Developmental Biology 368 (2012) 415–426

Contents lists available at SciVerse ScienceDirect

Developmental Biology

journal homepage: www.elsevier.com/locate/developmentalbiology

Genomes and Developmental Control

The in vivo dissection of direct RFX-target gene promoters in *C. elegans* reveals a novel *cis*-regulatory element, the C-box

Jan Burghoorn [1,2], Brian P. Piasecki [1,3], Filip Crona [4], Prasad Phirke, Kristian E. Jeppsson [5], Peter Swoboda *

Karolinska Institute, Center for Biosciences at NOVUM, Department of Biosciences and Nutrition, Hälsovägen 7, S-141 83 Huddinge, Sweden

ARTICLE INFO

Article history:
Received 31 December 2011
Received in revised form
23 April 2012
Accepted 25 May 2012
Available online 5 June 2012

Keywords:
Caenorhabditis elegans
Cis-regulatory motifs
Regulatory Factor X transcription factors
Ciliated sensory neurons
Cilia

ABSTRACT

At the core of the primary transcriptional network regulating ciliary gene expression in *Caenorhabditis elegans* sensory neurons is the RFX/DAF-19 transcription factor, which binds and thereby positively regulates 13–15 bp X-box promoter motifs found in the *cis*-regulatory regions of many ciliary genes. However, the variable expression of direct RFX-target genes in various sets of ciliated sensory neurons (CSNs) occurs through as of yet uncharacterized mechanisms. In this study the *cis*-regulatory regions of 41 direct RFX-target genes are compared using in vivo genetic analyses and computational comparisons of orthologous nematode sequences. We find that neither the proximity to the translational start site nor the exact sequence composition of the X-box promoter motif of the respective ciliary gene can explain the variation in expression patterns observed among different direct RFX-target genes. Instead, a novel enhancer element appears to co-regulate ciliary genes in a DAF-19 dependent manner. This cytosine- and thymidine-rich sequence, the C-box, was found in the *cis*-regulatory regions in close proximity to the respective X-box motif for 84% of the most broadly expressed direct RFX-target genes sampled in this study. Molecular characterization confirmed that these 8–11 bp C-box sequences act as strong enhancer elements for direct RFX-target genes. An artificial promoter containing only an X-box promoter motif and two of the C-box enhancer elements was able to drive strong expression of a GFP reporter construct in many *C. elegans* CSNs. These data provide a much-improved understanding of how direct RFX-target genes are differentially regulated in *C. elegans* and will provide a molecular model for uncovering the transcriptional network mediating ciliary gene expression in animals.

Introduction

The correct spatiotemporal expression of genes is required for the proper development of all cells. In multicellular organisms variations in gene expression patterns define cellular specialization and functional diversification (Prud'homme et al., 2007; Wray, 2007). Genes that are commonly regulated in specific cell types are often hardwired within a specific gene regulatory network. Whereby, similarly expressed genes share specific *cis*-regulatory DNA sequences that are bound and regulated by common transcriptional regulators, including transcription factors (Tuch et al., 2008). In the model organism *Caenorhabditis elegans* for example, numerous *cis*-regulatory sequences have been identified that promote the expression of genes broadly or specifically expressed in certain groups of neurons (Etchberger et al., 2007; Hobert et al., 2010; McCarroll et al., 2005; Nokes et al., 2009; Ruvinsky et al., 2007; Swoboda et al., 2000).

Cilia are complex microtubule-based organelles that facilitate a variety of motile- and sensory-specific functions in many eukaryotic organisms (Berbari et al., 2009; Rosenbaum and Witman, 2002; Silverman and Leroux, 2009). In animals, Regulatory Factor X (RFX) transcription factors (TFs) are key direct upstream regulators of ciliary-specific genes (Piasecki et al., 2010; Swoboda et al., 2000). In this transcriptional network, X-box promoter motif sequences residing in the *cis*-regulatory regions of more than 50 known ciliary genes are bound and positively regulated by RFX TFs (Blacque et al., 2005; Efimenko et al., 2005; El Zein et al., 2009; Laurencon et al., 2007). X-box promoter motifs are 13–15 bp imperfect inverted repeat sequences (Emery et al., 1996), which typically reside within 300 bp of the translational start site of the respective direct RFX-target genes (Blacque et al., 2005; Efimenko et al., 2005; Laurencon

* Corresponding author. Fax: +46 8 774 5538.
E-mail address: peter.swoboda@ki.se (P. Swoboda).
[1] JB and BPP contributed equally and should be considered joint first authors.
[2] Current address: Landgoed Ekenstein, Alberdaweg 70, NL-9901 TA Appingedam, The Netherlands.
[3] Current address: Lawrence University, Department of Biology, Appleton, WI 54911, USA.
[4] Current address: Stockholm University, Wenner-Gren Institute for Developmental Biology, S-106 91 Stockholm, Sweden.
[5] Current address: Karolinska Institute, Department of Cell and Molecular Biology, S-171 77 Stockholm, Sweden.

0012-1606/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ydbio.2012.05.033

et al., 2007). Phylogenetic footprinting of ciliary gene promoters has shown that this regulatory pathway is highly conserved in animals (Piasecki et al., 2010).

In *C. elegans*, cilia are exclusively formed on a subset of terminally differentiated neurons, so-called ciliated sensory neurons (CSNs; Perkins et al., 1986). In contrast to mammalian genomes, which encode at least seven RFX TF genes (Aftab et al., 2008), the *C. elegans* genome encodes only a single RFX TF gene, *daf-19* (Swoboda et al., 2000). However, multiple isoforms of *daf-19* occur in *C. elegans*, including A, B, C (cilia-specific), and M (male-specific) isoforms (Senti et al., 2009; Senti and Swoboda, 2008; Wang et al., 2010). Mutations in the *C. elegans* gene *daf-19* lead to the complete loss of all ciliated structures in an otherwise viable organism (Perkins et al., 1986; Swoboda et al., 2000). Bioinformatics-based comparisons and expression profiling experiments have revealed a large battery of putative and known direct RFX-target genes in *C. elegans* (Blacque et al., 2005; Chen et al., 2006; Efimenko et al., 2005; Phirke et al., 2011).

While about two-thirds of the more than 40 direct DAF-19/RFX-target genes identified to date in *C. elegans* are expressed in most or all 60 CSNs, approximately one-third are expressed in only a specific subset of CSNs. Of note, the CSN-specific expression is significantly reduced or eliminated when expressed in a *daf-19/rfx* mutant background or upon mutation of the respective X-box promoter motif sequences (Blacque et al., 2005; Burghoorn et al., 2007; Chen et al., 2006; Efimenko et al., 2006, 2005; Li et al., 2008; Murayama et al., 2005; Schafer et al., 2003; Swoboda et al., 2000; Williams et al., 2008; Winkelbauer et al., 2005). However, the mechanisms controlling direct RFX-target gene expression to either all or to only specific subsets of CSNs remain largely unknown. In this study, we conduct bioinformatics-based comparisons and in vivo dissections of direct RFX-target gene promoters in closely related *Caenorhabditis* species. These analyses reveal a novel *cis*-regulatory enhancer element, the C-box, which is located in close proximity (<60 bp) to the X-box promoter motif of direct RFX-target genes that are broadly expressed in CSNs.

Materials and methods

Strains and growth conditions

Unless otherwise indicated, the wild-type *C. elegans* strain, Bristol N2, was used for all experiments. The DAF-19/RFX dependent expression of *xbx-1*, *dyf-2*, *xbx-9*, and *nhr-44* was examined using the strains JT204 *daf-12(sa204)* and JT6924 *daf-19(m86); daf-12(sa204)* as previously described (Senti et al., 2009). All strains were cultured using standard procedures (Brenner, 1974). Germline transformations were conducted for all promoter-to-GFP fusion constructs using microinjection (Mello and Fire, 1995; Mello et al., 1991). Extrachromosomal arrays containing a GFP-tagged transgene injected at 100 ng/μl together with an *elt-2::mCherry* co-injection marker (gift from G. Jansen) were used for all expression analyses. All strains used and strain construction details are available upon request.

Generation of plasmid constructs for transgenesis

Construction of 2 kb promoter::GFP fusion constructs for the genes *xbx-1*, *dyf-2*, *mks-3* and *che-2* have been described previously (Efimenko et al., 2006; Efimenko et al., 2005; Phirke et al., 2011; Schafer et al., 2003). Promoter::GFP fusion constructs of variable lengths were generated for the genes *xbx-1*, *mks-3*, *bbs-8*, *che-2*, *dyf-2*, *xbx-9*, *nhr-4*, *nhr-44*, *nhr-194*, *nhr-219*, *nhr-258* and *pes-1* using PCR amplification followed by subsequent cloning of each of the respective regions into the pPD95.75, pPD95.77 or pPD95.79 GFP expression vectors (gift from A. Fire). The 2 kb promoter::GFP *xbx-1* construct was elongated to various lengths using Lambda phage DNA (#SMO0101/2/3, Fermentas GmbH, St. Leon-Rot, Germany). X-box promoter motif mutations for the 2 kb *xbx-1*, *dyf-2*, *xbx-9* and *nhr-44* promoter::GFP fusion constructs were generated using fusion PCR (Hobert, 2002). Minimal X-box motif promoters were constructed by separately cloning the 14 bp X-box promoter motif sequences from the genes *xbx-1*, *mks-3*, *bbs-8*, *che-2*, *dyf-2*, *xbx-9* and *nhr-44* into the pPD95.75 GFP expression vector exactly 109 bp upstream of the translational start site (ATG) of the reporter gene. Mutations of exactly 10 bp were introduced into the respective C-box sites in the *xbx-1*, *mks-3*, *bbs-8*, *che-2* and *xbx-9* GFP expression vectors using fusion PCR (Hobert, 2002). The *xbx-1* artificial promoter construct was generated by PCR amplifying Lambda phage DNA (#SMO0101/2/3) using modified oligonucleotides, which preserved the native configuration, distance, and composition of only the X-box promoter motif and C-box enhancer elements; the modified product was subsequently cloned into the pPD95.75 GFP expression vector.

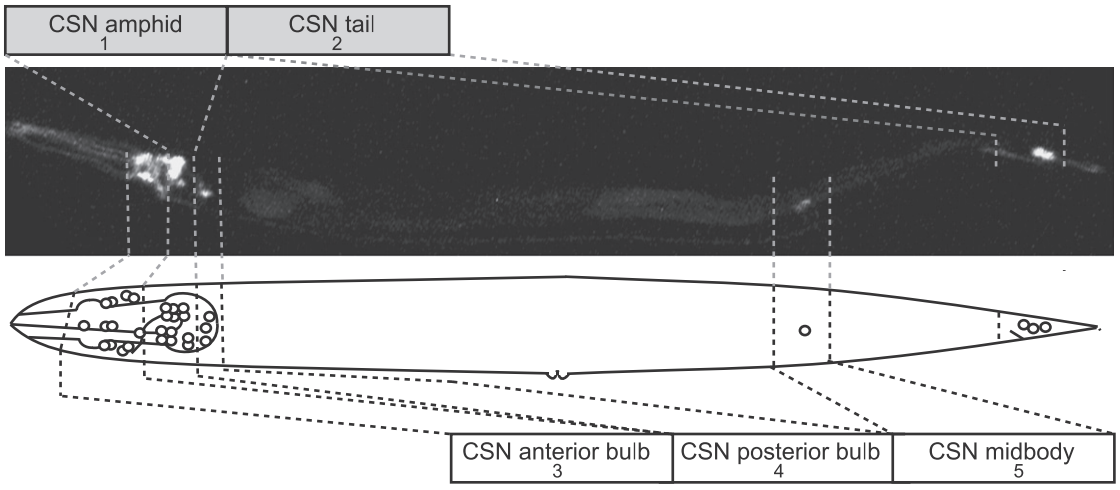
Microscopy and imaging

For live imaging of worms expressing GFP-tagged fusion proteins, worms were anesthetized using 20 mM sodium azide in M9 buffer and immobilized on a 2% agar pad. Worms were analyzed using conventional fluorescence microscopy using a Zeiss Axioplan 2 microscope at a total magnification of 630× (Carl Zeiss, Jena, Germany). GFP-tagged fusion proteins were localized and assigned to CSNs using a 40× objective lens on a Zeiss confocal microscope (CLSM510). Confocal projections were prepared using the Zeiss LSM image browser.

Analysis of GFP expression patterns

To significantly reduce possible experimental variability introduced by the mosaic expression pattern of a particular transgene, three of six independent transgenic lines with the highest penetrance were selected and analyzed for the respective promoter-to-GFP-fusion constructs. For all experiments, transgenic lines were grown at 20 °C for at least three generations prior to analysis. The 60 *C. elegans* CSNs present in an adult hermaphrodite worm (see also at www.wormatlas.org) were divided into 5 subgroups or anatomical regions including, neurons that reside in the amphids (region 1=24 CSNs: AWAL/R, AWBL/R, AWCL/R, AFDL/R, ASEL/R, ADFL/R, ASGL/R, ASHL/R, ASIL/R, ASJLR, ASKL/R, ADLL/R) or in the tail (region 2=5 CSNs: PHAL/R, PHBL/R, PQR), neurons surrounding the anterior bulb (region 3=24 CSNs: BAGL/R, CEPVL/R, CEPDL/R, IL1L/R, IL2L/R, IL1VL/R, IL2VL/R, IL1DL/R, IL2DL/R, OLLL/R, OLQVL/R, OLQDL/R) or the posterior bulb (region 4=5 CSNs: ADEL/R, FLPL/R, AQR) of the pharynx, and neurons in the midbody region of the worm (region 5=2 CSNs: PDEL/R). In the CSNs of the amphids (region 1) and in the tail (region 2) direct RFX-target genes are consistently expressed at high frequencies. In all other CSNs direct RFX-target gene expression patterns are typically more variable (cf. Fig. 1). The average number of cells expressing GFP in each of these respective anatomical regions was determined for at least 30 transgenic animals per construct. The unmodified 2 kb promoter-to-GFP-fusion construct of a given direct RFX-target gene was used as a baseline control for all comparisons and the expression levels of all mutated and variable length promoter-fusion constructs are depicted as fold-change relative to the level of the unmodified construct (cf. Figs. 4–7, Supplemental Figs. S1 and S2, and Supplemental Table S1). To account for the possible impact of photobleaching during the analyses, for all experiments the anatomical regions of all

A



B

nr	identity	gene name	description	X-box	position distance to ATG	RFX dependent expression pattern	references
1	Y110A7A.20	*ift-20*	IFT20	GTCTCTATAGCAAC	-61	1, 2, 3, 4, 5.	Blacque et al., 2005
2	Y38F2AL.2	*tza-1/mksr-2*	B9D2	GTTGCCGTGGCAAC	-65	1, 2, 3, 4, 5.	Williams et al., 2008
3	R01H10.6	*bbs-5*	BBS-5	GTCTCCATGGCAAC	-66	1, 2, 3, 4, 5.	Efimenko et al., 2005
4	R148.1	*mks-1/xbx-7*	MKS-1	GTCACCATAGGAAC	-70	1, 2, 3, 4, 5.	Efimenko et al., 2005
5	F35D2.4	*mks-3*	MKS-3	ATCACCATGACAAC	-75	1, 2, 3, 4, 5.	Phirke et al., 2011
6	F59C6.7	*che-13*	IFT57	GTTGCTATAGCAAC	-75	1, 2, 3, 4, 5.	Efimenko et al., 2005
7	M28.7	*nph-1*	NPHP1	GTTGCCAGGGGCAAC	-77	1, 2, 3, 4, 5.	Winkelbauer et al., 2005
8	F02D8.3	*xbx-1*	DYNC2LI1	GTTTCCATGGTAAC	-78	1, 2, 3, 4, 5.	Schafer et al., 2003
9	T25F10.5	*bbs-8*	BBS8	GTACCCATGGCAAC	-84	1, 2, 3, 4, 5.	Efimenko et al., 2005
10	K03E6.4	*tza-2/mksr-1*	B9D1	GTTCCCTTGGCAAC	-85	1, 2, 3, 4, 5.	Williams et al., 2008
11	T27B1.1	*osm-1*	IFT172	GCTACCATGGCAAC	-86	1, 2, 3, 4, 5.	Efimenko et al., 2005
12	ZK328.7	*ift-139*	IFT139	GTTACCATGGCAAT	-88	1, 2, 3, 4, 5.	Blacque et al., 2005
13	F20D12.3	*bbs-2*	BBS2	GTATCCATGGCAAC	-94	1, 2, 3, 4, 5.	Efimenko et al., 2005
14	Y105E8A.5	*bbs-1*	BBS1	GTTCCCATAGCAAC	-99	1, 2, 3, 4, 5.	Efimenko et al., 2005
15	R31.3	*osm-6*	IFT52	GTTACCATAGTAAC	-100	1, 2, 3, 4, 5.	Efimenko et al., 2005
16	C27H5.7	*dyf-13*	TTC26	GTCTCCATAGCAAC	-102	1, 2, 3, 4, 5.	Blacque et al., 2005
17	Y75B8A.12	*bbs-7/osm-12*	BBS7	GTTGCCATAGTAAC	-108	1, 2, 3, 4, 5.	Efimenko et al., 2005
18	C47E8.6	*xbx-10*	GAS8	GTTACCATGCCAAC	-108	1, 2, 3, 4, 5.	Blacque et al., 2005
19	Y41G9A.1	*osm-5*	IFT88	GTTACTATGGCAAC	-116	1, 2, 3, 4, 5.	Efimenko et al., 2005
20	T24A11.2	*xbx-5*	Unknown	GTCTCCATGACAAC	-122	1, 2, 3, 4, 5.	Efimenko et al., 2005
21	F38G1.1	*che-2*	IFT80	GTTGTCATGGTGAC	-130	1, 2, 3, 4, 5.	Efimenko et al., 2005
22	R13H4.1	*nph-4*	NPHP4	ATTTCCATGACAAC	-168	1, 2, 3, 4, 5.	Winkelbauer et al., 2005
23	F10B5.4	*tub-1*	TUB	ATCTCCATGACAAC	-183	1, 2, 3, 4, 5.	Efimenko et al., 2005
24	C02H7.1	*dyf-11*	TRAF3IP1	GTCTCCATGACAAC	-196	1, 2, 3, 4, 5.	Li et al., 2008
25	M04C9.5	*dyf-5*	MAPK	GTTACCATAGAAAC	-285	1, 2, 3, 4, 5.	Chen et al., 2006
26	ZK520.3	*dyf-2*	IFT144	GTTACCAAGGCAAC	-153	1, 2, 4.	Efimenko et al., 2006
27	C26B2.4	*nhr-258*	NHR	GTAACTATGGAAAC	-70	1, 2.	Current work
28	C27A7.4	*che-11*	IFT140	ATCTCCATGGCAAC	-86	1, 2.	Efimenko et al., 2005
29	C04C3.5	*dyf-3*	Qilin	GTTTCTATGGGAAC	-88	1, 2.	Murayama et al., 2005
30	M04D8.6	*xbx-3*	Unknown	GTTGTCTTGGCAAC	-98	1, 2.	Efimenko et al., 2005
31	C15C8.1	*xbx-9*	Unknown	GTAACCATAGCAAC	-124	1, 2.	Current work
32	F53A2.4	*nud-1*	NudC	GTATCCATGAAAAC	-263	1, 2.	Efimenko et al., 2005
33	T19A5.4	*nhr-44*	NHR	GTCTTCATGGCAAC	-77	1.	Efimenko et al., 2005
34	C23H5.3	*xbx-4*	Unknown	GTTGCCATGACAAC	-82	1.	Efimenko et al., 2005
35	F59E11.8	*nhr-194*	NHR	GCTGCCATAACAAC	-159	1.	Current work
36	T19A5.5	*nhr-219*	NHR	GTTTCCATAACAAT	-186	1.	Current work
37	Y102E9.1	*odr-4*	Membrane protein	ATCGTCATGGTAAC	-201	1.	Efimenko et al., 2005
38	F32B6.1	*nhr-4*	NHR	TTCGCCATGGAAAC	-260	1.	Current work
39	T28H11.4	*pes-1*	Forkhead TF	GTTTTCATAAGAAC	-271	1.	Current work
40	F40F9.1	*xbx-6*	FAS Inhibitor 2	GTTTCCATGGAAAC	-152	2.	Efimenko et al., 2005
41	T28F4.2	*asic-2*	Ion channel	GTATCCATGGGCAC	-62	3.	Phirke et al., 2011

Fig. 1. Expression patterns of direct RFX-target genes in ciliated sensory neurons (CSNs) of the worm *C. elegans*. (A) Fluorescence confocal projection (top) and schematic diagram (bottom) depicting the expression pattern of the *xbx-1* promoter fused to GFP, a fusion construct that is ubiquitously expressed in all 60 CSNs. The hermaphrodite worm shown here has been subdivided into five specific anatomical regions to be able to classify the variety in expression patterns often seen with direct RFX-target genes: CSNs of the amphids (region 1) and in the tail (region 2), surrounding the anterior bulb (region 3) and posterior bulb (region 4) of the pharynx, and in the midbody region of the worm (region 5). Open circles in the worm schematic illustrate the positions of all CSN cell bodies in a hermaphrodite animal (only the left CSNs of lateral pairs are shown). (B) Description of the 41 experimentally characterized X-box regulated genes in *C. elegans* analyzed in this study. The sequence composition and position of the X-box promoter motif relative to the translational start site (ATG) is depicted for each gene. Using the specific anatomical expression regions defined in the schematic above (cf. Fig. 1A), the CSN-specific expression pattern is listed for each gene.

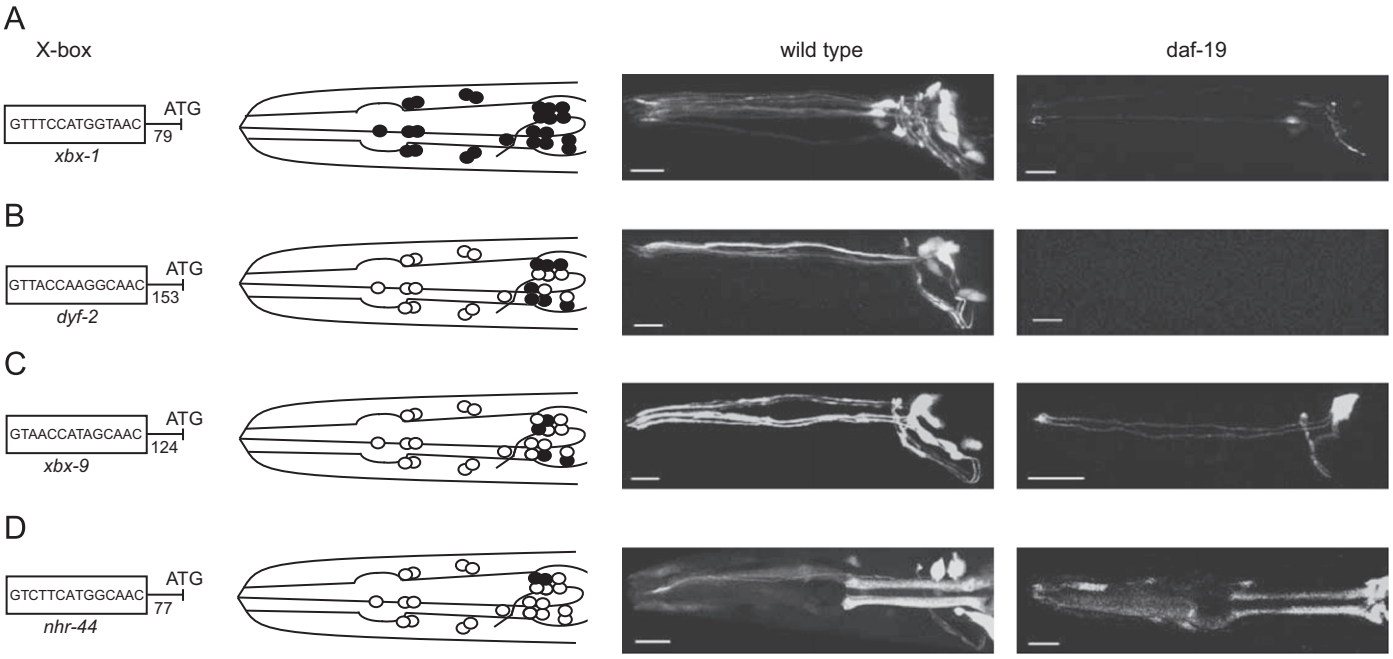


Fig. 2. Expression of direct RFX-target genes in wild type and *daf-19* mutant worms.(A–D) Schematic diagrams (left) and expression patterns of promoter-to-GFP fusion constructs (right) of representative direct RFX-target genes (*xbx-1*, *dyf-2*, *xbx-9*, *nhr-44*) are shown. Note that each gene is expressed in different numbers of CSNs, respectively. For each gene, the corresponding X-box promoter motif sequence composition and position relative to the translational start site (ATG) is shown. Schematic head diagrams illustrate the wild-type expression pattern of each gene: the presence (black ovals) or absence (white ovals) of GFP expression in neuronal cell bodies is depicted for each of the 48 CSNs (only the left CSNs of lateral pairs are shown) collectively found in the amphids and the group of CSNs surrounding the anterior bulb of the pharynx (cf. Fig. 1A, regions 1 and 3). Expression patterns observed in other CSNs (cf. Fig. 1B) are not depicted. Confocal projections of the head region (right) from transgenic worms expressing promoter-to-GFP fusion constructs are shown for each gene in both wild type and *daf-19* mutant backgrounds. In all panels the scale bars depict 10 μm.

transgenic animals were always analyzed in the same order of sequence: first region 4, then region 3, region 1, region 5, and region 2 last.

Bioinformatics

400 bp of promoter region from each of the 41 *C. elegans* direct RFX-target genes sampled in this study were first analyzed using a multiple expectation maximization for motif elicitation (MEME) algorithm (Bailey and Elkan, 1994). Subsequently, orthologous 400 bp promoter regions from *C. elegans*, *C. briggsae* and *C. remanei* were individually collected for each of the respective direct RFX-target genes sampled. Based on the GFP expression analysis comparisons carried out for each of the *C. elegans* direct RFX-target genes, all the corresponding orthologous gene promoter regions from the other two *Caenorhabditis* species (*C. briggsae* and *C. remanei*) were pooled into one of two groups: (i) genes that are broadly expressed in many or most CSNs or (ii) genes that are expressed in only specific CSNs. The pooled promoter regions were subsequently re-analyzed using MEME. The consensus motif illustration for the C-box was generated using WebLogo V2.8.2 (Crooks et al., 2004).

Results

C. elegans direct RFX-target genes show variable expression patterns within the CSN class of neurons

To begin exploring the dynamics of direct RFX-target gene regulation, the expression patterns of 41 known direct RFX-target gene promoters in *C. elegans* were compared (Fig. 1). To simplify these comparisons we first categorized each of the 60 CSNs of an adult hermaphrodite worm as being part of one of five defined anatomical regions, which were each selected to reflect the variation in expression commonly observed among various direct RFX-target genes (Fig. 1A). All but two direct RFX-target genes are expressed in CSNs of the amphids, while 32 of 41 such genes are expressed in both the amphid and tail CSNs (regions 1 and 2, respectively) (Fig. 1B). Direct RFX-target genes are more variably expressed in CSNs surrounding the anterior bulb (region 3) and posterior bulb (region 4) of the pharynx, and in the midbody (region 5) (Fig. 1B). We then selected seven direct RFX-target genes as a basis for our in-depth comparative analyses of gene expression pattern and regulation. Whereby, the genes *xbx-1/dync2li1*, *mks-3/mks3*, *bbs-8/bbs8* and *che-2/ift80* are broadly expressed in all CSNs (Fig. 2A and data not shown). In contrast, the expression of *dyf-2/ift144, xbx-9/novel* and *nhr-44/nhr* are restricted to a more targeted set of CSNs (Fig. 2B–D). As expected and as has been previously observed, for all seven genes the CSN-specific expression is significantly reduced or eliminated when expressed in a *daf-19/rfx* mutant background or when the respective X-box promoter motif sequences have been mutated and expressed in otherwise wild-type worms (Fig. 2; Supplemental Fig. S1A).

Direct RFX-target gene expression pattern specificity is independent of X-box proximity

To explore the possibility that direct RFX-target gene expression patterns are influenced by the position of the X-box promoter motif, the promoter regions of all 41 direct RFX-target genes sampled in this study were compared (Fig. 3A and B). First, the relative proximity of each X-box promoter motif from the translational start site of a direct RFX-target gene was determined (Fig. 1B). This data was then used to plot the number and position of X-box promoter motif sequences for genes that are broadly (Fig. 1B, nr 1–25) or more restrictedly (Fig. 1B, nr 26–41) expressed in CSNs

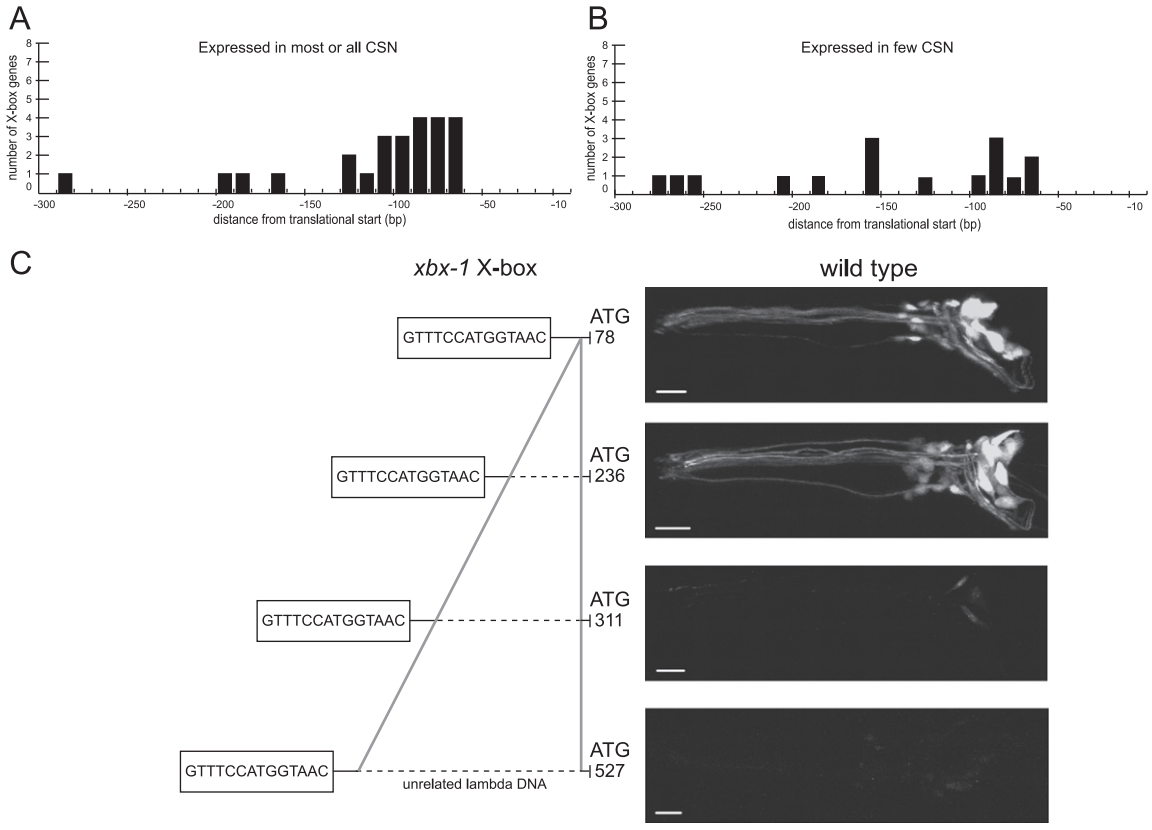


Fig. 3. Effect of X-box promoter motif position on the expression of direct RFX-target genes. (A, B) Comparisons of the number of X-box promoter motif sequences residing at specific distances from the translational start site (ATG) for all 41 direct RFX-target genes sampled. Separate bar graphs depict genes that are either (A) broadly expressed in most or all CSNs (cf. Fig. 1B, nr 1–25) or (B) show a more restricted expression pattern in specific CSNs (cf. Fig. 1B, nr 26–41). (C) Comparison of reporter gene expression (promoter-to-GFP fusion constructs) between unmodified and extended *xbx-1* gene promoter regions. The unmodified *xbx-1* promoter, which contains an X-box promoter motif at 78 bp from the translational start site, was compared to replicate promoter regions that had been extended to various lengths by inserting unrelated lambda-phage DNA. Schematic diagrams (left) depict the X-box promoter motif sequence composition, location, and when present, position of inserted DNA for each construct. Representative confocal projections show GFP expression patterns (right) in the head region from animals expressing each of the respective constructs. In all panels the scale bars depict 10 μm.

(Fig. 3A – B). These comparisons revealed that broadly expressed direct RFX-target genes typically contain an X-box promoter motif between 60 bp and 130 bp upstream of the translational start site (80% of 25). While direct RFX-target genes that are expressed in only a few CSNs can contain X-box sequences that also reside within 130 bp of the translational start site (50% of 16), the location of the X-box in these genes is not clustered toward the proximal end of this range. Interestingly, X-box promoter motif sequences were found to strictly reside within 300 bp upstream of the translational start site for all 41 direct RFX-target genes sampled in this study.

To experimentally examine the relationship between X-box promoter motif proximity and direct RFX-target gene expression pattern, unmodified and artificially elongated *xbx-1* promoter constructs were analyzed in vivo. Using the unmodified *xbx-1* promoter as a template, several GFP-fusion gene constructs were generated that each contained this promoter either unaltered or extended to various lengths away from the translational start site using unrelated Lambda phage DNA (Fig. 3C). Transgenic worms expressing unmodified and elongated constructs revealed the ubiquitous expression of the *xbx-1::gfp* fusion gene in all CSNs for constructs with an X-box distance of less than 300 bp. However, consistent with the observed X-box positions of all 41 direct RFX-target genes sampled in this study (Fig. 3A and B), an expression threshold appears to occur at approximately 300 bp upstream of the translational start site. Whereby, expression of a direct RFX-target gene is significantly reduced or eliminated when the X-box promoter motif resides at a distance greater than approximately 300 bp upstream of the translational start site of the respective gene. Thus, it appears that direct RFX-target gene expression pattern specificity is largely independent of X-box proximity provided the X-box promoter motif resides in relatively close proximity upstream of the translational start site (within 300 bp).

Direct RFX-target gene expression pattern specificity is largely independent of X-box sequence composition

Additional variables that may contribute to the expression pattern specificity of a particular direct RFX-target gene include either the X-box promoter motif sequence composition or the sequence composition of the promoter region flanking the X-box. To explore either of these possibilities, various direct RFX-target gene promoters were molecularly dissected, to determine the specific role that these different promoter components—both individually and in concert—have on the expression of a GFP reporter gene in vivo.

In the first set of experiments, we analyzed whether the X-box promoter motif sequence alone was sufficient to drive expression of a GFP reporter gene. Minimal promoters consisting of only the respective X-box promoter motif sequence from direct RFX-target genes with either a broad CSN expression pattern, including *xbx-1*, *mks-3*, *bbs-8* and *che-2* or a more restricted CSN expression pattern, including *dyf-2*, *xbx-9* and *nhr-44* were examined. Transgenic animals expressing any of these GFP-fusion constructs revealed either an undetectable level of reporter gene expression or a level of expression that was barely visible. Of note, this very weak visibility of expression was restricted to the amphid (region

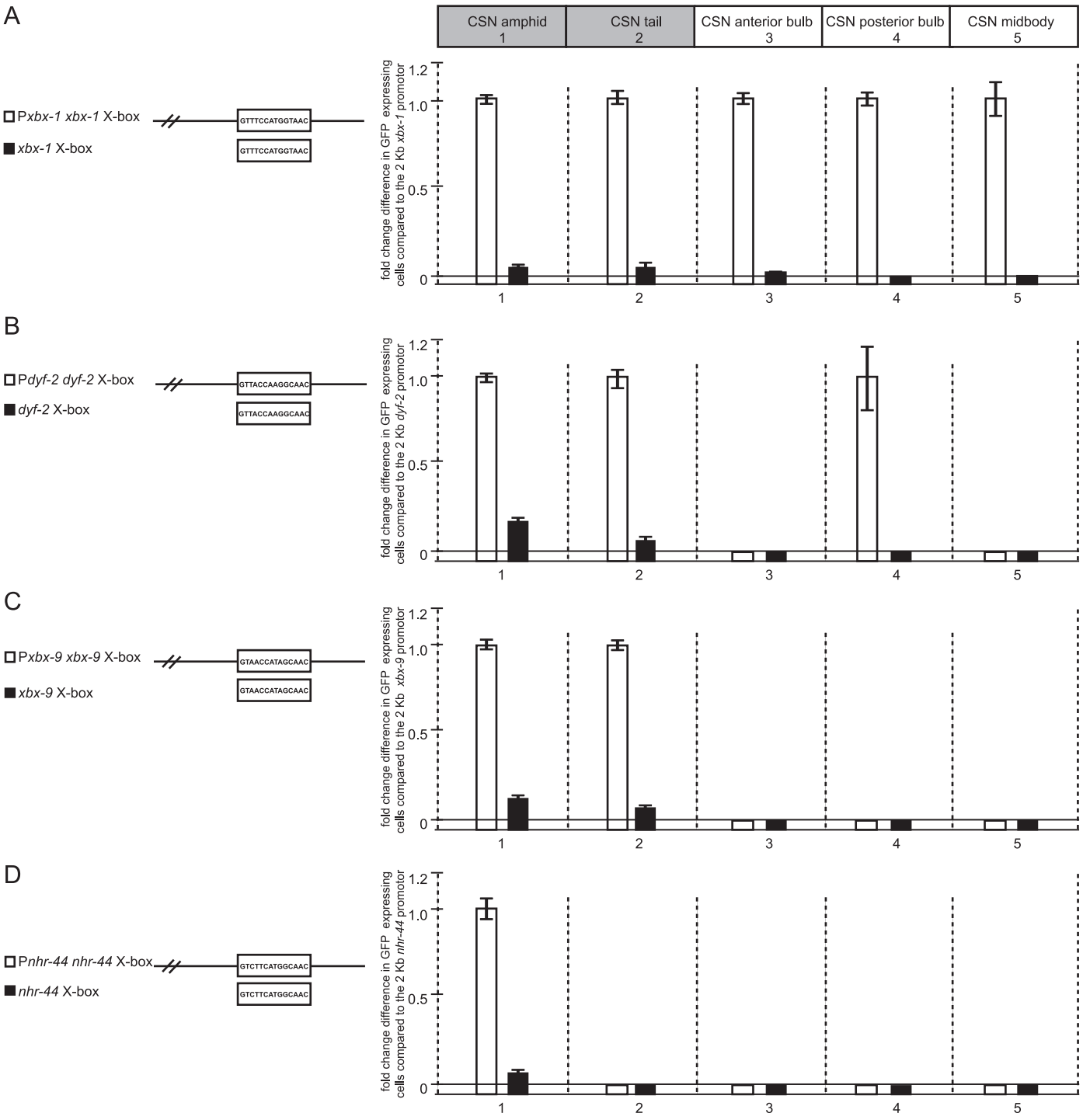


Fig. 4. Expression of direct RFX-target gene promoters with or without X-box flanking regions. (A–D) Comparisons of reporter gene expression (promoter-to-GFP fusion constructs) between unmodified promoter regions and minimal promoters consisting of only the respective X-box promoter motif sequence for each of the direct RFX-target gene sampled (*xbx-1*, *dyf-2*, *xbx-9*, *nhr-44*). Schematic diagrams (left) illustrate the promoter regions, whereby a double hash mark at the 5′ end of a construct indicates that an approximately 2 kb promoter region was used. Expression analyses (right) illustrate the specific reporter gene expression patterns observed in a particular anatomical region for each construct (cf. Fig. 1A and Materials and methods for details). Error bars represent the standard error of the mean.

1) and tail (region 2) CSNs (Fig. 4A–D; Supplemental Figs. S1B and S2A–C).

While the X-box promoter motif alone was unable to sufficiently drive expression of a reporter gene in *C. elegans* CSNs, it remained possible that—to varying degrees—both the X-box promoter motif sequence composition and the sequence composition of the X-box flanking regions affect the expression pattern specificity of a given direct RFX-target gene. To explore both of these possibilities, various reporter genes were constructed in which either the X-box or the X-box flanking regions of various different direct RFX-target genes were reciprocally exchanged. First, separate *xbx-1* reporter gene constructs were generated that each contained an X-box promoter motif modified to the corresponding motif sequence found in genes with more restricted expression patterns, including *dyf-2*, *xbx-9* or *nhr-44* (Fig. 5A–C). Transgenic worms expressing constructs that contained either the *dyf-2* or *xbx-9*

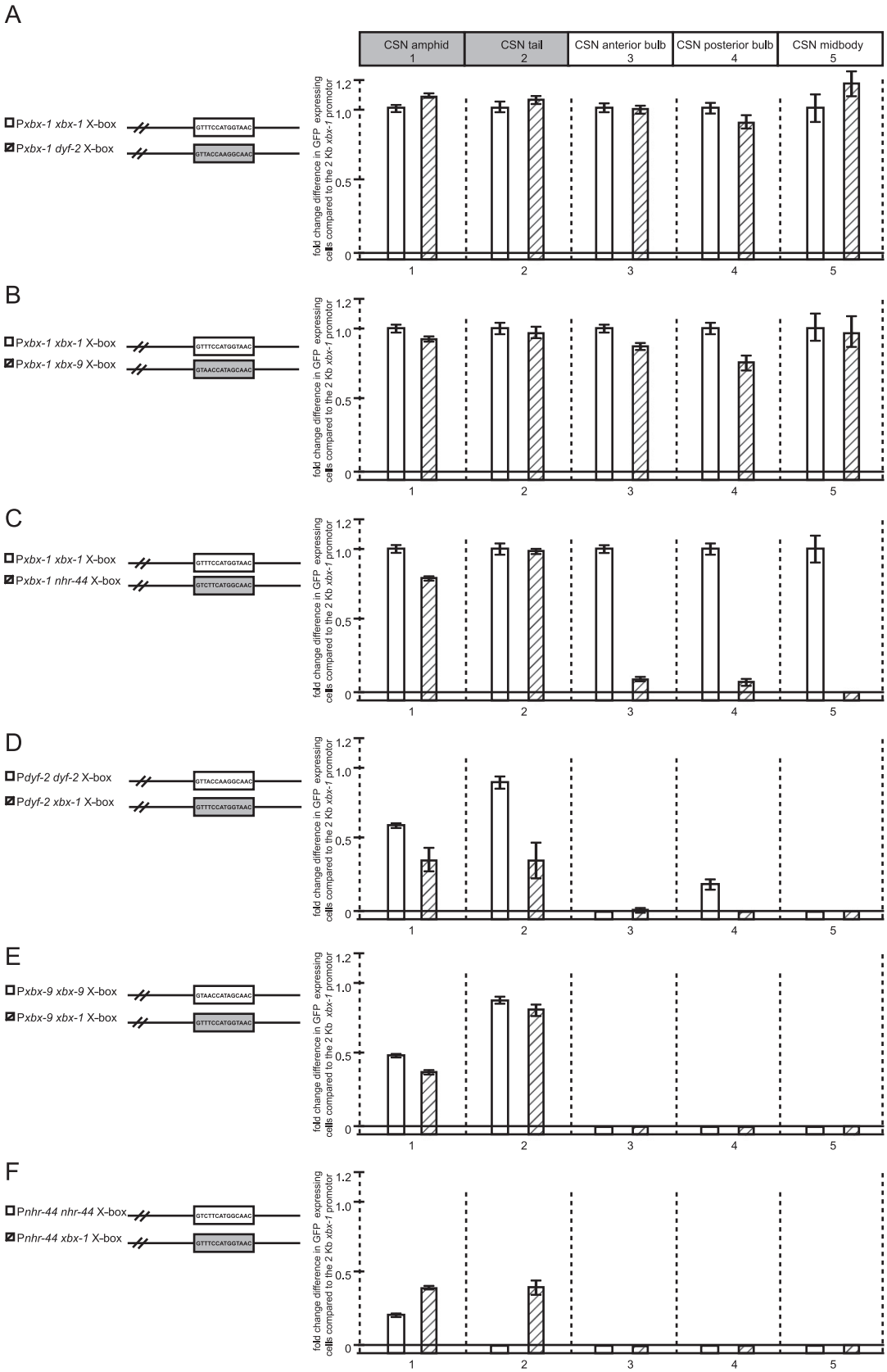


Fig. 5. Expression of direct RFX-target gene promoters with swapped-in X-box motif sequences. Reciprocal reporter gene comparisons of promoter-to-GFP fusion constructs expressing either (A–C) *xbx-1* promoter regions individually modified to contain a *dyf-2*, *xbx-9* or *nhr-44* X-box promoter motif sequence or (D–F) the promoter regions of *dyf-2*, *xbx-9* or *nhr-44* separately modified to contain the *xbx-1* X-box promoter motif sequence. Each schematic diagram (left) depicts the unmodified promoter region containing an endogenous X-box promoter motif sequence (top, white bar) and the promoter region modified to contain the corresponding X-box promoter motif sequence (bottom, dark bar). Double hash marks on the left of each promoter schematic illustrate that an approximately 2 kb promoter region was used for these analyses. Expression analyses (right) illustrate the specific reporter gene expression patterns observed in a particular anatomical region for each construct (cf. Fig. 1A and Materials and methods for details). Expression of GFP was normalized to the expression of the unmodified *xbx-1* X-box promoter construct. Error bars represent the standard error of the mean.

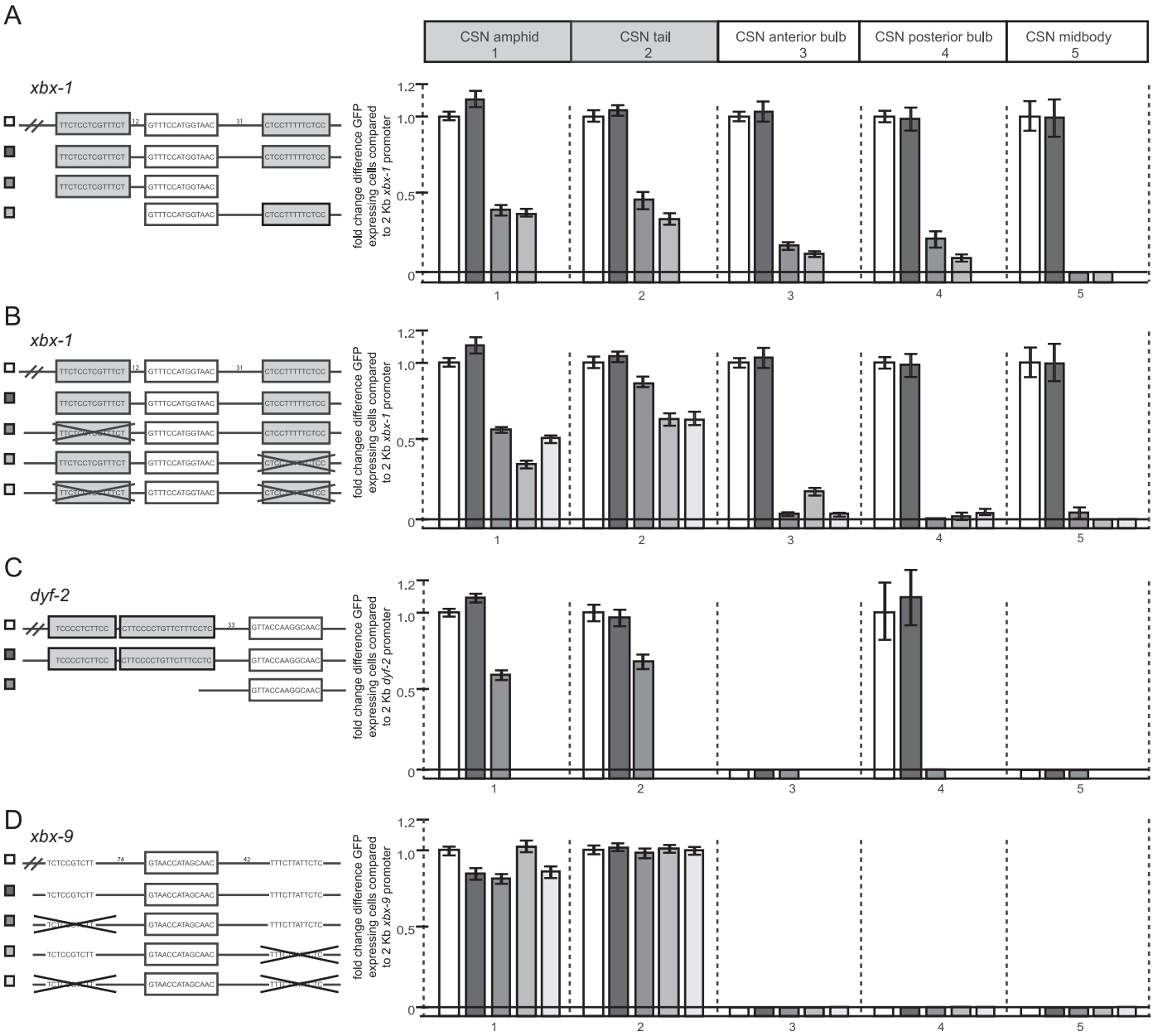


Fig. 6. In vivo dissection of direct RFX-target gene promoters. (A–D) Comparisons of reporter gene expression (promoter-to-GFP fusion constructs) between unmodified promoter regions and promoter regions that contain deletions or mutations. Schematic diagrams (left) illustrate key features in the respective unmodified and modified promoter regions from the direct RFX-target genes *xbx-1*, *dyf-2 and xbx-9*, including X-box promoter motif sequences (white boxes), C-box enhancer elements (gray boxes), and sequences that are very similar to C-box sequences (not boxed). A cross mark through a particular sequence indicates that it was replaced using a cross-means mutation. Deletions are represented by a gap. Double hash marks at the 5′ end of a construct indicate that an approximately 2 kb promoter region was used. Expression analyses (right) illustrate the specific reporter gene expression patterns observed in a particular anatomical region for each construct (cf. Fig. 1A and Materials and methods for details). Error bars represent the standard error of the mean.

X-box motif sequence embedded in the *xbx-1* promoter were both expressed in all CSNs, which are expression patterns that mirror the unmodified *xbx-1* constructs used for comparisons (Fig. 5A and B). Whereas, expression of either the unmodified *dyf-2* or *xbx-9* reporter gene constructs occurs in only a subset of CSNs (Fig. 5D and E). Transgenic worms expressing a construct containing the *nhr-44* X-box promoter motif sequence embedded in the *xbx-1* promoter resulted in a more restricted expression pattern than was observed for the unmodified *xbx-1* promoter region (Fig. 5C), but at the same time in a much broader expression pattern and at a higher expression level than that observed for the unmodified *nhr-44* reporter gene construct alone (Fig. 5F).

In a reciprocal set of experiments, the respective X-box motif sequences from the *dyf-2*, *xbx-9* and *nhr-44* gene promoters were each separately modified to the corresponding *xbx-1* X-box promoter motif sequence. Transgenic worms expressing reporter gene constructs containing the *xbx-1* X-box embedded in either the *dyf-2* or *xbx-9* promoter regions resulted in more restricted expression patterns that mirrored each of the respective unmodified *dyf-2* and *xbx-9* reporter gene constructs (Fig. 5D and E), which in both cases are more restricted expression patterns than those observed for each of the unmodified or reciprocally mutated *xbx-1* reporter gene constructs (compare Fig. 5D with A and E with B). Interestingly, transgenic worms expressing a reporter

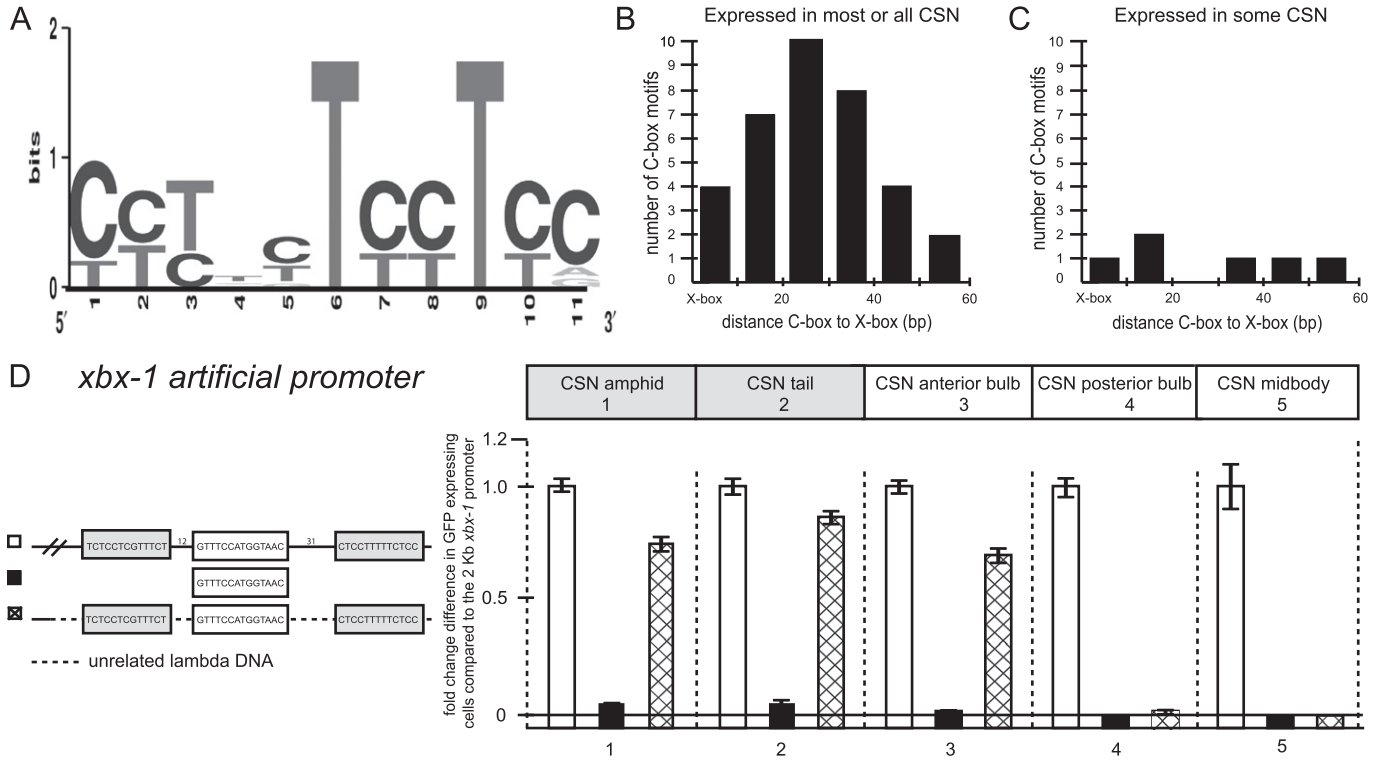


Fig. 7. Analyses of C-box enhancer elements in direct RFX-target genes. (A) Consensus sequence logo of the C-box enhancer element based on nine experimentally verified sequences. (B, C) Comparisons of the number and positions of C-box enhancer elements identified in direct RFX-target genes. The distance between each candidate or experimentally verified C-box (cf. Supplemental Fig. S3) in connection to its corresponding X-box promoter motif is depicted in separate bar graphs for direct RFX-target genes that are (B) broadly expressed in most or all CSNs (cf. Fig. 1B, nr 1–25) or that (C) show a more targeted or restricted expression in specific CSNs (cf. Fig. 1B, nr 26–41). (D) Expression analysis of an artificial promoter containing an X-box and C-boxes. Schematic diagrams (left) illustrate key features from the unmodified promoter region of the direct RFX-target gene *xbx-1*, including X-box promoter motif sequences (white boxes) and C-box enhancer elements (gray boxes). Unrelated lambda-phage DNA is depicted as dashed lines. Deletions are represented by a gap. Double hash marks at the 5′ end of a construct indicate that an approximately 2 kb promoter region was used. Expression analyses (right) illustrate the specific reporter gene expression patterns observed in a particular anatomical region for each construct (cf. Fig. 1A and Materials and methods for details). Error bars represent the standard error of the mean.

gene construct containing the *xbx-1* X-box promoter motif sequence embedded in the *nhr-44* promoter region resulted (i) in an increase of expression of this reporter gene as compared to the expression pattern observed for the unmodified *nhr-44* reporter gene construct (Fig. 5F), and (ii) in a highly decreased expression as compared to the unmodified or reciprocally mutated *xbx-1* reporter gene constructs (compare Fig. 5F with C). Thus, while the X-box sequence composition has a minor affect on the level and specificity of direct RFX-target gene expression patterns in CSNs, the above data indicate that X-box sequence composition is likely not the primary factor regulating direct RFX-target gene expression pattern specificity.

Cumulatively, the minimal X-box promoter motif expression analysis and reciprocal promoter motif comparisons indicate that while the X-box promoter motif is required for the efficient expression of direct RFX-target genes, expression pattern specificity is largely influenced by sequences flanking the X-box motif. Thus, we hypothesized that direct RFX-target gene promoters may contain additional regulatory elements that influence both the level and specificity of their gene expression patterns.

Identification of a novel enhancer element, the C-box, in direct RFX-target gene promoters

It has been previously demonstrated that the human RFX5 recognition site contains an additional enhancer element within 18 bp of the 3′ end of the X-box promoter motif of a direct target gene of human RFX5 (Boss and Jensen, 2003; Masternak et al., 2000). However, other than the conserved composition of X-box promoter motif sequences, no additional sequence homology has been observed when comparing human RFX5 enhancer elements and other sequences residing in *C. elegans* direct RFX-target gene promoters. To explore the possibility that a cryptic or divergent enhancer element resides in close proximity to the X-box promoter motifs of direct RFX-target genes in *C. elegans*, the expression patterns of *xbx-1* and *mks-3* X-box promoter motifs elongated by exactly 18 bp on both the 5′ and 3′ flanking regions were examined. These analyses revealed that constructs containing only an X-box motif and short flanking sequences were largely unable to drive expression of a reporter gene in *C. elegans* (Supplemental Fig. S1). We conclude that the expression of a direct RFX-target gene likely depends on *cis*-regulatory sequences that extend beyond the immediate X-box flanking region.

To provide a preliminary indication of the overall size and complexity of a direct RFX-target gene promoter in *C. elegans*, reporter genes expressing unmodified and truncated promoters were compared in vivo for select direct RFX-target genes with either a broad CSN expression pattern, including *xbx-1*, *mks-3*, *bbs-8* and *che-2*, or a more restricted CSN expression pattern, including *dyf-2* and *xbx-9*. These analyses revealed that promoters of 400 bp in length or less consistently retained an expression pattern similar to each of the respective 2 kb unmodified promoter constructs (Fig. 6A–D; Supplemental Fig. S2A–C). Each of these 400 bp direct RFX-target gene promoter sequences were then compared using the MEME algorithm, a bioinformatics-based method used to identify conserved regulatory elements in the promoters of similarly regulated genes (cf. Materials and methods; Bailey and Elkan, 1994). Other than identifying known

X-box motif sequences from each of these direct RFX-target gene promoters, these initial comparisons were unable to predict with a high degree of certainty the presence of additional candidate regulatory sequence motifs.

We next exploited the observation that enhancer motifs are often conserved in different *Caenorhabditis* species (Bigelow et al., 2004) by applying a comparative genomics-based approach in conjunction with a MEME search effort. For each of the 41 *C. elegans* direct RFX-target genes sampled in this study, the corresponding orthologous promoter regions were extracted from the *Caenorhabditis briggsae* and *Caenorhabditis remanei* genomes. When pooling all of the promoters together and analyzing them for the presence of conserved elements using MEME, with the exception of X-box promoter motif sequences, we were unable to identify additional conserved sequence motifs. Next, each of the respective orthologous, direct RFX-target gene promoter regions from *C. elegans*, *C. briggsae* and *C. remanei* were clustered according to the expression patterns observed for the *C. elegans* direct RFX-target gene. Direct RFX-target gene promoters and their orthologous counterparts were pooled into two groups: (i) gene promoters that are broadly (cf. Fig. 1B, nr 1–25) and (ii) that are more restrictedly (cf. Fig. 1B, nr 26–41) expressed in *C. elegans* CSNs. Subsequent MEME analyses of these pooled, direct RFX-target gene promoters, extracted from the genomes of three different *Caenorhabditis* species revealed an enrichment of a specific cytosine (C) and thymidine (T) rich sequences in the promoter regions of direct RFX-target genes that are broadly expressed in CSNs. When present, these C/T rich sequences were conserved in the respective orthologous gene promoters of *C. elegans*, *C. briggsae* and *C. remanei*. No C/T rich sequences or any other motifs were identified with MEME in the pooled direct RFX-target genes with more restricted CSN expression patterns. C/T rich sequences thus seem to be associated with broad expression in CSNs.

To determine experimentally if the C/T rich sequences identified in many broadly expressed, direct RFX target gene promoters enhance the expression of these promoters in vivo, we examined the expression patterns of various reporter gene constructs in which these C/T rich sequences were either present or absent (Fig. 6A–D; Supplemental Fig. S2A–C). We found that the *xbx-1* gene contained a single C/T rich sequence in close proximity (<31 bp) to both the 5′ and 3′ ends of its X-box promoter motif. Reporter gene constructs containing deletions of either one or both of these C/T rich sequences resulted in a partial (amphid and tail CSNs; anatomical regions 1 and 2) or complete (all other CSNs; anatomical regions 3, 4 and 5) reduction in CSN expression (Fig. 6A). Similarly, *xbx-1* reporter gene constructs containing mutations in either one or both of these C/T rich sequences mirrored the expression patterns observed for reporter gene constructs containing singly or doubly deleted C/T rich sequences, respectively (compare Fig. 6A with B).

As was observed for the *xbx-1* promoter, a series of deletion and mutation constructs eliminating the C/T rich sequences identified in the *mks-3*, *bbs-8*, *che-2* and *dyf-2* promoters resulted in similar pronounced decreases in reporter gene expression patterns for all of the respective direct RFX-target genes sampled (Fig. 6C; Supplemental Fig. S2A–C). In contrast, mutations in two C/T rich sequences found in relatively close proximity to the *xbx-9* X-box promoter motif did not alter the expression of reporter gene constructs. Thus, the *xbx−9* C/T rich sequences likely represent false positive hits (Fig. 6D). In addition, no C/T rich sequences could be identified in close proximity to the *nhr-44* X-box promoter motif. Interestingly, both the *xbx-9* and *nhr-44* genes have significantly more restricted CSN expression patterns than the other direct RFX-target genes examined.

These data collectively demonstrate that direct RFX-target gene promoters contain sequence elements in addition to X-box promoter motifs that are required for the efficient and specific expression of ciliary genes in the various *C. elegans* neurons of the CSN class. One such sequence element was identified, which was subsequently found to enhance the broad CSN expression of direct RFX-target genes in vivo. We hereby term this C/T rich sequence the "C-box" enhancer element. Based on the identification of nine experimentally verified C-box enhancer elements—and three false positive hits—a preliminary C-box consensus sequence was determined (Fig. 7A). All experimentally verified C-box sequence elements are comprised of either a continuous stretch of 9 C/T nucleotides or an 11 nucleotides C/T-rich stretch with—sometimes—a central guanosine base. All C-box enhancer elements were found to reside within 60 bp from the respective X-box promoter motif of each of the respective direct RFX-target genes examined (Fig. 7A–C; Supplemental Fig. S3).

Using the above criteria as a guide, 25 candidates and experimentally verified C-box enhancer elements were identified in the promoters of all 41 direct RFX-target genes sampled in this study (Supplemental Fig. S3). The number and location of C-box enhancer elements residing at specific distances with regard to their respective X-box promoter motif were then determined for direct RFX-target genes that are either broadly (Fig. 7B) or more restrictedly (Fig. 7C) expressed in CSNs. We found that C-box enhancer elements are significantly more abundant in direct RFX-target genes that are broadly expressed in many or all CSNs (84%) than in direct RFX-target genes that have a more restricted CSN expression (19%) (compare Fig. 7B with C; Supplemental Fig. S3). Thus, C-box enhancer elements may primarily facilitate the broad CSN expression patterns of direct RFX-target genes.

Expression pattern analyses of artificial xbx-1 gene promoters

To further reveal how the C-box enhancer element specifically affects direct RFX-target gene expression, a reporter gene with an artificial *xbx-1* promoter was constructed. This artificial reporter gene construct exclusively preserved the composition and spacing of the respective X-box promoter motif and C-box enhancer elements of the endogenous *xbx-1* promoter (Fig. 7D, left). Expression of this artificial reporter gene construct was then compared to transgenic worms expressing either an unmodified 2 kb or an *xbx-1* X-box minimal reporter gene construct (Fig. 7D, right). In contrast to the minimal construct that contained only the X-box promoter motif, a pronounced CSN-specific expression pattern was observed in transgenic animals carrying the artificial *xbx-1* reporter gene construct. However, this artificial reporter gene construct was not able to fully restore expression specificity or abundance to levels as broad or as high as those observed in the unmodified *xbx-1* reporter gene construct. Interestingly, CSN expression in the amphids, tail, and surrounding the anterior bulb of the pharynx (anatomical regions 1, 2 and 3) was restored to between 68% and 85% of the levels seen with the unmodified promoter constructs (Fig. 7D; Supplemental Table S1). While, CSN expression surrounding the posterior bulb of the pharynx and in the midbody (anatomical regions 4 and 5) was not restored. These data verify that the C-box enhancer element alone can enhance expression of direct RFX-target genes in vivo. However, collectively these data also indicate that there likely remain additional elements that facilitate expression of direct RFX-target genes in *C. elegans* CSNs (Figs. 6A–D and 7D; Supplemental Figs. S1A and B and S2A–C).

Discussion

In this study we demonstrate that in *C. elegans* neither the position nor the exact sequence composition of the X-box promoter motif is the primary factor determining the specificity of direct RFX-target gene expression patterns in the various neurons of the CSN class. Whereby, direct RFX-target genes are often ciliary genes. Instead,

sequences in adjacent *cis*-regulatory regions work in concert with the X-box promoter motif to regulate the expression of direct RFX-target genes in CSNs. One specific sequence, termed the C-box enhancer element, was identified in the most broadly and abundantly expressed direct RFX-target genes in *C. elegans*. These data suggest that direct RFX-target gene specificity is dependent on the X-box promoter motif and its flanking sequences. Recently, the modENCODE Consortium has revealed the genome-wide binding sites for more than 20 TFs in *C. elegans*, not including DAF-19/RFX (Niu et al., 2011). Another study determined which specific TFs bind a similar number of *cis*-regulatory regions (Gerstein et al., 2010). Both data sets collectively confirm that most TFs bind near the translational start site of their respective direct target genes. The results from our study are largely complementary to the modENCODE efforts in that we identify the C-box, a novel enhancer element for a relatively large set of genes, which all share a common mode of regulation, namely being regulated by DAF-19/RFX. We identify with a much higher degree of resolution the specific sites that are likely bound by an as-of-yet unidentified transcriptional regulator. Future work aims to identify this putative transcriptional regulator using methodology similar to those used by the modENCODE Consortium.

The sequence composition of the C-box enhancer element identified in this study is very similar to the sequence composition of the N1-box, which is a *cis*-regulatory element that was previously implicated in the expression of many neuronal genes in *C. elegans* (Ruvinsky et al., 2007). Comparisons between C-box and N1-box enhancer element consensus motifs reveal that both elements are composed primarily of C/T-rich sequences. However, N1-box *cis*-regulatory elements were found to contain a more stringent composition of nucleotides, including the presence of guanosine nucleotides at positions 3 and 5 of the N1-box consensus motif. In contrast to the C-box, which was found here in genes highly and exclusively expressed in CSNs, N1-box genes are typically expressed pan-neuronally, that is, in many different types of neurons, where the CSNs represent only one class. Subsequent experimental analyses will be required to determine whether or not the C-box enhancer element and the N1-box *cis*-regulatory element share a common mode of regulation. One possibility that cannot be excluded is that the C-box may represent a degenerate N1-box. This possibility appears to be in agreement with the observation that those direct RFX-target genes, which are expressed most highly and broadly within the CSN class of neurons, additionally contain C-box enhancer elements.

The X-box promoter motif, bound and regulated by RFX TFs, is a highly conserved *cis*-regulatory element found in the transcriptional network of many ciliary genes of animals (Chu et al., 2010; Piasecki et al., 2010). Unlike humans, which contain at least 7 RFX TFs (Aftab et al., 2008), only a single RFX TF, DAF-19, controls the expression of direct RFX-target genes in various CSNs in *C. elegans* (Swoboda et al., 2000). In contrast, C-box enhancer elements appear to be evolutionarily conserved exclusively in nematodes, and even within this clade only to varying degrees. Using computational-based approaches similar to the ones described in this study, we were unable to identify sequences that resemble C-box enhancer elements in a variety of human ciliary gene promoters, including *ift52*/*osm-6*, *ift172*/*osm-1*, *bbs1*/*bbs-1*, *bbs5*/*bbs-5* and *dync2li1*/*xbx-1*.

Interestingly, several additional *cis*-regulatory elements were previously identified in human RFX5 target gene promoters, which are exclusively found in MHC class II genes. In addition to the X-box promoter motif, RFX5 target genes contain W/S, X1 (X-box), X2 and Y *cis*-regulatory elements (Boss and Jensen, 2003; Masternak et al., 2000). Similar to the C-box enhancer elements in *C. elegans*, the *cis*-regulatory elements found in human RFX5 target gene promoters always reside in close proximity (<60 bp) to their respective X-box promoter motifs. Further, the number of additional enhancer elements identified for RFX5 target genes in humans is consistent with the idea that additional sequences likely contribute to the specific expression patterns of direct RFX-target genes in *C. elegans*.

Individual transcription factors are often functionally conserved over very large phylogenetic distances, as is the case for the RFX TF family (Chu et al., 2010; Piasecki et al., 2010). In contrast, *cis*-regulatory sequence evolution is a highly dynamic process, typically occurring at much higher rates than observed for gene coding regions (Ruvinsky and Ruvkun, 2003). In the case of the RFX TF family the X-box promoter motif sequence is conserved across animal species. However, additional *cis*-regulatory elements in the X-box flanking regions may be rapidly diverging. Even when comparing the promoters of closely related organisms, such as nematodes, the expression of neuronal genes has been found to occur through both conserved (Ruvinsky et al., 2007) and divergent (Nokes et al., 2009) mechanisms. Thus, it seems likely that the expression of direct RFX-target genes depends on a combination of promoter sequence elements that are evolutionarily conserved to varying degrees, especially when comparing between orthologous gene promoters of highly divergent animals.

Acknowledgments

This study was performed in parts at the microscope facilities at the Departments of Cell and Molecular Biology, and Biosciences and Nutrition, of the Karolinska Institute; supported by Grants from the Knut and Alice Wallenberg Foundation, the Swedish Research Council and the Center for Biosciences. JB was supported by a fellowship from the Carl Tryggers Stiftelse. BPP was supported by a Grant from the Lars Hiertas Minne Foundation and by a Fulbright Fellowship. Work in the laboratory of PS was supported by the Swedish Foundation for Strategic Research, the Swedish Research Council, the Marcus Borgström Foundation, and the NordForsk Nordic networks for *C. elegans* and cilia/centrosomes research. Elizabeth De Stasio critically evaluated the manuscript.

Appendix A. Supplementary materials

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.ydbio.2012.05.033.

References

Aftab, S., Semenec, L., Chu, J.S., Chen, N., 2008. Identification and characterization of novel human tissue-specific RFX transcription factors. BMC Evol. Biol. 8, 226.

Bailey, T.L., Elkan, C., 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc. Int. Conf. Intell. Syst. Mol. Biol. 2, 28–36.

Berbari, N.F., O'Connor, A.K., Haycraft, C.J., Yoder, B.K., 2009. The primary cilium as a complex signaling center. Curr. Biol. 19, R526–535.

Bigelow, H.R., Wenick, A.S., Wong, A., Hobert, O., 2004. CisOrtho: a program pipeline for genome-wide identification of transcription factor target genes using phylogenetic footprinting. BMC Bioinformatics 5, 27.

Blacque, O.E., Perens, E.A., Boroevich, K.A., Inglis, P.N., Li, C., Warner, A., Khattra, J., Holt, R.A., Ou, G., Mah, A.K., McKay, S.J., Huang, P., Swoboda, P., Jones, S.J., Marra, M.A., Baillie, D.L., Moerman, D.G., Shaham, S., Leroux, M.R., 2005. Functional genomics of the cilium, a sensory organelle. Curr. Biol. 15, 935–941.

Boss, J.M., Jensen, P.E., 2003. Transcriptional regulation of the MHC class II antigen presentation pathway. Curr. Opin. Immunol. 15, 105–111.

Brenner, S., 1974. The genetics of *Caenorhabditis elegans*. Genetics 77, 71–94.

Burghoorn, J., Dekkers, M.P., Rademakers, S., de Jong, T., Willemsen, R., Jansen, G., 2007. Mutation of the MAP kinase *dyf-5* affects docking and undocking of kinesin-2 motors and reduces their speed in the cilia of *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. USA 104, 7157–7162.

Chen, N., Mah, A., Blacque, O.E., Chu, J., Phgora, K., Bakhoum, M.W., Newbury, C.R., Khattra, J., Chan, S., Go, A., Efimenko, E., Johnsen, R., Phirke, P., Swoboda, P., Marra, M., Moerman, D.G., Leroux, M.R., Baillie, D.L., Stein, L.D., 2006. Identification of ciliary and ciliopathy genes in *Caenorhabditis elegans* through comparative genomics. Genome Biol. 7, R126.

Chu, J.S., Baillie, D.L., Chen, N., 2010. Convergent evolution of RFX transcription factors and ciliary genes predated the origin of metazoans. BMC Evol. Biol. 10, 130.

Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E., 2004. WebLogo: a sequence logo generator. Genome Res. 14, 1188–1190.

Efimenko, E., Blacque, O.E., Ou, G., Haycraft, C.J., Yoder, B.K., Scholey, J.M., Leroux, M.R., Swoboda, P., 2006. *Caenorhabditis elegans dyf-2*, an orthologue of human WDR19, is a component of the intraflagellar transport machinery in sensory cilia. Mol. Biol. Cell 17, 4801–4811.

Efimenko, E., Bubb, K., Mak, H.Y., Holzman, T., Leroux, M.R., Ruvkun, G., Thomas, J.H., Swoboda, P., 2005. Analysis of xbx genes in *C. elegans*. Development 132, 1923–1934.

El Zein, L., Ait-Lounis, A., Morle, L., Thomas, J., Chhin, B., Spassky, N., Reith, W., Durand, B., 2009. RFX3 governs growth and beating efficiency of motile cilia in mouse and controls the expression of genes involved in human ciliopathies. J. Cell Sci. 122, 3180–3189.

Emery, P., Strubin, M., Hofmann, K., Bucher, P., Mach, B., Reith, W., 1996. A consensus motif in the RFX DNA binding domain and binding domain mutants with altered specificity. Mol. Cell Biol. 16, 4486–4494.

Etchberger, J.F., Lorch, A., Sleumer, M.C., Zapf, R., Jones, S.J., Marra, M.A., Holt, R.A., Moerman, D.G., Hobert, O., 2007. The molecular signature and cis-regulatory architecture of a *C. elegans* gustatory neuron. Genes Dev. 21, 1653–1674.

Gerstein, M.B., Lu, Z.J., Van Nostrand, E.L., Cheng, C., Arshinoff, B.I., Liu, T., Yip, K.Y., Robilotto, R., Rechtsteiner, A., Ikegami, K., Alves, P., Chateigner, A., Perry, M., Morris, M., Auerbach, R.K., Feng, X., Leng, J., Vielle, A., Niu, W., Rhrissorrakrai, K., Agarwal, A., Alexander, R.P., Barber, G., Brdlik, C.M., Brennan, J., Brouillet, J.J., Carr, A., Cheung, M.S., Clawson, H., Contrino, S., Dannenberg, L.O., Dernburg, A.F., Desai, A., Dick, L., Dose, A.C., Du, J., Egelhofer, T., Ercan, S., Euskirchen, G., Ewing, B., Feingold, E.A., Gassmann, R., Good, P.J., Green, P., Gullier, F., Gutwein, M., Guyer, M.S., Habegger, L., Han, T., Henikoff, J.G., Henz, S.R., Hinrichs, A., Holster, H., Hyman, T., Iniguez, A.L., Janette, J., Jensen, M., Kato, M., Kent, W.J., Kephart, E., Khivansara, V., Khurana, E., Kim, J.K., Kolasinska-Zwierz, P., Lai, E.C., Latorre, I., Leahey, A., Lewis, S., Lloyd, P., Lochovsky, L., Lowdon, R.F., Lubling, Y., Lyne, R., MacCoss, M., Mackowiak, S.D., Mangone, M., McKay, S., Mecenas, D., Merrihew, G., Miller 3rd, D.M., Muroyama, A., Murray, J.I., Ooi, S.L., Pham, H., Phippen, T., Preston, E.A., Rajewsky, N., Ratsch, G., Rosenbaum, H., Rozowsky, J., Rutherford, K., Ruzanov, P., Sarov, M., Sasidharan, R., Sboner, A., Scheid, P., Segal, E., Shin, H., Shou, C., Slack, F.J., Slightam, C., Smith, R., Spencer, W.C., Stinson, E.O., Taing, S., Takasaki, T., Vafeados, D., Voronina, K., Wang, G., Washington, N.L., Whittle, C.M., Wu, B., Yan, K.K., Zeller, G., Zha, Z., Zhong, M., Zhou, X., Mod, E.C., Ahringer, J., Strome, S., Gunsalus, K.C., Micklem, G., Liu, X.S., Reinke, V., Kim, S.K., Hillier, L.W., Henikoff, S., Piano, F., Snyder, M., Stein, L., Lieb, J.D., Waterston, R.H., 2010. Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. Science 330, 1775–1787.

Hobert, O., 2002. PCR fusion-based approach to create reporter gene constructs for expression analysis in transgenic *C. elegans*. Biotechniques 32, 728–730.

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

Laurencon, A., Dubruille, R., Efimenko, E., Grenier, G., Bissett, R., Cortier, E., Rolland, V., Swoboda, P., Durand, B., 2007. Identification of novel regulatory factor X (RFX) target genes by comparative genomics in *Drosophila* species. Genome Biol. 8, R195.

Li, C., Inglis, P.N., Leitch, C.C., Efimenko, E., Zaghloul, N.A., Mok, C.A., Davis, E.E., Bialas, N.J., Healey, M.P., Heon, E., Zhen, M., Swoboda, P., Katsanis, N., Leroux, M.R., 2008. An essential role for DYF-11/MIP-T3 in assembling functional intraflagellar transport complexes. PLoS Genet. 4, e1000044.

Masternak, K., Muhlethaler-Mottet, A., Villard, J., Peretti, M., Reith, W., 2000. Molecular genetics of the Bare lymphocyte syndrome. Rev. Immunogenet. 2, 267–282.

McCarroll, S.A., Li, H., Bargmann, C.I., 2005. Identification of transcriptional regulatory elements in chemosensory receptor genes by probabilistic segmentation. Curr. Biol. 15, 347–352.

Mello, C., Fire, A., 1995. DNA transformation. Methods Cell Biol. 48, 451–482.

Mello, C.C., Kramer, J.M., Stinchcomb, D., Ambros, V., 1991. Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. EMBO J. 10, 3959–3970.

Murayama, T., Toh, Y., Ohshima, Y., Koga, M., 2005. The *dyf-3* gene encodes a novel protein required for sensory cilium formation in *Caenorhabditis elegans*. J. Mol. Biol. 346, 677–687.

Niu, W., Lu, Z.J., Zhong, M., Sarov, M., Murray, J.I., Brdlik, C.M., Janette, J., Chen, C., Alves, P., Preston, E., Slightham, C., Jiang, L., Hyman, A.A., Kim, S.K., Waterston, R.H., Gerstein, M., Snyder, M., Reinke, V, 2011. Diverse transcription factor binding features revealed by genome-wide ChIP-seq in *C. elegans*. Genome Res. 21, 245–254.

Nokes, E.B., Van Der Linden, A.M., Winslow, C., Mukhopadhyay, S., Ma, K., Sengupta, P., 2009. Cis-regulatory mechanisms of gene expression in an olfactory neuron type in *Caenorhabditis elegans*. Dev. Dyn. 238, 3080–3092.

Perkins, L.A., Hedgecock, E.M., Thomson, J.N., Culotti, J.G., 1986. Mutant sensory cilia in the nematode *Caenorhabditis elegans*. Dev. Biol. 117, 456–487.

Phirke, P., Efimenko, E., Mohan, S., Burghoorn, J., Crona, F., Bakhoum, M.W., Trieb, M., Schuske, K., Jorgensen, E.M., Piasecki, B.P., Leroux, M.R., Swoboda, P., 2011. Transcriptional profiling of *C. elegans* DAF-19 uncovers a ciliary base-associated protein and a CDK/CCRK/LF2p-related kinase required for intraflagellar transport. Dev. Biol. 357, 235–247.

Piasecki, B.P., Burghoorn, J., Swoboda, P., 2010. Regulatory factor X (RFX)-mediated transcriptional rewiring of ciliary genes in animals. Proc. Natl. Acad. Sci. USA 107, 12969–12974.

Prud'homme, B., Gompel, N., Carroll, S.B., 2007. Emerging principles of regulatory evolution. Proc. Natl. Acad. Sci. USA 104 (Suppl. 1), 8605–8612.

Rosenbaum, J.L., Witman, G.B., 2002. Intraflagellar transport. Nat. Rev. Mol. Cell Biol. 3, 813–825.

Ruvinsky, I., Ohler, U., Burge, C.B., Ruvkun, G., 2007. Detection of broadly expressed neuronal genes in *C. elegans*. Dev. Biol. 302, 617–626.

Ruvinsky, I., Ruvkun, G., 2003. Functional tests of enhancer conservation between distantly related species. Development 130, 5133–5142.

Schafer, J.C., Haycraft, C.J., Thomas, J.H., Yoder, B.K., Swoboda, P., 2003. XBX-1 encodes a dynein light intermediate chain required for retrograde intraflagellar transport and cilia assembly in *Caenorhabditis elegans*. Mol. Biol. Cell 14, 2057–2070.

Senti, G., Ezcurra, M., Lobner, J., Schafer, W.R., Swoboda, P., 2009. Worms with a single functional sensory cilium generate proper neuron-specific behavioral output. Genetics 183, 595–605.

Senti, G., Swoboda, P., 2008. Distinct isoforms of the RFX transcription factor DAF-19 regulate ciliogenesis and maintenance of synaptic activity. Mol. Biol. Cell 19, 5517–5528.

Silverman, M.A., Leroux, M.R., 2009. Intraflagellar transport and the generation of dynamic, structurally and functionally diverse cilia. Trends Cell Biol. 19, 306–316.

Swoboda, P., Adler, H.T., Thomas, J.H., 2000. The RFX-type transcription factor DAF-19 regulates sensory neuron cilium formation in *C. elegans*. Mol. Cell 5, 411–421.

Tuch, B.B., Li, H., Johnson, A.D., 2008. Evolution of eukaryotic transcription circuits. Science 319, 1797–1799.

Wang, J., Schwartz, H.T., Barr, M.M., 2010. Functional specialization of sensory cilia by an RFX transcription factor isoform. Genetics 186, 1295–1307.

Williams, C.L., Winkelbauer, M.E., Schafer, J.C., Michaud, E.J., Yoder, B.K., 2008. Functional redundancy of the B9 proteins and nephrocystins in *Caenorhabditis elegans* ciliogenesis. Mol. Biol. Cell 19, 2154–2168.

Winkelbauer, M.E., Schafer, J.C., Haycraft, C.J., Swoboda, P., Yoder, B.K., 2005. The *C. elegans* homologs of nephrocystin-1 and nephrocystin-4 are cilia transition zone proteins involved in chemosensory perception. J. Cell Sci. 118, 5575–5587.

Wray, G.A., 2007. The evolutionary significance of cis-regulatory mutations. Nat. Rev. Genet. 8, 206–216.