Establishment and maintenance of motor neuron identity via temporal modularity in terminal selector function

Yinan Li [1, 2], Anthony Osuma [1, 2], Edgar Correa [1, 3], Munachi Okebalama[1], Pauline Dao[1], Olivia Gaylord [4], Jihad Aburas [1], Priota Islam [5, 6], André E.X. Brown [5, 6], Paschalis Kratsios [1, 2, 3, 4, 7]

[1] Department of Neurobiology, University of Chicago, Chicago, IL, USA

[2] Committee on Neurobiology, University of Chicago, Chicago, IL, USA

[3] Cell and Molecular Biology Program, University of Chicago, Chicago, IL, USA

[4] Committee on Development, Regeneration and Stem Cell Biology, University of Chicago, Chicago, IL, USA

[5] MRC London Institute of Medical Sciences, London, UK

[6] Institute of Clinical Sciences, Imperial College London, London, UK

[7] The Grossman Institute for Neuroscience, Quantitative Biology, and Human Behavior, University of Chicago, Chicago, IL, USA

Correspondence: pkratsios@uchicago.edu

ABSTRACT

Terminal selectors are transcription factors (TFs) that establish during development and maintain throughout life the identity of individual neuron types, but the underlying mechanisms remain poorly defined. Here, we use ChIP-Sequencing to globally identify the direct targets of UNC-3/Ebf, the terminal selector of cholinergic motor neurons (MNs) in *C. elegans*. Unexpectedly, we find two groups of target genes with distinct temporal requirements for UNC-3. One group encodes multiple classes of proteins (e.g., receptors, secreted molecules) and requires UNC-3 for both embryonic initiation and post-embryonic maintenance of expression. The second group solely encodes TFs critical for MN identity, whose maintenance, but not initiation, requires UNC-3. Hence, the suite of UNC-3 targets in cholinergic MNs is modified across different life stages, revealing “temporal modularity” in terminal selector function. Removing the ability of UNC-3 to maintain expression of a single TF (*cfi-1/Arid3a*) in MNs led to locomotion defects, indicating minimal disruption of temporal modularity affects animal behavior. Another terminal selector (UNC-30/Pitx) also exhibits temporal modularity, supporting the generality of this mechanism for the control of neuronal identity.

INTRODUCTION

The development of multicellular organisms is a multi-step process that culminates in the generation of distinct cell types. How individual cell types acquire and maintain their unique identity represents a central goal in the field of developmental biology. Seminal studies across many model systems have shown that certain transcription factors (TFs) play a central role in the control of cell type identity. These cell type-specific factors define cellular identity through the activation of cell type-specific effector gene batteries [1-3]. Because they are often continuously expressed throughout the life of the cell, such TFs are required for not only the initial specification, but also the maintenance of a unique cellular identity [4].

In the vertebrate nervous system, the advent of single-cell RNA-Sequencing has enabled the identification of multiple TFs with continuous expression in individual neuron types [5-8]. Despite the presence of these neuron type-specific TFs throughout the brain, our understanding of their function across different life stages remains rudimentary, in part due to the difficulty of tracking individual neuron types in the complex vertebrate nervous system throughout embryonic and post-natal life. However, such tracking is possible with single-cell resolution in the simple nervous system of the nematode *Caenorhabditis elegans,* comprised of 302 neurons that fall into 118 neuron types, offering an opportunity to study how individual neuron types acquire and maintain their identity.

Over the past two decades, a comprehensive genetic analysis of continuously expressed, neuron type-specific TFs has been performed in *C. elegans*. This analysis gave rise to the concept of “terminal selectors” [2, 9-11]. A defining feature of terminal selectors is the ability to broadly co-regulate effector genes, whose protein products determine the terminally differentiated state, and thereby identity, 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 [2, 9, 10]. Mechanistically, terminal selectors are thought to act directly, via their cognate binding sites, to activate the expression of terminal identity genes. To date, terminal selectors have been identified via genetic analyses for more than half of the 118 *C. elegans* neuron types [12], indicating their prominent role in this model organism as

critical determinants of neuronal identity. Multiple cases of terminal selectors have also been identified in flies, chordates and mice [4, 13, 14], 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 [4].

However, our understanding of the molecular mechanisms through which terminal selectors establish and maintain neuronal identity remains rudimentary, in part due to two major challenges. First, the unbiased identification of their downstream targets is a mandatory first step towards elucidating terminal selector function. While this is becoming a reality for a limited number of terminal selectors [15-19], the majority of studies follow a candidate approach focused on a select subset of terminal identity genes [12, 20, 21]. Hence, the extent of terminal identity features and the breadth of biological processes controlled by these important regulators remain largely unexplored. Second, terminal selectors are continuously expressed – from development through adulthood – in specific neuron types. However, whether their function in a particular neuron type remains the same, or changes at different life stages is not known. Hence, longitudinal studies 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 target genes can change over time.

To address the aforementioned challenges, we focused on UNC-3, the sole *C. elegans* ortholog of the Collier/Olf/Ebf (COE) family of TFs [22]. UNC-3 acts as a terminal selector in cholinergic motor neurons (MNs) of the *C. elegans* ventral nerve cord [23]. Importantly, mutations in EBF3, a human ortholog of UNC-3, cause a neurodevelopmental syndrome characterized by motor developmental delay [24-27]. Using a candidate approach, it has been previously shown that UNC-3 controls cholinergic MN identity by co-regulating expression of various terminal identity genes (e.g., acetylcholine biosynthesis components, ion channels, NT receptors, neuropeptides) [23]. Animals lacking *unc-3* gene activity display severe locomotion defects [28, 29], suggesting that UNC-3 may act as a master regulator and 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 provide a map of *cis*-regulatory regions bound by UNC-3 in the *C.*

elegans genome. This map enabled the unbiased identification of ~3,500 protein-coding genes as putative direct targets of UNC-3. Protein class ontology analysis suggests that UNC-3, besides neuronal terminal identity, also controls additional biological processes, such as metabolic pathways and gene regulatory networks (24.07% of UNC-3 targets code for TFs and nucleic acid-binding proteins). These findings bring us one step closer to obtaining a comprehensive understanding of how terminal selectors control neuronal identity and 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 the same cell type (cholinergic MNs). One group encodes for multiple classes of proteins (e.g., receptors, secreted molecules, TFs) that require UNC-3 both for embryonic initiation and post-embryonic maintenance of their expression. We refer to this group as “initiation and maintenance” module. 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. We refer to this group as “maintenance-only” module. Hence, the suite of UNC-3 targets in cholinergic MNs is modified across different life stages, a phenomenon we term “temporal modularity” in UNC-3 function.

Furthermore, we find that different mechanisms control gene expression in each module. Genes from the “initiation and maintenance” module are activated directly and continuously by UNC-3. Genes from the “maintenance-only” module (e.g., *cfi-1/Arid3a)* utilize distinct initiation and maintenance strategies. In brief, we find that the Hox proteins LIN-39 (Scr/Dfd/Hox4-5) and MAB-5 (Antp/Hox6-8) control initiation of *cfi-1* in cholinergic MNs independently of UNC-3, but *cfi-1* maintenance in these neurons depends on both Hox and UNC-3. Importantly, CRISPR/Cas9-mediated removal of the ability of UNC-3 to selectively maintain expression of a single target gene (*cfi-1*) in MNs led to locomotion defects, indicating that 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* [30, 31]. In summary, we propose that terminal selectors establish and maintain neuronal identity by modifying the suite of their target genes at different life stages. We

hypothesize that temporal modularity in TF function may be a broadly applicable principle for the control of cellular identity.

RESULTS

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 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 1A**), indicating that this reporter faithfully recapitulates the endogenous *unc-3* expression pattern [32, 33]. 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 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) (**Figure 1C** and **S1A**). The majority of UNC-3 binding peaks (91.95%) are predominantly located between 0 and 3 kb upstream of a transcription start site (TSS), 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 (**Figure 1D** and **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, which highly resembles the binding site of UNC-3 vertebrate orthologs (**Figure 1E**) [35-38]. 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, which are the majority of *unc-3*-expressing cells (53 out of 72 cells). Previous studies identified 10 terminal identity genes as direct UNC-3 targets and 43 terminal identity genes whose expression is affected by genetic removal of *unc-3* [23, 39]. 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 (**Figure 2A** and **S1B, Supplemental Table 1**). Moreover, ChIP-Seq for UNC-3 appears

highly sensitive, as it identified peaks in *unc-3*-dependent genes expressed in a limited number neurons (e.g., *glr-4/GluR* is expressed in 4 out of 72 *unc-3+* neurons) (**Figure 2A**). In conclusion, our ChIP-Seq experiment generated a comprehensive map of UNC-3 binding in the *C. elegans* genome, which provides biochemical evidence 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 [40]. This analysis revealed three broad categories (**Figure 2B, Supplemental Table 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 highlights 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 [32]. 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 [41]. We found that 52.95% of these MN-expressed genes are bound by UNC-3 and fall in the aforementioned gene categories (**Supplemental Table 3**), thereby constituting putative new UNC-3 targets in cholinergic MNs.

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 2B-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 15 elements located upstream of 15 randomly selected TFs of different families (e.g., homeobox, nuclear hormone receptors, Zn finger) (**Figure 2C, Supplemental Table 4**). We generated transgenic reporter animals and found that 10 of these *cis*-regulatory elements were sufficient to drive RFP expression in ventral cord cholinergic MNs (**Supplemental Table 4**). Expression of five of the 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*, and *ceh-44/Cux1,* are known to be expressed in subsets of *unc-3* positive MNs [42-44]. Of note, some of these TF reporters are also expressed in *unc-3*-negative MNs of the nerve cord, namely the GABAergic DD and VD neurons (**Supplemental Table 4**). Next, we tested for *unc-3* dependency at the L4 stage, and found that 9 of these 10 TF reporters depend on *unc-3* for proper expression in cholinergic MNs (**Supplemental Table 4**). Eight reporters (*nhr-1, nhr-40, zfh-2, ztf-26, ceh-44, cfi-1, bnc-1, mab-9)* are positively controlled by *unc-3,* wherea*s* one (*nhr-49*) is negatively regulated by *unc-3* (**Figure S2A-B, Supplemental Table 4**). 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 [23, 39], these findings suggest that UNC-3 can also act directly to either activate or repress expression of multiple TF-encoding genes, uncovering an extensive gene regulatory network downstream of UNC-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 question because its direct targets (terminal identity genes and newly

identified TFs [**Supplemental Table 4**]) are continuously expressed in cholinergic MNs. However, *unc-3* dependency of terminal identity genes was mostly tested at a single larval stage (L4) [23]. 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 3A**). However, a different picture emerged after testing the 8 TF reporters that are positively regulated by UNC-3 (**Supplemental Table 4**). 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 S2A**), 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 3B**). However, maintenance of their expression during late larval and/or adult stages does depend on UNC-3 (**Figure 3B**). Hence, this longitudinal analysis revealed two groups of targets genes 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, **Figure 3A, C**). 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, **Figure 3B-C**).

Collectively, these findings suggest that, in cholinergic MNs, the suite of UNC-3 target genes can change at different life stages; distinct UNC-3 targets are temporally controlled via either "initiation and maintenance" or "maintenance-only" modules. To describe this phenomenon, we use the term "temporal modularity in UNC-3 function" given that the function of a cell type-specific TF at a particular life stage is determined by the suite of targets it controls at that stage (**Figure 3C**). 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.

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 3B**). 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 (**Figure 3B** and **4A**). 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 4A**). Surprisingly, this analysis did not reveal a separate initiation element. Instead, it identified a 2.5 kb distal element (reporter #7) demarcated by the UNC-3 binding peak (**Figure 4A**) that drives RFP expression in subsets of *unc-3*-expressing cholinergic MNs (DA, DB, VA, VB subtypes) at embryonic (3-fold), larval and adult stages (**Figure 4A, 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 (**Figure 4A, C**), which will be discussed later in Results. We conclude that this 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,* that 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 (**Figure 4A-C**), a pattern consistent with previous studies describing *cfi-1* expression [42, 45]. We determined the onset of *mNG::AID::cfi-1* expression in MNs to be at the 3-fold embryonic stage, coinciding with the onset of transgenic reporters containing the distal enhancer (**Figure 4B**). 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 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) (**Figure 4A-B**, quantification specifically for cholinergic MNs shown in **4D**).

We conclude that a distal enhancer (located ~12 kb upstream of *cfi-1*) is both necessary and sufficient for *cfi-1* expression in nerve cord MNs. Genome editing suggests that a 769 bp sequence within this enhancer is required for both initiation and maintenance of *cfi-1* in nerve cord MNs (**Figure 4D**). 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.

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 4A**). However, the UNC-3 peak is spread across several hundred of 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 4E**). 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 [35, 46] (**Figure 4E**). 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 4F**). Intriguingly, *mNG* expression is significantly down-regulated in cholinergic MNs at later larval (L4) and adult (day 1) stages, recapitulating the phenotype of *unc-3* null mutants (**Figure 4F**). 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 [47-49]. 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 S5**), excluding a potential involvement of CFI-1 in its own maintenance.

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 [50]. 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*.

By performing a longitudinal analysis on *lin-39; mab-5* double null mutants [51], we 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 5A-E**). Since onset of *cfi-1* expression occurs at the 3-fold embryonic stage (**Figure 4B**), these results suggest LIN-39 and MAB-5 are required for *cfi-1* initiation. Conversely, initiation of *cfl-1* expression is not affected in a null mutant of *unc-3* (**Figure 5B-C**). Moreover, available ChIP-Seq data from modENCODE [52] indicate that LIN-39 and MAB-5 bind to the *cfi-1* distal enhancer (**Figure 5A**). 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 5A, 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 5A, E**). This decrease was weaker compared to *lin-39; mab-5* double mutants (**Figure 5C**), 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 5F**). We conclude that Hox proteins LIN-39 and MAB-5 are necessary for *cfi-1* initiation in cholinergic MNs (left panel, **Figure 5G**), and act through the same distal enhancer utilized by UNC-3 to maintain *cfi-1* (right panel, **Figure 5G**).

If Hox proteins were only required for initiation, then endogenous *cfi-1* expression should not be affected at later stages (L2, L4, D1 adults) in *lin-39; mab-5* double

mutants. However, this is not the case (**Figure 5B-C**), suggesting that Hox proteins play an additional role at later stages. One possibility is that these 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 [29, 53]. However, expression of *cfi-1* was unaffected in the adult (**Figure S3B**). 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. 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 (**Figure 5C**). 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 5E**).

Together, our findings suggest 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), which is utilized by UNC-3 to also maintain *cfi-1* (**Figure 5G**).

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

Expression of *cfi-1* in cholinergic (DA, DB, VA, VB) MNs is maintained throughout life by Hox and UNC-3 (**Figure 5G**). 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 transcriptional repressor [42]. 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 S4A**). Hence, early removal of *cfi-1* leads to DA and DB neurons adopting a mixed identity. 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 4A**), which also serves as a conditional allele as it carries the auxin-inducible degron (AID) [53]. 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 S4B-C**). At the L3 stage (2 days upon continuous auxin treatment), we observed ectopic expression of *glr-4* in DA and DB neurons (**Figure S4C**). These results suggest that CFI-1 is required post-embryonically to prevent DA and DB neurons from adopting mixed identity, and underscore the critical role of UNC-3 and Hox in maintaining *cfi-1* expression (**Figure 5G**).

Disruption of temporal modularity in UNC-3 function leads to locomotion defects

Animals carrying *unc-3* null alleles display severe locomotion defects [29], likely due to combined defects in the expression of UNC-3 targets from both "initiation and maintenance" and "maintenance-only" modules (**Figure 3C**). Genes from the "initiation and maintenance" module include (among others) terminal identity genes coding for ACh biosynthesis components (**Figure 3A**). 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 the expression of UNC-3 targets from the "maintenance-only" module is critical for locomotion. To test this, we focused on c*fi-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 4E-F**). As controls, we used animals carrying: (**a**) the endogenous *cfi-1* reporter allele (*mNG::AID::cfi-1),* (**b**) a putative null *cfi-1* allele (*ot786*) [42], in which *cfi-1* activity is affected in MNs and other neuron types of the motor circuit [32, 45], and (**c**) a deletion of the distal enhancer (*mNG::AID::cfi-1*$^{\Delta enhancer(769\ bp)}$), in which both initiation and maintenance of *cfi-1* are abrogated in nerve cord MNs (**Figure 6A**). We performed a high-resolution behavioral analysis of freely moving adult (day 2) animals of the above genotypes using automated multi-worm tracking technology [54, 55]. We found several features related to *C. elegans* locomotion (*e.g.,* body curvature, velocity) severely affected in *cfi-1 (ot786)* putative null animals (**Figure 6B-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 6B-G**). As expected, these defects were also present in animals carrying the *mNG::AID::cfi-1* $^{\Delta enhancer\ (769\ bp)}$ 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) and observed locomotion defects. This analysis suggests that minimal disruption of temporal modularity in UNC-3 function can affect animal behavior.

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 3B**). 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 (**Figure 3-5**). 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 S6A**) [42]. 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 S6C-D**), 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 S6C-D**). Similar to *cfi-1*, these findings strongly suggest that endogenous *bnc-1* expression depends on Hox and UNC-3.

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 question, we focused on UNC-30/PITX, the terminal selector of GABAergic MN (DD, VD) identity in the *C. elegans* nerve cord (**Figure 7A**) [31]. 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])[30], 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]) (**Figure 7B**), suggesting a requirement for initiation and maintenance. As mentioned earlier (**Figure 4C**), *cfi-1* is also expressed in GABAergic MNs (**Figure 7A**). 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 embryonic stage (3-fold), but progressively stronger effects at larval (L2, L4) and adult (day 1) stages (**Figure 7C**), 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 3C**), the function of UNC-30 in GABAergic MNs is organized into two temporal modules (module #1: initiation and maintenance; module #2: maintenance-only) (**Figure 7D-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 [19]. UNC-30 binds to the *cis*-regulatory region of terminal identity genes (*unc-25*/GAD, *unc-47*/VAGT) (**Figure 7B**), confirming previous observations [30]. UNC-30 also binds to the same distal enhancer of *cfi-1* in GABAergic MNs, as UNC-3 does in cholinergic MNs (**Figure 7C**). However, the UNC-30 binding sites are distinct from the UNC-3 sites in this enhancer (**Figure 7C**). CRISPR-mediated deletion of this enhancer abolished *cfi-1* expression in both cholinergic and GABAergic MNs (**Figure 4A-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” [56], in which the same *cis*-regulatory element is used to control gene expression in different neuron types.

DISCUSSION

Terminal selectors are continuously expressed TFs that determine the identity and function of individual neuron types [2, 9]. 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 this knowledge gap may help us understand how cellular identity is established during development and maintained throughout life, a fundamental question in the field of developmental biology. In this study, we use ChIP-Seq to identify the direct targets of UNC-3/Ebf, the terminal selector of cholinergic MNs in the *C. elegans* nerve cord. Through a longitudinal analysis of *unc-3* mutants at various life stages (embryo, larva, adult), we found two groups of target genes with distinct temporal requirements for UNC-3 in cholinergic MNs. One group requires UNC-3 for both embryonic initiation and post-embryonic maintenance of expression, whereas a second group requires UNC-3 only for post-embryonic maintenance. Hence, the suite of UNC-3 targets in cholinergic MNs is modified across different life stages, a phenomenon we term "temporal modularity in UNC-3 function". Minimal disruption of this modularity by selectively removing the ability of UNC-3 to maintain expression of a single target (*cfi-1/Arid3a* from the second group) led to locomotion defects, highlighting the necessity of temporal modularity for animal behavior.

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 4C**) [32]. However, most of UNC-3 targets from either module ("initiation and maintenance" or "maintenance-only") are expressed is some, but not all, of these subtypes. By controlling both initiation and maintenance of terminal identity genes (e.g., *acr-2/AchR, glr-4/GluR*) and TFs (*zfh-2/Zfhx3, ztf-26, nhr-1*) (module #1, **Figure 3C**) [[23] and this study], UNC-3 consolidates the unique identity of each MN subtype. However, target genes from module #2 ("maintenance-only") escape initiation by UNC-3, but do require UNC-3 for maintenance (**Figure 3C**). Interestingly, all UNC-3 targets in module #2 code for TFs (*cfi-1/Arid3a, bnc-1/BNC1-2,*

mab-9/Tbx20, ceh-44/Cux1, nhr-40), specifically expressed in some of the six MN subtypes (SAB, DA, DB, VA, VB, AS). Three of these factors (*cfi-1/Arid3a, bnc-1/Bnc1, mab-9/Tbx20)* are thought to act as transcriptional repressors that prevent the adoption of mixed MN identity [42]. For example, loss of *cfi-1* during early development results in "mixed" identity of DA and DB subtypes, as these neurons, in addition to their normal molecular signature, also acquire expression of genes (e.g., *glr-4*/GluR) normally expressed in other MNs (**Figure S4A**). 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 *cfi-1* expression during larval and adult stages. In summary, the temporal modularity in UNC-3 function can be envisioned as a "double safety mechanism" for the generation and maintenance of MN diversity. That is, UNC-3 initiates and maintains the expression of MN subtype-specific genes (module #1), but also selectively maintains the expression of TFs critical for MN subtype identity (module #2) (**Figure 3C**).

The observed temporal modularity in UNC-3 function 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 expressed in multiple subtypes of a given neuronal class [13]. In mice, for example, the terminal selectors Nurr1 and Pet1 are expressed in several subtypes of dopaminergic and serotonergic neurons, respectively [5, 34]. Future studies on Nurr1, Pet1 and other terminal selectors could determine whether temporal modularity in the function of these continuously expressed TFs is a broadly applicable strategy for establishing and maintaining neuronal subtype identity.

New mechanistic insights into how terminal selectors control neuronal identity

Genes classified into module #1, such as terminal identity genes, require UNC-3 for initiation and maintenance, whereas genes from module #2 selectively require UNC-3 for maintenance (**Figure 3C**). These temporally distinct requirements for UNC-3 suggest that different mechanisms control gene expression in each module, significantly advancing our understanding of how terminal selectors establish and maintain neuronal identity.

Although ChIP-Seq for UNC-3 was performed at a single life stage (L2), our longitudinal genetic analysis together with a previous study that selectively removed UNC-3 activity in adult MNs [42] support a model where UNC-3 is constantly binding on the *cis*-regulatory regions of module #1 target genes to initiate and maintain their expression. On the other hand, the analysis of module #2 target genes (e.g., *cfi-1/Arid3a)* provides strong evidence of distinct initiation and maintenance strategies (**Figure 5G**). We find that 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 Hox and UNC-3 (**Figure 5G**). Mechanistically, we propose that Hox-dependent initiation and Hox/UNC-3-dependent maintenance of *cfi-1* are "funneled" through the same distal enhancer. This finding suggests initiation and maintenance of expression of a particular gene, in a specific cell type, can be achieved by distinct TFs acting upon the same *cis*-regulatory region (enhancer) (**Figure 5G**). Such notion may be relevant for the field of gene regulation because to date a similar mechanism has not been reported. Instead, previous studies found that distinct and physically separated *cis*-regulatory regions are responsible for either initiation or maintenance of expression of a particular gene in the same cell type [58-62].

Hox proteins are continuously expressed in multiple neuron types of the nerve cord [29]. How do they achieve the required specificity to initiate and maintain *cfi-1* only in subsets of cholinergic MNs (DA, DB, VA, VB; **Figure 4C**)? Specificity likely arises through collaboration with distinct TFs responsible for either initiation or maintenance of *cfi-1* (**Figure 5G**). This model is supported by our analysis: (a) UNC-3, LIN-39 and MAB-5 are co-expressed in DA, DB, VA, and VB neurons [29], and (b) UNC-3 is selectively required for *cfi-1* maintenance in these neurons (**Figure 5G**). Since onset of *cfi-1* is *unc-3* independent (**Figure 5C**), 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 will likely act through the same enhancer because its deletion completely abolished *cfi-1* expression in cholinergic MNs at early (and late) stages (**Figure 4D, 5G**). We propose that Hox proteins collaborate with distinct TFs over time, that is, initiation-specific factors and the maintenance-specific factor UNC-3 to ensure continuous and specific expression of *cfi-1/Arid3a*, *bnc-1/Bnc1* (**Figure S6**) and other targets from module #2 in distinct MN subtypes (**Figure 5G**). Future identification of the aforementioned initiation-specific

factors may provide key insights into the fundamental problem of how neuronal subtype identity is generated in the nervous system.

Limitations of this study

In the present paper, we conducted a longitudinal analysis for 14 UNC-3 target genes, which led to the identification of two temporal modules. It is likely though that UNC-3 temporally controls its targets though additional modules. For example, animals lacking *unc-3* display severe axon guidance defects in cholinergic MNs [33], 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.

In addition, the breadth of biological processes controlled by terminal selectors remains largely unknown. Through ChIP-Seq, we identified in an unbiased manner ~3,500 protein coding genes as putative direct targets of UNC-3, potentially implicating UNC-3 in a range of biological processes. First, our findings corroborate and significantly extend previous reports on the role of UNC-3 in neuronal terminal identity [23, 63] by identifying hundreds of terminal identity genes (42.18% of total ChIP-Seq hits) as putative UNC-3 targets. Second, we propose 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 UNC-3 control.

Temporal modularity may be a shared feature among continuously expressed, cell type-specific TFs

Our *C. elegans* study revealed that the suite of target genes of two terminal selectors (UNC-3/Ebf, UNC-30/Pitx) can be modified over time, providing evidence for temporal modularity in the function of two continuously expressed TFs that control neuronal identity. Temporal modularity may also be a feature shared by neuron type-specific TFs in vertebrates. A recent study in the mouse brainstem reported that Pet-1, the terminal selector of serotonergic neurons, activates expression of serotonin pathway genes

during development, but switches transcriptional targets at later stages [18]. Given that terminal selectors have been described in several species, it will be interesting to determine whether they control aspects of neuronal identity by modifying the suite of their target genes over time. Beyond the nervous system, cell type-specific TFs with continuous expression have been described in multiple tissues in flies and mice [64, 65]. It is possible that such TFs control cellular identity over time by utilizing the principle of temporal modularity. Our findings may provide a conceptual framework for future studies on the establishment and maintenance of cell type identity.

MATERIALS AND METHODS

C. elegans strain culture

Worms were grown at 20˚C or 25˚C on nematode growth media (NGM) plates supplied with *E. coli* OP50 as food source [28]. A list of all *C. elegans* strains used is provided in **Supplemental Table 5**.

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 [66]. 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).

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 ul 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.

ChIP-sequencing data analysis

Unique reads were mapped to the *C. elegans* genome (ce10) with bowtie2 [67]. 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 [68], and peak signals were shown in IGV. Heatmap of peak coverage in regard to UNC-3 enrichment center was generated with NGSplot [69]. The average profile of peaks binding to TSS region was generated with ChIPseeker [70]. To study the distribution of peaks genome-wide, the peaks were annotated using annotatePeaks.pl provided by Homer [71], 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.

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.

Targeted genome editing

The *cfi-1* endogenous reporter strain *kas16 [mNG::AID::cfi-1]* was generated by employing CRISPR/Cas9 genome editing and 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.

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 [72].

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.

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.

Temporally controlled protein degradation

Temporally controlled protein degradation was achieved with the auxin-inducible degradation system [53]. 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 LIN-39 and CFI-1 proteins, the following alleles were used: *lin-39 (kas9 [lin-39::mNG::AID]), and cfi-1 (kas16 [mNG::AID::cfi-1]).* L1 or L4 worms were grown at 20 ˚C 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.

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°C for at least two days before use. Imaging plates are seeded with 50µl of a 1:10 dilution of OP50 in M9 the day before tracking and left to dry overnight with the lid on at room temperature.

Behavioral feature extraction and analysis

All videos were analyzed using Tierpsy Tracker to extract each worm's position and posture over time [73]. These postural data were then converted into a set of behavioral features selected from a large set of features as previously described [54]. 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.

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.graph- pad.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 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., Data curation, Investigation, Visualization, Methodology, Writing—original draft, review and editing; A.O., E. C., Formal analysis, Validation, Investigation; M.O., P. D., O. G., J. A., P.I., Investigation; A.E.X.B., Project administration, Writing—editing; P. K., Conceptualization, Supervision, Investigation, Funding acquisition, Project administration, Writing— original draft, review and editing.

DECLARATION OF INTERESTS

The authors declare no competing interests.

FIGURE LEGENDS

Figure 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.

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.

Figure 2. 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.

Figure 3. 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 ⩾ 12). UNC-3 is required for both initiation and maintenance of all 4 terminal identity genes. *** $p < 0.001$.

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 ⩾ 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 (**Figure S2**), 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.

Figure 4. UNC-3 acts through a distal enhancer to maintain *cfi-1* expression in cholinergic motor neurons.

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 $\geqslant$ 12). N.S.: not significant, *** $p < 0.001$.

Figure 5. UNC-3 and Hox control *cfi-1* expression in cholinergic MNs.

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 ($N \geqslant 12$). 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 ($N \geqslant 13$). * $p < 0.05$.

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 ($N \geqslant 12$). 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.

Figure 6. Disruption of temporal modularity in UNC-3 function leads to locomotion defects

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.

Figure 7. 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.

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 ($N \geqslant 12$). 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, *** $p < 0.001$.

LEGENDS OF SUPPLEMENTARY FIGURES

Figure S1. 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,* and *twk-40*).

Figure S2. UNC-3 directly controls the expression of several TF reporters in MNs.

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)**. $N \geqslant 15$. 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.

Figure S3. 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: Knock-down of *lin-39::mNG::AID* starting at the L4 stage (onset of auxin treatment at L4) did not affect *cfi-1*$^{2.5kb}$*::RFP* expression in adult animals 4 days later (N ⩾ 13). N.S.: not significant. The number of MNs expressing *cfi-1*$^{2.5kb}$*::RFP* was quantified.

Figure S4. CFI-1 is required post-embryonically to maintain DA and DB neuronal identities.
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 ⩾ 20. N.S.: not significant, ** $p < 0.01$.

Figure S5: 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-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-1*$^{2.5kb}$*::RFP*) is not altered upon genetic removal of *cfi-1*. N.S.: not significant.

Figure S6: 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 ($N \geqslant 12$). *** $p < 0.001$.

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.

LEGENDS OF SUPPLEMENTARY TABLES

Supplemental Table 1: Table summarizing UNC-3 ChIP-Seq signal distribution at the *cis*-regulatory regions of previously identified UNC-3 targets.

Supplemental Table 2: Table summarizing the results of protein class ontology analysis over novel target genes of UNC-3. In total, 1,425 genes are classified into 25 protein classes.

Supplemental Table 3: UNC-3 binds to the *cis*-regulatory region of numerous genes expressed in cholinergic MNs.

Supplemental Table 4: Table summarizing novel transcription factor (TF) targets of UNC-3 identified by ChIP-Seq. Reporter alleles of each TF were built and examined for expression pattern. TFs that are expressed in MNs were further tested for UNC-3 dependency. Ten of 16 tested TFs (62.5%) show expression in MNs, of which 9 require UNC-3 activity for normal expression. Two reporters were generated for *nhr-1* and *nhr-40* because two distinct UNC-3 ChIP-Seq peaks were found in the *cis*-regulatory region of these genes. N/A: Not applicable.

Supplemental Table 5: A list of all *C. elegans* strains described in this study.

REFERENCES

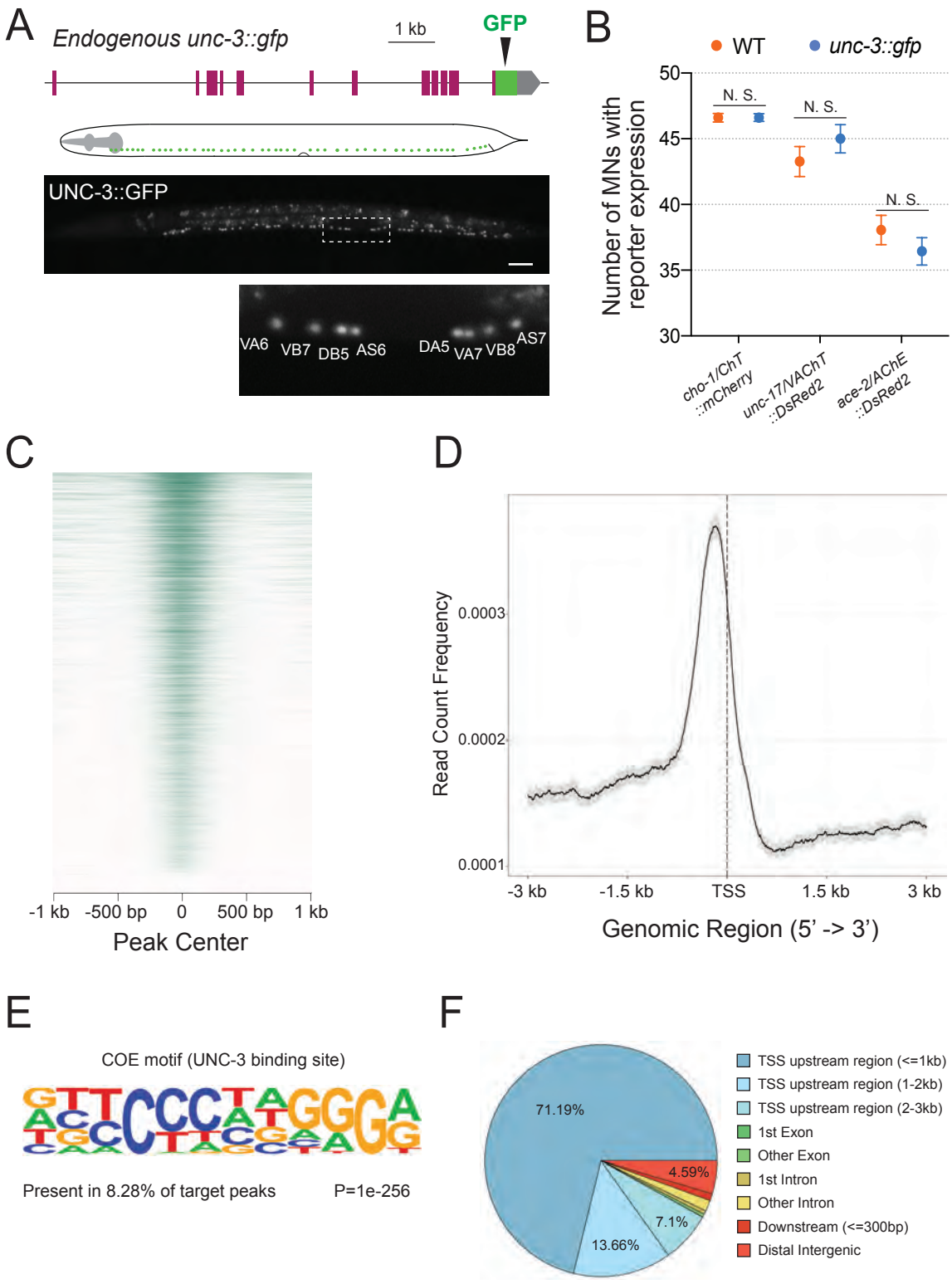
1. Davidson, E.H., McClay, D.R., and Hood, L. (2003). Regulatory gene networks and the properties of the developmental process. Proc Natl Acad Sci U S A *100*, 1475-1480.
2. Hobert, O. (2016). Terminal Selectors of Neuronal Identity. Curr Top Dev Biol *116*, 455-475.
3. Levine, M., and Tjian, R. (2003). Transcription regulation and animal diversity. Nature *424*, 147-151.
4. Deneris, E.S., and Hobert, O. (2014). Maintenance of postmitotic neuronal cell identity. Nat Neurosci *17*, 899-907.
5. Okaty, B.W., Freret, M.E., Rood, B.D., Brust, R.D., Hennessy, M.L., deBairos, D., Kim, J.C., Cook, M.N., and Dymecki, S.M. (2015). Multi-Scale Molecular Deconstruction of the Serotonin Neuron System. Neuron *88*, 774-791.
6. Sathyamurthy, A., Johnson, K.R., Matson, K.J.E., Dobrott, C.I., Li, L., Ryba, A.R., Bergman, T.B., Kelly, M.C., Kelley, M.W., and Levine, A.J. (2018). Massively Parallel Single Nucleus Transcriptional Profiling Defines Spinal Cord Neurons and Their Activity during Behavior. Cell Rep *22*, 2216-2225.
7. Saunders, A., Macosko, E.Z., Wysoker, A., Goldman, M., Krienen, F.M., de Rivera, H., Bien, E., Baum, M., Bortolin, L., Wang, S., et al. (2018). Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain. Cell *174*, 1015-1030 e1016.
8. Zeisel, A., Hochgerner, H., Lonnerberg, P., Johnsson, A., Memic, F., van der Zwan, J., Haring, M., Braun, E., Borm, L.E., La Manno, G., et al. (2018). Molecular Architecture of the Mouse Nervous System. Cell *174*, 999-1014 e1022.
9. Hobert, O. (2008). Regulatory logic of neuronal diversity: terminal selector genes and selector motifs. Proc Natl Acad Sci U S A *105*, 20067-20071.
10. Hobert, O. (2011). Regulation of terminal differentiation programs in the nervous system. Annu Rev Cell Dev Biol *27*, 681-696.
11. Garcia-Bellido, A. (1975). Genetic control of wing disc development in Drosophila. Ciba Found Symp *0*, 161-182.
12. Hobert, O. (2016). A map of terminal regulators of neuronal identity in Caenorhabditis elegans. Wiley Interdiscip Rev Dev Biol *5*, 474-498.
13. Hobert, O., and Kratsios, P. (2019). Neuronal identity control by terminal selectors in worms, flies, and chordates. Curr Opin Neurobiol *56*, 97-105.
14. 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 e613.
15. Corbo, J.C., Lawrence, K.A., Karlstetter, M., Myers, C.A., Abdelaziz, M., Dirkes, W., Weigelt, K., Seifert, M., Benes, V., Fritsche, L.G., et al. (2010). CRX ChIP-seq reveals the cis-regulatory architecture of mouse photoreceptors. Genome Res *20*, 1512-1525.
16. Housset, M., Samuel, A., Ettaiche, M., Bemelmans, A., Beby, F., Billon, N., and Lamonerie, T. (2013). Loss of Otx2 in the adult retina disrupts retinal pigment epithelium function, causing photoreceptor degeneration. J Neurosci *33*, 9890-9904.
17. Kadkhodaei, B., Alvarsson, A., Schintu, N., Ramskold, D., Volakakis, N., Joodmardi, E., Yoshitake, T., Kehr, J., Decressac, M., Bjorklund, A., et al. (2013). Transcription factor Nurr1 maintains fiber integrity and nuclear-encoded mitochondrial gene expression in dopamine neurons. Proc Natl Acad Sci U S A *110*, 2360-2365.

18. Wyler, S.C., Spencer, W.C., Green, N.H., Rood, B.D., Crawford, L., Craige, C., Gresch, P., McMahon, D.G., Beck, S.G., and Deneris, E. (2016). Pet-1 Switches Transcriptional Targets Postnatally to Regulate Maturation of Serotonin Neuron Excitability. J Neurosci *36*, 1758-1774.
19. Yu, B., Wang, X., Wei, S., Fu, T., Dzakah, E.E., Waqas, A., Walthall, W.W., and Shan, G. (2017). Convergent Transcriptional Programs Regulate cAMP Levels in C. elegans GABAergic Motor Neurons. Dev Cell *43*, 212-226 e217.
20. Flames, N., and Hobert, O. (2009). Gene regulatory logic of dopamine neuron differentiation. Nature *458*, 885-889.
21. Lopes, R., Verhey van Wijk, N., Neves, G., and Pachnis, V. (2012). Transcription factor LIM homeobox 7 (Lhx7) maintains subtype identity of cholinergic interneurons in the mammalian striatum. Proc Natl Acad Sci U S A *109*, 3119-3124.
22. Dubois, L., and Vincent, A. (2001). The COE--Collier/Olf1/EBF--transcription factors: structural conservation and diversity of developmental functions. Mech Dev *108*, 3-12.
23. Kratsios, P., Stolfi, A., Levine, M., and Hobert, O. (2012). Coordinated regulation of cholinergic motor neuron traits through a conserved terminal selector gene. Nat Neurosci *15*, 205-214.
24. Blackburn, P.R., Barnett, S.S., Zimmermann, M.T., Cousin, M.A., Kaiwar, C., Pinto, E.V.F., Niu, Z., Ferber, M.J., Urrutia, R.A., Selcen, D., et al. (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.
25. Chao, H.T., Davids, M., Burke, E., Pappas, J.G., Rosenfeld, J.A., McCarty, A.J., Davis, T., Wolfe, L., Toro, C., Tifft, C., et al. (2017). A Syndromic Neurodevelopmental Disorder Caused by De Novo Variants in EBF3. Am J Hum Genet *100*, 128-137.
26. Harms, F.L., Girisha, K.M., Hardigan, A.A., Kortum, F., Shukla, A., Alawi, M., Dalal, A., Brady, L., Tarnopolsky, M., Bird, L.M., et al. (2017). Mutations in EBF3 Disturb Transcriptional Profiles and Cause Intellectual Disability, Ataxia, and Facial Dysmorphism. Am J Hum Genet *100*, 117-127.
27. Sleven, H., Welsh, S.J., Yu, J., Churchill, M.E., Wright, C.F., Henderson, A., Horvath, R., Rankin, J., Vogt, J., Magee, A., et al. (2017). De Novo Mutations in EBF3 Cause a Neurodevelopmental Syndrome. Am J Hum Genet *100*, 138-150.
28. Brenner, S. (1974). The genetics of Caenorhabditis elegans. Genetics *77*, 71-94.
29. 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*.
30. Eastman, C., Horvitz, H.R., and Jin, Y. (1999). Coordinated transcriptional regulation of the unc-25 glutamic acid decarboxylase and the unc-47 GABA vesicular transporter by the Caenorhabditis elegans UNC-30 homeodomain protein. J Neurosci *19*, 6225-6234.
31. Jin, Y., Hoskins, R., and Horvitz, H.R. (1994). Control of type-D GABAergic neuron differentiation by C. elegans UNC-30 homeodomain protein. Nature *372*, 780-783.
32. 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., et al. (2015). A cellular and regulatory map of the cholinergic nervous system of C. elegans. Elife *4*.

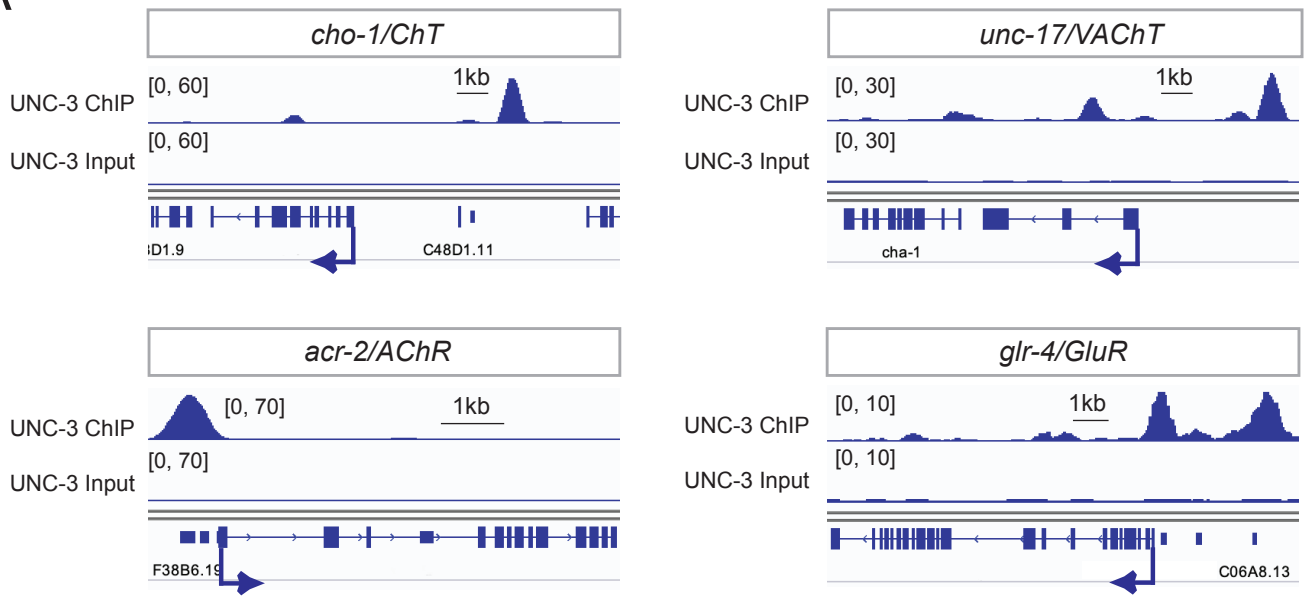
33. Prasad, B.C., Ye, B., Zackhary, R., Schrader, K., Seydoux, G., and Reed, R.R. (1998). unc-3, a gene required for axonal guidance in Caenorhabditis elegans, encodes a member of the O/E family of transcription factors. Development *125*, 1561-1568.
34. Kadkhodaei, B., Ito, T., Joodmardi, E., Mattsson, B., Rouillard, C., Carta, M., Muramatsu, S., Sumi-Ichinose, C., Nomura, T., Metzger, D., et al. (2009). Nurr1 is required for maintenance of maturing and adult midbrain dopamine neurons. J Neurosci *29*, 15923-15932.
35. Treiber, N., Treiber, T., Zocher, G., and Grosschedl, R. (2010). Structure of an Ebf1:DNA complex reveals unusual DNA recognition and structural homology with Rel proteins. Genes Dev *24*, 2270-2275.
36. Treiber, T., Mandel, E.M., Pott, S., Gyory, I., Firner, S., Liu, E.T., and Grosschedl, R. (2010). Early B cell factor 1 regulates B cell gene networks by activation, repression, and transcription- independent poising of chromatin. Immunity *32*, 714-725.
37. Wang, M.M., and Reed, R.R. (1993). Molecular cloning of the olfactory neuronal transcription factor Olf-1 by genetic selection in yeast. Nature *364*, 121-126.
38. Wang, S.S., Tsai, R.Y., and 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*, 4149-4158.
39. Kratsios, P., Pinan-Lucarre, B., Kerk, S.Y., Weinreb, A., Bessereau, J.L., and Hobert, O. (2015). Transcriptional coordination of synaptogenesis and neurotransmitter signaling. Curr Biol *25*, 1282-1295.
40. Mi, H., Muruganujan, A., Casagrande, J.T., and Thomas, P.D. (2013). Large-scale gene function analysis with the PANTHER classification system. Nat Protoc *8*, 1551-1566.
41. 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.
42. 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.
43. Pocock, R., Mione, M., Hussain, S., Maxwell, S., Pontecorvi, M., Aslam, S., Gerrelli, D., Sowden, J.C., and Woollard, A. (2008). Neuronal function of Tbx20 conserved from nematodes to vertebrates. Dev Biol *317*, 671-685.
44. Brozova, E., Simeckova, K., Kostrouch, Z., Rall, J.E., and Kostrouchova, M. (2006). NHR-40, a Caenorhabditis elegans supplementary nuclear receptor, regulates embryonic and early larval development. Mech Dev *123*, 689-701.
45. 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.
46. Wang, M.M., Tsai, R.Y., Schrader, K.A., and 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*, 5805-5813.
47. Leyva-Diaz, E., and Hobert, O. (2019). Transcription factor autoregulation is required for acquisition and maintenance of neuronal identity. Development *146*.
48. Scott, M.M., Krueger, K.C., and 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*, 2628-2636.
49. Xue, D., Finney, M., Ruvkun, G., and Chalfie, M. (1992). Regulation of the mec-3 gene by the C.elegans homeoproteins UNC-86 and MEC-3. EMBO J *11*, 4969-4979.

50. 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*.
51. Liu, J., and 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*, 5179-5190.
52. Boyle, A.P., Araya, C.L., Brdlik, C., Cayting, P., Cheng, C., Cheng, Y., Gardner, K., Hillier, L.W., Janette, J., Jiang, L., et al. (2014). Comparative analysis of regulatory information and circuits across distant species. Nature *512*, 453-456.
53. Zhang, L., Ward, J.D., Cheng, Z., and Dernburg, A.F. (2015). The auxin-inducible degradation (AID) system enables versatile conditional protein depletion in C. elegans. Development *142*, 4374-4384.
54. Javer, A., Ripoll-Sanchez, L., and Brown, A.E.X. (2018). Powerful and interpretable behavioural features for quantitative phenotyping of Caenorhabditis elegans. Philos Trans R Soc Lond B Biol Sci *373*.
55. Yemini, E., Jucikas, T., Grundy, L.J., Brown, A.E., and Schafer, W.R. (2013). A database of Caenorhabditis elegans behavioral phenotypes. Nat Methods *10*, 877-879.
56. Sabaris, G., Laiker, I., Preger-Ben Noon, E., and Frankel, N. (2019). Actors with Multiple Roles: Pleiotropic Enhancers and the Paradigm of Enhancer Modularity. Trends Genet *35*, 423-433.
57. Hoxha, E., Tonini, R., Montarolo, F., Croci, L., Consalez, G.G., and Tempia, F. (2013). Motor dysfunction and cerebellar Purkinje cell firing impairment in Ebf2 null mice. Mol Cell Neurosci *52*, 51-61.
58. Ellmeier, W., Sunshine, M.J., Maschek, R., and Littman, D.R. (2002). Combined deletion of CD8 locus cis-regulatory elements affects initiation but not maintenance of CD8 expression. Immunity *16*, 623-634.
59. Johnson, S.A., Harmon, K.J., Smiley, S.G., Still, F.M., and Kavaler, J. (2011). Discrete regulatory regions control early and late expression of D-Pax2 during external sensory organ development. Dev Dyn *240*, 1769-1778.
60. Manzanares, M., Bel-Vialar, S., Ariza-McNaughton, L., Ferretti, E., Marshall, H., Maconochie, M.M., Blasi, F., and 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*, 3595-3607.
61. Pfeffer, P.L., Payer, B., Reim, G., di Magliano, M.P., and Busslinger, M. (2002). The activation and maintenance of Pax2 expression at the mid-hindbrain boundary is controlled by separate enhancers. Development *129*, 307-318.
62. Rhee, H.S., Closser, M., Guo, Y., Bashkirova, E.V., Tan, G.C., Gifford, D.K., and Wichterle, H. (2016). Expression of Terminal Effector Genes in Mammalian Neurons Is Maintained by a Dynamic Relay of Transient Enhancers. Neuron *92*, 1252-1265.
63. Kim, K., Colosimo, M.E., Yeung, H., and Sengupta, P. (2005). The UNC-3 Olf/EBF protein represses alternate neuronal programs to specify chemosensory neuron identity. Dev Biol *286*, 136-148.
64. Pikkarainen, S., Tokola, H., Kerkela, R., and Ruskoaho, H. (2004). GATA transcription factors in the developing and adult heart. Cardiovasc Res *63*, 196-207.
65. Zhou, Q., Liu, M., Xia, X., Gong, T., Feng, J., Liu, W., Liu, Y., Zhen, B., Wang, Y., Ding, C., et al. (2017). A mouse tissue transcription factor atlas. Nat Commun *8*, 15089.
66. Hobert, O. (2002). PCR fusion-based approach to create reporter gene constructs for expression analysis in transgenic C. elegans. Biotechniques *32*, 728-730.

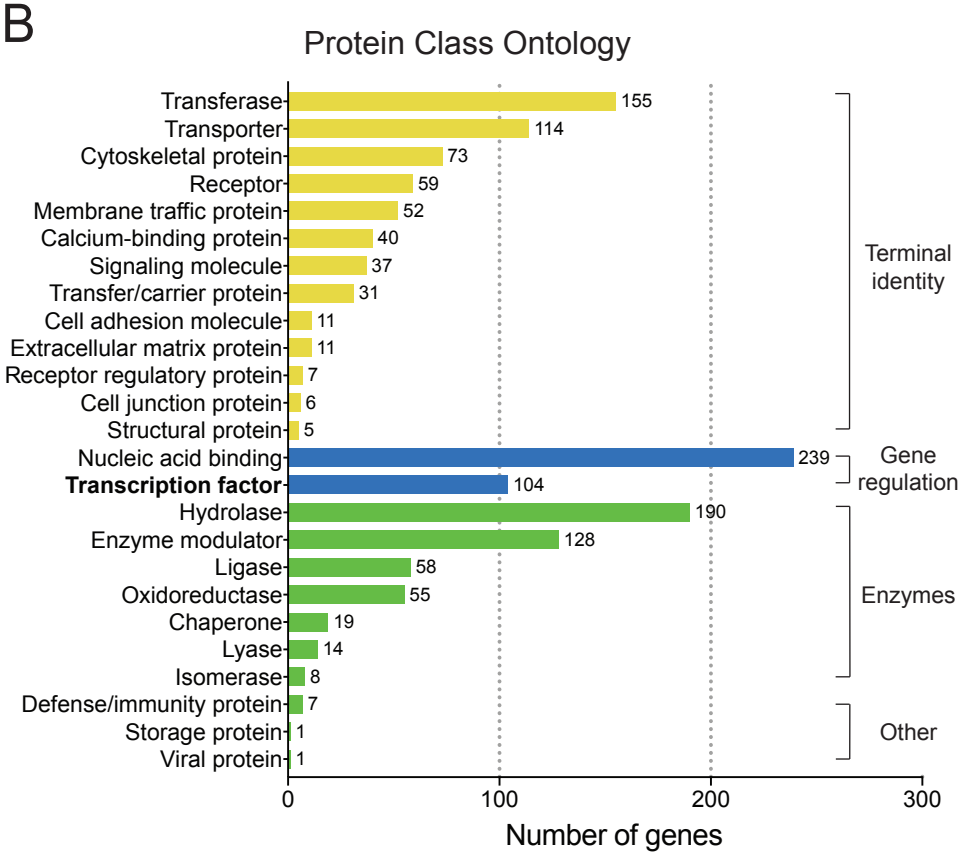
67. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359.
68. Ramirez, F., Ryan, D.P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dundar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res *44*, W160-165.
69. Shen, L., Shao, N., Liu, X., and Nestler, E. (2014). ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. BMC Genomics *15*, 284.
70. Yu, G., Wang, L.G., and He, Q.Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics *31*, 2382-2383.
71. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and 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*, 576-589.
72. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat Methods *9*, 676-682.
73. Javer, A., Currie, M., Lee, C.W., Hokanson, J., Li, K., Martineau, C.N., Yemini, E., Grundy, L.J., Li, C., Ch'ng, Q., et al. (2018). An open-source platform for analyzing and sharing worm-behavior data. Nat Methods *15*, 645-646.
74. Federici, T., and Boulis, N.M. (2006). Gene-based treatment of motor neuron diseases. Muscle Nerve *33*, 302-323.


A
Endogenous unc-3::gfp
1 kb
GFP
UNC-3::GFP
VA6
VB7
DB5
AS6
DA5
VA7
VB8
AS7
B
WT
unc-3::gfp
Number of MNs with reporter expression
50
45
40
35
30
N. S.
N. S.
N. S.
cho-1/ChT ::mCherry
unc-17/VAChT ::DsRed2
ace-2/AChE ::DsRed2
C
-1 kb
-500 bp
0
500 bp
1 kb
Peak Center
D
Read Count Frequency
0.0003
0.0002
0.0001
-3 kb
-1.5 kb
TSS
1.5 kb
3 kb
Genomic Region (5' -> 3')
E
COE motif (UNC-3 binding site)
Present in 8.28% of target peaks
P=1e-256
F
71.19%
4.59%
7.1%
13.66%
TSS upstream region (<=1kb)
TSS upstream region (1-2kb)
TSS upstream region (2-3kb)
1st Exon
Other Exon
1st Intron
Other Intron
Downstream (<=300bp)
Distal Intergenic

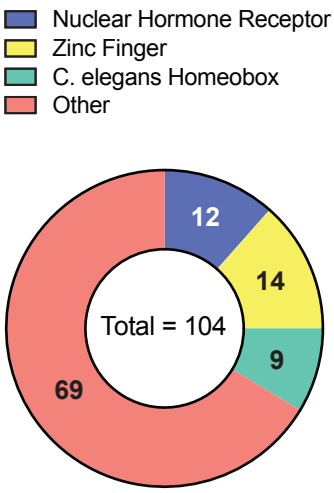
A

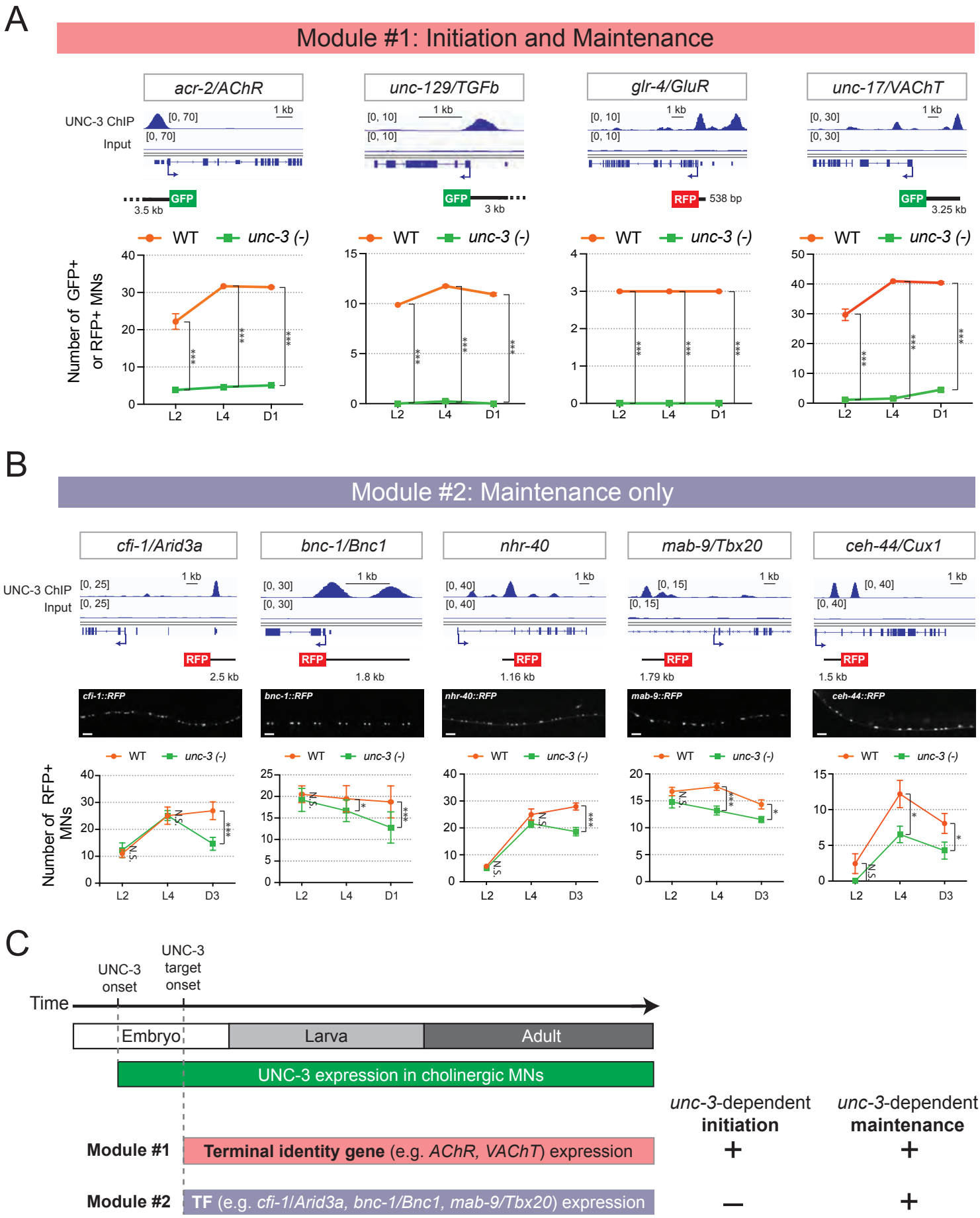

cho-1/ChT
UNC-3 ChIP
UNC-3 Input
[0, 60]
[0, 60]
1kb
C48D1.11
unc-17/VAChT
UNC-3 ChIP
UNC-3 Input
[0, 30]
[0, 30]
1kb
cha-1
acr-2/AChR
UNC-3 ChIP
UNC-3 Input
[0, 70]
[0, 70]
1kb
glr-4/GluR
UNC-3 ChIP
UNC-3 Input
[0, 10]
[0, 10]
1kb
C06A8.13

B

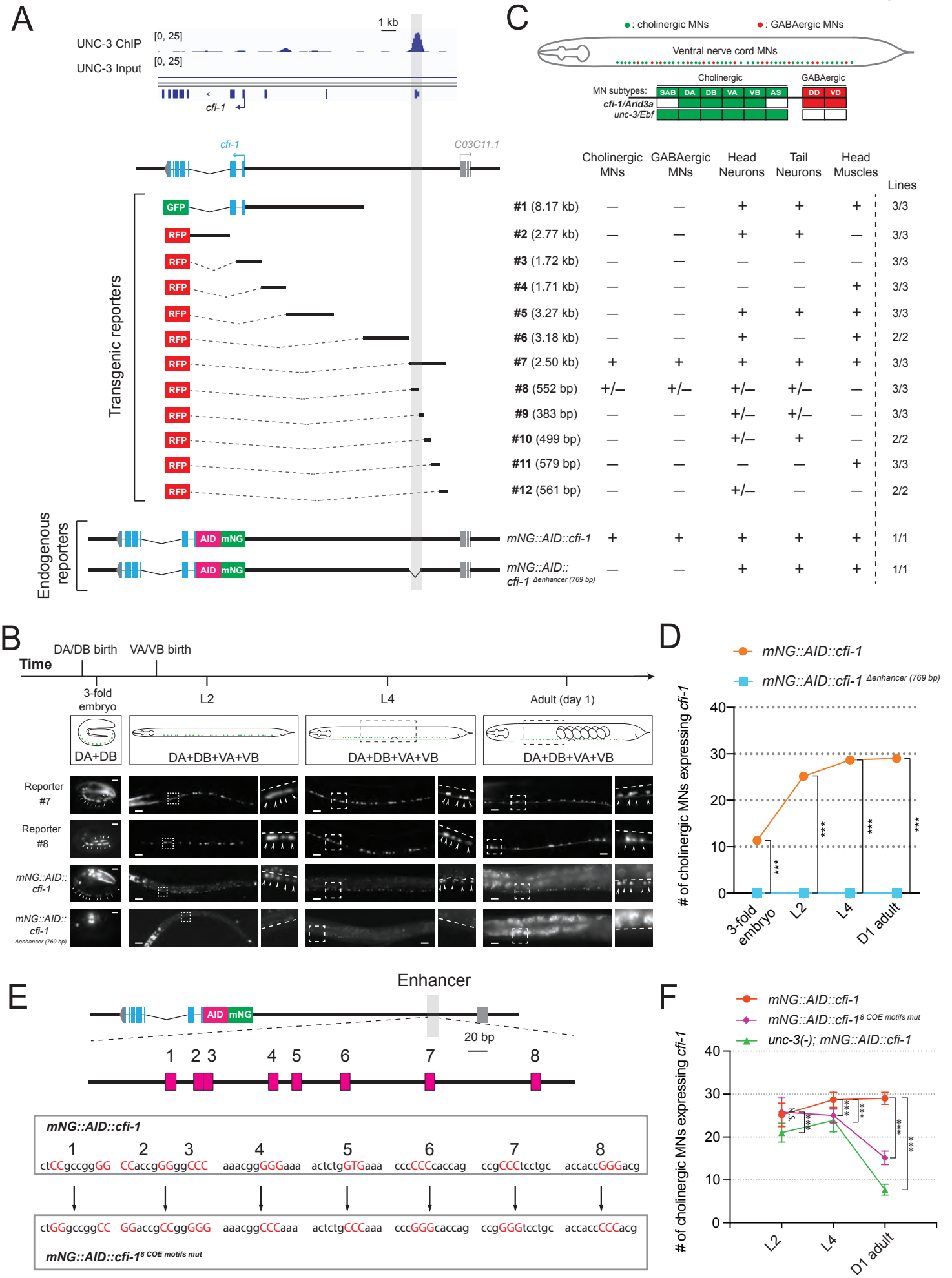

Protein Class Ontology
Transferase 155
Transporter 114
Cytoskeletal protein 73
Receptor 59
Membrane traffic protein 52
Calcium-binding protein 40
Signaling molecule 37
Transfer/carrier protein 31
Cell adhesion molecule 11
Extracellular matrix protein 11
Receptor regulatory protein 7
Cell junction protein 6
Structural protein 5
Nucleic acid binding 239
Transcription factor 104
Hydrolase 190
Enzyme modulator 128
Ligase 58
Oxidoreductase 55
Chaperone 19
Lyase 14
Isomerase 8
Defense/immunity protein 7
Storage protein 1
Viral protein 1
Terminal identity
Gene regulation
Enzymes
Other
0
100
200
300
Number of genes

C

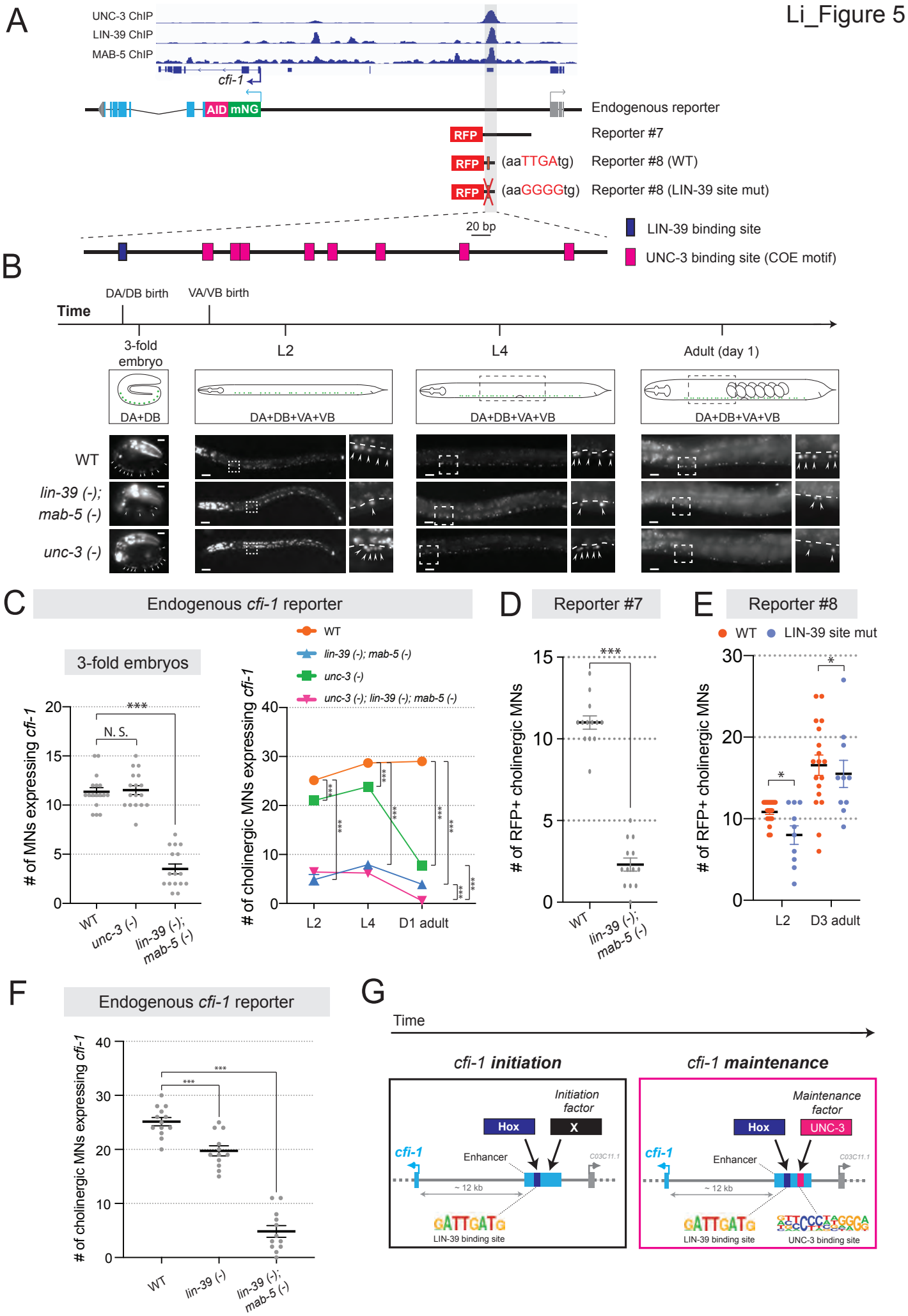

Nuclear Hormone Receptor
Zinc Finger
C. elegans Homeobox
Other
12
14
9
69
Total = 104


A
Module #1: Initiation and Maintenance
acr-2/AChR
unc-129/TGFb
glr-4/GluR
unc-17/VAChT
UNC-3 ChIP
Input
1 kb
GFP
RFP
538 bp
3.5 kb
3 kb
3.25 kb
WT
unc-3 (-)
Number of GFP+ or RFP+ MNs
L2
L4
D1
B
Module #2: Maintenance only
cfi-1/Arid3a
bnc-1/Bnc1
nhr-40
mab-9/Tbx20
ceh-44/Cux1
2.5 kb
1.8 kb
1.16 kb
1.79 kb
1.5 kb
cfi-1::RFP
bnc-1::RFP
nhr-40::RFP
mab-9::RFP
ceh-44::RFP
Number of RFP+ MNs
L2
L4
D3
C
UNC-3 onset
UNC-3 target onset
Time
Embryo
Larva
Adult
UNC-3 expression in cholinergic MNs
unc-3-dependent initiation
unc-3-dependent maintenance
Module #1
Terminal identity gene (e.g. AChR, VAChT) expression
Module #2
TF (e.g. cfi-1/Arid3a, bnc-1/Bnc1, mab-9/Tbx20) expression

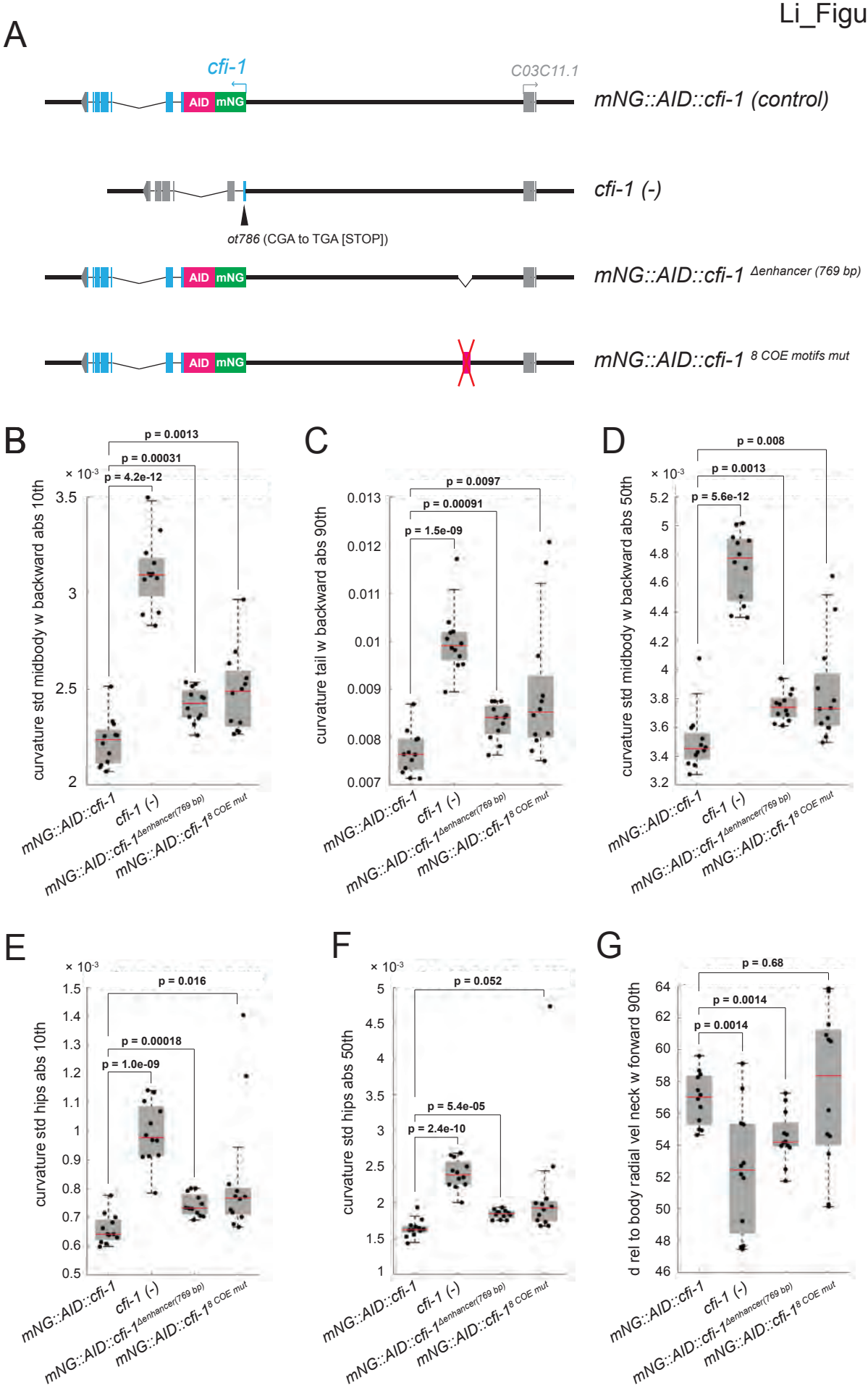
Li_Figure 4


A
1 kb
UNC-3 ChIP
[0, 25]
UNC-3 Input
[0, 25]
cfi-1
cfi-1
C03C11.1
Transgenic reporters
GFP
RFP
Endogenous reporters
AID
mNG
C
: cholinergic MNs
: GABAergic MNs
Ventral nerve cord MNs
Cholinergic
GABAergic
MN subtypes:
SAB
DA
DB
VA
VB
AS
DD
VD
cfi-1/Arid3a
unc-3/Ebf
Cholinergic MNs
GABAergic MNs
Head Neurons
Tail Neurons
Head Muscles
Lines
#1 (8.17 kb) — — + + + 3/3
#2 (2.77 kb) — — + + — 3/3
#3 (1.72 kb) — — — — — 3/3
#4 (1.71 kb) — — — — + 3/3
#5 (3.27 kb) — — + + + 3/3
#6 (3.18 kb) — — + — + 2/2
#7 (2.50 kb) + + + + + 3/3
#8 (552 bp) +/– +/– +/– +/– — 3/3
#9 (383 bp) — — +/– +/– — 3/3
#10 (499 bp) — — +/– + — 2/2
#11 (579 bp) — — — — + 3/3
#12 (561 bp) — — +/– — — 2/2
mNG::AID::cfi-1 + + + + + 1/1
mNG::AID::cfi-1 Δenhancer (769 bp) — — + + + 1/1
B
Time
DA/DB birth
VA/VB birth
3-fold embryo
L2
L4
Adult (day 1)
DA+DB
DA+DB+VA+VB
DA+DB+VA+VB
DA+DB+VA+VB
Reporter #7
Reporter #8
mNG::AID::cfi-1
mNG::AID::cfi-1 Δenhancer (769 bp)
D
mNG::AID::cfi-1
mNG::AID::cfi-1 Δenhancer (769 bp)
# of cholinergic MNs expressing cfi-1
40
30
20
10
0
***
3-fold embryo
L2
L4
D1 adult
E
Enhancer
AID
mNG
20 bp
1 2 3 4 5 6 7 8
mNG::AID::cfi-1
1 2 3 4 5 6 7 8
ctCCgccggGG CCaccgGGggCCC aaacggGGGaaa actctgGTGaaa cccCCCcaccag ccgCCCtcctgc accaccGGGacg
ctGGgccggCC GGaccgCCggGGG aaacggCCCaaa actctgCCCaaa cccGGGcaccag ccgGGGtcctgc accaccCCCacg
mNG::AID::cfi-1 8 COE motifs mut
F
mNG::AID::cfi-1
mNG::AID::cfi-1 8 COE motifs mut
unc-3(-); mNG::AID::cfi-1
# of cholinergic MNs expressing cfi-1
40
30
20
10
0
NS
***
L2
L4
D1 adult

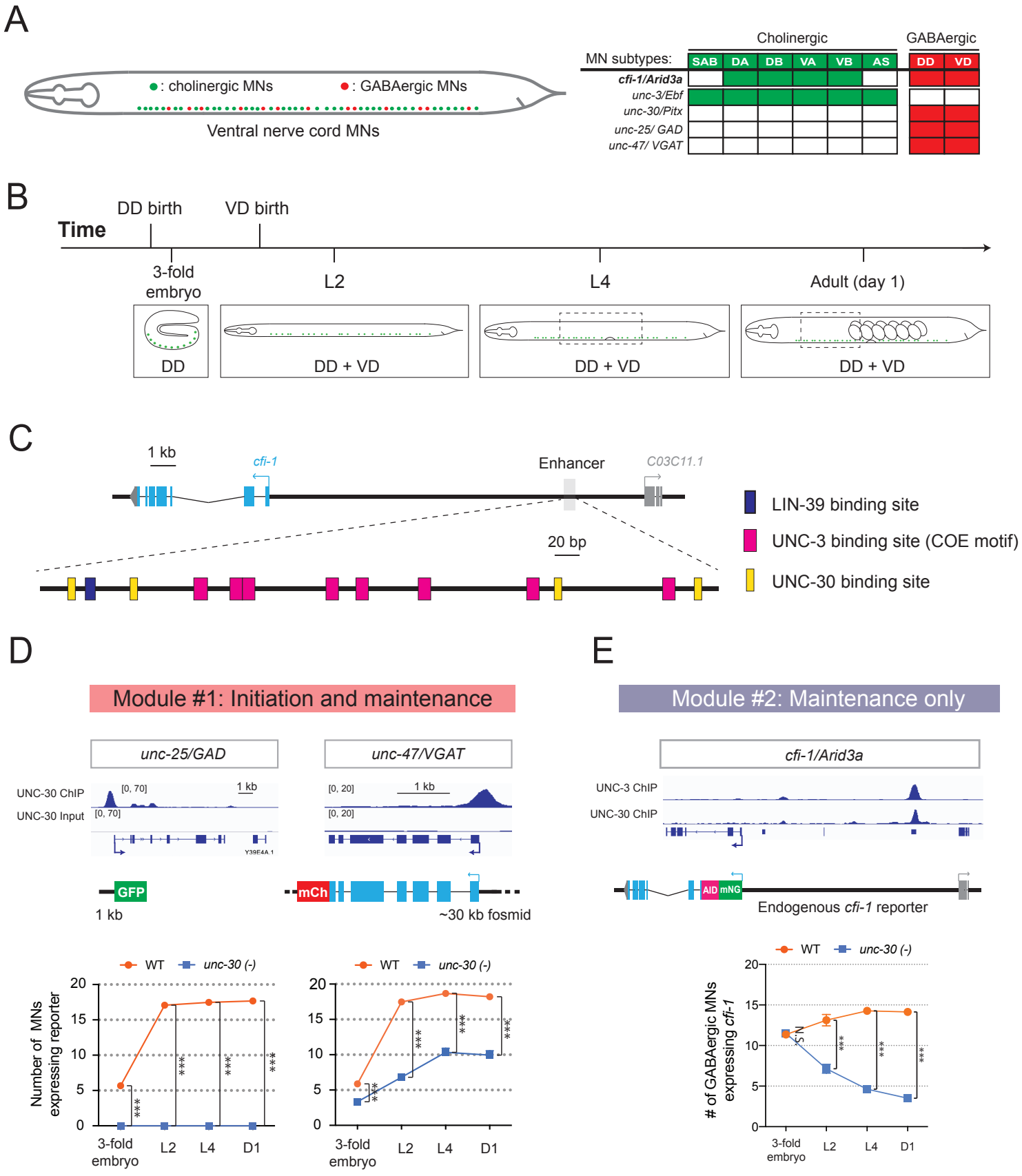
Li_Figure 5


A
UNC-3 ChIP
LIN-39 ChIP
MAB-5 ChIP
cfi-1
AID
mNG
Endogenous reporter
RFP
Reporter #7
(aaTTGAtg)
Reporter #8 (WT)
(aaGGGGtg)
Reporter #8 (LIN-39 site mut)
20 bp
LIN-39 binding site
UNC-3 binding site (COE motif)
B
Time
DA/DB birth
VA/VB birth
3-fold embryo
L2
L4
Adult (day 1)
DA+DB
DA+DB+VA+VB
WT
lin-39 (-); mab-5 (-)
unc-3 (-)
C
Endogenous cfi-1 reporter
3-fold embryos
# of MNs expressing cfi-1
N. S.
***
WT
unc-3 (-)
lin-39 (-); mab-5 (-)
# of cholinergic MNs expressing cfi-1
WT
lin-39 (-); mab-5 (-)
unc-3 (-)
unc-3 (-); lin-39 (-); mab-5 (-)
L2
L4
D1 adult
D
Reporter #7
# of RFP+ cholinergic MNs
E
Reporter #8
WT
LIN-39 site mut
D3 adult
F
Endogenous cfi-1 reporter
# of cholinergic MNs expressing cfi-1
lin-39 (-)
G
Time
cfi-1 initiation
cfi-1 maintenance
Initiation factor
Hox
X
Enhancer
C03C11.1
~ 12 kb
GATTGATG
LIN-39 binding site
Maintenance factor
UNC-3
UNC-3 binding site

Li_Figure 6

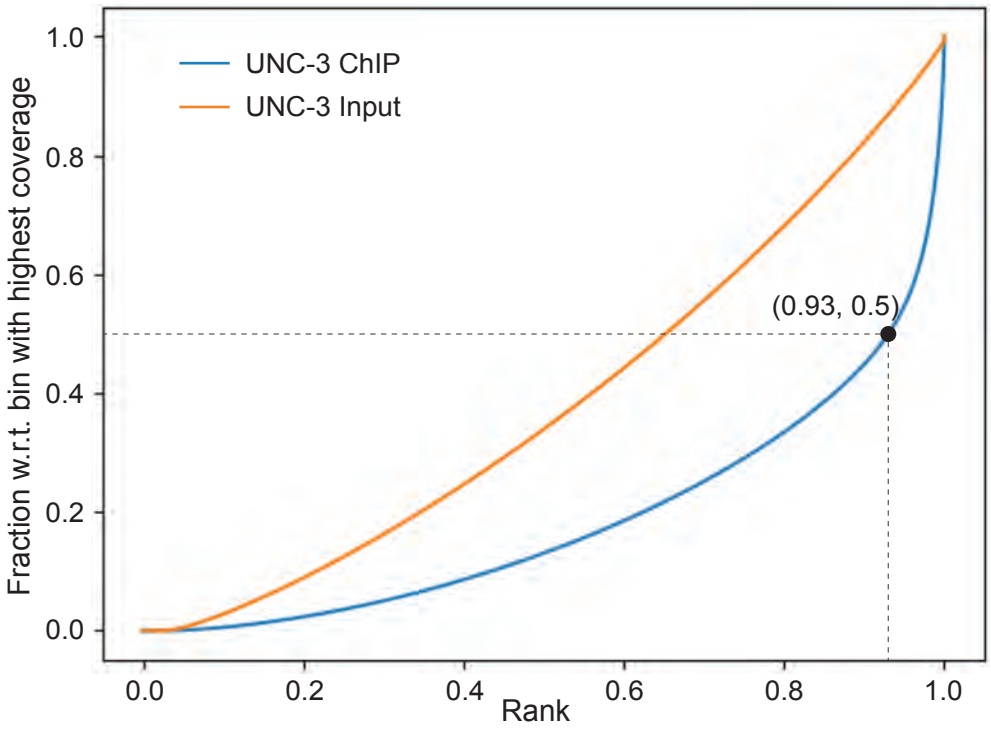

A
cfi-1
C03C11.1
AID
mNG
mNG::AID::cfi-1 (control)
cfi-1 (-)
ot786 (CGA to TGA [STOP])
mNG::AID::cfi-1 Δenhancer (769 bp)
mNG::AID::cfi-1 8 COE motifs mut
B
curvature std midbody w backward abs 10th
× 10-3
p = 0.0013
p = 0.00031
p = 4.2e-12
C
curvature tail w backward abs 90th
p = 0.0097
p = 0.00091
p = 1.5e-09
D
curvature std midbody w backward abs 50th
p = 0.008
p = 0.0013
p = 5.6e-12
E
curvature std hips abs 10th
p = 0.016
p = 0.00018
p = 1.0e-09
F
curvature std hips abs 50th
p = 0.052
p = 5.4e-05
p = 2.4e-10
G
d rel to body radial vel neck w forward 90th
p = 0.68
p = 0.0014
p = 0.0014
mNG::AID::cfi-1
cfi-1 (-)
mNG::AID::cfi-1Δenhancer(769 bp)
mNG::AID::cfi-18 COE mut

Li_Figure 7

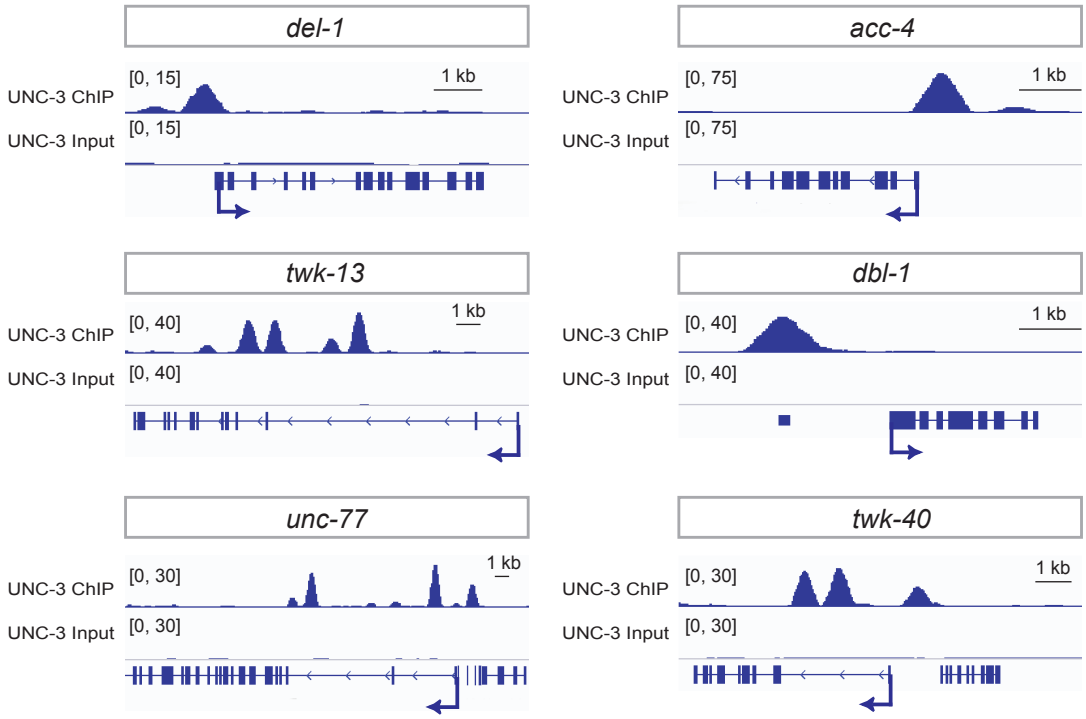

A
: cholinergic MNs
: GABAergic MNs
Ventral nerve cord MNs
Cholinergic
GABAergic
MN subtypes:
SAB
DA
DB
VA
VB
AS
DD
VD
cfi-1/Arid3a
unc-3/Ebf
unc-30/Pitx
unc-25/ GAD
unc-47/ VGAT
B
Time
DD birth
VD birth
3-fold embryo
L2
L4
Adult (day 1)
DD
DD + VD
DD + VD
DD + VD
C
1 kb
cfi-1
Enhancer
C03C11.1
20 bp
LIN-39 binding site
UNC-3 binding site (COE motif)
UNC-30 binding site
D
Module #1: Initiation and maintenance
unc-25/GAD
unc-47/VGAT
UNC-30 ChIP
UNC-30 Input
[0, 70]
[0, 70]
1 kb
[0, 20]
[0, 20]
1 kb
GFP
1 kb
mCh
~30 kb fosmid
WT
unc-30 (-)
Number of MNs expressing reporter
3-fold embryo
L2
L4
D1
E
Module #2: Maintenance only
cfi-1/Arid3a
UNC-3 ChIP
UNC-30 ChIP
AID
mNG
Endogenous cfi-1 reporter
WT
unc-30 (-)
# of GABAergic MNs expressing cfi-1
N.S.
3-fold embryo
L2
L4
D1

Supplementary Figure S1

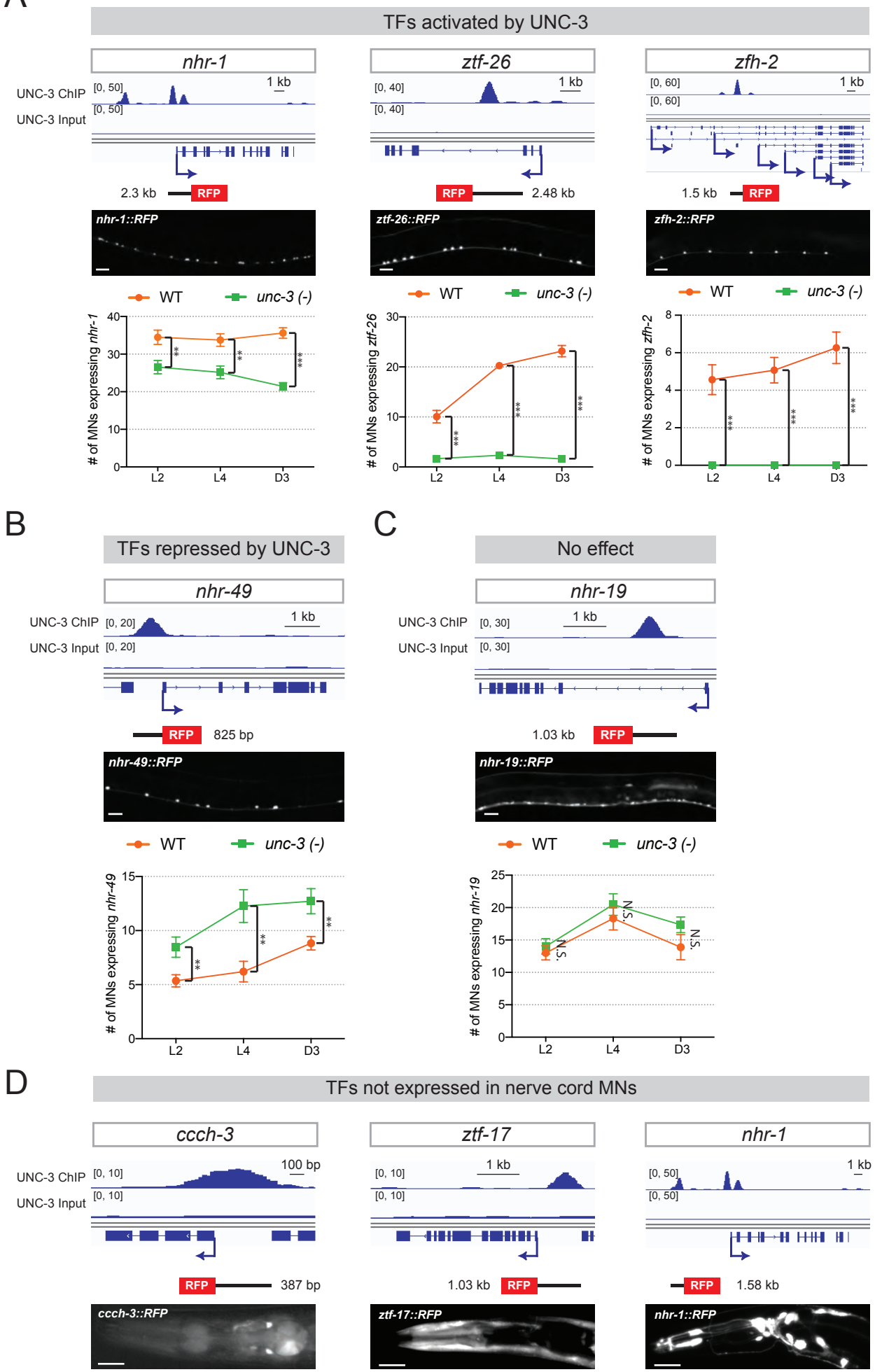
A



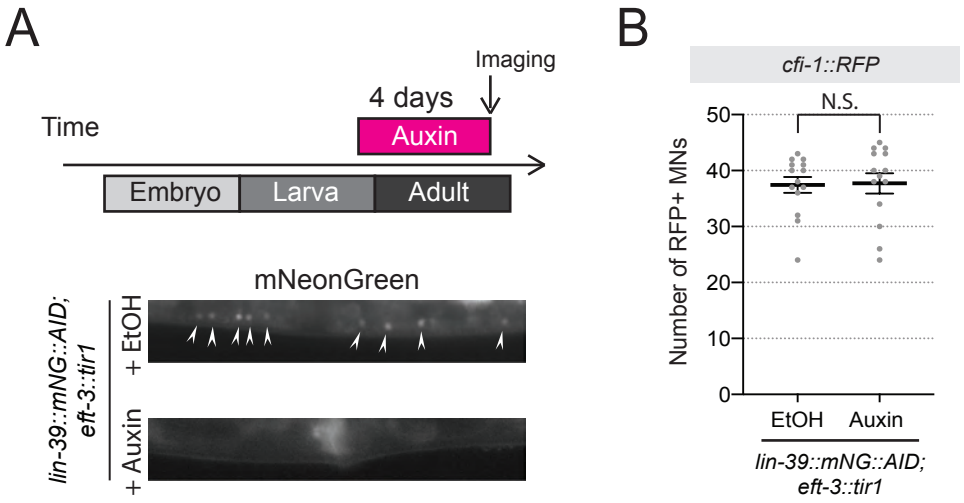
B



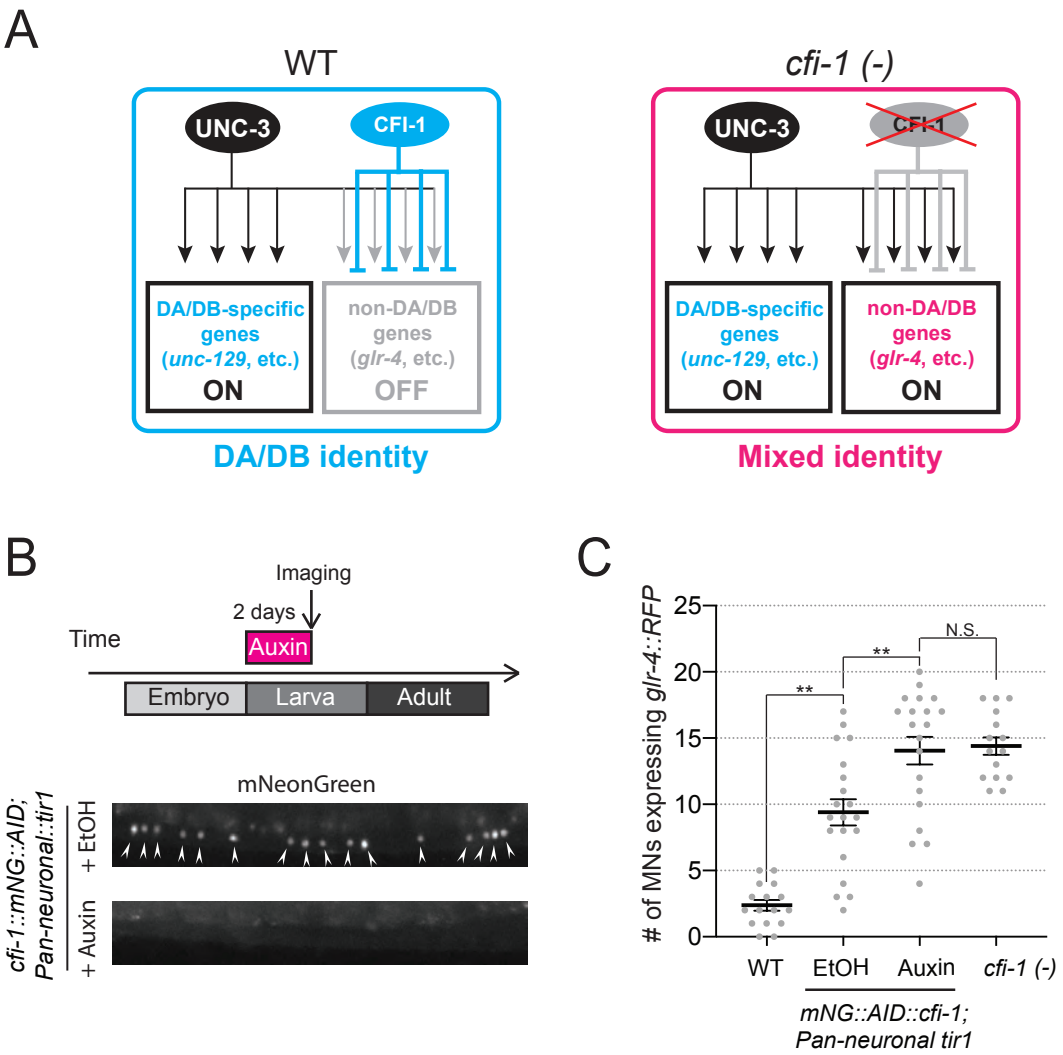
Supplementary Figure S2



Supplementary Figure S3

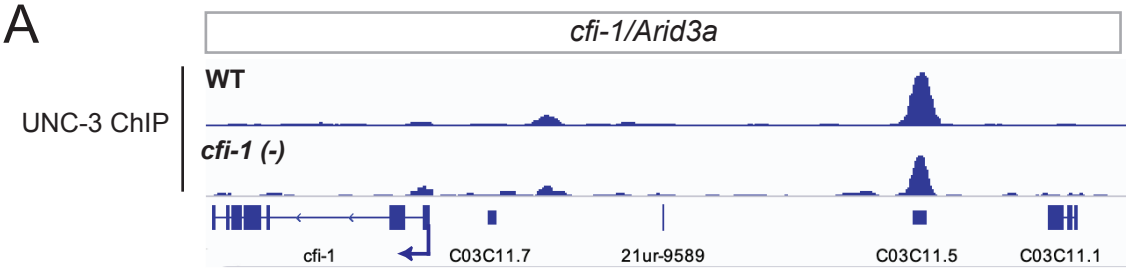


Supplementary Figure S4

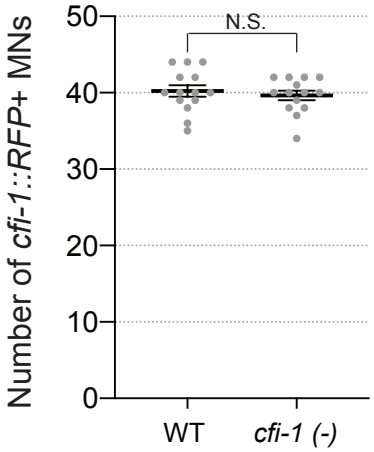


Supplementary Figure S5

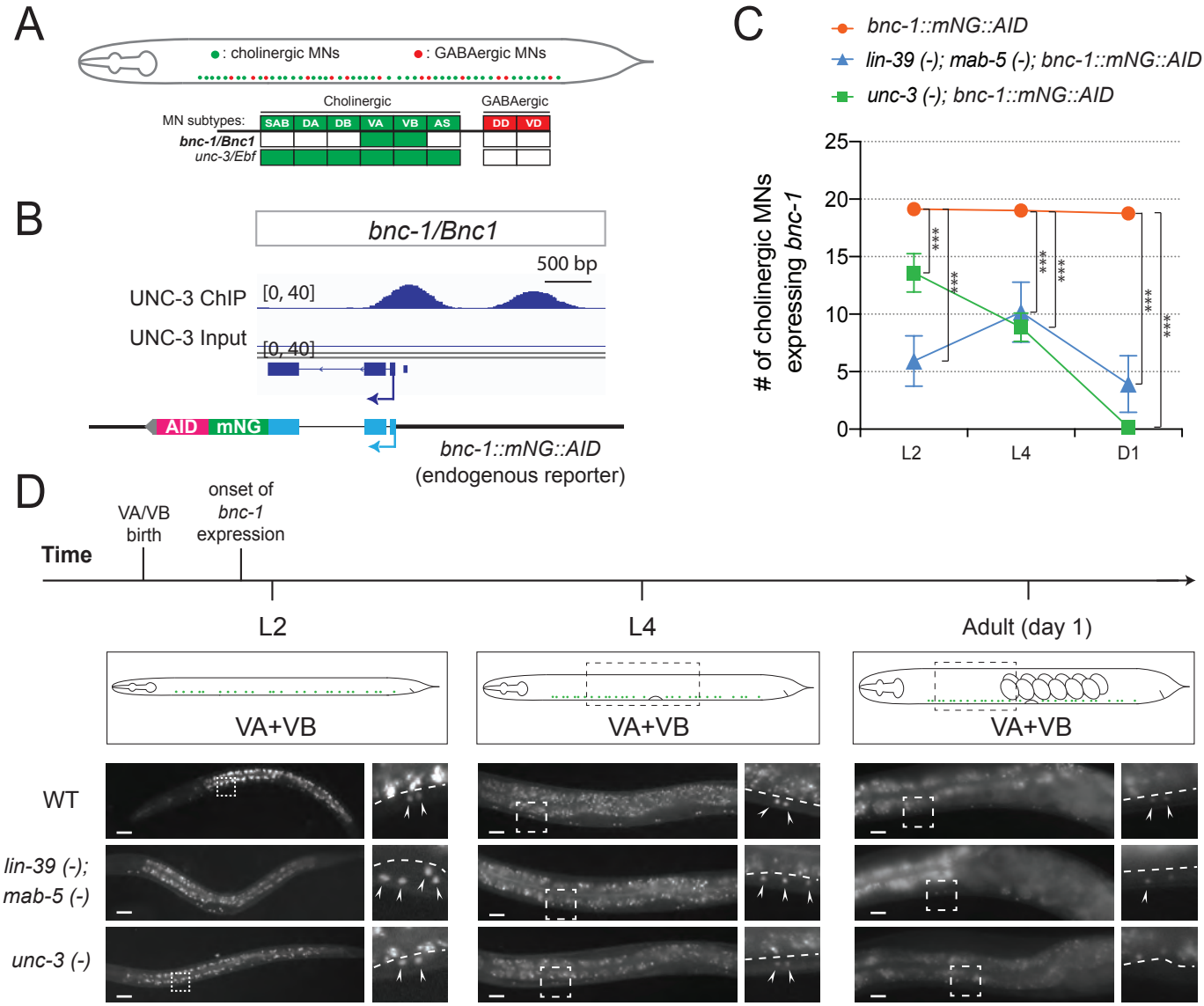
A



B



Supplementary Figure S6



Supplemental Table 1: Analysis of UNC-3 targets for direct UNC-3 binding.

	Terminal identity gene	Previously known direct target	ChIP-Seq peaks	Intron	ATG upstream (<1kb)	ATG upstream (1-2kb)	ATG upstream (2-3kb)	ATG upstream (>3kb)
1	*unc-17*	Yes	Yes	+	-	-	+	+
2	*cho-1*	Yes	Yes	+	-	-	-	+
3	*ace-2*	Yes	Yes	-	-	-	-	+
4	*acr-2*	Yes	Yes	-	+	-	-	-
5	*acr-14*	Yes	Yes	-	+	-	-	-
6	*del-1*	Yes	Yes	-	+	-	+	+
7	*glr-4*	Yes	Yes	-	+	+	-	+
8	*madd-4*	Yes	Yes	+	-	-	+	-
9	*unc-129*	Yes	Yes	-	+	-	-	-
10	*acr-16*	Yes	No	-	-	-	-	-
Summary		**9/10** previously known direct targets show UNC-3 binding based on ChIP-Seq						
11	*acr-5*	No	Yes	-	+	-	-	-
12	*acr-15*	No	Yes	-	+	-	-	-
13	*unc-63*	No	Yes	-	-	+	-	-
14	*gar-2*	No	Yes	+	-	+	-	-
15	*trp-1*	No	Yes	-	-	+	-	-
16	*unc-77*	No	Yes	+	-	+	-	-
17	*unc-8*	No	Yes	+	-	-	-	-
18	*inx-12*	No	Yes	-	-	-	-	+
19	*acc-4*	No	Yes	-	+	+	-	-
20	*gbb-1*	No	Yes	-	+	-	-	+
21	*kvs-1*	No	Yes	+	-	-	-	+
22	*dbl-1*	No	Yes	-	+	+	-	+
23	*tig-2*	No	Yes	-	+	+	-	-
24	*nlp-21*	No	Yes	+	-	+	-	-
25	*unc-40*	No	Yes	+	+	-	-	-
26	*max-1*	No	Yes	+	+	+	-	-
27	*F39B2.8*	No	Yes	+	+	-	-	-
28	*dop-1*	No	Yes	-	-	-	-	+
29	*tsp-7*	No	Yes	-	+	-	-	-
30	*rig-4*	No	Yes	-	-	-	+	-
31	*F29G6.2*	No	Yes	-	+	-	-	-
32	*F55C12.4*	No	Yes	-	+	-	+	+
33	*tmc-1*	No	Yes	-	-	+	+	-
34	*twk-7*	No	Yes	+	+	+	+	-
35	*twk-13*	No	Yes	+	-	-	-	+
36	*twk-40*	No	Yes	+	+	-	-	-
37	*twk-43*	No	Yes	+	-	-	-	-
38	*acr-21*	No	Yes	-	+	-	-	-
39	*glr-5*	No	Yes	-	+	+	-	-
40	*exp-1*	No	Yes	+	+	-	-	-
41	*mig-13*	No	Yes	+	+	-	+	+
42	*ace-4*	No	Yes	+	-	-	-	-
43	*ddr-2*	No	Yes	-	+	-	-	-
44	*unc-53*	No	Yes	+	+	+	-	-
45	*flp-18*	No	Yes	+	+	-	-	-
46	*lgc-36*	No	Yes	-	+	-	-	-
47	*slo-2*	No	Yes	+	+	-	-	-
48	*nrx-1*	No	Yes	+	-	-	-	+
49	*lgc-55*	No	No	-	-	-	-	-
50	*itr-1*	No	No	-	-	-	-	-
51	*avr-15*	No	No	-	-	-	-	-
52	*ace-3*	No	No	-	-	-	-	-
53	*cog-1*	No	No	-	-	-	-	-
Summary		**38/43** previously known *unc-3*-dependent genes show UNC-3 binding based on ChIP-Seq						

Supplemental Table 4.

	Reporter	TF family	*cis*-regulatory region included	RFP expression	*unc-3* dependence	Expression	Expression in Cholinergic MNs	Expression in GABAergic MNs	# of lines
1	*nhr-1* (second peak)	Nuclear hormone receptor	-838 bp ~ 1,535 bp	YES	Positively regulated	VNC, Head and Tail Neurons	YES	YES	2
2	*nhr-40* (second peak)	Nuclear hormone receptor	4,356 bp ~ 5,525 bp	YES	Positively regulated	VNC, Head and Tail Neurons	YES	YES	2
3	*mab-9*	T-box	-5,561 bp ~ -3,773 bp	YES	Positively regulated	VNC, Head and Tail Neurons	YES	YES	2
4	*ztf-26*	Zinc finger	910 bp ~ 3,394 bp	YES	Positively regulated	VNC and Head Neurons	YES	NO	2
5	*ceh-44*	C. elegans homeobox	1,605 bp ~ 3,111 bp	YES	Positively regulated	VNC, Head and Tail Neurons	YES	Not determined	2
6	*zfh-2*	Zinc finger	12,048 bp ~ 13,549 bp	YES	Positively regulated	VNC, Head and Tail Neurons	YES	NO	2
7	*cfi-1*	AT-rich interaction domain	-13,824 bp ~ -11,329 bp	YES	Positively regulated	VNC, Head and Tail Neurons	YES	YES	2
8	*bnc-1*	Zinc finger	-1,800 bp ~ -1 bp	YES	Positively regulated	VNC and Tail Neurons	YES	NO	2
9	*nhr-49*	Nuclear hormone receptor	-750 bp ~ 75 bp	YES	Negatively regulated	VNC, Head and Tail Neurons	NO	YES	2
10	*nhr-1* (first peak)	Nuclear hormone receptor	-7,452 bp ~ -5,871 bp	YES	Not tested	Head and Tail Neurons	N/A	N/A	2
11	*nhr-19*	Nuclear hormone receptor	1,261 bp ~ 2,294 bp	YES	NO	VNC, Head and Tail Neurons	Not determined	Not determined	2

12	*nhr-40* (first peak)	Nuclear hormone receptor	919 bp ~ 1,839 bp	YES	NO	VNC, Head and Tail Neurons	Not determined	Not determined	2
13	*ccch-3*	Zinc finger	-385 bp ~ -1 bp	YES	Not tested	Head Neurons	N/A	N/A	2
14	*ztf-17*	Zinc finger	-1,031 bp ~ 3 bp	YES	Not tested	Head Neurons, Muscles, and Intestine	N/A	N/A	2
15	*nhr-47*	Nuclear hormone receptor	-675 bp ~ 171 bp	NO	N/A	N/A	N/A	N/A	2
16	*unc-55*	Nuclear receptor	-2,105 bp ~ -1,327 bp	NO	N/A	N/A	N/A	N/A	2
17	*ztf-13*	Zinc finger	-1,392 bp ~ -1 bp	NO	N/A	N/A	N/A	N/A	2
18	*ztf-14*	Zinc finger	-2,087 bp ~ -531 bp	NO	N/A	N/A	N/A	N/A	2

Genotype

unc-3 (n3435) X
unc-3 (e151) X
lin-39(n1760) mab-5(e1239)/ht2 III; lgIs58 [[gcy-32::gfp] V
lin-39 (n1760)/dpy-17(e164) unc-32(e189) III
unc-30 (e191) IV
cfi-1 (ot786) I
unc-119(ed3); ieSi57 [Peft-3::TIR1::mRuby::unc-54 3' UTR, cb-unc-119(+)] II
otTi28 [unc-11prom8+ehs-1prom7+rgef-1prom2::TIR1::mTurquoise2::unc-54 3'UT
unc-3 (ot839 [unc-3::gfp]) X
cfi-1 (kas16 [mNG::AID::cfi-1]) I
cfi-1 (syb1812 [A4e_enhancer deletion -13,264 -12,495 from cfi-1 ATG in kas16]
cfi-1 (syb1856 [8 COE motifs mut in A4e] kas16[cfi-1::mNG::AID]) I
bnc-1(ot845[bnc-1::+mNG+AID]) V
lin-39 (kas9 [lin-39::mNG::AID]) III
otIs544 [cho-1(fosmid)::SL2::mCherry::H2B + pha-1(+)]
pha-1 (e2123) III; otEx4750 [unc-17_1kb prom::DsRed2]
pha-1 (e2123)III; otEx4746 [ace-2_1.8kb prom::DsRed2]
juIs14 [acr-2p::GFP + lin-15(+)] IV.
evIs82b [unc-129::GFP + dpy-20(+)] IV.
otIs476[glr-4 prom::TagRFP + pha-1(+)]
vsIs48 [unc-17::GFP].
pha-1 (e2123); evIs82b [unc-129::GFP + dpy-20(+)] IV; otEx5966 [[Pbnc-1_1.8kbp
nsEx37 [cfi-1_8.17kb prom::GFP + rol-6(su1006)].
pha-1(e2123); kasEx12 [cfi-1_-11,329 bp -13,824 bp::RFP + pha-1(+)] line 1
pha-1(e2123); kasEx182 [cfi-1_-11,329 bp -13,824 bp::RFP + pha-1(+)] line 2
pha-1(e2123); kasEx183 [cfi-1_-11,329 bp -13,824 bp::RFP + pha-1(+)] line 3
pha-1 (e2123) III; kasEx184 Ex[cfi-1_2.77 kb(993 bp to 3,764 bp)::rfp] Line 1
pha-1 (e2123) III; kasEx185 [cfi-1_2.77 kb(993 bp to 3,764 bp)::rfp] Line 2
pha-1 (e2123) III; kasEx186 [cfi-1_2.77 kb(993 bp to 3,764 bp)::rfp] Line 3
pha-1 (e2123) III; kasEx187 [cfi-1_1.72 kb (547 bp to -1,173 bp)::rfp] Line 1
pha-1 (e2123) III; kasEx188 [cfi-1_1.72 kb (547 bp to -1,173 bp)::rfp] Line 2
pha-1 (e2123) III; kasEx189 [cfi-1_1.72 kb (547 bp to -1,173 bp)::rfp] Line 3
pha-1 (e2123) III; kasEx190 [cfi-1_1.71 kb (-1,164 bp to -2,875 bp) ::rfp] Line 1
pha-1 (e2123) III; kasEx191 [cfi-1_1.71 kb (-1,164 bp to -2,875 bp) ::rfp] Line 2
pha-1 (e2123) III; kasEx192 [cfi-1_1.71 kb (-1,164 bp to -2,875 bp) ::rfp] Line 3

pha-1 (e2123) III; kasEx193 [cfi-1_3.27 kb(-2,865 bp to -6,141 bp)::rfp] Line 1
pha-1 (e2123) III; kasEx194 [cfi-1_3.27 kb(-2,865 bp to -6,141 bp)::rfp] Line 2
pha-1 (e2123) III; kasEx195 [cfi-1_3.27 kb(-2,865 bp to -6,141 bp)::rfp] Line 3
pha-1 (e2123) III; kasEx196 [cfi-1_3.18 kb(-8,162 bp to -11,346 bp) ::rfp] Line 1
pha-1 (e2123) III; kasEx197 [cfi-1_3.18 kb(-8,162 bp to -11,346 bp) ::rfp] Line 2
pha-1 (e2123) III; kasEx198 [cfi-1_552 bp(-11,329 bp to -11,881 bp) ::rfp] Line 1
pha-1 (e2123) III; kasEx199 [cfi-1_552 bp(-11,329 bp to -11,881 bp) ::rfp] Line 2
pha-1 (e2123) III; kasEx200 [cfi-1_552 bp(-11,329 bp to -11,881 bp) ::rfp] Line 3
pha-1 (e2123) III; kasEx201 [cfi-1_383 bp(-11,851 bp to -12,234 bp) ::rfp] Line 1
pha-1 (e2123) III; kasEx202 [cfi-1_383 bp(-11,851 bp to -12,234 bp) ::rfp] Line 2
pha-1 (e2123) III; kasEx203 [cfi-1_383 bp(-11,851 bp to -12,234 bp) ::rfp] Line 3
pha-1 (e2123) III; kasEx204 [cfi-1_499 bp(-12,223 bp to -12,722 bp)::rfp] Line 1
pha-1 (e2123) III; kasEx205 [cfi-1_499 bp(-12,223 bp to -12,722 bp)::rfp] Line 2
pha-1 (e2123) III; kasEx206 [cfi-1_579 bp(-12,705 bp to -13,284 bp)::rfp] Line 1
pha-1 (e2123) III; kasEx207 [cfi-1_579 bp(-12,705 bp to -13,284 bp)::rfp] Line 2
pha-1 (e2123) III; kasEx208 [cfi-1_579 bp(-12,705 bp to -13,284 bp)::rfp] Line 3
pha-1 (e2123) III; kasEx209 [cfi-1_561 bp(-13,263 bp to -13,824 bp)::rfp] Line 1
pha-1 (e2123) III; kasEx210 [cfi-1_561 bp(-13,263 bp to -13,824 bp)::rfp] Line 2
kasEx211 [nhr-1 (peak 1, -7455 to -5853)::RFP::unc-54 3'UTR] Line 1
kasEx212 [nhr-1 (peak 2, -893 to +1535)::RFP::unc-54 3'UTR] Line 1
kasEx213 [nhr-1 (peak 2, -893 to +1535)::RFP::unc-54 3'UTR] Line 2
kasEx214 [nhr-40 (peak 1, +938 to +1846)::RFP::unc-54 3'UTR] Line 1
kasEx215 [nhr-40 (peak 1, +938 to +1846)::RFP::unc-54 3'UTR] Line 2
kasEx216 [nhr-40 (peak 2, +4360 to +5522)::RFP::unc-54 3'UTR] Line 1
kasEx217 [nhr-40 (peak 2, +4360 to +5522)::RFP::unc-54 3'UTR] Line 2
pha-1 (e2123) III; kasEx218 [zfh-2::rfp (12,048 bp to 13,549 bp)] Line 1
pha-1 (e2123) III; kasEx219 [zfh-2::rfp (12,048 bp to 13,549 bp)] Line 2
kasEx220 [nhr-49 (-803 to +58)::RFP::unc-54 3'UTR] Line 1
kasEx221 [nhr-49 (-803 to +58)::RFP::unc-54 3'UTR] Line 2
kasEx222 [nhr-19 (+1250 to +2302)::RFP::unc-54 3'UTR] Line 1
kasEx220 [nhr-19 (+1250 to +2302)::RFP::unc-54 3'UTR] Line 2
pha-1 (e2123) III; kasEx224 [ccch-3 (Peak 1, 0 to -387)::RFP::unc-54 3’UTR] Line
pha-1 (e2123) III; kasEx225 [ccch-3 (Peak 1, 0 to -387)::RFP::unc-54 3’UTR] Line
pha-1 (e2123) III; kasEx226 [ztf-17 (Peak 1, 0 to -1034)::RFP::unc-54 3’UTR] Line
pha-1 (e2123) III; kasEx227 [ztf-17 (Peak 1, 0 to -1034)::RFP::unc-54 3’UTR] Line
pha-1 (e2123) III; kasEx228 [ztf-26 (Peak 1, -1341 to +753)::RFP::unc-54 3’UTR] L

pha-1 (e2123) III; kasEx229 [ztf-26 (Peak 1, -1341 to +753)::RFP::unc-54 3'UTR] L
pha-1 (e2123) III; kasEx230 [ztf-26 (Peak 2, +909 to +3394)::RFP::unc-54 3'UTR]
pha-1 (e2123) III; kasEx231 [ztf-26 (Peak 2, +909 to +3394)::RFP::unc-54 3'UTR]
kasEx232[mab-9 (-5569 to -3768)::RFP::unc-54 3'UTR] Line 1
kasEx233 [mab-9 (-5569 to -3768)::RFP::unc-54 3'UTR] Line 2
pha-1 (e2123) III; kasEx234 [ceh-44 (1,605 bp to 3,111 bp) ::rfp] Line1
pha-1 (e2123) III; kasEx235[ceh-44 (1,605 bp to 3,111 bp) ::rfp] Line2