HHS Public Access

Author manuscript

Cell Rep. Author manuscript; available in PMC 2024 May 14.

Published in final edited form as:

Cell Rep. 2024 March 26; 43(3): 113857. doi:10.1016/j.celrep.2024.113857.

A molecular atlas of adult *C. elegans* motor neurons reveals ancient diversity delineated by conserved transcription factor codes

Jayson J. Smith[1,2,7], **Seth R. Taylor**[3,4,7], **Jacob A. Blum**[5], **Weidong Feng**[1,2], **Rebecca Collings**[3], **Aaron D. Gitler**[5], **David M. Miller III**[3,6,*], **Paschalis Kratsios**[1,2,8,*]

[1]Department of Neurobiology, University of Chicago, Chicago, IL 60637, USA

[2]University of Chicago Neuroscience Institute, Chicago, IL 60637, USA

[3]Department of Cell and Developmental Biology, Vanderbilt University School of Medicine, Nashville, TN 37240, USA

[4]Department of Cell Biology and Physiology, Brigham Young University, Provo, UT 84602, USA

[5]Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA

[6]Program in Neuroscience, Vanderbilt University, Nashville, TN 37240, USA

[7]These authors contributed equally

[8]Lead contact

SUMMARY

Motor neurons (MNs) constitute an ancient cell type targeted by multiple adult-onset diseases. It is therefore important to define the molecular makeup of adult MNs in animal models and extract organizing principles. Here, we generate a comprehensive molecular atlas of adult *Caenorhabditis elegans* MNs and a searchable database. Single-cell RNA sequencing of 13,200 cells reveals that ventral nerve cord MNs cluster into 29 molecularly distinct subclasses. Extending *C. elegans* Neuronal Gene Expression Map and Network (CeNGEN) findings, all MN subclasses are delineated by distinct expression codes of either neuropeptide or transcription factor gene families. Strikingly, combinatorial codes of homeodomain transcription factor genes succinctly delineate adult MN diversity in both *C. elegans* and mice. Further, molecularly defined MN subclasses in *C. elegans* display distinct patterns of connectivity. Hence, our study couples the

*Correspondence: david.miller@vanderbilt.edu (D.M.M.), pkratsios@uchicago.edu (P.K.).

AUTHOR CONTRIBUTIONS

J.J.S., S.R.T., D.M.M., and P.K. were responsible for conceptualization. J.J.S., S.R.T., W.F., and R.C. were responsible for investigation. J.J.S. and S.R.T. were responsible for analysis. J.A.B. and A.D.G. integrated data into http://celegans.spinalcordatlas.org. D.M.M. and P.K. provided resources, funding, and supervision. J.J.S. and P.K. drafted the manuscript. All authors read and edited the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

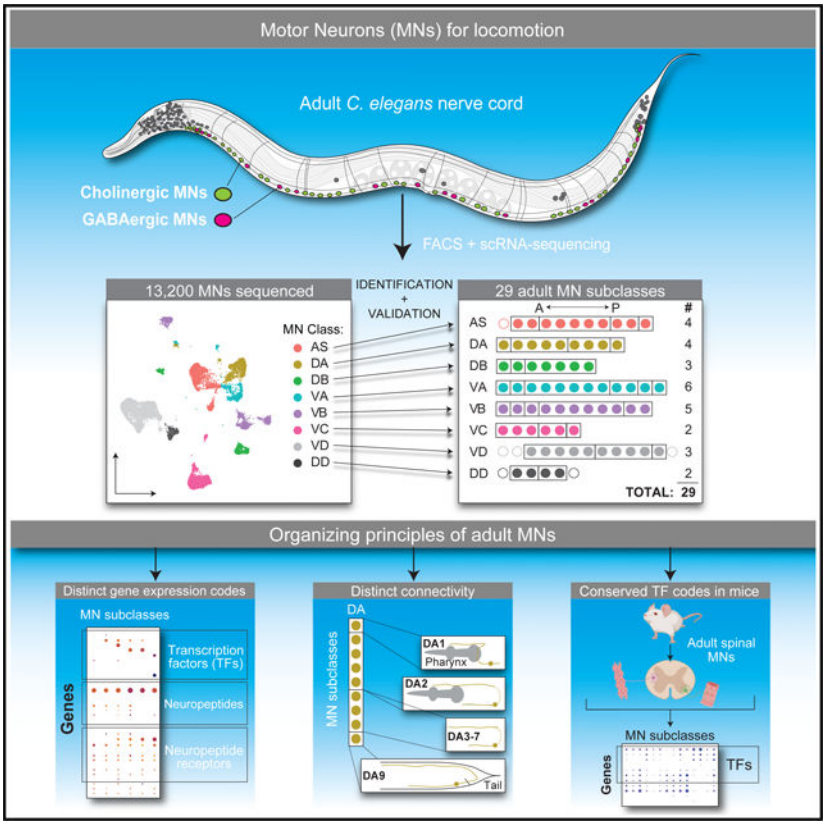
Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2024.113857.

connectivity map of the *C. elegans* motor circuit with a molecular atlas of its constituent MNs and uncovers organizing principles and conserved molecular codes of adult MN diversity.

In brief

Smith et al. provide a gene expression map of adult *C. elegans* motor neurons, uncovering striking diversity and conserved molecular codes of transcription factor genes.

Graphical Abstract



INTRODUCTION

Motor neurons (MNs) constitute the primary output of the central nervous system. Due to their stereotyped cell body positions and axonal projections, MNs have been extensively studied. Much of our current knowledge describes developmental events, including MN birth, migration, axodendritic morphogenesis, and synapse formation.[1–4] However, our understanding of how these long-lived, post-mitotic cells maintain their identity and function during adult life is limited, in part due to a lack of comprehensive molecular characterizations of adult MN diversity. A deeper understanding of adult MN molecular profiles may aid the development of effective treatments for certain adult-onset diseases, such as amyotrophic lateral sclerosis (ALS), which are characterized by progressive MN dysfunction and degeneration.

MNs represent a diverse cell population traditionally divided into subtypes (or subclasses) based on qualitative criteria, such as cell lineage, cell body position, morphology, target muscle, and selected molecular markers.[5,6] Recent advances in single-cell RNA sequencing (scRNA-seq) technology have revolutionized our ability to classify neurons across species.[7,8] In the nematode *Caenorhabditis elegans*, scRNA-seq at larval stages revealed distinct molecular signatures for individual neuron classes, including 13 transcriptionally distinct MN populations in the late larval nerve cord (analogous to the vertebrate spinal cord).[9,10] Profiling studies performed more recently generated gene expression atlases for most adult *C. elegans* cell types, including multiple neuron classes.[11,12] However, the extent of molecular diversity within individual adult neuronal classes, including MN classes, remains incompletely defined.

Similarly, in the fly *Drosophila melanogaster*,[13,14] zebrafish *Danio rerio*,[15–18] and mouse *Mus musculus*,[19–21] scRNA-seq has been used to profile MNs and their progenitors. To date, single-cell profiling in these models has been performed primarily in developmental stages. These studies revealed substantial molecular diversity within developing MNs across species, in agreement with previous work that used qualitative criteria for MN classification. However, the focus on early development in these studies has left the extent of molecular diversity within adult MNs largely uncharacterized. Elucidating the diversity in adult nervous systems is highly relevant to adult-onset MN diseases because distinct MN subpopulations in the hindbrain and spinal cord differ in disease susceptibility.[22,23] Recent studies in the mouse spinal cord support the idea that the molecular diversity of developing MNs is simplified or “trimmed down” in the adult.[19,20,24,25] However, the complexity of the mammalian spinal cord and rarity of MNs, as they make up only 0.4% of the total cells of the spinal cord, remain as major challenges toward attaining a comprehensive and spatially resolved profile of all spinal MNs.

Seminal studies in worms, flies, and mice have uncovered critical roles for transcription factors (TFs) in MN development. In particular, members of specific subfamilies of homeodomain (HD) TFs, such as LIM, HOX, and TALE, are known to affect various facets of MN development (e.g., specification, axon guidance).[4,26,27] Emerging evidence in *C. elegans* suggests that HOX genes are expressed in adult MNs and required to maintain cholinergic MN identity.[28,29] However, a comprehensive description of TF families expressed in adult *C. elegans* MNs is currently needed. Such knowledge will lay the groundwork for future studies to dissect TF functional roles in adult MNs.

In this study, we employ scRNA-seq to profile adult *C. elegans* MNs of the ventral nerve cord (VNC) and its flanking ganglia. By leveraging the simple and precisely defined anatomy of the *C. elegans* motor system (composed of 75 MNs), we generated a comprehensive molecular profile of adult VNC and ganglionic MNs. To allow interspecies comparisons, our data are deposited at http://celegans.spinalcordatlas.org, a database that also contains adult MN profiles from mouse and human. In *C. elegans*, we find that the eight cardinal MN classes, originally defined by anatomical criteria (e.g., axonal morphology, target muscle), subdivide into 29 molecularly distinct subclasses, uncovering a far greater degree of adult MN diversity than previously estimated. Consistent with CeNGEN project (https://www.cengen.org) findings on larval neuronal classes,[10] these 29 MN subclasses are

defined by the combinatorial expression of conserved genes encoding TFs from various families (e.g., nuclear hormone receptors [NHRs], HD proteins, zinc-finger [ZF] proteins), neuropeptides, and neuropeptide receptors. Importantly, analysis of available mouse MN transcriptomes revealed that multiple HD and NHR proteins are expressed in these cells in the adult. Strikingly, we found—in both *C. elegans* and mice—that combinatorial codes of HD TFs constitute the "minimal molecular descriptors" of adult MN diversity, consistent with recent studies showing that combinatorial HD codes delineate all the neuron classes of *C. elegans*.[10,30,31] Last, by using the *C. elegans* wiring diagram, we found that molecularly defined subclasses display distinct patterns of connectivity. Hence, the connectivity map of the adult *C. elegans* motor circuit is now coupled with a molecular atlas of its constituent MNs, paving the way for future investigations of MN development and function.

RESULTS

Single-cell profiling of adult *C. elegans* MNs reveals striking molecular diversity

The *C. elegans* motor circuit in hermaphrodites consists of six cholinergic (AS, DA, DB, VA, VB, VC) and two GABAergic (DD, VD) MN classes that control locomotion (DA, DB, VA, VB, AS, DD, VD) or egg-laying (VC) (Figure 1A).[32,33] Each class is defined by its distinct morphology and contains a fixed number of neurons (AS, 11 neurons; DA, nine; DB, seven; VA, 12; VB, 11; VC, six; DD, six; VD, 13) that intermingle along the VNC (58 cells) and in its flanking ganglia (17 cells), totaling 75 cells. Here, we developed a genetic labeling strategy to isolate, in day 1 adult *C. elegans* hermaphrodites, all 58 MNs located in the VNC, as well as the majority (11 of 17) of ganglionic MNs located in the retrovesicular (RVG) and preanal (PAG) ganglia that flank the VNC.

We used fluorescence-activated cell sorting (FACS) to isolate cells[10,34] marked with either *acr-2p::gfp* or *lin-39p::rfp* transgene expression and then profiled their transcriptomes by independent scRNA-seq experiments (Figures 1B–1D; File S1). After quality control, computational analysis, and integration (see STAR Methods), we identified 13,200 *bona fide* ganglionic and VNC MNs representing all eight cardinal MN classes (Figures 1D and 1E), each exhibiting known class-specific gene expression and neurotransmitter identity (Figures 1F and 1H; File S1). In total, we profiled 69 of 75 (92%) adult ganglionic and VNC MNs (Figures 1B and 1C), with a minimum of 420 cells (DD), a maximum of 3,758 (VD), and an average of 1,650 cells captured per MN class (File S1). The six cells not captured are located either in the retrovesicular (AS1, DD1, VD1, VD2) or preanal (DD6, VD13) ganglia and were not labeled by our genetic strategy (Figure 1C).

Consistent with the *acr-2p::gfp* and *lin-39p::rfp* expression patterns (Figure 1C), MNs of the DA, DB, VA, VB, AS, VC, DD, and VD classes were derived from *lin-39p::rfp*-expressing cells, whereas MNs from the DA, DB, VA, and VB classes were derived from cells expressing both reporters (Figure 1G). Importantly, although certain DA, DB, VA, and VB cells were independently isolated from either *acr-2p::gfp* or *lin-39p::rfp* animals, these cells cluster together after computational integration (Figure 1G), suggesting that neither transgene significantly perturbs MN gene expression. In addition, coherent cell clusters that are exclusively derived from the *lin-39p::RFP* strain (AS, VC, DD, VD) represent distinct MN classes (Figures 1E and 1G), indicating that the small fraction of L4-stage larval cells

(~20%) in the *lin-39p::rfp* preparation (see STAR Methods for asynchronous growth of *lin-39p::rfp* animals) substantially overlaps with adult MN profiles. Overall, we achieved high sequencing depth across all eight cardinal MN classes, with a median of 1,300 unique molecular identifiers (UMIs) and a median of 603 genes detected per cell (File S1). Notably, this day 1 adult dataset revealed molecular separation between DD and VD MNs (Figures S1A and S1B), which were inseparable in previous scRNA-seq datasets from developmental (larval stage 4 [L4])[10] and young adult[11] stages, possibly due to lower sequencing depth, variable MN sampling efficiency in the L4 dataset, and/or lack of FACS-based enrichment in the previously published adult dataset.

To investigate the extent of molecular diversity within these adult MNs (VNC and ganglionic), we generated separate uniform manifold approximation and projection (UMAP) representations from the cells belonging to each cardinal MN class (AS, DA, DB, VA, VB, VC, DD, VD). This approach revealed transcriptionally distinct subsets, which we describe as MN subclasses (See STAR Methods for a detailed description of subclass annotation, Figures S2–S4; File S1). At the level of the eight cardinal MN classes (AS, VA, VB, VD, VC, DA, DB, DD), there is considerable agreement in gene expression between our study and a previously published adult *C. elegans* dataset (Tables S1–S3).[11] However, at the level of adult MN subclasses, our analysis identified a striking degree of molecular diversity. In day 1 adults, we found 29 molecularly distinct MN subclasses; five are GABAergic (DD, two subclasses; VD, three) and 24 are cholinergic (AS, four; DA, four; DB, three; VA, six; VB, five; VC, two) (Figures 2A, S4A, and S4B). The VA class is the most diverse (six subclasses), whereas the VC and DD classes are the least diverse (two subclasses). Differential gene expression revealed that subclasses within each cardinal class can be distinguished based on enrichment of dozens of genes, ranging from 20 (DD2–3 subclass) to 421 (VC4–5) (File S2). Overall, our single-cell analysis revealed strong molecular diversity within all eight MN classes of the adult *C. elegans* nerve cord and its flanking ganglia, subdividing them into 29 distinct subclasses.

Comparison of larval and adult MN transcriptomes

Using scRNA-seq results from L4 larvae, the *C. elegans* Neuronal Gene Expression Map and Network (CeNGEN) consortium previously described 11 distinct subclasses within cholinergic MNs (e.g., DA9, DB1, VA12, VB1, VB2, VC4/5) and no subclasses within GABAergic MNs[10] of the VNC and its flanking ganglia (RVG, PAG). In fact, MNs in the two cardinal GABAergic classes, DD and VD, were combined into a single cluster at L4.[10] The current day 1 adult single-cell dataset almost triples (11 vs. 29 subclasses) the degree of previously described ganglionic and VNC MN diversity in *C. elegans* (Table S1). Thus, we asked whether the increase in adult MN diversity reflected changes in detection (i.e., MN sampling, sequencing depth) or MN development/maturation.

Because we identified many genes with enriched expression in each of the 29 adult MN subclasses (File S2), we used this knowledge to re-annotate the L4 MN dataset from CeNGEN. This analysis identified 18 subclasses of cholinergic MNs at L4 (File S2), an increase of seven (11 vs. 18) over the original annotation.[10] Specifically, we observed two AS (AS1–10, AS11), four DA (DA1, DA2–5, DA6–8, DA9), three DB (DB1, DB2,

DB3–7), four VA (VA1, VA2–8, VA9–11, VA12), three VB (VB1, VB2, VB3–11), and two VC (VC1–3/6, VC4–5) subclasses in the L4 dataset. Fourteen of these subclasses corresponded to the same anatomical cells both in L4 and adult datasets (Table S1; File S2). Upon re-annotation, we also resolved the two GABAergic classes of MNs (DD, VD) at L4, indicating they are indeed molecularly distinct prior to the adult stage.

At least three non-mutually exclusive factors could explain the detection of increased MN diversity in adults (29 subclasses) relative to larvae (18 subclasses): (1) increased cell sampling (i.e., more MNs were sequenced in our adult versus the L4 dataset), (2) increased gene detection (i.e., higher number of UMIs and genes per cell detected in adult MNs), and/or (3) our analysis may have detected a *bona fide* increase in molecular diversity of MNs as the animals transition from larval (L4) to adult life. For most anatomical classes and subclasses, we sequenced a greater number of adult cells and detected more UMIs and genes/cells than the L4 data (File S1).

The detection of 18 larval and 29 adult MNs subclasses suggests that the molecular diversity that characterizes developing MNs is not trimmed down but is either maintained or elaborated in the adult stage. Consistent with the hypothesis that MN diversity is elaborated in the adult, behavioral analyses have shown differences in locomotor patterns between L4 and adult worms.[35,36]

Gene Ontology analysis of the adult MN transcriptome in *C. elegans*

To increase our confidence in distinguishing gene expression from noise, we used an established dynamic thresholding procedure based on a high-confidence ground-truth dataset composed of 169 genes expressed in the *C. elegans* nervous system (see STAR Methods; Figure S4C; File S3).[10] To obtain a comprehensive picture of the molecular makeup of adult *C. elegans* MNs, we conducted Gene Ontology (GO) analysis with WormCat[37] on the thresholded list of transcripts (9,275 in total) expressed across the 29 ganglionic and VNC MN subclasses (Figure 2B; File S4). The top three categories over-represented in adult MNs relative to the whole genome contain genes encoding proteins necessary for (1) metabolism, with 1,050 genes (11.3%) involved in lipid metabolism, glycolysis, etc.; (2) neuronal signaling, with 724 genes (7.8%) involved in calcium signaling, small GTPase pathways, etc.; and (3) transmembrane transport, with 435 genes (4.7%) encoding ion channels, neurotransmitter receptors, etc. (Figure 2B; File S4). These categories were immediately followed by genes involved in RNA biology (348 genes, 3.8%; e.g., mRNA binding, processing, and methylation), proteolysis (322 genes, 3.4%), and gene regulation (234 TFs, 134 chromatin factors) (Figure 2B; File S4).

Distinct codes of homeodomain proteins and NHRs delineate all MN subclasses

We next aimed to identify molecular descriptors for each ganglionic and VNC MN subclass. We specifically focused on TF-en-coding genes (Figure 2B; Files S4 and S5) as TFs tend to be reliable classifiers of neuronal identity across species.[5] The 234 TF genes expressed in adult MN subclasses belong to 13 conserved families (Figures 2C and 2D; File S5). Most (186 of the 234 [79.5%]) encode proteins from five TF families: 86 NHR, 39 HD (Table S2), 28 ZF, 19 basic leucine zipper (bZIP), and 14 basic-helix-loop-helix (bHLH)

proteins (Figure 2C). Strikingly, 206 of the 234 TF genes (88%) are expressed in specific MN subclasses. The remaining 28 TF genes are expressed in all 29 MN subclasses (Figures 2C and 2D; File S5). Each subclass expresses an average of 122 TF genes, with a minimum of 86 (VB3 subclass) and a maximum of 154 TF genes (VC1–3 and 6). Importantly, we found that each MN subclass expresses a distinct combination of TF genes (Figure 2D; File S5). As an independent validation, we cross-referenced our data with a previously published adult *C. elegans* dataset.[11] Although that study had different scope and did not employ FACS for MN enrichment, we found a significant agreement in TF gene expression across the eight cardinal MN classes (Tables S2 and S3).

A key observation is that combinatorial expression codes of either HD or NHR TF genes are sufficient to delineate all 29 adult MN subclasses (Figure 2D; File S5). Interestingly, the HD TF gene codes constitute the minimal (succinct) descriptors of adult MN diversity, as we detected subclass-specific expression of only 35 HD TF-encoding genes compared to 81 NHR-encoding genes. Importantly, endogenous fluorescent reporters (direct protein fusions) for several of these HD TF genes (e.g., *vab-7*, *unc-4*, *ceh-20*, *ceh-58*, *ceh-89*) revealed expression in MNs during larval stages,[30,31] suggesting that these HD TF codes are established before the adult stage. Last, this analysis not only validates previously reported TF expression patterns (e.g., *unc-3*, *cfi-1*, *mab-9*, *alr-1*, *dve-1*, *lin-11*, *irx-1*) (Figure 2D, right panel) but also offers dozens of additional molecular markers for adult *C. elegans* MN classes and subclasses.

Four *C. elegans* Hox genes delineate most adult MN subclasses

Four of the 39 HD TF genes encode Hox proteins (Figures 2D and 3A), traditionally described as fundamental regulators of anterior-posterior patterning during early animal development.[38–40] Due to the surprisingly persistent Hox gene expression in adult *C. elegans* MNs, we validated our scRNA-seq data with endogenous reporters (protein fusions) for all six *C. elegans* Hox genes (Figure 3B; Tables S3 and S4; Figure S5). We employed CRISPR-Cas9 to generate endogenous fluorescent (mNeonGreen) reporters for three Hox genes (*ceh-13* [*Lab/Hox1*], *mab-5* [*Antp/Hox6-8*], and *egl-5* [*Abd-A/Abd-B/Hox9-13*]) and also used available endogenous reporters for the remaining three Hox genes (*lin-39* [*Scr/Dfd/Hox3-5*], *php-3* [*Abd-B/Hox9-13*],[41] and *nob-1* [*Abd-B/Hox9-13*]).[41,42] To generate a complete Hox gene expression map in adult MNs with single-cell resolution, we assessed the colocalization of each Hox gene reporter with available RFP fluorescent markers for each MN class in day 1 adult animals (see STAR Methods) (Figures 3B and S5). Consistent with the scRNA-seq data, only four endogenous Hox gene reporters are detected in adult MNs (c*eh-13*, *lin-39*, *mab-5*, *egl-5*). Further, single-molecule RNA fluorescence *in situ* hybridization (FISH) and scRNA-seq data for *lin-39* largely agree with endogenous LIN-39::mNG protein expression (Figures 3C and S5; Table S4). Importantly, this analysis spatially resolved all 29 MN subclasses *in vivo* with single-cell resolution (Figure 3C).

We found that a combinatorial Hox code fully delineates MN diversity only in three (DA, DD, VC) of the eight classes (Figure 3C; Tables S2–S4). In the DA class, the anterior subclass DA1 (composed of a single neuron called DA1) exclusively expresses *ceh-13*; the midbody DA2–5 subclass (three neurons: DA2–5) expresses both *ceh-13* and *lin-39*;

the posterior subclass DA6–8 expresses *ceh-13*, *lin-39*, and *mab-5*; and the most posterior subclass, DA9, exclusively expresses *egl-5* (Figure 3C). In the VC class, the VC1–3/6 subclass expresses *ceh-13* and *lin-39*, whereas the VC4–5 subclass expresses *lin-39* and *mab-5* (Figure 3C). In the DD class, subclasses DD2–3 and DD4–5 show distinct Hox gene expression, and this is also the case for the two single neurons DD1 and DD6 (both not captured by our genetic labeling) (Figures 3B and 3C).

In the remaining MN classes (cholinergic and GABAergic), we found that Hox expression codes are insufficient to fully delineate MN diversity. For example, in the VA MN class, only three of the six VA subclasses express a distinct combination of Hox genes (VA2, *ceh-13*; VA3–8, *ceh-13* and *lin-39*; VA12, *egl-5*) (Figures 3C–3E). The most anterior subclass VA1 does not express any Hox genes (Figures 3C–3E). Instead, the TF *elt-1*/GATA1–3 is specifically expressed in VA1 (Figures 3F and 3G). Thus, only three (VA2, VA3–8, and VA12) of the six VA subclasses are distinctly delineated by a combinatorial Hox code (Figures 3C and 3E).

Like VA, four other MN classes (VB, DB, AS, VD) are partially delineated by Hox gene expression (three of five VB subclasses, two of three DB subclasses, three of four AS subclasses, two of three VD subclasses) (Figure 3C). Like VA1, the anterior subclasses DB1, VB1, and VB2 do not express any Hox genes but distinctly express other TFs. Based on our scRNA-seq data, we found within the DB and VB classes that (1) adult DB1 and VB1 distinctly express the TF *sptf-1* (human SP7), consistent with *sptf-1* expression at L4[10]; and (2) DB2 and VB2 distinctly express the bHLH gene *hlh-32* (human BHLHE22), a finding validated *in vivo* with a CRISPR-Cas9-engineered endogenous reporter (Figure S6A; Table S3).

Four important conclusions emerge from this analysis. First, congruence between mRNA (scRNA-seq) and protein (endogenous reporters) detection methods firmly establish that four Hox genes are expressed in adult *C. elegans* MNs. Second, three Hox genes (*lin-39*, *mab-5*, *egl-5*) exhibit spatial collinearity in adult MNs; i.e., the anterior boundary of *lin-39* expression in MNs is located anterior to the *mab-5* boundary (Figures 3A and 3B). Third, Hox genes in adult *C. elegans* are not expressed in the anterior ganglionic MNs (e.g., VA1, DB1, VB1, VB2), reminiscent of vertebrate nervous systems where Hox expression is restricted to the hindbrain and spinal cord.[26] Last, combinatorial expression of four Hox genes is insufficient to fully delineate MN diversity, but combinatorial expression of 35 HD TF-encoding genes, including the four Hox genes, is sufficient to delineate each of the 29 adult MN subclasses (Figure 2D; File S5).

New markers for adult MN subclasses

Next, we used fluorescent reporters to validate additional scRNA-seq data and identify markers for individual ganglionic and VNC MN subclasses (Figure S6; Table S3). For example, the *elt-1*/GATA1–3 endogenous reporter that marks VA1 (Figure 3F) is also selectively expressed in DA1 among DA MNs, all 11 MNs of the AS class, as well as DD and VD MNs (Figure 3G; Table S3). We did not detect *elt-1* endogenous reporter expression in any other MN of the adult nerve cord. Similarly, we used endogenous reporters for the TFs *egl-44* (human TEAD2/4) and *vab-3* (human PAX6) to detect highly

restricted expression: EGL-44 is in one anterior MN subclass (AS2–3), whereas VAB-3 is in three posterior MN subclasses (AS11, VA11, VD12) (Figures S6B and S6C; Table S3), consistent with previously reported *vab-3* expression.[31] Further, we identified *ilys-4*, a gene predicted to enable lysozyme activity and expressed in DD and VD neurons, as an additional marker for DB1, VB1, and VC1–3/6 subclasses (Figure S6D; Table S3). Finally, 15 of the 29 MN subclasses distinctly express at least one gene. We provide a list of the 128 genes identified as exclusively expressed in single subclasses (File S3). Intriguingly, 64 of the 128 genes (50%) are only expressed in the VC4–5 subclass, perhaps a reflection of its distinct functional role and dual neurotransmitter identity; VC4 and VC5 are the only nerve cord MNs that predominantly innervate vulva muscle cells.[33] VC4 and VC5 are not only cholinergic but also serotonergic as they express genes (e.g., *cat-1*, *tph-1*) involved in the catecholamine pathway (File S3).[43,44] We also provide a list of 720 genes selectively expressed in a single subclass within a given MN class (File S3). Thus, we provide additional markers, validated for subclass-specific expression, as well as lists of genes with expression in specific MN subclasses, enabling genetic access to previously inaccessible subsets of *C. elegans* MNs.

Neuropeptides and neuropeptide receptors delineate all MN subclasses

Each MN subclass in the *C. elegans* nerve cord and flanking ganglia expresses a distinct combination of genes encoding HD and NHR TFs (Figure 2D). We next wondered whether other gene families can be used as succinct descriptors of adult MN diversity. We therefore investigated neuropeptides and neuropeptide receptors because (1) these gene families are found within the top three GO categories (e.g., signaling, transmembrane proteins) of genes differentially expressed within MNs (Figure 2B; Files S2–S4), and (2) the CeNGEN project discovered that each *C. elegans* neuron class at L4 expresses a distinct code of neuropeptides and neuropeptide receptors.[10] Our scRNA-seq analysis revealed that adult cholinergic and GABAergic MNs express a total of 76 genes encoding neuropeptides from two families: 24 FMRFamide-like peptide (*flp*) genes and 52 neuropeptide-like protein (*nlp*) genes (Figure 4A). Moreover, we detected expression of 55 genes encoding neuropeptide receptors from four families: 30 neuropeptide receptors (*npr*), 16 FMRFamide peptide receptors (*frpr*), eight dromyosuppressin receptor-related (*dmsr*) genes, and one tachykinin receptor (*tkr*) gene (Figure 4B). In total, 131 genes (76 neuropeptides and 55 neuropeptide receptors) were detected; 96.2% of them (126 of 131) were expressed in specific MN subclasses, with a minimum of 24 (in VA2 and VD3–7 subclasses) and a maximum of 53 (VC4–5 subclass) genes expressed per subclass (Figures 4A and 4B). Each MN subclass expresses on average 14 neuropeptides and 13 neuropeptide receptors, suggesting that extra-synaptic signaling may be important for MN function in adult *C. elegans*. Remarkably, every one of the 29 MN subclasses is molecularly defined by the expression of a distinct combination of neuropeptides and neuropeptide receptors.

We used fluorescent reporters for five neuropeptide genes (*flp-7*, *flp-11*, *nlp-7*, *nlp-11*, *nlp-13*) to validate their scRNA-seq expression *in vivo* (Figures 4C and S6E–S6H). For example, an endogenous *nlp-11* reporter is expressed in 74% (23 out of 31) of cholinergic MNs in which *nlp-11* transcripts were detected by scRNA-seq (Figure 4C). Similarly, an endogenous *nlp-13* reporter shows 72% (18 out of 25 cells) concurrence with the scRNA-seq

results (Figure S6E). We provide a detailed expression analysis in individual MN subclasses (Figures 4C and S6E; Table S3). Next, we used transgenic (not endogenous) reporter lines for *flp-7*, *flp-11*, and *nlp-7* to monitor expression *in vivo*. The *flp-7* and *nlp-7* transgenic reporters show strong agreement with our scRNA-seq results with expression in 66.7% (2 out of 3) and 88.9% (8 out of 9) of MNs subclasses, respectively, but were also detected in MNs not predicted by our scRNA-seq data. A transgenic reporter for *flp-11* showed poor agreement with expression in only 2.2% (1 out of 44) of predicted MNs (Figures S6F–S6H; Table S3). These disparities could be due to incomplete gene regulatory regions included in the transgenic reporter lines for *flp-7*, *flp-11*, and *nlp-7*. As an independent validation, we interrogated a previously published scRNA-seq dataset of adult *C. elegans*[11] and a study conducted at L4[45] and found considerable agreement in terms of MN expression for several neuropeptide genes and receptors (*flp-7*, *flp-11*, *nlp-7*, *nlp-11*, *nlp-13*, *flp-18*, *flp-27*, *flp-28*, *flp-32*, *nlp-45*, *frpr-19*, *dmsr-2*, *dmsr-6*, *tkr-1*) (Table S3). We conclude that our analysis has identified additional molecular markers for a total of 12 MN subclasses (VA1, VA9–10, VA11, VA12, DA2–5, DA6–8, DA9, DB1, AS9–10, AS11, VC1–3/6, VC4–5) (Table S3). Altogether, our analysis provides a comprehensive map of neuropeptide and neuropeptide receptor gene expression in adult *C. elegans* MNs at single-cell resolution, complementing previous studies in larval and adult *C. elegans* neurons.[10–12,46]

Molecularly defined MN subclasses display differences in synaptic connectivity

The high degree of molecular diversity within adult MNs (75 MNs subdivided into 29 subclasses) is striking, but what is its biological significance? These 75 cells were historically organized into eight cardinal classes (AS, DA, DB, DD, VA, VB, VC, VD) based on qualitative criteria, such as axonal morphology and target muscle innervation.[33] Our re-annotation of the CeNGEN data identified 18 MN subclasses at L4, whereas our adult dataset revealed 29 molecularly distinct subclasses. We therefore hypothesized that the molecular differences within MNs of a given class correlate with subclass-specific traits, such as distinct connectivity patterns. To investigate this idea, we used available connectivity data (https://nemanode.org) derived from electron microscopy (EM) reconstructions of all 75 adult MNs in the *C. elegans* hermaphrodite.[33,47] This comprehensive dataset includes both chemical and electrical synapses for each MN. A consistent principle revealed by our analysis is that the degree of molecular diversity among MNs correlates with connectivity differences, as exemplified below for DA, DB, and VC neurons.

Consistent with their original classification,[33] all members of the DA class (nine neurons in total) share basic anatomical features. All DAs possess anteriorly directed neurites, innervate dorsal body wall muscles (dBWMs), and receive input from the same pre-motor interneurons (e.g., AVA) (Figure 5A; File S6). However, close analysis of each DA neuron's connectivity revealed differences that correlate with cell body position along the rostrocaudal axis. DA1, the only DA neuron located in the anterior ganglion (RVG), displays a distinct connectivity pattern (DA1 synapses with FLP, AS1, DD1, VB11) and specifically innervates dBWM cell 8 (Figures 5A and 5B; File S6). No other DA neuron shares this connectivity pattern. DA9, located in the posterior ganglion (PAG), also displays distinct connectivity among DA neurons (DA9 connects to DVA, DB7, DD6, DA8, AS11, VA11, VA12, dBWM 22–24) (Figures 5A and 5B; File S6). Notably, both DA1 and DA9 neurons

are molecularly distinct based on our molecular profiling (Figure 2A). Similarly, DA2–5 neurons, located in the anterior half of the VNC, show a distinct pattern of connectivity compared to other DA neurons (Figures 5A and 5B; File S6) and are also molecularly different (Figure 2A). The remaining DA neurons (DA6–8) constitute a single molecularly defined subclass (Figure 2A), but they do show connectivity differences; i.e., DA8 adopts connections that are not observed for DA6 and DA7. Because a handful of published reporter genes specifically label DA8 among the DAs,[29] we surmise that cluster analysis of our scRNA-seq analysis was insensitive to subtle molecular differences.

Like members of the DA class, DB neurons are united by a set of shared criteria: posteriorly directed neurites, innervation of dBWMs, and electrical synapses with the AVB pre-motor interneuron (Figure 5C; File S6). Our detailed analysis of synaptic data for DB MNs revealed distinct connectivity patterns for DB1 (connects to FLPR, RID, PVDL, VD1, VD2, DB2, VB3, DA1, DA2, DA6) and DB2 (connects to DA3, DA4, DB1, DB3, VB1, VB4, AS3) (Figures 5C and 5D; File S6). DB1 and DB2 are the only DBs with cell bodies in the anterior ganglion (RVG). Both DB1 and DB2 neurons are molecularly distinct based on our scRNA-seq profiling (Figure 2A). Based on connectivity patterns, the remaining DB neurons fall into two categories: DB3–6 and DB7 (Figures 5C and 5D; File S6). However, our molecular profiling grouped these five DB neurons as one subclass (DB3–7) (Figure 2A).

We extended this analysis to the VC class of hermaphrodite-specific MNs. VC1–3 and VC6 extend long processes into the VNC, where they connect to body wall muscle and other neurons (Figures 5E and 5F). On the other hand, VC4 and VC5 send short processes and predominantly innervate vulva muscle cells (vm2) (Figures 5E and 5F). These connectivity differences match our molecular analysis, as the VC class is divided into two molecularly distinct subclasses (VC1–3/6 and VC4–5) (Figure 2A).

We conducted a similar analysis for the remaining classes of cholinergic MNs (VA, VB, AS) (File S6). Overall, we observe that the majority of the molecularly defined MN subclasses display connectivity differences. This remarkable congruence of molecular and anatomical classification in MN subclasses likely extends to many other *C. elegans* neurons, as originally proposed.[5]

Codes of homeodomain TFs delineate adult MN diversity in mice

We next sought to determine whether the organizing principles we describe for adult MNs in *C. elegans* extend to the adult mammalian nervous system. We focused on the mouse spinal cord, which contains two cardinal MN classes: (1) skeletal MNs, which innervate skeletal muscles and control voluntary movement; and (2) visceral MNs, which synapse onto peripheral ganglia of the sympathetic chain and control involuntary movement of smooth muscles[49] (Figure 6A). We specifically asked whether the simple organizing principle of combinatorial HD or NHR gene expression codes we describe in *C. elegans* can also delineate adult MN diversity in the mouse spinal cord.

To this end, we leveraged a single-cell transcriptomic atlas that revealed molecular diversity of adult mouse spinal MNs (skeletal, 14 subclasses; visceral, 15 subclasses).[25]

We conducted an independent analysis of the mouse transcriptomic dataset (see STAR Methods) and found a similar degree of MN diversity to that reported previously.[25] Next, we assembled a list of all mouse HD- and NHR-encoding genes (see STAR Methods) and examined their expression in adult spinal MN subclasses. To increase stringency, we only considered transcripts detected in >5% of the MNs in at least one subclass. In total, we detected 76 out of the 240 mouse HD genes (31.6%) (Figure 6B; File S5) and noted expression patterns ranging from restricted (one or more subclasses) to broad (all subclasses) in adult mouse MNs (Figure 6C). We independently confirmed expression in adult MNs by analyzing available RNA *in situ* hybridization data (Allen Brain Atlas) (Figure 6C). Twenty-two of the 76 HD genes (28.9%) encode HOX TFs (Figures 6B and 6C), reiterating the importance of sustained HOX expression in adult MNs. Two of the 22 Hox genes, *Hoxc4* and *Hoxc5*, are expressed in all 29 mouse MN subclasses. Similar to our observations in *C. elegans* (Figure 2D), the remaining 20 Hox genes are expressed selectively, ranging from a single (e.g., *Hoxa11*, *Hoxc13*) to 23 (e.g., *Hoxa7*) mouse MN subclasses. Altogether, we found distinct Hox codes for all 14 skeletal subclasses and six of the 15 visceral subclasses (Figure 6C). Hence, our analysis suggests that combinatorial expression of Hox genes in adult MN subclasses is conserved from *C. elegans* to mice. Hox genes and other HD TFs (e.g., LIM family) are important determinants of spinal MN identity during development.[51–53] Hence, the expression patterns of Hox, LIM, and other HD genes we detect in adult MNs are likely established during development (Figure 6C).

We also detected expression of 43 out of the 119 mouse NHR genes (36.1%) (Figure 6D). Compared to HD genes, most NHR genes are expressed broadly in all adult MN subclasses in mice, albeit at variable levels (Figure 6D). Interestingly, NHR genes tend to be expressed in specific MN subclasses in *C. elegans* (Figure 2D). This observation together with an explosive expansion of NHR genes (284 in total) in the worm[54] point toward divergent functional roles for NHR TFs in MNs. Because hormones play crucial roles in the sympathetic nervous system,[55,56] it is conceivable that the detected expression of NHR genes in mouse visceral MNs is biological meaningful.

Finally, we asked whether the combinatorial expression of either HD or NHR genes is sufficient to delineate adult MN diversity in the mouse spinal cord. For this, we averaged the expression of HD or NHR genes across all cells in each mouse MN subclass and then compared all subclasses (see STAR Methods). Clustering mouse MN subclasses on HD gene expression alone is sufficient to correctly cluster individual MN subclasses into skeletal and visceral groups (Figure 6E), suggesting that combinatorial HD codes may constitute the minimal (succinct) descriptors of adult MN diversity in mice. However, this was not the case when we clustered mouse MN subclasses based solely on NHR gene expression (Figure 6F). Altogether, our analysis of *C. elegans* and mouse MN transcriptomes uncovered an evolutionarily conserved organizing principle: individual classes of adult MNs can be molecularly delineated by the combinatorial expression of HD TF genes.

DISCUSSION

In this study, we leveraged the simplicity of the *C. elegans* motor system to profile adult VNC and ganglionic MNs in a comprehensive and spatially resolved manner. We

discovered a striking degree of molecular diversity in the adult *C. elegans* nerve cord, with eight MN classes previously defined by anatomy now subdividing into 29 molecularly defined subclasses. Our analysis suggests that the molecular diversity that characterizes developing *C. elegans* MNs is either maintained or elaborated in the adult stage. Importantly, we extracted several organizing principles of the adult motor system. First, we identified conserved gene families (TFs, neuropeptides, and neuropeptide receptors), whose combinatorial expression delineates each of the 29 MN subclasses. Second, combinatorial codes of HD TFs define adult MN diversity both in *C. elegans* and mice. Third, we leveraged the available *C. elegans* connectome[33] to determine that molecularly defined subclasses display distinct patterns of connectivity. Hence, the connectivity map of the adult *C. elegans* motor circuit is now coupled with a molecular atlas of its constituent MNs, paving the way for new opportunities to investigate molecular mechanisms of adult motor circuit function.

Codes of HD TFs delineate adult MN diversity in *C. elegans* and mice

In total, we identified 206 genes encoding TFs expressed in specific MN subclasses (Figure 2D; Files S4–S5). Three important conclusions emerge from this analysis. First, although combinations of TFs from various families are known to define distinct identities of developing MNs across species,[4] it is not clear whether distinct TF gene expression codes exist for distinct MN subpopulations in the adult. Extending previous results for MN classes,[10,11,30,31] we found that each of the 29 adult MN subclasses in the *C. elegans* VNC and flanking ganglia expresses a distinct TF code. Second, most of the identified TFs belong to five conserved families: 86 NHR, 39 HD, 28 ZF, 19 basic leucine zipper (bZIP), and 14 basic-helix-loop-helix (bHLH) proteins. Because several members of these families are expressed in adult spinal MNs in mice,[57] similar TF codes may delineate adult MN diversity in mammals. This is consistent with our findings in the mouse spinal cord, where combinatorial codes of HD TFs, as in *C. elegans*, could define adult MN diversity (Figure 6), substantiating the recent hypothesis that combinatorial HD codes of neuronal diversity are ancient and conserved across species.[30,31] Third, a recent study showed that each of the 118 *C elegans* neuron classes is defined by a distinct combination of HD proteins.[30] Our findings suggest that this principle also applies at the level of neuronal subclasses. Each of the 29 MN subclasses expresses a distinct combination of HD TF-encoding genes (Figure 2D). Notably, NHR TF genes are also expressed in distinct combinations across *C. elegans* MN subclasses, but the number of NHR-encoding genes expressed in adult MNs is double that of HD-encoding genes (81 NHRs versus 35 HDs) (Figure 2D; File S5). Hence, combinatorial codes of 35 HD TF genes constitute the “minimal descriptor” of adult MN diversity in *C. elegans*.

Future functional studies are needed to determine whether the TF codes we describe constitute functional TF codes essential for adult MN identity and function. In *C. elegans*, combinations of TFs have been described for several neuron types,[30,31,58–62] but studies of TF function in adult neurons remain rare.[63] In the case of nerve cord MNs, only a handful (e.g., *unc-3/EBF*, *cfi-1/ARID*, *bnc-1/BNC*, *mab-9/TBX*, *lin-39/HOX*)[28,64–66] of the 234 TFs we found to be expressed in adult MNs are known to have *bona fide* roles on adult MN

identity and function. This can be addressed by auxin-inducible TF degradation in adult stages to bypass developmental effects caused by constitutive TF alleles.

Hox codes in adult MNs: What is their functional role?

Our study revealed that four of the six *C. elegans* Hox genes are expressed in adult MNs. Across species, Hox genes have been studied extensively during early developmental patterning,[26,67–69] but their expression and function in the adult stage are incompletely defined. Using scRNA-seq profiling and endogenous reporters, we show that four Hox genes —*ceh-13* (*Lab/Hox1*), *lin-39* (*Scr/Dfd/Hox4–5*), *mab-5* (*Antp/Hox6–8*), and *egl-5* (*Abd-A/Abd-B/Hox9–13*)—are expressed in adult MNs along the rostrocaudal axis. Our findings on *lin-39*, *mab-5*, and *egl-5* are largely consistent with recent *C. elegans* studies that used transgenic reporters to characterize Hox gene expression.[29,31,62] We note that some rostrally located MN subclasses (e.g., VA1, VB1, VB2, DB1) do not express any Hox genes based on scRNA-seq and endogenous reporter data. This finding parallels Hox gene expression in mammalian systems where their rostral-most expression boundary terminates in the hindbrain.[26,70]

What roles do Hox genes play in adult neurons? In *C. elegans*, Hox genes *ceh-13*, *lin-39*, *mab-5*, and *egl-5* are expressed both in developing and adult MNs (Figure 2).[28,71] In mice, Hox genes are expressed in developing mouse MNs,[26,72] and we found that their expression persists into adulthood (Figure 6C). The sustained Hox expression, from embryo to adult, in specific MNs along the rostrocaudal axis points toward non-canonical roles in the adult nervous system. Recent work showed that *lin-39* (*Scr/Dfd/Hox4–5*) is required in the adult to maintain the neuro-transmitter identity of nerve cord MNs.[28] The continuous expression of *ceh-13*, *mab-5*, and *egl-5* suggest that these Hox genes, such as *lin-39*, may also play critical roles in maintenance of adult MN identity. Recent findings suggest parallel roles for Hox genes in other species. In mice, *Hoxc8* is required to maintain spinal MN identity[73] and *Hoxa5* is necessary to maintain neuronal identity and connectivity in the brainstem.[74–76] In head MNs of the fly *D. melanogaster*, the Hox gene *Dfd* is required for maintenance of neuromuscular synapse function.[77] Altogether, our findings generate testable hypotheses on possible Hox gene functions in the adult nervous system, supporting an emerging theme from recent Hox studies in *C. elegans*, flies, and mice.[71,78–80]

Distinct codes of neuropeptides and receptors delineate each MN subclass

Extra-synaptic signaling (e.g., neuropeptides) is an ancient feature of animal nervous systems. Neurons secrete peptides into the extracellular environment, which then can activate G-protein-coupled receptors (GPCRs) (e.g., neuropeptide receptors) on other neurons. As in other animals, extra-synaptic signaling modulates many aspects of *C. elegans* development and behavior, including learning and memory,[81] olfaction,[82,83] and locomotion.[84–88] A fundamental challenge in neuropeptide biology is to deorphanize neuropeptide receptors, i.e., to identify neuropeptide and GPCR pairs.[89] A key step toward addressing this challenge is to build gene expression maps of neuropeptides and receptors with single-cell resolution.[45] Here, we provide a comprehensive expression map of neuropeptides and neuropeptide receptors in adult *C. elegans* MNs (Figures 4A and 4B).

There are over 300 predicted neuropeptide-encoding genes embedded in the *C. elegans* genome from three major families: insulin-like peptides (*ins*), FMRFamide-related peptides (*flp*), and non-insulin/non-FMRFamide-related neuropeptide-like proteins (*nlp*).[89,90] We focused on *flp*- and *nlp*-genes, which represent the most well-characterized neuropeptide families in *C. elegans*, totaling 113 genes.[90] Consistent with CeNGEN project findings on larval neuronal classes,[10] adult MN subclasses collectively express 67.3% (76 out of 113) of these neuropeptide genes, raising the hypothesis that adult *C. elegans* MNs exhibit significant neuropetidergic activity. Adult MN subclasses also express 55 of the 150 (36.7%) predicted *C. elegans* neuropeptide receptor genes,[90] suggesting that MNs also respond to neuropeptide signaling. Interestingly, *C. elegans* MNs exhibit a higher degree of autocrine signaling (i.e., co-expression of neuropeptide-GPCR pairs) relative to other neuronal types.[45] Genes related to *C. elegans* neuropeptides and neuropeptide receptors are also expressed in *Drosophila* and mouse MNs,[20,25,91] highlighting their potentially conserved role in the motor system.

Functional significance of adult MN diversity in *C. elegans*

What is the biological significance of the observed MN diversity in the adult? First, it is suggestive of a widespread network of "wireless," extra-synaptic signaling in the adult *C. elegans* VNC. Second, molecularly defined MN subclasses (VNC and ganglionic) display distinct patterns of connectivity. Such correlation may enable individual MN subclasses to receive distinct inputs. For example, the DA9 subclass, unlike any other DA subclass, receives input from two glutamatergic neurons (PHC, PVR) (Figure 5B; File S6). The glutamate ionotropic receptor-encoding gene *glr-4* (human GRIK) is expressed in DA9,[29] likely enabling receipt of such glutamatergic input. Third, subclass diversity may also enable MNs to functionally multitask. Cholinergic MNs in *C. elegans*, apart from stimulating muscle contraction, are known to perform additional functions: neurons of the DB and VB classes transduce proprioceptive signals during forward locomotion, whereas neurons of the DA and VA classes act as local oscillators for backward locomotion.[92,93] Hence, the observed adult MN diversity is consistent with the idea of "compression."[92] That is, essential circuit properties (muscle contraction, proprioception, motor rhythm generation) in a small locomotor network, such as the *C. elegans* motor circuit, are executed by, and thereby compressed into, one population of neurons, MNs. We propose that these important properties of a simple locomotory circuit may be achieved by diversifying MN classes into molecularly distinct subclasses.

Utility and accessibility of this molecular resource

1. We provide a wealth of additional molecular markers for specific subsets of adult VNC and ganglionic MNs in *C. elegans*, which should facilitate experimental access to neuronal populations that can now be investigated at the genetic (e.g., Cre/*loxP* system), molecular, and functional (e.g., GCaMP, optogenetics) levels.
2. Molecularly defined MN subclasses display distinct connectivity. The connectivity map of the adult *C. elegans* motor circuit is now coupled with a molecular atlas of its constituent MNs.

3. Our data can be accessed at WormBase (https://wormbase.org/) and http://celegans.spinalcordatlas.org, which also hosts scRNA-seq data for adult spinal MNs of the mouse and human, thus offering opportunities for cross-species molecular comparisons.

Limitations of this study

We provide a molecular blueprint of adult MNs in *C. elegans*, paving the way for future studies on molecular mechanisms that maintain MN identity and function. Although we captured all adult VNC and most ganglionic MNs in *C. elegans* (Figure 1C), our dataset does not include six ganglionic MNs (AS1, DD1, DD6, VD1, VD2, VD13). In addition, closely related MN clusters may have remained unresolved by our computational analysis. Thus, the true number of adult MN subclasses may be greater than 29. In addition, our gene expression datasets do not include non-coding RNAs. Alternative library preparation methods combined with different approaches, such as single-nucleus RNA-seq, could be used in the future to obtain a more complete molecular description of adult *C. elegans* MNs. Last, future studies (e.g., cell ablation, optogenetics) are needed to characterize these 29 molecularly distinct MN subclasses for potential functional roles in the motor circuit.

STAR★METHODS

RESOURCE AVAILABILITY

Lead contact—Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Paschalis Kratsios (pkratsios@uchicago.edu).

Materials availability—Some *C. elegans* strains generated in this study have been submitted to and are commercially available from the *Caenorhabditis* Genetics Center (CGC); otherwise, strains are available upon request.

Data and code availability

- Raw sequencing data have been deposited at GEO and are available under the accession code GSE234962. All data generated for or analyzed in this study are contained in the manuscript and supporting files. Microscopy data will be shared by the lead contact upon request.
- All original code has been deposited at the Zenodo repository under the https://doi.org/10.5281/zenodo.10567646.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Maintenance of *C. elegans* strains—All *C. elegans* strains were cultured at 20°C or 25°C on nematode growth media (NGM) seeded with standard *E. coli* OP50 as a food source. Embryonic, larval, and day 1 adult stage hermaphrodites were assessed as described

in the main text, figures, and figure legends. All strains used or generated for this study are listed in the key resources table.

METHOD DETAILS

Preparation of day 1 adult stage *C. elegans* and dissociation—Worms were grown on 8P nutrient agar 150 mm plates seeded with *E. coli* strain NA22. To obtain synchronized cultures of early adult worms, embryos obtained by hypochlorite treatment of adult hermaphrodites were incubated in M9 buffer at room temperature for 16 h and then grown on NA22-seeded plates for 54 h (*acr-2p::GFP* worms) or 62 h (*lin-39p::RFP* worms). The developmental age of each culture was determined by scoring vulval morphology (>90 worms). For *acr-2p::GFP*, 96.8% of the worms were young adults. Adult *acr-2p::GFP* worms were collected and separated from bacteria by washing twice with ice-cold M9 and centrifuging at 150 rcf for 2.5 min. Even after synchronization, the *lin-39p::RFP* strain showed variability in age, and therefore young adults were further isolated by size exclusion using a 35 μm nylon mesh. Worms were washed off plates with M9 then placed on a 35 μm nylon mesh suspended over M9 containing NA22 bacteria for 25 min. Larval worms, but not adults, are able to crawl through the mesh. Young adult worms remaining on the mesh were collected into M9 and spun at 150 rcf for 2.5 min. After the mesh isolation, 81.5% of the *lin-39p::RFP* worms were young adults and the remaining worms were in the late L4 stage.

Worms were transferred to a 1.6 mL centrifuge tube and pelleted at 16,000 rcf for 1 min. 250 μL pellets of packed worms were treated with 500 μL of SDS-DTT solution (20 mM HEPES, 0.25% SDS, 200 mM DTT, 3% sucrose, pH 8.0) for 6 min, and the reaction was quenched by adding 750 μL of egg buffer. Worms were pelleted at 16,000 rcf, then washed five additional times by diluting with 1 mL egg buffer and pelleting at 16,000 rcf for 30 s *acr-2p::GFP* worms were incubated in pronase (15 mg/mL, Sigma-Aldrich P8811) in egg buffer for 23 min at room temperature. *Lin-39p::RFP* worms were digested in cold-active protease (10 mg/mL, Protease from Bacillus licheniformis, Sigma-Aldrich P4860, with DNAse I 35 U/mL) in egg buffer for 35 min at 4°C. During the protease incubations, solutions were triturated by pipetting through a P1000 pipette tip for four sets of 80 repetitions. The status of dissociation was monitored under a fluorescence dissecting microscope at 5-min intervals. Digestions were stopped by adding 750 μL L-15 media supplemented with 10% fetal bovine serum (L-15–10), and cells were pelleted by centrifuging at 530 rcf for 5 min at 4°C. The pellet was resuspended in 1 mL L-15–10, and single-cells were separated from whole worms and debris by centrifuging at 100 rcf for 2 min at 4°C. The supernatant was then passed through a 35-μmmicron strainer cap into a 5 mL collection tube (Falcon 352235). The pellet was resuspended a second time in an additional 1 mL of L-15–10, spun at 100 rcf for 2 min at 4°C, and the resulting supernatant was added to the collection tube.

FACS isolation of *C. elegans* motor neurons for single-cell RNA-seq—Sorting of single-cell suspensions of both strains was performed on BD FACSAria III cell sorters equipped with 70-μmmicron diameter nozzles. DAPI was added to the samples (final concentration of 1 μg/mL) to label dead and dying cells. Non-fluorescent N2 standards were used to set gates to exclude auto-fluorescent cells. We set FACS gates to encompass

a wide range of fluorescent intensities to ensure capture of all targeted cell types. This approach may contribute to the presence of unlabeled neuronal and non-neuronal cells in our dataset (these non-targeted cells were excluded from our analysis of VNC MNs. See downstream processing). Cells were sorted for 3 h under the “4-way Purity” mask into 1.5 mL microcentrifuge tubes containing 200 μL L-15–33 (L-15 medium containing 33% fetal bovine serum), and subsequently concentrated by centrifugation at 1200 rcf for 12 min at 4°C. Fluorescent cells were counted on a hemocytometer. Single-cell suspensions used for 10x Genomics single-cell sequencing ranged from 180 to 350 cells/μL. Two samples from each strain (a total of four samples) were prepared for encapsulation using the 10X Genomics Chromium system. We estimate the time from harvesting worms off plates to encapsulation in the 10X Genomics Chromium Controller to be ~6 h (~1.5 h for dissociation, 3 h for sorting, 1–1.5 h for concentrating, counting and encapsulation). Once dissociated, cells were kept at 4°C for all subsequent procedures.

Single-cell RNA sequencing—Each sample (targeting 10,000 cells per sample) was processed for single cell 3’ RNA sequencing utilizing the 10X Chromium system, v3 chemistry. Libraries were prepared using P/N 1000075, 1000073, and 120262 following the manufacturer’s protocol. The libraries were sequenced using the Illumina NovaSeq 6000 with 150 bp paired end reads. Real-Time Analysis software (RTA, version 2.4.11; Illumina) was used for base calling.

Downstream processing—FASTQ files were run through the 10X Genomics Cell Ranger software (v6.0.2 for *acr-2p::GFP*, v6.1.1 for *lin-39p::RFP*) using a custom reference genome based on WS273 with 3′ UTR extensions for several genes,[10] available at GEO, accession number GSE234962. The Emptydrops function from the R package DropletUtils was used to determine droplets containing cells, using all droplets with fewer than 50 UMIs to construct the ambient RNA profile. We then corrected for background RNA content using the SoupX R package,[96] with a threshold of 25 UMIs for the background expression profile. The genes used to estimate contamination for each sample are found in File S1, column K. Background subtracted counts were rounded to the nearest integer, and the corrected gene-by-barcode matrices were used for further processing. Quality control metrics were calculated for each dataset with the scater package for R,[97] using the percentage of UMIs from the mitochondrial genes *nduo-1, nduo-2, nduo-3, nduo-4, nduo-5, nduo-6, ctc-1, ctc-2, ctc-3, ndfl-4, atp-6, ctb-1, MTCE.7 and MTCE.33*. Droplets with greater than 20% of UMIs coming from mitochondrial genes were removed. Each of the four samples was processed individually up to this point. The four samples were combined into one cell_data_set object in monocle3[9,98–100] for dimensionality reduction and clustering. For dimensionality reduction on the full dataset, 75 principle components were used, and 2-dimensional UMAPs were generated using parameters umap.min_dist: 0.3 and umap.n_neighbors: 75, since these parameters clearly separated most neuron classes in prior work.[10] All other parameters were used at default values.

Motor neuron class and subclass identification in *C. elegans*—We identified clusters containing VNC MNs by expression of known marker genes, including *unc-3, bnc-1, unc-4, unc-129, vab-7, unc-25,* and *unc-47.* A separate UMAP containing these cells

was generated using 40 principal components. Batch correction between strains was run using the align_cds function in monocle3, based on the batchelor R package,[101] with the alignment_k parameter set to 5 (Figure S1C). This correction was performed to remove possible strain effects, but only at the level of UMAP reduction. No changes were made to underlying gene expression data. UMAP dimensionality reduction was re-run with umap.min_dist (0.2) and umap.n_neighbors (50) to increase the separation of closely-related classes and subclasses. We assigned motor neuron class IDs to single-cell clusters based on expression of known class marker genes (File S1). We noticed that all the anatomical classes had multiple distinct groups of cells in UMAP space. Separate UMAPs were similarly created for the cells corresponding to each anatomical class, including GABAergic MNs (marked by expression of *unc-25, unc-30, unc-47* and *ttr-39*) and cholinergic MNs.

Generally, subclasses were annotated using a combination of separation in UMAP space and comparison to either previously known or newly generated expression patterns of cluster-enriched marker genes. To illustrate in detail how subclasses were identified, we will use the VA neuron subclasses as an example. Clusters containing VA MNs were identified in the UMAP of all VNC MNs by the co-expression of bnc-1 and unc-4 (Figure S2A), which are known to be co-expressed in only VA neurons among VNC MNs10,33,46,. To identify possible VA subclasses, we re-ran the dimensionality reduction and clustering on all the *bnc-1/unc-4+* clusters, resulting in the UMAP in Figure S2B. Plotting expression of *bnc-1* revealed that cluster 10 expressed very little bnc-1 (Figure S2C). We identified the top 75 enriched marker genes for each cluster using the top_marker function in monocle3. We selected the top 20 most selective marker genes for each cluster and compared them to the CeNGEN L4 neuronal expression by plotting them using the ‘Heatmaps of gene expression’ tab on the CeNGEN(cengen.shinyapps.io/CengenApp). The most specific genes for cluster 10, which included *itr-1* (Figure S3D), were most consistently expressed in the DA9 cluster from the CeNGEN L4 data (Figure S2E). The most selective genes for cluster 7, which included *cog-1* (Figure S2F), were most consistently expressed in VA12 in the CeNGEN L4 dataset (Figure S2G). Thus, we annotated cluster 10 as DA9 and cluster 7 as VA12.

To further resolve the VA neurons, we removed DA9 and re-ran dimensionality reduction and clustering (Figure S2H). Clustering yielded 9 clusters. Cluster 6 in this plot was VA12. We generated cluster-enriched marker genes and identified several candidate genes to allow for assigning anatomical identities to individual clusters. Among these was *elt-1,* a TF enriched in cluster 5 (Figure 3F). An endogenous *elt-1* reporter strain showed *elt-1* expression in only VA1, among VA neurons (Figure 3G). This same cluster had very low levels of any *C. elegans* HOX gene. This pattern of expression is most consistent with an identify of VA1 for cluster 5 (Figure 3B). The expression of the homeodomain TF *vab-3* was restricted to cluster 8 (Figure S2I), and it was recently shown with an endogenous marker that among VA neurons, *vab-3* is only expressed in VA11 (Figure S2I). Cluster 8 also strongly expresses *mab-5*, with very little *lin-39*, consistent with VA11 (Figure 3D). Cluster 7 and part of cluster 1 express C24H12.1 (Figure S2J). Single-molecule FISH of C24H12.1 in L2 worms revealed that this gene is primarily expressed in VA2 (Figure S2K–S2M). The C24H12.1-expressing clusters also express *ceh-13*, very little *lin-39,* and no *mab-5,* or *egl-5* (Figure 3D). According to the endogenous expression patterns of HOX genes that we describe in the work, this combination strongly suggests cluster 7 is VA2.

Cluster 3 expresses *mab*-5 and *ceh-13*, with very little *lin-39* and no *egl-5*. By comparison to the endogenous HOX expression and process of elimination vs. the other VA clusters, we annotated cluster 3 as VA9–10. Clusters 1,2,4, and 9 shared similar HOX expression (strong *ceh-13+/lin-39+,* Figure 3D) and no clear marker genes strongly distinguished these clusters from each other. Therefore, we annotated these clusters as VA3–8.

As an additional example, Figure S3 illustrates the annotation of DB subclasses. File S1 contains the genes used to annotate subclasses for each of the anatomical classes in this work, and Figure S4 contains both heatmap and dotplot representations of the expression of the genes used to assign anatomical identity to all the subclasses in the dataset.

Differential expression in *C. elegans* motor neurons—Differential expression analyses between subclasses within each MN subclass and between L4 and adult data were performed using the FindMarkers function in Seurat v4.1.1,[102] using default parameters. L4 CeNGEN scRNA-seq data were re-analyzed and re-annotated using monocle3.

Thresholding—We detected low levels of unexpected gene expression in some subclasses that we anticipated could be due to noise and/or uncorrected contamination from ambient RNA (e.g., expression of the cholinergic marker *unc-17* in GABAergic neurons, Figures S4A and S4B). To make an informed decision regarding whether a gene is truly expressed in each neuron class or subclass, we used a dynamic thresholding procedure,[10] as described below. All of the code for this procedure has been deposited at Zenodo under the https://doi.org/10.5281/zenodo.10567646. Importantly, this thresholding was performed after cell annotation. We used a ground truth dataset of 169 genes with expression patterns across the nervous system previously determined with high confidence fosmid fluorescent reporters, CRISPR strains or other methods (File S3).[30,31,94,103–105] To include as much of the ground truth data and have the most accurate thresholding, we included the 48 neuronal cell types in the single cell data which contained >20 cells (File S3). We calculated two forms of aggregate expression data for each cell subtype; 1) normalized expression values and 2) the percent of cells in which a gene is detected (with a UMI >0). We used the percent of cells as the thresholding metric.

We first set initial thresholds to retain ubiquitously expressed genes and to remove non-neuronal genes. Genes detected in ≥ 1% of the cells in every neuron cluster were considered expressed in all neuron types (383 genes), whereas transcripts detected in ≤ 2% of the cells in every neuron cluster were considered not expressed at all (1885 genes; no genes were detected in ≥ 1% and ≤ 2% of the cells in every neuron. As most genes displayed varying levels of expression, we found that a single threshold failed to reliably capture expression for all genes. To illustrate, for *unc-17/VAChT* the VC4–5 cluster had the highest proportion of cells expressing (91%), whereas for the HOX gene *ceh-13,* the cluster with the highest proportion of cells (AS2–3) had only 55.1% of cells expressing. Thus, we applied percentile thresholding for each gene individually. Thresholds were calculated as a fraction of the highest proportion of cells for each individual gene. For example, a threshold of 0.06 (0.06 * the highest proportion of cells for each gene) results in different absolute cut-offs for each gene. For example, for *unc-17*/VAChT, with a maximal proportion of 91%, the 0.06 percentile threshold was 5.46% of cells (0.06 * 91 = 5.46). For *ceh*-13, with a maximal

proportion of 55.1, the 0.06 percentile threshold was 3.306% of cells (0.06 * 55.1 = 3.306). For each threshold percentile, we generated receiver operating characteristic (ROC) and precision-recall (PR) curves by comparing to the ground truth. Wegenerated 5,000 stratified bootstraps of the ground truth genes using the R package boot (Canty and Ripley, 2019; Davison and Hinkley, 1997) and computed the True Positive Rate (TPR), False Positive Rate (FPR) and False Discovery Rate (FDR) for the entire dataset in relation to ground truth expression. We estimated 95% confidence intervals with the adjusted percentile (BCa) method. Finally, we selected a balanced threshold (0.06 * the highest proportion of cells for each gene) to use for analyses profiling gene expression across all neuron types and across gene families. This threshold yielded a TPR of 0.858, an FPR of 0.0935 and an FDR of 0.141. If the expression of a gene in a certain cell type was below the threshold, we set that expression to zero. We retained the continuous normalized expression values for all instances of gene expression above the threshold. A full list of threshold gene expression among neuron classes with greater than 20 cells is contained in File S3.

Microscopy—Larval and adult stage animals were anesthetized with sodium azide (NaN_3, 100 mM) and mounted on a 4% agarose pad on imaging slides. Images were recorded with an automated fluorescence microscope (Zeiss, Axio Imager Z2). Z stacks (0.50–1.0 μm step size) were acquired using a Zeiss Axiocam 503 mono (ZEN Blue software, version 2.3.69.1000), a Zeiss LSM 880, or a Zeiss LSM 900, or a Leica SP8 laser scanning confocal microscope (Leica Application Suite [LAS] X confocal software, version 3.5.7.23225). Representative images shown are max intensity z-projections. All image reconstruction was performed in Fiji (version 2.9.0/1.53t). Fluorescent overlays and cropping for figures were generated using Adobe Photoshop 2022 (23.2.2 Release). Images of strains containing the fluorescent Neuronal Polychromatic Atlas of Landmarks (NeuroPAL) transgene[94] for cell identification were acquired on a Zeiss LSM880 confocal microscope with 0.5 μm step size.

Single molecule RNA FISH—smFISH was performed with custom C24H12.1 and *lin-39* probes linked to Quasar 670 (Biosearch Technologies). Synchronized L2 larvae were washed from plates with M9, fixed with 4% paraformaldehyde in 1X PBS for 35 min, then permeabilized in 70% ethanol for 54. Hybridization followed manufacturer's instructions (http://www.biosearchtech.com/stellarisprotocols) and was performed at 37C for 16h in Stellaris RNA FISH hybridization buffer (Biosearch Technologies Cat# SMF-HB1–10) containing C24H12.1 probe at 1:100. VA and VB neurons were marked with *otIs707 [bnc-1::GFP]* and cell nuclei were stained with DAPI. Z-stacks capturing the complete VA cell body were collected in 0.2 mm steps using the 405nm, 488nm, and 647nm lasers on a Nikon spinning disk confocal microscope. Images were 3D-deconvolved in NIS elements. smFISH puncta were manually counted if they corresponded to circular fluorescent spots that exceeded the Quasar 670 background signal and were located within a GFP-labeled VA cell body. At least 17 worms were scored per genotype.

Identification of motor neuron classes and subclasses in C. elegans—Motor neuron classes and subclasses were identified based on the following criteria: (1) co-localization with or exclusion from cells which express reporters with previously characterized expression patterns[106]; Anatomically invariant positioning of neuronal bodies

in the retrovesicular ganglion (RVG), ventral nerve cord (VNC), and preanal ganglion (PAG); (3) class-specific birth order (i.e., embryonic versus post-embryonic emergence); (4) total cell numbers in each motor neuron class or subclass; (5) co-expression with the NeuroPAL transgene.[94]

Analysis of transcriptomes from adult mouse motor neurons—Single-nucleus motor neuron data was downloaded from spinalcordatlas.org and analyzed using Seurat (v4.1.1). The data were sub-set to only include visceral and skeletal motor neuron clusters generated by the Gitler lab[25]; these annotations were present in the metadata of the downloaded file. As the subtype classificiations were not present in the downloaded file, we re-ran dimensionality reduction on subsets of both visceral and skeletal motor neurons independently to recreate the subclasses recreate the subclasses as in the original publication[25] as follows. For the visceral subset, 30 principal components were calculated and used for UMAP reduction (additional parameters: seed.use: 5, n.neighbors: 40L). The Seurat FindNeighbors function was used with 30 pcs, and FindClusters was run with resolution (0.5). This approach identified 18 clusters. Three pairs of clusters were manually merged to match the description of subclusters as shown in figure 2e of Blum et al. 2021.[25] Specifically, cluster 17 was merged with cluster 3, cluster 16 was merged with cluster 11, and cluster 12 was merged with cluster 4. The following marker genes were used to verify the annotation of subclasses: Trhr, Nts, Gm10248, Chodl, Mdga1, Frem1, Slc26a7, Sst, Bnc2, Fras1, Gpc3, Slc16a12, Gas7, Lypd6, Tll2, Col24a1, Etv1, Lgr5, Mamdc2, Gm20754, Gm26911, Gm43948, Col12a1, Fn1, Mpped2, Dscaml1, Fstl4, Pde11a, Sox5, Cdh20, Tnc.

For the skeletal subset, 20 principal components were used for UMAP reduction, and FindNeighbors. FindClusters was run with resolution (0.5). This revealed clear populations matching the γ and γ* MNs, as characterized by expression of marker genes for γ MNs (Creb5, Klhl1, Pard3b, Sema3c, Nrp2) and for γ* MNs (Crtac1, Kitl, Stxbp6, Plch1). The remaining α skeletal MNs were subset and processed as follows (RunUMAP and FindNeighbors with 20 principal components, FindClusters with resolution: 0.5). These set-tings generated 14 clusters, 2 pairs of which were merged to best match a comparison with figure 4c in Blum et al., based on distinct expression patterns of the following marker genes (Rreb1, Sdk2, Htr1f, Pdgfd, Ptchd4, Etl4, Grm5, Cacna1e, Kcnh7, Cpne4, Grik1, Hcrtr2, Megf11, Kcnq5, Chodl, Erbb4, Gm2164, Cdh9, Gab1, Prom1, Clca3a1, Cdh8, A330008L17Rik, Sema3e, Ppp1r1c, Mpped2, Shisa9, Hmcn1, Slc44a5). Specifically, cluster 13 was merged with cluster 10, and cluster 8 was merged with cluster 0. Together, this gave a dataset with 29 subtypes of mouse spinal cord MNs (15 visceral MNs, denoted v0 – v14, 2 γ skeletal MNs, denoted γ and γ*, and 12 α skeletal MNs, denoted α0– α11).

Generation of mouse HD and NHR gene lists—The list of HD TF genes in mouse was derived from the Animal TFDB database (version 3.0) (http://bioinfo.life.hust.edu.cn/AnimalTFDB/).[107] By filtering for species *Mus musculus*, we obtained a list of 240 genes annotated to encode HD TF proteins. We assessed all 240 of these genes for expression in the adult mouse MN dataset. The list of 119 NHR TF encoding genes in mouse was derived from the Jackson Laboratory Mouse Genome Informatics tool (https://www.informatics.jax.org/) via a search for genes associated with the *nuclear*

receptor activity (GO: 0004879) GO term in the Gene Ontology Browser (https://www.informatics.jax.org/vocab/gene_ontology/GO:0004879).

Generation of clustering dendrograms based on mouse HD and NHR gene expression—To probe the expression of homeodomain (HD) and nuclear hormone receptor (NHR) TFs, we required a given gene to be detected in >5% of the nuclei in at least one of the 29 mouse spinal cord MN subtypes. We identified 76 HD and 43 NHR TFs that met this criterion. We averaged expression of these genes across the nuclei within each cell type, and clustered the averaged expression matrices of HD and NHR expression independently using the R package hclust (using the Euclidean distance and ward.D2 method). We generated clustering dendrograms using the R package hclust and dot plots using Seurat.

QUANTIFICATION AND STATISTICAL ANALYSIS

Details of sample size, center, and dispersion are contained in R analysis scripts, figure legends, and in the relevant STAR Methods sections above. All statistical calculations were performed in R (v.4.2.0). The Wilcoxon rank-sum test in Seurat was used for calculating differential expression, with n representing the number of cells corresponding to an anatomical cell type or subtype. The resulting p values were adjusted using the Bonferroni correction, and only genes with an adjusted p value <0.05 were considered significant. For microscopy data in Figures 1B, 3G, 4C, S2K, S2L, S5A–S5E, S5G, S5H, S6A–S6H, n represents the number of animals assessed for reporter expression and is provided in the figure (Figures S5A–H and S6A–S6H) and in all related figure legends.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

Strains were provided by the *Caenorhabditis* Genetics Center (CGC), funded by NIH grant P40 OD010440). We thank members of the Kratsios lab and O. Hobert for input on the manuscript. We thank the UC Integrated Light Microscopy Core, funded by NCI grant P30CA014599, and the Vanderbilt Flow Cytometry Shared Resource, supported by Ingram Cancer Center (P30 CA68485), DDRC (DK058404). For scRNA-seq, we thank the VANTAGE Core Facility, supported by CTSA (5UL1 RR024975-03), the Ingram Cancer Center (P30 CA68485), the Vision Center (P30 EY08126), and NIH/NCRR (G20 RR030956). We also thank the Vanderbilt Cell Imaging Shared Resource (NIH CA68485, DL20593, DK58404, DK59637, EY08126) for access to facilities. This study was funded by NIH grants to D.M.M. (R01NS113559, R01NS100547, R01NS106951) and P.K. (R21 NS108505, R01 NS118078, R01 NS116365).

REFERENCES

1. Dalla Torre di Sanguinetto SA, Dasen JS, and Arber S. (2008). Transcriptional mechanisms controlling motor neuron diversity and connectivity. Curr. Opin. Neurobiol. 18, 36–43. 10.1016/j.conb.2008.04.002. [PubMed: 18524570]
2. Dasen JS (2009). Transcriptional networks in the early development of sensory-motor circuits. Curr. Top. Dev. Biol. 87, 119–148. 10.1016/S0070-2153(09)01204-6. [PubMed: 19427518]
3. Lee SK, and Pfaff SL (2001). Transcriptional networks regulating neuronal identity in the developing spinal cord. Nat. Neurosci. 4, 1183–1191. 10.1038/nn750nn750. [PubMed: 11687828]

4. Catela C, and Kratsios P. (2021). Transcriptional mechanisms of motor neuron development in vertebrates and invertebrates. Dev. Biol. 475, 193–204. 10.1016/j.ydbio.2019.08.022. [PubMed: 31479648]
5. Hobert O, Glenwinkel L, and White J. (2016). Revisiting Neuronal Cell Type Classification in Caenorhabditis elegans. Curr. Biol. 26, R1197–R1203. 10.1016/j.cub.2016.10.027. [PubMed: 27875702]
6. Stifani N. (2014). Motor neurons and the generation of spinal motor neuron diversity. Front. Cell. Neurosci. 8, 293. 10.3389/fncel.2014.00293. [PubMed: 25346659]
7. Holguera I, and Desplan C. (2018). Neuronal specification in space and time. Science 362, 176–180. 10.1126/science.aas9435. [PubMed: 30309944]
8. Mu Q, Chen Y, and Wang J. (2019). J. Deciphering Brain Complexity Using Single-cell Sequencing. Dev. Reprod. Biol. 17, 344–366. 10.1016/j.gpb.2018.07.007.
9. Cao J, Packer JS, Ramani V, Cusanovich DA, Huynh C, Daza R, Qiu X, Lee C, Furlan SN, Steemers FJ, et al. (2017). Comprehensive single-cell transcriptional profiling of a multicellular organism. Science 357, 661–667. 10.1126/science.aam8940. [PubMed: 28818938]
10. Taylor SR, Santpere G, Weinreb A, Barrett A, Reilly MB, Xu C, Varol E, Oikonomou P, Glenwinkel L, McWhirter R, et al. (2021). Molecular topography of an entire nervous system. Cell 184, 4329–4347.e23. 10.1016/j.cell.2021.06.023. [PubMed: 34237253]
11. Ghaddar A, Armingol E, Huynh C, Gevirtzman L, Lewis NE, Waterston R, and O'Rourke EJ (2023). Whole-body gene expression atlas of an adult metazoan. Sci. Adv. 9, eadg0506. 10.1126/sciadv.adg0506.
12. Roux AE, Yuan H, Podshivalova K, Hendrickson D, Kerr R, Kenyon C, and Kelley D. (2023). Individual cell types in C. elegans age differently and activate distinct cell-protective responses. Cell Rep. 42, 112902. 10.1016/j.celrep.2023.112902.
13. Seroka A, Lai SL, and Doe CQ (2022). Transcriptional profiling from whole embryos to single neuroblast lineages in Drosophila. Dev. Biol. 489, 21–33. 10.1016/j.ydbio.2022.05.018. [PubMed: 35660371]
14. Velten J, Gao X, Van Nierop Y Sanchez P, Domsch K, Agarwal R, Bognar L, Paulsen M, Velten L, and Lohmann I. (2022). Single-cell RNA sequencing of motoneurons identifies regulators of synaptic wiring in Drosophila embryos. Mol. Syst. Biol. 18, e10255. 10.15252/msb.202110255. [PubMed: 35225419]
15. Farnsworth DR, Saunders LM, and Miller AC (2020). A single-cell transcriptome atlas for zebrafish development. Dev. Biol. 459, 100–108. 10.1016/j.ydbio.2019.11.008. [PubMed: 31782996]
16. Scott K, O'Rourke R, Winkler CC, Kearns CA, and Appel B. (2021). Temporal single-cell transcriptomes of zebrafish spinal cord pMN progenitors reveal distinct neuronal and glial progenitor populations. Dev. Biol. 479, 37–50. 10.1016/j.ydbio.2021.07.010. [PubMed: 34303700]
17. Tambalo M, Mitter R, and Wilkinson DG (2020). A single cell transcriptome atlas of the developing zebrafish hindbrain. Development 147, dev184143. 10.1242/dev.184143.
18. Xing L, Chai R, Wang J, Lin J, Li H, Wang Y, Lai B, Sun J, and Chen G. (2022). Expression of myelin transcription factor 1 and lamin B receptor mediate neural progenitor fate transition in the zebrafish spinal cord pMN domain. J. Biol. Chem. 298, 102452. 10.1016/j.jbc.2022.102452.
19. Delile J, Rayon T, Melchionda M, Edwards A, Briscoe J, and Sagner A. (2019). Single cell transcriptomics reveals spatial and temporal dynamics of gene expression in the developing mouse spinal cord. Development 146, dev173807. 10.1242/dev.173807.
20. Liau ES, Jin S, Chen YC, Liu WS, Calon M, Nedelec S, Nie Q, and Chen JA (2023). Single-cell transcriptomic analysis reveals diversity within mammalian spinal motor neurons. Nat. Commun. 14, 46. 10.1038/s41467-022-35574-x(2023. [PubMed: 36596814]
21. Rosenberg AB, Roco CM, Muscat RA, Kuchina A, Sample P,Yao Z, Graybuck LT, Peeler DJ, Mukherjee S, Chen W, et al. (2018). Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. Science 360, 176–182. 10.1126/science.aam8999. [PubMed: 29545511]
22. Blum JA, and Gitler AD (2022). Singling out motor neurons in the age of single-cell transcriptomics. Trends Genet. 38, 904–919. 10.1016/j.tig.2022.03.016. [PubMed: 35487823]

23. Ragagnin AMG, Shadfar S, Vidal M, Jamali MS, and Atkin JD (2019). Motor Neuron Susceptibility in ALS/FTD. Front. Neurosci. 13, 532. 10.3389/fnins.2019.00532. [PubMed: 31316328]
24. Alkaslasi MR, Piccus ZE, Hareendran S, Silberberg H, Chen L, Zhang Y, Petros TJ, and Le Pichon CE (2021). Single nucleus RNA-sequencing defines unexpected diversity of cholinergic neuron types in the adult mouse spinal cord. Nat. Commun. 12, 2471. 10.1038/s41467-021-22691-2(2021. [PubMed: 33931636]
25. Blum JA, Klemm S, Shadrach JL, Guttenplan KA, Nakayama L, Kathiria A, Hoang PT, Gautier O, Kaltschmidt JA, Greenleaf WJ, and Gitler AD (2021). Single-cell transcriptomic analysis of the adult mouse spinal cord reveals molecular diversity of autonomic and skeletal motor neurons. Nat. Neurosci. 24, 572–583. 10.1038/s41593-020-00795-0(2021. [PubMed: 33589834]
26. Philippidou P, and Dasen JS (2013). Hox genes: choreographers in neural development, architects of circuit organization. Neuron 80, 12–34. 10.1016/j.neuron.2013.09.020. [PubMed: 24094100]
27. Thor S, and Thomas JB (2002). Motor neuron specification in worms, flies and mice: conserved and 'lost' mechanisms. Curr. Opin. Genet. Dev. 12, 558–564. [PubMed: 12200161]
28. Feng W, Destain H, Smith JJ, and Kratsios P. (2022). Maintenance of neurotransmitter identity by Hox proteins through a homeostatic mechanism. Nat. Commun. 13, 6097. 10.1038/s41467-022-33781-0(2022. [PubMed: 36243871]
29. Kratsios P, Kerk SY, Catela C, Liang J, Vidal B, Bayer EA, Feng W, De La Cruz ED, Croci L, Consalez GG, et al. (2017). An intersectional gene regulatory strategy defines subclass diversity of C. elegans motor neurons. Elife 6, e25751. 10.7554/eLife.25751. [PubMed: 28677525]
30. Reilly MB, Cros C, Varol E, Yemini E, and Hobert O. (2020). Unique homeobox codes delineate all the neuron classes of C. elegans. Nature 584, 595–601. 10.1038/s41586-020-2618-9. [PubMed: 32814896]
31. Reilly MB, Tekieli T, Cros C, Aguilar GR, Lao J, Toker IA, Vidal B, Leyva-Díaz E, Bhattacharya A, Cook SJ, et al. (2022). Widespread employment of conserved C. elegans homeobox genes in neuronal identity specification. PLoS Genet. 18, e1010372. 10.1371/journal.pgen.1010372.
32. Von Stetina SE, Treinin M, and Miller DM 3rd. (2006). The motor circuit. Int. Rev. Neurobiol. 69, 125–167. 10.1016/S0074-7742(05)69005-8. [PubMed: 16492464]
33. White JG, Southgate E, Thomson JN, and Brenner S. (1986). The structure of the nervous system of the nematode Caenorhabditis elegans. Philos. Trans. R. Soc. Lond. B Biol. Sci. 314, 1–340. [PubMed: 22462104]
34. Spencer WC, McWhirter R, Miller T, Strasbourger P, Thompson O, Hillier LW, Waterston RH, and Miller DM 3rd. (2014). Isolation of specific neurons from C. elegans larvae for gene expression profiling. PLoS One 9, e112102. 10.1371/journal.pone.0112102.
35. Stern S, Kirst C, and Bargmann CI (2017). Neuromodulatory Control of Long-Term Behavioral Patterns and Individuality across Development. Cell 171, 1649–1662.e10. 10.1016/j.cell.2017.10.041. [PubMed: 29198526]
36. Sun H, and Hobert O. (2021). Temporal transitions in the post-mitotic nervous system of Caenorhabditis elegans. Nature 600, 93–99. 10.1038/s41586-021-04071-4(2021. [PubMed: 34759317]
37. Higgins DP, Weisman CM, Lui DS, D'Agostino FA, and Walker AK (2022). Defining characteristics and conservation of poorly annotated genes in Caenorhabditis elegans using WormCat 2.0. Genetics 221, iyac085. 10.1093/genetics/iyac085.
38. Krumlauf R, Marshall H, Studer M, Nonchev S, Sham MH, and Lumsden A. (1993). Hox homeobox genes and regionalisation of the nervous system. J. Neurobiol. 24, 1328–1340. 10.1002/neu.480241006. [PubMed: 7901322]
39. Mallo M, Wellik DM, and Deschamps J. (2010). Hox genes and regional patterning of the vertebrate body plan. Dev. Biol. 344, 7–15. 10.1016/j.ydbio.2010.04.024. [PubMed: 20435029]
40. Van Auken K, Weaver DC, Edgar LG, and Wood WB (2000). Caenorhabditis elegans embryonic axial patterning requires two recently discovered posterior-group Hox genes. Proc. Natl. Acad. Sci. USA 97, 4499–4503. 10.1073/pnas.97.9.4499. [PubMed: 10781051]
41. Murray JI, Preston E, Crawford JP, Rumley JD, Amom P, Anderson BD, Sivaramakrishnan P, Patel SD, Bennett BA, Lavon TD, et al. (2022). The anterior Hox gene ceh-13 and elt-1/GATA activate

the posterior Hox genes nob-1 and php-3 to specify posterior lineages in the C. elegans embryo. PLoS Genet. 18, e1010187. 10.1371/journal.pgen.1010187.

42. Feng W, Li Y, Dao P, Aburas J, Islam P, Elbaz B, Kolarzyk A, Brown AE, and Kratsios P. (2020). A terminal selector prevents a Hox transcriptional switch to safeguard motor neuron identity throughout life. Elife 9, e50065. 10.7554/eLife.50065. [PubMed: 31902393]
43. Duerr JS, Gaskin J, and Rand JB (2001). Identified neurons in C. elegans coexpress vesicular transporters for acetylcholine and monoamines. Am. J. Physiol. Cell Physiol. 280, C1616–C1622. 10.1152/ajpcell.2001.280.6.C1616. [PubMed: 11350757]
44. Pereira L, Kratsios P, Serrano-Saiz E, Sheftel H, Mayo AE, Hall DH, White JG, LeBoeuf B, Garcia LR, Alon U, and Hobert O. (2015). A cellular and regulatory map of the cholinergic nervous system of C. elegans. Elife 4, e12432. 10.7554/eLife.12432. [PubMed: 26705699]
45. Ripoll-Sánchez L, Watteyne J, Sun H, Fernandez R, Taylor SR, Weinreb A, Bentley BL, Hammarlund M, Miller DM 3rd, Hobert O, et al. (2023). The neuropeptidergic connectome of C. elegans. Neuron 111, 3570–3589.e5. 10.1016/j.neuron.2023.09.043. [PubMed: 37935195]
46. Fox RM, Von Stetina SE, Barlow SJ, Shaffer C, Olszewski KL, Moore JH, Dupuy D, Vidal M, and Miller DM 3rd. (2005). A gene expression fingerprint of C. elegans embryonic motor neurons. BMC Genom. 6, 42. 10.1186/1471-2164-6-42.
47. Cook SJ, Jarrell TA, Brittin CA, Wang Y, Bloniarz AE, Yakovlev MA, Nguyen KCQ, Tang LTH, Bayer EA, Duerr JS, et al. (2019). Whole-animal connectomes of both Caenorhabditis elegans sexes. Nature 571, 63–71. 10.1038/s41586-019-1352-7. [PubMed: 31270481]
48. Varshney LR, Chen BL, Paniagua E, Hall DH, and Chklovskii DB (2011). Structural properties of the Caenorhabditis elegans neuronal network. PLoS Comput. Biol. 7, e1001066. 10.1371/journal.pcbi.1001066.
49. Espinosa-Medina I, Saha O, Boismoreau F, Chettouh Z, Rossi F, Richardson WD, and Brunet JF (2016). The sacral autonomic outflow is sympathetic. Science 354, 893–897. 10.1126/science.aah5454. [PubMed: 27856909]
50. Holland PWH, Booth HAF, and Bruford EA (2007). Classification and nomenclature of all human homeobox genes. BMC Biol. 5, 47. 10.1186/1741-7007-5-47. [PubMed: 17963489]
51. Sharma K, Sheng HZ, Lettieri K, Li H, Karavanov A, Potter S, Westphal H, and Pfaff SL (1998). LIM homeodomain factors Lhx3 and Lhx4 assign subtype identities for motor neurons. Cell 95, 817–828. [PubMed: 9865699]
52. Thaler JP, Koo SJ, Kania A, Lettieri K, Andrews S, Cox C, Jessell TM, and Pfaff SL (2004). A postmitotic role for Isl-class LIM homeodomain proteins in the assignment of visceral spinal motor neuron identity. Neuron 41, 337–350. 10.1016/s0896-6273(04)00011-x. [PubMed: 14766174]
53. Thaler JP, Lee SK, Jurata LW, Gill GN, and Pfaff SL (2002). LIM factor Lhx3 contributes to the specification of motor neuron and interneuron identity through cell-type-specific protein-protein interactions. Cell 110, 237–249. [PubMed: 12150931]
54. Antebi A. (2006). Nuclear Hormone Receptors in C. elegans. WormBook 3, 1–13. 10.1895/wormbook.1.64.1.
55. Parry LJ, McGuane JT, Gehring HM, Kostic IGT, and Siebel AL (2005). Mechanisms of relaxin action in the reproductive tract: studies in the relaxin-deficient (Rlx−/−) mouse. Ann. N. Y. Acad. Sci. 1041, 91–103. 10.1196/annals.1282.013. [PubMed: 15956691]
56. Yoshimura M, Polosa C, and Nishi S. (1987). Slow EPSP and the depolarizing action of noradrenaline on sympathetic preganglionic neurons. Brain Res. 414, 138–142. 10.1016/0006-8993(87)91334-5. [PubMed: 2887238]
57. Patel T, Hammelman J, Aziz S, Jang S, Closser M, Michaels TL, Blum JA, Gifford DK, and Wichterle H. (2022). Transcriptional dynamics of murine motor neuron maturation in vivo and in vitro. Nat. Commun. 13, 5427. 10.1038/s41467-022-33022-4(2022. [PubMed: 36109497]
58. Duggan A, Ma C, and Chalfie M. (1998). Regulation of touch receptor differentiation by the Caenorhabditis elegans mec-3 and unc-86 genes. Development 125, 4107–4119. 10.1242/dev.125.20.4107. [PubMed: 9735371]

59. Zheng C, Diaz-Cuadros M, and Chalfie M. (2015). Hox Genes Promote Neuronal Subtype Diversification through Posterior Induction in Caenorhabditis elegans. Neuron 88, 514–527. 10.1016/j.neuron.2015.09.049. [PubMed: 26539892]
60. Zheng C, Jin FQ, and Chalfie M. (2015). Hox Proteins Act as Transcriptional Guarantors to Ensure Terminal Differentiation. Cell Rep. 13, 1343–1352. 10.1016/j.celrep.2015.10.044. [PubMed: 26547238]
61. Zhang F, Bhattacharya A, Nelson JC, Abe N, Gordon P, Lloret-Fernandez C, Maicas M, Flames N, Mann RS, Colón-Ramos DA, and Hobert O. (2014). The LIM and POU homeobox genes ttx-3 and unc-86 act as terminal selectors in distinct cholinergic and serotonergic neuron types. Development 141, 422–435. 10.1242/dev.099721. [PubMed: 24353061]
62. Zheng C, Lee HMT, and Pham K. (2022). Nervous system-wide analysis of Hox regulation of terminal neuronal fate specification in Caenorhabditis elegans. PLoS Genet. 18, e1010092. 10.1371/journal.pgen.1010092.
63. Destain H, Prahlad M, and Kratsios P. (2024). Maintenance of neuronal identity in C. elegans and beyond: Lessons from transcription and chromatin factors. Semin. Cell Dev. Biol. 154, 35–47. 10.1016/j.semcdb.2023.07.001. [PubMed: 37438210]
64. Li Y, Smith JJ, Marques F, Osuma A, Huang HC, and Kratsios P. (2023). Cell context-dependent CFI-1/ARID3 functions control neuronal terminal differentiation. Cell Rep. 42, 112220. 10.1016/j.celrep.2023.112220.
65. Li Y, Osuma A, Correa E, Okebalama MA, Dao P, Gaylord O, Aburas J, Islam P, Brown AE, and Kratsios P. (2020). Establishment and maintenance of motor neuron identity via temporal modularity in terminal selector function. Elife 9, e59464. 10.7554/eLife.59464. [PubMed: 33001031]
66. Kerk SY, 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. 10.1016/j.neuron.2016.11.036. [PubMed: 28056346]
67. Estacio-Gómez A, and Díaz-Benjumea FJ (2014). Roles of Hox genes in the patterning of the central nervous system of Drosophila. Fly (Austin) 8, 26–32. 10.4161/fly.27424. [PubMed: 24406332]
68. Krumlauf R. (2016). Hox Genes and the Hindbrain: A Study in Segments. Curr. Top. Dev. Biol. 116, 581–596. 10.1016/bs.ctdb.2015.12.011. [PubMed: 26970643]
69. Lawrence PA, and Morata G. (1994). Homeobox genes: their function in Drosophila segmentation and pattern formation. Cell 78, 181–189. 10.1016/0092-8674(94)90289-5. [PubMed: 7913879]
70. Parker HJ, and Krumlauf R. (2020). A Hox gene regulatory network for hindbrain segmentation. Curr. Top. Dev. Biol. 139, 169–203. 10.1016/bs.ctdb.2020.03.001. [PubMed: 32450960]
71. Feng W, Li Y, and Kratsios P. (2021). Emerging Roles for Hox Proteins in the Last Steps of Neuronal Development in Worms, Flies, and Mice. Front. Neurosci. 15, 801791. 10.3389/fnins.2021.801791.
72. Dasen JS, and Jessell TM (2009). Hox networks and the origins of motor neuron diversity. Curr. Top. Dev. Biol. 88, 169–200. 10.1016/S0070-2153(09)88006-X. [PubMed: 19651305]
73. Catela C, Chen Y, Weng Y, Wen K, and Kratsios P. (2022). Control of spinal motor neuron terminal differentiation through sustained Hoxc8 gene activity. Elife 11, e70766. 10.7554/eLife.70766. [PubMed: 35315772]
74. Lizen B, Hutlet B, Bissen D, Sauvegarde D, Hermant M, Ahn MT, and Gofflot F. (2017). HOXA5 localization in postnatal and adult mouse brain is suggestive of regulatory roles in postmitotic neurons. J. Comp. Neurol. 525, 1155–1175. 10.1002/cne.24123. [PubMed: 27650319]
75. Lizen B, Moens C, Mouheiche J, Sacré T, Ahn MT, Jeannotte L, Salti A, and Gofflot F. (2017). Conditional Loss of Hoxa5 Function Early after Birth Impacts on Expression of Genes with Synaptic Function. Front. Mol. Neurosci. 10, 369. 10.3389/fnmol.2017.00369. [PubMed: 29187810]
76. Maheshwari U, Kraus D, Vilain N, Holwerda SJB, Cankovic V, Maiorano NA, Kohler H, Satoh D, Sigrist M, Arber S, et al. (2020). Postmitotic Hoxa5 Expression Specifies Pontine Neuron Positional Identity and Input Connectivity of Cortical Afferent Subsets. Cell Rep. 31, 107767. 10.1016/j.celrep.2020.107767.

77. Friedrich J, Sorge S, Bujupi F, Eichenlaub MP, Schulz NG, Wittbrodt J, and Lohmann I. (2016). Hox Function Is Required for the Development and Maintenance of the Drosophila Feeding Motor Unit. Cell Rep. 14, 850–860. 10.1016/j.celrep.2015.12.077. [PubMed: 26776518]

78. Allen AM, Neville MC, Birtles S, Croset V, Treiber CD, Waddell S, and Goodwin SF (2020). A single-cell transcriptomic atlas of the adult Drosophila ventral nerve cord. Elife 9, e54074. 10.7554/eLife.54074. [PubMed: 32314735]

79. Hutlet B, Theys N, Coste C, Ahn MT, Doshishti-Agolli K, Lizen B, and Gofflot F. (2016). Systematic expression analysis of Hox genes at adulthood reveals novel patterns in the central nervous system. Brain Struct. Funct. 221, 1223–1243. 10.1007/s00429-014-0965-8. [PubMed: 25527350]

80. Takahashi Y, Hamada J.i., Murakawa K, Takada M, Tada M, Nogami I, Hayashi N, Nakamori S, Monden M, Miyamoto M, et al. (2004). Expression profiles of 39 HOX genes in normal human adult organs and anaplastic thyroid cancer cell lines by quantitative real-time RT-PCR system. Exp. Cell Res. 293, 144–153. [PubMed: 14729064]

81. De Fruyt N, Yu AJ, Rankin CH, Beets I, and Chew YL (2020). The role of neuropeptides in learning: Insights from C. elegans. Int. J. Biochem. Cell Biol. 125, 105801. 10.1016/j.biocel.2020.105801.

82. Chalasani SH, Kato S, Albrecht DR, Nakagawa T, Abbott LF, and Bargmann CI (2010). Neuropeptide feedback modifies odor-evoked dynamics in Caenorhabditis elegans olfactory neurons. Nat. Neurosci. 13, 615–621. 10.1038/nn.2526. [PubMed: 20364145]

83. Fadda M, De Fruyt N, Borghgraef C, Watteyne J, Peymen K, Vandewyer E, Naranjo Galindo FJ, Kieswetter A, Mirabeau O, Chew YL, et al. (2020). NPY/NPF-Related Neuropeptide FLP-34 Signals from Serotonergic Neurons to Modulate Aversive Olfactory Learning in Caenorhabditis elegans. J. Neurosci. 40, 6018–6034. 10.1523/JNEUROSCI.2674-19.2020. [PubMed: 32576621]

84. Chang YJ, Burton T, Ha L, Huang Z, Olajubelo A, and Li C. (2015). Modulation of Locomotion and Reproduction by FLP Neuropeptides in the Nematode Caenorhabditis elegans. PLoS One 10, e0135164. 10.1371/journal.pone.0135164.

85. Chen D, Taylor KP, Hall Q, and Kaplan JM (2016). The Neuropeptides FLP-2 and PDF-1 Act in Concert To Arouse Caenorhabditis elegans Locomotion. Genetics 204, 1151–1159. 10.1534/genetics.116.192898. [PubMed: 27585848]

86. Chew YL, Grundy LJ, Brown AEX, Beets I, and Schafer WR (2018). Neuropeptides encoded by nlp-49 modulate locomotion, arousal and egg-laying behaviours in Caenorhabditis elegans via the receptor SEB-3. Philos. Trans. R. Soc. Lond. B Biol. Sci. 373, 20170368. 10.1098/rstb.2017.0368.

87. Ramachandran S, Banerjee N, Bhattacharya R, Lemons ML, Florman J, Lambert CM, Touroutine D, Alexander K, Schoofs L, Alkema MJ, et al. (2021). A conserved neuropeptide system links head and body motor circuits to enable adaptive behavior. Elife 10, e71747. 10.7554/eLife.71747. [PubMed: 34766905]

88. Stawicki TM, Takayanagi-Kiya S, Zhou K, and Jin Y. (2013). Neuropeptides function in a homeostatic manner to modulate excitation-inhibition imbalance in C. elegans. PLoS Genet. 9, e1003472. 10.1371/journal.pgen.1003472.

89. Beets I, Zels S, Vandewyer E, Demeulemeester J, Caers J, Baytemur E, Courtney A, Golinelli L, Hasakioğulları I, Schafer WR, et al. .(2023). System-wide mapping of peptide-GPCR interactions in C. elegans. Cell Rep. 42, 113058. 10.1016/j.celrep.2023.113058.

90. Van Bael S, Zels S, Boonen K, Beets I, Schoofs L, and Temmerman L. (2018). A Caenorhabditis elegans Mass Spectrometric Resource for Neuropeptidomics. J. Am. Soc. Mass Spectrom. 29, 879–889. 10.1007/s13361-017-1856-z. [PubMed: 29299835]

91. Hsu IU, Linsley JW, Zhang X, Varineau JE, Berkhoudt DA, Reid LE, Lum MC, Orzel AM, Leflein A, Xu H, et al. (2020). Stac protein regulates release of neuropeptides. Proc. Natl. Acad. Sci. USA 117, 29914–29924. 10.1073/pnas.2009224117. [PubMed: 33168737]

92. Gao S, Guan SA, Fouad AD, Meng J, Kawano T, Huang YC, Li Y, Alcaire S, Hung W, Lu Y, et al. (2018). Excitatory motor neurons are local oscillators for backward locomotion. Elife 7, e29915. 10.7554/eLife.29915. [PubMed: 29360035]

93. Wen Q, Po MD, Hulme E, Chen S, Liu X, Kwok SW, Gershow M, Leifer AM, Butler V, Fang-Yen C, et al. (2012). Proprioceptive coupling within motor neurons drives C. elegans forward locomotion. Neuron 76, 750–761. 10.1016/j.neuron.2012.08.039. [PubMed: 23177960]

94. Yemini E, Lin A, Nejatbakhsh A, Varol E, Sun R, Mena GE, Samuel ADT, Paninski L, Venkatachalam V, and Hobert O. (2021). NeuroPAL: A Multicolor Atlas for Whole-Brain Neuronal Identification in C. elegans. Cell 184, 272–288.e11. 10.1016/j.cell.2020.12.012. [PubMed: 33378642]

95. Holdorf AD, Higgins DP, Hart AC, Boag PR, Pazour GJ, Walhout AJM, and Walker AK (2020). WormCat: An Online Tool for Annotation and Visualization of Caenorhabditis elegans Genome-Scale Data. Genetics 214, 279–294. [PubMed: 31810987]

96. Young MD, and Behjati S. (2020). SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. GigaScience 9, giaa151. 10.1093/gigascience/giaa151.

97. McCarthy DJ, Campbell KR, Lun ATL, and Wills QF (2017). Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. Bioinformatics 33, 1179–1186. 10.1093/bioinformatics/btw777. [PubMed: 28088763]

98. Qiu X, Hill A, Packer J, Lin D, Ma YA, and Trapnell C. (2017). Single-cell mRNA quantification and differential analysis with Census. Nat. Methods 14, 309–315. 10.1038/nmeth.4150. [PubMed: 28114287]

99. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, and Trapnell C. (2017). Reversed graph embedding resolves complex single-cell trajectories. Nat. Methods 14, 979–982. 10.1038/nmeth.4402. [PubMed: 28825705]

100. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, and Rinn JL (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. 32, 381–386. 10.1038/nbt.2859. [PubMed: 24658644]

101. Haghverdi L, Lun ATL, Morgan MD, and Marioni JC (2018). Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat. Biotechnol. 36, 421–427. 10.1038/nbt.4091. [PubMed: 29608177]

102. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, Hao Y, Stoeckius M, Smibert P, and Satija R. (2019). Comprehensive Integration of Single-Cell Data. Cell 177, 1888–1902.e21. 10.1016/j.cell.2019.05.031. [PubMed: 31178118]

103. Bhattacharya A, Aghayeva U, Berghoff EG, and Hobert O. (2019). Plasticity of the Electrical Connectome of C. elegans. Cell 176, 1174–1189.e16. 10.1016/j.cell.2018.12.024. [PubMed: 30686580]

104. Harris TW, Arnaboldi V, Cain S, Chan J, Chen WJ, Cho J, Davis P, Gao S, Grove CA, Kishore R, et al. (2020). WormBase: a modern Model Organism Information Resource. Nucleic Acids Res. 48, D762–D767. 10.1093/nar/gkz920. [PubMed: 31642470]

105. Stefanakis N, Carrera I, and Hobert O. (2015). Regulatory Logic of Pan-Neuronal Gene Expression in C. elegans. Neuron 87, 733–750. 10.1016/j.neuron.2015.07.031. [PubMed: 26291158]

106. Osseward PJ 2nd and Pfaff SL (2019). Cell type and circuit modules in the spinal cord. Curr. Opin. Neurobiol. 56, 175–184. 10.1016/j.conb.2019.03.003. [PubMed: 30954861]

107. Hu H, Miao YR, Jia LH, Yu QY, Zhang Q, and Guo AY (2019).AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. Nucleic Acids Res. 47, D33–D38. 10.1093/nar/gky822. [PubMed: 30204897]

Highlights

- Nerve cord motor neurons (MNs) in adult *C. elegans* are organized into 29 subclasses
- Each MN subclass expresses a distinct code of transcription factors (TFs) and neuropeptides
- Molecularly distinct MN subclasses display connectivity differences
- Codes of homeodomain TFs delineate adult MN diversity in *C. elegans* and mice

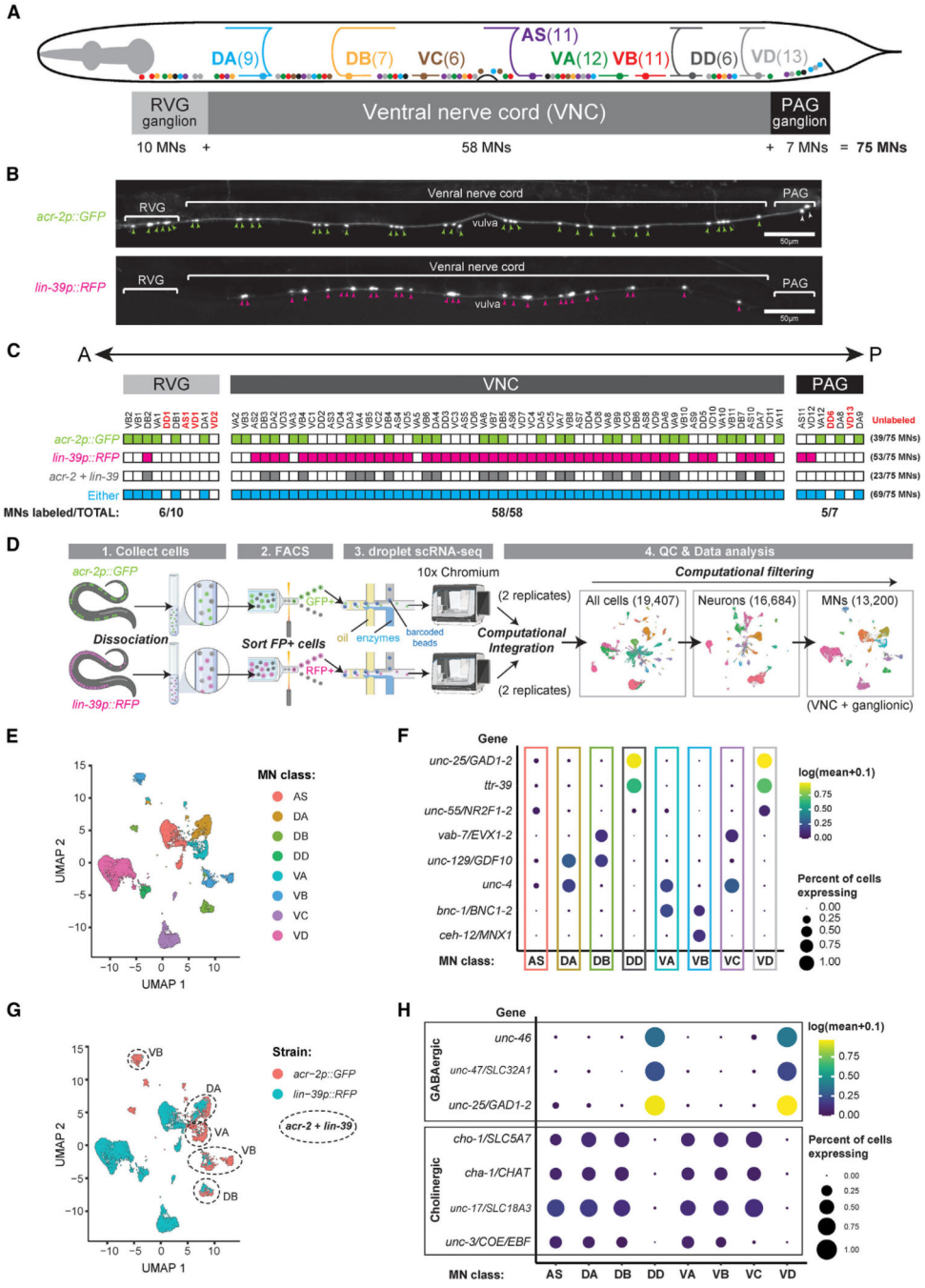


Figure 1. Strategy for scRNA-seq of adult *C. elegans* MNs

(A) Schematic of *C. elegans* MNs in RVG, VNC, and PAG.

(B) Fluorescence micrographs: *acr-2p::GFP* and *lin-39p::RFP* transgene expression. Scale bar, 50 μm.

(C) Single-neuron reporter data for each transgene depicted in (B). n = 10.

(D) Workflow for scRNA-seq. Ganglionic and VNC MNs were isolated separately from *acr-2::gfp* and *lin-39::rfp* transgenic animals. Two biological replicates.

(E) UMAP plot: molecular separation of all eight MN classes.

(F) Dot plot showing log expression and percentage of cells (VNC and ganglionic MNs) expressing known MN class-specific genes.
(G) UMAP plot in (E), but colors depict strain of origin.
(H) Dot plot showing log expression and percentage of cells expressing neurotransmitter identity genes.

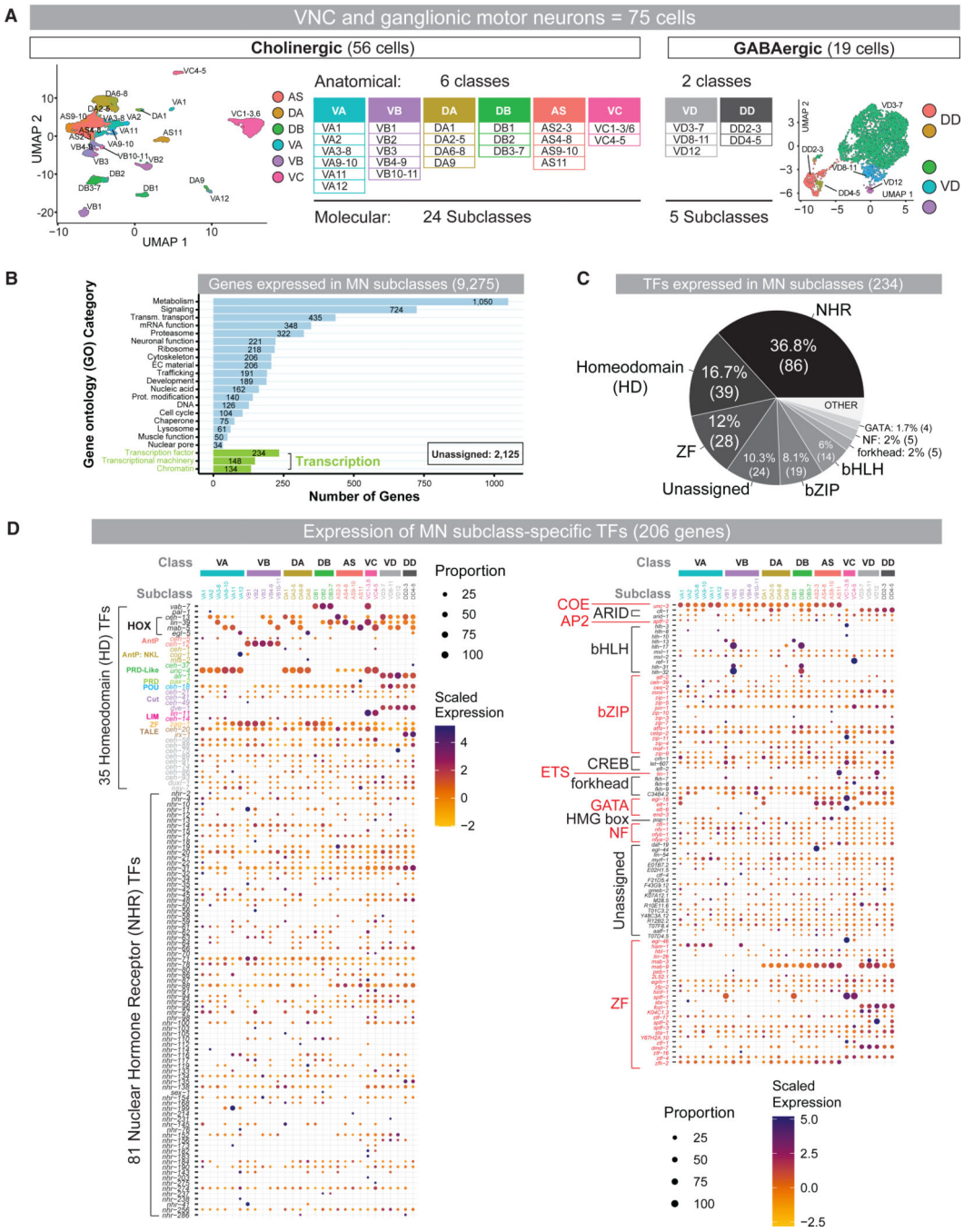


Figure 2. scRNA-seq identifies striking MN diversity and a subclass-specific TF code

(A) UMAP and table showing 24 cholinergic MN subclasses (left) and five GABAergic MN subclasses (right) in adult *C. elegans* VNC and ganglia.

(B) Gene Ontology (GO) categories (wormcat.com) over-represented (Fisher's exact test, $p < 0.01$) in adult MN subclasses.

(C) Chart depicting TF numbers detected in adult MNs.

(D) Dot plots showing scaled expression and proportion of cells expressing MN subclass-specific TFs. TF families (left).

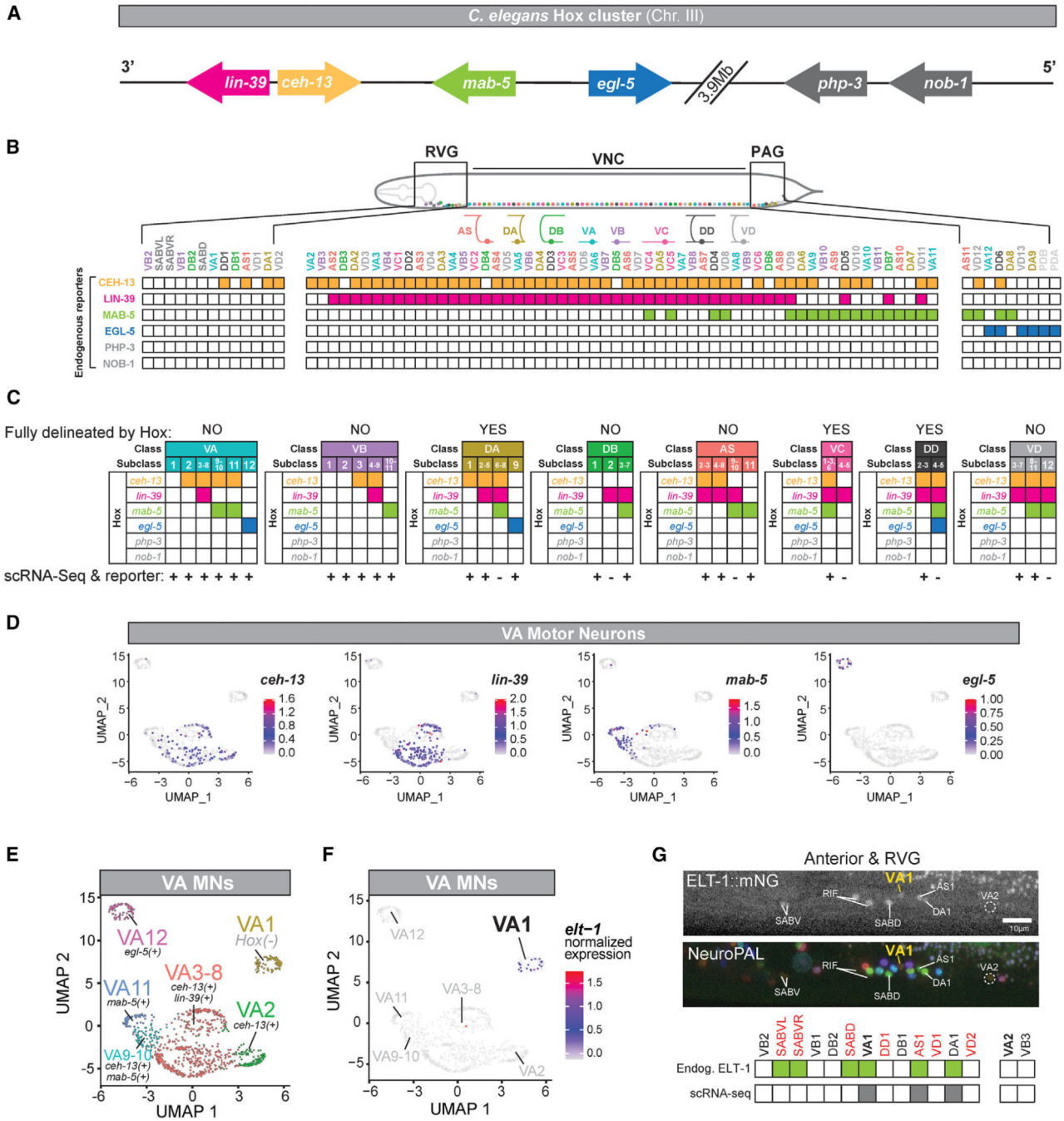


Figure 3. Hox gene expression delineates adult MN subclasses

(A) *C. elegans* Hox cluster.

(B) Expression of endogenous fluorescent reporters for all six Hox genes in every adult MN (VNC and ganglionic).

(C) Tables showing strong concordance for Hox gene expression between scRNA-seq and endogenous reporter analysis. Colored boxes denote scRNA-seq expression. +, complete agreement between the two methods; −, disagreement. “Yes” or “no” indicates whether or not all subclasses of a given class can be solely defined by Hox codes.

(D) Feature plots showing log Hox expression in VA MNs.
(E) UMAP: all VA subclasses (color coded) and Hox expression.
(F) UMAP: *elt-1*/GATA1–3 in VA.
(G) Endogenous *elt-1::mNG* and NeuroPAL expression in RVG and anterior VNC MNs, *elt-1*/GATA1–3 scRNA-seq data (bottom). Red: MNs not captured by FACS. Expression pattern: average of nine biological replicates. Scale bar, 10 μm.

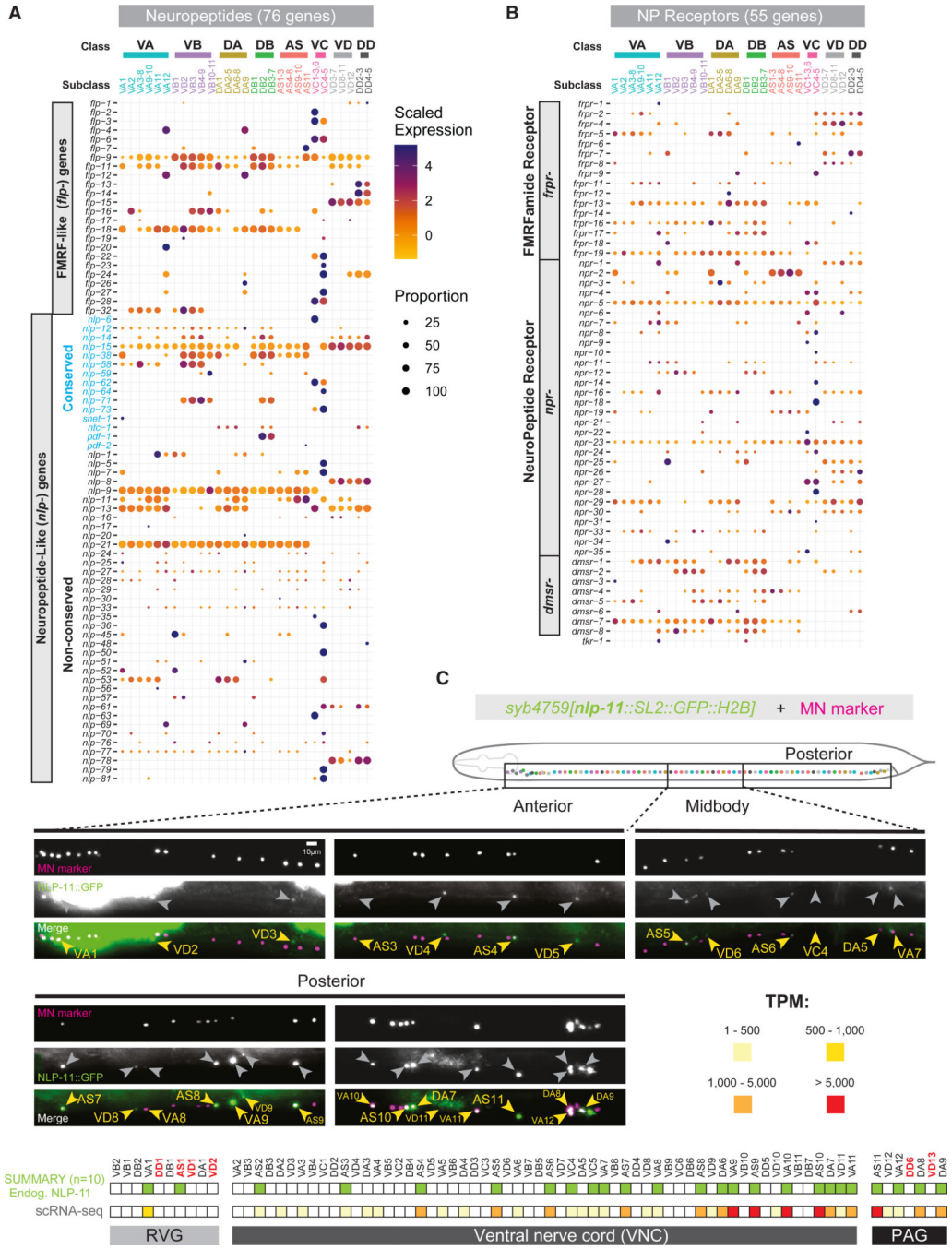


Figure 4. Combinatorial expression of extra-synaptic signaling genes delineates adult MN subclasses

(A and B) Dot plots: scaled expression and proportion of cells expressing genes encoding neuropeptides (A) or neuropeptide receptors (B) in adult MNs (ganglionic and VNC). (C) Fluorescence micrographs and endogenous reporter expression for *nlp-11(syb4759)*. TPM, transcripts per million. scRNA-seq transcript detection displayed as color code (light yellow, 1–500 TPM; red, >5,000 TPM). Red, MNs not captured by FACS. Expression pattern: average of nine biological replicates. Scale bar, 10 μm.

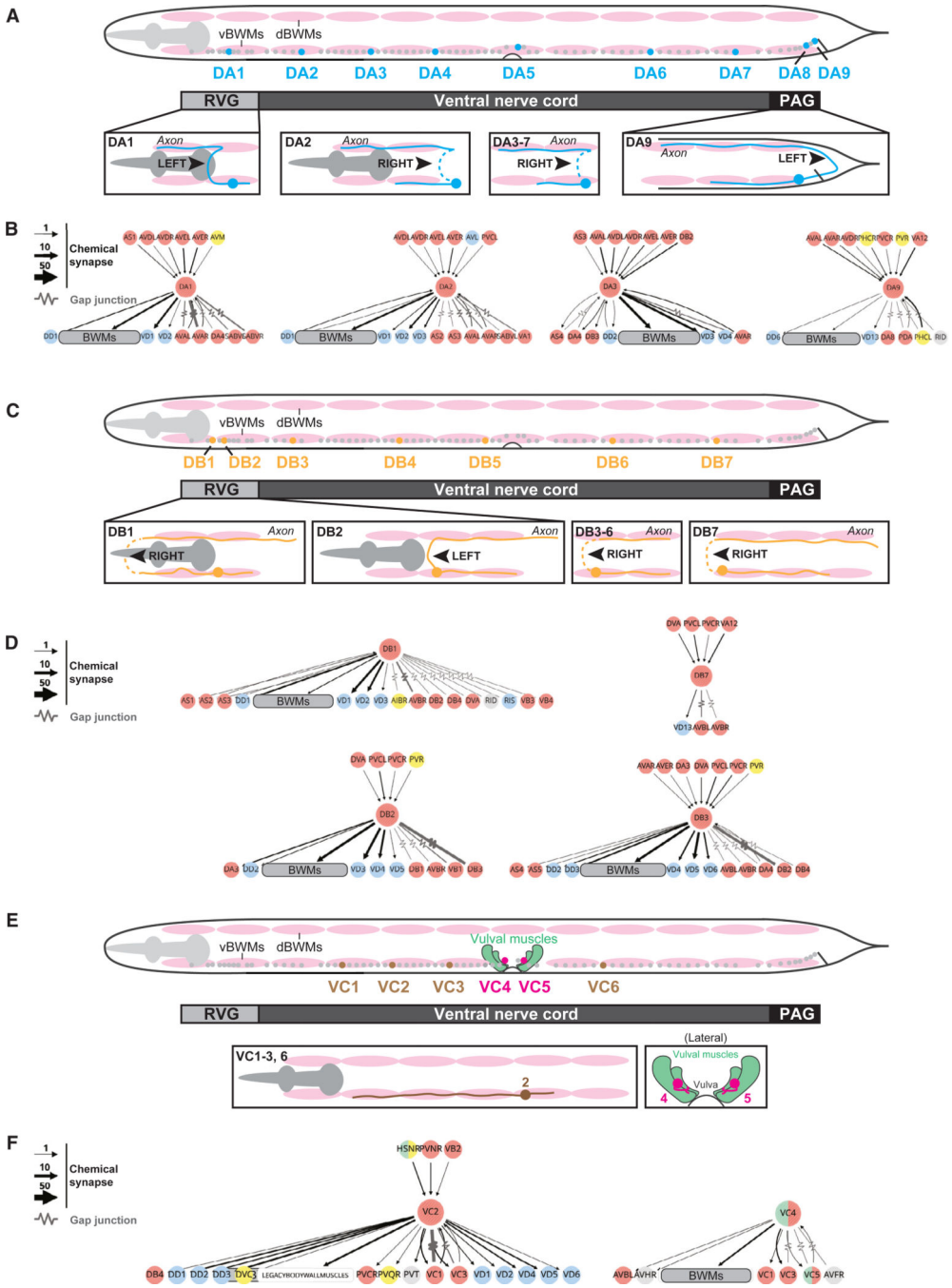


Figure 5. DA and DB MN subclasses display distinct connectivity patterns

(A, C, and E) Schematics: soma location for DA (A), DB (C), and VC (E) MNs and their respective morphologies derived from EM reconstructions (wormatlas.org; wormwiring.org). v/dBWMs, ventral/dorsal body wall muscle.

(B, D, F) Neural network diagrams of DA (B), DB (D), and VC (F). Network diagrams from nemanode.org using complete adult EM data.[33,48] Connections depicted with at least one chemical or electrical synapse. Red, cholinergic; blue, GABAergic; yellow, glutamatergic; gray, unknown.

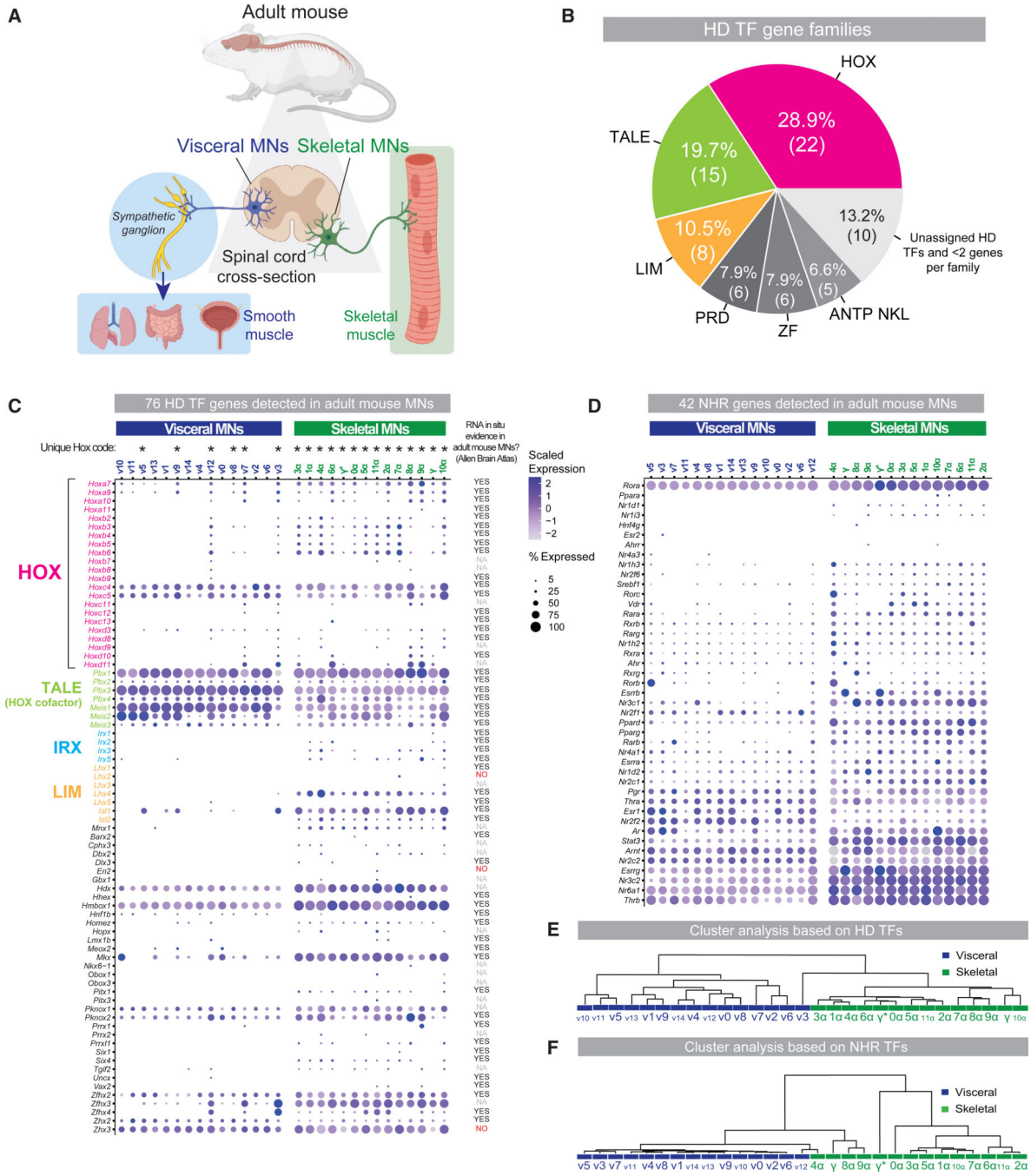


Figure 6. HD TFs delineate adult mouse MN diversity

(A) Schematic: cholinergic MN populations of the adult mouse spinal cord. Blue, visceral; green, skeletal.[25]

(B) Pie chart: number of HD TF genes from each family detected in adult mouse MNs.[50]

(C and D) Dot plot: scaled expression and proportion (%) of cells expressing the 76 HD TF genes (C) or the 42 NHR TF genes (D) detected in either visceral or skeletal MNs. Scaled expression: centered, scaled expression for each gene across all MN subclasses. Values and corresponding colors reflect standard deviations below (negative values) or above (positive

values) mean expression. Left: HD TF families. Right: *in situ* hybridization data from Allen Brain Atlas for each HD TF gene. YES, clear MN expression; NA, no data available; NO, no MN expression.

(E and F) Clustering dendrogram based on averaged expression of each HD TF gene (E) or each NHR TF gene (F) across all visceral and skeletal MN subclasses.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
E. coli: OP-50–1	Caenorhabditis Genetics Center (CGC) (https://cgc.umn.edu/)	WormBase ID: WBStrain00041971
E. coli: NA22	Caenorhabditis Genetics Center (CGC) (https://cgc.umn.edu/)	WormBase ID: WBStrain00041948
Chemicals, peptides, and recombinant proteins		
Pronase, Protease from *Streptomyces griseus*	Sigma-Aldrich	Cat# P8811
Leibovitz L-15 medium, no phenol red	ThermoFisher	Cat# 21083027
Dithiothreitol	Fisher	Cat# BP172–5
Sodium dodecyl sulfate	Research Organics, Inc.	Cat# 9010L
DAPI	Sigma	Cat# D9542
Fetal Bovine Serum, Heat Inactivated	Fisher	Cat# SH3007103HI
Critical commercial assays		
10x Chromium Single Cell 3′ GEM, Library & Gel Bead Kit v3	10x Genomics	Cat#1000269
10X Chromium Single Cell A Chip Kit	10x Genomics	Cat#1000009
High sensitivity DNA reagents (used with Agilent Bioanalyzer 2100 system)	Agilent Technologies	Cat#5067–4626
Deposited data		
R analysis code for scRNA-seq for *C. elegans* and mouse	This study	Zenodo Repository https://doi.org/10.5281/zenodo.10567646
Raw data scRNA sequencing	This study	NCBI GEO: GSE234962
Online visualization tool for adult *C. elegans* motor neurons	This study	celegans.spinalcordatlas.org
Experimental models: Organisms/strains		
C. elegans: Wild-type N2	Caenorhabditis Genetics Center (CGC) N2 (https://cgc.umn.edu/)	WormBase ID: WBStrain00000001
C. elegans: kasIs2[lin-39_intron1::tagRFP]	Feng et al.[28]	KRA257
C. elegans: juIs14[acr-2p::GFP + lin-15(+)]	CGC	CZ631

REAGENT or RESOURCE	SOURCE	IDENTIFIER
C. elegans: ceh-13(syb2307[ceh-13::3xFLAG::mNG::AID]); otIs544[cho-1(fosmid)::SL2::mChOpti::H2B + pha-1(+)]	This study	KRA831
C. elegans: mab-5(syb6730[mab-5::3xFLAG::mNG::AID]);otIs544[cho-1(fosmid)::SL2::mChOpti::H2B + pha-1(+)]	This study	KRA832
C. elegans: mab-5(syb6730[mab-5::3xFLAG::mNG::AID]);otIs564[unc-47(fosmid)::SL2::H2B::mChOpti + pha-1(+)]	This study	KRA833
C. elegans: egl-5(syb2361[egl-5::mNG::AID])	This study	PHX2361
C. elegans: php-3(st12213[php-3::TY1::EGFP::3xFLAG])	Gift from lab of Dr. Amanda Zacharias (Cincinnati Children's Hospital Medical Center)	RW12213
C. elegans: nob-1(syb2679[nob-1::GFP])	Gift from lab of Dr. John Murray (University of Pennsylvania)	PHX2679
C. elegans: elt-1(dev217[elt-1::mNeonGreen]); otIs669 him-5(1490) NeuroPAL (Neuronal Polychromatic Atlas of Landmarks) transgene (Yemini et al.[94]).	This study	NC4097
C. elegans: nlp-11(syb4759[nlp-11::SL2::GFP::H2B]); otIs544[cho-1(fosmid)::SL2::mChOpti::H2B + pha-1(+)]	This study	KRA834
C. elegans: hlh-32(st12256[hlh-32::TY1::EGFP]);otIs544[cho-1(fosmid)::SL2::mChOpti::H2B + pha-1(+)]	This study	KRA835
C. elegans: egl-44(st12173[egl-44::GFP + loxP + unc-119(+) + loxP]) II; otIs669 him-5(e1490) NeuroPAL	This study	NC4099
C. elegans: vab-3(dev190[vab-3::mNeonGreen]); otIs544[cho-1(fosmid)::SL2::mChOpti::H2B + pha-1(+)]	This study	KRA836
C. elegans: eEx650[ilys-4p::GFP + unc-119(+)]; otIs544[cho-1(fosmid)::SL2::mChOpti::H2B + pha-1(+)]	This study	KRA837
C. elegans: nlp-13(syb3411[nlp-13::T2A::3xNLS::GFP]);otIs544[cho-1(fosmid)::SL2::mChOpti::H2B + pha-1(+)]	This study	KRA838
C. elegans: ynIs66[flp-7p::GFP]; otIs544 [cho-1(fosmid)::SL2::mChOpti::H2B + pha-1(+)]	This study	KRA839
C. elegans: ynIs40[flp-11p::GFP]; otIs544 [cho-1(fosmid)::SL2::mChOpti::H2B + pha-1(+)]	This study	KRA840
C. elegans: rtEx224[F18E9.2(nlp-7)::GFP + lin-15(+)];otIs544[cho-1(fosmid)::SL2::mChOpti::H2B + pha-1(+)]	This study	KRA841
Software and algorithms		
Adobe Illustrator 2022 (23.2.2 Release) −2024	Adobe.com	Adobe Suite
Adobe Photoshop 2022 (23.2.2 Release) −2024	Adobe.com	Adobe Suite
(Fiji Is Just) ImageJ, Fiji (version 2.9.0/1.53t)	Fiji.sc	ImageJ
Jackson Laboratory Mouse Genome Informatics tool	https://www.informatics.jax.org/	Mouse Genomic Informatics Tool
Rstudio (2023.6.0)	https://posit.co/download/rstudio-desktop/	RStudio
R version 4.2.0	https://cran.r-project.org/	The Comprehensive R Archive Network
TFDB database (version 3.0)	http://bioinfo.life.hust.edu.cn/AnimalTFDB4/	AnimalTFDB3.0
WormCat	Wormcat.com	WormCat 2.0 (Holdorf et al.[95])
ZEN Blue	https://www.zeiss.com/microscopy/en/home.html	ZEN Blue version 2.3.69.1000

REAGENT or RESOURCE	SOURCE	IDENTIFIER
ZEN 2 Blue	https://www.zeiss.com/microscopy/en/home.html	ZEN 2 Blue version 3.8.99.02000
Other		
Zeiss, Axio Imager Z2	Carl Zeiss Microscopy	https://www.zeiss.com/microscopy/
Zeiss LSM 880 Confocal Microscope	Carl Zeiss Microscopy	https://www.zeiss.com/microscopy/
Zeiss, LSM 900 Confocal Microscope	Carl Zeiss Microscopy	https://www.zeiss.com/microscopy/