Joint actions of diverse transcription factor families ensure enhancer selectivity and robust neuron terminal differentiation

Angela Jimeno-Martín, Noemi Daroqui, Erick Sousa, Rebeca Brocal-Ruiz, Miren Maicas, Nuria Flames*

Developmental Neurobiology Unit, Instituto de Biomedicina de Valencia IBV-CSIC, Valencia, 46010, Spain

* Correspondence: nflames@ibv.csic.es

Running title: Specific TF families drive neuron specification

Keywords

neuron specification, *C. elegans*, terminal differentiation, transcription factor, regulatory genome, robustness, gene regulatory network, RNAi screen

ABSTRACT

Identifying general principles for neuron-type specification programs remains a major challenge. Here we performed RNA interference (RNAi) screen against all 875 transcription factors (TFs) encoded in *Caenorhabditis elegans* genome to systematically investigate the complexity of neuron-specification regulatory networks. Using this library, we screened for defects in nine different neuron types of the monoaminergic (MA) superclass and two cholinergic motoneurons to search for common themes underlying neuron specification.

We identified more than 90 TF candidates to be involved in neuron specification, of which 28 were also confirmed by mutant analysis. Correct specification of each individual neuron type requires at least nine different TFs. Individual neuron types do not usually share TFs involved in their specification but share a common pattern of involved TF families, composed mainly by five out of more than fifty in *C. elegans*: Homeodomain (HD), basic Helix Loop Helix (bHLH), Zinc Finger (ZF), Basic Leucine Zipper Domain (bZIP) and Nuclear Hormone Receptors (NHR). In addition, we studied terminal differentiation complexity focusing on the dopaminergic terminal regulatory program. We found three TFs (UNC-62, VAB-3 and MEF-2) that work together with the three known dopaminergic terminal selectors (AST-1, CEH-43, CEH-20). This complex combination of TFs provides genetic robustness and imposes a specific gene regulatory signature enriched in the regulatory regions of dopamine effector genes. Our results provide new insights on neuron-type regulatory programs in *C. elegans* that could help better understand neuron specification and the evolution of neuron types.

INTRODUCTION

Organisms' display of a wide range of complex biological functions requires cellular division of labor that results in multicellularity. Cell diversity is particularly extensive in nervous systems because the complexity of neural function demands a remarkable degree of cellular specialization. Neuron cell-types can be classified based on different criteria including morphology, physiology, neurotransmitter synthesis, molecular markers or whole transcriptomes (Zeng and Sanes 2017). Transcription factors (TFs) are the main orchestrators of neuron-type specification programs and induced expression of small combinations of TFs is sufficient for direct reprogramming of non-neuronal cells into specific neuron types (Masserdotti et al. 2016). However, the complete gene regulatory networks that implement specific neuron fates, either during development or induced by reprogramming, are still poorly understood. How many different TFs are involved in the specification of each neuron type? Are there common rules shared by all neuron-type specification programs?

Previous studies in many different organisms allowed the identification of a few conserved features of neuron specification programs. Some examples are: 1) the importance of signal-regulated TFs that translate morphogens and intercellular signaling into gene expression patterns that drive lineage commitment and neuronal progenitor patterning (Borello and Pierani 2010; Andrews et al. 2019; Liu and Niswander 2005; Angerer et al. 2011; Nuez and Félix 2012; Rentzsch et al. 2017), 2) the central role of basic Helix Loop Helix (bHLH) TFs as proneural factors (Bertrand et al. 2002; Guillemot and Hassan 2017), 3) the prevalent role of homeodomain (HD) TFs in neuron subtype

specification (Briscoe et al. 2000; Shirasaki and Pfaff 2002; Thor et al. 1999) and 4) the terminal selector model for neuronal terminal differentiation, in which, at the last steps of development specific TFs, termed terminal selectors, directly co-regulate expression of most neuron-type specific effector genes (Hobert, 2008).

Here we took advantage of the amenability of *Caenorhabditis elegans* for unbiased large-scale screens to study the complexity of neuronal gene regulatory networks and to identify common principles underlying the specification of different types of neurons. To this end, we performed an RNA interference (RNAi) screen against all 875 transcription factors (TFs) encoded by the *C. elegans* genome and systematically assessed their contribution in the specification of nine different types of neurons of the monoaminergic (MA) superclass and two cholinergic motoneurons. We focused mainly on MA neurons not only because they are evolutionary conserved and clinically relevant in humans (Flames and Hobert 2011), but also because the MA superclass comprises a set of neuronal types with very diverse developmental origins and functions in both worms and humans (Flames and Hobert 2011). Importantly, we have previously shown that gene regulatory networks directing the terminal fate of two types of MA neurons (dopaminergic and serotonergic neurons) are conserved in worms and mammals (Lloret-Fernández et al. 2018; Doitsidou et al. 2013; Flames and Hobert 2009; Remesal et al. 2020). Thus, the identification of common principles underlying MA specification could help unravel general rules for neuron-type specification.

Our results, that combine both RNAi and mutant analysis, unveil five major conclusions. First, we did not find evidence for global regulators of MA

specification. Second, each of the studied neuron types is regulated by a complex combination of at least nine different TFs likely acting at different developmental times. Third, in spite of this individual TF diversity, identified TFs consistently belong to only five out of the more than fifty *C. elegans* TF families: HD, bHLH, Zinc Finger (ZF), basic Leucine Zipper Domain (bZIP) and phylogenetically conserved members of the Nuclear Hormone Receptors (NHR). Forth, the study of dopaminergic neuronal terminal differentiation identified a combination of six TFs from different families likely acting as a terminal selector collective. Functionally, this TF complexity provides robustness to gene expression and enhancer selectivity at the genome wide level. And fifth, in agreement to previous reports, we find evidence of several regulatory routines acting in parallel to regulate neuron-type specific transcriptomes. In summary, our results provide new insights into neuronal gene regulatory networks involved in the generation and evolution of neuron diversity.

RESULTS

A whole genome transcription factor RNAi screen identifies new TFs required for neuron-type specification

To increase our understanding of the gene regulatory networks controlling neuron-type specification used an RNAi library against all the putative TFs encoded in the *C. elegans* genome. For RNAi library preparation we followed (Narasimhan et al. 2015) complete *C. elegans* list of 875 TFs (**Table S1**), which include 745 TFs present in CisBP 0.9 data base, 18 additional TFs manually curated from (Reece-Hoyes et al. 2005) TFv2 list not present in CisBP but supported by PFAM and SMART databases, plus 112 medium-confidence TFs from TFv2 list which show partial evidence for TF activity (Narasimhan et al., 2015). RNAi clones were obtained either from the two available whole genome RNAi libraries (Rual et al. 2004; Kamath et al. 2003) or were newly generated (See Material and Methods and **Table S1**).

We used the *rff-3(pk1426)* mutant background to sensitize neurons for RNAi effects as previously done (Simmer et al. 2003) and combined this mutation with three different fluorescent reporters that label the MA system in the worm (**Figure 1A**): the vesicular monoamine transporter, *cat-1::gfp* (*otIs224*), expressed in all MA neurons; the dopamine transporter, *dat-1::mcherry* (*otIs181*), expressed in dopaminergic neurons only; and the tryptophan hydroxylase enzyme *tph-1::dsred* (*vsIs97*), expressed in serotonergic neurons only. Altogether our strategy labels nine different MA neuronal classes (ADE, CEPV, CEPD, NSM, ADF, RIC, RIM, HSN, PDE) and the two cholinergic VC4 and VC5 motoneurons, which are not MA but express *cat-1* reporter for unknown reasons. Two additional MA neurons, AIM and RIH are not labeled by

117 *cat-1::gfp* (*otIs224*) and thus were not considered in our study. Analyzed
 118 neurons are developmentally, molecularly and functionally very diverse: (1) in
 119 embryonic development, they arise from very different branches of the AB
 120 lineage (**Figure 1B**), (2) they include motoneurons, sensory neurons, and
 121 interneurons, that altogether use five different neurotransmitters (**Figure 1A**)
 122 and (3) each neuronal type expresses very different transcriptomes, having in
 123 common only the minimal fraction of genes related to MA metabolism (**Figure**

Figure 1

Figure 1. A genome wide transcription factor RNAi screen reveals specific TF-families are required for the specification of eleven neuron types.

A) *cat-1::gfp* (*otIs224*) reporter strain labels nine classes of MA neurons (dopaminergic in blue, serotonergic in green and tyraminergetic and octopaminergic in brown) and the VC4 and VC5 cholinergic motoneurons. AIM and RIH serotonergic neurons are not labeled by this reporter. HSN is both serotonergic and cholinergic. For the RNAi screen the dopaminergic *dat-1::mcherry* (*otIs181*) and serotonergic *tph-1::dsred* (*vsIs97*) reporters were also used together with *otIs224*. *ttx-3::mcherry* reporter, labeling AIY is co-integrated in *otIs181* but was not scored.

B) Developmental *C. elegans* hermaphrodite lineage showing the diverse origins of neurons analyzed in this study.

C) Heatmap showing the disparate transcriptomes of the different MA neurons. Dopaminergic neurons (DA) cluster in one category because they are molecularly very similar. Data obtained from Larval L2 single-cell RNA-seq experiments (Cao et al. 2017). PDE, HSN, VC4 and VC5 are not yet mature at this larval stage.

D) Phenotype distribution of TF RNAi screen results. 91 RNAi clones produce 141 phenotypes as some TF RNAi clones are assigned to more than one cell type and/or phenotypic category. Most neuron types are affected by knock down of at least 9 different TFs. We could not differentiate between RIC and RIM due to proximity and morphological similarity, thus they were scored as a unique category.

E) TF family distribution of TFs in *C. elegans* genome, TFs retrieved in our RNAi screen, TFs confirmed by mutant analysis and mutant alleles with any assigned neuronal phenotype in WormBase. Homeodomain TFs are over-represented and NHR TFs decreased compared to the genome distribution.

F) TF family distribution of TFs confirmed by allele analysis and expressed in the affected neuron and/or its lineage. Phenotypes are divided in three categories according to TF expression: in the postmitotic neuron only, in the early lineage only or in both.

1C). Of note, the dopaminergic (DA) system, composed by four anatomically defined neuronal types (CEPV, CEPD, ADE and PDE), constitutes an exception to the MA-system diversity. Although dopaminergic neurons are developmentally diverse, they are functionally and molecularly homogeneous and share their terminal differentiation program (Doitsidou et al. 2013; Flames and Hobert 2009) (**Figure 1C**). In summary, considering the diversity of labeled neurons we reasoned their global study could unravel shared principles of *C. elegans* neuron specification.

Based on reporter expression in worms fed with negative control RNAi (L4440 empty vector), we set an initial threshold of 10% penetrance in at least two out of three replicates to consider a positive hit. Under these conditions, we found 141 different phenotypes assigned to 91 of the 875 TFs (**Table S2**). Missing or

reduced reporter expression was the most frequent phenotype although we also observed ectopic fluorescent cells and migration, axon guidance or morphology defects (**Figure 1D** and **Table S2**).

We retrieved phenotypes for 24 out of 31 known regulators of MA neuron fate, including TFs involved in MA lineage specification, axon guidance and known MA terminal selectors (**Table S3**). Thus, we estimated a false negative rate of around 23%. To estimate the false positive rate of the screen we decided to focus on the validation of the set of 38 TFs displaying missing fluorescence phenotypes with penetrance higher than 20% (comprising 59 phenotypes considering each neuron type individually, **Table 1** and **Table S2**). We combined published data for 14 TFs and new mutant analysis for 19 additional TFs. 40 out of 46 tested RNAi phenotypes were verified by the corresponding mutant (**Table 1** and **Table S4**) estimating a false positive rate of 14% for RNAi clones producing missing phenotypes of >20% penetrance. See **Table 2** for a summary of RNAi screen hits, validations and error rates.

In total, 38 mutant-confirmed TFs (20 already known, 11 newly characterized by us and 7 false negative TFs) produce 52 missing reporter expression phenotypes affecting different MA and VC4/VC5 neurons. Available single-cell RNAseq or reporter expression data (Reilly et al. 2020; Cao et al. 2017; Packer et al. 2019; Murray et al. 2012) shows detectable TF expression either in the affected neuron or the corresponding developmental lineage for 48 of the 52 phenotypes (92%) (**Table S5**), suggesting a cell autonomous role for most of these TFs in neuron specification.

We observed that each neuron type is affected by 9 to 15 different TF RNAi clones (**Figure 1D** and **Table S2**) with the exception of the NSM neurons, for

Table 1. TF RNAi clones associated to missing reporter expression with > 20 % penetrance.

TF	TF family	Neuron type	RNAi penetrance. % cat-1::gfp OFF	Mutant phenotype	Allele	Source of mutant data
<i>lin-32</i>	bHLH	PDE	90	Yes	<i>e1926, gm239</i>	(Zhao and Emmons, 1995)
<i>lag-1</i>	CSL	ADF	86	Yes	<i>om13 q385</i>	This work
<i>unc-4</i>	HD	VCs	83	Yes	<i>e120, e2323</i>	(Zheng et al., 2013)
<i>lin-28</i>	CSD	PDE	82	Yes	<i>n719</i>	This work and (Ambros and Horvitz 1984)
<i>lin-39</i>	HD	VCs	78	Yes	<i>n1760</i>	(Potts et al., 2009)
<i>lin-40</i>	ZF - GATA	VCs	73	Yes	<i>ku285</i>	This work
<i>tra-1</i>	ZF - C2H2	VCs	67	Yes	<i>e1488</i>	This work
<i>ceh-20</i>	HD	VCs	65	Yes	<i>ay42</i>	(Liu et al., 2006)
<i>sem-2</i>	HMG box	HSN	65	Not tested		
<i>hlh-14</i>	bHLH	ADF	64	Yes	<i>ok780, gm34</i>	This work
<i>unc-86</i>	HD	HSN	64	Yes	<i>n846</i>	(Lloret-Fernandez et al., 2018; Sze et al., 2002)
<i>hbl-1</i>	ZF - C2H2	PDE	63	Yes	<i>mg285</i>	This work
<i>ham-1</i>	WH	CEPV	54	Yes	<i>gt1984, ot339</i>	(Offenburger et al., 2017)
<i>ham-1</i>	WH	CEPD	51	Yes	<i>gt1984, ot339</i>	(Offenburger et al., 2017)
<i>nob-1</i>	HD	HSN	51	No	<i>os6</i>	This work
<i>pax-3</i>	HD	VCs	50	Not tested		
<i>egl-5</i>	HD	HSN	48	Yes	<i>n945</i>	(Singhvi et al., 2008)
<i>sex-1</i>	ZF - NHR	HSN	43	Yes	<i>y263</i>	This work
<i>hbl-1</i>	ZF - C2H2	HSN	41	Yes	<i>mg285</i>	This work
<i>ceh-43</i>	HD	CEPV	40	Yes	<i>ot406, ot340, tm480</i>	(Doitsidou et al., 2013)
<i>ceh-43</i>	HD	CEPD	40	Yes	<i>ot406, ot340, tm480</i>	(Doitsidou et al., 2013)
<i>vab-15</i>	HD	PDE	40	Yes	<i>u781</i>	This work
<i>unc-62</i>	HD	ADE	39	Yes	<i>e917</i>	This work
<i>unc-62</i>	HD	PDE	39	Yes	<i>e917</i>	This work
<i>ast-1</i>	WH - ETS	PDE	38	Yes	<i>ot417, hd1, rh300, gk463</i>	(Flames and Hobert, 2009)
<i>ast-1</i>	WH - ETS	HSN	38	Yes	<i>ot417, hd1, rh300, gk463</i>	(Lloret-Fernandez et al., 2018)
<i>vab-3</i>	HD	CEPV	38	Yes	<i>ot266, ot292, ot346</i>	(Doitsidou et al., 2008) and this work
<i>ceh-20</i>	HD	ADE	36	Yes	<i>mu290, ay42</i>	(Doitsidou et al., 2013)
<i>ast-1</i>	WH - ETS	ADE	35	Yes	<i>ot417, hd1, rh300, gk463</i>	(Flames and Hobert, 2009)
<i>ceh-20</i>	HD	PDE	35	Yes	<i>mu290, ay42</i>	(Doitsidou et al., 2013)
<i>bed-3</i>	ZF - BED	PDE	33	Yes	<i>gk996</i>	This work
<i>lin-14</i>	Unknown	VCs	33	Yes	<i>n179</i>	This work
<i>zip-5</i>	bZIP	RIC/RIM	33	Yes	<i>gk646</i>	This work
<i>vab-3</i>	HD	ADE	33	Yes	<i>ot346</i>	This work
<i>lin-14</i>	Unknown	HSN	32	Yes	<i>e912, n179</i>	(Olsson-Carter and Slack, 2010)
<i>sex-1</i>	ZF - NHR	ADF	31	No	<i>y263</i>	This work
<i>ceh-43</i>	HD	ADE	29	Yes	<i>ot406, ot340, tm480</i>	(Doitsidou et al., 2013)
<i>ceh-43</i>	HD	PDE	29	Yes	<i>ot406, ot340, tm480</i>	(Doitsidou et al., 2013)
<i>lin-28</i>	CSD	HSN	29	Yes	<i>n719</i>	(Olsson-Carter and Slack, 2010)
<i>ref-2</i>	ZF - C2H2	VCs	29	Not tested		
<i>unc-86</i>	HD	NSM	29	Yes	<i>n846</i>	(Sze et al., 2002)
<i>lin-14</i>	Unknown	NSM	28	No	<i>n179</i>	This work
<i>sem-4</i>	ZF - C2H2	HSN	28	Yes	<i>n2654</i>	(Lloret-Fernandez et al., 2018)
<i>zip-10</i>	bZIP	ADF	28	No	<i>ok3462</i>	This work
<i>ceh-44</i>	HD	PDE	27	No	<i>tm919</i>	This work
<i>egl-18</i>	ZF - GATA	HSN	26	Yes	<i>ok290</i>	(Lloret-Fernandez et al., 2018)
<i>C32E8.1</i>	bZIP	CEPD	25	Not tested		
<i>nhr-61</i>	ZF - NHR	VCs	25	Not tested		
<i>C32E8.1</i>	bZIP	CEPV	24	Not tested		
<i>lin-26</i>	ZF - C2H2	PDE	23	Yes	<i>n156</i>	This work
<i>vab-3</i>	HD	PDE	23	Yes	<i>ot346</i>	This work
<i>hlh-3</i>	bHLH	HSN	23	Yes	<i>tm1688</i>	(Lloret-Fernandez et al., 2018)
<i>lim-4</i>	HD	ADF	23	Yes	<i>yz3, yz12</i>	(Zheng et al., 2005)
<i>nhr-2</i>	ZF - NHR	RIC/RIM	22	Yes	<i>tm860</i>	This work
<i>C32E8.1</i>	bZIP	PDE	22	Not tested		
<i>dfo-1</i>	CBF	CEPD	22	No	<i>tm4702</i>	This work
<i>C32E8.1</i>	bZIP	ADE	21	Not tested		
<i>ceh-12</i>	HD	RIC/RIM	21	No	<i>tm1619</i>	This work
<i>pqn-21</i>	ZF - C2H2	ADF	21	Not tested		

which only two clones produced missing fluorescence phenotypes. We noticed that NSM neurons showed weak phenotypes upon *gfp* RNAi treatment (Supplementary Figure 1) implying that, even in the *rrf-3(pk1426)* sensitized background NSM could be particularly refractory to RNAi. Thus, we excluded NSM for further analysis. None of the TF RNAi clones affect all MA neurons,

suggesting the absence of global regulators of MA fate, despite the fact that all MA neurons share *cat-1*/VMAT expression. Alternatively, global regulators of MA fate could show redundant functions or produce early embryonic lethalties and thus would not be recovered from our RNAi screen.

In summary, our RNAi screen followed by mutant allele validation identified a new role for 11 TFs in MA and VC4/VC5 neuron specification (**Table 1**, **Table 2** and **Table S4**), including a role for *zip-5* Basic Leucine Zipper Domain (bZIP) and *nhr-2* Nuclear Hormone Receptor (NHR) in RIC and RIM specification. To our knowledge, these are the first TFs known to participate in *C. elegans* tyraminerpic and octopaminergic differentiation.

Table 2. RNAi screen summary

RNAi screen summary	
# of RNAi phenotypes (all)	141
# of RNAi phenotypes (missing fluorescence)	113
# of RNAi phenotypes >20% penetrance (missing fluorescence)	59
# of TFs with RNAi phenotype (all)	91
# of TFs with RNAi phenotype (missing fluorescence)	78
# of TFs with RNAi phenotype >20% penetrance (missing fluorescence)	39
# of phenotype assigned to mutant alleles (missing fluorescence)	52
# of TFs with phenotypes confirmed by mutant alleles (missing fluorescence)	38
# of TFs with newly assigned neuronal phenotypes confirmed by mutant alleles (missing fluorescence)	11
False negative rate (TFs)	23%
False positive rate (referred to phenotypes, >20% penetrance)	14%
False positive rate (referred to TFs, >20% penetrance)	15%

A specific set of transcription factor families controls neuron specification

Next, to try to infer common rules for neuron specification, we focused on the 113 missing reporter expression RNAi phenotypes (assigned to 78 RNAi clones) and analyzed the screen results based on TF families instead of individual TF members. According to their DNA binding domain, *C. elegans* TFs can be classified into more than fifty different TF families (Stegmaier et al. 2004; Narasimhan et al. 2015) (**Table S1**). bHLH, HD, ZF, bZIP and NHR families comprise 75% of the TFs in the *C. elegans* genome. RNAi clones targeting these families also generate 75% of the missing reporter phenotypes (**Figure**

1E). However, we noticed that the prevalence of two TF families, HD and NRH TFs, largely differs from what would be expected from the number of TF members encoded in the genome (**Figure 1E**). HD family, in spite of representing only 12% of total TFs in the *C. elegans* genome, corresponds to 27% of RNAi clones showing a phenotype (21/78) ($p < 0.005$ Fisher exact test). In contrast, NHR TFs are under-represented when considering the total number of NHR TFs encoded in the genome (**Figure 1E**). We found that NHR-TF RNAi clones represent only 9% of missing reporter phenotypes, even though 30% of all *C. elegans* TFs belong to this family ($p < 0.0001$, Fisher exact test). A roughly similar TF family distribution is observed when considering MA neuron types individually (**Supplementary Figure 2**). The specific TF family distribution found in our RNAi screen could be biased by our false positive and negative rates, thus we focused only on the set of 38 TFs confirmed by mutant analysis. Mutant-verified TFs show a very similar TF family distribution, with 30% HD TFs and 5% NHR TFs ($p < 0.0001$, $p = 0.0017$ respectively compared to whole genome, Fisher exact test), suggesting a prevalent role for HD family but not for NHRs in neuron specification (**Figure 1E**).

Next, to explore if this TF family distribution is particular to specification of the MA system and VC4, VC5 neurons or it could generally apply to neuron specification networks, we analyzed the TF family distribution of the 93 *C. elegans* TFs for which mutant alleles display any neuronal phenotype according to WormBase data (neuron specification defects, migration or axon guidance) (**Table S6**). Of note, TF family distribution for these 93 TFs is highly similar to the one observed in our RNAi screen and our mutant analysis (**Figure 1E**)

suggesting it could constitute a general rule of neuron-type specification programs.

NHR TF family has expanded in *C. elegans*, where it is composed by 272 members compared to less than 50 NRH TFs encoded in the human genome. Only 8% of the 272 *C. elegans* NHRs have orthologs in non-nematode species (Maglich et al. 2001; Taubert et al. 2011; Bodofsky et al. 2017). Importantly, we found phylogenetically conserved NHRs are enriched for neuronal functions both in the RNAi screen (3 out of 7 NHR TFs) and in the global WormBase neuronal mutants (5 out of 6 NHR TFs). This observation suggests that, among NHR members, those phylogenetically conserved could have a prevalent role in neuron specification. Alternatively, lack of phenotypes for nematode specific NHRs could be attributed to more redundant actions among them.

Defects either in early lineage specification or in neuronal terminal differentiation can produce similar reporter expression defects, thus early and late functions of TFs cannot be discerned *a priori*. In principle, TFs expressed at early stages in development but absent from the postmitotic differentiating neuron are more likely to affect lineage specification while TFs with onset of expression at the postmitotic neuroblast might be involved in terminal differentiation. As already mentioned 48 of the 52 missing-reporter phenotypes verified by mutant alleles show detectable expression for the corresponding TF either in the affected neuron or the corresponding developmental lineage. We divided TF expression data into three categories: expressed only in the postmitotic cell, expressed only in the lineage prior to the last division or both (**Figure 1F** and **Table S5**). Interestingly, TF families are differentially distributed among categories. bHLH TF expression is mostly restricted to lineage

expression only, what is in agreement with known early but not late roles of bHLH proneural factors (Guillemot and Hassan 2017). On the contrary, most HD TFs are expressed both in the lineage progenitors as well as in the postmitotic neuron, what could indicate dual roles for this TF family both as early lineage determinants and as terminal selectors.

In summary, our results show a stereotypic TF family distribution associated to specification of different neuronal types, with high prevalence of HD TFs. Moreover, we find differential early or late expression patterns for these families that suggest they could fulfill different roles in neuron specification.

Identification of new transcription factors involved in dopaminergic terminal differentiation

Our RNAi screen retrieved known and novel TFs likely required at different steps of neuron specification and differentiation. Next, we aimed to use our TF RNAi screen to specifically study the complexity of terminal differentiation programs. Terminal differentiation is regulated by specific TFs, termed terminal selectors, that in the immature postmitotic neuron directly activate the expression of effector genes such as ion channels, neurotransmitter receptors, biosynthesis enzymes, etc, that define the neuron-type specific transcriptome (Hobert 2008). To date, at least one terminal selector has been assigned to 76 of the 118 *C. elegans* hermaphrodite neuron types (Hobert 2016). TFs, including terminal selectors, act in combinations to activate target enhancers, to date, a maximum of three terminal selectors have been identified to act together in the regulation of a neuron-type. An exception to this rule is the HSN serotonergic neuron. We found that a combination of six transcription factors

work together as a terminal selector collective to control terminal differentiation of the HSN serotonergic neuron (Lloret-Fernández et al. 2018), revealing that the HSN terminal differentiation program appears considerably more complex than any of the other neuron types in which terminal selectors have been identified. Thus, we aimed to discern if HSN regulatory complexity constitutes the exception or the rule in *C. elegans* neuron-type terminal specification programs.

As mentioned, we cannot distinguish whether TFs identified in the RNAi screen act as early lineage regulators or at the terminal differentiation step. To circumvent this limitation, we decided to focus on the four dopaminergic neuron types (CEPV, CEPD, ADE and PDE) that arise from different lineages but converge to the same terminal regulatory program. Accordingly, known early lineage determinants affect unique dopaminergic neuron types, for example, CEH-16/HD TF is required for V5 lineage asymmetric divisions (Huang et al. 2009) and thus controls PDE generation but not other dopaminergic neuron types (**Table S2**). Conversely, all mature dopaminergic neurons are functionally and molecularly similar and thus converge in the same terminal differentiation program (Doitsidou et al. 2013; Flames and Hobert 2009) and RNAi clones targeting the three already known dopaminergic terminal selectors (*ast-1*, *ceh-43* and *ceh-20*) (**Figure 2A**) show phenotypes for at least two out of the four dopaminergic neuronal pairs (**Supplementary Figure 3**). Accordingly, we reasoned that RNAi clones leading to similarly broad dopaminergic phenotypes could also play a role in dopaminergic terminal differentiation.

In addition to the three already known dopaminergic terminal selectors, we found ten TF RNAi clones affecting two or more dopaminergic neuron types

(**Table S2**). To prioritize among these ten TFs, we first focused on *unc-62*/MEIS-HD and *vab-3*/PAIRED-HD that show high penetrant RNAi phenotypes that we had already confirmed by mutant analysis (**Table 1**).

***unc-62*/MEIS-HD and *vab-3*/PAIRED-HD TFs have dual roles in dopaminergic lineage specification and terminal differentiation**

We used mutant alleles to further characterize the role of these TFs in dopaminergic terminal fate. *unc-62* has multiple functions in development and null alleles are embryonic lethal precluding analysis of dopaminergic differentiation defects (Van Auken et al. 2002). Three viable hypomorphic alleles *e644*, *mu232* and *e917* show expression defects of a *cat-2*/tyrosine hydroxylase reporter, the rate-limiting enzyme for dopamine synthesis (**Figure 2B, E and Supplementary Figure 4**). Of note, in addition to missing expression, *unc-62(e644)* allele displays also a small percentage of ectopic CEPs (**Supplementary Figure 4**), similar to RNAi experiments (**Table S2**), suggesting *unc-62* could have a role in CEPs lineage determination. For further characterization, we focused on *unc-62(e917)* allele as it showed higher penetrance of missing reporter expression and no ectopic cells. *unc-62(e917)* mutant shows broad defects in expression of the dopamine pathway genes (*bas-1*/dopamine decarboxylase, *cat-2*/tyrosine hydroxylase, *cat-4*/GTP cyclohydrolase and *cat-1*/vesicular MA transporter reporter) in ADE and PDE while CEPV shows small but significant defects in *cat-2* and *cat-1* reporter expression (**Figure 2C-G**). Expression of additional effector genes not directly related to dopaminergic biosynthesis, such as the ion channel *asic-1*, is also affected in *unc-62(e917)* mutants (**Figure 2H**). Importantly, *dat-1* reporter

310 expression in the ADE is unaffected in *unc-62(e917)* mutants revealing the
311 presence of the cell and suggesting the ADE lineage is unaffected in this
312 particular allele. In contrast, *unc-62(e917)* shows loss of PDE expression for all
313 analyzed reporters raising the possibility of lineage specification defects. Two
314 additional observations support this idea: 1) *unc-62(e917)* mutants show
315 correlated *asic-1* reporter expression defects in PDE and its sister cell PVD

316 (26/60 animals lose *asic-1::gfp* expression in both cells, 34/60 show unaffected
317 expression in both and 0/60 animals only PDE or PVD expression affected),
318 suggesting the lineage is affected; and 2) PDE loss of *cat-2* or *dat-1* reporter
319 expression in *unc-62(e917)* mutant animals correlate with expression defects
320 for the ciliated marker *ift-20* and the panneuronal reporter *rab-3*
321 (Supplementary Figure 4), which are usually unaffected in terminal

Figure 2

dopaminergic specification.

A) AST-1/ETS, CEH-43/DLL HD and CEH-20/PBX HD are known terminal selector of all four dopaminergic neuron types and directly activate expression of the dopamine pathway genes. CAT-1/VMAT1/2: vesicular monoamine transporter, CAT-2/TH: tyrosine hydroxylase, CAT-4/GCH1: GTP cyclohydrolase, BAS-1/DDC: dopamine decarboxylase, DAT-1/DAT: Dopamine transporter, DA: dopamine, Tyr: tyrosine.

B) Schematic representation of *unc-62* and *vab-3* gene loci and alleles used in the analysis.

C-G) Dopamine pathway gene reporter expression analysis in *unc-62(e971)* and *vab-3(ot346)* alleles. For *cat-2* and *dat-1* analysis, disorganization of *vab-3(ot346)* head neurons precluded us from distinguishing CEPV from CEPD and thus are scored as a unique CEP category. For *bas-1*, *cat-4* and *cat-1* analysis, disorganization of *vab-3(ot346)* head precluded the identification of CEPs among other GFP expressing neurons in the region and thus only ADE and PDE scoring is shown. *Wild type otIs199(cat-2::gfp)* expression is similar to *nIs118* and not shown in the figure. n>50 each condition, *: p<0.05 compared to *wild type*. Fisher exact test. Scale: 10 μ m.

H) *asic-1* sodium channel reporter expression analysis in *unc-62(e971)* and *vab-3(ot346)* alleles.

differentiation mutants (Stefanakis et al. 2015; Flames and Hobert 2009).

Altogether, our results are consistent with a role for *unc-62* in ADE terminal differentiation and in lineage specification of PDE and CEPs. The early *unc-62* role in CEPs and PDE lineage formation precludes the assignment of later functions, however adult animals express *unc-62* both in ADE and PDE (Reilly et al. 2020) what could indicate a role in terminal differentiation for both cell types.

To characterize the role of *vab-3* in dopaminergic terminal differentiation we analyzed *vab-3(ot346)*, a deletion allele originally isolated from a forward genetic screen for dopaminergic mutants (**Figure 2B**) (Doitsidou et al. 2008). Similar to our RNAi results, reported defects in *vab-3(ot346)* consists of a mixed phenotype of extra and missing *dat-1::gfp* CEPs and accordingly it was proposed to act as an early determinant of CEP lineages (Doitsidou et al. 2008). In addition to this phenotype, expression analysis of all dopamine pathway gene reporters in *vab-3(ot346)* unravels significant *cat-4* and *bas-1* expression defects in ADE and *cat-2* expression defects in ADE and PDE

(**Figure 2D, E, F**). Importantly, *dat-1* and *cat-1* reporter expression is unaffected in ADE and PDE which strongly suggests the missing expression phenotype in these two neuron types is not due to lineage specification defects (**Figure 2C, G**). We confirmed CEP lineage defects in *vab-3(ot346)* mutants with the analysis of *gcy-36* and *mod-5* expression, effector genes expressed in URX (sister cell of CEPD) and AIM (cousin of CEPV) respectively. We find both are ectopically expressed in *vab-3(ot346)* animals (**Supplementary Figure 4**).

Thus, similar to *unc-62*, *vab-3* seems to have a dual role in dopaminergic specification: it is required for proper ADE and PDE terminal differentiation and for correct CEP lineage generation. A potential role for *vab-3* in CEPs terminal differentiation could be masked by its earlier requirement in lineage determination. Interestingly, adult worms maintain *vab-3* expression in CEPS while adult ADE and PDE expression has not been reported (Reilly et al. 2020).

***mef-2*/MADS provides robustness to the dopaminergic terminal differentiation program**

Our results expand to five the number of TFs involved in dopaminergic terminal differentiation with strong predominance of HD TFs: *ast-1*/ETS TF, *ceh-43*/HD, *ceh-20*/HD-PBX, *unc-62*/HD-MEIS and *vab-3*/HD-PAIRED. However, RNAi for additional non-HD TFs displayed reporter expression defects in several dopaminergic neuron types (**Table S2**) albeit at lower penetrance. We analyzed dopamine pathway gene reporter expression in *cep-1(ep347)* P53 TF, *dro-1(tm4702)* CBF TF, *dnj-17(ju1234)* ZF TF, *mef-2(gk633)* MADS TF, *nhr-223(tm1172)* and *unc-55(e1170)* NHR TFs. A small but significant decrease in *cat-1* reporter expression in ADE neuron was found in *mef-2(gk633)* mutants

(**Figure 3A, B**). *mef-2* fosmid reporter analysis shows broad neuronal expression at first larval and adult stages that include all dopaminergic neurons (**Figure 3C** and **Supplementary Figure 5**). However, the expression of other dopamine pathway genes is unaffected in *mef-2(gk633)* (**Supplementary Figure 5**) and *cat-1* reporter expression defects were not observed using an additional allele *mef-2(gv1)* (**Supplementary Figure 5**). To rule out *mef-2(gk633)* *cat-1* reporter expression defects are due to an unknown background mutation we drove *mef-2* cDNA expression under the *bas-1* monoaminergic promoter what we found it was sufficient to rescue *cat-1* expression defects in the ADE (**Figure 3B**). Although the mechanisms underlying the discrepancy between *gk633* and *gv1* alleles is uncertain, *mef-2(gk633)* deletes the promoter and first exon completely abolishing transcription, while *mef-2(gv1)* produce a truncated transcript that could still display some functionality (**Figure 3A**).

Next, considering the extremely subtle phenotype of *mef-2(gk633)* mutants we wondered if it has any biological significance. In our previous characterization of the HSN terminal regulatory program we found extensive redundancy among TFs (Lloret-Fernández et al. 2018). Thus we hypothesized that other TFs could compensate for the lack of *mef-2*, masking its role in DA specification. To test this hypothesis we performed double mutant analysis of *mef-2(gk633)* with a hypomorphic missense mutation for *ast-1*, the main dopaminergic terminal selector (Flames and Hobert 2009). We found that *mef-2(gk633)* and *ast-1(hd1)* act synergistically in the regulation of *cat-1::gfp* expression not only in ADE but also both in CEPD and CEPV (**Figure 3D**) demonstrating a *bona fide* requirement for *mef-2* in dopaminergic terminal specification.

Figure 3

Figure 3. *mef-2*/MADS provides robustness to the dopaminergic differentiation program.

A) Schematic representation of *mef-2* locus and alleles used in the analysis.

B) *mef-2(gk633)* shows *cat-1* expression defects in the ADE neuron, supporting a role for this TF in dopaminergic terminal differentiation. These defects are rescued by specific expression of *mef-2* cDNA under *bas-1* promoter. *: $p < 0.05$, compared to *wild type*, n.s: not statistically significant compared to *wild type*. Fisher exact test with Bonferroni correction. $n > 50$ each condition.

C) Confocal microscopy L4 expression analysis of MEF-2::GFP reporter fosmid (*wgls301*). Low magnification picture is the maximal projection of a Z stack of the head showing broad neuronal expression. High magnification is a unique Z image showing MEF-2 expression in all dopaminergic neurons labeled with *dat-1::mcherry* (*otls181*). Scale: 20 μm , 5 μm .

D) Analysis of double *ast-1(hd1); mef-2(gk633)* mutants unravels synergistic effects. $n > 50$ each condition, *: $p < 0.05$. Fisher exact test. Comparison of double phenotypes to the addition of both single phenotypes. Scale: 10 μm .

Similar to what we found for HSN terminal selectors (Lloret-Fernández et al. 2018), synergistic effects of *mef-2(gk633)* and *ast-1(hd1)* are reporter-specific, *mef-2(gk633); ast-1(hd1)* double mutants show no defects in *bas-1::gfp* expression, revealing that ADE and CEPs are not absent in these animals (Supplementary Figure 5). These results expand the number of TFs and TF families involved in DA terminal differentiation and underscore the importance of genetic redundancy in neuron terminal differentiation programs.

The dopaminergic terminal differentiation program is phylogenetically conserved

Next we performed neuron-type specific rescue experiments to test for cell autonomous effects of these TFs. *unc-62* codes for eight different coding isoforms, *unc-62 isoform a* (*unc-62a*) is expressed neuronally and it is the only isoform commonly affected in *e917*, *mu232* and *e644* alleles, which all show dopaminergic defects. We expressed *unc-62a* under the *dat-1* promoter as *dat-1* expression is unaffected in CEPs and ADE in *unc-62(e971)* mutants (Figure

2C), however, we find that none of the two transgenic lines produce significant rescue of the *unc-62(e971)* phenotype (**Figure 4A**). As we have determined that *unc-62* has a dual role in terminal differentiation and lineage specification, failure to rescue could indicate *unc-62* might be required in ADE also earlier than the onset of *dat-1* expression, alternatively different isoforms or specific TF levels might be needed for rescue.

A similar strategy was used for *vab-3(ot346)* rescue experiments. *vab-3* codes for three isoforms, but only the long isoform *vab-3a* contains both the PAIRED and the HD DNA binding protein domains. Expression of *vab-3a* under the *dat-1* promoter is sufficient to rescue *cat-2* reporter expression defects in ADE, demonstrating a cell autonomous and terminal role for *vab-3* in neuron specification (**Figure 4B**). As expected, *vab-3(ot346)* CEP lineage defects were not rescued, as the promoter used for *vab-3* rescue (*dat-1prom*) is only expressed in terminally differentiated CEPs.

For *mef-2* rescue experiments we used double mutants with *ast-1(hd1)*, that show stronger dopaminergic expression defects. *mef-2* codes for a single isoform, we find *mef-2* cDNA expression under the *bas-1* promoter rescues loss of CEPV and CEPD *cat-1* reporter expression in *ast-1(hd1); mef-2(gk633)* double mutants, demonstrating a cell autonomous and terminal role for *mef-2* in dopaminergic neuron specification (**Figure 4C**).

Figure 4

Figure 4. The dopaminergic terminal differentiation program is phylogenetically conserved.

A-C) Cell autonomous rescue experiments of *unc-62*, *vab-3* single mutants and *ast-1;mef-2* double mutants with *C. elegans* and mouse ortholog cDNAs. *dat-1* promoter was used to drive expression in dopaminergic neurons in *unc-62* and *vab-3* mutants and *bas-1* promoter to drive expression in *ast-1(hd1); mef-2(gk633)* double mutants. L1 and L2 represent two independent transgenic lines. Arrowheads in C point CEPD, the additional neuron is the ADF, unaffected in these mutants. Scale: 10 μ m n>50 each condition, *: p<0.05 compared to mutant phenotype. Fisher exact test with Bonferroni correction for multiple comparisons.

Mouse orthologs for *ast-1*, *ceh-43* and *ceh-20* (Etv1, Dlx2 and Pbx1 respectively) are required for mouse olfactory bulb dopaminergic terminal differentiation, thus the dopaminergic terminal differentiation program seems to be phylogenetically conserved (Brill et al., 2008; Cave et al., 2010; Flames and Hobert, 2009; Remesal et al., 2020). Remarkably, Meis2, the mouse ortholog of *unc-62* and Pax6, the mouse ortholog of *vab-3* are also necessary for olfactory bulb dopaminergic specification (Agoston et al., 2014; Brill et al., 2008). Thus, we performed similar rescue experiments using mouse orthologs for the new dopaminergic TFs. Mouse Pax6 and Mef2a are able to rescue *vab-3* and *mef-2*

phenotypes respectively (**Figure 4B, C**). However, Meis2 did not produce significant rescue of *unc-62* mutants, similar to *unc-62* cDNA failure to rescue, suggesting MEIS factors might be also required at earlier time points, specific concentrations or isoforms (**Figure 4A**). Our data further supports the phylogenetic conservation of the gene regulatory network controlling dopaminergic terminal differentiation.

Cis-regulatory module of *cat-2/tyrosine hydroxylase* dopaminergic effector gene contains functional binding sites for UNC-62, VAB-3 and MEF-2

We have previously reported functional binding sites (BS) for *ast-1*, *ceh-43* and *ceh-20* terminal selectors in the *cis*-regulatory modules of the dopamine pathway genes (Flames and Hobert 2009; Doitsidou et al. 2013). To analyze if the new regulators of dopaminergic terminal fate could also directly activate dopaminergic effector gene expression we focused on the *cis*-regulatory analysis of *cat-2/tyrosine hydroxylase* a gene exclusively expressed in the dopaminergic neurons.

The *cat-2* minimal *cis*-regulatory module (*cat-2p21*), in addition to the previously described ETS, PBX and HD BS (Flames and Hobert 2009; Doitsidou et al. 2013) also contains predicted MADS, MEIS and PAIRED BS (**Figure 5B**). VAB-3 protein contains two DNA binding domains, PAIRED and HD where each DNA domain fulfills specific functions (Brandt et al. 2019). Interestingly, we find that one of the two already identified functional HD BS matches a PAIRED-type HD consensus (HTAATTR, labeled as HD* in **Figure 5**) suggesting it could be recognized by VAB-3 and/or CEH-43.

Point mutation in PAIRED and MEIS BS show strong *gfp* expression defects while MADS BS mutation had only a small effect in the PDE (**Figure 5** and **Supplementary Figure 6**). These data reinforce a model in which known dopaminergic terminal selectors *ast-1/Etv1*, *ceh-43/Dlx2* and *ceh-20/Pbx1* could act together with *unc-62/Meis2*, *vab-3/Pax6* and *mef-2/Mef2a* to regulate dopaminergic terminal differentiation.

Figure 5

Figure 5. *cis*-regulatory analysis of *cat-2*/tyrosine hydroxylase effector gene reveals functional binding sites for UNC-62, VAB-3 and MEF-2.

A) *cat-2* minimal dopaminergic *cis*-regulatory module (*cat-2p21*) mutational analysis. In addition to published functional AST-1/ETS, CEH-20/PBX HD, CEH-43/DLL HD, binding sites (Doitsidou et al. 2013; Flames and Hobert 2009), UNC-62/MEIS, VAB-3/PAIRED and MEF-2/MADS binding sites are also required for correct GFP reporter expression. HD* represents a PAIRED-type HD consensus (HTAATTR). Black crosses represent point mutations to disrupt the corresponding TFBS. +: > 70% of mean *wild type* construct values; +/-: expression values 70–30% lower than mean *wild type* expression values; -: values are less than 30% of mean *wild type* values. --: less than 5% of GFP expression. n > 60 cells per line.

The dopaminergic regulatory signature is preferentially associated to dopaminergic neuron effector genes

Our results show that at least six TFs seem to be required for correct *C. elegans* dopaminergic terminal specification. A seemingly complex combination of TFs activates expression of the mature HSN transcriptome (Lloret-Fernández et al. 2018). Our findings suggest this TF complexity might be a common theme in neuronal terminal differentiation programs. Then, why are these complex combinations of TFs required for gene expression? We previously described that transcription factor binding site (TFBS) clusters for the HSN TF collective are preferentially found in regulatory regions near HSN expressed genes and can be used to *de novo* identify enhancers active in the HSN neuron (Lloret-Fernández et al. 2018). Thus, in addition to robustness of gene expression, a function for this complex terminal differentiation programs might be providing enhancer selectivity.

To test this hypothesis, we asked if similar to HSN, the dopaminergic regulatory program imposes a defining regulatory signature in dopaminergic-expressed genes. Published single-cell RNA-seq data (Cao et al. 2017) was used to identify additional genes differentially expressed in dopaminergic neurons (**Supplementary Figure 7**). We found 86 genes whose expression is enriched in dopaminergic neurons compared to other clusters of ciliated sensory neurons. As expected, this gene list includes all dopamine pathway genes and other known dopaminergic effector genes, but not pancilia expressed genes (**Table S7**). In analogy to our previous analysis of the HSN regulatory genome (Lloret-Fernández et al. 2018), we analyzed the upstream and intronic sequences of these genes. For comparison purposes, we built ten thousand

sets of 86 random genes with similar upstream and intronic length distribution to dopaminergic expressed genes.

First, we focused our analysis only on the three already published dopaminergic terminal selectors (*ast-1/ETS*, *ceh-43/DLL HD* and *ceh-20/PBX HD*). We found this regulatory signature (presence of ETS+HD+PBX binding sites clustered in 700 bp or smaller regions) lacks enough specificity. All genes, either dopaminergic expressed genes or random sets, contain DNA windows with matches for all three TFs (100% of dopaminergic expressed genes compared to 100% in random sets). Reducing DNA-window search length from 700 bp to 300 bp or 150 bp did not increase specificity. Next, we expanded our analysis to search for a regulatory signature containing one or more matches for each of the seven position weight matrices associated to the dopaminergic TF terminal regulatory program (ETS+HD+HD*+PBX+MEIS+PAIRED+MADS binding sites clustered in 700bp or smaller region) (**Figure 6A**). We found that 87% of dopaminergic expressed genes contain at least one associated dopaminergic regulatory signature window, a significantly higher percentage compared to the random sets of genes (mean random sets 79%, $p < 0.001$) (**Figure 6B**). Thus, dopaminergic signature is significantly enriched in dopaminergic-expressed genes. Next, we built similar sets of differentially expressed genes for five randomly picked non-dopaminergic neuron categories (RIA, ASE, Touch Receptor neurons, GABAergic neurons and ALN/PLN/SDQ cluster) (**Supplementary Figure 7 and Table S7**). We found that the percentage of expressed genes containing the dopaminergic signature is smaller in non-dopaminergic neurons compared to dopaminergic neurons (**Figure 6B**). This difference is statistically significant for all neuron types except the

ALN/PLN/SDQ neuron cluster (**Figure 6B**). In addition, none of the non-dopaminergic neuronal types show a significant enrichment of the dopaminergic signature with respect to their respective background of ten thousand random sets of comparable genes (**Figure 6B**, $p > 0.05$). The reason why ALN/PLN/SDQ expressed genes show a higher association to dopaminergic regulatory signature compared to other non-dopaminergic neurons is uncertain. Terminal selectors for ALN/PLN/SDQ neurons are yet unknown, thus a similar combination of TF families controlling terminal differentiation of these neurons could explain the higher presence of the dopaminergic regulatory signature. Dopaminergic regulatory signature enrichment in dopaminergic-expressed genes is even more pronounced when considering only the promoter sequence (1.5 Kb upstream of the ATG) (**Figure 6C**). These data suggest that proximal regulation has a major role in neuronal terminal differentiation in *C. elegans*. Four out of five dopamine pathway genes contain at least one associated dopaminergic regulatory signature window. Remarkably, previously isolated *cis*-regulatory modules overlap with predicted dopaminergic regulatory signature windows (**Figure 6D**) (Flames and Hobert 2009). Experimentally determined enhancers for two genes, *dat-1* and *cat-2*, contain all seven types of TFBS, while *cat-1* and *bas-1* enhancers contain only 6 and 4 types of motifs respectively. Thus we performed similar dopaminergic regulatory signature analysis considering the presence of at least six or at least five motifs. We find that the full complement of seven dopaminergic TFBS is required to provide specificity to the dopaminergic regulatory signature as regulatory windows containing only six or five types of dopaminergic TF motifs are not preferentially associated to dopaminergic expressed genes compared to the random

background (**Supplementary Figure 7**), moreover, although the percentage of
genes with partial dopaminergic regulatory signature is still higher in
dopaminergic expressed genes compared to gene sets expressed in other
neuron types, this difference is smaller than with the full complement of TFBS
(**Supplementary Figure 7**).

Figure 6

Figure 6. The Dopaminergic Regulatory Signature is preferentially associated to dopaminergic expressed genes.

A) Position weight matrix logos assigned to each member of the dopaminergic TFs. The dopaminergic regulatory signature is defined by the presence of one or more matches for each of the seven PWM in less than 700 bp DNA window. Color code for the sites is used in panel D.

B) Dopaminergic regulatory signature is more prevalent in the upstream and intronic sequences of the set of 86 genes with enriched expression in dopaminergic neurons (red dot) compared to the distribution in 10,000 sets of random comparable genes (blue violin plot). Analysis of five additional gene sets with enriched expression in non-dopaminergic neurons (RIA, ASE, Touch receptor neurons, GABAergic neurons and ALN/PLN/SDQ) does not show enrichment compared to their random sets (red dots and grey violin plots) and show lower percentage of genes with dopaminergic regulatory signature. %: percentage of genes with assigned dopaminergic regulatory signature. PCTL: percentile of the real value (red dot) in the 10,000 random set value distribution. Brunner-Munzel test. *: $p < 0.05$, **: $p < 0.01$.

C) Dopaminergic regulatory signature in promoter regions (1.5 kb upstream ATG) is also enriched in dopaminergic expressed genes compared to random sets or to other non-dopaminergic expressed genes, suggesting proximal regulation has a major role in dopaminergic terminal differentiation.

D) Experimentally isolated minimal enhancers for four out of the five dopamine pathway genes (Flames and Hobert 2009) overlap with predicted dopaminergic regulatory signature windows. Black lines represent the coordinates covered by bioinformatically predicted dopaminergic regulatory signature windows. Green lines mark published minimal enhancers for the respective gene (Flames and Hobert 2009; Doitsidou et al. 2013). Specific matches for all seven TFBS classes are represented with the color code indicated in (A). Experimentally identified sites not retrieved in the bioinformatic analysis due to small sequence mistakes are marked with an asterisk. Dark blue bar profiles represent sequence conservation in *C. briggsae*, *C. brenneri*, *C. remanei* and *C. japonica*. Dopaminergic regulatory signature does not necessarily coincide with conserved regions.

545 Altogether, our data suggest that the dopaminergic TF collective acts through
546 the dopaminergic regulatory signature to activate transcription of dopaminergic
547 effector genes. Indirect recruitment of TFs by protein-protein interactions or
548 redundant actions of the TFs could explain why partial complement of TFBS
549 can be sufficient, in some context, to activate gene expression in dopaminergic

cells. Finally, the presence of the dopaminergic signature in some genes not expressed in dopaminergic neurons indicates that the signature itself is not sufficient to induce dopaminergic expression. Additional TFBS, gene repression mechanisms, or chromatin accessibility could further regulate dopaminergic regulatory signature specificity. It is also possible that specific syntactic rules (TFBS order, distance and disposition) discriminate functional from non-functional dopaminergic regulatory signature windows.

Additional parallel gene routines expressed in the dopaminergic neurons do not show enrichment in dopaminergic regulatory signature

Next, we examined the distribution of the dopaminergic regulatory signature across the entire *C. elegans* genome. We found it is preferentially present in putative regulatory sequences of neuronally expressed genes compared to the rest of the genome (**Figure 7A**). This enrichment is also present when analyzing only promoter sequences (**Figure 7A**).

Gene ontology analysis of all genes in the *C. elegans* genome with dopaminergic regulatory signature revealed enrichment of many neuronal processes related to dopaminergic differentiation and function, including regulation of locomotion, learning and memory, response to stimuli or dopamine secretion (**Figure 7B**). Similar gene ontology categories are enriched when only genes with dopaminergic regulatory signature present in their promoters are considered (**Figure 7B**).

Thus we next aim to explore to what extent dopaminergic regulator signature is present in the whole terminal transcriptome of dopaminergic neurons. Different hierarchies of gene expression co-exist in any given cell, mechanosensory

dopaminergic neurons co-express at least four types of genes: 1) Dopaminergic effector genes such as dopaminergic pathway genes, neuropeptides, neurotransmitter receptors, etc., that are preferentially (but not exclusively) expressed by this neuron type; 2) Genes coding for structural components of cilia which are expressed by all sixty ciliated sensory neurons in *C. elegans*; 3) Panneuronal genes, such as components of the synaptic machinery or cytoskeleton, expressed by all 302 neurons in *C. elegans* hermaphrodite and 4) Ubiquitous genes expressed by all cells of the organism, such as ribosomal or heat shock proteins (**Figure 7C**). Thus, we next studied the distribution of the dopaminergic regulatory signature in these additional gene categories.

We find that the percentage of genes with dopaminergic signature is highest for dopaminergic enriched effector genes compared to other sets of genes (**Figure 7C**). This difference is also present when only promoter sequences are analyzed (**Figure 7C**). Of note, although lower than dopaminergic effector gene-set, panneuronal genes also show increased presence of dopaminergic regulatory signature compared to other regulatory routines of the cell (**Figure 7C**) suggesting dopaminergic TFs could also contribute to panneuronal gene control, which is consistent with previous reports on panneuronal gene regulation (Stefanakis et al. 2015). The frequency of dopaminergic regulatory signature in panciliated and ubiquitous genes is more similar to genes not expressed in neurons. Thus, our data suggest that the main function for the dopaminergic TF collective is the regulation of neuron-type specific gene expression and they could be involved to a less extent in panneuronal gene regulation.

Figure 7

Figure 7. Dopaminergic regulatory signature is enriched in dopaminergic effector genes and panneuronally expressed genes but not in other gene sets expressed in dopaminergic neurons.

A) Genomic search of the dopaminergic regulatory signature reveals enrichment in neuron-expressed genes compared to genes expressed in non-neuronal tissues. Enrichment is maintained when only promoter sequences are analyzed. Chi-squared with Yates correction test. ***: $p < 0.001$, **: $p < 0.01$. Expression data for neuronal and non-neuronal tissues obtained from (Packer et al., 2019).

B) Gene ontology analysis of genes with associated dopaminergic regulatory signature. p-values in the corresponding GO category are represented.

C) Four sets of gene associated to different levels of gene expression specificity coexist in the dopaminergic neurons: 1) dopaminergic effector genes mostly specific of dopaminergic neurons; 2) pancilia genes expressed by all sensory ciliated neurons; 3) panneuronal genes expressed by all neurons and 4) ubiquitous genes expressed by all cells. Dopaminergic regulatory signature is enriched in dopaminergic effector genes and to a less extent in panneuronal expressed genes but it is not present in other panciliated or ubiquitous genes. Quantification of dopaminergic signature in non-neuronal genes is shown as negative control and marked with a dashed line. Expression data for pancilia, panneuronal, ubiquitous genes are obtained from (Packer et al., 2019).

DISCUSSION

Neuron-type specification is mainly controlled by five transcription factor families

Focusing mainly on the MA superclass of neurons we provide a comprehensive view of the TFs required at any developmental step in the specification of different neuronal types. We did not find any global regulator of MA fate, which is likely due to the molecular and functional diversity found in this superclass of neurons. Each neuron type is regulated by different sets of TFs, however, we uncovered general rules shared by all studied neuron types that might also apply to other non-MA neurons. First, we identified at least 9 different TFs required in the specification of each neuron type (with the exception of the NSM neuron, which might be particularly unresponsive to RNAi), evidencing that complex gene regulatory networks underlie neuron type specification. Second, in spite of this TF diversity, most TFs involved in the specification of all MA neurons consistently belong to only five out of the more than fifty *C. elegans* TF families: HD, bHLH, ZF, bZIP and phylogenetically conserved members of the NHR family. Importantly, analysis of genetic mutants displaying diverse neuronal phenotypes reveals the same TF family distribution, which validate results obtained from the RNAi screen and expands these findings to other neuronal types.

As other fundamental processes in biology, basic principles of neuron specification are expected to be deeply evolutionary conserved. Of note, a recent unbiased CRISPR screen in mouse embryonic stem cells identified 64 TFs that are able to induce neuronal phenotypes (Liu et al. 2018). Eighty percent of these factors also belong to the same five transcription factor families

identified in our study. The ancestral role of bHLH TFs as proneural factors has been established in a wide range of metazoans including vertebrates, *Drosophila*, *C. elegans* and the cnidaria *Nematostella vectensis* which has one of the most simple nervous systems (Poole et al. 2011; Lloret-Fernández et al. 2018; Guillemot and Hassan 2017; Layden et al. 2012). HD TFs are involved in many developmental processes and their role in neuron terminal differentiation is also conserved in different animal groups including cnidaria (Hobert 2016; Thor et al. 1999; Tournière et al. 2020; Babonis and Martindale 2017; Briscoe et al. 2000; Shirasaki and Pfaff 2002). Interestingly, our results suggest HD functions are not limited to neuron terminal differentiation but do also play important roles as lineage determinants. We find *vab-3/PAIRED* HD and *unc-62/MEIS* HD mutants show both dopaminergic lineage and terminal differentiation defects. Similar dual roles have been also reported for *ceh-43/HD* and *ceh-20/PBX* HD (Doitsidou et al. 2013).

Several NHRs, bZIP and ZF TFs are known to regulate neuron-type specification in mammals, such as *Couptf1*, *Nurr1*, *Tlx* and *Nr2e3* NHR TFs (Bovetti et al. 2013; Haider et al. 2000; Zetterström et al. 1997; Roy et al. 2004), *Nrl*, *cMaf* and *Mafb* bZIP TFs (Wende et al. 2012; Bianchi et al. 2003; Mears et al. 2001) or *Myt1l*, *Gli1*, *Sp8*, *Ctip2*, *Fezf1/2* ZF TFs (Mall et al. 2017; Hynes et al. 1997; Waclaw et al. 2006; Arlotta et al. 2005; Shimizu et al. 2010). However, in contrast to bHLH and HD, a general role in neuron specification for these families has been poorly studied in any model organism. Functional characterization of the NHR, bZIP and ZF TF candidates retrieved from our RNAi screen will help better understand the role of these TF families in neuron specification.

Complexity of terminal differentiation programs provide genetic robustness and enhancer selectivity

Here we characterize six TFs from different TF families that work together in the direct activation of dopaminergic effector gene expression. We find this TF complexity has at least two functional consequences: it provides genetic robustness and enhancer selectivity.

Regarding robustness, compensatory actions among TFs have been mostly described for paralogous genes, however, we find non-paralogous TFs can also act redundantly in the regulation of terminal gene expression. A few other examples of non-paralogous TF compensation are also found in embryogenesis, larval development or muscle specification in *C. elegans* (Walton et al. 2015; Kuntz et al. 2012; Baugh et al. 2005). Synergistic effects among terminal selectors have been previously reported in other neuronal types, such as HSN and NSM serotonergic neurons or glutamatergic neurons (Zhang et al. 2014a; Serrano-Saiz et al. 2013; Lloret-Fernández et al. 2018), suggesting it might be a general feature of neuron-type terminal regulatory programs.

In addition, we find TF complexity might also be important for enhancer selectivity. The sequence determinants that differentiate active regulatory regions from other non-coding regions of the genome are currently largely unknown. Interestingly, one of the best sequence predictors of active enhancers is the number of putative TF binding sites for different TFs found in a region (Kheradpour et al. 2013; Tewhey et al. 2016; Grossman et al. 2017). We find binding site clusters of the dopaminergic TFs are preferentially located in

putative regulatory regions of dopaminergic-expressed genes compared to other genes. Although HD TFs represent a large number of the dopaminergic terminal selector collective, additional TF families (such as ETS and MEF) are also required. This might be particularly relevant during terminal differentiation to increase diversity of TFBS that help distinguish dopaminergic specific enhancers. Thus TF complexity of neuronal terminal regulatory programs might not only provide genetic robustness but also facilitate enhancer selection.

Parallel gene regulatory routines co-exist in the cell

The active transcriptome of a mature neuron can be divided in different gene categories depending on specificity of expression, ranging from cell-type specific genes to ubiquitous gene expression. While it is well established that terminal selectors have a direct role in neuron-type-specific gene expression, their role in other gene categories is less clear (Hobert 2016). In our study we find the dopaminergic regulatory signature is predominantly enriched in dopamine effector genes, supporting a restricted role for terminal selectors in the regulation of neuron-type effector genes. Dopaminergic regulatory signature is also associated, to a less extent, to panneuronal genes, which is consistent with the known minor contributions for some terminal selectors in panneuronal gene expression (Stefanakis et al. 2015). In contrast, genes coding for cilia components (panncilia genes) and ubiquitous genes are mostly devoid of dopaminergic regulatory signature, suggesting that other TFs activate these gene categories. Transcriptional regulation of ubiquitous gene has been poorly studied. Broadly expressed TFs, such as Sp1, GABP/ETS and YY1 seem to play a role in ubiquitous gene expression in mammals (Bellora et al. 2007;

Farré et al. 2007). Cilia structural genes are directly regulated by *daf-19*/RFX TF (Swoboda et al. 2000), and this role is conserved in vertebrates (Choksi et al. 2014). However, *daf-19* is expressed in all neurons, thus additional unknown TFs must control cilia gene expression. In summary, and consistent with previous reports based on the analysis of a limited number of genes (Stefanakis et al. 2015), terminal selector collectives seem mostly devoted to neuron-type specific gene expression and to a less extend to panneuronal genes.

Neuron type evolution and deep homology

Mouse TF orthologs of the *C. elegans* dopaminergic terminal regulatory program are also required for olfactory bulb dopaminergic terminal specification, which is the most ancestral dopaminergic population of the mammalian brain (Agoston et al. 2014; Brill et al. 2008; Flames and Hobert 2009; Bovetti et al. 2013; Remesal et al. 2020; Qiu et al. 1995). Here we find mouse TFs can, in most cases, functionally substitute their worm orthologs suggesting both dopaminergic neuronal populations share deep homology, which refers to the relationship between cells in distant species that share their genetic regulatory programs (Shubin et al. 2009). Similar conservation of terminal regulatory programs is found for other neurons in *C. elegans* (Lloret-Fernández et al. 2018; Kratsios et al. 2012).

Identification of deep homology is an emerging strategy in evolutionary biology used to assign homologous neuronal types in distant species (Arendt 2008; Arendt et al. 2016, 2019). The study of neuronal regulatory programs in different animal groups and the description of homologous relationships will help us better understand the origin of neurons and the nervous system

725 evolution.

METHODS

C. elegans Strains and Genetics

C. elegans culture and genetics were performed as described (Brenner 1974).

Strains used in this study are listed in **Table S9**.

Generation of the Transcription factor RNAi library

To generate the complete TF RNAi library we used *C. elegans* TF list from (Narasimhan et al. 2015). 702 clones were extracted from published genome libraries Dr. Ahringer library (BioScience) and Dr. Vidal library (BioScience) (Kamath et al. 2003; Rual et al. 2004) clones were verified by Sanger sequencing. 170 clones were newly generated. TF clones were generated from genomic N2 DNA by PCR amplification of the target gene, subcloning into L4440 plasmid (pPD129.36, Addgene) and transformed into *E. coli* HT115 (ED3) strain (from CGC). The complete list of TF RNAi feeding clones and primers used to generate new clones are listed in **Table S1**.

RNAi feeding experiments

RNAi feeding experiments were performed following standard protocols (Kamath et al. 2001). Briefly, Isopropyl β -D-1-thiogalactopyranoside (IPTG) was added at 6 mM to NGM medium to prepare RNAi plates. RNAi clones were cultured overnight and induced with IPTG (4 mM) three hours before seeding. Adult gravid hermaphrodites were transferred to each different seeded IPTG plates within a drop of alkaline hypochlorite solution. After overnight incubation at 20°C, 10-15 newly hatched larvae were picked into a fresh IPTG plates seeded with the same RNAi clone and considered the parental generation (P0).

Approximately 7 days later young adult F1 generation was scored. Lethal RNAi clones, that precluded F1 analysis, were scored at P0 as young adults (**Table S2**). In each replicate, a minimum of 30 worms per clone, coming from three distinct plates were scored. All experiments were performed at 20°C. Each clone was scored in two independent replicates, a third replicate was performed when results from the first and second replicates did not coincide. In all scorings we include L4440 empty clone as negative control and *gfp* RNAi clone as positive control.

Mutant strains and genotyping

Strains used in this study are listed in **Table S9**. Deletion alleles were genotyped by PCR. Point mutations were genotyped by sequencing (**Table S10**). Alleles *vab-3(ot346)*, *unc-62(e644)* and *lag-1(q385)* were determined by visual mutant phenotype. The mutation *unc-62(mu232)* was followed through its link with the fluorescent reporter *muls35*.

Generation of *C. elegans* transgenic lines

Gene constructs for *cis*-regulatory analysis were generated by cloning into the pPD95.75 vector. For identification of the putative binding sites the following consensus sequences were used: GACA for UNC-62/MEIS HD (Campbell and Walthall 2016); GRAGBA for VAB-3/PAIRED HD (Holst et al. 1997; Kim et al. 2008) and CTAWWWWTAG for MEF-2/MADS (Boyle et al. 2014). Directed mutagenesis was performed with Quickchange II XL site-directed mutagenesis kit (Stratagene). For microinjections the plasmid of interest (50 ng/μl) was injected together with *rol-6* co-marker pRF4 [*rol-6*(su1006), 100 ng/μl]. *vab-*

3::*gfp* fosmid injection DNA mix contained fosmid (40 ng/μl) and two comarkers, pRF4 (50 ng/μl) and pNF101 (*ttx-3prom::mcherry*) (50 ng/μl). For rescue experiments, commercially available cDNA of the TF candidates was obtained from Dharmacon Inc and Invitrogen. cDNAs corresponding to the entire coding sequence of *unc-62*, *vab-3*, *mef-2*, *Meis2*, *Pax6*, and *Mef2a* were amplified by PCR and cloned into *dat-1* and *bas-1* promoter reporter plasmids replacing *gfp* cDNA (See primers in **Table S10**). cDNA plasmid (25 ng/μl) was injected directly into the corresponding mutant background as complex arrays together with digested *E. coli* genomic DNA (50 ng/μl) and pNF417 (*unc-122::rfp*, 25 ng/μl) fluorescent co-marker.

Scoring and statistics

Scoring and micrographs were performed using 40X objective in a Zeiss Axioplan 2 microscope. For RNAi screening experiments and *cis*-analysis, 30 young adult animals per line or per RNAi clone and replicate were analyzed; For mutant analysis 50 individuals were scored. RNAi experiments were performed at 20°C while *cis*-regulatory and mutant analyses were performed at 25°C. Fisher exact test, two tailed was used for statistical analysis.

For mutant analysis lack of GFP signalling was considered OFF, if GFP expression was substantially weaker than WT, 'FAINT' category was included. For *cis*-regulatory analysis two or three independent lines for each transgenic construct were analyzed. Mean expression value of three *wild type* lines is considered the reference value and mutated constructs are considered to show normal expression when GFP expression was 100-70% of average expression of the corresponding *wild type* construct ("+" phenotype). 70-30% reduction

compared to *wild type* reference value was considered a partial phenotype “+/-”. GFP expression below 30% was considered as great loss of expression (“-” phenotype) and if no GFP was detected (less than 5% of the neurons) was assigned as total loss of GFP (“--” phenotype).

For MEF-2 expression analysis *wgls301* *mef-2* fosmid reporter was crossed into *otls181* (*dat-1::mcherry*, *ttx-3::mcherry*). Images were taken with confocal TCS-SP8 Leica microscope and processed with ImageJ 1.50i software.

For cross regulation experiments, the expression of *unc-62*, *vab-3* and *mef-2* reporters were scored in 10 L1 larvae, in *wild type* and *ast-1(hd92)* background. Z stack pictures (Zeiss microscope) of all the length of the animal were taken keeping the same distance between z-planes. Images were analyzed with ImageJ software to score cells in the head of the animals co-expressing *ift-20* reporter with each TF.

Bioinformatic Analysis

For the differential expression heatmap, single-cell RNA-seq data from L2 larval stage were used (Cao et al., 2017). Cells were classified into four different categories: 1) serotonergic neurons if they expressed *tph-1* or *mod-5* genes, 2) dopaminergic (DA) neurons if they expressed *dat-1* or *cat-2* genes, 3) RIC neurons if they expressed *tbh-1* and *tdc-1* genes, and 4) RIM neurons if they expressed *tdc-1* but no *tbh-1* genes. Cells that fell into the serotonergic category were further subdivided into ADF neurons or NSM neurons by refinement of t-SNE coordinates. A total of 140 cells were used for the heatmap (15 ADF, 74 DA, 16 NSM, 10 RIC, 25 RIM). Genes with zero count across all the neuronal categories were excluded and z-scores were calculated to

normalize data. The heatmap and clustering was performed using the package
heatmaply (Galili et al. 2018).

TFs associated to neuronal phenotypes were obtained retrieving WormBase
annotated alleles associated to phenotypes matching the terms "neuron",
"nervous system" or "axon". For expression analysis, embryonic, L2 and L4 sc-
RNA-seq data (Cao et al. 2017; Packer et al. 2019; Taylor et al. 2020) and
fluorescence reporter lineage data (<https://epic.gs.washington.edu>) was used to
determine if TFs are expressed in each neuron type or its corresponding
lineage.

Unless otherwise indicated, all regulatory signature analyses were performed
using the software R (Zhang et al. 2014b) and packages from Bioconductor
(Huber et al. 2015). sc-RNA-seq data from (Cao et al. 2017) was downloaded
from the author's website (<http://atlas.gs.washington.edu/worm-rna/>). These
data correspond to nearly 50,000 cells coming from *Caenorhabditis elegans* at
the L2 larval stage. Cells with failed QC, doublets and unclassified cells were
filtered and excluded from subsequent analysis. Differential expression analysis
between dopaminergic neurons cluster and clusters for the other ciliated
sensory neurons was performed using Monocle (Qiu et al. 2017b, 2017a;
Trapnell et al. 2014) and results were filtered by q-value (≤ 0.05) in order to get
a list of differentially expressed genes in dopaminergic neurons. For gene
enrichment of neuronal clusters corresponding to non-dopaminergic neurons
(RIA, ASE, Touch receptor neurons, GABAergic neurons and ALN/PLN/SDQ
neurons) we followed a similar strategy and performed differential expression
tests between each neuronal cluster and all cells from the dataset annotated as
neurons (**Table S7**). Specificity of these six gene sets was checked by the

enrichment of its anatomical association in *C. elegans* using the web tool WormEnrichr (<https://amp.pharm.mssm.edu/WormEnrichr/>) (Chen et al. 2013). Gene lists for neuronal, non-neuronal, ubiquitous, panneuronal and panciliated categories were inferred from a more comprehensive sc-RNA-seq dataset (Packer et al. 2019). We retrieved gene expression data (log2 transcripts per million) from all the genes that were expressed in at least one annotated terminal cell bin, getting a final matrix of 15,813 genes x 409 terminal cell bins. Following authors' original approach, genes were ordered by hierarchical clustering and cell bins were ordered by tissues; resulting in differential gene clusters which marked sites of predominant expression [Figure S32 from (Packer et al. 2019)]. Using these data, we manually curated five gene lists: non-neuronal tissue-specific, ubiquitous, panneuronal, panciliated, and neuron-type specific expression (**Table S7**). For genes located in operons, only the gene located at the 5' end of the cluster, and thus subdued to *cis*-regulation, were considered for dopaminergic regulatory signature analysis. For hybrid operons, additional promoters were also included (Blumenthal, Davis & Garrido-Lecca, 2015). For genomic search of dopaminergic regulatory signature similar exclusion of downstream genes from operon was performed, removing a total of 2,083 genes from the analysis. The final curated gene lists used in this study are listed in **Table S7**.

For *C. elegans* regulatory signature analysis, we downloaded PWMs from Cis-BP version 1.02 (Weirauch et al. 2014) corresponding to the TF binding sites of the six transcription factors that compose the dopaminergic regulatory signature in *C. elegans*. If the exact match for *C. elegans* was not available, we selected the PWM from the *M. musculus* or *H. sapiens* orthologous TFs (HD, ref. M5340;

876 ETS, ref. M0709; MADS, ref. M6342; MEIS, ref. M6048; PAIRED, ref. M1500;
877 PBX, ref. M1898), plus an additional hybrid PAIRED HD site (represented as
878 HD*, ref. M6189). Following published methodology (Lloret-Fernández et al.
879 2018) we downloaded upstream and intronic gene regions of protein-coding
880 genes from WormBase version 262 and then classified genes using the gene
881 lists mentioned above. Upstream regions were trimmed to a maximum of 10 kb.
882 PWMs were aligned to genomic sequences and we retrieved matches with a
883 minimum score of 70%. To increase specificity, we removed all matches that
884 did not bear an exact consensus sequence for the corresponding TF family
885 (consensus sequence for HD: TAATT, ETS: VMGGAWR, MADS: WWWDTAG,
886 MEIS: DTGTCD, PAIRED: GGARSA, HD*: HTAATTR, PBX: GATNNAT).
887 Sliding window search with a maximum length of 700 bp was performed to find
888 regions that included at least one match for the 7 TF binding motifs, allowing
889 flexible motif composition. To assess signature enrichment in the set of
890 dopaminergic expressed genes 10,000 sets of 86 genes random genes were
891 built considering that: 1) they were not differentially expressed in dopaminergic
892 neurons, 2) at least one ortholog had been described in other *Caenorhabditis*
893 species (*C. brenneri*, *C. briggsae*, *C. japonica* or *C. remanei*), and 3) their
894 upstream and intronic regions were similar in length, on average, to those of the
895 86 dopamine-expressed genes (Mann-Whitney U test, p-value > 0.05). For
896 each one of the non-dopaminergic neuronal groups (RIA, ASE, Touch receptor
897 neurons, GABAergic neurons and ALN/PLN/SDQ neurons), similar sets of
898 random genes were built. We considered the enrichment in signature to be
899 significant when the percentile of the neuronal group in regard to the internal
900 random control was above 95. Differences between dopaminergic expressed

genes and other neuronal groups were assessed by Brunner-Munzel test performed with R package `brunnermunzel` (Toshiaki 2019), *: p-value < 0.05, **: p-value < 0.01, ***: p-value < 0.001).

Gene ontology analysis of the genes assigned dopaminergic regulatory signature were carried out using the web tool GOrilla (<http://cbl-gorilla.cs.technion.ac.il/>) (Eden et al. 2009), using *C. elegans* coding genome as control list.

Data access

Data used for the analysis and raw scorings are accessible in supplementary materials.

Competing interest statement

Authors declare no competing interests

Acknowledgments

We thank CGC (P40 OD010440) for providing strains. Elia García and Francisco Anguix for technical help. Ana Pilar Gómez-Escribano and Carla Lloret-Fernández for helping in the generation of the RNAi library. Bioinformatics and Biostatistics Unit from Principe Felipe Research Center (CIPF) for providing access to the cluster, co-funded by European Regional Development Funds (FEDER). Dr Peter Askjaer and Dr. Marta Artal for sharing reagents. Dr Guillermo Ayala for advice on statistical analysis. Dr Oscar Marín, Dr Beatriz Rico and Dr Luisa Cochella labs for scientific discussion. Dr Oscar Marín, Dr Luisa Cochella, Dr Inés Carrera, Dr Roger Pocock, Dr Arantza Barrios

926 and Dr Oliver Hobert for comments on the manuscript.

927 **Author Contributions**

928 A.J., N.D., M.M., R.B., conducted the experiments, E.S. performed the
929 bioinformatics analysis, N.F. designed the experiments and wrote the paper.

930

REFERENCES

- Agoston Z, Heine P, Brill MS, Grebbin BM, Hau AC, Kallenborn-Gerhardt W, Schramm J, Götz M, Schulte D. 2014. Meis2 is a Pax6 co-factor in neurogenesis and dopaminergic periglomerular fate specification in the adult olfactory bulb. *Dev.*
- Ambros V, Horvitz R. 1984. Heterochronic Mutants of the Nematode *Caenorhabditis elegans* Author (s): Victor Ambros and H . Robert Horvitz Reviewed work (s): Source : Science , New Series , Vol . 226 , No . 4673 (Oct . 26 , 1984), pp . 409-416 Published by : American Associatio. *Science* (80-) **226**: 409–416.
- Andrews MG, Kong J, Novitch BG, Butler SJ. 2019. New perspectives on the mechanisms establishing the dorsal-ventral axis of the spinal cord. In *Current Topics in Developmental Biology*.
- Angerer LM, Yaguchi S, Angerer RC, Burke RD. 2011. The evolution of nervous system patterning: Insights from sea urchin development. *Development*.
- Arendt D. 2008. The evolution of cell types in animals: Emerging principles from molecular studies. *Nat Rev Genet*.
- Arendt D, Bertucci PY, Achim K, Musser JM. 2019. Evolution of neuronal types and families. *Curr Opin Neurobiol*.
- Arendt D, Musser JM, Baker CVH, Bergman A, Cepko C, Erwin DH, Pavlicev M, Schlosser G, Widder S, Laubichler MD, et al. 2016. The origin and evolution of cell types. *Nat Rev Genet*.
- Arlotta P, Molyneaux BJ, Chen J, Inoue J, Kominami R, MacKlis JD. 2005. Neuronal subtype-specific genes that control corticospinal motor neuron development in vivo. *Neuron* **45**: 207–221.
- Babonis LS, Martindale MQ. 2017. PaxA, but not PaxC, is required for cnidocyte development in the sea anemone *Nematostella vectensis*. *Evodevo* **8**: 1–20.
- Baugh LR, Wen JC, Hill AA, Slonim DK, Brown EL, Hunter CP. 2005. Synthetic lethal analysis of *Caenorhabditis elegans* posterior embryonic patterning genes identifies conserved genetic interactions. *Genome Biol* **6**.
- Bellora N, Farré D, Mar MM. 2007. Positional bias of general and tissue-specific regulatory motifs in mouse gene promoters. *BMC Genomics* **8**: 1–13.
- Bertrand N, Castro DS, Guillemot F. 2002. Proneural genes and the specification of neural cell types. *Nat Rev Neurosci*.
- Blanchi B, Kelly LM, Viemari JC, Lafon I, Burnet H, Bévengut M, Tillmanns S, Daniel L, Graf T, Hilaire G, et al. 2003. MafB deficiency causes defective respiratory rhythmogenesis and fatal central apnea at birth. *Nat Neurosci* **6**: 1091–1099.
- Bodofsky S, Koitz F, Wightman B. 2017. Conserved and Exapted Functions of Nuclear Receptors in Animal Development. *Nucl Recept Res* **4**: 1–33.
- Borello U, Pierani A. 2010. Patterning the cerebral cortex: Traveling with morphogens. *Curr Opin Genet Dev*.
- Bovetti S, Bonzano S, Garzotto D, Giannelli SG, Iannielli A, Armentano M, Studer M, De Marchis S. 2013. COUP-TFI controls activity-dependent tyrosine hydroxylase expression in adult dopaminergic olfactory bulb interneurons. *Dev*.
- Boyle AP, Araya CL, Brdlik C, Cayting P, Cheng C, Cheng Y, Gardner K, Hillier LW, Janette J, Jiang L, et al. 2014. Comparative analysis of regulatory

information and circuits across distant species. *Nature*.

Brandt JP, Rossillo M, Du Z, Ichikawa D, Barnes K, Chen A, Noyes M, Bao Z, Ringstad N. 2019. Lineage context switches the function of a *C. elegans* Pax6 homolog in determining a neuronal fate. *Dev*.

Brenner S. 1974. The genetics of *Caenorhabditis elegans*. *Genetics*.

Brill MS, Snapyan M, Wohlfrom H, Ninkovic J, Jawerka M, Mastick GS, Ashery-Padan R, Saghatelian A, Berninger B, Götz M. 2008. A Dlx2- and Pax6-dependent transcriptional code for periglomerular neuron specification in the adult olfactory bulb. *J Neurosci*.

Briscoe J, Pierani A, Jessell TM, Ericson J. 2000. A homeodomain protein code specifies progenitor cell identity and neuronal fate in the ventral neural tube. *Cell*.

Campbell RF, Walthall WW. 2016. Meis/UNC-62 isoform dependent regulation of CoupTF-II/UNC-55 and GABAergic motor neuron subtype differentiation. *Dev Biol*.

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* (80-).

Cave JW, Akiba Y, Banerjee K, Bhosle S, Berlin R, Baker H. 2010. Differential regulation of dopaminergic gene expression by Er81. *J Neurosci*.

Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles G V., Clark NR, Ma'ayan A. 2013. Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*.

Choksi SP, Lauter G, Swoboda P, Roy S. 2014. Switching on cilia: Transcriptional networks regulating ciliogenesis. *Dev*.

Doitsidou M, Flames N, Lee AC, Boyanov A, Hobert O. 2008. Automated screening for mutants affecting dopaminergic-neuron specification in *C. elegans*. *Nat Methods*.

Doitsidou M, Flames N, Topalidou I, Abe N, Felton T, Remesal L, Popovitchenko T, Mann R, Chalfie M, Hobert O. 2013. A combinatorial regulatory signature controls terminal differentiation of the dopaminergic nervous system in *C. elegans*. *Genes Dev*.

Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. 2009. GOrilla: A tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*.

Farré D, Bellora N, Mularoni L, Messeguer X, Albà MM. 2007. Housekeeping genes tend to show reduced upstream sequence conservation. *Genome Biol* 8: 1–10.

Flames N, Hobert O. 2009. Gene regulatory logic of dopamine neuron differentiation. *Nature*.

Flames N, Hobert O. 2011. Transcriptional Control of the Terminal Fate of Monoaminergic Neurons. *Annu Rev Neurosci*.

Galili T, O'Callaghan A, Sidi J, Sievert C. 2018. Heatmaply: An R package for creating interactive cluster heatmaps for online publishing. *Bioinformatics* 34: 1600–1602.

Grossman SR, Zhang X, Wang L, Engreitz J, Melnikov A, Rogov P, Tewhey R, Isakova A, Deplancke B, Bernstein BE, et al. 2017. Systematic dissection of genomic features determining transcription factor binding and enhancer function. *Proc Natl Acad Sci U S A* 114: E1291–E1300.

Guillemot F, Hassan BA. 2017. Beyond proneural: emerging functions and

- regulations of proneural proteins. *Curr Opin Neurobiol*.
- Haider NB, Jacobson SG, Cideciyan A V., Swiderski R, Streb LM, Searby C, Beck G, Hockey R, Hanna DB, Gorman S, et al. 2000. Mutation of a nuclear receptor gene, NR2E3, causes enhanced S cone syndrome, a disorder of retinal cell fate. *Nat Genet*.
- Hoebert O. 2016. A map of terminal regulators of neuronal identity in *Caenorhabditis elegans*. *Wiley Interdiscip Rev Dev Biol*.
- Hoebert O. 2008. Regulatory logic of neuronal diversity: Terminal selector genes and selector motifs. *Proc Natl Acad Sci U S A*.
- Holst BD, Wang Y, Jones FS, Edelman GM. 1997. A binding site for Pax proteins regulates expression of the gene for the neural cell adhesion molecule in the embryonic spinal cord. *Proc Natl Acad Sci U S A*.
- Huang X, Tian E, Xu Y, Zhang H. 2009. The *C. elegans* engrailed homolog *ceh-16* regulates the self-renewal expansion division of stem cell-like seam cells. *Dev Biol*.
- Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto L, Girke T, et al. 2015. Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods*.
- Hynes M, Stone DM, Dowd M, Pitts-Meek S, Goddard A, Gurney A, Rosenthal A. 1997. Control of cell pattern in the neural tube by the zinc finger transcription factor and oncogene *Gli-1*. *Neuron* **19**: 15–26.
- Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R, Gotta M, Kanapin A, Le Bot N, Moreno S, Sohrmann M, et al. 2003. Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature*.
- Kamath RS, Martinez-Campos M, Zipperlen P, Fraser AG, Ahringer J. 2001. Effectiveness of specific RNA-mediated interference through ingested double-stranded RNA in *Caenorhabditis elegans*. *Genome Biol*.
- Kheradpour P, Ernst J, Melnikov A, Rogov P, Wang L, Zhang X, Alston J, Mikkelsen TS, Kellis M. 2013. Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. *Genome Res* **23**: 800–811.
- Kim DS, Matsuda T, Cepko CL. 2008. A core paired-type and POU homeodomain-containing transcription factor program drives retinal bipolar cell gene expression. *J Neurosci*.
- Kratsios P, Stolfi A, Levine M, Hoebert O. 2012. Coordinated regulation of cholinergic motor neuron traits through a conserved terminal selector gene. *Nat Neurosci*.
- Kuntz SG, Williams BA, Sternberg PW, Wold BJ. 2012. Transcription factor redundancy and tissue-specific regulation: Evidence from functional and physical network connectivity. *Genome Res* **22**: 1907–1919.
- Layden MJ, Boekhout M, Martindale MQ. 2012. *Nematostella vectensis* achaete-scute homolog *NvashA* regulates embryonic ectodermal neurogenesis and represents an ancient component of the metazoan neural specification pathway. *Development* **139**: 1013–1022.
- Liu A, Niswander LA. 2005. Bone morphogenetic protein signalling and vertebrate nervous system development. *Nat Rev Neurosci*.
- Liu H, Strauss TJ, Potts MB, Cameron S. 2006. Direct regulation of *egl-1* of programmed cell death by the Hox protein MAB-5 and CEH-20, a *C. elegans* homolog of *Pbx1*. *Development*.
- Liu Y, Yu C, Daley TP, Wang F, Cao WS, Bhate S, Lin X, Still C, Liu H, Zhao D,

- et al. 2018. CRISPR Activation Screens Systematically Identify Factors that Drive Neuronal Fate and Reprogramming. *Cell Stem Cell* **23**: 758-771.e8. <https://doi.org/10.1016/j.stem.2018.09.003>.
- Lloret-Fernández C, Maicas M, Mora-Martínez C, Artacho A, Jimeno-Martín Á, Chirivella L, Weinberg P, Flames N. 2018. A transcription factor collective defines the HSN serotonergic neuron regulatory landscape. *Elife*.
- Maglich JM, Sluder A, Guan X, Shi Y, McKee DD, Carrick K, Kamdar K, Willson TM, Moore JT. 2001. Comparison of complete nuclear receptor sets from the human, *Caenorhabditis elegans* and *Drosophila* genomes. *Genome Biol*.
- Mall M, Kareta MS, Chanda S, Ahlenius H, Perotti N, Zhou B, Grieder SD, Ge X, Drake S, Euong Ang C, et al. 2017. Myt1l safeguards neuronal identity by actively repressing many non-neuronal fates. *Nature*.
- Masserdotti G, Gascón S, Götz M. 2016. Direct neuronal reprogramming: Learning from and for development. *Dev*.
- Mears AJ, Kondo M, Swain PK, Takada Y, Bush RA, Saunders TL, Sieving PA, Swaroop A. 2001. Nrl is required for rod photoreceptor development. *Nat Genet*.
- Murray JI, Boyle TJ, Preston E, Vafeados D, Mericle B, Weisdepp P, Zhao Z, Bao Z, Boeck M, Waterston RH. 2012. Multidimensional regulation of gene expression in the *C. elegans* embryo. *Genome Res* **22**: 1282–1294.
- Narasimhan K, Lambert SA, Yang AWH, Riddell J, Mnaimneh S, Zheng H, Albu M, Najafabadi HS, Reece-Hoyes JS, Fuxman Bass JI, et al. 2015. Mapping and analysis of *Caenorhabditis elegans* transcription factor sequence specificities. *Elife*.
- Nuez I, Félix MA. 2012. Evolution of susceptibility to ingested double-stranded rnas in *Caenorhabditis* nematodes. *PLoS One*.
- Offenburger SL, Bensaddek D, Murillo AB, Lamond AI, Gartner A. 2017. Comparative genetic, proteomic and phosphoproteomic analysis of *C. elegans* embryos with a focus on ham-1/STOX and pig-1/MELK in dopaminergic neuron development. *Sci Rep*.
- Olsson-Carter K, Slack FJ. 2010. A developmental timing switch promotes axon outgrowth independent of known guidance receptors. *PLoS Genet*.
- Packer JS, Zhu Q, Huynh C, Sivaramakrishnan P, Preston E, Dueck H, Stefanik D, Tan K, Trapnell C, Kim J, et al. 2019. A lineage-resolved molecular atlas of *C. elegans* embryogenesis at single-cell resolution . *Science* (80-) **365**.
- Poole RJ, Bashllari E, Cochella L, Flowers EB, Hobert O. 2011. A Genome-Wide RNAi screen for factors involved in neuronal specification in *Caenorhabditis elegans*. *PLoS Genet*.
- Potts MB, Wang DP, Cameron S. 2009. Trithorax, Hox, and TALE-class homeodomain proteins ensure cell survival through repression of the BH3-only gene *egl-1*. *Dev Biol*.
- Qiu M, Bulfone A, Martinez S, Meneses JJ, Shimamura K, Pedersen RA, Rubenstein JLR. 1995. Null mutation of *Dlx-2* results in abnormal morphogenesis of proximal first and second branchial arch derivatives and abnormal differentiation in the forebrain. *Genes Dev*.
- Qiu X, Hill A, Packer J, Lin D, Ma YA, Trapnell C. 2017a. Single-cell mRNA quantification and differential analysis with Census. *Nat Methods*.
- Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, Trapnell C. 2017b.

- 1131 Reversed graph embedding resolves complex single-cell trajectories. *Nat*
1132 *Methods*.
- 1133 Reece-Hoyes JS, Deplancke B, Shingles J, Grove CA, Hope IA, Walhout AJM.
1134 2005. A compendium of *Caenorhabditis elegans* regulatory transcription
1135 factors: A resource for mapping transcription regulatory networks. *Genome*
1136 *Biol* **6**.
- 1137 Reilly MB, Cros C, Varol E, Yemini E, Hobert O. 2020. Unique homeobox codes
1138 delineate all *C. elegans* neuron classes. *Press* 1–28.
1139 <http://dx.doi.org/10.1038/s41586-020-2618-9>.
- 1140 Remesal L, Roger-Banyat I, Chirivella L, Maicas M, Brocal-Ruiz R, Pérez-
1141 Villalva A, Cucarella C, Casado M, Flames N. PBX1 is a terminal selector
1142 of olfactory bulb dopaminergic neurons. *Dev*.
- 1143 Remesal L, Roger-Baynat I, Chirivella L, Maicas M, Brocal-Ruiz R, Pérez-
1144 Villalva A, Cucarella C, Casado M, Flames N. 2020. PBX1 acts as terminal
1145 selector for olfactory bulb dopaminergic neurons. *Development*
1146 dev.186841.
- 1147 Rentzsch F, Layden M, Manuel M. 2017. The cellular and molecular basis of
1148 cnidarian neurogenesis. *Wiley Interdiscip Rev Dev Biol* **6**: 1–19.
- 1149 Roy K, Kuznicki K, Wu Q, Sun Z, Bock D, Schutz G, Vranich N, Monaghan AP.
1150 2004. The *tlx* gene regulates the timing of neurogenesis in the cortex. *J*
1151 *Neurosci*.
- 1152 Rual JF, Ceron J, Koreth J, Hao T, Nicot AS, Hirozane-Kishikawa T,
1153 Vandenhoute J, Orkin SH, Hill DE, van den Heuvel S, et al. 2004. Toward
1154 improving *Caenorhabditis elegans* phenome mapping with an ORFeome-
1155 based RNAi library. *Genome Res*.
- 1156 Serrano-Saiz E, Poole RJ, Felton T, Zhang F, De La Cruz ED, Hobert O. 2013.
1157 XModular control of glutamatergic neuronal identity in *C. elegans* by
1158 distinct homeodomain proteins. *Cell*.
- 1159 Shimizu T, Nakazawa M, Kani S, Bae YK, Shimizu T, Kageyama R, Hibi M.
1160 2010. Zinc finger genes *Fezf1* and *Fezf2* control neuronal differentiation by
1161 repressing *Hes5* expression in the forebrain. *Development* **137**: 1875–
1162 1885.
- 1163 Shirasaki R, Pfaff SL. 2002. Transcriptional Codes and the Control of Neuronal
1164 Identity. *Annu Rev Neurosci*.
- 1165 Shubin N, Tabin C, Carroll S. 2009. Deep homology and the origins of
1166 evolutionary novelty. *Nature*.
- 1167 Simmer F, Moorman C, Van Der Linden AM, Kuijk E, Van Den Berghe PVE,
1168 Kamath RS, Fraser AG, Ahringer J, Plasterk RHA. 2003. Genome-wide
1169 RNAi of *C. elegans* using the hypersensitive *rff-3* strain reveals novel gene
1170 functions. *PLoS Biol*.
- 1171 Singhvi A, Frank CA, Garriga G. 2008. The T-box gene *tbx-2*, the homeobox
1172 gene *egl-5* and the asymmetric cell division gene *ham-1* specify neural fate
1173 in the HSN/PHB lineage. *Genetics* **179**: 887–898.
- 1174 Stefanakis N, Carrera I, Hobert O. 2015. Regulatory Logic of Pan-Neuronal
1175 Gene Expression in *C. elegans*. *Neuron*.
- 1176 Stegmaier P, Kel AE, Wingender E. 2004. Systematic DNA-binding domain
1177 classification of transcription factors. *Genome Inform*.
- 1178 Swoboda P, Adler HT, Thomas JH. 2000. The RFX-type transcription factor
1179 DAF-19 regulates sensory neuron cilium formation in *C. Elegans*. *Mol Cell*.
- 1180 Sze JY, Zhang S, Li J, Ruvkun G. 2002. The *C. elegans* POU-domain

transcription factor UNC-86 regulates the tph-1 tryptophan hydroxylase gene and neurite outgrowth in specific serotonergic neurons. *Development*.
 Taubert S, Ward JD, Yamamoto KR. 2011. Nuclear hormone receptors in nematodes: Evolution and function. *Mol Cell Endocrinol*.
 Taylor SR, Santpere G, Weinreb A, Barrett A, Reilly MB, Xu C, Varol E, Oikonomou P, Glenwinkel L, McWhirter R, et al. 2020. Molecular topography of an entire nervous system. *bioRxiv*.
 Tewhey R, Kotliar D, Park DS, Liu B, Winnicki S, Reilly SK, Andersen KG, Mikkelsen TS, Lander ES, Schaffner SF, et al. 2016. Direct identification of hundreds of expression-modulating variants using a multiplexed reporter assay. *Cell* **165**: 1519–1529. <https://doi.org/10.1016/j.cell.2016.04.027>.
 Thor S, Andersson SGE, Tomlinson A, Thomas JB. 1999. A LIM-homeodomain combinatorial code for motor- neuron pathway selection system , although expression was also observed in precursors of the. *Nature* **397**: 76–80.
 Toshiaki A. 2019. Brunnermunzel: (Permuted) Brunner-Munzel Test. *R Packag version 135* <https://CRAN.R-project.org/package=brunnermunzel>.
 Tournière O, Dolan D, Richards GS, Sunagar K, Columbus-Shenkar YY, Moran Y, Rentzsch F. 2020. NvPOU4/Brain3 Functions as a Terminal Selector Gene in the Nervous System of the Cnidarian Nematostella vectensis. *Cell Rep* **30**: 4473–4489.e5.
 Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL. 2014. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol*.
 Van Auken K, Weaver D, Robertson B, Sundaram M, Saldi T, Edgar L, Elling U, Lee M, Boese Q, Wood WB. 2002. Roles of the homothorax/Meis/Prep homolog UNC-62 and the Exd/Pbx homologs CEH-20 and CEH-40 in C. elegans embryogenesis. *Development*.
 Wacław RR, Allen ZJ, Bell SM, Erdélyi F, Szabó G, Potter SS, Campbell K. 2006. The zinc finger transcription factor Sp8 regulates the generation and diversity of olfactory bulb interneurons. *Neuron* **49**: 503–516.
 Walton T, Preston E, Nair G, Zacharias AL, Raj A, Murray JI. 2015. The Bicoid Class Homeodomain Factors ceh-36/OTX and unc-30/PITX Cooperate in C. elegans Embryonic Progenitor Cells to Regulate Robust Development. *PLoS Genet* **11**: 1–28.
 Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, Najafabadi HS, Lambert SA, Mann I, Cook K, et al. 2014. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell*.
 Wende H, Lechner SG, Cheret C, Bourane S, Kolanczyk ME, Pattyn A, Reuter K, Munier FL, Carroll P, Lewin GR, et al. 2012. The transcription factor c-Maf controls touch receptor development and function. *Science (80-)* **335**: 1373–1376.
 Zeng H, Sanes JR. 2017. Neuronal cell-type classification: Challenges, opportunities and the path forward. *Nat Rev Neurosci*.
 Zetterström RH, Solomin L, Jansson L, Hoffer BJ, Olson L, Perlmann T. 1997. Dopamine neuron agenesis in Nurr1-deficient mice. *Science (80-)*.
 Zhang F, Bhattacharya A, Nelson JC, Abe N, Gordon P, Lloret-Fernandez C, Maicas M, Flames N, Mann RS, Colón-Ramos DA, et al. 2014a. The LIM and POU homeobox genes ttx-3 and unc-86 act as terminal selectors in distinct cholinergic and serotonergic neuron types. *Dev*.

1231 Zhang Z, Ersoz E, Lai C-Q, Todhunter RJ, Tiwari HK, Gore M a, Bradbury PJ,
1232 Yu J, Arnett DK, Ordovas JM, et al. 2014b. R: A language and environment
1233 for statistical computing. R Foundation for Statistical Computing, Vienna,
1234 Austria. URL <http://www.R-project.org/>. *Nat Genet*.
1235 Zhao C, Emmons SW. 1995. A transcription factor controlling development of
1236 peripheral sense organs in *C. elegans*. *Nature* **373**: 74–78.
1237 Zheng C, Karimzadegan S, Chiang V, Chalfie M. 2013. Histone Methylation
1238 Restrains the Expression of Subtype-Specific Genes during Terminal
1239 Neuronal Differentiation in *Caenorhabditis elegans*. *PLoS Genet*.
1240 Zheng X, Chung S, Tanabe T, Sze JY. 2005. Cell-type specific regulation of
1241 serotonergic identity by the *C. elegans* LIM-homeodomain factor LIM-4.
1242 *Dev Biol*.
1243