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 motorneurons 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

51 specification (Briscoe et al. 2000; Shirasaki and Pfaff 2002; Thor et al. 1999)
52 and 4) the terminal selector model for neuronal terminal differentiation, in which,
53 at the lasts steps of development specific TFs, termed terminal selectors,
54 directly co-regulate expression of most neuron-type specific effector genes
55 (Hobert, 2008).
56 Here we took advantage of the amenability of *Caenorhabditis elegans* for
57 unbiased large-scale screens to study the complexity of neuronal gene
58 regulatory networks and to identify common principles underlying the
59 specification of different types of neurons. To this end, we performed an RNA
60 interference (RNAi) screen against all 875 transcription factors (TFs) encoded
61 by the *C. elegans* genome and systematically assessed their contribution in the
62 specification of nine different types of neurons of the monoaminergic (MA)
63 superclass and two cholinergic motorneurons. We focused mainly on MA
64 neurons not only because they are evolutionary conserved and clinically
65 relevant in humans (Flames and Hobert 2011), but also because the MA
66 superclass comprises a set of neuronal types with very diverse developmental
67 origins and functions in both worms and humans (Flames and Hobert 2011).
68 Importantly, we have previously shown that gene regulatory networks directing
69 the terminal fate of two types of MA neurons (dopaminergic and serotonergic
70 neurons) are conserved in worms and mammals (Lloret-Fernández et al. 2018;
71 Doitsidou et al. 2013; Flames and Hobert 2009; Remesal et al. 2020). Thus, the
72 identification of common principles underlying MA specification could help
73 unravel general rules for neuron-type specification.
74 Our results, that combine both RNAi and mutant analysis, unveil five major
75 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.

92 **RESULTS**

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

95 To increase our understanding of the gene regulatory networks controlling
96 neuron-type specification used an RNAi library against all the putative TFs
97 encoded in the *C. elegans* genome. For RNAi library preparation we followed
98 (Narasimhan et al. 2015) complete *C. elegans* list of 875 TFs (**Table S1**), which
99 include 745 TFs present in CisBP 0.9 data base, 18 additional TFs manually
100 curated from (Reece-Hoyes et al. 2005) TFv2 list not present in CisBP but
101 supported by PFAM and SMART databases, plus 112 medium-confidence TFs
102 from TFv2 list which show partial evidence for TF activity (Narasimhan et al.,
103 2015). RNAi clones were obtained either from the two available whole genome
104 RNAi libraries (Rual et al. 2004; Kamath et al. 2003) or were newly generated
105 (See Material and Methods and **Table S1**).

106 We used the *rrf-3(pk1426)* mutant background to sensitize neurons for RNAi
107 effects as previously done (Simmer et al. 2003) and combined this mutation
108 with three different fluorescent reporters that label the MA system in the worm
109 (**Figure 1A**): the vesicular monoamine transporter*, cat-1::gfp (otIs224),*
110 expressed in all MA neurons; the dopamine transporter, *dat-1::mcherry*
111 *(otIs181),* expressed in dopaminergic neurons only; and the tryptophan
112 hydroxylase enzyme *tph-1::dsred* (*vsIs97),* expressed in serotonergic neurons
113 only. Altogether our strategy labels nine different MA neuronal classes (ADE,
114 CEPV, CEPD, NSM, ADF, RIC, RIM, HSN, PDE) and the two cholinergic VC4
115 and VC5 motoneurons, which are not MA but express *cat-1* reporter for
116 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**

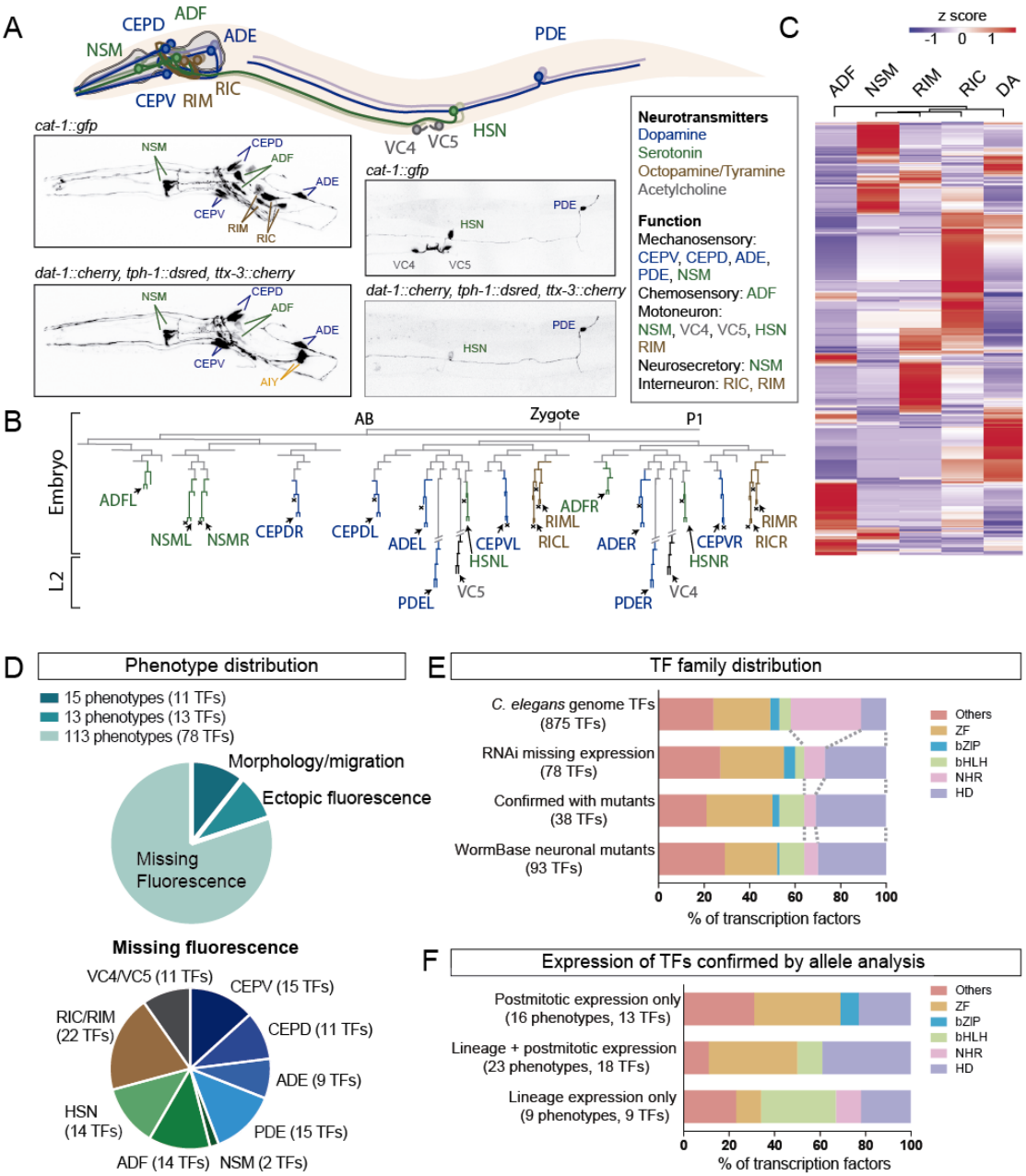

Figure 1

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 tyraminergic 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.

124 **1C**). Of note, the dopaminergic (DA) system, composed by four anatomically
125 defined neuronal types (CEPV, CEPD, ADE and PDE), constitutes an exception
126 to the MA-system diversity. Although dopaminergic neurons are
127 developmentally diverse, they are functionally and molecularly homogeneous
128 and share their terminal differentiation program (Doitsidou et al. 2013; Flames
129 and Hobert 2009) (**Figure 1C**). In summary, considering the diversity of labeled
130 neurons we reasoned their global study could unravel shared principles of *C.*
131 *elegans* neuron specification.
132 Based on reporter expression in worms fed with negative control RNAi (L4440
133 empty vector), we set an initial threshold of 10% penetrance in at least two out
134 of three replicates to consider a positive hit. Under these conditions, we found
135 141 different phenotypes assigned to 91 of the 875 TFs (**Table S2**). Missing or

136 reduced reporter expression was the most frequent phenotype although we also
137 observed ectopic fluorescent cells and migration, axon guidance or morphology
138 defects (**Figure 1D** and **Table S2**).
139 We retrieved phenotypes for 24 out of 31 known regulators of MA neuron fate,
140 including TFs involved in MA lineage specification, axon guidance and known
141 MA terminal selectors (**Table S3**). Thus, we estimated a false negative rate of
142 around 23%. To estimate the false positive rate of the screen we decided to
143 focus on the validation of the set of 38 TFs displaying missing fluorescence
144 phenotypes with penetrance higher than 20% (comprising 59 phenotypes
145 considering each neuron type individually, **Table 1** and **Table S2**). We
146 combined published data for 14 TFs and new mutant analysis for 19 additional
147 TFs. 40 out of 46 tested RNAi phenotypes were verified by the corresponding
148 mutant (**Table 1** and **Table S4**) estimating a false positive rate of 14% for RNAi
149 clones producing missing phenotypes of >20% penetrance. See **Table 2** for a
150 summary of RNAi screen hits, validations and error rates.
151 In total, 38 mutant-confirmed TFs (20 already known, 11 newly characterized by
152 us and 7 false negative TFs) produce 52 missing reporter expression
153 phenotypes affecting different MA and VC4/VC5 neurons. Available single-cell
154 RNAseq or reporter expression data (Reilly et al. 2020; Cao et al. 2017; Packer
155 et al. 2019; Murray et al. 2012) shows detectable TF expression either in the
156 affected neuron or the corresponding developmental lineage for 48 of the 52
157 phenotypes (92%) (**Table S5**), suggesting a cell autonomous role for most of
158 these TFs in neuron specification.
159 We observed that each neuron type is affected by 9 to 15 different TF RNAi
160 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
lin-32	bHLH	PDE	90	Yes	*e1926, gm239*	(Zhao and Emmons, 1995)
lag-1	CSL	ADF	86	Yes	*om13 q385*	This work
unc-4	HD	VCs	83	Yes	e120, e2323	(Zheng et al., 2013)
lin-28	CSD	PDE	82	Yes	*n719*	This work and (Ambros and Horvitz 1984)
lin-39	HD	VCs	78	Yes	*n1760*	(Potts et al., 2009)
lin-40	ZF - GATA	VCs	73	Yes	*ku285*	This work
tra-1	ZF - C2H2	VCs	67	Yes	*e1488*	This work
ceh-20	HD	VCs	65	Yes	*ay42*	(Liu et al., 2006)
sem-2	HMG box	HSN	65	Not tested		
hlh-14	bHLH	ADF	64	Yes	*ok780, gm34*	This work
unc-86	HD	HSN	64	Yes	*n846*	(Lloret-Fernandez et al., 2018; Sze et al., 2002)
hbl-1	ZF - C2H2	PDE	63	Yes	*mg285*	This work
ham-1	WH	CEPV	54	Yes	*gt1984, ot339*	(Offenburger et al., 2017)
ham-1	WH	CEPD	51	Yes	*gt1984, ot339*	(Offenburger et al., 2017)
nob-1	HD	HSN	51	No	*os6*	This work
pax-3	HD	VCs	50	Not tested		
egl-5	HD	HSN	48	Yes	*n945*	(Singhvi et al., 2008)
sex-1	ZF - NHR	HSN	43	Yes	*y263*	This work
hbl-1	ZF - C2H2	HSN	41	Yes	*mg285*	This work
ceh-43	HD	CEPV	40	Yes	*ot406, ot340, tm480*	(Doitsidou et al., 2013)
ceh-43	HD	CEPD	40	Yes	*ot406, ot340, tm480*	(Doitsidou et al., 2013)
vab-15	HD	PDE	40	Yes	*u781*	This work
unc-62	HD	ADE	39	Yes	*e917*	This work
unc-62	HD	PDE	39	Yes	*e917*	This work
ast-1	WH - ETS	PDE	38	Yes	ot417 , hd1, rh300, gk463	(Flames and Hobert, 2009)
ast-1	WH - ETS	HSN	38	Yes	ot417 , hd1, rh300, gk463	(Lloret-Fernandez et al., 2018)
vab-3	HD	CEPV	38	Yes	ot266,ot292, ot346	(Doitsidou et al., 2008) and this work
ceh-20	HD	ADE	36	Yes	*mu290, ay42*	(Doitsidou et al., 2013)
ast-1	WH - ETS	ADE	35	Yes	ot417 , hd1, rh300, gk463	(Flames and Hobert, 2009)
ceh-20	HD	PDE	35	Yes	*mu290, ay42*	(Doitsidou et al., 2013)
bed-3	ZF - BED	PDE	33	Yes	*gk996*	This work
lin-14	Unknown	VCs	33	Yes	*n179*	This work
zip-5	bZIP	RIC/RIM	33	Yes	*gk646*	This work
vab-3	HD	ADE	33	Yes	ot346	This work
lin-14	Unknown	HSN	32	Yes	*e912,n179*	(Olsson-Carter and Slack, 2010)
sex-1	ZF - NHR	ADF	31	No	*y263*	This work
ceh-43	HD	ADE	29	Yes	*ot406, ot340, tm480*	(Doitsidou et al., 2013)
ceh-43	HD	PDE	29	Yes	*ot406, ot340, tm480*	(Doitsidou et al., 2013)
lin-28	CSD	HSN	29	Yes	*n719*	(Olsson-Carter and Slack, 2010)
ref-2	ZF - C2H2	VCs	29	Not tested		
unc-86	HD	NSM	29	Yes	*n846*	(Sze et al., 2002)
lin-14	Unknown	NSM	28	No	*n179*	This work
sem-4	ZF - C2H2	HSN	28	Yes	*n2654*	(Lloret-Fernandez et al., 2018)
zip-10	bZIP	ADF	28	No	*ok3462*	This work
ceh-44	HD	PDE	27	No	*tm919*	This work
egl-18	ZF - GATA	HSN	26	Yes	*ok290*	(Lloret-Fernandez et al., 2018)
C32E8.1	bZIP	CEPD	25	Not tested		
nhr-61	ZF - NHR	VCs	25	Not tested		
C32E8.1	bZIP	CEPV	24	Not tested		
lin-26	ZF - C2H2	PDE	23	Yes	*n156*	This work
vab-3	HD	PDE	23	Yes	ot346	This work
hlh-3	bHLH	HSN	23	Yes	*tm1688*	(Lloret-Fernandez et al., 2018)
lim-4	HD	ADF	23	Yes	*yz3, yz12*	(Zheng et al., 2005)
nhr-2	ZF - NHR	RIC/RIM	22	Yes	*tm860*	This work
C32E8.1	bZIP	PDE	22	Not tested		
dro-1	CBF	CEPD	22	No	*tm4702*	This work
C32E8.1	bZIP	ADE	21	Not tested		
ceh-12	HD	RIC/RIM	21	No	*tm1619*	This work
pqn-21	ZF - C2H2	ADF	21	Not tested		

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

166 suggesting the absence of global regulators of MA fate, despite the fact that all
167 MA neurons share *cat-1*/VMAT expression. Alternatively, global regulators of
168 MA fate could show redundant functions or produce early embryonic lethalities
169 and thus would not be recovered from our RNAi screen.
170 In summary, our RNAi screen followed by mutant allele validation identified a
171 new role for 11 TFs in MA and VC4/VC5 neuron specification (**Table 1, Table 2**
172 and **Table S4**), including a role for *zip-5* Basic Leucine Zipper Domain (bZIP)
173 and *nhr-2* Nuclear Hormone Receptor (NHR) in RIC and RIM specification. To
174 our knowledge, these are the first TFs known to participate in *C. elegans*
175 tyraminergic 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%

176

177 **A specific set of transcription factor families controls neuron specification**
178 Next, to try to infer common rules for neuron specification, we focused on the
179 113 missing reporter expression RNAi phenotypes (assigned to 78 RNAi
180 clones) and analyzed the screen results based on TF families instead of
181 individual TF members. According to their DNA binding domain, *C. elegans* TFs
182 can be classified into more than fifty different TF families (Stegmaier et al. 2004;
183 Narasimhan et al. 2015) (**Table S1**). bHLH, HD, ZF, bZIP and NHR families
184 comprise 75% of the TFs in the *C. elegans* genome. RNAi clones targeting
185 these families also generate 75% of the missing reporter phenotypes (**Figure**

186 **1E**). However, we noticed that the prevalence of two TF families, HD and NRH
187 TFs, largely differs from what would be expected from the number of TF
188 members encoded in the genome (**Figure 1E**). HD family, in spite of
189 representing only 12% of total TFs in the *C. elegans* genome, corresponds to
190 27% of RNAi clones showing a phenotype (21/78) ($p<0.005$ Fisher exact test).
191 In contrast, NHR TFs are under-represented when considering the total number
192 of NHR TFs encoded in the genome (**Figure 1E**). We found that NHR-TF RNAi
193 clones represent only 9% of missing reporter phenotypes, even though 30% of
194 all *C. elegans* TFs belong to this family ($p<0.0001$, Fisher exact test). A roughly
195 similar TF family distribution is observed when considering MA neuron types
196 individually (**Supplementary Figure 2**). The specific TF family distribution found
197 in our RNAi screen could be biased by our false positive and negative rates,
198 thus we focused only on the set of 38 TFs confirmed by mutant analysis.
199 Mutant-verified TFs show a very similar TF family distribution, with 30% HD TFs
200 and 5% NHR TFs ($p<0.0001$, $p=0.0017$ respectively compared to whole
201 genome, Fisher exact test), suggesting a prevalent role for HD family but not for
202 NHRs in neuron specification (**Figure 1E**).
203 Next, to explore if this TF family distribution is particular to specification of the
204 MA system and VC4, VC5 neurons or it could generally apply to neuron
205 specification networks, we analyzed the TF family distribution of the 93 *C.*
206 *elegans* TFs for which mutant alleles display any neuronal phenotype according
207 to WormBase data (neuron specification defects, migration or axon guidance)
208 (**Table S6**). Of note, TF family distribution for these 93 TFs is highly similar to
209 the one observed in our RNAi screen and our mutant analysis (**Figure 1E**)

210 suggesting it could constitute a general rule of neuron-type specification
211 programs.
212 NHR TF family has expanded in *C. elegans,* where it is composed by 272
213 members compared to less than 50 NRH TFs encoded in the human genome.
214 Only 8% of the 272 *C. elegans* NHRs have orthologs in non-nematode species
215 (Maglich et al. 2001; Taubert et al. 2011; Bodofsky et al. 2017). Importantly, we
216 found phylogenetically conserved NHRs are enriched for neuronal functions
217 both in the RNAi screen (3 out of 7 NHR TFs) and in the global WormBase
218 neuronal mutants (5 out of 6 NHR TFs). This observation suggests that, among
219 NHR members, those phylogenetically conserved could have a prevalent role in
220 neuron specification. Alternatively, lack of phenotypes for nematode specific
221 NHRs could be attributed to more redundant actions among them.
222 Defects either in early lineage specification or in neuronal terminal
223 differentiation can produce similar reporter expression defects, thus early and
224 late functions of TFs cannot be discerned *a priori.* In principle, TFs expressed at
225 early stages in development but absent from the postmitotic differentiating
226 neuron are more likely to affect lineage specification while TFs with onset of
227 expression at the postmitotic neuroblast might be involved in terminal
228 differentiation. As already mentioned 48 of the 52 missing-reporter phenotypes
229 verified by mutant alleles show detectable expression for the correspoding TF
230 either in the affected neuron or the corresponding developmental lineage. We
231 divided TF expression data into three categories: expressed only in the
232 postmitotic cell, expressed only in the lineage prior to the last division or both
233 (**Figure 1F** and **Table S5**). Interestingly, TF families are differentially distributed
234 among categories. bHLH TF expression is mostly restricted to lineage

235 expression only, what is in agreement with known early but not late roles of
236 bHLH proneural factors (Guillemot and Hassan 2017). On the contrary, most
237 HD TFs are expressed both in the lineage progenitors as well as in the
238 postmitotic neuron, what could indicate dual roles for this TF family both as
239 early lineage determinants and as terminal selectors.
240 In summary, our results show a stereotypic TF family distribution associated to
241 specification of different neuronal types, with high prevalence of HD TFs.
242 Moreover, we find differential early or late expression patterns for these families
243 that suggest they could fulfill different roles in neuron specification.
244
245 **Identification of new transcription factors involved in dopaminergic**
246 **terminal differentiation**
247 Our RNAi screen retrieved known and novel TFs likely required at different
248 steps of neuron specification and differentiation. Next, we aimed to use our TF
249 RNAi screen to specifically study the complexity of terminal differentiation
250 programs. Terminal differentiation is regulated by specific TFs, termed terminal
251 selectors, that in the immature postmitotic neuron directly activate the
252 expression of effector genes such as ion channels, neurotransmitter receptors,
253 biosynthesis enzymes, etc, that define the neuron-type specific transcriptome
254 (Hobert 2008). To date, at least one terminal selector has been assigned to 76
255 of the 118 *C. elegans* hermaphrodite neuron types (Hobert 2016). TFs,
256 including terminal selectors, act in combinations to activate target enhancers, to
257 date, a maximum of three terminal selectors have been identified to act together
258 in the regulation of a neuron-type. An exception to this rule is the HSN
259 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

285 (**Table S2**). To prioritize among these ten TFs, we first focused on *unc-*
286 *62*/MEIS-HD and *vab-3*/PAIRED-HD that show high penetrant RNAi phenotypes
287 that we had already confirmed by mutant analysis (**Table 1**).
288
289 ***unc-62*/MEIS-HD and *vab-3*/PAIRED-HD TFs have dual roles in**
290 **dopaminergic lineage specification and terminal differentiation**
291 We used mutant alleles to further characterize the role of these TFs in
292 dopaminergic terminal fate. *unc-62* has multiple functions in development and
293 null alleles are embryonic lethal precluding analysis of dopaminergic
294 differentiation defects (Van Auken et al. 2002). Three viable hypomorphic
295 alleles *e644*, *mu232* and *e917* show expression defects of a *cat-2*/tyrosine
296 hydroxylase reporter, the rate-limiting enzyme for dopamine synthesis (**Figure**
297 **2B, E** and **Supplementary Figure 4**). Of note, in addition to missing
298 expression, *unc-62(e644)* allele displays also a small percentage of ectopic
299 CEPs (**Supplementary Figure 4**), similar to RNAi experiments (**Table S2**),
300 suggesting *unc-62* could have a role in CEPs lineage determination. For further
301 characterization, we focused on *unc-62(e917)* allele as it showed higher
302 penetrance of missing reporter expression and no ectopic cells. *unc-62(e917)*
303 mutant shows broad defects in expression of the dopamine pathway genes
304 (*bas-1/dopamine decarboxylase, cat-2/tyrosine hydroxylase, cat-4/GTP*
305 *cyclohydrolase* and *cat-1/vesicular MA transporter* reporter) in ADE and PDE
306 while CEPV shows small but significant defects in *cat-2* and *cat-1* reporter
307 expression (**Figure 2C-G**). Expression of additional effector genes not directly
308 related to dopaminergic biosynthesis, such as the ion channel *asic-1*, is also
309 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

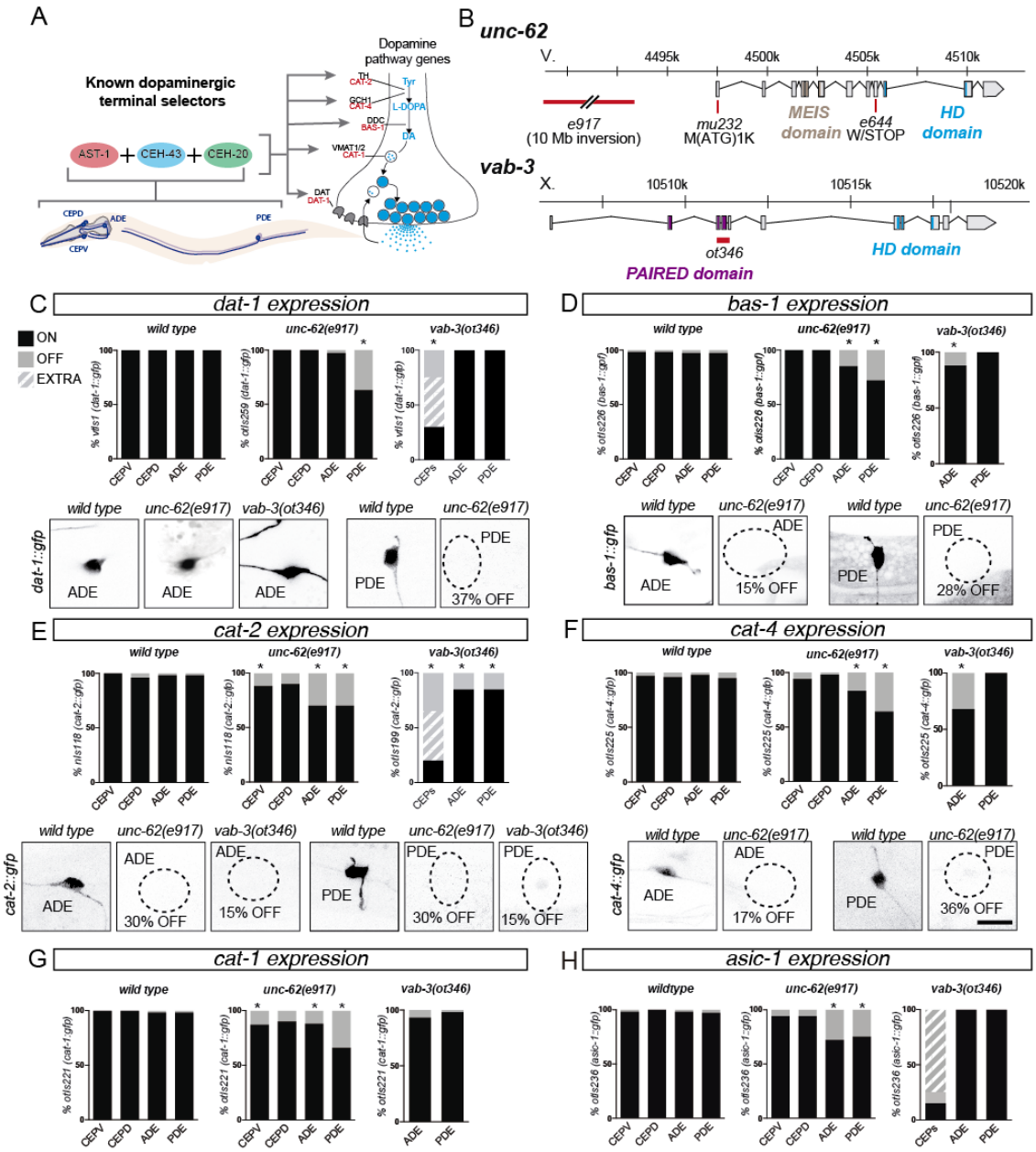

Figure 2

Figure 2. *unc-62*/MEIS HD and *vab-3*/PAIRED HD are required for correct 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 µm.

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

338 (**Figure 2D, E, F**). Importantly, *dat-1* and *cat-1* reporter expression is unaffected
339 in ADE and PDE which strongly suggests the missing expression phenotype in
340 these two neuron types is not due to lineage specification defects (**Figure 2C,**
341 **G**). We confirmed CEP lineage defects in *vab-3(ot346)* mutants with the
342 analysis of *gcy-36* and *mod-5* expression, effector genes expressed in URX
343 (sister cell of CEPD) and AIM (cousin of CEPV) respectively. We find both are
344 ectopically expressed in *vab-3(ot346)* animals (**Supplementary Figure 4**).
345 Thus, similar to *unc-62*, *vab-3* seems to have a dual role in dopaminergic
346 specification: it is required for proper ADE and PDE terminal differentiation and
347 for correct CEP lineage generation. A potential role for *vab-3* in CEPs terminal
348 differentiation could be masked by its earlier requirement in lineage
349 determination. Interestingly, adult worms maintain *vab-3* expression in CEPS
350 while adult ADE and PDE expression has not been reported (Reilly et al. 2020).
351

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

354 Our results expand to five the number of TFs involved in dopaminergic terminal
355 differentiation with strong predominance of HD TFs: *ast-1*/ETS TF, *ceh-43*/HD,
356 *ceh-20*/HD-PBX, *unc-62*/HD-MEIS and *vab-3*/HD-PAIRED. However, RNAi for
357 additional non-HD TFs displayed reporter expression defects in several
358 dopaminergic neuron types (**Table S2**) albeit at lower penetrance. We analyzed
359 dopamine pathway gene reporter expression in *cep-1(ep347*) P53 TF, *dro-*
360 *1(tm4702)* CBF TF, *dnj-17(ju1234)* ZF TF, *mef-2(gk633)* MADS TF, *nhr-*
361 *223(tm1172) and unc-55(e1170)* NHR TFs. A small but significant decrease in
362 *cat-1* reporter expression in ADE neuron was found in *mef-2(gk633)* mutants

363 (**Figure 3A, B**). *mef-2* fosmid reporter analysis shows broad neuronal
364 expression at first larval and adult stages that include all dopaminergic neurons
365 (**Figure 3C** and **Supplementary Figure 5**). However, the expression of other
366 dopamine pathway genes is unaffected in *mef-2(gk633)* (**Supplementary**
367 **Figure 5**) and *cat-1* reporter expression defects were not observed using an
368 additional allele *mef-2(gv1)* (**Supplementary Figure 5**). To rule out *mef-*
369 *2(gk633) cat-1* reporter expression defects are due to an unknown background
370 mutation we drove *mef-2* cDNA expression under the *bas-1* monoaminergic
371 promoter what we found it was sufficient to rescue *cat-1* expression defects in
372 the ADE (**Figure 3B**). Although the mechanisms underlying the discrepancy
373 between *gk633 and gv1* alleles is uncertain, *mef-2(gk633)* deletes the promoter
374 and first exon completely abolishing transcription, while *mef-2(gv1)* produce a
375 truncated transcript that could still display some functionality (**Figure 3A**).
376 Next, considering the extremely subtle phenotype of *mef-2(gk633)* mutants we
377 wondered if it has any biological significance. In our previous characterization of
378 the HSN terminal regulatory program we found extensive redundancy among
379 TFs (Lloret-Fernández et al. 2018). Thus we hypothesized that other TFs could
380 compensate for the lack of *mef-2,* masking its role in DA specification. To test
381 this hypothesis we performed double mutant analysis of *mef-2(gk633)* with a
382 hypomorphic missense mutation for *ast-1,* the main dopaminergic terminal
383 selector (Flames and Hobert 2009). We found that *mef-2(gk633)* and *ast-1(hd1)*
384 act synergistically in the regulation of *cat-1::gfp* expression not only in ADE but
385 also both in CEPD and CEPV (**Figure 3D**) demonstrating a *bona fide*
386 requirement for *mef-2* in dopaminergic terminal specification.

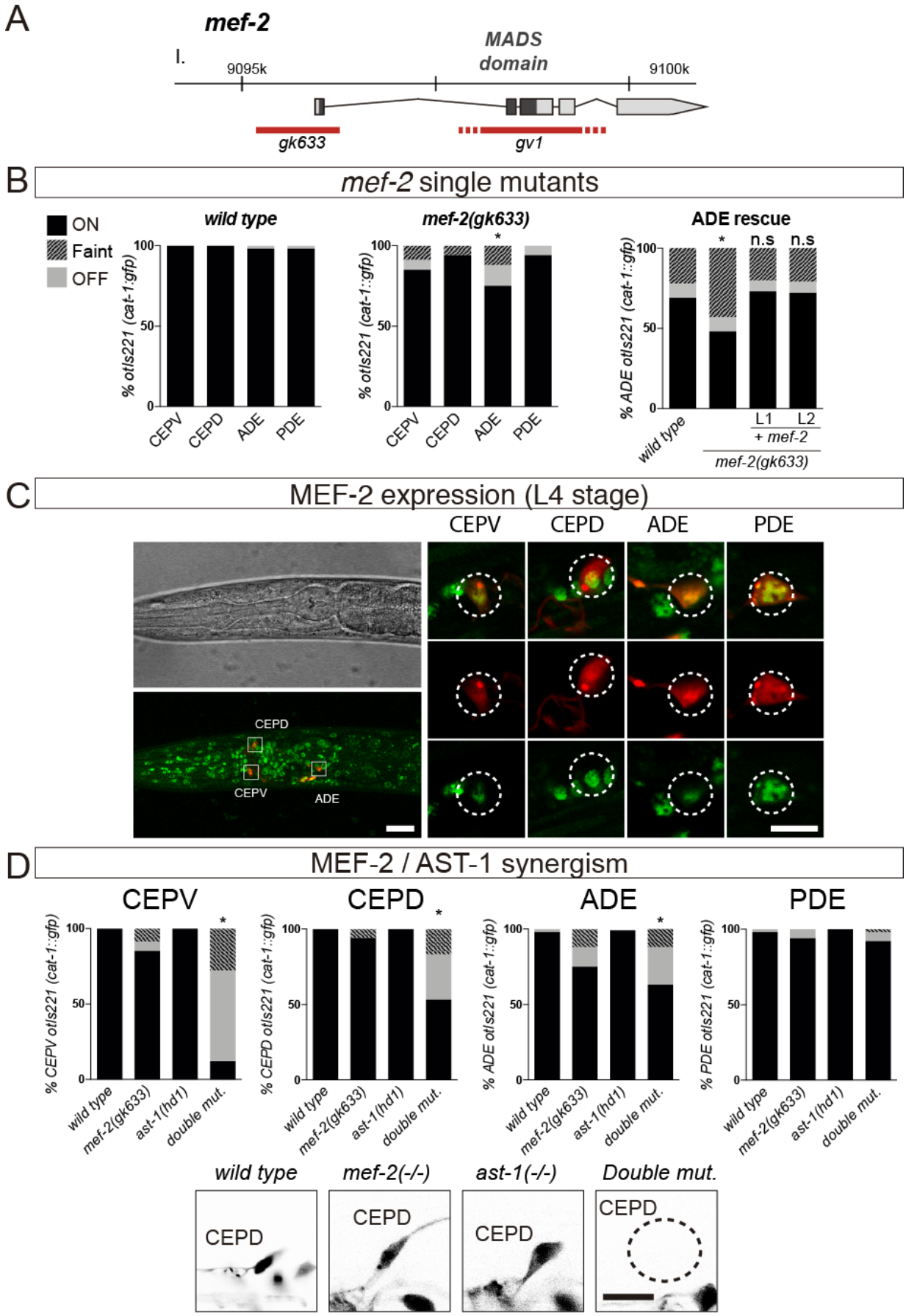
A
mef-2
I.
9095k
MADS domain
9100k
gk633
gv1
B
mef-2 single mutants
ON
Faint
OFF
wild type
mef-2(gk633)
ADE rescue
% otIs221 (cat-1::gfp)
% ADE otIs221 (cat-1::gfp)
CEPV
CEPD
ADE
PDE
n.s
L1
L2
+ mef-2
C
MEF-2 expression (L4 stage)
D
MEF-2 / AST-1 synergism
% CEPV otIs221 (cat-1::gfp)
% CEPD otIs221 (cat-1::gfp)
% ADE otIs221 (cat-1::gfp)
% PDE otIs221 (cat-1::gfp)
wild type
mef-2(gk633)
ast-1(hd1)
double mut.
mef-2(-/-)
ast-1(-/-)
Double mut.

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 (*wgIs301*). 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* (*otIs181*). 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.

388 Similar to what we found for HSN terminal selectors (Lloret-Fernández et al.
389 2018), synergistic effects of *mef-2(gk633)* and *ast-1(hd1)* are reporter-specific,
390 *mef-2(gk633)*; *ast-1(hd1)* double mutants show no defects in *bas-1::gfp*
391 expression, revealing that ADE and CEPs are not absent in these animals
392 (**Supplementary Figure 5**). These results expand the number of TFs and TF
393 families involved in DA terminal differentiation and underscore the importance
394 of genetic redundancy in neuron terminal differentiation programs.
395

396 **The dopaminergic terminal differentiation program is phylogenetically**
397 **conserved**

398 Next we performed neuron-type specific rescue experiments to test for cell
399 autonomous effects of these TFs. *unc-62* codes for eight different coding
400 isoforms, *unc-62 isoform a (unc-62a)* is expressed neuronally and it is the only
401 isoform commonly affected in *e917, mu232* and *e644* alleles, which all show
402 dopaminergic defects. We expressed *unc-62a* under the *dat-1* promoter as *dat-*
403 *1* expression is unaffected in CEPs and ADE in *unc-62(e971)* mutants (**Figure**

404 **2C**), however, we find that none of the two transgenic lines produce significant
405 rescue of the *unc-62(e971)* phenotype (**Figure 4A**). As we have determined
406 that *unc-62* has a dual role in terminal differentiation and lineage specification,
407 failure to rescue could indicate *unc-62* might be required in ADE also earlier
408 than the onset of *dat-1* expression, alternatively different isoforms or specific TF
409 levels might be needed for rescue.
410 A similar strategy was used for *vab-3(ot346)* rescue experiments. *vab-3* codes
411 for three isoforms, but only the long isoform *vab-3*a contains both the PAIRED
412 and the HD DNA binding protein domains. Expression of *vab-3a* under the *dat-1*
413 promoter is sufficient to rescue *cat-2* reporter expression defects in ADE,
414 demonstrating a cell autonomous and terminal role for *vab-3* in neuron
415 specification (**Figure 4B**). As expected, *vab-3(ot346)* CEP lineage defects were
416 not rescued, as the promoter used for *vab-3* rescue (*dat-1prom*) is only
417 expressed in terminally differentiated CEPs.
418 For *mef-2* rescue experiments we used double mutants with *ast-1(hd1),* that
419 show stronger dopaminergic expression defects. *mef-2* codes for a single
420 isoform, we find *mef-2* cDNA expression under the *bas-1* promoter rescues loss
421 of CEPV and CEPD *cat-1* reporter expression in *ast-1(hd1); mef-2(gk633)*
422 double mutants, demonstrating a cell autonomous and terminal role for *mef-2* in
423 dopaminergic neuron specification (**Figure 4C**).

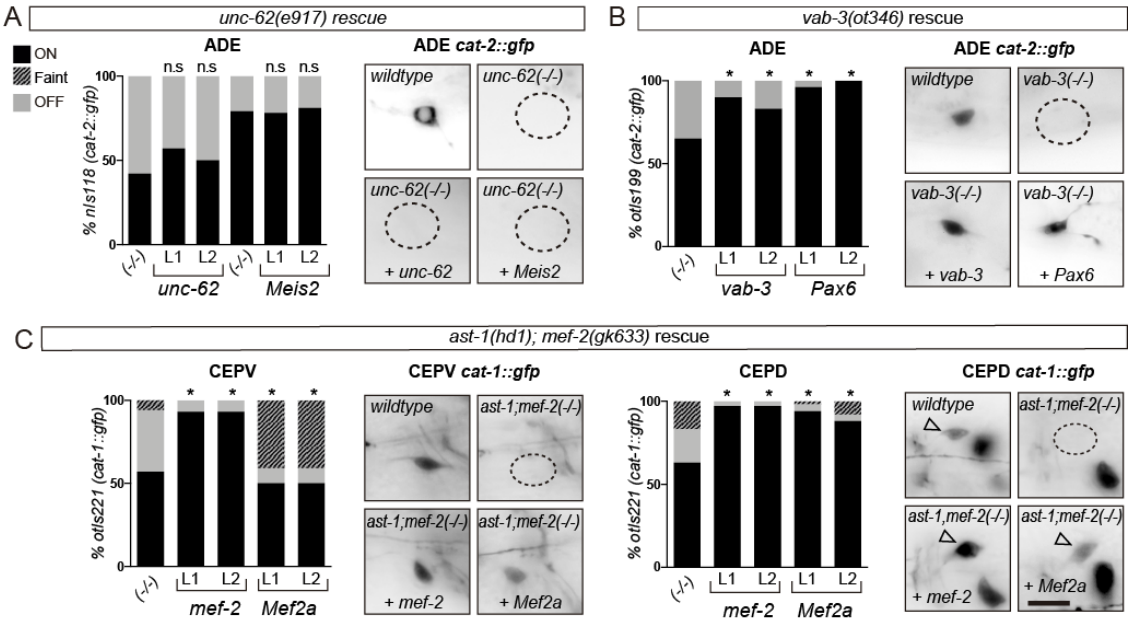

Figure 4

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 µm 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*

433 phenotypes respectively (**Figure 4B, C**). However, Meis2 did not produce
434 significant rescue of *unc-62* mutants, similar to *unc-62* cDNA failure to rescue,
435 suggesting MEIS factors might be also required at earlier time points, specific
436 concentrations or isoforms (**Figure 4A**). Our data further supports the
437 phylogenetic conservation of the gene regulatory network controlling
438 dopaminergic terminal differentiation.
439

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

443 We have previously reported functional binding sites (BS) for *ast-1, ceh-43* and
444 *ceh-20* terminal selectors in the *cis*-regulatory modules of the dopamine
445 pathway genes (Flames and Hobert 2009; Doitsidou et al. 2013). To analyze if
446 the new regulators of dopaminergic terminal fate could also directly activate
447 dopaminergic effector gene expression we focused on the *cis*-regulatory
448 analysis of *cat-2/tyrosine hydroxylase a gene* exclusively expressed in the
449 dopaminergic neurons.
450 The *cat-2* minimal *cis*-regulatory module (*cat-2p21*), in addition to the previously
451 described ETS, PBX and HD BS (Flames and Hobert 2009; Doitsidou et al.
452 2013) also contains predicted MADS, MEIS and PAIRED BS (**Figure 5B**). VAB-
453 3 protein contains two DNA binding domains, PAIRED and HD where each
454 DNA domain fulfills specific functions (Brandt et al. 2019). Interestingly, we find
455 that one of the two already identified functional HD BS matches a PAIRED-type
456 HD consensus (HTAATTR, labeled as HD* in **Figure 5**) suggesting it could be
457 recognized by VAB-3 and/or CEH-43.

458 Point mutation in PAIRED and MEIS BS show strong *gfp* expression defects
459 while MADS BS mutation had only a small effect in the PDE (**Figure 5** and
460 **Supplementary Figure 6**). These data reinforce a model in which known
461 dopaminergic terminal selectors *ast-1*/Etv1, *ceh-43*/Dlx2 and *ceh-20*/Pbx1 could
462 act together with *unc-62*/Meis2, *vab-3*/Pax6 and *mef-2*/Mef2a to regulate
463 dopaminergic terminal differentiation.
464

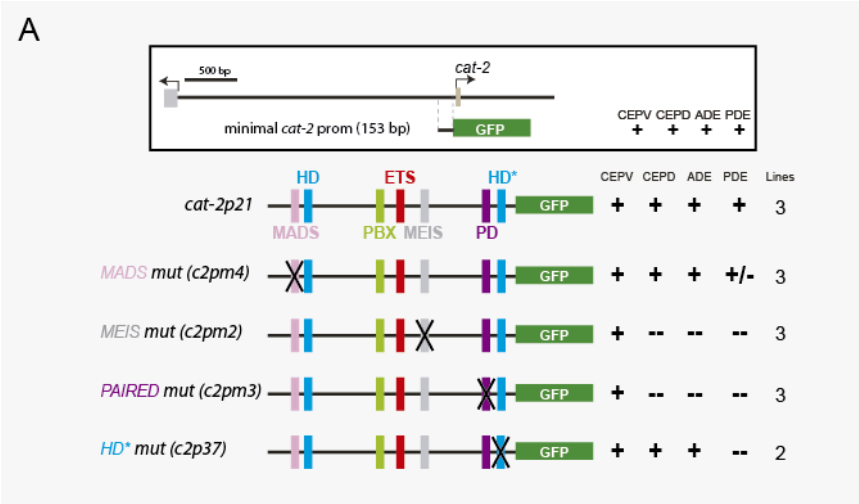

Figure 5

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.

465 **The dopaminergic regulatory signature is preferentially associated to**
466 **dopaminergic neuron effector genes**

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

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

490 sets of 86 random genes with similar upstream and intronic length distribution to
491 dopaminergic expressed genes.
492 First, we focused our analysis only on the three already published dopaminergic
493 terminal selectors (*ast-1/ETS, ceh-43/DLL HD* and *ceh-20/PBX HD*). We found
494 this regulatory signature (presence of ETS+HD+PBX binding sites clustered in
495 700 bp or smaller regions) lacks enough specificity. All genes, either
496 dopaminergic expressed genes or random sets, contain DNA windows with
497 matches for all three TFs (100% of dopaminergic expressed genes compared to
498 100% in random sets). Reducing DNA-window search length from 700 bp to
499 300 bp or 150 bp did not increase specificity. Next, we expanded our analysis to
500 search for a regulatory signature containing one or more matches for each of
501 the seven position weight matrices associated to the dopaminergic TF terminal
502 regulatory program (ETS+HD+HD*+PBX+MEIS+PAIRED+MADS binding sites
503 clustered in 700bp or smaller region) (**Figure 6A**). We found that 87% of
504 dopaminergic expressed genes contain at least one associated dopaminergic
505 regulatory signature window, a significantly higher percentage compared to the
506 random sets of genes (mean random sets 79%, $p<0.001$) (**Figure 6B**). Thus,
507 dopaminergic signature is significantly enriched in dopaminergic-expressed
508 genes. Next, we built similar sets of differentially expressed genes for five
509 randomly picked non-dopaminergic neuron categories (RIA, ASE, Touch
510 Receptor neurons, GABAergic neurons and ALN/PLN/SDQ cluster)
511 (**Supplementary Figure 7** and **Table S7**). We found that the percentage of
512 expressed genes containing the dopaminergic signature is smaller in non-
513 dopaminergic neurons compared to dopaminergic neurons (**Figure 6B**). This
514 difference is statistically significant for all neuron types except the

515 ALN/PLN/SDQ neuron cluster (**Figure 6B**). In addition, none of the non-
516 dopaminergic neuronal types show a significant enrichment of the dopaminergic
517 signature with respect to their respective background of ten thousand random
518 sets of comparable genes (**Figure 6B**, p>0.05). The reason why ALN/PLN/SDQ
519 expressed genes show a higher association to dopaminergic regulatory
520 signature compared to other non-dopaminergic neurons is uncertain. Terminal
521 selectors for ALN/PLN/SDQ neurons are yet unknown, thus a similar
522 combination of TF families controlling terminal differentiation of these neurons
523 could explain the higher presence of the dopaminergic regulatory signature.
524 Dopaminergic regulatory signature enrichment in dopaminergic-expressed
525 genes is even more pronounced when considering only the promoter sequence
526 (1.5 Kb upstream of the ATG) (**Figure 6C**). These data suggest that proximal
527 regulation has a major role in neuronal terminal differentiation in *C. elegans*.
528 Four out of five dopamine pathway genes contain at least one associated
529 dopaminergic regulatory signature window. Remarkably, previously isolated *cis*-
530 regulatory modules overlap with predicted dopaminergic regulatory signature
531 windows (**Figure 6D**) (Flames and Hobert 2009). Experimentally determined
532 enhancers for two genes, *dat-1* and *cat-2,* contain all seven types of TFBS ,
533 while *cat-1* and *bas-1* enhancers contain only 6 and 4 types of motifs
534 respectively. Thus we performed similar dopaminergic regulatory signature
535 analysis considering the presence of at least six or at least five motifs. We find
536 that the full complement of seven dopaminergic TFBS is required to provide
537 specificity to the dopaminergic regulatory signature as regulatory windows
538 containing only six or five types of dopaminergic TF motifs are not preferentially
539 associated to dopaminergic expressed genes compared to the random

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

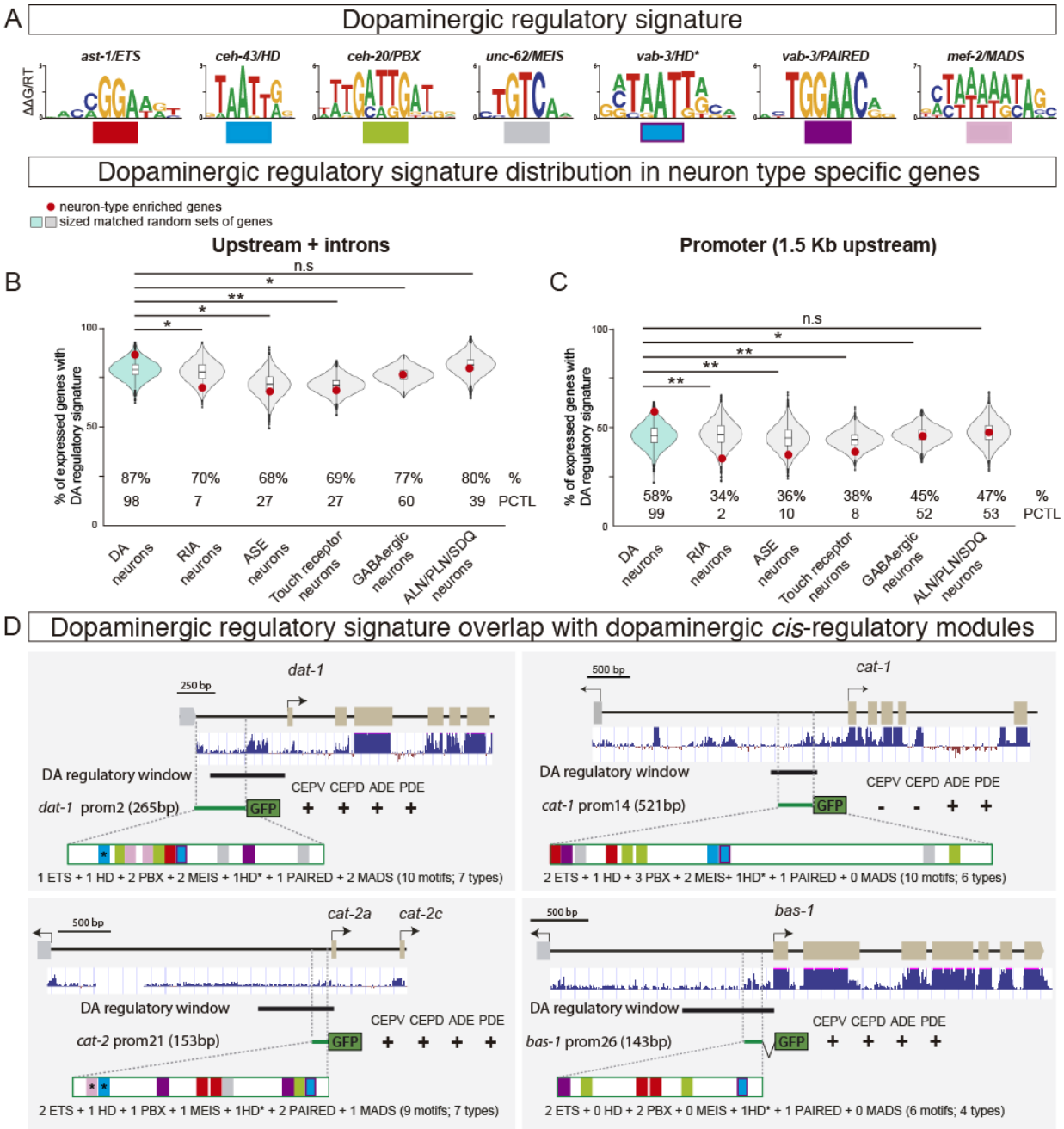

Figure 6

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 mistaches 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

550 cells. Finally, the presence of the dopaminergic signature in some genes not
551 expressed in dopaminergic neurons indicates that the signature itself is not
552 sufficient to induce dopaminergic expression. Additional TFBS, gene repression
553 mechanisms, or chromatin accessibility could further regulate dopaminergic
554 regulatory signature specificity. It is also possible that specific syntactic rules
555 (TFBS order, distance and disposition) discriminate functional from non-
556 functional dopaminergic regulatory signature windows.
557
558 **Additional parallel gene routines expressed in the dopaminergic neurons**
559 **do not show enrichment in dopaminergic regulatory signature**
560 Next, we examined the distribution of the dopaminergic regulatory signature
561 across the entire *C. elegans* genome. We found it is preferentially present in
562 putative regulatory sequences of neuronally expressed genes compared to the
563 rest of the genome (**Figure 7A**)**.** This enrichment is also present when
564 analyzing only promoter sequences (**Figure 7A**).
565 Gene ontology analysis of all genes in the *C. elegans* genome with
566 dopaminergic regulatory signature revealed enrichment of many neuronal
567 processes related to dopaminergic differentiation and function, including
568 regulation of locomotion, learning and memory, response to stimuli or dopamine
569 secretion (**Figure 7B**). Similar gene ontology categories are enriched when only
570 genes with dopaminergic regulatory signature present in their promoters are
571 considered (**Figure 7B**).
572 Thus we next aim to explore to what extend dopaminergic regulator signature is
573 present in the whole terminal transcriptiome of dopaminergic neurons. Different
574 hierarchies of gene expression co-exist in any given cell, mechanosensory

575 dopaminergic neurons co-express at least four types of genes: 1) Dopaminergic
576 effector genes such as dopaminergic pathway genes, neuropeptides,
577 neurotransmitter receptors, etc., that are preferentially (but not exclusively)
578 expressed by this neuron type; 2) Genes coding for structural components of
579 cilia which are expressed by all sixty ciliated sensory neurons in *C. elegans*; 3)
580 Panneuronal genes, such as components of the synaptic machinery or
581 cytoskeleton, expressed by all 302 neurons in *C. elegans* hermaphrodite and 4)
582 Ubiquitous genes expressed by all cells of the organism, such as ribosomal or
583 heat shock proteins (**Figure 7C**). Thus, we next studied the distribution of the
584 dopaminergic regulatory signature in these additional gene categories.
585 We find that the percentage of genes with dopaminergic signature is highest for
586 dopaminergic enriched effector genes compared to other sets of genes (**Figure**
587 **7C**). This difference is also present when only promoter sequences are
588 analyzed (**Figure 7C**). Of note, although lower than dopaminergic effector
589 gene-set, panneuronal genes also show increased presence of dopaminergic
590 regulatory signature compared to other regulatory routines of the cell (**Figure**
591 **7C**) suggesting dopaminergic TFs could also contribute to panneuronal gene
592 control, which is consistent with previous reports on panneuronal gene
593 regulation (Stefanakis et al. 2015). The frequency of dopaminergic regulatory
594 signature in panciliated and ubiquitous genes is more similar to genes not
595 expressed in neurons. Thus, our data suggest that the main function for the
596 dopaminergic TF collective is the regulation of neuron-type specific gene
597 expression and they could be involved to a less extent in panneuronal gene
598 regulation.

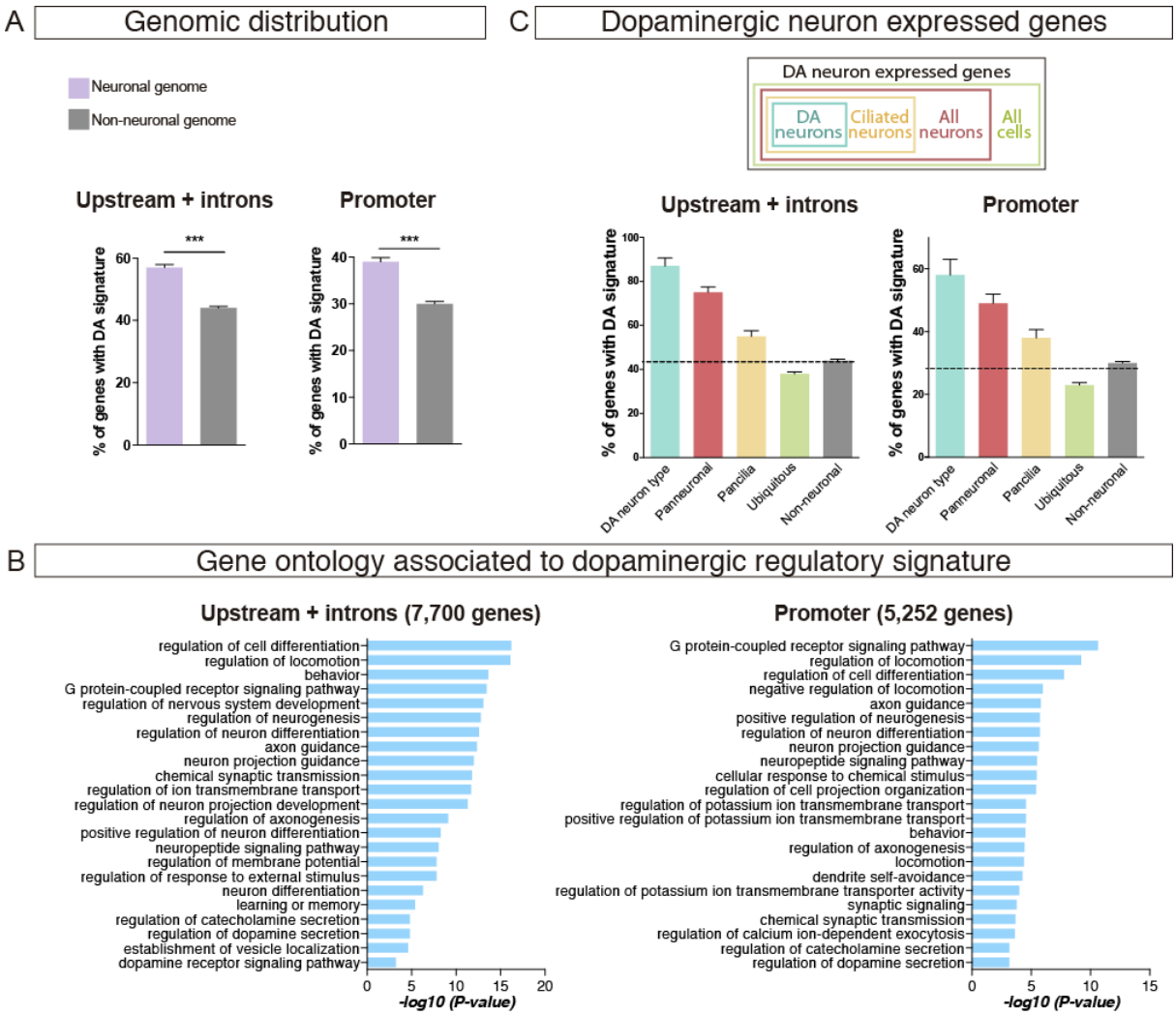

Figure 7

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).

600 **DISCUSSION**

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

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

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

625 identified in our study. The ancestral role of bHLH TFs as proneural factors has
626 been established in a wide range of metazoans including vertebrates,
627 Drosophila, *C. elegans* and the cnidaria *Nematostella vectensis* which has one
628 of the most simple nervous systems (Poole et al. 2011; Lloret-Fernández et al.
629 2018; Guillemot and Hassan 2017; Layden et al. 2012). HD TFs are involved in
630 many developmental processes and their role in neuron terminal differentiation
631 is also conserved in different animal groups including cnidaria (Hobert 2016;
632 Thor et al. 1999; Tournière et al. 2020; Babonis and Martindale 2017; Briscoe et
633 al. 2000; Shirasaki and Pfaff 2002). Interestingly, our results suggest HD
634 functions are not limited to neuron terminal differentiation but do also play
635 important roles as lineage determinants. We find vab-3/PAIRED HD and unc-
636 62/MEIS HD mutants show both dopaminergic lineage and terminal
637 differentiation defects. Similar dual roles have been also reported for ceh-43/HD
638 and ceh-20/PBX HD (Doitsidou et al. 2013).
639 Several NHRs, bZIP and ZF TFs are known to regulate neuron-type
640 specification in mammals, such as Couptf1, Nurr1, Tlx and Nr2e3 NHR TFs
641 (Bovetti et al. 2013; Haider et al. 2000; Zetterström et al. 1997; Roy et al. 2004),
642 Nrl, cMaf and Mafb bZIP TFs (Wende et al. 2012; Blanchi et al. 2003; Mears et
643 al. 2001) or Myt1l, Gli1, Sp8, Ctip2, Fezf1/2 ZF TFs (Mall et al. 2017; Hynes et
644 al. 1997; Waclaw et al. 2006; Arlotta et al. 2005; Shimizu et al. 2010). However,
645 in contrast to bHLH and HD, a general role in neuron specification for these
646 families has been poorly studied in any model organism. Functional
647 characterization of the NHR, bZIP and ZF TF candidates retrieved from our
648 RNAi screen will help better understand the role of these TF families in neuron
649 specification.

650

651 **Complexity of terminal differentiation programs provide genetic**
652 **robustness and enhancer selectivity**

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

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

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

675 putative regulatory regions of dopaminergic-expressed genes compared to
676 other genes. Although HD TFs represent a large number of the dopaminergic
677 terminal selector collective, additional TF families (such as ETS and MEF) are
678 also required. This might be particularly relevant during terminal differentiation
679 to increase diversity of TFBS that help distinguish dopaminergic specific
680 enhancers. Thus TF complexity of neuronal terminal regulatory programs might
681 not only provide genetic robustness but also facilitate enhancer selection.
682
683 **Parallel gene regulatory routines co-exist in the cell**
684 The active transcriptome of a mature neuron can be divided in different gene
685 categories depending on specificity of expression, ranging from cell-type
686 specific genes to ubiquitous gene expression. While it is well established that
687 terminal selectors have a direct role in neuron-type-specific gene expression,
688 their role in other gene categories is less clear (Hobert 2016). In our study we
689 find the dopaminergic regulatory signature is predominantly enriched in
690 dopamine effector genes, supporting a restricted role for terminal selectors in
691 the regulation of neuron-type effector genes. Dopaminergic regulatory signature
692 is also associated, to a less extent, to panneuronal genes, which is consistent
693 with the known minor contributions for some terminal selectors in panneuronal
694 gene expression (Stefanakis et al. 2015). In contrast, genes coding for cilia
695 components (panncilia genes) and ubiquitous genes are mostly devoid of
696 dopaminergic regulatory signature, suggesting that other TFs activate these
697 gene categories. Transcriptional regulation of ubiquitous gene has been poorly
698 studied. Broadly expressed TFs, such as Sp1, GABP/ETS and YY1 seem to
699 play a role in ubiquitous gene expression in mammals (Bellora et al. 2007;

700 Farré et al. 2007). Cilia structural genes are directly regulated by *daf-19*/RFX TF
701 (Swoboda et al. 2000), and this role is conserved in vertebrates (Choksi et al.
702 2014). However, *daf-19* is expressed in all neurons, thus additional unknown
703 TFs must control cilia gene expression. In summary, and consistent with
704 previous reports based on the analysis of a limited number of genes (Stefanakis
705 et al. 2015), terminal selector collectives seem mostly devoted to neuron-type
706 specific gene expression and to a less extend to panneuronal genes.
707
708 **Neuron type evolution and deep homology**
709 Mouse TF orthologs of the *C. elegans* dopaminergic terminal regulatory
710 program are also required for olfactory bulb dopaminergic terminal specification,
711 which is the most ancestral dopaminergic population of the mammalian brain
712 (Agoston et al. 2014; Brill et al. 2008; Flames and Hobert 2009; Bovetti et al.
713 2013; Remesal et al. 2020; Qiu et al. 1995). Here we find mouse TFs can, in
714 most cases, functionally substitute their worm orthologs suggesting both
715 dopaminergic neuronal populations share deep homology, which refers to the
716 relationship between cells in distant species that share their genetic regulatory
717 programs (Shubin et al. 2009). Similar conservation of terminal regulatory
718 programs is found for other neurons in *C. elegans* (Lloret-Fernández et al.
719 2018; Kratsios et al. 2012).
720 Identification of deep homology is an emerging strategy in evolutionary biology
721 used to assign homologous neuronal types in distant species (Arendt 2008;
722 Arendt et al. 2016, 2019). The study of neuronal regulatory programs in
723 different animal groups and the description of homologous relationships will
724 help us better understand the origin of neurons and the nervous system

725 evolution.

726 **METHODS**

727 ***C. elegans* Strains and Genetics**

728 *C. elegans* culture and genetics were performed as described (Brenner 1974).
729 Strains used in this study are listed in **Table S9**.

730

731 **Generation of the Transcription factor RNAi library**

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

741

742 **RNAi feeding experiments**

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

751 Approximately 7 days later young adult F1 generation was scored. Lethal RNAi
752 clones, that precluded F1 analysis, were scored at P0 as young adults (**Table**
753 **S2**). In each replicate, a minimum of 30 worms per clone, coming from three
754 distinct plates were scored. All experiments were performed at 20°C. Each
755 clone was scored in two independent replicates, a third replicate was performed
756 when results from the first and second replicates did not coincide. In all scorings
757 we include L4440 empty clone as negative control and *gfp* RNAi clone as
758 positive control.
759
760 **Mutant strains and genotyping**
761 Strains used in this study are listed in **Table S9**. Deletion alleles were
762 genotyped by PCR. Point mutations were genotyped by sequencing (**Table**
763 **S10**). Alleles *vab-3(ot346)*, *unc-62(e644) and lag-1(q385)* were determined by
764 visual mutant phenotype. The mutation *unc-62(mu232)* was followed through its
765 link with the fluorescent reporter *muIs35.*
766
767 **Generation of *C. elegans* transgenic lines**
768 Gene constructs for *cis*-regulatory analysis were generated by cloning into the
769 pPD95.75 vector. For identification of the putative binding sites the following
770 consensus sequences were used: GACA for UNC-62/MEIS HD (Campbell and
771 Walthall 2016); GRAGBA for VAB-3/PAIRED HD (Holst et al. 1997; Kim et al.
772 2008) and CTAWWWWTAG for MEF-2/MADS (Boyle et al. 2014). Directed
773 mutagenesis was performed with Quickchange II XL site-directed mutagenesis
774 kit (Stratagene). For microinjections the plasmid of interest (50 ng/µl) was
775 injected together with *rol-6* co-marker pRF4 [rol-6(su1006), 100 ng/µl]. *vab-*

776 *3::gfp* fosmid injection DNA mix contained fosmid (40 ng/µl) and two
777 comarkers, pRF4 (50 ng/µl) and pNF101 (*ttx-3prom::mcherry*) (50 ng/µl). For
778 rescue experiments, commercially available cDNA of the TF candidates was
779 obtained from Dharmacon Inc and Invitrogen. cDNAs corresponding to the
780 entire coding sequence of *unc-62, vab-3, mef-2, Meis2, Pax6, and Mef2a* were
781 amplified by PCR and cloned into *dat-1* and *bas-1* promoter reporter plasmids
782 replacing *gfp* cDNA (See primers in **Table S10**). cDNA plasmid (25 ng/µl) was
783 injected directly into the corresponding mutant background as complex arrays
784 together with digested *E. coli* genomic DNA (50 ng/µl) and pNF417 (*unc-*
785 *122::rfp,* 25 ng/µl) fluorescent co-marker.
786
787 **Scoring and statistics**
788 Scoring and micrographs were performed using 40X objective in a Zeiss
789 Axioplan 2 microscope. For RNAi screening experiments and *cis*-analysis, 30
790 young adult animals per line or per RNAi clone and replicate were analyzed;
791 For mutant analysis 50 individuals were scored. RNAi experiments were
792 performed at 20ºC while *cis*-regulatory and mutant analyses were performed at
793 25ºC. Fisher exact test, two tailed was used for statistical analysis.
794 For mutant analysis lack of GFP signalling was considered OFF, if GFP
795 expression was substantially weaker than WT, 'FAINT' category was included.
796 For *cis*-regulatory analysis two or three independent lines for each transgenic
797 construct were analyzed. Mean expression value of three *wild type* lines is
798 considered the reference value and mutated constructs are considered to show
799 normal expression when GFP expression was 100-70% of average expression
800 of the corresponding *wild type* construct ("+" phenotype). 70-30% reduction

801 compared to *wild type* reference value was considered a partial phenotype "+/-".
802 GFP expression below 30% was considered as great loss of expression ("-"
803 phenotype) and if no GFP was detected (less than 5% of the neurons) was
804 assigned as total loss of GFP ("--" phenotype).
805 For MEF-2 expression analysis *wgIs301* mef-2 fosmid reporter was crossed into
806 *otIs181* (*dat-1::mcherry, ttx-3::mcherry*) .Images were taken with confocal TCS-
807 SP8 Leica microscope and processed with ImageJ 1.50i software.
808 For cross regulation experiments, the expression of *unc-62*, *vab-3* and *mef-2*
809 reporters were scored in 10 L1 larvae, in *wild type* and *ast-1(hd92)* background.
810 Z stack pictures (Zeiss microscope) of all the length of the animal were taken
811 keeping the same distance between z-planes. Images were analyzed with
812 ImageJ software to score cells in the head of the animals co-expressing *ift-20*
813 reporter with each TF.
814
815 **Bioinformatic Analysis**
816 For the differential expression heatmap, single-cell RNA-seq data from L2 larval
817 stage were used (Cao et al., 2017). Cells were classified into four different
818 categories: 1) serotonergic neurons if they expressed *tph-1* or *mod-5* genes, 2)
819 dopaminergic (DA) neurons if they expressed *dat-1* or *cat-2* genes, 3) RIC
820 neurons if they expressed *tbh-1* and *tdc-1* genes, and 4) RIM neurons if they
821 expressed *tdc-1* but no *tbh-1* genes. Cells that fell into the serotonergic
822 category were further subdivided into ADF neurons or NSM neurons by
823 refinement of t-SNE coordinates. A total of 140 cells were used for the heatmap
824 (15 ADF, 74 DA, 16 NSM, 10 RIC, 25 RIM). Genes with zero count across all
825 the neuronal categories were excluded and z-scores were calculated to

826 normalize data. The heatmap and clustering was performed using the package
827 heatmaply (Galili et al. 2018).
828 TFs associated to neuronal phenotypes were obtained retrieving WormBase
829 annotated alleles associated to phenotypes matching the terms "neuron",
830 "nervous system" or "axon". For expression analysis, embryonic, L2 and L4 sc-
831 RNA-seq data (Cao et al. 2017; Packer et al. 2019; Taylor et al. 2020) and
832 fluorescence reporter lineage data (https://epic.gs.washington.edu) was used to
833 determine if TFs are expressed in each neuron type or its corresponding
834 lineage.
835 Unless otherwise indicated, all regulatory signature analyses were performed
836 using the software R (Zhang et al. 2014b) and packages from Bioconductor
837 (Huber et al. 2015). sc-RNA-seq data from (Cao et al. 2017) was downloaded
838 from the author's website (http://atlas.gs.washington.edu/worm-rna/). These
839 data correspond to nearly 50,000 cells coming from *Caenorhabditis elegans* at
840 the L2 larval stage. Cells with failed QC, doublets and unclassified cells were
841 filtered and excluded from subsequent analysis. Differential expression analysis
842 between dopaminergic neurons cluster and clusters for the other ciliated
843 sensory neurons was performed using Monocle (Qiu et al. 2017b, 2017a;
844 Trapnell et al. 2014) and results were filtered by q-value (≤ 0.05) in order to get
845 a list of differentially expressed genes in dopaminergic neurons. For gene
846 enrichment of neuronal clusters corresponding to non-dopaminergic neurons
847 (RIA, ASE, Touch receptor neurons, GABAergic neurons and ALN/PLN/SDQ
848 neurons) we followed a similar strategy and performed differential expression
849 tests between each neuronal cluster and all cells from the dataset annotated as
850 neurons (**Table S7**). Specificity of these six gene sets was checked by the

851 enrichment of its anatomical association in *C. elegans* using the web tool
852 WormEnrichr (https://amp.pharm.mssm.edu/WormEnrichr/) (Chen et al. 2013).
853 Gene lists for neuronal, non-neuronal, ubiquitous, panneuronal and panciliated
854 categories were inferred from a more comprehensive sc-RNA-seq dataset
855 (Packer et al. 2019). We retrieved gene expression data (log2 transcripts per
856 million) from all the genes that were expressed in at least one annotated
857 terminal cell bin, getting a final matrix of 15,813 genes x 409 terminal cell bins.
858 Following authors' original approach, genes were ordered by hierarchical
859 clustering and cell bins were ordered by tissues; resulting in differential gene
860 clusters which marked sites of predominant expression [Figure S32 from
861 (Packer et al. 2019)]. Using these data, we manually curated five gene lists:
862 non-neuronal tissue-specific, ubiquitous, panneuronal, panciliated, and neuron-
863 type specific expression (**Table S7**). For genes located in operons, only the
864 gene located at the 5' end of the cluster, and thus subdued to *cis*-regulation,
865 were considered for dopaminergic regulatory signature analysis. For hybrid
866 operons, additional promoters were also included (Blumenthal, Davis & Garrido-
867 Lecca, 2015). For genomic search of dopaminergic regulatory signature similar
868 exclusion of downstream genes from operon was performed, removing a total of
869 2,083 genes from the analysis. The final curated gene lists used in this study
870 are listed in **Table S7**.
871 For *C. elegans* regulatory signature analysis, we downloaded PWMs from Cis-
872 BP version 1.02 (Weirauch et al. 2014) corresponding to the TF binding sites of
873 the six transcription factors that compose the dopaminergic regulatory signature
874 in *C. elegans*. If the exact match for *C. elegans* was not available, we selected
875 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

901 genes and other neuronal groups were assessed by Brunner-Munzel test
902 performed with R package brunnermunzel (Toshiaki 2019), *: p-value < 0.05, **:
903 p-value < 0.01, ***: p-value < 0.001).
904 Gene ontology analysis of the genes assigned dopaminergic regulatory
905 signature were carried out using the web tool GOrilla (http://cbl-
906 gorilla.cs.technion.ac.il/) (Eden et al. 2009), using *C. elegans* coding genome as
907 control list.
908
909 **Data access**
910 Data used for the analysis and raw scorings are accessible in supplementary
911 materials.
912
913 **Competing interest statement**
914 Authors declare no competing interests
915

916 **Acknowledgments**

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

and Dr Oliver Hobert for comments on the manuscript.

Author Contributions

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

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, Saghatelyan 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

1031 regulations of proneural proteins. *Curr Opin Neurobiol*.
1032 Haider NB, Jacobson SG, Cideciyan A V., Swiderski R, Streb LM, Searby C,
1033 Beck G, Hockey R, Hanna DB, Gorman S, et al. 2000. Mutation of a
1034 nuclear receptor gene, NR2E3, causes enhanced S cone syndrome, a
1035 disorder of retinal cell fate. *Nat Genet*.
1036 Hobert O. 2016. A map of terminal regulators of neuronal identity in
1037 Caenorhabditis elegans. *Wiley Interdiscip Rev Dev Biol*.
1038 Hobert O. 2008. Regulatory logic of neuronal diversity: Terminal selector genes
1039 and selector motifs. *Proc Natl Acad Sci U S A*.
1040 Holst BD, Wang Y, Jones FS, Edelman GM. 1997. A binding site for Pax
1041 proteins regulates expression of the gene for the neural cell adhesion
1042 molecule in the embryonic spinal cord. *Proc Natl Acad Sci U S A*.
1043 Huang X, Tian E, Xu Y, Zhang H. 2009. The C. elegans engrailed homolog ceh-
1044 16 regulates the self-renewal expansion division of stem cell-like seam
1045 cells. *Dev Biol*.
1046 Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo
1047 HC, Davis S, Gatto L, Girke T, et al. 2015. Orchestrating high-throughput
1048 genomic analysis with Bioconductor. *Nat Methods*.
1049 Hynes M, Stone DM, Dowd M, Pitts-Meek S, Goddard A, Gurney A, Rosenthal
1050 A. 1997. Control of cell pattern in the neural tube by the zinc finger
1051 transcription factor and oncogene Gli-1. *Neuron* **19**: 15–26.
1052 Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R, Gotta M, Kanapin A, Le
1053 Bot N, Moreno S, Sohrmann M, et al. 2003. Systematic functional analysis
1054 of the Caenorhabditis elegans genome using RNAi. *Nature*.
1055 Kamath RS, Martinez-Campos M, Zipperlen P, Fraser AG, Ahringer J. 2001.
1056 Effectiveness of specific RNA-mediated interference through ingested
1057 double-stranded RNA in Caenorhabditis elegans. *Genome Biol*.
1058 Kheradpour P, Ernst J, Melnikov A, Rogov P, Wang L, Zhang X, Alston J,
1059 Mikkelsen TS, Kellis M. 2013. Systematic dissection of regulatory motifs in
1060 2000 predicted human enhancers using a massively parallel reporter
1061 assay. *Genome Res* **23**: 800–811.
1062 Kim DS, Matsuda T, Cepko CL. 2008. A core paired-type and POU
1063 homeodomain-containing transcription factor program drives retinal bipolar
1064 cell gene expression. *J Neurosci*.
1065 Kratsios P, Stolfi A, Levine M, Hobert O. 2012. Coordinated regulation of
1066 cholinergic motor neuron traits through a conserved terminal selector gene.
1067 *Nat Neurosci*.
1068 Kuntz SG, Williams BA, Sternberg PW, Wold BJ. 2012. Transcription factor
1069 redundancy and tissue-specific regulation: Evidence from functional and
1070 physical network connectivity. *Genome Res* **22**: 1907–1919.
1071 Layden MJ, Boekhout M, Martindale MQ. 2012. Nematostella vectensis
1072 achaete-scute homolog NvashA regulates embryonic ectodermal
1073 neurogenesis and represents an ancient component of the metazoan
1074 neural specification pathway. *Development* **139**: 1013–1022.
1075 Liu A, Niswander LA. 2005. Bone morphogenetic protein signalling and
1076 vertebrate nervous system development. *Nat Rev Neurosci*.
1077 Liu H, Strauss TJ, Potts MB, Cameron S. 2006. Direct regulation of egl-1 of
1078 programmed cell death by the Hox protein MAB-5 and CEH-20, a C.
1079 elegans homolog of Pbx1. *Development*.
1080 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.

Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods*.

Reece-Hoyes JS, Deplancke B, Shingles J, Grove CA, Hope IA, Walhout AJM. 2005. A compendium of Caenorhabditis elegans regulatory transcription factors: A resource for mapping transcription regulatory networks. *Genome Biol* **6**.

Reilly MB, Cros C, Varol E, Yemini E, Hobert O. 2020. Unique homeobox codes delineate all C . elegans neuron classes. *Press* 1–28. http://dx.doi.org/10.1038/s41586-020-2618-9.

Remesal L, Roger-Banyat I, Chirivella L, Maicas M, Brocal-Ruiz R, Pérez-Villalva A, Cucarella C, Casado M, Flames N. PBX1 is a terminal selector of olfactory bulb dopaminergic neurons. *Dev*.

Remesal L, Roger-Baynat I, Chirivella L, Maicas M, Brocal-Ruiz R, Pérez-Villalva A, Cucarella C, Casado M, Flames N. 2020. PBX1 acts as terminal selector for olfactory bulb dopaminergic neurons. *Development* dev.186841.

Rentzsch F, Layden M, Manuel M. 2017. The cellular and molecular basis of cnidarian neurogenesis. *Wiley Interdiscip Rev Dev Biol* **6**: 1–19.

Roy K, Kuznicki K, Wu Q, Sun Z, Bock D, Schutz G, Vranich N, Monaghan AP. 2004. The tlx gene regulates the timing of neurogenesis in the cortex. *J Neurosci*.

Rual JF, Ceron J, Koreth J, Hao T, Nicot AS, Hirozane-Kishikawa T, Vandenhaute J, Orkin SH, Hill DE, van den Heuvel S, et al. 2004. Toward improving Caenorhabditis elegans phenome mapping with an ORFeome-based RNAi library. *Genome Res*.

Serrano-Saiz E, Poole RJ, Felton T, Zhang F, De La Cruz ED, Hobert O. 2013. XModular control of glutamatergic neuronal identity in C. elegans by distinct homeodomain proteins. *Cell*.

Shimizu T, Nakazawa M, Kani S, Bae YK, Shimizu T, Kageyama R, Hibi M. 2010. Zinc finger genes Fezf1 and Fezf2 control neuronal differentiation by repressing Hes5 expression in the forebrain. *Development* **137**: 1875–1885.

Shirasaki R, Pfaff SL. 2002. Transcriptional Codes and the Control of Neuronal Identity. *Annu Rev Neurosci*.

Shubin N, Tabin C, Carroll S. 2009. Deep homology and the origins of evolutionary novelty. *Nature*.

Simmer F, Moorman C, Van Der Linden AM, Kuijk E, Van Den Berghe PVE, Kamath RS, Fraser AG, Ahringer J, Plasterk RHA. 2003. Genome-wide RNAi of C. elegans using the hypersensitive rrf-3 strain reveals novel gene functions. *PLoS Biol*.

Singhvi A, Frank CA, Garriga G. 2008. The T-box gene tbx-2, the homeobox gene egl-5 and the asymmetric cell division gene ham-1 specify neural fate in the HSN/PHB lineage. *Genetics* **179**: 887–898.

Stefanakis N, Carrera I, Hobert O. 2015. Regulatory Logic of Pan-Neuronal Gene Expression in C. elegans. *Neuron*.

Stegmaier P, Kel AE, Wingender E. 2004. Systematic DNA-binding domain classification of transcription factors. *Genome Inform*.

Swoboda P, Adler HT, Thomas JH. 2000. The RFX-type transcription factor DAF-19 regulates sensory neuron cilium formation in C. Elegans. *Mol Cell*.

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//CRANR-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 homethorax/Meis/Prep homolog UNC-62 and the Exd/Pbx homologs CEH-20 and CEH-40 in C. elegans embryogenesis. *Development*.

Waclaw 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*.

Zhang Z, Ersoz E, Lai C-Q, Todhunter RJ, Tiwari HK, Gore M a, Bradbury PJ, Yu J, Arnett DK, Ordovas JM, et al. 2014b. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/. *Nat Genet*.

Zhao C, Emmons SW. 1995. A transcription factor controlling development of peripheral sense organs in C. elegans. *Nature* **373**: 74–78.

Zheng C, Karimzadegan S, Chiang V, Chalfie M. 2013. Histone Methylation Restrains the Expression of Subtype-Specific Genes during Terminal Neuronal Differentiation inCaenorhabditis elegans. *PLoS Genet*.

Zheng X, Chung S, Tanabe T, Sze JY. 2005. Cell-type specific regulation of serotonergic identity by the C. elegans LIM-homeodomain factor LIM-4. *Dev Biol*.