A combinatorial regulatory signature controls terminal differentiation of the dopaminergic nervous system in *C. elegans*

Maria Doitsidou, Nuria Flames, Irini Topalidou, et al.

Genes Dev. 2013 27: 1391-1405
Access the most recent version at doi:10.1101/gad.217224.113

A combinatorial regulatory signature controls terminal differentiation of the dopaminergic nervous system in *C. elegans*

Maria Doitsidou,[1,2,3,4,8,9] Nuria Flames,[1,2,5,8,9] Irini Topalidou,[6,7] Namiko Abe,[1] Terry Felton,[1,2] Laura Remesal,[1,2,5] Tatiana Popovitchenko,[1,2,3,4] Richard Mann,[1] Martin Chalfie,[6] and Oliver Hobert[1,2]

[1]Department of Biochemistry and Molecular Biophysics, [2]Howard Hughes Medical Institute, Columbia University Medical Center, New York, New York 10032, USA; [3]Norwegian Center for Movement Disorders, Stavanger University Hospital, Stavanger N-4068, Norway; [4]Center for Organelle Research, University of Stavanger, Stavanger N-4036, Norway; [5]Instituto de Biomedicina de Valencia-Consejo Superior de Investigaciones Científicas (IBV-CSIC), 46010 Valencia, Spain; [6]Department of Biological Sciences, Columbia University, New York, New York 10027, USA; [7]Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA

Terminal differentiation programs in the nervous system are encoded by *cis*-regulatory elements that control the expression of terminal features of individual neuron types. We decoded the regulatory information that controls the expression of five enzymes and transporters that define the terminal identity of all eight dopaminergic neurons in the nervous system of the *Caenorhabditis elegans* hermaphrodite. We show that the tightly coordinated, robust expression of these dopaminergic enzymes and transporters ("dopamine pathway") is ensured through a combinatorial *cis*-regulatory signature that is shared by all dopamine pathway genes. This signature is composed of an Ets domain-binding site, recognized by the previously described AST-1 Ets domain factor, and two distinct types of homeodomain-binding sites that act in a partially redundant manner. Through genetic screens, we identified the sole *C. elegans* Distalless/Dlx ortholog, *ceh-43,* as a factor that acts through one of the homeodomain sites to control both induction and maintenance of terminal dopaminergic fate. The second type of homeodomain site is a Pbx-type site, which is recognized in a partially redundant and neuron subtype-specific manner by two Pbx factors, *ceh-20* and *ceh-40*, revealing novel roles of Pbx factors in the context of terminal neuron differentiation. Taken together, we revealed a specific regulatory signature and cognate, terminal selector-type transcription factors that define the entire dopaminergic nervous system of an animal. Dopaminergic neurons in the mouse olfactory bulb express a similar combinatorial transcription factor collective of Ets/Dlx/Pbx factors, suggesting deep phylogenetic conservation of dopaminergic regulatory programs.

[*Keywords*: *cis*-regulatory motif; differentiation; dopamine; elegans; homeodomain; neuron]

Supplemental material is available for this article.

Received May 6, 2013; revised version accepted May 16, 2013.

The underlying basis of the functional and anatomical diversity of cell types in the nervous system is the differential expression of neuron-type-specific gene batteries, which are composed of terminal differentiation genes whose products define the specific properties of a mature neuron throughout its lifetime (Hobert et al. 2010). These "terminal differentiation" or "effector" genes include, for example, neurotransmitter-synthesizing enzymes, neurotransmitter receptors, transporters, the many ion channels that tune the electric properties of a neuron, specific synaptic adhesion molecules, and many others. With this notion in mind, one approach to understand the generation of neuronal diversity is a "bottom-up" approach that focuses on studying the *cis*-regulatory principles by which the expression of terminal gene batteries is controlled. In spite of the conceptual promise of such an approach, the *cis*-regulatory architecture of neuronal gene batteries is generally poorly described.

The neurotransmitter dopamine controls a number of diverse behaviors across the animal kingdom (Iversen and Iversen 2007). Dopaminergic neurons are molecularly defined through the coordinated expression of five genes

[8]These authors contributed equally.
[9]Corresponding authors
E-mail maria.doitsidou@uis.no
E-mail nflames@ibv.csic.es
Article is online at http://www.genesdev.org/cgi/doi/10.1101/gad.217224.113.

("dopamine pathway genes") that encode proteins involved in the synthesis of dopamine (tyrosine hydroxylase [TH], GTP-cyclohydrolase [GTPCH], and aromatic amino acid decarboxylase [AAAD]), packaging of dopamine into synaptic vesicles (vesicular monoaminergic transporter [VMAT]), and reuptake of dopamine after synaptic release (dopamine transporter [DAT]) (Fig. 1A; Flames and Hobert 2011). Surprisingly little is known about the *cis*-regulatory elements and *trans*-acting factors that ensure coordinated expression of these specific genes. Understanding in detail how expression of these genes is regulated will provide insights into general regulatory principles of nervous system development, and such insights may also instruct our ability to generate this neuron type through reprogramming approaches in vitro or in vivo.

Dopaminergic neurons as well as dopamine pathway genes are phylogenetically conserved. The nematode *Caenorhabditis elegans* contains eight dopaminergic neurons in the hermaphroditic nervous system that fall into four distinct classes: the embryonically generated ADE, CEPD, and CEPV classes in the head and the post-embryonically generated PDE class in the midbody (Fig. 1A; Sulston et al. 1975; Flames and Hobert 2011). Using the amenability of *C. elegans* to transgenic reporter gene studies, we previously dissected the *cis*-regulatory control regions of the five dopamine pathway genes (*cat-2*/TH, *cat-4*/GTPCH, *bas-1*/AAAD, *cat-1*/VMAT, and *dat-1*/DAT) and identified a shared *cis*-regulatory motif—termed the dopaminergic (DA) motif—that is present in all five dopamine pathway genes (Flames and Hobert 2009). Through mutational analysis, we found that the DA motif is required for all five genes to be expressed in all eight dopaminergic neurons (Fig. 1B). We showed that this *cis*-regulatory motif is recognized by the Ets domain factor AST-1, a terminal selector-type transcription factor (Flames and Hobert 2009). Terminal selectors are transcription factors that define the terminal identity state of individual neuron types through direct regulation of many identity features of a specific neuron type (Hobert 2008, 2011). Genetic removal of *ast-1* results in a failure of all dopaminergic neurons to properly differentiate. A dopaminergic differentiation defect was

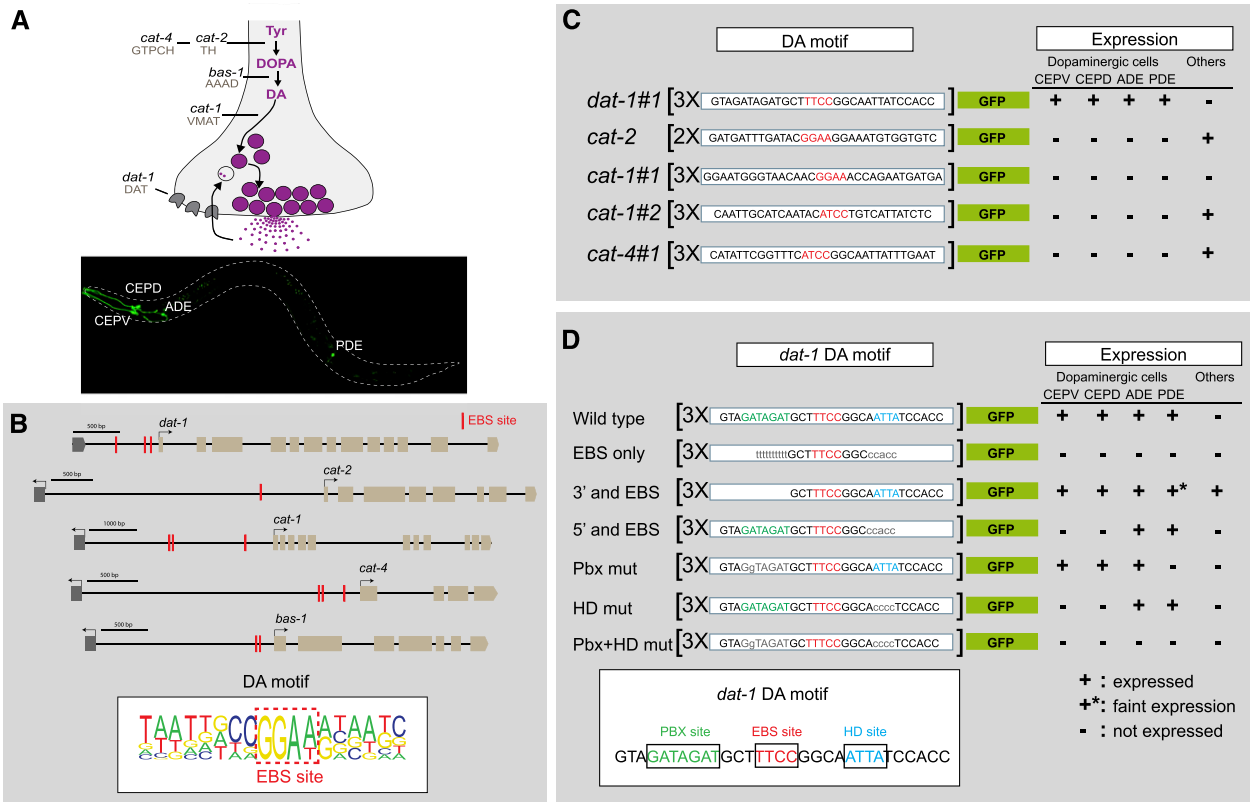

Figure 1. ETS-, Pbx-, and homeodomain-binding sites are required for *dat-1* expression. (*A*) Schematic of the dopamine synthesis pathway genes and a transgenic worm expressing *dat-1::gfp*, providing an overview of all dopaminergic neurons in *C. elegans*. (*B*) Summary of previous *cis*-regulatory analysis of dopamine pathway genes, illustrating with red bars the functionally required ETS domain-binding sites (EBSs) and sequence logo generated with all functional EBSs (Flames and Hobert 2009). (*C*) Analysis of the sufficiency of the EBS (labeled in red) from different dopamine pathway genes, flanked by >10 bp on either side, to drive expression in all dopaminergic neurons. Two (2×) or three (3×) copies of each sequence were cloned into the GFP reporter vector. For each construct, at least two independent lines were analyzed ($n > 30$ for each line). (+) More than 30% of the cells express GFP; (−) no GFP expression. (*D*) Mutational analysis of the predicted HD (labeled in blue) and Pbx (labeled in green) sites in the context of the trimerized *dat-1* DA motif. For each construct, two independent lines were analyzed. (+) More than 30% of the cells express GFP; (−) no GFP expression; (+*) faint expression.

also observed in olfactory bulb dopaminergic neurons of mice lacking the AST-1 homolog Etv1 (Flames and Hobert 2009). One key issue that remained unresolved by our previous studies is the question of specificity. While the Ets domain transcription factor AST-1 and mouse Etv1 are required to generate dopaminergic neurons, they are not sufficient to do so, since both genes are expressed in multiple other, nondopaminergic neuron types. Similar specificity issues apply to many other terminal selector-type transcription factors that are required to define the identity of specific neuron types in vertebrate and invertebrate nervous systems but are often expressed in many other cell types as well (Hobert 2011; Holmberg and Perlmann 2012).

Here we investigated this issue of specificity using a combination of *cis*-regulatory mutational analysis and genetic screening approaches. We found that *ast-1* indeed does not act in isolation but rather through a combinatorial *cis*-regulatory signature that is present in the *cis*-regulatory regions of all dopamine pathways genes. We identified three distinct types of *trans*-acting factors that recognize this *cis*-regulatory signature and show that these factors act together as a "transcription factor collective" (Junion et al. 2012) to ensure robust execution of the terminal differentiation program of dopaminergic neurons. Mouse homologs of this transcription factor collective are also expressed in a specific population of dopaminergic neurons in vertebrates, suggesting that our findings may also apply to mammals.

Results

A combinational cis*-regulatory signature required for gene expression in dopaminergic neurons*

We previously showed that a construct containing in triplicate a 31-base-pair (bp) element from the *dat-1* promoter is sufficient to drive expression of a reporter gene in all eight dopaminergic neurons of the *C. elegans* hermaphrodite (Fig. 1C; Flames and Hobert 2009). This element contains an Ets domain-binding site (EBS), which we showed through deletion analysis to be essential for expression of *dat-1* in dopaminergic neurons. We found functional EBSs in all of the dopamine pathway genes and therefore called the EBS the "DA motif" (Fig. 1B; Flames and Hobert 2009). Unexpectedly, when we tested the DA motifs from other dopamine pathway genes in a manner similar to our testing of the DA motif from *dat-1* (i.e., EBS plus 13–14 bp of flanking sequences), we found that, unlike in the *dat-1* case, DA motifs from none of the other four dopamine pathway genes were sufficient to drive expression in dopaminergic neurons (Fig. 1C) even though each of the DA motifs is required to drive expression in all dopaminergic neurons (Flames and Hobert 2009).

We therefore examined the 31-bp DA motif from the *dat-1* promoter in more detail. Using MatInspector software analysis, we noted the presence of a predicted Pbx-type homeodomain (HD)-binding site ["GAT(N)$_{1-2}$GAT"] and a canonical HD-binding site ("TAAT") flanking the EBS. Mutating either site alone had partial effects on the expression of the reporter gene in dopaminergic neurons, while mutating both sites simultaneously—leaving, at the same time, the EBS intact—completely abolished expression of *dat-1* in all dopaminergic neuron types (Fig. 1D).

The DA motif-containing 30- to 32-bp elements from each of the other four dopamine pathway genes do not contain a combination of predicted Pbx- and canonical HD-binding sites, thereby providing a potential explanation for their insufficiency to drive expression in dopaminergic neurons. However, each of the minimal regions from all other dopamine pathway genes driving dopaminergic neuron expression contained a set of predicted Pbx- and HD-binding sites similar to those observed in the *dat-1* promoter. We systematically mutated these predicted Pbx and HD sites in four of the five dopamine pathway genes and found that they are required for dopaminergic neuron expression (Fig. 2A–D). Unlike the EBS, which is essential for expression of all dopamine pathway genes (Flames and Hobert 2009), the Pbx and HD sites act in a partially redundant manner, as detailed below.

dat-1/DAT The minimal *dat-1 cis*-regulatory element that drives expression in all dopaminergic neurons (400 bp) contains two predicted HD sites, one predicted Pbx site, and three functional EBSs (previously shown to be essential for expression) (Fig. 2A). Mutating a single HD site has no noticeable effect on *dat-1* reporter gene expression in transgenic animals (data not shown), and mutating both HD sites has partial effects, similar to the effect of the HD mutation in the *dat-1* DA motif construct (Figs. 1D, 2A; Supplemental Fig. 1). Mutating the one Pbx site has no effect (data not shown). However, mutating both HD sites and the one Pbx site together completely eliminates reporter gene expression in all dopaminergic neurons (even though the Ets sites are still present) (Fig. 2A; Supplemental Fig. 1), again in concordance with the effect seen in the *dat-1* DA motif construct analysis.

cat-2/TH The minimal *cat-2 cis*-regulatory element that drives expression in all dopaminergic neurons (153 bp) also contains two potential canonical HD sites, one potential Pbx site, and one functional EBS (previously shown to be essential) (Fig. 2B). The effects of the mutation of individual sites are similar to the *dat-1* case. Mutation of the first HD site results in partial effects, whereas mutation of the second HD site does not have any effect. The combined mutation of both sites did not significantly increase the loss of expression (Fig. 2B). Interestingly, the second HD site that shows no effect when mutated did not match the ATAAT consensus sequence found for the *dat-1* HD sites (Supplemental Fig. 1). Similarly, mutating the single Pbx site has a very mild effect. However, mutating the ATAAT HD site concomitantly with the Pbx site completely eliminates reporter gene expression in all dopaminergic neurons (even though the Ets site is still present) (Fig. 2B; Supplemental Fig. 1).

cat-4/GTPCH In addition to the functionally required EBS, the minimal *cat-4 cis*-regulatory element that drives

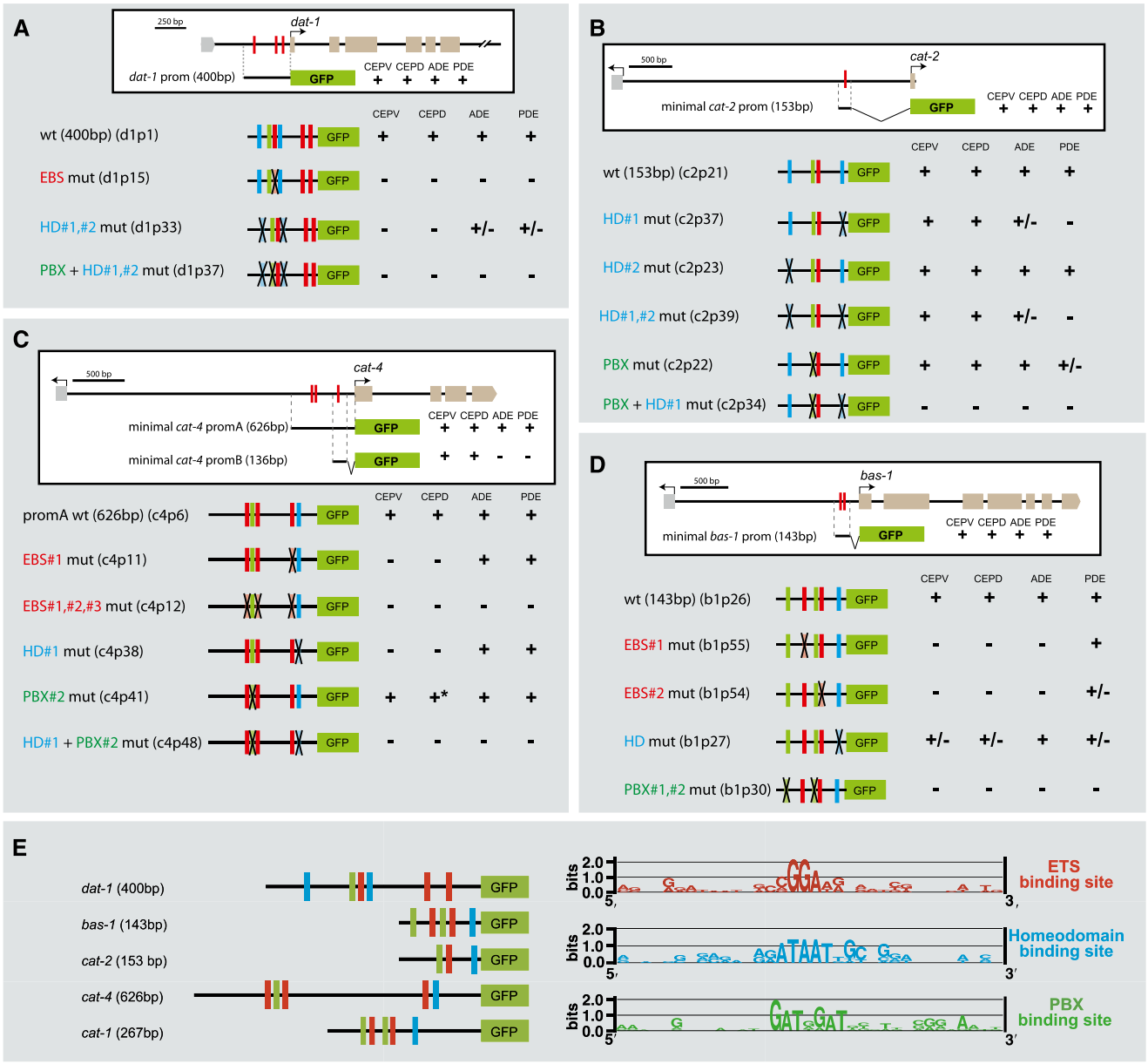

Figure 2. The *cis*-regulatory modules of the dopamine pathway genes. (*A–D*) Mutational analysis of the *cis*-regulatory modules of dopamine pathway genes. (Red bar) Functional EBS; (green bar) predicted Pbx site; (blue bar) predicted HD site. Mutation of the site is indicated by a black cross and by light shading of the respective bar. For each construct, at least two lines were analyzed ($n > 30$ for each line). Wild-type constructs set as 100% expression. (+) Twenty percent expression up or down the wild-type value; (+*) 20% up or down the wild-type value but faint expression; (+/−) >20% decrease compared with wild-type value; (−) 0%–15% of the cells expressing GFP in absolute numbers. The primary data for this analysis as well as the nature of each mutation are shown in Supplemental Figure 1. (*E*) Summary of the *cis*-regulatory signature present in the dopamine pathway genes and sequence logo of the functional EBS-, Pbx-, and HD-binding sites in each minimal *cis*-regulatory motif.

expression in all dopaminergic neurons (626 bp) has a substantial number of predicted HD- and Pbx-binding sites (Supplemental Fig. 1). Through analysis of a smaller *cat-4* promoter region (136 bp) that is active only in a subset of dopaminergic neurons and subsequent mutation of all HD sites in this small region, we identified one HD site that matches the ATAAT consensus and whose mutation resulted in expression defects (Supplemental Fig. 1). Similar defects where seen when the mutation in this HD site was introduced in the context of the larger *cat-4* element (which is active in all dopaminergic neurons) (Fig. 2C; Supplemental Fig. 1). If combined with a mutation in a single Pbx site, expression of the reporter is completely eliminated (Fig. 2C; Supplemental Fig. 1).

bas-1/AAAD The minimal *bas-1 cis*-regulatory element that drives expression in all dopaminergic neurons (143 bp) also contains one predicted HD site that matches the

ATAAT consensus, two predicted Pbx sites, and two functional EBSs (previously shown to be essential) (Fig. 2D). Mutating the single HD site has partial effects on *bas-1* reporter gene expression in transgenic animals. Mutating either Pbx site alone has no or just partial effects, but if both Pbx sites are mutated, reporter gene expression is eliminated in all dopaminergic neurons (even though the EBS is still present) (Fig. 2D; Supplemental Fig. 1).

In conclusion, members of the dopamine pathway are controlled by a common *cis*-regulatory signature composed of one or multiple essential EBSs and redundantly operating HD- and Pbx-binding sites (Fig. 2E). Importantly, the combination of these three motifs is sufficient to drive specific expression in all dopaminergic neurons (Fig. 1D). Notably, there is no defined "motif grammar" (Spitz and Furlong 2012); that is, no specific number, relative orientation, or spacing of the three motifs defines this *cis*-regulatory signature.

The sole C. elegans *Distalless/Dlx ortholog, ceh-43, controls dopaminergic neuron differentiation*

To identify the *trans*-acting factors that act through the ATAAT HD site, we turned to a collection of mutant animals that show abnormal expression of the *dat-1::gfp* dopaminergic fate marker and that we isolated previously through automated sorting of EMS-mutagenized animals (Doitsidou et al. 2008). One locus that we identified in this screen is the previously uncloned *dopy-2* gene, represented by four mutant alleles: *ot340*, *ot479*, *ot406*, and *ot345* (Doitsidou et al. 2008). We mapped this mutant to a small interval on chromosome III by SNP linkage analysis and three-factor mapping (see Fig. 4A, below). Through a combination of transformation rescue analysis (Table 1; see Fig. 4A, below), whole-genome sequencing (Supplemental Table 1), RNAi phenocopy (Table 1), and analysis of a genetic deletion in the locus (kindly provided by the *C. elegans* knockout facility at Tokyo Women's Medical School) (Fig. 3B,D), we found that *dopy-2* corresponds to the *ceh-43* locus (Fig. 4A,B). Mutations in the *ceh-43* locus had not previously been described. From here on, we refer to *dopy-2* as *ceh-43*.

ceh-43 encodes the sole *C. elegans* ortholog of the fly *Distalless* and vertebrate *Dlx* homeobox genes (Aspock and Burglin 2001). In vertebrates, *Dlx* genes have been implicated in neuronal patterning (Panganiban and Rubenstein 2002), but their roles in terminal neuron differentiation have not previously been described. The viability of the four alleles isolated from our screen (Table 1) and the nature of the mutations (Fig. 4B) suggest that they are hypomorphic mutations. Thus, we further analyzed the involvement of *ceh-43* in dopaminergic neuron differentiation using the *tm480* deletion allele, a likely null allele. Since animals carrying this allele die as embryos, we analyzed mosaic animals that specifically lost *ceh-43* expression in dopaminergic neurons (see the Material and Methods for details on mosaic analysis). These animals show defects in the expression of all dopamine pathway genes and in all dopaminergic neuron types in both the hermaphrodite (Fig. 3B,D) and the male (data not shown).

The function of *ceh-43* is not restricted to controlling the dopamine pathway genes—it also controls the expression of other terminal identity markers of dopaminergic neurons; namely, the *trp-4* and *asic-1* ion channel

Table 1. ceh-43 *mutant phenotypes*

Genotype and reporter transgenes[a]	*gfp(−)* CEPV	*gfp(−)* CEPD	*gfp(−)* ADE	*gfp(−)* PDE	Animals showing *gfp(−)* cells	Animals with extra *gfp(+)* cells	Animals with any phenotype	*n*
dat-1::gfp	0% ± 0%	0% ± 0%	0% ± 0%	0% ± 0%	0%	0%	0%	>100
cat-4::gfp	0% ± 0%	0% ± 0%	0% ± 0%	0% ± 0%	0%	0%	0%	>100
cat-2::gfp	0% ± 0%	0% ± %0	0% ± 0%	0% ± 0%	0%	0%	0%	>100
bas-1::gfp	0% ± 0%	0% ± 0%	0% ± 0%	0% ± 0%	0%	0%	0%	>100
asic-1::gfp	0% ± 0%	0% ± 0%	0% ± 0%	0% ± 0%	0%	0%	0%	>100
ceh-43(ot345); dat-1::gfp	0% ± 0%	18% ± 4%	6% ± 2%	0% ± 0%	41%	31%	58%	41
ceh-43(ot406); dat-1::gfp	0% ± 0%	65.9% ± 5%	50% ± 5%	0% ± 0%	95%	16%	97%	60
ceh-43(ot479); dat-1::gfp	100% ± 0%	100% ± 0%	0% ± 0%	0% ± 0%	100%	nd	100%	60
ceh-43(ot340); dat-1::gfp	100% ± 0%	100% ± 0%	0% ± 0%	0% ± 0%	100%	nd	100%	>100
ceh-43(ot340); cat-4::gfp	100% ± 0%	100% ± 0%	0% ± 0%	0% ± 0%	100%	nd	100%	40
ceh-43(ot340); asic-1::gfp	100% ± 0%	100% ± 0%	0% ± 0%	0% ± 0%	100%	nd	100%	>100
ceh-43(ot340); cat-2::gfp	100% ± 0%	100% ± 0%	0% ± 0%	0% ± 0%	100%	nd	100%	40
ceh-43(ot340); bas-1::gfp	8% ± 4%	6% ± 3%	0% ± 0%	0% ± 0%	8%	0%	14%	25
ceh-43(ot340); dat-1::gfp	100% ± 0%	100% ± 0%	0% ± 0%	0% ± 0%	100%	0%	0%	40
ceh-43(ot340); dat-1::gfp; Ex[ceh-43^{fosmid}]	2.5% ± 5%	5% ± 6%	0% ± 0%	0% ± 0%	0%	nd	0% ± 0%	40
ceh-43(ot340); dat-1::gfp; Ex[ceh-43^{DEL_fosmid}]	84.3% ± 9%	100% ± 0%	0% ± 0%	0% ± 0%	75%	nd	75%	35
eri-1;lin-15b; dat-1::gfp empty vector RNAi	0% ± 0%	0% ± 0%	0% ± 0%	12% ± 3.3%	18%	nd	18%	50
eri-1;lin-15b; dat-1::gfp ceh-43 RNAi	21% ± 4%	30% ± 4.5%	36% ± 4.7%	33% ± 4.6%	79%	nd	79%	52

gfp(−) refers to dopaminergic neurons showing absence of *gfp* expression. Adult animals were scored. (nd) Not determined.
[a]Reporters were as follows: *vtIs1* (*dat-1::gfp*); *otIs225* (*cat-4::gfp*); *otIs254* (*cat-2::gfp*); *otIs226* (*bas-1::gfp*); *otIs236* (*asic-1::gfp*).

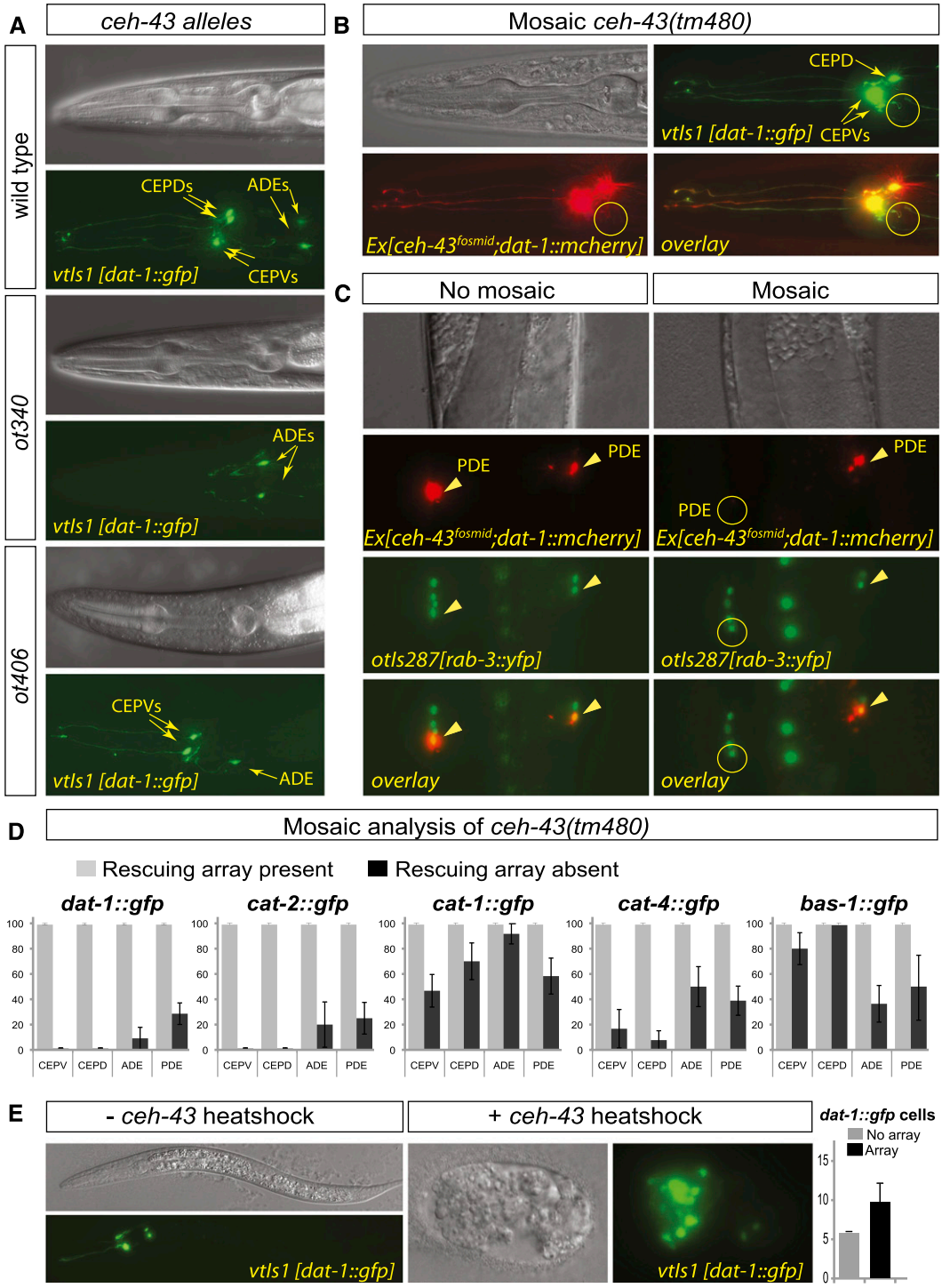

Figure 3. Loss-of-function and gain-of-function analysis of *ceh-43*. (*A*) *ceh-43* mutant alleles retrieved from screens for mutants with loss of *dat-1::gfp* expression. (*B–D*) Mosaic analysis of *ceh-43*-null mutant animals, showing loss of dopamine pathway genes (*B*,*D*) and no effect on expression of a panneuronal marker (*C*). For mosaic analysis, inviable null mutant *ceh-43(tm480)* animals were balanced with an extrachromosomal array that harbors a rescuing fosmid that contains the *ceh-43* locus and a *dat-1::mCherry* expression construct that allows assessing for the presence or absence of the array in specific dopaminergic neurons. For phenotypic output, expression of a chromosomally integrated, *gfp*-based fate marker was examined. (*E*) Ectopic expression of *ceh-43* under control of the heat shock leads to embryonic arrest and induction of *dat-1::gfp* expression in additional cells. The heat shock was applied at the end of gastrulation (~300 min after fertilization), and animals were scored ~16 h after heat shock.

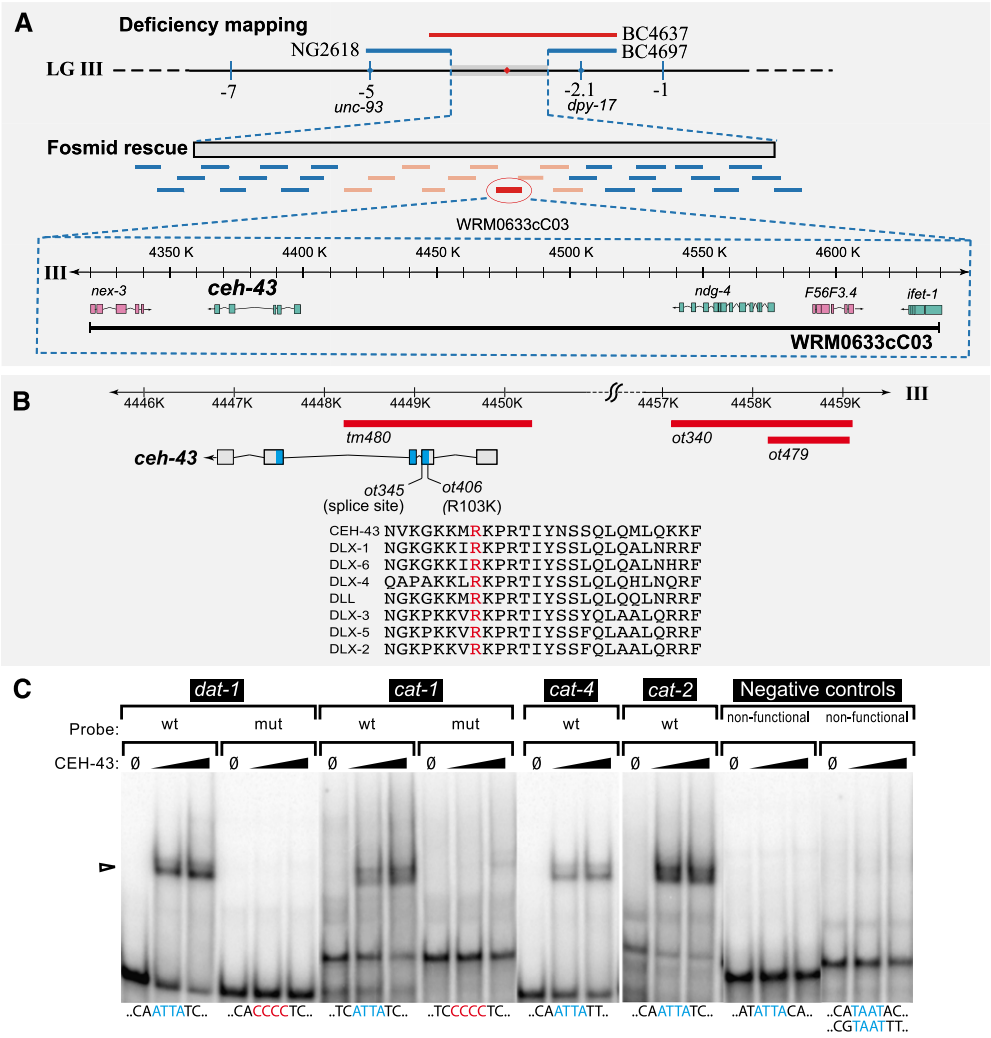

Figure 4. Molecular characterization of *ceh-43*. (*A*) Mapping and rescue of the *dopy-2* locus. *ot340* was mapped with high-throughput SNP mapping between −7 and −1 cM on LGIII. This interval was further narrowed down to five genes using three-factor mapping (*ot340* was located between *unc-93* and *dpy-17*), deficiency mapping (deficiency strain BC4637 failed to complement, whereas NG2618 and BC4697 complemented *ot340*), and fosmid rescue (fosmids were injected in three pools of 10; red pool rescued, and subsequently, fosmids in the rescuing pool were injected as single fosmids; only fosmid WRM0633cC03 rescued). RNAi against *ceh-43* recapitulated the phenotype. (*B*) Mutant alleles of the *ceh-43* locus. Red lines indicate the deletion alleles. Blue boxes of the gene structure denote the homeodomain. The sequence alignment shows part of the homeodomain, which is affected in the *ot406* allele. (*C*) EMSAs with CEH-43 protein on *dat-1*, *cat-1*, *cat-4*, and *cat-2* regulatory elements. CEH-43 binds to ATAAT-containing elements in *dat-1*, *cat-1*, *cat-4*, and *cat-2*, in which mutation of ATAAT affects reporter expression in vivo (arrowhead). Mutation of the ATAAT motifs in *dat-1* and *cat-1* abolishes binding in vitro. CEH-43 does not bind to other TAAT-containing sequences in *cat-2* (negative control; *left*) or *cat-4* (negative control; *right*), in which mutation of TAAT has no effect on reporter activity in vivo.

genes (Table 1; data not shown). The expression of a panneuronal marker, *rab-3*, is not affected (Fig. 3C), indicating that *ceh-43* does not affect the generation of dopaminergic neurons but affects the adoption of a specific neuronal identity. This phenotype is similar to the loss of the *trans*-acting factor for the Ets-binding site, AST-1 (Flames and Hobert 2009).

Two missense mutations in the *ceh-43* locus, identified through our EMS screen, revealed weaker defects (Figs. 3A, 4B; Table 1). Notably, two other alleles (*ot479* and *ot340*) are partially overlapping, <2-kb deletions that reside >7 kb upstream of the *ceh-43* locus (Fig. 4B; Supplemental Table 1). Both alleles fail to complement the missense alleles (Doitsidou et al. 2008). In these deletion alleles, a highly penetrant loss of dopaminergic neuron fate is restricted entirely to the CEPD and CEPV neuron types, while ADE and PDE neuron types are unaffected (Fig. 3A; Table 1). As we show below, both deletions eliminate *cis*-regulatory elements required for expression of *ceh-43* in the CEPD and CEPV neurons.

We further corroborated the notion that *ceh-43* acts through the ATAAT motifs by using an in vitro approach in which we tested whether bacterially produced CEH-43 protein can bind to regulatory elements from four of the

dopamine pathway genes. We readily detected such binding in gel shift assays (Fig. 4C). This in vitro binding was abrogated when the same ATAAT motifs that disrupted activity in the in vivo reporter gene assay were mutated (Fig. 4C). This binding is not the result of unspecific binding to the TAAT core motif, since other TAAT motifs in the minimal regulatory regions that do not match to the ATAAT consensus and showed no activity in the in vivo reporter assay do not bind CEH-43 (Fig. 4C, negative controls). Two other *ceh-43* targets that we identified in other cellular contexts also contain functional ATAAT-binding sites and bind CEH-43 in vitro (L Cochella, J Etchberger, N Abe, and O Hobert, unpubl. data).

ceh-43 not only is required to express terminal fate markers of dopaminergic neurons, but is sufficient to do so, at least in some cellular contexts. When misexpressed under control of a ubiquitous and inducible heat-shock promoter, up to twice as many dopamine marker-positive cells can be generated (Fig. 3E). The extent of ectopic dopaminergic neuron induction upon heat-shocked-induced *ceh-43* expression is similar to the extent of ectopic dopaminergic neuron induction upon heat-shocked induction of *ast-1* (Flames and Hobert 2009).

ceh-43 is expressed in dopaminergic neurons, and its function is continuously required to maintain the differentiated state of dopaminergic neurons

To examine *ceh-43* expression, we generated a fosmid-based reporter in which we recombineered *gfp* at the C-terminal end of the *ceh-43* locus in the context of an ~32-kb fosmid that contains several genes upstream of and downstream from *ceh-43* (Fig. 5A). Through colabeling with a dopaminergic neuron-specific marker, we found the fosmid reporter to be expressed in all dopaminergic neurons throughout the life of the neurons (Fig. 5B). Expression can also be observed in some additional head and body neurons as well as nonneuronal cells (Supplemental Fig. 2a). This expression was corroborated with immunostaining of endogenous CEH-43 protein using a pan-species anti-Distalless antibody (data not shown). As assessed with immunostaining for CEH-43 and reporter transgene for *ast-1* expression, despite the broad neuronal expression of both *ceh-43* and *ast-1*, they uniquely overlap in dopaminergic neurons plus one additional pair of nondopaminergic neurons in the head and one additional neuron in the midbody region (Supplemental Fig. 2b).

Introducing the smallest of the overlapping deletions present in the *ot340* and *ot479* alleles in the context of the fosmid reporter results in the loss of expression of the reporter specifically in the CEPD/V neuron types, demonstrating that these deletions affect relevant *cis*-regulatory elements (Fig. 5C). Moreover, we found that *ot340* mutants can be rescued with a wild-type fosmid containing the *ceh-43* locus but not by a fosmid in which these regulatory elements are deleted (Fig. 5D), corroborating the notion that *ceh-43* acts autonomously at least in the CEP neurons to affect their terminal differentiation.

We noted that animals carrying a missense mutation in the homeobox of *ceh-43* (*ot406* allele) show a significantly more pronounced defect in the expression of dopamine markers in adults compared with young larvae (Fig. 5E). This progressive loss of dopaminergic neuron identity indicates that *ceh-43*, like *ast-1* (Flames and Hobert 2009), is continuously required to maintain the differentiated state of dopaminergic neurons, a typical feature of terminal selector-type transcription factors (Hobert 2008).

Cell type- and target gene-specific interactions of ceh-43 and ast-1

The coexpression of CEH-43 and AST-1 in mature dopaminergic neurons as well as the presence of cognate and functionally required CEH-43- and AST-1-binding sites in dopamine pathway genes suggest that both proteins cooperate to activate dopamine pathway genes. The level of cooperation may differ in distinct dopaminergic neuron types, as suggested by the observation that in *ast-1*-null mutant animals (*gk463* allele), the *dat-1* gene is still normally expressed in one dopaminergic neuron type: the CEPV neurons (Table 2). In *ceh-43(ot406)* hypomorphic animals, *dat-1* expression in CEPV is also unaffected. In *ast-1(gk463); ceh-43(ot406)* double mutants, expression of *dat-1* is now strongly affected in the CEPV neurons (Table 2). This finding suggests that, in the absence of *ast-1*, correct *dat-1* expression in the CEPV neurons can still be ensured by *ceh-43*, but under such circumstances, even weak disruptions of *ceh-43* activity will severely affect *dat-1* expression. The genetic synergism between *ceh-43* and *ast-1* was further confirmed by the analysis of double-hypomorph mutants (Table 2). In *ceh-43(ot406)* mutants, *dat-1* expression is only partially affected in the ADE and CEPD neurons, but this phenotype is greatly enhanced when the *ceh-43(ot406)* allele is combined with the *ast-1(hd1)* hypomorphic allele, which alone has no defects in *dat-1* expression (Table 2).

These results not only corroborate the genetic interaction of these two factors, but also make the point that individual dopamine pathway genes may display a differentially tuned requirement for individual *trans*-acting factors in specific cell types. In CEPVs, the loss of *ast-1* can be tolerated in regard to *dat-1* expression because *ceh-43* can ensure robust *dat-1* expression, but for other dopamine pathway genes, loss of *ast-1* alone results in strong defects in expression. Moreover, this is not a general compensatory role for *ceh-43* because in the other dopaminergic neurons, the presence of *ceh-43* cannot compensate for a complete *ast-1* loss in regard to *dat-1* expression.

Two distinct Pbx genes constitute the third component of the dopamine regulatory signature

To identify the factors that operate through the third *cis*-regulatory motif required for dopaminergic neuron expression, the predicted Pbx HD site, we turned to a candidate gene approach. The *C. elegans* genome codes for three Pbx genes: *ceh-20*, *ceh-40*, and *ceh-60* (Van Auken

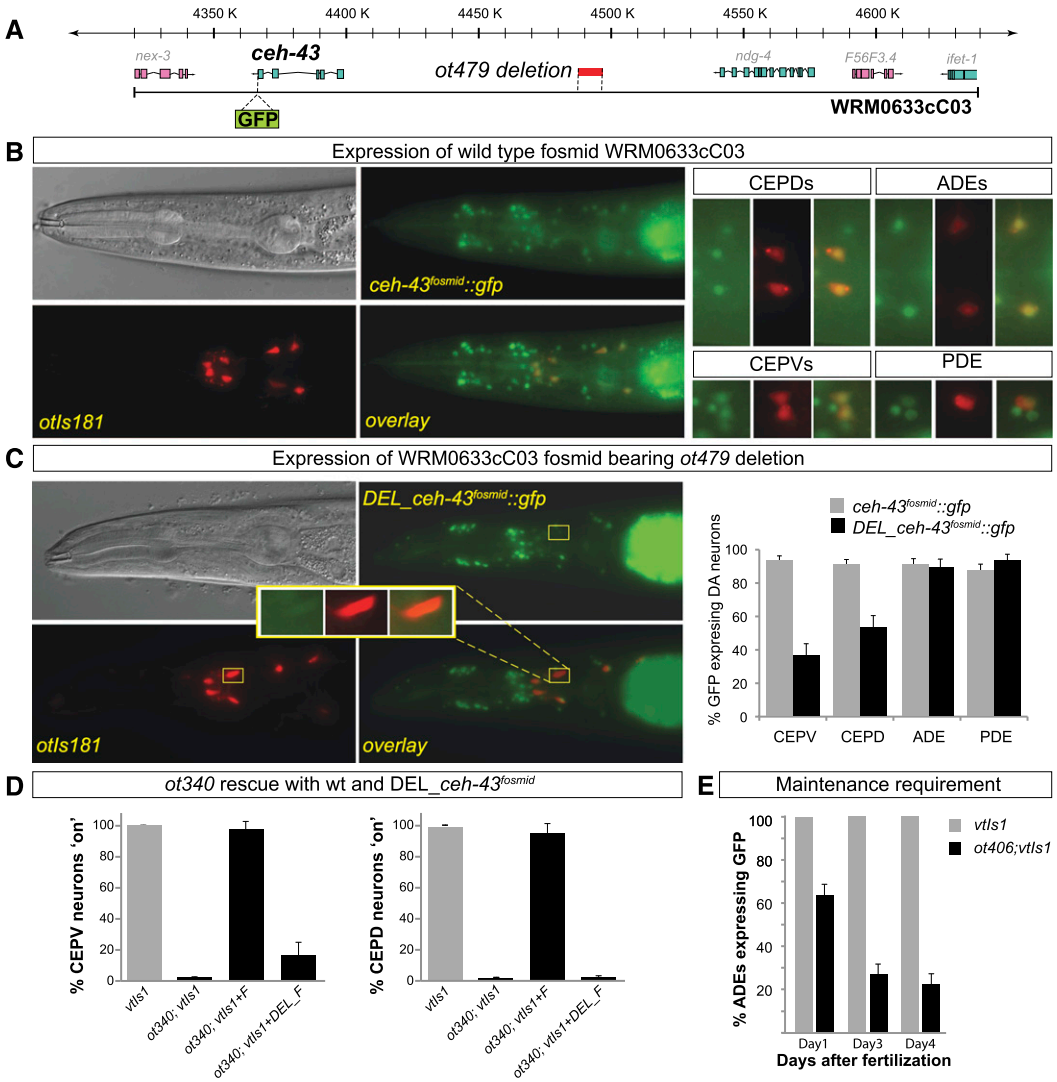

Figure 5. Expression pattern of *ceh-43*. (*A*) Structure of the fosmid reporter for *ceh-43*. (*B*) Expression pattern of the fosmid reporter (*ceh-43^{fosmid}::gfp*) and its overlap with dopaminergic markers in adult animals. (*C*,*D*) The *ot479* allele is a regulatory allele. The allele is shown in *A*. (*C*) Introduction of this deletion into the fosmid reporter abolishes expression of *ceh-43* in CEP dopaminergic neurons. (*D*) The ability of the wild-type *ceh-43* fosmid to rescue the *ceh-43(ot340)* mutant phenotype is abolished upon introduction of the *ot479* deletion in the fosmid. (*E*) Sustained expression of *ceh-43* in dopaminergic neurons is likely functionally relevant for maintaining dopaminergic fate. The *ot406* allele shows a progressive loss in expression of the *dat-1::gfp* marker in ADEs, suggesting a continuous requirement of *ceh-43* gene activity.

et al. 2002; Mukherjee and Burglin 2007). Functions for these genes in the nervous system have not previously been reported.

We examined functions for each of the three Pbx genes in the context of dopaminergic neuron development using deletion or missense alleles available for each of the loci (Fig. 6A). We found that *ceh-20*-null mutants (*ok541*) show defects in *dat-1::gfp* expression exclusively in the midbody dopaminergic PDE neuron class (Fig. 6B). Similar *dat-1::gfp* expression defects in the PDE neurons are observed in multiple *ceh-20* alleles (Table 3). Apart from *dat-1::gfp*, expression of all other members of the dopamine pathway is also defective in the PDE neurons of *ceh-20* mutants (Fig. 6B). The effect of *ceh-20* on some of the members of the pathway (e.g., *cat-2*) is stronger than expected from the mutation of the Pbx *cis*-regulatory site. This may either be indicative of cross-regulation of the *trans*-acting factors controlling the dopamine pathway genes (e.g., *ceh-20* regulating dopamine pathway genes and also regulating *ceh-43*) or hint toward the presence of atypical *cis*-regulatory sites for *ceh-20*/*Pbx*.

As in the case of *ast-1* and *ceh-43*, the function of *ceh-20* is not restricted to controlling the dopamine pathway genes. *ceh-20* also controls the expression of other terminal identity markers of dopaminergic neurons; namely, the *trp-4* and *asic-1* ion channel genes [12% of animals show missing *trp-4::gfp*, and 80% show missing *asic-1::gfp* in *ceh-20(mu290)* mutants; $n > 30$]. Consistent

Table 2. *Genetic interactions between* ceh-43/Dll *and* ast-1

Genotype	*dat-1::gfp(−)* CEPV	*dat-1::gfp(−)* CEPD	*dat-1::gfp(−)* ADE	*n*
Wild type	0% ± 0%	0% ± 0%	0% ± 0%	>100
ast-1(gk463)	3% ± 4%	100% ± 0%	100% ± 0%	70
ceh-43(ot406)	4% ± 4%	46% ± 5%	35% ± 5%	78
ast-1(gk463); ceh-43(ot406)	100% ± 0%	100% ± 0%	100% ± 0%	50
ast-1(hd1)	0% ± 0%	0% ± 0%	0% ± 0%	100
ceh-43(ot406)	0% ± 0%	59% ± 4%	39% ± 3.9%	152
ast-1(hd1); ceh-43(ot406)	0% ± 0%	92.8% ± 2.3%	46% ± 4.4%	124

Animals were scored at 25°C at the L1 stage. Numbers refer to the percentage of respective neuron that failed to express *dat-1::gfp* (*vtIs1*).

with a role in initiating and maintaining the PDE differentiation program, a *ceh-20* reporter gene construct is expressed in the PDE neurons throughout their life (Fig. 6C).

In a fraction of *ceh-20(mu290)* mutant animals, no neurons appear to be generated from the post-deirid lineage, which generates the dopaminergic PDE and glutamatergic PVD sensory neurons (10% of the animals

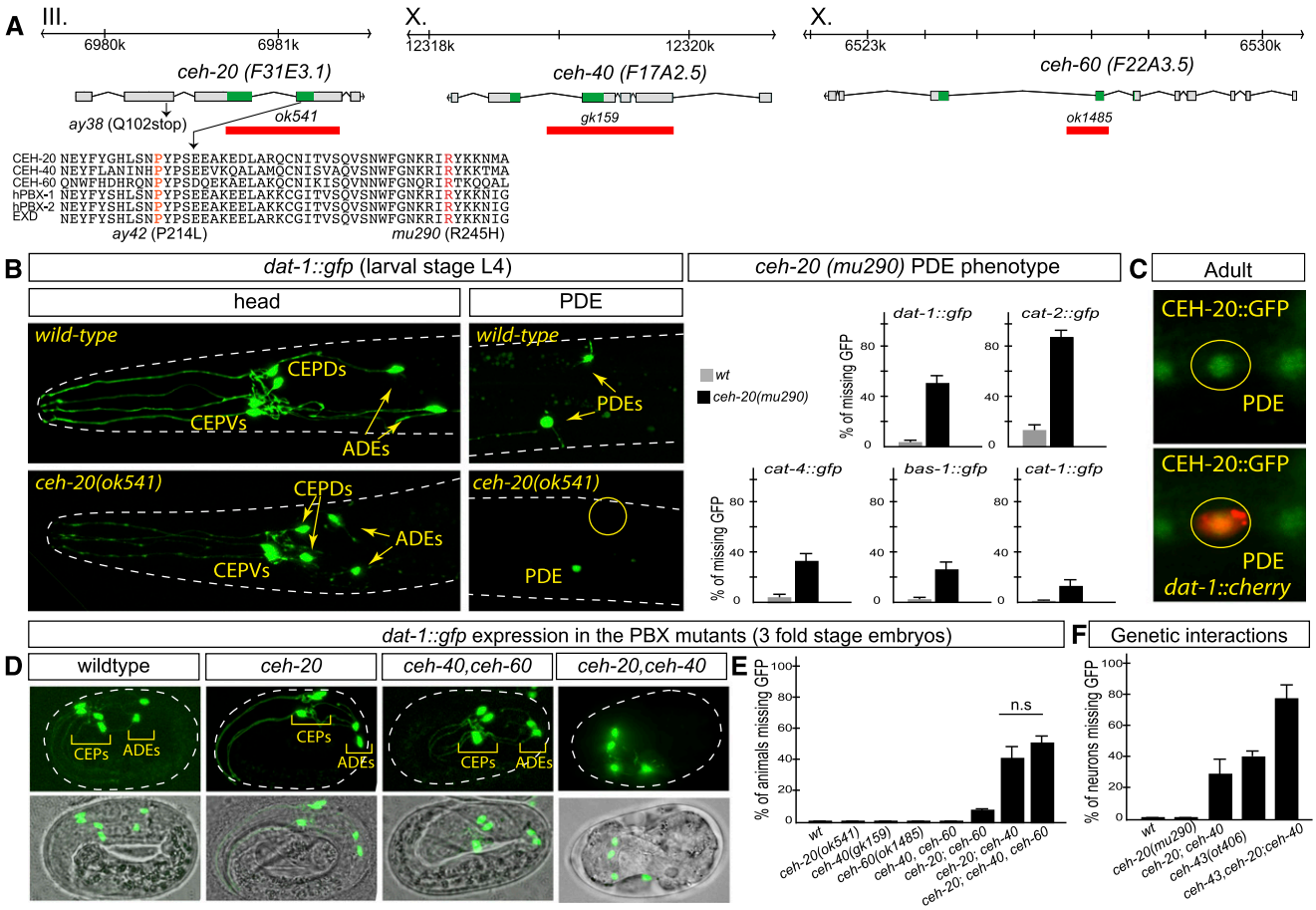

Figure 6. Pbx genes controlling dopamine fate. (*A*) Schematic drawing of alleles used in this study. Red bars indicate the region covered by the deletion allele. Green boxes of the gene structure denote the homeodomain. Neither *ceh-60* nor *ceh-40* alleles had been characterized before. (*B*) *ceh-20* PDE mutant phenotype. Images of *dat-1::gfp* expression showing no phenotype for the head neurons and missing PDE in the *ceh-20(ok541)*-null allele. Bar graphs show the characterization of the PDE phenotype for all dopamine pathway genes in *ceh-20(mu290)* mutants. (*C*) *ceh-20* is expressed in the PDE, and its expression is maintained throughout the life of the worm. The picture shows expression of a CEH-20::GFP construct. (*D*) *dat-1::gfp* expression in the different Pbx mutant combinations. Due to the embryonic lethality of *ceh-20(ok541); ceh-40(gk159)* double mutants, scoring was done in threefold embryos. (*E*) Quantification of *dat-1::gfp* expression in the different Pbx mutant combinations. The double mutant *ceh-20(ok541); ceh-40(gk159)* shows a dopaminergic neuron differentiation defect that is not significantly enhanced in the triple mutant *ceh-20(ok541); ceh-40(gk159) ceh-60(ok1485)*. Apart from animals that lack *dat-1::gfp* expression, we also frequently observed animals in which additional cells express *dat-1::gfp*. We have not further pursued this phenotype. (*F*) Genetic interaction between Pbx genes and *ceh-43*. Quantification of cells expressing *dat-1::gfp*. The defect in *dat-1::gfp* expression in a Pbx double-mutant combination of *ceh-20(mu290);ceh-40(gk159)* alleles is enhanced by the presence of a hypomorphic *ceh-43* allele (*ot406*). The effect is more than expected from simple additivity.

Table 3. *Pbx genes affecting dopaminergic neuron differentiation*

	Loss of *dat-1::gfp* expression in all or a subset of head neurons (ADE and CEPs)	Loss of *dat-1::gfp* expression of midbody neurons (PDE)	*n*
Wild type	0% ± 0%	3% ± 2%	55
ceh-20(ok541)	0% ± 0%	55% ± 7%[a]	56
ceh-20(ay42)	0% ± 0%	54% ± 8%	39
ceh-20(ay38)	0% ± 0%	40% ± 8%[b]	39
ceh-20(mu290)	0% ± 0%	50% ± 7%[c]	56
ceh-40(gk159)	0% ± 0%	0% ± 0%	50
ceh-60(ok1485)	0% ± 0%	0% ± 0%	50
ceh-40(gk159), ceh-60(ok1485)	0% ± 0%	0% ± 0%	52
ceh-20(ok541); ceh-40(gk159)	38% ± 5%[d]	Not scorable due to embryonic lethality	86
ceh-20(ok541); ceh-60(ok1485)	9% ± 4%[e]	50% ± 6%	66
ceh-20(ok541); ceh-40(gk159), ceh-60(ok1485)	53% ± 4%[f]	Not scorable due to embryonic lethality	126

[a]*ok541* allele 5% of the escapers show extra GFP cells in PDE region
[b]*ay38* allele 11% of the escapers show extra GFP cells in PDE region
[c]*mu290* allele 5% of animals show extra GFP cells in PDE region
[d]Thirty-six percent of the embryos show extra GFP cells
[e]Fourteen percent of the embryos show extra GFP cells
[f]Twenty-four percent of the embryos show extra GFP cells

generate no neurons in the post-deirid, 14% generate only one neuron, 37% generate the normal number of two neurons, and 39% generate more than two neurons; n = 49), suggesting lineage defects. However, the PDE differentiation defects in *ceh-20* mutants are not merely due to PDE lineage defects because *dat-1* expression is also lost in animals with an unaffected PDE lineage, as assessed by correct expression of the pan-neuronal *rab-3* marker in PDE (Supplemental Fig. 3).

In contrast to *ast-1* and *ceh-43* (Fig. 3E), *ceh-20* alone is not able to induce the production of additional dopaminergic neurons upon heat-shock promoter-mediated misexpression (as assessed by ectopic *dat-1::gfp* expression) (data not shown). Animals that coexpress heat-shock promoter-driven *ceh-43*, *ast-1*, and *ceh-20* constructs show no more ectopic dopaminergic neuron production than animals expressing *ceh-43* or *ast-1* alone (data not shown).

The effect of *ceh-20* loss of function is restricted to the PDE neurons, since the expression of several dopamine pathway genes is completely unaffected in head dopaminergic neurons (Fig. 6B; data not shown). Since our *cis*-regulatory analysis indicates that putative Pbx-binding sites are required for the expression of dopamine pathway genes in not just PDE neurons but also all dopaminergic head neurons, we tested whether removal of other Pbx genes affected dopaminergic head neurons. *ceh-40*-null mutants, *ceh-60*-null mutants, and *ceh-40 ceh-60* double-null mutants did not show defects in *dat-1::gfp* expression in any dopaminergic neuron type (Fig. 6D; Table 3). However, removal of *ceh-40* in a *ceh-20* mutant background, which alone has no effect on head dopaminergic neurons, results in defects in *dat-1::gfp* expression in head dopaminergic neurons (Fig. 6D; Table 3). In contrast, combining the *ceh-60* mutation with the *ceh-20* mutation only shows very mild (<10% penetrant) differentiation defects of head dopaminergic neurons. The defect of the *ceh-20; ceh-40* double mutants are not further enhanced in *ceh-20; ceh-40 ceh-60* triple-null mutants (Fig. 6E; Table 3). However, the *ceh-20; ceh-40* double-mutant defects are significantly enhanced by removal of *ceh-43/Distalless* (*ceh-43; ceh-20; ceh-40* triple mutants) (Fig. 6F), which corroborates the genetic interactions among distinct members of the dopaminergic regulatory ensemble. Taken together, *ceh-20* functions nonredundantly in the PDE neurons to control their differentiation but acts redundantly with *ceh-40* in head dopaminergic neurons. Consistent with this notion, we could not detect expression of *ceh-40* in the midbody PDE neuron but detected expression of both *ceh-20* and *ceh-40* in the head dopaminergic neurons (Supplemental Fig. 4).

We examined the overlap of *ast-1*, *ceh-43*, and Pbx gene expression to assess whether their coexpression uniquely defines dopaminergic neurons. We detected coexpression of *ast-1*, *ceh-43*, and *ceh-20* in at least one nondopaminergic neuron (SDQL). The two most likely possibilities that could explain this lack of specificity are that either there are still other components of the dopaminergic transcription factor collective that remain to be identified and are not expressed in SDQL (those are unlikely to be DNA-binding factors, since the ETS, HD, and Pbx sites are sufficient to exclusive drive gene expression in dopaminergic neurons) or, alternatively, repressive mechanisms could operate in SDQL to inhibit the expression of dopaminergic neuron identity.

We also investigated whether *ast-1*, *ceh-43*, and *ceh-20* cross-regulate each other's expression. Focusing on the PDE neuron class, we found that *ast-1* expression is unaffected in *ceh-43* and *ceh-20* mutants (data not shown). However, *ast-1* is required for *ceh-43* expression (Supplemental Fig. 5) but is not required for *ceh-20* expression (data not shown). These results indicate that these factors act to some extent independently of one another.

Taken together, our reverse genetic analysis identified a third component of the regulatory signature of dopaminergic neuron terminal differentiation programs. As

predicted by the *cis*-regulatory analysis, this third component is a Pbx gene. In one dopaminergic neuron type, only one Pbx gene is required; in other dopaminergic neuron types, a combination of Pbx genes act redundantly.

AST-1, CEH-43, and CEH-20 cooperate to activate dopamine pathway genes in a heterologous, cell-based cis*-activation assay*

To further explore the mechanistic basis of the interaction of AST-1, CEH-43, and a Pbx factor in controlling and maintaining dopaminergic neuron identity, we turned to a yeast transcription assay (Topalidou et al. 2011). We inserted the regulatory regions of two of the dopamine pathway genes, *bas-1* and *cat-2*, upstream of the yeast minimal *cyc-1* promoter to drive the expression of β-galactosidase in the presence of different combinations of AST-1, CEH-43/Dll, and CEH-20/Pbx. We found that expression of neither factor alone is able to induce gene expression, but expression of all three factors together results in very strong induction of gene expression (Fig. 7). In double combinations, only combinations with AST-1 yield induction of gene expression, albeit at lower levels than the triple combination (Fig. 7). These findings support a model in which all three factors work in a cooperative manner to control gene expression.

The partial activity of AST-1 in combination with either CEH-43 or CEH-20 observed in the yeast assay matches with the general theme revealed by the *cis*-regulatory analysis: While Ets-binding sites are always essential for reporter gene expression, the loss of either HD or Pbx *cis*-regulatory sites can sometimes be tolerated, likely because AST-1 can still operate with the remaining cofactor.

Discussion

We described here a regulatory signature that specifies the terminal differentiation program of all dopaminergic neurons in *C. elegans*. We found that a combination of three *cis*-regulatory motifs is sufficient to dictate exclusive gene expression in all dopaminergic neurons of the *C. elegans* nervous system. This *cis*-regulatory signature is recognized by a combination of three types of *trans*-acting factors: an Ets domain factor, a Dlx-type HD, and Pbx-type HD proteins. These proteins work as terminal selectors to initiate and maintain the terminally differentiated state of *C. elegans* dopaminergic neurons. The mutant phenotypes of all three types of regulatory factors are similar to each other and are also similar to the phenotypic consequences of removal of terminal selector transcription factors that control terminal differentiation of other *C. elegans* neuronal cell types (Hobert 2008, 2011): Neurons are generated and still express panneuronal features but fail to adopt neuron-type-specific properties (in this case, the terminal features that define dopaminergic identity).

Early patterning roles have been described for Pbx genes in several different contexts in both vertebrates and invertebrates (Laurent et al. 2008), but a function of Pbx genes in directly controlling post-mitotic, terminal differentiation programs has not yet, to our knowledge, been described in any system to date. Notably, the role of *C. elegans* Pbx genes in controlling dopaminergic neuron fate appears to be independent of the function of Hox genes, which are common interaction partners for Pbx genes in many cellular contexts (Moens and Selleri 2006; Laurent et al. 2008). The anterior-most expression boundary of the most anterior HOX cluster gene, *ceh-13*, does not cover the region where the dopaminergic neuron classes ADE, CEPD, and CEPV are located (Wittmann et al. 1997). As expected, *ceh-13*-null mutants do not affect dopaminergic neuron marker gene expression (our unpublished observations). Even though the role of Pbx as a Hox cofactor has been well documented, Hox-independent functions of Pbx family members have been

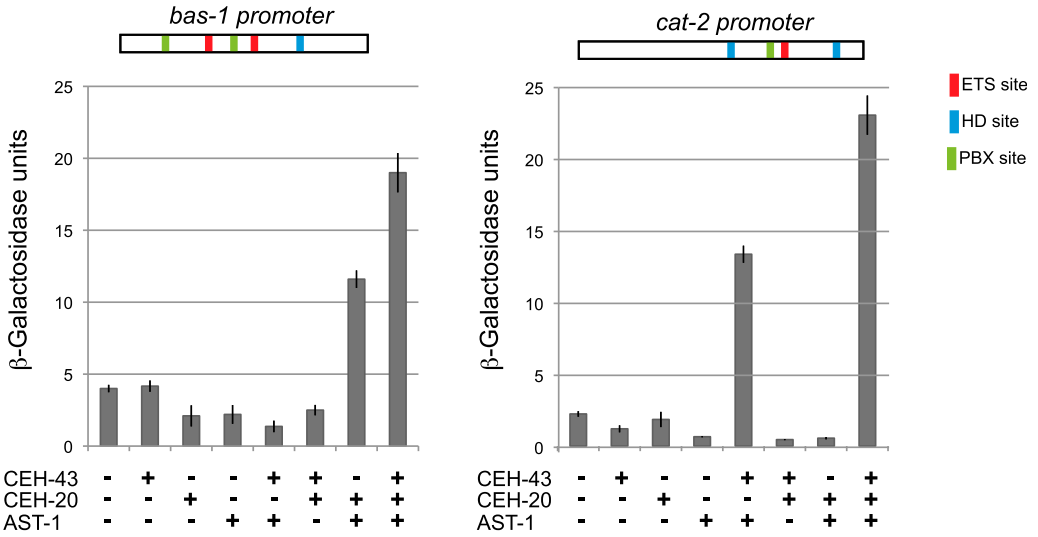

Figure 7. Cooperative gene activation by the AST-1/CEH-43/CEH-20 transcription factor collective. AST-1, CEH-43, and CEH-20 cooperate to induce transcription in yeast cells. AST-1, CEH-43, and CEH-20 or their combinations were expressed in yeast cells together with the *bas-1p::lacZ* (*left* panel) or *cat-2p::lacZ* (*right* panel) reporters. Values are the mean ± SEM of two to five replicates. $P < 0.02$, Student's *t*-test.

described in several species, including *C. elegans* (Yang et al. 2005), vertebrates (Ferretti et al. 2011), and flies (e.g., Casares and Mann 2001; Bessa et al. 2002). Intriguingly, one study in *Drosophila* has shown that the *ceh-43* ortholog Distalless genetically interacts with the fly Pbx homolog Exd (Dong et al. 2000). The function of worm Pbx genes in controlling dopaminergic neuron differentiation is also independent of another common cofactor of Pbx gene, the Meis-type homeobox gene. There is a single canonical Meis homolog in the *C. elegans*, encoded by the *unc-62* loucs (Van Auken et al. 2002), but animals carrying a null mutant allele of *unc-62* show no reduction in the production of dopaminergic neurons (our unpublished observations).

As revealed through the use of a heterologous yeast transcriptional assay system, the Ets/Dlx/Pbx regulatory ensemble acts in a cooperative manner to activate gene expression. Cooperativity is also suggested by genetic interaction tests that show synergy (rather than additivity) in loss-of-function scenarios. The cooperativity is reminiscent of the terminal differentiation program executed by the cholinergic AIY interneuron class in *C. elegans*. Here again, two factors (the HD transcription factors *ttx-3* and *ceh-10*), each expressed in a number of distinct neuron types, uniquely overlap in their expression in the AIY interneurons, where they cooperatively activate scores of terminal differentiation genes that define AIY interneuron identity (Wenick and Hobert 2004). The key conceptual and mechanistic difference between the cooperativity of *ttx-3*/*ceh-10* and *ast-1*/*ceh-43*/*Pbx* lies in what generally is referred to as the "grammar" of *cis*-regulatory logic; that is, the overall organization of the individual *cis*-regulatory motifs. In the case of the AIY-expressed, *ttx-3*/*ceh-10*-dependent regulatory elements, the two binding sites for TTX-3 and CEH-10 are precisely spaced and oriented to allow for cooperative binding. Contrasting such fixed grammar, the *cis*-regulatory regions that control dopaminergic neuron expression are composed of *cis*-regulatory sites (ETS domain, HD, and Pbx-binding site) that display no fixed spacing, relative orientation, or overall number of individual sites, yet the activity of the *cis*-acting factors is nevertheless cooperative, as determined by mutant analysis and heterologous transcription assays in yeast. A recently published study described a similar regulatory architecture for cardiac gene expression in *Drosophila* and proposed the term "transcription factor collective" (Junion et al. 2012). In that case, five cardiogenic transcriptional regulators cooperatively activate target genes in the absence of a defined motif grammar, and in some target cases, only a subset of the these factors are bound. Our studies broaden this concept to a different cell type and organism, suggesting universality of this regulatory mechanism.

We further extend the transcription factor collective model by our ability to examine its activity in four distinct neuronal subclasses: ADE, CEPD, CEPV, and PDE. Even though these four dopaminergic classes share many molecular and functional features, these neurons have distinct lineage histories and distinct axo/dendritic projections and are located in different parts of the nervous system. Even though our *cis*-regulatory analysis in combination with our genetic analysis clearly shows that all four neuron classes employ this combination of regulatory factors, the extent of the involvement of individual factors differs for each individual dopamine pathway gene in individual dopaminergic neuron subtypes. For example, *dat-1* expression critically depends on *ast-1* expression in all neurons except for the CEPV neurons, in which *ast-1* function can be compensated for by *ceh-43* gene activity. Head dopaminergic neuron types rely on multiple Pbx genes, as a mutant phenotype is only evident in double-mutant combinations, while midbody dopaminergic neurons critically depend on only one Pbx gene (*ceh-20*). The underlying common theme of all of these interactions may be that the presence of multiple factors ensures robust expression of the target genes of these transcription factors and that each cell and target gene may use slightly different although related means to ensure this robustness.

The same Ets/Dlx/Pbx regulatory signature that we describe here may also function in dopaminergic neurons in vertebrates. We showed previously that Etv1, a mouse homolog of *ast-1*, is required for the appropriate differentiation of olfactory dopaminergic neurons (Flames and Hobert 2009). The defects of Etv1 mutants, however, are not as pronounced as *ast-1* mutant defects are; for example, while TH expression is affected by Etv1, DAT expression is not (Cave et al. 2010). One potential explanation for this partial effect can be seen in the regulatory logic described here: Other factors that cooperate with Etv1 may partially compensate for its loss. In fact, the failure to see a loss of *Dat* expression in Etv1 mouse mutants (Cave et al. 2010) may precisely mirror the absence of a phenotype of *dat-1* expression in the CEPV neurons of *ast-1*-null mutants; that is, the *ast-1* phenotype is only revealed if a cooperating factor (in this case, *ceh-43*) is disabled.

The factors that cooperate with vertebrate Etv1 to ensure dopaminergic neuron differentiation in the olfactory bulb could be the same as we defined here in *C. elegans.* A null mutation of Dlx2 results in early specification defects of olfactory bulb neurons and a loss of TH-positive neurons (Qiu et al. 1995). Dlx2 expression is maintained in adult dopaminergic neurons, and later function of Dlx2 in terminal differentiation specifically of the dopamine olfactory bulb neurons is suggested through expression of a dominant-negative form of Dlx2, which resulted in dopaminergic neuron specification defects (Brill et al. 2008). To assess whether the third component of the worm Ets/Dlx/Pbx regulatory signature is also expressed in mouse olfactory bulb dopaminergic neurons, we stained mouse olfactory bulbs with a Pbx1/2/3 antibody and observed Pbx immunoreactivity in TH-positive neurons (Supplemental Fig. 6). Intriguingly, Pbx1 is expressed in the midbrain dopaminergic neurons, and Pbx1 mutants show dopaminergic axon pathfinding defects (Sgado et al. 2012). These observations suggest that the regulatory code between olfactory bulb dopaminergic neuron specification and worm dopaminergic neuron specification could, at least in part, be phylogenetically conserved.

Materials and methods

DNA constructs and cis*-regulatory analysis*

All *gfp*-based reporter constructs were generated using the pPD95.75 vector as backbone and by subcloning into the multiple cloning site. Mutagenesis reactions were performed using the QuickChangeII XL site-directed mutagenesis kit (Stratagene). Constructs were injected or crossed into *otIs181* (*dat-1::cherry; ttx-3::cherry*) to allow easy identification of dopaminergic neurons. All reporter constructs were injected at 50 ng/μL using *rol-6(su1006)* as a coinjection marker (100 ng/μL). For each construct, two or three independent lines were scored (at least 30 animals per line). Sequences of the wild-type minimal constructs are in the Supplemental Material, and the nature of the mutations introduced is indicated in Supplemental Figure 1.

Cloning of dopy-2/ceh-43

ot340 was mapped with high-throughput SNP mapping (Davis et al. 2005) between −7 and −1 cM on LGIII. Three-factor mapping mapped *ot340* between *unc-93* and *dpy-17*. Deficiency mapping placed the *ot340* locus within deficiency BC4637 but outside deficiencies NG2618 and BC4697. Fosmid rescue of pools of fosmids were injected, followed by single-fosmid injections of the fosmids included in the rescuing pool. Single-fosmid injection of WRM0633cC03 rescued the *ot340* phenotype. Only five genes were included in the rescuing fosmid. RNAi against each of these genes was performed using a bacterial feeding protocol (Kamath and Ahringer 2003) in an *eri-1;lin-15b;vtIs1* mutant background (Kennedy et al. 2004). RNAi against *ceh-43* recapitulated the phenotype. Sanger sequencing revealed point mutations in the *ceh-43* locus for *ot406* and *ot345* alleles but no mutation in alleles *ot340* and *ot479*. Whole-genome sequencing using an Illumina platform followed by data analysis using MAQGene (Bigelow et al. 2009) of *ot340* and *ot479* revealed two overlapping deletions >7 kb upstream of *ceh-43* CDS (Supplemental Table 1). Deletions were confirmed by Sanger sequencing to span genomic regions on LGIII from 4457789 to 4459806 (2018 bp) for *ot340* and from 4458889 to 4459754 (866 bp) for *ot479*.

Fosmid recombineering and ceh-43 expression analysis

ceh-43-containing fosmid WRM0633cC03 was tagged as described before (Tursun et al. 2009) to generate OH9993 (*otEx4439[ceh-43^{fosmid}::gfp;ttx-3::dsred;rol-6]*) and integrated to generate OH10447 (*otIs339[ceh-43^{fosmid}::gfp;ttx-3::dsred;rol-6]*). The *ot479* deletion was engineered in the recombineered fosmid as described before (Tursun et al. 2009) to generate *DEL_ceh-43^{fosmid}::gfp*. For determining the neuronal expression of *ceh-43*, strain MDH33 (*otIs339[ceh-43^{fosmid}::gfp;ttx-3::dsred;rol-6];otIs355[rab-3::NLS::tagRFP]*) was used. To determine *ceh-43* expression in *ast-1* mutants, strain MDH38 (*ast-1(gk463) bli-2(e768) unc-4(e120);Ex[ast-1^{cosmid};ttx-3::gfp;dat-1::mCherry];otIs339; otIs355*) was used.

The ability of the recombineered fosmid with and without the deletion to rescue *ceh-43(ot340)* mutants was assessed using strains OH10412 (*ceh-43(ot340);vtIs1; otEx4439[ceh-43^{fosmid}::gfp;ttx-3::rfp;rol-6]*) and OH10424 (*ceh-43(ot340); vtIs1;[ExDEL_ceh-43^{fosmid}::gfp;ttx-3::rfp;rol-6]*), respectively.

Mosaic analysis

The strain CH1890 (*ceh-43(tm480)/qC1 e1259q339[gIs26]*) was injected with *ceh-43* fosmid WRM0633cC03 (15 ng/μL) and *dat-1::mCherry* (40 ng/μL) to generate the strain *ceh-43 (tm480);Ex[ceh-43^{fosmid};dat-1::mCherry]*. This balanced strain was then crossed with animals bearing chromosomally integrated, *gfp*-based transgenes that monitor expression of dopamine pathway genes. Mosaic animals were identified by the loss of *dat-1::mCherry* in one or more dopaminergic neurons, and the fate of these neurons was assessed by scoring the expression of the integrated transgene. Note that the experimental design of this mosaic analysis does not allow distinguishing between loss of the array and inability of the array to rescue the mutant phenotype; that is, animals that fail to show *dat-1::mCherry* may carry the array containing *dat-1::mcherry* and the *ceh-43* fosmid, but *ceh-43* fails to rescue the *dat-1* expression defect, and this is why no *dat-1::mCherry* expression is observed in the cell. This very failure of *ceh-43* to rescue the expression of the *dat-1::mCherry* phenotype implies that *ceh-43* mutation has an effect on *dat-1* expression to begin with. Thus, any dopaminergic phenotype in *ceh-43*-rescued mutant worms attests to a role of *che-43* in dopaminergic fate whether this phenotype is caused by loss of the array per se or its inability to rescue with 100% penetrance.

Ectopic expression of ceh-43

ceh-43 under control of the heat-shock promoter was injected into *vtIs1[dat-1::gfp]* to generate the strain *Ex[hsp-16.2::ceh-43;hsp-16.2::NLS-mCherry;ttx-3::ds-red];vtIs1*. The heat shock was applied at the end of gastrulation (~300 min after fertilization), and animals were scored ~16 h after heat shock.

Protein purification, EMSA, and yeast transcription assay

Full-length His-tagged CEH-43 was expressed in BL21 cells and purified using Co^{2+} chromatography. EMSAs were performed as previously described (Gebelein et al. 2002). For all EMSAs, CEH-43 was tested at two concentrations: 60 nM and 180 nM. Probe sequences are listed in the Supplemental Material.

For the yeast assays, the reporter plasmids *bas-1p::lacZ* and *cat-2p::lacZ* were constructed by cloning the 279 bp and 354 bp upstream of the ATG of the bas-1 and cat-2 genes, respectively, into the HindIII/XmaI sites of plasmid pXCZ55 (a gift from Mark Ptashne). Yeast strains expressing *bas-1p:: lacZ* and *cat-2p::lacZ* were constructed by digesting plasmids with these DNAs with ApaI and integrating them in the URA locus of the yeast strain YPH499 (Stratagene). Positive colonies were identified using single-colony PCR. Yeast transformation was performed using the Liac/SS carrier DNA/PEG method (Gietz and Woods 2002). Induction was achieved as described previously (Topalidou et al. 2011), and cells were harvested when OD_{600} = 1.5. Liquid β-galactosidase assays were performed as described by Reynolds et al. (2001).

Acknowledgments

We thank Qi Chen for expert assistance in generating transgenic strains, Alexander Boyanov for expert assistance in whole-genome sequencing, Shohei Mitani at Tokyo Women's Medical University School of Medicine for the *tm480* allele, the Oklahoma and Vancouver knockout consortia for *gk* and *ok* alleles, Bob Waterston for the *ceh-40* fosmid reporter *stIs11399*, and the Servei Central de Suport a la Investigació Experimental (SCSIE) from Universidad de Valencia for microscope assistance. This work was funded by EMBO post-doctoral fellowships and Marie Curie Funds (to M.D. and N.F.), the New York Stem Cell Foundation Fellowships and the Spanish Government (SAF2011-

26273) (to N.F), the NIH (R01NS039996-05; R01NS050266-03 to O.H.; R01GM30997 to M.C.; R01GM054510 to R.S.M.; and F32GM099160 to N.A.), and the Stavanger University Hospital (to M.D.). N.F is a NARSAD Young Investigator. O.H. is an Investigator of the Howard Hughes Medical Institute.

References

Aspock G, Burglin TR. 2001. The *Caenorhabditis elegans* distal-less ortholog ceh-43 is required for development of the anterior hypodermis. *Dev Dyn* **222:** 403–409.

Bessa J, Gebelein B, Pichaud F, Casares F, Mann RS. 2002. Combinatorial control of *Drosophila* eye development by eyeless, homothorax, and teashirt. *Genes Dev* **16:** 2415–2427.

Bigelow H, Doitsidou M, Sarin S, Hobert O. 2009. MAQGene: Software to facilitate *C. elegans* mutant genome sequence analysis. *Nat Methods* **6:** 549.

Brill MS, Snapyan M, Wohlfrom H, Ninkovic J, Jawerka M, Mastick GS, Ashery-Padan R, Saghatelyan A, Berninger B, Gotz M. 2008. A dlx2- and pax6-dependent transcriptional code for periglomerular neuron specification in the adult olfactory bulb. *J Neurosci* **28:** 6439–6452.

Casares F, Mann RS. 2001. The ground state of the ventral appendage in *Drosophila*. *Science* **293:** 1477–1480.

Cave JW, Akiba Y, Banerjee K, Bhosle S, Berlin R, Baker H. 2010. Differential regulation of dopaminergic gene expression by Er81. *J Neurosci* **30:** 4717–4724.

Davis MW, Hammarlund M, Harrach T, Hullett P, Olsen S, Jorgensen EM. 2005. Rapid single nucleotide polymorphism mapping in *C. elegans*. *BMC Genomics* **6:** 118.

Doitsidou M, Flames N, Lee AC, Boyanov A, Hobert O. 2008. Automated screening for mutants affecting dopaminergic-neuron specification in *C. elegans*. *Nat Methods* **5:** 869–872.

Dong PD, Chu J, Panganiban G. 2000. Coexpression of the homeobox genes Distal-less and homothorax determines *Drosophila* antennal identity. *Development* **127:** 209–216.

Ferretti E, Li B, Zewdu R, Wells V, Hebert JM, Karner C, Anderson MJ, Williams T, Dixon J, Dixon MJ, et al. 2011. A conserved Pbx–Wnt–p63–Irf6 regulatory module controls face morphogenesis by promoting epithelial apoptosis. *Dev Cell* **21:** 627–641.

Flames N, Hobert O. 2009. Gene regulatory logic of dopamine neuron differentiation. *Nature* **458:** 885–889.

Flames N, Hobert O. 2011. Transcriptional control of the terminal fate of monoaminergic neurons. *Annu Rev Neurosci* **34:** 153–184.

Gebelein B, Culi J, Ryoo HD, Zhang W, Mann RS. 2002. Specificity of Distalless repression and limb primordia development by abdominal Hox proteins. *Dev Cell* **3:** 487–498.

Gietz RD, Woods RA. 2002. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol* **350:** 87–96.

Hobert O. 2008. Regulatory logic of neuronal diversity: Terminal selector genes and selector motifs. *Proc Natl Acad Sci* **105:** 20067–20071.

Hobert O. 2011. Regulation of terminal differentiation programs in the nervous system. *Annu Rev Cell Dev Biol* **27:** 681–696.

Hobert O, Carrera I, Stefanakis N. 2010. The molecular and gene regulatory signature of a neuron. *Trends Neurosci* **33:** 435–445.

Holmberg J, Perlmann T. 2012. Maintaining differentiated cellular identity. *Nat Rev Genet* **13:** 429–439.

Iversen SD, Iversen LL. 2007. Dopamine: 50 years in perspective. *Trends Neurosci* **30:** 188–193.

Junion G, Spivakov M, Girardot C, Braun M, Gustafson EH, Birney E, Furlong EE. 2012. A transcription factor collective defines cardiac cell fate and reflects lineage history. *Cell* **148:** 473–486.

Kamath RS, Ahringer J. 2003. Genome-wide RNAi screening in *Caenorhabditis elegans*. *Methods* **30:** 313–321.

Kennedy S, Wang D, Ruvkun G. 2004. A conserved siRNA-degrading RNase negatively regulates RNA interference in *C. elegans*. *Nature* **427:** 645–649.

Laurent A, Bihan R, Omilli F, Deschamps S, Pellerin I. 2008. PBX proteins: Much more than Hox cofactors. *Int J Dev Biol* **52:** 9–20.

Moens CB, Selleri L. 2006. Hox cofactors in vertebrate development. *Dev Biol* **291:** 193–206.

Mukherjee K, Burglin TR. 2007. Comprehensive analysis of animal TALE homeobox genes: New conserved motifs and cases of accelerated evolution. *J Mol Evol* **65:** 137–153.

Panganiban G, Rubenstein JL. 2002. Developmental functions of the Distal-less/Dlx homeobox genes. *Development* **129:** 4371–4386.

Qiu M, Bulfone A, Martinez S, Meneses JJ, Shimamura K, Pedersen RA, Rubenstein JL. 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* **9:** 2523–2538.

Reynolds A, Lundblad V, Dorris D, Keaveney M. 2001. Yeast vectors and assays for expression of cloned genes. *Curr Protoc Mol Biol* **39:** 13.6.1–13.6.6.

Sgado P, Ferretti E, Grbec D, Bozzi Y, Simon HH. 2012. The atypical homeoprotein Pbx1a participates in the axonal pathfinding of mesencephalic dopaminergic neurons. *Neural Dev* **7:** 24.

Spitz F, Furlong EE. 2012. Transcription factors: From enhancer binding to developmental control. *Nat Rev Genet* **13:** 613–626.

Sulston J, Dew M, Brenner S. 1975. Dopaminergic neurons in the nematode *Caenorhabditis elegans*. *J Comp Neurol* **163:** 215–226.

Topalidou I, van Oudenaarden A, Chalfie M. 2011. *Caenorhabditis elegans* aristaless/Arx gene alr-1 restricts variable gene expression. *Proc Natl Acad Sci* **108:** 4063–4068.

Tursun B, Cochella L, Carrera I, Hobert O. 2009. A toolkit and robust pipeline for the generation of fosmid-based reporter genes in *C. elegans*. *PLoS ONE* **4:** e4625.

Van Auken K, Weaver D, Robertson B, Sundaram M, Saldi T, Edgar L, Elling U, Lee M, Boese Q, Wood WB. 2002. Roles of the Homothorax/Meis/Prep homolog UNC-62 and the Exd/Pbx homologs CEH-20 and CEH-40 in *C. elegans* embryogenesis. *Development* **129:** 5255–5268.

Wenick AS, Hobert O. 2004. Genomic *cis*-regulatory architecture and *trans*-acting regulators of a single interneuron-specific gene battery in *C. elegans*. *Dev Cell* **6:** 757–770.

Wittmann C, Bossinger O, Goldstein B, Fleischmann M, Kohler R, Brunschwig K, Tobler H, Muller F. 1997. The expression of the *C. elegans* labial-like Hox gene ceh-13 during early embryogenesis relies on cell fate and on anteroposterior cell polarity. *Development* **124:** 4193–4200.

Yang L, Sym M, Kenyon C. 2005. The roles of two *C. elegans* HOX co-factor orthologs in cell migration and vulva development. *Development* **132:** 1413–1428.