EBF-like HLH transcription factor family: implications in olfactory gene regulation and neuronal development. J Neurosci, 17, 4149-58.

WIESENFAHRT, T., BERG, J. Y., OSBORNE NISHIMURA, E., ROBINSON, A. G., GOSZCZYNSKI, B., LIEB, J. D. & MCGHEE, J. D. 2016. The function and regulation of the GATA factor ELT-2 in the C. elegans endoderm. Development, 143, 483-91.

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.

XUE, D., FINNEY, M., RUVKUN, G. & CHALFIE, M. 1992. Regulation of the mec-3 gene

by the C.elegans homeoproteins UNC-86 and MEC-3. EMBO J, 11, 4969-79.

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.

YU, G., WANG, L. G. & HE, Q. Y. 2015. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics, 31, 2382-3.

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.

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

ZHOU, Q., LIU, M., XIA, X., GONG, T., FENG, J., LIU, W., LIU, Y., ZHEN, B., WANG, Y., DING, C. & QIN, J. 2017. A mouse tissue transcription factor atlas. Nat Commun, 8, 15089.

CHAPTER 4

CONCLUSIONS AND PERSPECTIVES

4.1 Delineating the spatial and temporal regulatory mechanisms underlying neuronal terminal identity

In this thesis, we report that in the *C. elegans* VNC, MN identity is defined by the spatial and temporal expression pattern of both terminal identity genes and their transcriptional regulators.

From the spatial perspective, each terminal identity gene is expressed in a unique subset of MNs. The combination of different terminal identity genes, whether expressed or not in any given MN, represents a signature that gives rise to the specific terminal properties of that particular MN. Based on the work in this thesis, we find that the neuron type-specific expression of the terminal identity gene *glr-4* directly results from transcription factors expressed in a region-specific manner. That is, the two Hox genes *lin-39* and *mab-5* are specifically expressed in MNs located along midbody region of the VNC, the terminal selector UNC-3 is expressed in all cholinergic MNs necessary for locomotion, and the novel ARID transcription factor CFI-1 that counteracts UNC-3 is expressed only in some (DA, DB, VA, and VB) but not all cholinergic MNs.

From the temporal perspective, we highlight that *continuously expressed* terminal identity genes tightly related to the properties of a neuron must be maintained to make the neuron remain functioning throughout life. Although it is possible that certain terminal identity genes are only *transiently expressed* during early development and may not need to be maintained during later stages, this thesis makes an effort in filling our knowledge gap in late development of post-mitotic neurons. Therefore, we focus on the terminal identity

gene *glr-4*, whose expression is maintained at low levels in cholinergic MNs. To this end, we identified that the ARID TF CFI-1 is also continuously required in adulthood to maintain repression of *glr-4*. Our results argue that aside from the spatial features of TF expression patterns (i.e. the cell type-specificity in their expression), the temporal features (i.e. the developmental stage-specificity in their expression) are as important, if not more so, to a well understanding of neuronal identity regulation. Naturally, this leads to the study of the regulatory mechanisms underlying the temporal expression features of TFs in neurons. In this thesis, we follow up with this question by investigating the upstream regulation of *cfi-1*.

One particularly interesting observation from our findings is the existence of a gene regulatory network featuring *lin-39/Scr/Dfd/Hox3-5a*, *mab-5/Antp/Hox6-8*, *unc-3/Ebf*, *cfi-1/Arid3a*, and *glr-4/GRIK1*. In a MN where all 5 genes are expressed, there are several events happening which are worth highlighting here. First, CFI-1 requires different mechanisms for the initiation (regulated by Hox proteins LIN-39 and MAB-5) and maintenance (regulated by UNC-3) of its expression. This is interesting but not surprising considering CFI-1 may be a critical player in neuronal identity maintenance and its continuous expression must be ensured by UNC-3 with an extra layer of safety apart from the Hox genes. The importance of continuous requirement of CFI-1 is strongly suggested by the locomotion defects observed when we disrupt maintenance of *cfi-1* expression by mutating the binding sites of UNC-3. This implies that CFI-1 is required to maintain a broader range of terminal identity genes in addition to *glr-4*. The ChIP-Seq analysis of CFI-1 described in this work provides an excellent entry point for identifying global CFI-1 targets in an unbiased way. Second, LIN-39, MAB-5, and UNC-3 all function as activators of both *cfi-1* and *glr-4* expression. Because CFI-1 represses *glr-4* expression, this can be summarized as the midbody Hox proteins and UNC-3 activate directly and repress indirectly the expression of *glr-4* at the same time. This interesting gene regulatory pattern has been discovered in many species

including bacteria (Eichenberger et al., 2004), yeast (Lee et al., 2002), and animal cells (e.g. human embryonic stem cells (Boyer et al., 2005) and hematopoietic stem cells (Swiers et al., 2006)), and it is termed the incoherent feedforward loop (I1-FFL). A previous study reports that I1-FFL provides a mechanism for fold-change detection in gene regulation (Goentoro et al., 2009). This suggests that the I1-FFL we identified in *C. elegans* MNs may be a practical strategy for keeping the expression level of terminal identity genes in check. Supporting this possibility, we observed in *cfi-1* mutants higher expression levels of *glr-4* in cells that already express *glr-4* in wildtype animals (Figure. 2.4D). Whether these high *glr-4* levels will lead to toxicity remains an open question and is worth exploring in the future. A straightforward starting point would be testing the effect of *glr-4* overexpression in MNs.

The fact that both positive and negative regulation (albeit indirect) of *glr-4* can be achieved by the Hox proteins and UNC-3 may imply a general mechanism for transcription factor functions from the evolutionary perspective. That is, evolution has found a way to economize the number of terminal selectors required to define neuronal identities. Since the responsibilities of terminal selectors usually include the activation of a large pool of downstream target genes, having intermediate gene regulators rather than additional terminal selectors to take care of the fine tuning of terminal identity gene expression level may be an economical way to regulate and maintain a variety of neuron types in the nervous system. Because the transcription factors studied here are expressed in the nervous system of invertebrate and vertebrate animals (Green and Vetter, 2011; Joliot et al., 1991; Kratsios et al., 2011; Lou et al., 1995; Shaham and Bargmann, 2002), we propose that our findings may hint towards a general mechanism for the control of neuronal identity possibly adopted by other species as well.

4.2 Antagonistic functions of transcription factors directly contribute to the maintenance of neuronal identity

4.2.1 Transcription factor activators and repressors occupy the same cis-regulatory genomic regions and counteract each other to regulate gene expression

Our work reveals that the activator UNC-3 and repressor CFI-1 act via the same promoter of *glr-4* to regulate its expression. There are two hypothetical models to explain this phenomenon. First, the two transcription factors may compete for binding to the promoter. This is a common regulatory mechanism for gene regulation that has been described in other work (Darieva et al., 2010; Karreth et al., 2014; Zhang et al., 2021). Alternatively, the binding of one transcription factor may not necessarily prevent the other factor from occupying the promoter region but rather inhibit the function of the other factor.

Regarding the first model, although the binding sites of UNC-3 and CFI-1 within the *glr-4* promoter are not immediately adjacent to each other, we cannot rule out this possibility of binding competition. To further explore the antagonistic roles between UNC-3 and CFI-1, we compared their genome-wide binding patterns from the ChIP-Seq analyses. Surprisingly, we found all UNC-3 binding peaks overlap with CFI-1 binding peaks, indicating specific co-occupation of the two TFs over the whole *C. elegans* genome. This interesting observation inspires a new for studying gene regulation by transcription factors. We note however that while our ChIP-Seq analyses use whole worm lysates we are not able to distinguish the binding signal with cell type resolution. When studying a novel transcription factor, it is worth trying to obtain its genome-wide binding landscape and then either comparing it with available binding signals of known TFs or performing motif analyses over the binding landscape, which may yield abundance of binding sites of known TFs.

4.2.2 Maintenance effect from transcription factor may require binding events to take place during early development

Our investigation for the initiation and maintenance requirements for the expression of the ARID transcription factor *cfi-1* revealed temporal modularity in the function of the terminal selector UNC-3. While UNC-3 is required to initiate a group of cholinergic MN terminal identity genes during early development, it is dispensable for the initiation of *cfi-1* but indispensable for maintaining its expression in adulthood. Furthermore, we characterized that UNC-3 directly maintains *cfi-1* via binding to a distal enhancer. An interesting mechanistic question immediately arising from this observation is when does the binding need to take place for a terminal selector to maintain its targets if it is not required for their initial expression? Although it is reasonable to hypothesize that UNC-3 does not need to bind to the *cfi-1* enhancer in early stages, our ChIP-Seq results suggest otherwise. Our ChIP-Seq analysis was performed with worm extracts from the L2 stage, and it shows UNC-3 binding at this developmental stage. However, *cfi-1* expression is unaffected in unc-3 mutants at L2. The fact that UNC-3 binding takes place much earlier than when it is required to regulate *cfi-1* expression is intriguing. It suggests that even if transcription factors are required only for the maintenance of downstream target gene expression, the maintenance effect may still require preparation and it takes time for the transcription factor to "prepare the enhancer" for later action. Such phenomenon has been described as "chromatin priming" in the epigenetic field of study (Bonifer and Cockerill, 2017). This preparation is probably necessary for the transcription factor to have enough time to recruit co-factors and form a functioning complex. Indeed, transcriptional regulation is commonly achieved through collaboration between transcription factors and co-factors (Reiter et al., 2017; Wang et al., 2021). It is possible that the co-factors required for *cfi-1* maintenance are expressed and recruited to

the *cfi-1* enhancer later and only in adults. If this is the case, it aligns well with our discussion in section 4.1 and suggests that a combination of TF functions from the temporal perspective may play a critical role in regulating neural identity at different developmental stages. Alternatively, the reason that UNC-3 is required to maintain *cfi-1* expression could be that the activity of Hox proteins in adults is no longer needed to regulate *cfi-1*. Our inducible protein degradation experiment showed that LIN-39 has no maintenance effect on *cfi-1* expression. To rigorously test this hypothesis, however, binding events of Hox genes at the *cfi-1* enhancer need to be investigated in adult animals. Due to technical challenges, we were not able to perform this analysis.

4.3 Useful ChIP-Seq resources provided from this study for future work

In this thesis, two ChIP-Seq datasets for the TFs CFI-1/Arid3a and UNC-3/Ebf are described. These datasets are available to the community via open source repository (see Materials and Methods in Chapters 2 and 3 for details). We believe these datasets are a valuable asset for researchers studying gene regulation, as both TFs are highly conserved across species and little is known about their neuronal functions. It is important to emphasize that the ChIP-Seq analyses were performed by pulling down the endogenous UNC-3 and CFI-1 proteins which were tagged using the CRISPR/Cas9 system. Homozygous animals carrying alleles for these tagged fusion proteins were carefully examined and no obvious phenotypes were found, suggesting the function of UNC-3 and CFI-1 was not affected by tagging. When combined with other valuable resources, for example the RNA-seq data from the CeNGEN project, the ChIP-Seq datasets presented here should be helpful for further investigations in gene regulation and the control of neuronal identity.

4.4 References

Bonifer, C., and Cockerill, P.N. (2017). Chromatin priming of genes in development: Concepts, mechanisms and consequences. Exp. Hematol. 49, 1–8. Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell 122, 947–956.

Darieva, Z., Clancy, A., Bulmer, R., Williams, E., Pic-Taylor, A., Morgan, B.A., and Sharrocks, A.D. (2010). A competitive transcription factor binding mechanism determines the timing of late cell cycle-dependent gene expression. Mol. Cell 38, 29–40.

Eichenberger, P., Fujita, M., Jensen, S.T., Conlon, E.M., Rudner, D.Z., Wang, S.T., Ferguson, C., Haga, K., Sato, T., Liu, J.S., et al. (2004). The program of gene transcription for a single differentiating cell type during sporulation in Bacillus subtilis. PLoS Biol. 2, e328.

Goentoro, L., Shoval, O., Kirschner, M.W., and Alon, U. (2009). The incoherent feedforward loop can provide fold-change detection in gene regulation. Mol. Cell 36, 894–899.

Green, Y.S., and Vetter, M.L. (2011). EBF factors drive expression of multiple classes of target genes governing neuronal development. Neural Dev. 6, 19.

Joliot, A., Pernelle, C., Deagostini-Bazin, H., and Prochiantz, A. (1991). Antennapedia homeobox peptide regulates neural morphogenesis. Proc. Natl. Acad. Sci. USA 88, 1864–1868.

Karreth, F.A., Tay, Y., and Pandolfi, P.P. (2014). Target competition: transcription factors enter the limelight. Genome Biol. 15, 114.

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

Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I., et al. (2002). Transcriptional regulatory

networks in Saccharomyces cerevisiae. Science 298, 799–804.

Lou, L., Bergson, C., and McGinnis, W. (1995). Deformed expression in the Drosophila central nervous system is controlled by an autoactivated intronic enhancer. Nucleic Acids Res. 23, 3481–3487.

Reiter, F., Wienerroither, S., and Stark, A. (2017). Combinatorial function of transcription factors and cofactors. Curr. Opin. Genet. Dev. 43, 73–81.

Shaham, S., and Bargmann, C.I. (2002). Control of neuronal subtype identity by the *C. elegans* ARID protein CFI-1. Genes Dev. 16, 972–983.

Swiers, G., Patient, R., and Loose, M. (2006). Genetic regulatory networks programming hematopoietic stem cells and erythroid lineage specification. Dev. Biol. 294, 525–540.

Wang, Z., Wang, P., Li, Y., Peng, H., Zhu, Y., Mohandas, N., and Liu, J. (2021). Interplay between cofactors and transcription factors in hematopoiesis and hematological malignancies. Signal Transduct. Target. Ther. 6, 24.

Zhang, Y., Ho, T.D., Buchler, N.E., and Gordân, R. (2021). Competition for DNA binding between paralogous transcription factors determines their genomic occupancy and regulatory functions. Genome Res.

APPENDIX A

TRANSGENIC REPORTER ANALYSIS OF CHIP-SEQ-DEFINED ENHANCERS IDENTIFIES NOVEL TARGET GENES FOR THE TERMINAL SELECTOR UNC-3/COLLIER/EBF

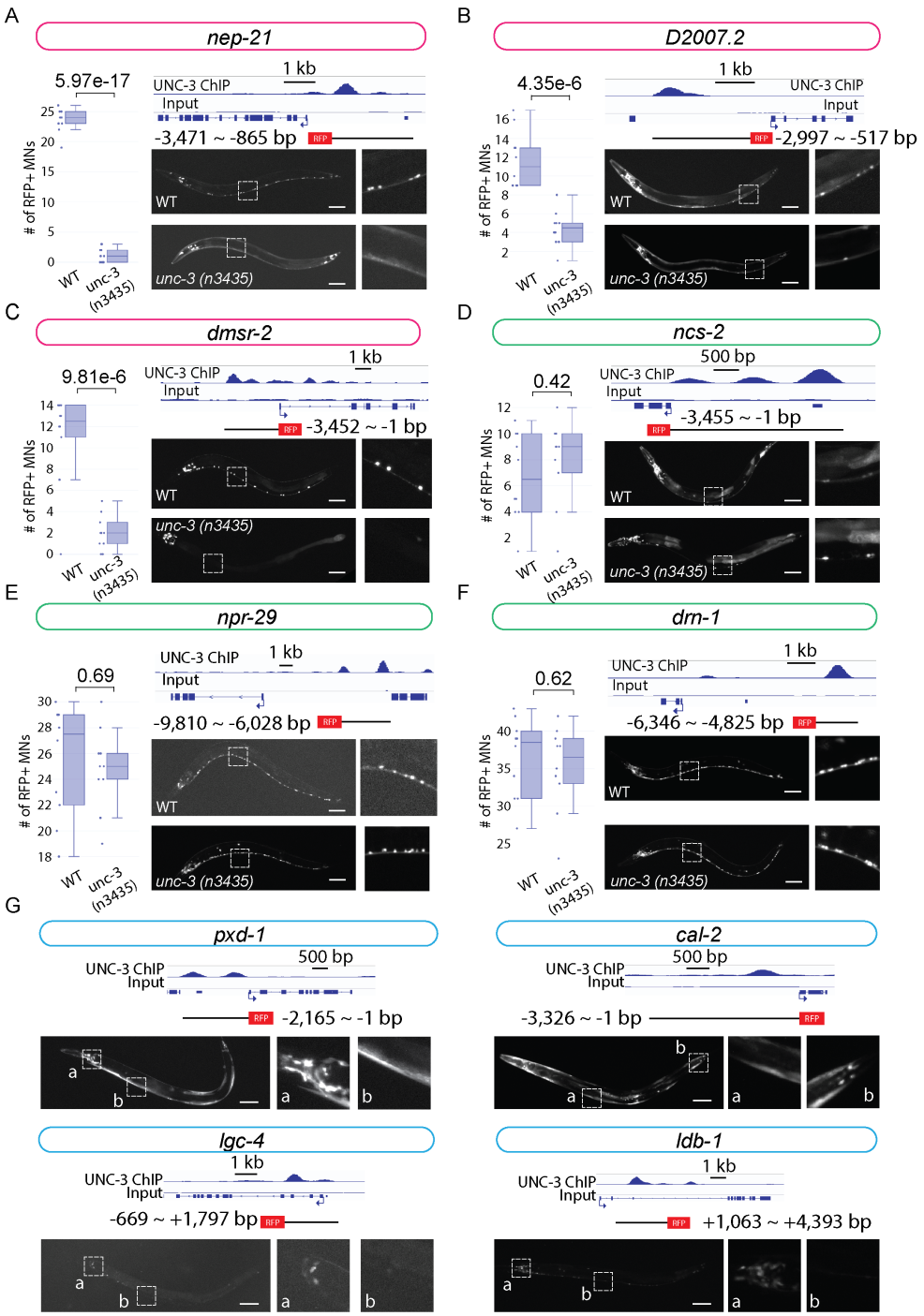


Figure A.1: **Survey for novel target genes of the terminal selector UNC-3.** UNC-3 ChIP-Seq signal is aligned to specific gene loci of interest. Reporter constructs carrying the genomic region bound by UNC-3 followed by tagRFP are used to build transgenic animals, and the expression images are shown (scale bar: 50 µm).

Figure A.1, continued.

A-F, Six reporters are expressed in MNs and their *unc-3* dependency is assessed. The expression of fluorescent reporters is imaged and compared between wildtype and *unc-3 (n3435)* mutant animals. The number of ventral nerve cord MNs showing tagRFP expression is quantified for each genotype (N=10) and shown in boxplots. Student t test is performed and statistical significance is determined by p-value. Three reporter genes (*nep-21, D2007.2, dmsr-2*) show decrease of expression in *unc-3 (n3435)* mutants (**A-C**), while the remaining three genes (*ncs-2, npr-29, drn-1*) do not show changes in their expression (**D-F**).
G, Four reporter genes (*pxd-1, cal-2, lgc-4*, and *ldb-1*) are not expressed in MNs of the ventral cord but are expressed in head neurons (all 4 genes), tail neurons (*pxd-1* and *cal-2*), and muscle or epidermal cells (*pxd-1* and *cal-2*). Zoomed insets are shown on the right of each image.