Transcription factors from multiple families ensure enhancer selectivity and robust neuron terminal differentiation

Angela Jimeno-Martín, Erick Sousa, Noemi Daroqui, 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

SUMMARY

To search for general principles underlying neuronal regulatory programs we built an RNA interference library against all transcription factors (TFs) encoded in *C. elegans* genome and systematically screened for specification defects in ten different neuron types of the monoaminergic (MA) superclass.

We identified over 90 TFs involved in MA specification, with at least ten different TFs controlling differentiation of each individual neuron type. These TFs belong predominantly to five TF families (HD, bHLH, ZF, bZIP and NHR). Next, focusing on the complexity of terminal differentiation, we identified and functionally characterized the dopaminergic terminal regulatory program. We found that seven TFs from four different families act in a TF collective to provide genetic robustness and to impose a specific gene regulatory signature enriched in the regulatory regions of dopamine effector genes. Our results provide new insights on neuron-type regulatory programs that could help better understand specification and evolution of neuron types.

Keywords

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

The display of a wide range of complex biological functions requires cellular division of labor that results in multicellular organisms. 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 in development or by induced 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 have identified conserved features of neuron specification programs, such as the importance of morphogens and intercellular signaling for lineage commitment and neuronal progenitor patterning (Andrews et al., 2019; Angerer et al., 2011; Borello and Pierani, 2010; Liu and Niswander, 2005; Nuez and Félix, 2012; Rentzsch et al., 2017), the central role of basic Helix Loop Helix (bHLH) TFs as proneural factors (Bertrand et al., 2002; Guillemot and Hassan, 2017), the prevalent role of homeodomain (HD) TFs in neuron subtype specification (Briscoe et al., 2000; Shirasaki and Pfaff, 2002; Thor et al., 1999) and the terminal selector model for terminal differentiation, in which TFs directly regulate expression of most neuron-type specific effector genes (Hobert, 2008).

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

Our results unveiled four main conclusions. First, specification of each type of MA neuron is regulated by a complex combination of at least ten different TFs. These TFs could be acting at any developmental step and are specific for each neuronal type, despite the shared expression of genes coding for enzymes and transporters required in monoaminergic metabolism. Second, in spite of this TF diversity, the TFs involved in the specification of all MA neurons consistently belong to only five out of the more than fifty *C. elegans* TF families: HD, bHLH,

Zinc Finger (ZF), basic Leucine Zipper Domain (bZIP) and non-nematode-specific members of the Nuclear Hormone Receptors (NHR). Importantly, analysis of all previously published TF mutant alleles that produce any neuronal phenotype in *C. elegans* reveals the same TF family distribution, which is consistent with a general involvement of these TF families in neuron specification. Third, we specifically focused on neuronal terminal differentiation programs and found that a complex combination of seven TFs from different families mediates the establishment of dopaminergic terminal fate. Functionally, this TF complexity provides robustness to gene expression and enhancer selectivity. And fourth, the dopaminergic terminal regulatory program plays a major role in neuron-type specific gene expression and also contributes to a less extent to panneuronal gene expression in the dopaminergic neurons, but it is not involved in the regulation of more general routines in the cell or ubiquitous gene expression. In summary, our results provide new insights into neuronal gene regulatory networks involved in the generation and evolution of neuron diversity.

RESULTS

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

To increase our understanding of the gene regulatory networks controlling neuron-type specification we built a complete and fully verified RNAi library against all the putative TFs encoded in the *C. elegans* genome. This list of RNAi clones include 763 high confidence TFs plus 112 additional RNAi clones against medium confidence TFs (Narasimhan et al., 2015; Reece-Hoyes et al., 2005) (**Data source 1**). 702 clones were extracted from published genome libraries (Kamath et al., 2003; Rual et al., 2004) and 173 clones were newly generated. All clones were verified by Sanger sequencing (**Data source 1 and Methods**). This is a new resource for the *C. elegans* community available upon request.

We used the *rrf-3(pk1426)* mutant background to sensitize neurons for RNAi effects (Simmer et al., 2003) and combined this mutation with three different fluorescent reporters that label the MA system in the worm: the vesicular monoamine transporter, *otIs224(cat-1::gfp)* expressed in all MA neurons; the dopamine transporter, *otIs181(dat-1::mcherry),* expressed in dopaminergic neurons only; and the tryptophan hydroxylase enzyme *vsIs97(tph-1::dsred),* expressed in serotonergic neurons only. Altogether our strategy labels nine different MA neuronal classes (ADE, CEPV, CEPD, NSM, ADF, RIC, RIM, HSN, PDE) and the cholinergic VC4 and VC5 motoneurons, which are not MA but express *cat-1* for unknown reasons (**Figure 1A**). The MA system in *C. elegans* is developmentally, molecularly and functionally very diverse: (1) in embryonic development, it arises from very different branches of the AB lineage (**Figure**

1B), (2) it is composed by motoneurons, sensory neurons, and interneurons, that altogether use five different neurotransmitters (**Figure 1A**) and (3) each neuronal type expresses very different transcriptomes, having in common only the minimal fraction of genes related to MA metabolism (**Figure 1C**). Of note, the dopaminergic system, composed by four anatomically defined neuronal types (CEPV, CEPD, ADE and PDE), constitutes an exception to the MA-system diversity. Although dopaminergic neurons are developmentally diverse, they are functionally and molecularly homogeneous and share their terminal differentiation program (Doitsidou et al., 2013; Flames and Hobert, 2009) (**Figure 1C**). In summary, considering the neuronal diversity of the MA system, we reasoned their global study would unravel shared principles of *C. elegans* neuronal specification.

Based on reporter expression in worms fed with negative control RNAi we set a threshold of 10% penetrance for positive hits. Under these conditions, 91 of the 875 clones produced defects in the specification of the MA system. These 91 TF candidates could be acting at any time along development. Missing or reduced reporter expression was the most frequent phenotype (78 RNAi clones, **Table 1**) although we also observed ectopic fluorescent cells and migration or morphology defects (**Figure 1D and Data source 1**). Some clones produce more than one type of phenotypes depending on the neuron type.

From our list of 91 factors we retrieved phenotypes for 22 out of 30 known regulators of MA neuron fate, including TFs involved in early lineage specification, neuron migration and known MA terminal selectors (**Data source 1**). Thus we estimated a false negative rate of around 27% which is similar to previous RNAi screens (Simmer et al., 2003). We analyzed randomly selected

mutant alleles for 12 candidates determined by RNAi that display various degrees of penetrance (from more than 80% of animals with defects to less than 15%, **Table 1**). Eight of the 12 alleles reproduce reporter defects induced by RNAi (**Table 1**), which revealed a false discovery rate (FDR) of approximately 33%, similar to published RNAi screens on neuronal functions (Liachko et al., 2019; Sieburth et al., 2005). Most known regulators of MA fate showed RNAi penetrances of at least 20% (**Table 1**). Conversely, most false positives displayed low penetrance RNAi phenotypes (<20%). In total we found 40 "high confidence" TFs displaying at least 20% penetrance by RNAi. 18 out of the 40 "high confidence" TFs have reported roles in MA specification described by mutant analysis (**Table 1**). Our mutant allele analysis identified new factors involved in MA specification, for example, a role for *zip-5*/bZIP TF in RIC and RIM specification (**Figure 1E**), which is the first reported TF involved in *C. elegans* tyraminergic and octopaminergic differentiation. We also unravel a role for *hlh-14/*bHLH TF and *lag-1*/CSL in ADF specification (**Figure 1E**) [also (Poole et al., 2011) and M.M., A.J., N.F. manuscript in preparation] or *unc-62*/MEIS HD in dopaminergic specification (**Figure 1E**).

We observed that each neuron type is affected by 10 to 15 different TF RNAi clones (**Table 1** and **Figure 1D**) with the exception of the NSM neurons, for which only two clones produced missing fluorescence phenotypes. We noticed that NSM neurons showed weak phenotypes upon *gfp* RNAi treatment (**Figure 1- Figure supplement 1**) suggesting that, even in the *rrf-3(pk1426)* sensitized background NSM could be particularly refractory to RNAi.

None of the TF RNAi clones affect all MA neurons, suggesting the absence of global regulators of MA fate, despite the fact that all MA neurons share *cat-*

164 *1*/VMAT expression. Nevertheless, a few RNAi clones show phenotypes in
165 more than one neuron type, this is most prominent in dopaminergic neurons
166 where 15 TFs RNAi clones affect more than one dopaminergic neuron type
167 (**Table1** and **Data Source 1**).
168 Altogether, our RNAi screen provided a list of 91 TFs that could play a role in
169 the specification of ten different *C. elegans* neuron-types at any developmental
170 stage. Most neuron types require at least 10 different TFs for their correct
171 specification. TFs often affect a unique neuron type with the exception of
172 dopaminergic neurons, that share some, but not all, fate regulators.

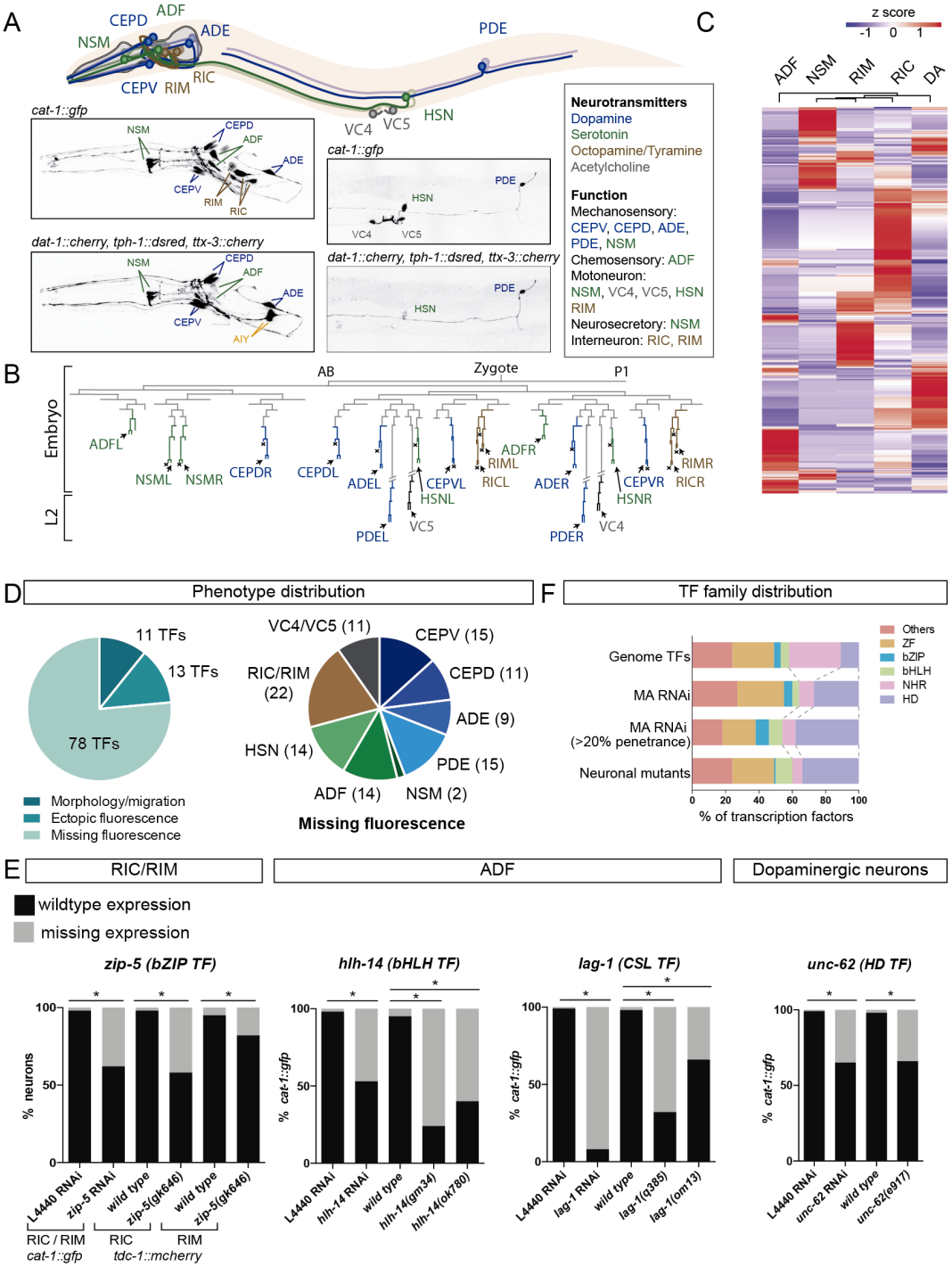

Figure 1

173

Figure 1. A genome wide transcription factor RNAi screen reveals specific TF-families are required for the specification of all monoaminergic 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, labelling 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 102 phenotypes as some TF RNAi clones are assigned to more than one cell and phenotypic category. Most neuron types are affected by knock down of at least 10 different TFs. We could not differentiate between RIC and RIM due to proximity and morphological similarity, thus they were scored as a unique category. See also **Table 1, Figure 1-S1 and Data Source 1** for details on TF RNAi library and raw data of RNAi screen results.
E) Mutant allele analysis confirms a developmental role for *zip-5* in RIC/RIM, *hlh-14* and *lag-1* in ADF and *unc-62* in dopaminergic neurons. n>50 each condition, *: p<0.05 compared to wild type, Fisher exact test with Bonferroni correction for multiple comparisons.
F) TF family distribution of TFs associated to neuronal phenotypes in our MA RNAi screen and in published mutant alleles (wormbase). Homeodomain TFs are over-represented and of NHR TFs decreased compared to the genome distribution. See also **Table 1** and **Figure 1-S2**.

174

175 **Table 1. TF RNAi screen results ordered by penetrance**
176

TF	TF family	Neuron type	RNA penetrance (%)	Mutant phenotype	Allele	Source
lin-32	bHLH	PDE	90	Yes	*e1926 gm239*	(Portman and Emmons, 2000)
lag-1	CSL	ADF	87	Yes	*om13 q385*	This work
lin-28	COLD BOX	PDE	82			
lin-39	HD	VCs	78	Yes	*n1760*	(Potts et al., 2009)
unc-4	HD	VCs	74	Yes	*e120, e2323*	(Zheng et al., 2013)
lin-40	ZF	VCs	74			
unc-86	HD	HSN	69	Yes	*n846*	(Lloret-Fernández et al., 2018; Sze et al., 2002)
hlh-14	bHLH	ADF	69	Yes	*ok780 gm34*	This work
sem-2	HMG	HSN	67			
tra-1	ZF	VCs	67			
hbl-1	ZF	PDE	63			
pax-3	HD	VCs	57			
ceh-20	HD	VCs	52	Yes	*ay42*	(Liu et al., 2006)
ham-1	WH	CEPs	51	Yes	*gt1984 ot339*	(Offenburger et al., 2017)
egl-5	HD	HSN	48	Yes	*n945*	(Singhvi et al., 2008)
ast-1	WH	HSN	42	Yes	*ot417 hd1 rh300 gk463*	(Lloret-Fernández et al., 2018)
vab-15	HD	PDE	41			
sex-1	NHR	HSN	41			
ast-1	ETS	ADE PDE	41	Yes	*ot417 hd1 rh300 gk463*	(Flames and Hobert, 2009)
nob-1	HD	HSN	40			
hbl-1	ZF	HSN	40			
unc-62	HD	ADE PDE	38	Yes	*e917*	This work
lin-14		HSN	35	Yes	*e912 n179*	(Olsson-Carter and Slack, 2010)
zip-5	bZIP	RIC/RIM	33	Yes	*ok646*	This work
bed-3	ZF	PDE	33			
lin-14		VCs	33	Yes	*n179*	This work
ceh-20	HD	ADE PDE	33	Yes	*mu290, ay42*	(Doitsidou et al., 2013)
sex-1	NHR	ADF	32			
sem-4	ZF	HSN	32	Yes	*n2654*	(Lloret-Fernández et al., 2018)
unc-86	HD	NSM	31	Yes	*n846*	(Sze et al., 2002)
ref-2	ZF	VCs	29			
lin-28	COLD BOX	HSN	29	Yes	*n719*	(Olsson-Carter and Slack, 2010)
lim-4	HD	ADF	28	Yes	*yz3 yz12*	(Zheng et al., 2005)
ceh-43	HD	CEPs ADE PDE	28	Yes	*ot406 ot340 tm480*	(Doitsidou et al., 2013)
hlh-3	bHLH	HSN	27	Yes	*tm1688*	(Lloret-Fernández et al., 2018)
lin-14		NSM	27	No	*n179*	This work
ceh-44	HD	PDE	27			
C32E8.1	bZIP	CEPs ADE PDE	25			
lin-26	ZF	PDE	23			

nhr-2	NHR	RIC/RIM	22			
vab-3	HD	CEPV ADE PDE	21	Yes	*ot266* *ot292* *ot346*	(Doitsidou et al., 2008)
egl-18	GATA	HSN	21	Yes	*n475*	(Lloret-Fernández et al., 2018)
ceh-16	HD	PDE	21	Yes	*ok841*	(Huang et al., 2009)
ceh-12	HD	RIC/RIM	21			
unc-62	HD	CEPs	21	Yes	*e917*	This work
wrm-1	β-catenin /Armadillo	VCs	20			
nhr-61	NHR	VCs	20			
egl-46	ZF	HSN	20	Yes	*sy628*	(Lloret-Fernández et al., 2018)
zip-10	bZIP	ADF	20	No	*ok3462*	This work
egl-43	ZF	RIC/RIM	19			
flh-1	ZF	ADF	19			
ceh-27	HD	RIC/RIM	19			
nhr-223	NHR	CEPs	19			
row-1	ZF	CEPV	18			
tbx-30	T - box	RIC/RIM	18			
ztf-8	ZF	CEPV	18			
unc-130	Fork Head	ADF	18			
dnj-17	ZF	CEPs	18	Yes	*ju1234*	This work
R05D3.3	ZF	ADF	17			
pqm-1	ZF	RIC/RIM	17	No	*ok485*	This work
nrh-148	NHR	RIC/RIM	17			
sdc-1	ZF	RIC/RIM	17			
sma-3	SMAD	RIC/RIM	17			
dxbp-1	ZF	CEPV	17			
lin-28	COLD BOX	VCs	17			
pqn-21	ZF	ADF	16			
scrt-1	ZF	RIC/RIM	16			
ceh-37	HD	ADF	16			
pha-4	GATA	HSN	15			
pax-3	HD	ADF	15			
ceh-27	HD	ADF	15			
mef-2	MADS BOX	CEPs ADE	15	Yes	*gk633*	This work
cep-1	p53	CEPs ADE	14	No	*ep347*	This work
vab-7	HD	RIC/RIM	14			
nfyb-1	CBF	RIC/RIM	14			
dro-1	CBF	CEPs ADE	14			
flh-2	ZF	ADF	14			
mls-2	HD	PDE	14	Yes	*cc615*	(Jiang et al., 2005)
odd-2	ZF	ADF	14			
ceh-13	HD	RIC/RIM	13			
zip-6	bZIP	RIC/RIM	13			
ceh-36	HD	RIC/RIM	13			
C27A12.2	ZF	RIC/RIM	13			
lin-48	ZF	CEPD	13			
nfx-1	NF - X1	RIC/RIM	12			
unc-55	NHR	CEPs	12	Yes	*e1170*	This work
sta-1	STAT	RIC/RIM	12			
sup-35	ZF	CEPV	11			
R05D3.3	ZF	RIC/RIM	11			
nhr-157	NHR	RIC/RIM	10			

Only missing fluorescence phenotypes are shown. Grey background highlights allele analysis performed in this work. Red font indicates RNAi results not reproduced by mutant analysis. 31 out of 35 tested RNAi phenotypes show similar results by mutant analysis (9 from this study, 22 from others). Most TFs with known functions on MA specification show RNAi penetrance of at least 20%. While most RNAi clones not validated by mutant analysis show low penetrance.

A specific set of transcription factor families controls monoaminergic neuron specification.

We next focused our analysis on TF families instead of individual TF members. According to their DNA binding domain, *C. elegans* TFs can be classified into more than fifty different TF families (Narasimhan et al., 2015; Stegmaier et al., 2004) (**Data source 1**). bHLH, HD, ZF, bZIP and NHR comprise 75% of the TFs in the *C. elegans* genome. RNAi clones targeting these families also generate 75% of the MA phenotypes (**Figure 1F**). However, we noticed that the prevalence of two TF families, HD and NRH TFs, differs from what would be expected from the number of TF members encoded in the genome (**Figure 1F**). In spite of representing only 12% of total TFs in the *C. elegans* genome, we found that 27% of RNAi clones showing a phenotype (21/78) belong to the HD family ($p<0.005$ Fisher exact test). In contrast, NHR TFs are under-represented when considering the total number of NHR TFs encoded in the genome (**Figure 1F**). We found that only 9% of NHR-TF RNAi clones produce MA phenotypes, even though 30% of all *C. elegans* TFs belong to this family ($p<0.0001$, Fisher exact test). NHR TF family has expanded in *C. elegans,* it is composed by 272 members compared to less than 50 NRH TFs encoded in the human genome. Only 8% of the 272 *C. elegans* NHRs have orthologs in non-nematode species (Maglich et al., 2001; Taubert et al., 2011). Importantly, we found NHRs with non-nematode ortholgs are enriched for neuronal functions (43% of NHR TFs associated to a MA RNAi phenotype are conserved, $p=0.04$, Fisher exact test). This observation suggests that among NHR members, those phylogenetically conserved have a prevalent role in neuron specification. The set of "high confidence" TFs with high penetrance phenotypes showed a similar TF family

distribution, with 37% HD TFs and 8% NHR TFs ($p<0.0001$, $p=0.0017$ respectively, Fisher exact test) (**Figure 1F**).

Strikingly, we noticed a roughly similar TF family distribution when considering each (MA) neuron type separately (**Figure 1- Figure Supplement 2**). This distribution is also present when considering TFs involved in the specification of the ASE glutamatergic gustatory neuron, a non-MA neuron for which a whole genome RNAi screen has been performed (Poole et al., 2011), but not in the regulation of other processes such as innate immune response or muscle specification (**Figure 1- Figure Supplement 2** and **Data source 1**).

Thus, our results suggest there is a specific TF family distribution associated to neuron specification, however, false negative and false positive RNAi rates could bias our interpretation. Thus we next turned into genetic mutant analysis, mutant alleles for 95 *C. elegans* TFs display reported neuronal phenotypes (annotated in Wormbase), including neuron specification defects, migration, axon guidance or behavioral deficits. Of note, TF family distribution for these 95 TFs is similar to the one observed in our MA RNAi screen (**Figure 1F**). HD family is also over-represented ($p<0.0001$, Fisher exact test) while NHR TFs are less present than what would be expected from the number of NHRs in the genome ($p<0.0001$, Fisher exact test). Five out of six NHR TFs producing known neuronal phenotypes have orthologs in non-nematode species, which constitutes a significant enrichment in conserved NHRs ($p<0.0001$ Fisher exact test). The exception being *odr-7*, a nematode specific NHR required for the terminal differentiation of AWA sensory neuron (Sengupta et al., 1994). In summary, this global mutant allele analysis validates the specific TF family distribution observed in our RNAi screen. Moreover, it also indicates that our

finding is not limited to the MA system and could apply generally in neuron-type specification programs.

Identification of new transcription factors involved in dopaminergic terminal differentiation

Next, we aimed to use our TF RNAi screen to specifically study the complexity of terminal differentiation programs. To date, at least one terminal selector has been assigned to 76 of the 118 *C. elegans* neuron types (Hobert, 2016). Of these, a maximum of three terminal selectors have been identified for each neuron. We recently found that a combination of six transcription factors work together as a TF 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 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.

It is important to notice that we cannot distinguish *a priori* what of the identified TFs act as early lineage regulators or at the terminal differentiation step, as both roles lead to the loss of fluorescence cells. To circumvent this limitation, we decided to focus on the four dopaminergic neuron types that arise from different lineages. Early lineage determinants affect unique dopaminergic neuron types, for example *ceh-16*/HD TF is required for V5 lineage asymmetric divisions and thus controls PDE generation but not other dopaminergic neuron types (Huang et al., 2009) (**Table 1**). Conversely, all mature dopaminergic neurons converge in the same terminal differentiation program (Doitsidou et al., 2013; Flames and

Hobert, 2009). Consistently, RNAi clones targeting the three known dopaminergic terminal selectors (*ast-1, ceh-43* and *ceh-20*) show phenotypes for at least two out of the four dopaminergic neuronal pairs (**Figure 2 - Figure Supplement 1**). Accordingly, we reasoned that RNAi clones leading to similarly broad dopaminergic phenotypes could also play a role in dopaminergic terminal differentiation.

In addition to the three already known dopaminergic terminal selectors, we found ten TF RNAi clones affecting two or more dopaminergic neuron types (**Table 1**). To prioritize among these ten TFs, we focused on *unc-62*/MEIS-HD, *vab-3*/PAIRED-HD, *unc-55*/NHR and *mef-2*/MADS. These four TFs have reported roles on neuron specification and many TFs involved in terminal differentiation have pleiotropic effects in several neuronal types (Hobert, 2016) (**Figure 2 - Figure supplement 1 and Data source 1**). Thus, we next aimed to study if *unc-62*/MEIS-HD, *vab-3*/PAIRED HD, *unc-55*/NHR and *mef-2*/MADS play a role in dopaminergic terminal differentiation.

unc-62/MEIS-HD and _vab-3_/PAIRED-HD TFs are required for dopaminergic lineage specification and dopaminergic terminal differentiation

We used mutant alleles to verify RNAi phenotypes and further characterize the role of these TFs in dopaminergic fate. *unc-62* has multiple functions in development and null alleles are embryonic lethal precluding analysis of dopaminergic differentiation defects (Van Auken et al., 2002). Three viable hypomorphic alleles *e644*, *mu232* and *e917* show expression defects of a *cat-2*/tyrosine hydroxylase reporter, the rate-limiting enzyme for dopamine synthesis (**Figure 2A, B, E** and **Figure 2 - Figure Supplement 2**). For further

characterization, we focused on *unc-62(e917)* allele as it showed higher penetrance. *unc-62(e917)* mutant shows broad defects in expression of the dopamine pathway genes in ADE and PDE (*bas-1, cat-2, cat-4* and *cat-1* reporter expression affected) while CEPV shows small but significant defects in *cat-2* and *cat-1* reporter expression (**Figure 2C-G**). Expression of additional effector genes not directly related to dopaminergic biosynthesis, such as the ion channel *asic-1*, is also affected in *unc-62(e917)* mutants (**Figure 2H**).

Despite the broad ADE phenotypes of *unc-62(e917), dat-1* gene expression is unaffected in this neuron revealing the presence of the cell and discarding that the lineage is affected. In contrast, *unc-62(e917)* shows loss of PDE expression for all analyzed reporters raising the possibility of lineage specification defects. Three additional observations support this idea: 1) *unc-62(e917)* mutants show correlated *asic-1* reporter expression defects in PDE and its sister cell PVD (26/60 animals *asic-1::gfp* lost in both cells, 34/60 unaffected expression in both, 0/60 animals only PDE or PVD expression affected), suggesting either *unc-62* is required for terminal differentiation of both cells or that the lineage is affected; 2) *unc-62(e917)* mutant animals show PDE expression defects for the ciliated marker *ift-20* and the panneuronal reporter *rab-3* (**Figure 2 - Figure supplement 2**), which are usually unaffected in terminal differentiation mutants (Flames and Hobert, 2009; Stefanakis et al., 2015). These results also suggest the PDE neuron itself is absent in *unc-62(e917)* mutants; 3) Finally, *ceh-20*/PBX HD dopaminergic terminal selector is required for correct PDE lineage formation (Doitsidou et al., 2013). MEIS and PBX factors are known to physically and genetically interact in different tissues in several model organisms (Jiang et al., 2009; Knoepfler et al., 1997; Maeda et al., 2002;

Noman et al., 2017; Potts et al., 2009; Rieckhof et al., 1997; Vlachakis et al., 2001). Thus one possibility is that CEH-20 and UNC-62 work together in the PDE lineage determination. Altogether, our results are consistent with a dual role for *unc-62* in ADE and CEPV terminal differentiation and in PDE lineage specification. The early role in PDE lineage formation precludes the assignment of later roles for *unc-62* in this dopaminergic neuron type.

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). 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). Our RNAi experiments reproduce the mixed phenotype of extra and missing CEPs but also displays missing reporter expression for ADE and PDE (**Figure 2 - Figure supplement 1**). Thus, to investigate if, as suggested by RNAi experiments, *vab-3* is also required for dopaminergic terminal fate, we analyzed the expression of all dopamine pathway gene reporters in *vab-3(ot346).* We found that, in addition to the already reported mixed phenotype of extra and missing *dat-1::gfp* expressing CEPs, *vab-3(ot346)* animals show significant *cat-4* and *bas-1* expression defects in ADE and *cat-2* expression defects in ADE and PDE (**Figure 2D, E, F**), supporting a broad role for *vab-3* in dopaminergic terminal differentiation. Importantly, *dat-1* and *cat-1* reporter expression is unaffected in ADE and PDE which rules out lineage defects as the underlying cause of the phenotype for these neuron types (**Figure 2C, G**). We confirmed CEP lineage defects in *vab-3(ot346)* mutants with the analysis of *gcy-36* and *mod-5*

expression, effector genes expressed in URX (sister cell of CEPD) and AIM (cousin of CEPV) respectively (**Figure 2 - Figure supplement 2**).

Thus, similar to *ceh-20* and *unc-62*, *vab-3* seems to have a dual role in dopaminergic specification: it is required for proper ADE and PDE terminal differentiation and for correct CEP lineage generation. A potential role for *vab-3* in CEPs terminal differentiation could be masked by the earlier requirement in lineage determination.

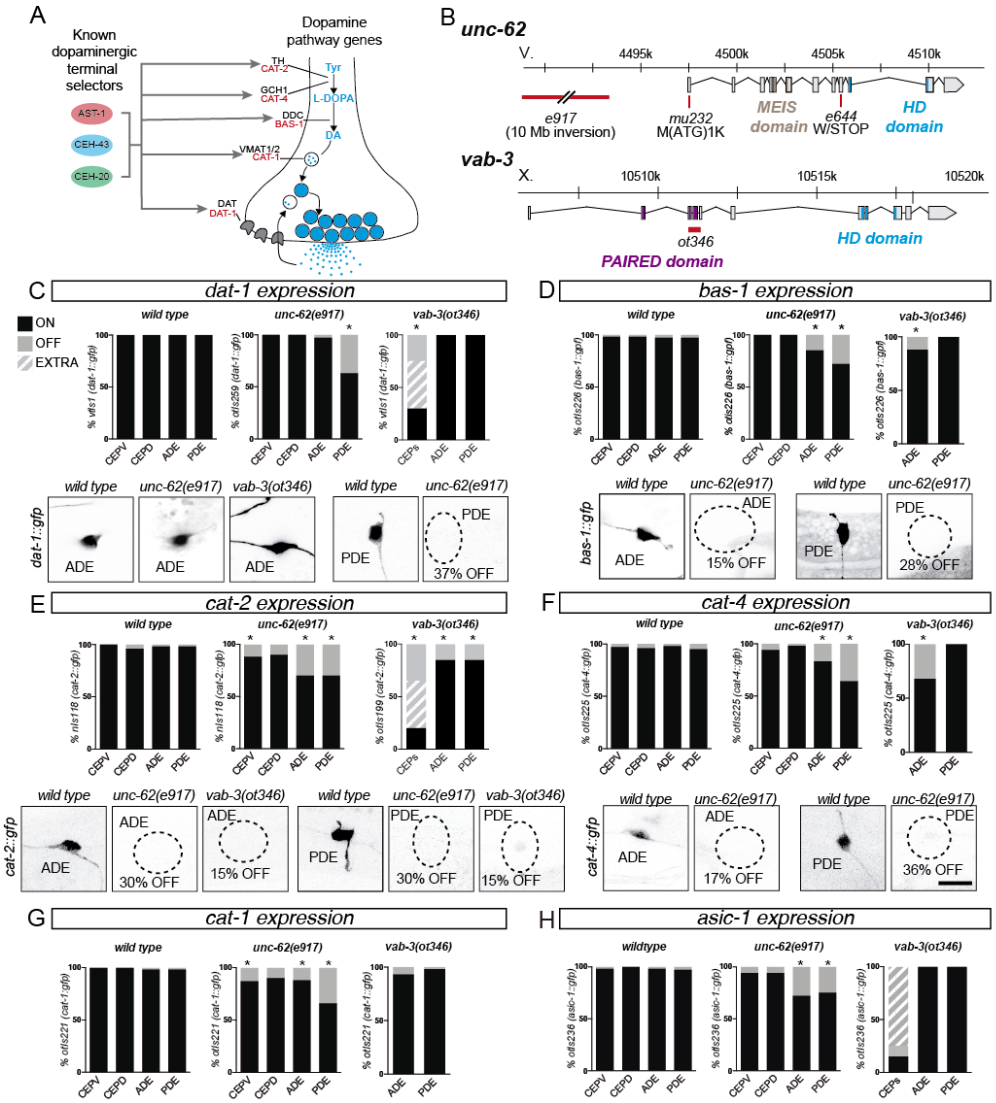

Figure 2

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 dopaminergic neurons 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 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 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 wildtype. Fisher exact test. Scale: 10 µm. See **Figure 2-S2** for analysis of additional alleles, markers and lineage defects.

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

***unc-55*/NHR and *mef-2*/MADS TFs provide robustness to the dopaminergic terminal differentiation program**

Next, we analyzed dopamine pathway gene expression in two different mutant alleles for *unc-55* and *mef-2* (**Figure 3A**). Similar to RNAi effects (**Figure 2 - Figure Supplement 1**), *mef-2(gk633)* displays significant defects in *cat-1::gfp* expression, however, these defects are not observed in *mef-2(gv1)* mutants (**Figure 3B**). Both alleles are predicted nulls, *mef-2(gk633)* deletes the promoter and first exon abolishing transcription, while *mef-2(gv1)* produce a truncated transcript that would be degraded by non-sense mediated decay (**Figure 3A**). Some alleles that activate the non-sense mediated decay in zebrafish and mouse have been recently shown to activate expression of compensatory genes inducing what has been termed transcriptional adaptation (El-Brolosy et al., 2019). Transcriptional adaptation is also present in *C. elegans* (Serobyan et al., 2020), and then it is possible that the lack of dopaminergic phenotype in *mef-2(gv1)* mutants compared to *mef-2(gk633)* is due to transcriptional adaptation.

In contrast, neither of the two predicted *unc-55* null alleles, the early stop *unc-55(e1170)* that induce non-sense mediated decay and the *unc-55(gk818)* deletion allele that likely abolish all *unc-55* transcription, reproduce the *cat-1* expression defects detected by RNAi experiments (**Figure 3B**). Similarly, expression of other dopaminergic effector genes is also unaffected in *unc-55(e1170)* and *unc-55(gk818*) mutants (**Figure 3 - Figure Supplement 1**). This discrepancy suggests that *unc-55* RNAi phenotype could be a false positive due to off target RNAi effects, alternatively the total absence of *unc-55* could be genetically compensated in both mutant alleles while, for currently unknown

366 reasons, *unc-55* RNAi in *rrf-3* background cannot be totally compensated. Thus,
367 we hypothesized single mutations in *mef-2* and *unc-55* genes could display
368 small or no-phenotypes due to mutational robustness of the dopaminergic gene
369 regulatory network. Double mutant analysis shows this to be the case: *unc-*
370 *55(e1170)* shows strong synergism with both *ast-1(hd1)* and *ceh-43(ot406)*
371 mutants in the regulation of *dat-1::gfp* expression in PDE (**Figure 3C-D**), while
372 *mef-2(gk633)* and *ast-1(hd1)* act synergistically in the regulation of *cat-1::gfp*
373 expression in CEPs and ADE **(Figure 3E-F and Figure 3 - Figure supplement**
374 **2**).
375 In summary, we find *unc-55* and *mef-2* RNAi experiments and mutant analysis
376 show differing results, the nature of these discrepancies is currently not well
377 understood and could be explained in part by transcriptional adaptation
378 mechanisms. Single mutant analysis reveals that the lack of these two factors
379 individually can be compensated, however, synergistic effects found with known
380 dopaminergic terminal selectors suggest a *bona fide* requirement for *unc-55*
381 and *mef-2* in dopaminergic terminal specification.

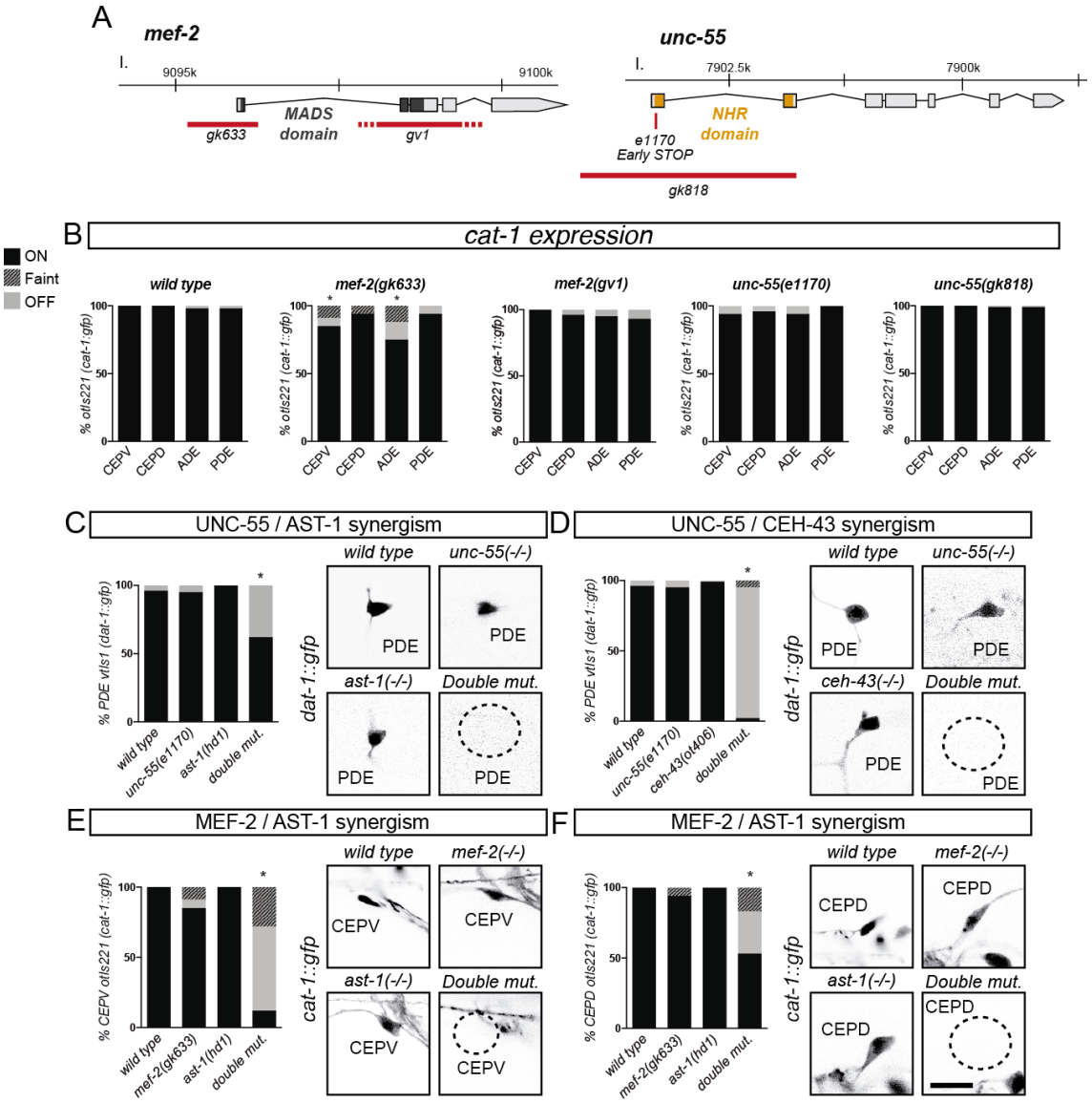

Figure 3

Figure 3. *mef-2*/MADS and *unc-55*/NHR TFs provide robustness to the dopaminergic differentiation program.

A) Schematic representation of *mef-2* and *unc-55* gene loci and alleles used in the analysis. All alleles are predicted nulls, *mef-2(gv1)* and *unc-55(e1170)* alleles are predicted to induce RNA decay response while *mef-2(gk633)* and *unc-55(gk818)* produce no mRNA transcript. *unc-55(gk818)* deletion also removes F55D12.6 gene that is located inside *unc-55* first intron.

B) *cat-1* expression defects produced by *mef-2* and *unc-55* RNAi are not reproduced in *mef-2(gv1)*, *unc-55(e1170)* and *unc-55(gk818)*. *mef-2(gk633),* shows similar *cat-1* expression defects to RNAi experiments, supporting a role for this TF in dopaminergic terminal differentiation. n>50 each condition, *: p<0.05 compared to *wildtype*. Fisher exact test. See also **Figure 3-S1,** for analysis of additional markers in *unc-55* and *mef-2* single mutants.

C-F) Double mutant analysis of *unc-55* and *mef-2* with known dopaminergic terminal selectors *ast-1* and *ceh-43* reveals synergistic effects. n>50 each condition, *: p<0.05. Fisher exact test. Scale: 10 µm. See also **Figure 3-S2** for additional double mutant analysis.

***Cis*-regulatory modules of dopaminergic effector genes contain functional binding sites for UNC-62, VAB-3, UNC-55 and MEF-2**

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

The *dat-1* minimal *cis*-regulatory module (*dat-1p2*) contains previously described functional binding sites (BS) for AST-1 (ETS BS), CEH-20 (PBX BS) and CEH-43 (HD BS) (**Figure 4A**) (Doitsidou et al., 2013; Flames and Hobert, 2009). In addition, we found consensus binding sites for MEF-2 (MADS BS), UNC-55 (COUP-TF BS), UNC-62 (MEIS BS) and VAB-3 (PAIRED BS) (**Figure 4A**). Site directed mutagenesis of predicted MADS, COUP-TF, MEIS and PAIRED BS leads to GFP expression defects, mainly in the PDE (**Figure 4A** and **Figure 4 - Figure supplement 1**). VAB-3 protein contains two DNA binding domains, PAIRED and HD where each DNA domain fulfills specific functions (Brandt et al., 2019). Interestingly, we find that one of the two already identified functional HD BS matches a PAIRED-type HD consensus (HTAATTR, labeled as HD* in **Figure 4**) suggesting it could be recognized by VAB-3 and/or CEH-43. Finally, combined COUP-TF and HD* BS mutation shows synergistic effects in the ADE (**Figure 4A and Figure 4 - Figure supplement 1**).

The *cat-2* minimal *cis*-regulatory module (*cat-2p21*), in addition to the previously described ETS, PBX and HD BS (Doitsidou et al., 2013; Flames and Hobert,

2009) also contains predicted MADS, MEIS and PAIRED BS but lacks any predicted COUP-TF site (**Figure 4B**). Point mutation in PAIRED and MEIS BS show strong GFP expression defects while MADS BS mutation had only a small effect in the PDE (**Figure 4B** and **Figure 4 - Figure supplement 1**). Similar to minimal *dat-1 cis*-regulatory module, one of the predicted HD BS matches the PAIRED HD consensus (HTAATTR, labeled as HD* in **Figure 4**).

Our results suggest that the four additional dopaminergic TFs (UNC-62, VAB-3, MEF-2 AND UNC-55) may act together with other known terminal selectors (AST-1, CEH-20 and CEH-43) to directly activate dopaminergic effector gene expression.

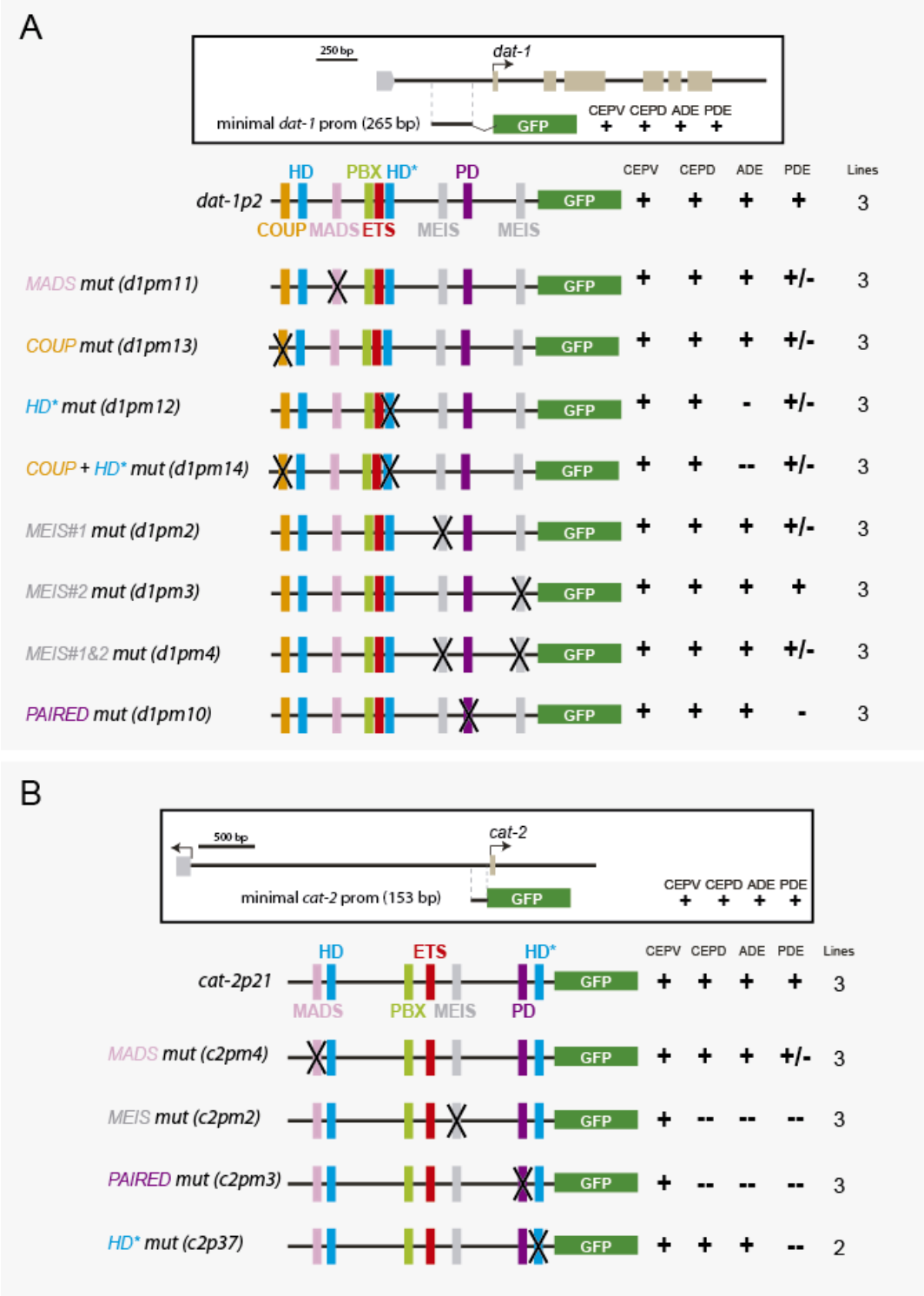

Figure 4

418

Figure 4. *cis*-regulatory analysis of dopamine pathway genes reveals functional binding sites for UNC-62, VAB-3, UNC-55 and MEF-2.
A) *dat-1* minimal dopaminergic *cis*-regulatory module (*dat-1p2*) mutational analysis. *dat-1p2* contains predicted AST-1/ETS, CEH-20/PBX HD, CEH-43/DLL HD, UNC-62/MEIS, VAB-3/PAIRED, UNC-55/COUP-TF and MEF-2/MADS binding sites. Functionality of ETS, PBX and HD sites has been previously shown (Doitsidou et al., 2013; Flames and Hobert, 2009). Point mutation of MADS, COUP-TF, MEIS and PAIRED (PD) BS also affect GFP expression in dopaminergic cells. 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. See **Figure 5 - Figure Supplement 1** for raw values and nature of the mutations
B) *cat-2* minimal dopaminergic *cis*-regulatory module (*cat-2p21*) mutational analysis. In addition to published functional ETS, PBX and HD binding sites (Doitsidou et al., 2013; Flames and Hobert, 2009), MADS, MEIS and PAIRED (PD) BS are also required for correct GFP reporter expression. 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. See **Figure 5 - Figure Supplement S1** for raw values and nature of the mutations

VAB-3, UNC-55 and MEF-2 act cell-autonomously in dopaminergic neurons

Next, we aimed to explore if the new identified factors act cell autonomously for dopaminergic fate induction. We found *unc-62, vab-3, mef-2* fosmid reporters and a transcriptional reporter for *unc-55* are expressed in dopaminergic neurons and expression is maintained throughout the life of the animal (**Figure 5A**). At young adult stage *unc-62* expression was not detected in CEPs, which is consistent with the stronger ADE and PDE phenotype of *unc-62* mutants.

Next, we investigated epistatic relationships among members of the dopaminergic terminal regulatory program. AST-1 is the TF with the strongest dopaminergic phenotype, and all analyzed dopaminergic effector genes are lost in *ast-1(hd92)* null mutants (Flames and Hobert, 2009). We thus aimed to study if AST-1 acts upstream of any of the newly identified dopaminergic TFs. *ast-1(hd92)* animals arrest at fist larval stage before PDE neuron birth and show morphological head disorganization and loss of expression for all dopaminergic

pathway genes. *unc-62, vab-3* and *mef-2* TFs are expressed in many neurons in the head in addition to dopaminergic neurons, thus, to identify dopaminergic head neurons we used a *ift-20* pancilia reporter for which we detected expression in approximately 30 neurons in the head, including the six dopaminergic neurons. We first confirmed that the total number of *ift-20* positive cells is similar in *ast-1(hd92)* and *wild type* worms, this was expected as *ast-1* does not regulate expression of cilia components (Flames and Hobert, 2009). Next we found *ast-1(hd92)* mutant animals show similar numbers of double positive *ift-20*/*unc-62* and *ift-20*/*mef-2* cells to controls, suggesting *ast-1* is not required for the expression of these TFs in dopaminergic neurons (**Figure 5B**). In contrast, *ast-1(hd92)* shows a small but significant decrease in the number of *ift-20* cells co-expressing *vab-3* (**Figure 5B**). A mean of two *ift-20* cells lose *vab-3* expression in *ast-1(hd92)* mutants indicating either a low penetrance for the phenotype or that *ast-1* is upstream *vab-3* only in one of the three pairs of dopaminergic neurons in the head. We could not distinguish between the two possibilities due to morphological disorganization of the head in *ast-1(hd92)* mutants. Finally, *unc-55* expression in the head was too faint to perform this quantitative analysis. In summary, it seems that, *unc-62, vab-3* and *mef-2* act mostly in parallel to *ast-1.*

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

460 **2C**) however we find none of the two transgenic lines produce significant rescue
461 of the *unc-62(e971)* phenotype (**Figure 5C**), failure to rescue could indicate
462 *unc-62* might be required in the ADE earlier than the onset of *dat-1* expression,
463 alternatively it could act non cell autonomously or require additional isoforms.
464 A similar strategy was used for *vab-3(ot346)* rescue experiments. *vab-3* codes
465 for three isoforms, but only the long isoform *vab-3*a contains both the PAIRED
466 and the HD protein domains. Expression of *vab-3a* under the *dat-1* promoter is
467 sufficient to rescue *cat-2* expression defects in ADE, demonstrating a cell
468 autonomous and terminal role for *vab-3* in neuron specification (**Figure 5D**). As
469 expected, *vab-3(ot346)* CEP lineage defects were not rescued, as the promoter
470 used for *vab-3* rescue (*dat-1prom*) is only expressed in terminally differentiated
471 CEPs.
472 For *mef-2* and *unc-55* rescue experiments we used double mutants with *ast-*
473 *1(hd1),* as single mutants do not show obvious dopaminergic expression
474 defects. *unc-55* codes for two isoforms. Expression of *unc-55* long isoform (*unc-*
475 *55a*) did not produce significant rescue of *dat-1::gfp* expression in the PDE
476 (**Figure 5E**). Although the other isoform, *unc-55b*, has been reported to be
477 expressed only in males (Shan and Walthall, 2008), it is possible that *unc-55b*
478 could be required in hermaphrodite dopaminergic neuron specification.
479 Alternatively, earlier *unc-55* expression or non-cell autonomous actions for *unc-*
480 *55* could explain failure to rescue with our experimental approach. Nonetheless,
481 as will be explained in the following section, the mouse ortholog of *unc-55*,
482 Nr2f1 (Coup-tf1) is able to rescue *ast-1(hd1), unc-55(e1170)* double mutant
483 phenotype (**Figure 5E).** Nr2f1 rescue of the dopaminergic phenotype supports

484 a cell autonomous and terminal role in dopaminergic specification for these
485 NHR TFs.
486 Finally, *mef-2* codes for a single isoform, we find *mef-2* cDNA expression under
487 the *bas-1* promoter rescues loss of CEPV and CEPD *cat-1::gfp* expression in
488 *ast-1(hd1); mef-2(gk633)* double mutants, demonstrating a cell autonomous
489 and terminal role for *mef-2* in dopaminergic neuron specification (**Figure 5F**).

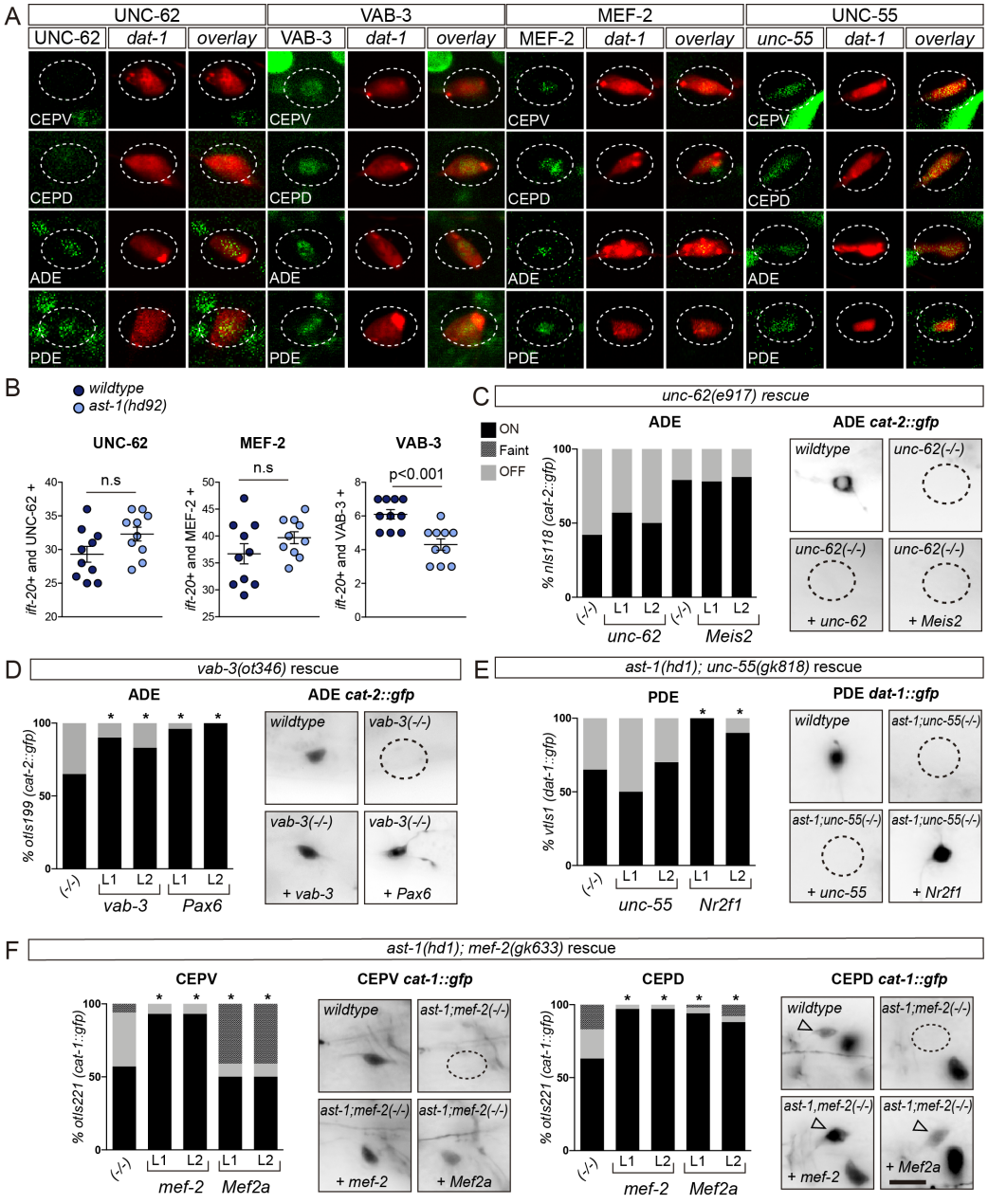

Figure 5

Figure 5. The dopaminergic terminal differentiation program is phylogenetically conserved.

A) Expression analysis of UNC-62 (*wgIs600),* VAB-3 (*vlcEx1046*) and MEF-2 (*wgIs301)* GFP tagged fosmid reporter and *unc-55::gfp* (*vlcEx498*) transcriptional reporter at L4 larval stage shows co-localization with *dat-1::cherry (otIs181)* dopaminergic reporter. UNC-62 expression was only detected in ADE and PDE at this developmental stage.

B) Expression analysis of head sensory cells (*ift-20* positive) co-expressing UNC-62, VAB-3 and MEF-2 in wildtype and *ast-1(hd92)* null mutants. n=10 each genotype. Two-tail unpaired T-test.

C-F) Cell autonomous rescue of *unc-62, vab-3* single mutants and *ast-1;mef-2* and *ast-1; unc-55* 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;mef-2* and *ast-1;unc-55* double mutants. L1 and L2 represent two independent transgenic lines. Arrowheads in D 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.

The dopaminergic terminal differentiation program is phylogenetically conserved

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*, Pax6, the mouse ortholog of *vab-3* and Nr2f1, the mouse ortholog of *unc-55* are also necessary for olfactory bulb dopaminergic specification (Agoston et al., 2014; Bovetti et al., 2013; Brill et al., 2008). Thus we next performed similar rescue experiments using mouse orthologs for the new dopaminergic TFs. Pax6, Mef2a and Nr2f1 are able to rescue *vab-3, mef-2* and *unc-55* mutant phenotypes respectively. However, Meis2 did not produce significant rescue of *unc-62* mutants, similar to *unc-62* cDNA failure to rescue, suggesting MEIS factors might be also required at earlier time points (**Figure 5 C-F**).

Altogether, our data reinforce a model in which known dopaminergic terminal selectors *ast-1*/Etv1, *ceh-43*/Dlx2 and *ceh-20*/Pbx1, act together with *unc-62*/Meis2, *vab-3*/Pax6, *mef-2*/Mef2a and *unc-55*/Nr2f1 to regulate dopaminergic terminal differentiation.

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

Thus at least seven TFs seem to be required for correct *C. elegans* dopaminergic terminal specification. A seemingly complex combination of TFs

activates expression of the mature HSN transcriptome (Lloret-Fernández et al., 2018). Our findings suggest this TF complexity might be a common theme in neuronal terminal differentiation programs. Then, why are these complex combinations of TFs required for gene expression? Our *unc-55* and *mef-2* analysis shows that one possible reason might be ensuring robustness of expression. In addition, we previously described that TFBS clusters for the HSN TF collective are preferentially found in HSN expressed genes and can be used to *de novo* identify HSN active enhancers (Lloret-Fernández et al., 2018). Thus, a second function for this complex terminal differentiation programs might be providing enhancer selectivity.

To test this hypothesis we next asked if the dopaminergic regulatory program imposes a defining regulatory signature in dopaminergic expressed genes. Published single-cell RNA-seq data (Cao et al., 2017) was used to identify additional genes differentially expressed in dopaminergic neurons (**Figure 6 - Figure supplement 1**). We found 86 genes whose expression is enriched in dopaminergic neurons compared to other clusters of ciliated sensory neurons. As expected, this gene list includes all dopamine pathway genes and other known dopaminergic effector genes, but not pancilia expressed genes (**Data source 2**). In analogy to our previous analysis of the HSN regulatory genome (Lloret-Fernández et al., 2018), we analyzed the upstream and intronic sequences of these genes. For comparison purposes, we built ten thousand sets of 86 random genes with similar upstream and intronic length distribution to dopaminergic expressed genes.

First, we focused our analysis only on the three already published dopaminergic terminal selectors (*ast-1/ETS, ceh-43/DLL HD* and *ceh-20/PBX HD*). We found

this regulatory signature (ETS+HD+PBX binding sites) lacks enough specificity as all genes (either dopaminergic expressed genes or random sets) contain DNA windows with matches for all three TFs (100% of dopaminergic expressed genes compared to 100% in random sets). Reducing DNA-window search length from 700 bp to 300 bp or 150 bp did not increase specificity (100% of dopaminergic expressed genes compared to 100% in random sets). Next, we expanded our analysis to DNA regulatory windows containing at least one match for each of the eight position weight matrices associated to the dopaminergic TF terminal regulatory program (**Figure 6A**). We found that seventy-eight percent of dopaminergic expressed genes contain at least one associated dopaminergic regulatory signature window, a significantly higher percentage compared to the random sets of genes (mean random sets 60%, $p<0.001$) (**Figure 6B**). Thus, dopaminergic signature is significantly enriched in dopaminergic-expressed genes. Next, we built similar sets of differentially expressed genes for five randomly picked non-dopaminergic neuron categories (RIA, ASE, Touch Receptor neurons, GABAergic neurons and ALN/PLN/SDQ cluster) (**Figure 6 - Figure supplement 1** and **Data source 2**). We found that the percentage of genes containing the dopaminergic signature is smaller in non-dopaminergic neurons compared to dopaminergic neurons (**Figure 6B**). This difference is statistically significant for all neurons except the ALN/PLN/SDQ neuron cluster (**Figure 6B**). In addition, none of the non-dopaminergic neuronal types show a significant enrichment of the dopaminergic signature with respect to their respective background of ten thousand random sets of comparable genes (**Figure 6B**, $p>0.05$). The reason why ALN/PLN/SDQ expressed genes show a higher association to dopaminergic regulatory

signature compared to other non-dopaminergic neurons is uncertain. Terminal selectors for ALN/PLN/SDQ neurons are yet unknown, a similar combination of TF families controlling terminal differentiation of these neurons could explain the higher presence of the dopaminergic regulatory signature.

Importantly, dopaminergic expressed genes are specifically enriched for the dopaminergic regulatory signature, as the HSN regulatory signature, that is, the presence of TFBS for the six members of the HSN TF collective (Lloret-Fernández et al., 2018), is not enriched in dopaminergic expressed genes **(Figure 6 - Figure supplement 1**).

Dopaminergic regulatory signature enrichment in dopaminergic expressed genes is even more pronounced when considering only the promoter sequence (1.5 Kb upstream of the ATG) (**Figure 6C**). These data shows that proximal regulation has a major role in neuronal terminal differentiation in *C. elegans*. The full complement of dopaminergic TFBS is required to provide specificity to the dopaminergic regulatory signature as regulatory windows containing only seven or six types of dopaminergic TF motifs are not preferentially found in dopaminergic expressed genes (**Figure 6 - Figure supplement 1**).

Four out of the five dopamine pathway genes contain at least one associated dopaminergic regulatory signature window. Remarkably, predicted regulatory windows overlap with the previously isolated *cis*-regulatory modules (**Figure 6D**) (Flames and Hobert, 2009).

Altogether, our data suggest that the dopaminergic TF collective acts through the dopaminergic regulatory signature to activate transcription of dopaminergic effector genes. However, the presence of the dopaminergic signature in some genes not expressed in dopaminergic neurons indicates the signature itself is

591 not sufficient to induce dopaminergic expression. Additional TFs, gene
592 repression mechanisms, or chromatin accessibility could further regulate
593 dopaminergic regulatory signature specificity. It is also possible that specific
594 syntactic rules (TFBS order, distance and disposition) discriminate functional
595 from non-functional dopaminergic regulatory signature windows.

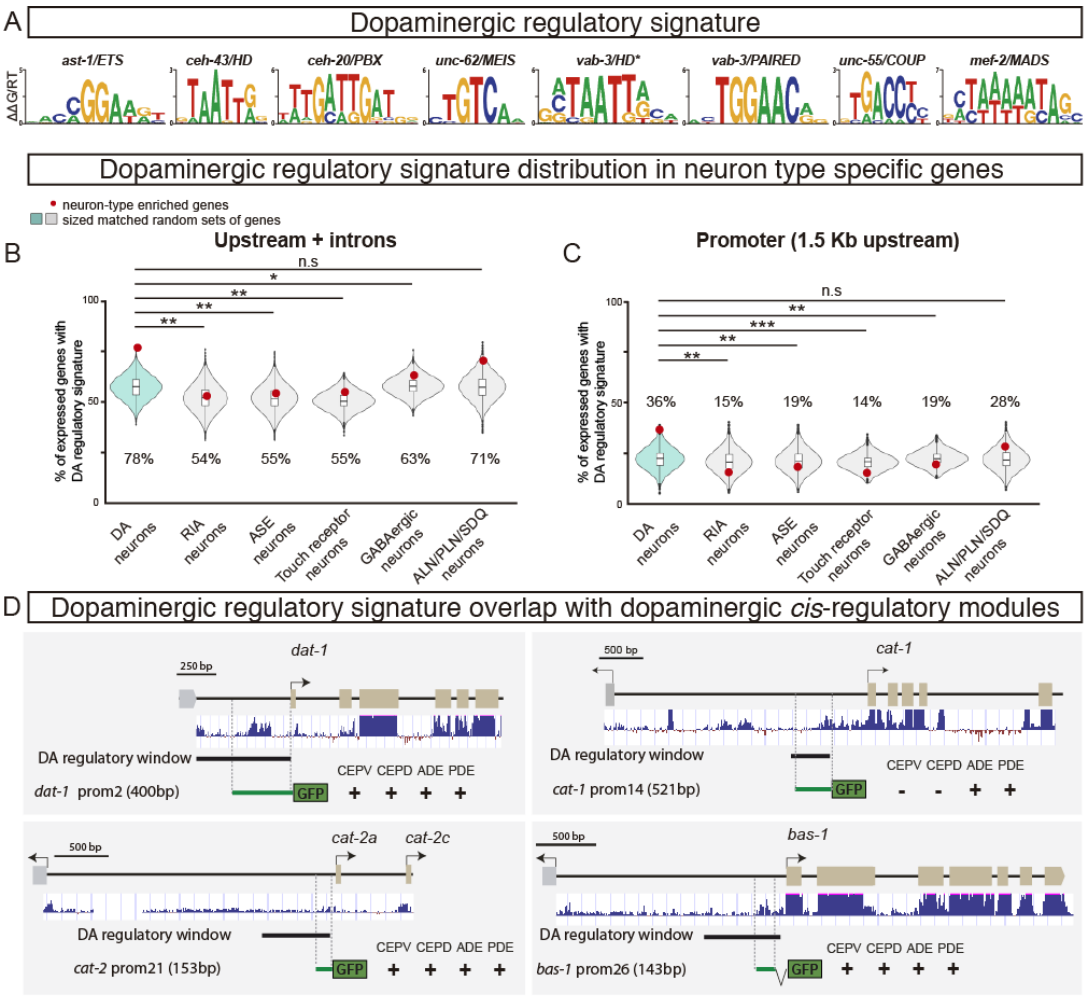

Figure 6

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 at least one match to each PWM in less than 700bp DNA window.

B) Dopaminergic regulatory signature is more prevalent in the upstream and intronic sequences of the set of 86 genes enriched in dopaminergic neurons (red dot) compared to the distribution in 10,000 sets of random comparable genes (blue violin plot) ($p<0.001$). 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. Brunner-Munzel test. *: $p<0.05$, **: $p<0.01$. See **Figure 6-S1** and **Data Source 2** for additional analysis of dopaminergic regulatory signature distribution and gene lists. Expression data obtained from (Cao et al., 2017).

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) Predicted dopaminergic regulatory signature windows overlap with the experimentally isolated minimal enhancers for four out of the five dopamine pathway genes (Flames and Hobert, 2009). Black lines represent the coordinates covered by bioinformatically predicted dopaminergic regulatory signature windows. Green lines mark published minimal enhancers for the respective gene. 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.

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

Next, we examined the distribution of the dopaminergic regulatory signature across the entire *C. elegans* genome. We found it is preferentially present in putative regulatory sequences of neuronally expressed genes compared to the rest of the genome (**Figure 7A** and **Data source 2**). This enrichment is also present when analyzing only promoter sequences (**Figure 7A**).

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

Different hierarchies of gene expression co-exist in any given cell, mechanosensory dopaminergic neurons co-express at least four types of genes: 1) Dopaminergic effector genes such as dopaminergic pathway genes, neuropeptides, neurotransmitter receptors, etc., that are preferentially expressed by this neuron type; 2) Genes coding for structural components of cilia which are expressed by all sixty ciliated sensory neurons in *C. elegans*; 3) Panneuronal genes, such as components of the synaptic machinery or cytoskeleton, expressed by all 302 neurons in *C. elegans* hermaphrodite and; 4) Ubiquitous genes expressed by all cells of the organism, such as ribosomal or heat shock proteins (**Figure 7C**). Thus, we next studied the distribution of the dopaminergic regulatory signature in these parallel gene categories.

We find that the percentage of genes with dopaminergic signature is higher for dopaminergic enriched effector genes compared to other parallel routines in the cell (**Figure 7C**). This difference is also present when only promoter sequences are analyzed (**Figure 7C**). Of note, although lower than dopaminergic effector genes, panneuronal genes also show increased presence of dopaminergic regulatory signature compared to other regulatory routines of the cell (**Figure 7C**) suggesting dopaminergic TFs could partially contribute to panneuronal gene control, consisting with previous reports on panneuronal gene regulation (Stefanakis et al., 2015). Interestingly, dopaminergic regulatory signature associated to panneuronal genes seems to be mostly located outside promoter regions (>1.5 Kb from ATG), which is different to dopamine effector genes. The frequency of dopaminergic regulatory signature in panciliated and ubiquitous genes is similar to genes not expressed in neurons. Thus, our data suggest that the main function for the dopaminergic TF collective is the regulation of neuron-type specific gene expression.

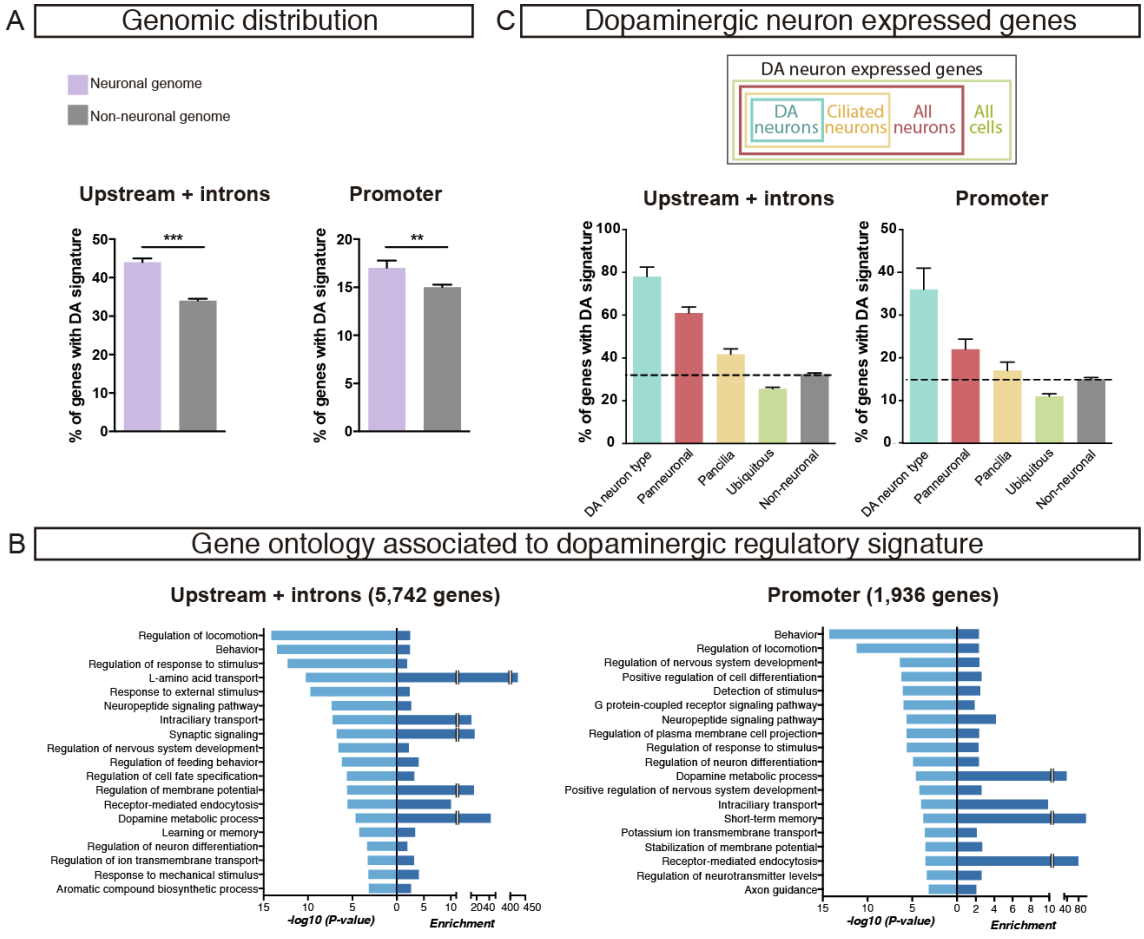

Figure 7

Figure 7. Dopaminergic regulatory signature is enriched in dopaminergic effector genes but not in other parallel regulatory routines of the cell.
A) Genomic search of the dopaminergic regulatory signature reveals enrichment in neuronal 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. See also **Data Source 2**.
B) Gene ontology analysis of genes with associated dopaminergic regulatory signature. p-values and enrichment of genes with dopaminergic signature in the corresponding GO category are represented.**C**) Parallel gene expression routines 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 distal regions of panneuronal genes but it is not present in other parallel routines of dopaminergic neurons. 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.

637

DISCUSSION

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

Focusing on the MA superclass of neurons we provide a comprehensive view of the TFs required at any developmental step in the specification of ten different neuronal types. We did not find any global regulator of MA fate, which is likely due to the molecular and functional diversity found in this superclass of neurons. Each neuron type is regulated by different sets of TFs, however, we uncovered general rules shared by all MA neuron types that might also apply to non-MA neurons. First, we identified at least 10 different TFs required in the specification of each neuron type (with the exception of the NSM neuron, which might be particularly unresponsive to RNAi), evidencing that complex gene regulatory networks underlie neuron type specification. Second, in spite of this TF diversity, most TFs involved in the specification of all MA neurons consistently belong to only five out of the more than fifty *C. elegans* TF families: HD, bHLH, ZF, bZIP and non-nematode-specific members of the NHR family. Importantly, analysis of genetic mutants displaying neuronal phenotypes reveals the same TF family distribution, which does not only validate results obtained from the RNAi screen but also expands these findings to other neuronal types.

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

identified in our study. The ancestral role of bHLH TFs as proneural factors has been established in a wide range of metazoans including vertebrates, Drosophila, *C. elegans* and the cnidaria *Nematostella vectensis* which has one of the most simple nervous systems (Guillemot and Hassan, 2017; Layden et al., 2012; Lloret-Fernández et al., 2018; Poole et al., 2011). HD TFs are involved in many developmental processes and their role in neuron terminal differentiation is also conserved in different animal groups including cnidaria (Babonis and Martindale, 2017; Briscoe et al., 2000; Hobert, 2016; Shirasaki and Pfaff, 2002; Thor et al., 1999; Tournière et al., 2020). Several NHRs, bZIP and ZF TFs are known to regulate neuron-type specification in mammals, such as Couptf1, Nurr1, Tlx and Nr2e3 NHR TFs (Bovetti et al., 2013; Haider et al., 2000; Roy et al., 2004; Zetterström et al., 1997), Nrl, cMaf and Mafb bZIP TFs (Blanchi et al., 2003; Mears et al., 2001; Wende et al., 2012) or Myt1l, Gli1, Sp8, Ctip2, Fezf1/2 ZF TFs (Arlotta et al., 2005; Hynes et al., 1997; Mall et al., 2017; Shimizu et al., 2010; Waclaw et al., 2006). However, in contrast to bHLH and HD, a general role in neuron specification for these families has been poorly studied in any model organism. Functional characterization of the NHR, bZIP and ZF TF candidates retrieved from our RNAI screen will help better understand the role of these TF families in neuron specification.

In addition to the results obtained from our TF RNAi screen, the complete and fully sequence-verified TF RNAi library is a new valuable resource for the community that could be used for the identification of gene regulatory networks involved in many different biological processes.

687 **Complexity of terminal differentiation programs provide genetic**
688 **robustness and enhancer selectivity**

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

693 Compensatory actions among TFs have been mostly described for paralogous
694 genes, however, we find non-paralogous TFs act redundantly in the regulation
695 of terminal gene expression. A few other examples of non-paralogous TF
696 compensation are also found in embryogenesis, larval development or muscle
697 specification (Baugh et al., 2005; Kuntz et al., 2012; Walton et al., 2015). We
698 find compensatory effects are very unique: they not only depend on individual
699 TF pairs (i.e. not all double mutant combinations are synergistic), but also are
700 specific for each analyzed reporter (i.e. synergistic TFs show synergy for one
701 reporter but not for others) and in each case, are limited to particular
702 dopaminergic neuron types (i.e. synergies affect one or more dopaminergic
703 neuron types but not others). These specific synergies might be the reflection of
704 flexible protein-protein interactions among TFs that will be modulated by the
705 specific TF binding sites dispositions in each enhancer and are consistent with
706 the model of TF collectives for enhancer functions (Spitz and Furlong, 2012).
707 Synergistic effects among terminal selectors have been previously reported in
708 other neuronal types, such as HSN and NSM serotonergic neurons or
709 glutamatergic neurons (Lloret-Fernández et al., 2018; Serrano-Saiz et al., 2013;
710 F. Zhang et al., 2014), suggesting it might be a general feature of neuron-type
711 terminal regulatory programs.

In addition, we find TF complexity might also be important for enhancer selectivity. The sequence determinants that differentiate active regulatory regions from other non-coding regions of the genome are currently largely unknown. Interestingly, one of the best sequence predictors of active enhancers is the number of putative TF binding sites for different TFs found in a region (Grossman et al., 2017; Kheradpour et al., 2013; Tewhey et al., 2016). We find binding site clusters of the dopaminergic TFs are preferentially located in putative regulatory regions of dopaminergic expressed genes compared to other genes. Thus TF complexity of neuronal terminal regulatory programs might not only provide genetic robustness but also facilitate enhancer selection.

Parallel gene regulatory routines co-exist in the cell

The active transcriptome of a mature neuron can be divided in different gene categories depending on specificity of expression, ranging from cell-type specific genes to ubiquitous gene expression. While it is well established that terminal selectors have a direct role in neuron-type-specific gene expression, their role in other gene categories is less clear (Hobert, 2016). In our study we find the dopaminergic regulatory signature is predominantly enriched in dopamine effector genes, supporting a restricted role for terminal selectors in the regulation of neuron-type effector genes. Dopaminergic regulatory signature is also associated, to a less extent, to the distal regulatory regions of panneuronal genes, which is consistent with the known minor contributions for some terminal selectors in panneuronal gene expression (Stefanakis et al., 2015). In contrast, pancilia genes and ubiquitous genes are mostly devoid of dopaminergic regulatory signature, suggesting that other TFs activate these

gene categories. Transcriptional regulation of ubiquitous gene has been poorly studied. Broadly expressed TFs, such as Sp1, GABP/ETS and YY1 seem to play a role in ubiquitous gene expression in mammals (Bellora et al., 2007; Farré et al., 2007). Cilia structural genes are directly regulated by *daf-19*/RFX TF (Swoboda et al., 2000), and this role is conserved in vertebrates (Choksi et al., 2014). However, *daf-19* is expressed in all neurons, thus additional unknown TFs must control cilia gene expression. Intriguingly, some terminal selectors mutants display morphological cilia defects detected by failure to stain with the lipophilic dye DiI (Hobert, 2016). Although to date direct ciliome targets have not been identified for any terminal selector, it is conceivable that individual TF members of terminal differentiation program, but not the whole TF collective, might participate together with *daf-19* in the transcriptional regulation of some ciliome genes in specific neuron types. This could explain why some terminal selector mutants show cilia defects but pancilia genes are not enriched for the dopaminergic regulatory signature. In summary, and consistent with previous reports based on the analysis of a limited number of genes (Stefanakis et al., 2015), terminal selector collectives seem mostly devoted to neuron-type specific gene expression and to a less extend to panneuronal genes.

Neuron type evolution and deep homology

Mouse TF orthologs of the *C. elegans* dopaminergic terminal regulatory program are also required for olfactory bulb dopaminergic terminal specification, which is the most ancestral dopaminergic population of the mammalian brain (Agoston et al., 2014; Bovetti et al., 2013; Brill et al., 2008; Flames and Hobert, 2009; Qiu et al., 1995; Remesal et al., 2020). Here we find mouse TFs can, in

762 most cases, functionally substitute their worm orthologs suggesting both
763 neuronal populations share deep homology, which refers to the relationship
764 between cells in distant species that share their genetic regulatory programs
765 (Shubin et al., 2009). Similar conservation of terminal regulatory programs is
766 found for other neurons in *C. elegans* (Kratsios et al., 2012; Lloret-Fernández et
767 al., 2018).
768 Identification of deep homology is an emerging strategy in evolutionary biology
769 used to assign homologous neuronal types in distant species (Arendt, 2008;
770 Arendt et al., 2019, 2016). The study of neuronal regulatory programs in
771 different animal groups and the description of homologous relationships will
772 help us better understand the origin of neurons and the nervous system
773 evolution.

774 **METHODS**

775 ***C. elegans* Strains and Genetics**

776 *C. elegans* culture and genetics were performed as described (Brenner, 1974).
777 Strains used in this study are listed in Data Source 3.

778

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

780 To generate the complete TF RNAi library we used *C. elegans* TF list from
781 (Narasimhan et al., 2015). TF RNAi library was generated selecting RNAi
782 bacterial clones targeting TFs from Dr. Ahringer library (BioScience) (Kamath et
783 al., 2003). All clones were verified by Sanger sequencing, incorrect or missing
784 clones were selected from Dr. Vidal library (BioScience) (Rual et al., 2004) and
785 verified by Sanger sequencing. Remaining missing TF clones were generated
786 from genomic N2 DNA by PCR amplification of the target gene, subcloning into
787 L4440 plasmid (pPD129.36, Addgene) and transformed into *E. coli* HT115
788 (ED3) strain (from CGC). Sanger sequencing was used to verify all clones. The
789 complete list of TF RNAi feeding clones is listed in Data Source 1.

790

791 **RNAi feeding experiments**

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

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

Mutant strains and genotyping

Strains used in this study are listed in Data Source 3. Deletion alleles were genotyped by PCR. Point mutations were genotyped by sequencing. Genotyping primers are included in Data Source 4. Alleles *vab-3(ot346)*,*unc-62(e644) and lag-1(q385)* were determined by visual mutant phenotype. The mutation *unc-62(mu232)* was followed through its link with the fluorescent reporter *muIs35.*

Generation of *C. elegans* transgenic lines

Gene constructs for *cis*-regulatory analysis were generated by cloning into the pPD95.75 vector. For identification of the putative binding sites the following consensus sequences were used: GACA for UNC-62/MEIS HD (Campbell and Walthall, 2016); GRAGBA for VAB-3/PAIRED HD (Holst et al., 1997; Kim et al., 2008); TGACCW for UNC-55/COUP-TF (Badis et al., 2009) and CTAWWWWTAG for MEF-2/MADS (Boyle et al., 2014). Directed mutagenesis was performed by Quickchange II XL site-directed mutagenesis kit

(Stratagene). For microinjections the plasmid of interest (50 ng/µl) was injected together with *rol-6* comarker pRF4 [rol-6(su1006), 100 ng/µl]. *unc-55* transcriptional reporter was generated cloning 4kb upstream the ATG of the gene into ppD95.75 and injected at 50 ng/µl concentration together with *rol-6(*su1006) co-marker (100 ng/µl). *vab-3::gfp* fosmid injection DNA mix contained fosmid (40 ng/µl) and two comarkers, pRF4 (50 ng/µl) and pNF101 (*ttx-3prom::mcherry*) (50 ng/µl). For rescue experiments, commercially available cDNA of the TF candidates was obtained from Dharmacon Inc and Invitrogen. cDNAs corresponding to the entire coding sequence of *unc-62, vab-3, unc-55, mef-2, Meis2, Pax6, Nr2f1 and Mef2a* were amplified by PCR and cloned into *dat-1* and *bas-1* promoter reporter plasmids replacing GFP cDNA (See primers in **Data Source 4**). cDNA plasmid (25 ng/µl) was injected directly into the corresponding mutant background as complex arrays together with digested *E. coli* genomic DNA (50 ng/µl) and *unc-122::rfp* (25 ng/µl*)* fluorescent co-marker (**Data Source 3**).

Scoring and statistics

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

For mutant analysis lack of GFP signalling was considered OFF, if GFP expression was substantially weaker than WT, 'FAINT' category was included.

For *cis*-regulatory analysis two or three independent lines for each transgenic construct were analyzed. Mean expression value of three wild type lines is considered the reference value and mutated constructs are considered to show normal expression when GFP expression was 100-70% of average expression of the corresponding wild type construct ("+" phenotype). 70-30% reduction compared to wild type reference value was considered a partial phenotype "+/-". GFP expression below 30% was considered as great loss of expression ("-" phenotype) and if no GFP was detected (less than 5% of the neurons) was assigned as total loss of GFP ("--" phenotype).

For transcription factor expression analysis, reporter lines were crossed into *otIs181* (*dat-1::mcherry, ttx-3::mcherry*) (See **Data Source 3** for strain list). Images were taken with confocal TCS-SP8 Leica microscope and processed with ImageJ 1.50i software.

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

Bioinformatic Analysis

Unless otherwise indicated, all the analyses were performed using the software R (Z. Zhang et al., 2014) and packages from Bioconductor (Huber et al., 2015). Single-cell (sc) RNA-seq data from (Cao et al., 2017) was downloaded from the author's website (http://atlas.gs.washington.edu/worm-rna/). These data

correspond to nearly 50,000 cells coming from *Caenorhabditis elegans* at the L2 larval stage. Cells with failed QC, doublets and unclassified cells were filtered and excluded from subsequent analysis. Differential expression analysis between dopaminergic neurons cluster and clusters for the other ciliated sensory neurons was performed using Monocle (Qiu et al., 2017b, 2017a; Trapnell et al., 2014) and results were filtered by q-value (≤ 0.05) in order to get a list of differentially expressed genes in dopaminergic neurons. For gene enrichment of neuronal clusters corresponding to non-dopaminergic neurons (RIA, ASE, Touch receptor neurons, GABAergic neurons and ALN/PLN/SDQ neurons) we followed a similar strategy and performed differential expression tests between each neuronal cluster and all cells from the dataset annotated as neurons (**Data Source 2**). Specificity of these six gene sets was checked by the enrichment of its anatomical association in *C. elegans* using the web tool WormEnrichr (https://amp.pharm.mssm.edu/WormEnrichr/) (Chen et al., 2013). Gene lists for neuronal, non-neuronal, ubiquitous, panneuronal and panciliated categories were inferred from a more comprehensive sc-RNA-seq dataset (Packer et al., 2019). We retrieved gene expression data (log2 transcripts per million) from all the genes that were expressed in at least one annotated terminal cell bin, getting a final matrix of 15,813 genes x 409 terminal cell bins. Following authors' original approach, genes were ordered by hierarchical clustering and cell bins were ordered by tissues; resulting in differential gene clusters which marked sites of predominant expression [Figure S32 from (Packer et al., 2019)]. Using these data, we manually curated 21 gene lists, which were classified into five categories: non-neuronal tissue-specific, ubiquitous, panneuronal, panciliated, and neuron-type specific expression (**Data**

899 **Source 2**). For genes located in operons, only the gene located at the 5' end of
900 the cluster, and thus subdued to *cis*-regulation, were considered for
901 dopaminergic regulatory signature analysis. For hybrid operons, additional
902 promoters were also included (Blumenthal, Davis & Garrido-Lecca, 2015). For
903 genomic search of dopaminergic regulatory signature similar exclusion of
904 downstream genes from operon was performed, removing a total of 2,083
905 genes from the analysis. The final curated gene lists used in this study are
906 listed in **Data Source 2**. .
907 For *C. elegans* regulatory signature analysis, we downloaded PWMs from Cis-
908 BP version 1.02 (Weirauch et al., 2014) corresponding to the TF binding sites of
909 the seven transcription factors that compose the dopaminergic regulatory
910 signature in *C. elegans*. If the exact match for *C. elegans* was not available, we
911 selected the PWM from the *M. musculus* or *H. sapiens* orthologous TF
912 (COUPTF, ref. M1457; HD, ref. M5340; ETS, ref. M0709; MADS, ref. M6342;
913 MEIS, ref. M6048; PAIRED, ref. M1500; PBX, ref. M1898), plus an additional
914 hybrid PAIRED HD site (represented as HD*, ref. M6189). Following published
915 methodology (Lloret-Fernández et al., 2018) we downloaded upstream and
916 intronic gene regions of protein-coding genes from WormBase version 262 and
917 then classified genes using the gene lists mentioned above. Upstream regions
918 were trimmed to a maximum of 10 kb. PWMs were aligned to genomic
919 sequences and we retrieved matches with a minimum score of 70%. To
920 increase specificity, we removed all matches that did not bear an exact
921 consensus sequence for the corresponding TF family (consensus sequence for
922 COUPTF: TGACC, HD: TAATT, ETS: VMGGAWR, MADS: WWWDTAG, MEIS:
923 DTGTCD, PAIRED: GGARSA, HD*: HTAATTR, PBX: GATNNAT). Sliding

window search with a maximum length of 700 bp was performed to find regions that included at least one match for 6 or more of the 8 different TF binding motifs, allowing flexible motif composition. Resulting windows were classified according to the number of different motifs that they bore (6 or more, 7 or more or 8). To assess signature enrichment in the set of dopaminergic expressed genes 10,000 sets of 86 genes random genes were built considering that 1) they were not differentially expressed in dopaminergic neurons, 2) at least one ortholog had been described in other *Caenorhabditis* species (*C. brenneri, C. briggsae, C. japonica* or *C. remanei*), and 3) their upstream and intronic regions were similar in length, on average, to those of the 86 dopamine-expressed genes (Mann-Whitney U test, p-value > 0.05). For each one of the non-dopaminergic neuronal groups (RIA, ASE, Touch receptor neurons, GABAergic neurons and ALN/PLN/SDQ neurons), similar sets of random genes were built. We considered the enrichment in signature to be significant when the percentile of the neuronal group in regard to the internal random control was above 95. Differences between dopaminergic expressed genes and other neuronal groups were assessed by Brunner-Munzel test performed with R package brunnermunzel (Toshiaki, 2019), *: p-value < 0.05, **: p-value < 0.01, ***: p-value < 0.001).

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

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.

Acknowledgments

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

963 **REFERENCES**
964
965 Agoston Z, Heine P, Brill MS, Grebbin BM, Hau AC, Kallenborn-Gerhardt W,
966 Schramm J, Götz M, Schulte D. 2014. Meis2 is a Pax6 co-factor in
967 neurogenesis and dopaminergic periglomerular fate specification in the
968 adult olfactory bulb. *Dev*. doi:10.1242/dev.097295
969 Andrews MG, Kong J, Novitch BG, Butler SJ. 2019. New perspectives on the
970 mechanisms establishing the dorsal-ventral axis of the spinal cordCurrent
971 Topics in Developmental Biology. doi:10.1016/bs.ctdb.2018.12.010
972 Angerer LM, Yaguchi S, Angerer RC, Burke RD. 2011. The evolution of nervous
973 system patterning: Insights from sea urchin development. *Development*.
974 doi:10.1242/dev.058172
975 Arlotta P, Molyneaux BJ, Chen J, Inoue J, Kominami R, MacKlis JD. 2005.
976 Neuronal subtype-specific genes that control corticospinal motor neuron
977 development in vivo. *Neuron* **45**:207–221.
978 doi:10.1016/j.neuron.2004.12.036
979 Babonis LS, Martindale MQ. 2017. PaxA, but not PaxC, is required for
980 cnidocyte development in the sea anemone Nematostella vectensis.
981 *Evodevo* **8**:1–20. doi:10.1186/s13227-017-0077-7
982 Badis G, Berger MF, Philippakis AA, Talukder S, Gehrke AR, Jaeger SA, Chan
983 ET, Metzler G, Vedenko A, Chen X, Kuznetsov H, Wang CF, Coburn D,
984 Newburger DE, Morris Q, Hughes TR, Bulyk ML. 2009. Diversity and
985 complexity in DNA recognition by transcription factors. *Science (80-)*.
986 doi:10.1126/science.1162327
987 Baugh LR, Wen JC, Hill AA, Slonim DK, Brown EL, Hunter CP. 2005. Synthetic
988 lethal analysis of Caenorhabditis elegans posterior embryonic patterning
989 genes identifies conserved genetic interactions. *Genome Biol* **6**.
990 doi:10.1186/gb-2005-6-5-r45
991 Bellora N, Farré D, Mar MM. 2007. Positional bias of general and tissue-specific
992 regulatory motifs in mouse gene promoters. *BMC Genomics* **8**:1–13.
993 doi:10.1186/1471-2164-8-459
994 Bertrand N, Castro DS, Guillemot F. 2002. Proneural genes and the
995 specification of neural cell types. *Nat Rev Neurosci*. doi:10.1038/nrn874
996 Blanchi B, Kelly LM, Viemari JC, Lafon I, Burnet H, Bévengut M, Tillmanns S,
997 Daniel L, Graf T, Hilaire G, Sieweke MH. 2003. MafB deficiency causes
998 defective respiratory rhythmogenesis and fatal central apnea at birth. *Nat*
999 *Neurosci* **6**:1091–1099. doi:10.1038/nn1129
1000 Borello U, Pierani A. 2010. Patterning the cerebral cortex: Traveling with
1001 morphogens. *Curr Opin Genet Dev*. doi:10.1016/j.gde.2010.05.003
1002 Bovetti S, Bonzano S, Garzotto D, Giannelli SG, Iannielli A, Armentano M,
1003 Studer M, De Marchis S. 2013. COUP-TFI controls activity-dependent
1004 tyrosine hydroxylase expression in adult dopaminergic olfactory bulb
1005 interneurons. *Dev*. doi:10.1242/dev.089961
1006 Boyle AP, Araya CL, Brdlik C, Cayting P, Cheng C, Cheng Y, Gardner K, Hillier
1007 LW, Janette J, Jiang L, Kasper D, Kawli T, Kheradpour P, Kundaje A, Li JJ,
1008 Ma L, Niu W, Rehm EJ, Rozowsky J, Slattery M, Spokony R, Terrell R,
1009 Vafeados D, Wang D, Weisdepp P, Wu YC, Xie D, Yan KK, Feingold EA,
1010 Good PJ, Pazin MJ, Huang H, Bickel PJ, Brenner SE, Reinke V, Waterston
1011 RH, Gerstein M, White KP, Kellis M, Snyder M. 2014. Comparative analysis
1012 of regulatory information and circuits across distant species. *Nature*.

doi:10.1038/nature13668
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*. doi:10.1242/dev.168153
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*. doi:10.1523/JNEUROSCI.0700-08.2008
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*. doi:10.1016/S0092-8674(00)80853-3
Campbell RF, Walthall WW. 2016. Meis/UNC-62 isoform dependent regulation of CoupTF-II/UNC-55 and GABAergic motor neuron subtype differentiation. *Dev Biol*. doi:10.1016/j.ydbio.2016.09.009
Cao J, Packer JS, Ramani V, Cusanovich DA, Huynh C, Daza R, Qiu X, Lee C, Furlan SN, Steemers FJ, Adey A, Waterston RH, Trapnell C, Shendure J. 2017. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science (80-)*. doi:10.1126/science.aam8940
Cave JW, Akiba Y, Banerjee K, Bhosle S, Berlin R, Baker H. 2010. Differential regulation of dopaminergic gene expression by Er81. *J Neurosci*. doi:10.1523/JNEUROSCI.0419-10.2010
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*. doi:10.1186/1471-2105-14-128
Choksi SP, Lauter G, Swoboda P, Roy S. 2014. Switching on cilia: Transcriptional networks regulating ciliogenesis. *Dev*. doi:10.1242/dev.074666
Doitsidou M, Flames N, Lee AC, Boyanov A, Hobert O. 2008. Automated screening for mutants affecting dopaminergic-neuron specification in C. elegans. *Nat Methods*. doi:10.1038/nmeth.1250
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*. doi:10.1101/gad.217224.113
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*. doi:10.1186/1471-2105-10-48
El-Brolosy MA, Kontarakis Z, Rossi A, Kuenne C, Günther S, Fukuda N, Kikhi K, Boezio GLM, Takacs CM, Lai SL, Fukuda R, Gerri C, Giraldez AJ, Stainier DYR. 2019. Genetic compensation triggered by mutant mRNA degradation. *Nature*. doi:10.1038/s41586-019-1064-z
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. doi:10.1186/gb-2007-8-7-r140
Flames N, Hobert O. 2011. Transcriptional Control of the Terminal Fate of Monoaminergic Neurons. *Annu Rev Neurosci*. doi:10.1146/annurev-neuro-061010-113824

1063 Flames N, Hobert O. 2009. Gene regulatory logic of dopamine neuron
1064 differentiation. *Nature*. doi:10.1038/nature07929
1065 Grossman SR, Zhang X, Wang L, Engreitz J, Melnikov A, Rogov P, Tewhey R,
1066 Isakova A, Deplancke B, Bernstein BE, Mikkelsen TS, Lander ES. 2017.
1067 Systematic dissection of genomic features determining transcription factor
1068 binding and enhancer function. *Proc Natl Acad Sci U S A* **114**:E1291–
1069 E1300. doi:10.1073/pnas.1621150114
1070 Guillemot F, Hassan BA. 2017. Beyond proneural: emerging functions and
1071 regulations of proneural proteins. *Curr Opin Neurobiol*.
1072 doi:10.1016/j.conb.2016.11.011
1073 Haider NB, Jacobson SG, Cideciyan A V., Swiderski R, Streb LM, Searby C,
1074 Beck G, Hockey R, Hanna DB, Gorman S, Duhl D, Carmi R, Bennett J,
1075 Weleber RG, Fishman GA, Wright AF, Stone EM, Sheffield VC. 2000.
1076 Mutation of a nuclear receptor gene, NR2E3, causes enhanced S cone
1077 syndrome, a disorder of retinal cell fate. *Nat Genet*. doi:10.1038/72777
1078 Hobert O. 2016. A map of terminal regulators of neuronal identity in
1079 Caenorhabditis elegans. *Wiley Interdiscip Rev Dev Biol*.
1080 doi:10.1002/wdev.233
1081 Hobert O. 2008. Regulatory logic of neuronal diversity: Terminal selector genes
1082 and selector motifs. *Proc Natl Acad Sci U S A*.
1083 doi:10.1073/pnas.0806070105
1084 Hobert O, Kratsios P. 2019. Neuronal identity control by terminal selectors in
1085 worms, flies, and chordates. *Curr Opin Neurobiol*.
1086 doi:10.1016/j.conb.2018.12.006
1087 Holst BD, Wang Y, Jones FS, Edelman GM. 1997. A binding site for Pax
1088 proteins regulates expression of the gene for the neural cell adhesion
1089 molecule in the embryonic spinal cord. *Proc Natl Acad Sci U S A*.
1090 doi:10.1073/pnas.94.4.1465
1091 Huang X, Tian E, Xu Y, Zhang H. 2009. The C. elegans engrailed homolog ceh-
1092 16 regulates the self-renewal expansion division of stem cell-like seam
1093 cells. *Dev Biol*. doi:10.1016/j.ydbio.2009.07.005
1094 Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo
1095 HC, Davis S, Gatto L, Girke T, Gottardo R, Hahne F, Hansen KD, Irizarry
1096 RA, Lawrence M, Love MI, MaCdonald J, Obenchain V, Oleš AK, Pagès H,
1097 Reyes A, Shannon P, Smyth GK, Tenenbaum D, Waldron L, Morgan M.
1098 2015. Orchestrating high-throughput genomic analysis with Bioconductor.
1099 *Nat Methods*. doi:10.1038/nmeth.3252
1100 Hynes M, Stone DM, Dowd M, Pitts-Meek S, Goddard A, Gurney A, Rosenthal
1101 A. 1997. Control of cell pattern in the neural tube by the zinc finger
1102 transcription factor and oncogene Gli-1. *Neuron* **19**:15–26.
1103 doi:10.1016/S0896-6273(00)80344-X
1104 Jiang Y, Horner V, Liu J. 2005. The HMX homeodomain protein MLS-2
1105 regulates cleavage orientation, cell proliferation and cell proliferation and
1106 cell fate specification in the C. elegans postembryonic mesoderm.
1107 *Development*. doi:10.1242/dev.01967
1108 Jiang Y, Shi H, Liu J. 2009. Two Hox cofactors, the Meis/Hth homolog UNC-62
1109 and the Pbx/Exd homolog CEH-20, function together during C. elegans
1110 postembryonic mesodermal development. *Dev Biol*.
1111 doi:10.1016/j.ydbio.2009.07.034
1112 Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R, Gotta M, Kanapin A, Le

Bot N, Moreno S, Sohrmann M, Welchman DP, Zipperien P, Ahringer J. 2003. Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. *Nature*. doi:10.1038/nature01278

Kamath RS, Martinez-Campos M, Zipperlen P, Fraser AG, Ahringer J. 2001. Effectiveness of specific RNA-mediated interference through ingested double-stranded RNA in Caenorhabditis elegans. *Genome Biol*. doi:10.1186/gb-2000-2-1-research0002

Kheradpour P, Ernst J, Melnikov A, Rogov P, Wang L, Zhang X, Alston J, Mikkelsen TS, Kellis M. 2013. Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. *Genome Res* **23**:800–811. doi:10.1101/gr.144899.112

Kim DS, Matsuda T, Cepko CL. 2008. A core paired-type and POU homeodomain-containing transcription factor program drives retinal bipolar cell gene expression. *J Neurosci*. doi:10.1523/JNEUROSCI.0397-08.2008

Knoepfler PS, Calvo KR, Chen H, Antonarakis SE, Kamps MP. 1997. Meis1 and pKnox1 bind DNA cooperatively with Pbx1 utilizing an interaction surface disrupted in oncoprotein E2a-Pbx1. *Proc Natl Acad Sci U S A*. doi:10.1073/pnas.94.26.14553

Kratsios P, Stolfi A, Levine M, Hobert O. 2012. Coordinated regulation of cholinergic motor neuron traits through a conserved terminal selector gene. *Nat Neurosci*. doi:10.1038/nn.2989

Kuntz SG, Williams BA, Sternberg PW, Wold BJ. 2012. Transcription factor redundancy and tissue-specific regulation: Evidence from functional and physical network connectivity. *Genome Res* **22**:1907–1919. doi:10.1101/gr.133306.111

Layden MJ, Boekhout M, Martindale MQ. 2012. Nematostella vectensis achaete-scute homolog NvashA regulates embryonic ectodermal neurogenesis and represents an ancient component of the metazoan neural specification pathway. *Development* **139**:1013–1022. doi:10.1242/dev.073221

Liachko NF, Saxton AD, McMillan PJ, Strovas TJ, Dirk Keene C, Bird TD, Kraemer BC. 2019. Genome wide analysis reveals heparan sulfate epimerase modulates TDP-43 proteinopathy. *PLoS Genet*. doi:10.1371/journal.pgen.1008526

Liu A, Niswander LA. 2005. Bone morphogenetic protein signalling and vertebrate nervous system development. *Nat Rev Neurosci*. doi:10.1038/nrn1805

Liu H, Strauss TJ, Potts MB, Cameron S. 2006. Direct regulation of egl-1 of programmed cell death by the Hox protein MAB-5 and CEH-20, a C. elegans homolog of Pbx1. *Development*. doi:10.1242/dev.02234

Liu Y, Yu C, Daley TP, Wang F, Cao WS, Bhate S, Lin X, Still C, Liu H, Zhao D, Wang H, Xie XS, Ding S, Wong WH, Wernig M, Qi LS. 2018. CRISPR Activation Screens Systematically Identify Factors that Drive Neuronal Fate and Reprogramming. *Cell Stem Cell* **23**:758-771.e8. doi: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*. doi:10.7554/eLife.32785

Maeda R, Ishimura A, Mood K, Park EK, Buchberg AM, Daar IO. 2002. Xpbx1b

and Xmeis1b play a collaborative role in hindbrain and neural crest gene expression in Xenopus embryos. *Proc Natl Acad Sci U S A*. doi:10.1073/pnas.082654899

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*. doi:10.1186/gb-2001-2-8-research0029

Mall M, Kareta MS, Chanda S, Ahlenius H, Perotti N, Zhou B, Grieder SD, Ge X, Drake S, Euong Ang C, Walker BM, Vierbuchen T, Fuentes DR, Brennecke P, Nitta KR, Jolma A, Steinmetz LM, Taipale J, Südhof TC, Wernig M. 2017. Myt1l safeguards neuronal identity by actively repressing many non-neuronal fates. *Nature*. doi:10.1038/nature21722

Masserdotti G, Gascón S, Götz M. 2016. Direct neuronal reprogramming: Learning from and for development. *Dev*. doi:10.1242/dev.092163

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*. doi:10.1038/ng774

Narasimhan K, Lambert SA, Yang AWH, Riddell J, Mnaimneh S, Zheng H, Albu M, Najafabadi HS, Reece-Hoyes JS, Fuxman Bass JI, Walhout AJM, Weirauch MT, Hughes TR. 2015. Mapping and analysis of Caenorhabditis elegans transcription factor sequence specificities. *Elife*. doi:10.7554/eLife.06967

Noman A, Liu Z, Aqeel M, Zainab M, Khan MI, Hussain A, Ashraf MF, Li X, Weng Y, He S. 2017. Basic leucine zipper domain transcription factors: the vanguards in plant immunity. *Biotechnol Lett* **39**:1779–1791. doi:10.1007/s10529-017-2431-1

Nuez I, Félix MA. 2012. Evolution of susceptibility to ingested double-stranded rnas in Caenorhabditis nematodes. *PLoS One*. doi:10.1371/journal.pone.0029811

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*. doi:10.1038/s41598-017-04375-4

Olsson-Carter K, Slack FJ. 2010. A developmental timing switch promotes axon outgrowth independent of known guidance receptors. *PLoS Genet*. doi:10.1371/journal.pgen.1001054

Packer JS, Zhu Q, Huynh C, Sivaramakrishnan P, Preston E, Dueck H, Stefanik D, Tan K, Trapnell C, Kim J, Waterston RH, Murray JI. 2019. A lineage-resolved molecular atlas of C. elegans embryogenesis at single-cell resolution . *Science (80-)* **365**. doi:10.1126/science.aax1971

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*. doi:10.1371/journal.pgen.1002109

Portman DS, Emmons SW. 2000. The basic helix-loop-helix transcription factors LIN-32 and HLH-2 function together in multiple steps of a C. elegans neuronal sublineage. *Development*.

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*. doi:10.1016/j.ydbio.2009.02.022

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*. doi:10.1101/gad.9.20.2523

Qiu X, Hill A, Packer J, Lin D, Ma YA, Trapnell C. 2017a. Single-cell mRNA quantification and differential analysis with Census. *Nat Methods*. doi:10.1038/nmeth.4150

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*. doi:10.1038/nmeth.4402

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**. doi:10.1186/gb-2005-6-13-r110

Remesal L, Roger-Banyat I, Chirivella L, Maicas M, Brocal-Ruiz R, Pérez-Villalva A, Cucarella C, Casado M, Flames N. n.d. 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. doi:10.1242/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. doi:10.1002/wdev.257

Rieckhof GE, Casares F, Ryoo HD, Abu-Shaar M, Mann RS. 1997. Nuclear translocation of extradenticle requires homothorax, which encodes an extradenticle-related homeodomain protein. *Cell*. doi:10.1016/S0092-8674(00)80400-6

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*. doi:10.1523/JNEUROSCI.1148-04.2004

Rual JF, Ceron J, Koreth J, Hao T, Nicot AS, Hirozane-Kishikawa T, Vandenhaute J, Orkin SH, Hill DE, van den Heuvel S, Vidal M. 2004. Toward improving Caenorhabditis elegans phenome mapping with an ORFeome-based RNAi library. *Genome Res*. doi:10.1101/gr.2505604

Sengupta P, Colbert HA, Bargmann CI. 1994. The C. elegans gene odr-7 encodes an olfactory-specific member of the nuclear receptor superfamily. *Cell*. doi:10.1016/0092-8674(94)90028-0

Serobyan V, Kontarakis Z, El-Brolosy MA, Welker JM, Tolstenkov O, Saadeldein AM, Retzer N, Gottschalk A, Wehman AM, Stainier DYR. 2020. Transcriptional adaptation in caenorhabditis elegans. *Elife*. doi:10.7554/eLife.50014

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*. doi:10.1016/j.cell.2013.09.052

Shan G, Walthall WW. 2008. Copulation in C. elegans males requires a nuclear hormone receptor. *Dev Biol* **322**:11–20. doi:10.1016/j.ydbio.2008.06.034

Shimizu Takeshi, Nakazawa M, Kani S, Bae YK, Shimizu Takashi, 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. doi:10.1242/dev.047167

Shirasaki R, Pfaff SL. 2002. Transcriptional Codes and the Control of Neuronal Identity. *Annu Rev Neurosci*. doi:10.1146/annurev.neuro.25.112701.142916

Shubin N, Tabin C, Carroll S. 2009. Deep homology and the origins of evolutionary novelty. *Nature*. doi:10.1038/nature07891

Sieburth D, Ch'ng Q, Dybbs M, Tavazoie M, Kennedy S, Wang D, Dupuy D, Rual JF, Hill DE, Vidal M, Ruvkun G, Kaplan JM. 2005. Systematic analysis of genes required for synapse structure and function. *Nature* **436**:510–516. doi:10.1038/nature03809

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*. doi:10.1371/journal.pbio.0000012

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. doi:10.1534/genetics.108.088948

Spitz F, Furlong EEM. 2012. Transcription factors: From enhancer binding to developmental control. *Nat Rev Genet*. doi:10.1038/nrg3207

Stefanakis N, Carrera I, Hobert O. 2015. Regulatory Logic of Pan-Neuronal Gene Expression in C. elegans. *Neuron*. doi:10.1016/j.neuron.2015.07.031

Stegmaier P, Kel AE, Wingender E. 2004. Systematic DNA-binding domain classification of transcription factors. *Genome Inform*. doi:10.11234/gi1990.15.2_276

Swoboda P, Adler HT, Thomas JH. 2000. The RFX-type transcription factor DAF-19 regulates sensory neuron cilium formation in C. Elegans. *Mol Cell*. doi:10.1016/S1097-2765(00)80436-0

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*. doi:10.1016/j.mce.2010.04.021

Tewhey R, Kotliar D, Park DS, Liu B, Winnicki S, Reilly SK, Andersen KG, Mikkelsen TS, Lander ES, Schaffner SF, Sabeti PC. 2016. Direct identification of hundreds of expression-modulating variants using a multiplexed reporter assay. *Cell* **165**:1519–1529. doi: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. doi:10.1016/j.celrep.2020.03.031

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*. doi:10.1038/nbt.2859

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*.

Vlachakis N, Choe SK, Sagerström CG. 2001. Meis3 synergizes with Pbx4 and Hoxb1b in promoting hindbrain fates in the zebrafish. *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. doi:10.1016/j.neuron.2006.01.018

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. doi:10.1371/journal.pgen.1005003

Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, Najafabadi HS, Lambert SA, Mann I, Cook K, Zheng H, Goity A, van Bakel H, Lozano JC, Galli M, Lewsey MG, Huang E, Mukherjee T, Chen X, Reece-Hoyes JS, Govindarajan S, Shaulsky G, Walhout AJM, Bouget FY, Ratsch G, Larrondo LF, Ecker JR, Hughes TR. 2014. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell*. doi:10.1016/j.cell.2014.08.009

Wende H, Lechner SG, Cheret C, Bourane S, Kolanczyk ME, Pattyn A, Reuter K, Munier FL, Carroll P, Lewin GR, Birchmeier C. 2012. The transcription factor c-Maf controls touch receptor development and function. *Science (80-)* **335**:1373–1376. doi:10.1126/science.1214314

Zeng H, Sanes JR. 2017. Neuronal cell-type classification: Challenges, opportunities and the path forward. *Nat Rev Neurosci*. doi:10.1038/nrn.2017.85

Zetterström RH, Solomin L, Jansson L, Hoffer BJ, Olson L, Perlmann T. 1997. Dopamine neuron agenesis in Nurr1-deficient mice. *Science (80-)*. doi:10.1126/science.276.5310.248

Zhang F, Bhattacharya A, Nelson JC, Abe N, Gordon P, Lloret-Fernandez C, Maicas M, Flames N, Mann RS, Colón-Ramos DA, Hobert O. 2014. The LIM and POU homeobox genes ttx-3 and unc-86 act as terminal selectors in distinct cholinergic and serotonergic neuron types. *Dev*. doi:10.1242/dev.099721

Zhang Z, et al. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/. *Nat Genet*. doi:10.1105/tpc.110.079111

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*. doi:10.1371/journal.pgen.1004017

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*. doi:10.1016/j.ydbio.2005.08.013