Analysis of *xbx* genes in *C. elegans*

Evgeni Efimenko[1], Kerry Bubb[2], Ho Yi Mak[4], Ted Holzman[3], Michel R. Leroux[5], Gary Ruvkun[4], James H. Thomas[2] and Peter Swoboda[1,*]

[1]Karolinska Institute, Department of Biosciences and Södertörn University College, Section of Natural Sciences, S-14189 Huddinge, Sweden
[2]Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA
[3]Department of Microbiology, University of Washington, Seattle, WA 98195, USA
[4]Department of Molecular Biology, Massachusetts General Hospital and Department of Genetics, Harvard Medical School, Boston, MA 02114, USA
[5]Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada
*Author for correspondence (e-mail: peter.swoboda@biosci.ki.se)

Accepted 10 February 2005

Development 132, 1923-1934
Published by The Company of Biologists 2005
doi:10.1242/dev.01775

Summary

Cilia and flagella are widespread eukaryotic subcellular components that are conserved from green algae to mammals. In different organisms they function in cell motility, movement of extracellular fluids and sensory reception. While the function and structural description of cilia and flagella are well established, there are many questions that remain unanswered. In particular, very little is known about the developmental mechanisms by which cilia are generated and shaped and how their components are assembled into functional machineries. To find genes involved in cilia development we used as a search tool a promoter motif, the X-box, which participates in the regulation of certain ciliary genes in the nematode *Caenorhabditis elegans*.

By using a genome search approach for X-box promoter motif-containing genes (*xbx* genes) we identified a list of about 750 *xbx* genes (candidates). This list comprises some already known ciliary genes as well as new genes, many of which we hypothesize to be important for cilium structure and function. We derived a *C. elegans* X-box consensus sequence by in vivo expression analysis. We found that *xbx* gene expression patterns were dependent on particular X-box nucleotide compositions and the distance from the respective gene start. We propose a model where DAF-19, the RFX-type transcription factor binding to the X-box, is responsible for the development of a ciliary module in *C. elegans*, which includes genes for cilium structure, transport machinery, receptors and other factors.

Key words: X-box, DAF-19, Ciliary genes, *C. elegans*

Introduction

Sensory behavior in higher animals depends on the correct recognition and processing of signals from the environment or from within the organism, and then on the appropriate reaction to those signals. The first step in this chain of events, signal recognition, is mediated by specialized cells of the nervous system, the sensory neurons. They create functional compartments for the localization and exposure of the signal reception and transduction machineries. There is considerable diversity in these compartments. Prominent examples are ciliated endings at the tip of dendrites of sensory neurons. Specialized mammalian sensory cilia include those in the photoreceptor cells in the eye, in the hair cells in the ear, and in olfactory neurons in the nose. The universal architecture of cilia and their close relatives, flagella, consists of a microtubular axonemal core enclosed by a membrane, exposed on the cell surface. A convergence of results from several systems has led to detailed structural descriptions of cilia and flagella (Perkins et al., 1986; Dutcher, 1995; Rosenbaum and Witman, 2002).

In the nematode *C. elegans* 60 of the 302 neurons of the hermaphrodite are ciliated sensory neurons (CSN) (Ward et al., 1975; White et al., 1986), forming many structurally distinct types of sensory cilia. Whereas many ciliary mutants are available in *C. elegans*, there is only one known gene mutation that completely eliminates all classes of sensory cilia and all functional components of cilium structure. This gene is *daf-19*, and it encodes the sole *C. elegans* member of the RFX-type transcription factors (Swoboda et al., 2000), found widely in the eukaryotic kingdom. All members of the RFX transcription factor family are characterized by the presence of a conserved DNA binding domain (DBD). The RFX-DBD binds to special motifs (X-boxes) in promoters of its target genes. Genes containing the X-box promoter motif are called *xbx* genes.

Apart from *C. elegans* DAF-19, the RFX family currently contains nine characterized members: five in mice and humans (RFX1-5) (Emery et al., 1996a), two in *Drosophila melanogaster* (dRFX1-2) (Dubruille et al., 2002; Otsuki et al., 2004), and one member each from *Schizosaccharomyces pombe* (Wu and McLeod, 1995) and *Saccharomyces cerevisiae* (Huang et al., 1998). The data obtained so far suggest diverse biological roles of RFX proteins. In yeasts they regulate some aspects of the cell cycle (Wu and McLeod, 1995; Huang et al., 1998). In humans RFX factors are involved in the transcriptional regulation of major histocompatibility complex class II genes (RFX5) (Reith and Mach, 2001) and in the

modulation of Ras signaling in epithelial cells (RFX3) (Maijgren et al., 2004).

The finding of X-boxes in promoters of ciliary genes in *C. elegans* has revealed an important role of the RFX family in the regulation of ciliogenesis (Swoboda et al., 2000). Since then, the conservation of RFX-binding elements has been reported in several distantly related species. For example, some ciliary genes in *D. melanogaster* contain X-box-like sequences in their promoters (Avidor-Reiss et al., 2004). However, an experimental demonstration of an RFX-dependence for *Drosophila* ciliary gene candidates exists so far only for the *nompB* gene (Han et al., 2003). Recently, data about the possible RFX regulation of ciliogenesis in mammals were also obtained. *Rfx3*-deficient mice exhibit frequent left-right asymmetry defects, which are caused by ciliary abnormalities in mutant embryos (Bonnafe et al., 2004). Mouse RFX3 regulates the expression of *D2lic*, the mouse ortholog of the *C. elegans* ciliary gene *xbx-1*, but does not affect the expression of *Tg737*, the mouse ortholog of the *C. elegans* ciliary gene *osm-5*. These observations suggest that RFX regulation of ciliogenesis in higher organisms is more complicated, and different subtypes of RFX proteins may be restricted to particular components of ciliary structure.

In our current work we first concentrated on the isolation of genes important for cilium structure and function using a genome-wide X-box promoter motif search in the nematode species *C. elegans* and *C. briggsae*. In this computational approach we focused our efforts only on 5′ flanking regions of genes, since X-box motifs have previously been shown to be functional in those regions (Swoboda et al., 2000). Subsequently, we performed expression analyses of the group of positive *C. elegans* matches in wild-type and *daf-19* mutant backgrounds, together with X-box mutagenesis experiments. Results of these analyses established the X-box consensus for *C. elegans,* the approximate number of *xbx* genes in the *C. elegans* genome and assigned already known and newly found ciliary genes to specific structural and functional groups. Because the organization of *C. elegans* sensory cilia is very similar to sensory cilia in mammals, the results obtained with the *C. elegans* model will have general significance.

Materials and methods

Worm strains

Growth and culture of *C. elegans* strains were carried out following standard procedures (Brenner, 1974). The following strains were used for this study: wild type N2 Bristol; JT8651 *daf-19(m86)/mnC1; lin-15(n765ts)*; JT6924 *daf-19(m86); daf-12(sa204)*; JT204 *daf-12(sa204)*; RB773 *nud-1(ok552)*; RB819 *xbx-4(ok635)*; RB857 *xbx-6(ok852)*; NL2099 *rrf-3(pk1426)*; CB3323 *che-13(e1805)*; CB1033 *che-2(e1033)*. Extrachromosomal arrays were used for all GFP expression analyses. All strains used and strain construction details are available on request.

Promoter motif search algorithm and sequence analyses

The X-box motif search was performed primarily with a Perl-based algorithm that searches through a given genome sequence for all possible matches. The algorithm first finds all sequences that match a defined consensus. After that step, the main module of the program implements a cross-match file (P. Green, personal communication), which compares a 3 kb window downstream of each match to a file containing the DNA sequences for all predicted genes, and a file containing assembled ESTs. Cross-match parameters – ‘minmatch’ and ‘minscore’ were set to 40. All other parameters were kept at default values. Minimal and maximal distances from positive matches to predicted genes were set to a range of 0-1000 nucleotides. To obtain a copy of the algorithm, contact kbubb@u.washington.edu. Cross-match must be obtained separately (see www.phrap.com for access/download information).

Subsequent to genome analyses using Cross-match, we made use of another program, DNA Motif Searcher, which takes a set of user-definable X-box sequences to search for additional motif instances in the genome. The set of motifs is interpreted into a position-specific score matrix (PSSM). Using this PSSM, the program can then identify the closest matching occurrences of the motif based on a score cutoff, or it can identify the top number of occurrences of the motif. For download or for more detailed explanations about Motif Searcher, please contact jht@u.washington.edu.

Genome sequence information, EST files, gene predictions and identities for X-box searches were obtained from the following sources: *C. elegans* complete genome sequence, WS122 release (ftp://ftp.sanger.ac.uk/pub/wormbase/WS122); *C. briggsae* draft genome sequence, cb25.agp8 version (ftp://ftp.wormbase.org/pub/wormbase/briggsae).

Generation and analysis of expression constructs

GFP expression constructs were designed by inserting about 2 kb of promoter regions and the first several codons of a gene of interest into the GFP expression vector pPD95.77 (gift from A. Fire). PCR fragments of promoter regions were obtained from wild-type N2 genomic DNA and cloned into appropriate sites of pPD95.77. For some genes, the wild-type X-boxes within promoters were mutated by overlap extension mutagenesis, replacing X-box sequences with nonspecific nucleotides containing indicative restriction enzyme sites. To check for correct translational reading frames and promoter regions, junctions between vector and amplified inserts were verified by sequencing for all constructs. For the XBX-2::GFP translational fusion, the entire coding sequence of the gene with about 1 kb of promoter were fused to the pPD95.77 vector.

The following worm strains were used for injections and expression analyses: JT8651, JT6924 and JT204. The strain JT8651 *daf-19(m86)/mnC1; lin-15(n765ts)* served as the wild-type background, since *daf-19(m86)* is fully recessive. *daf-19* mutants are strongly Daf-c (dauer larva formation – constitutive) across the normal temperature range. Therefore, segregating dauers were recovered at 15°C to obtain a *daf-19* homozygous background. Alternatively, the strain JT6924 *daf-19(m86); daf-12(sa204)* was used as a *daf-19* mutant background. Worms of this genotype exhibit a Daf-d (dauer larva formation – defective) phenotype and do not require the recovery of dauers. In this case, JT204 *daf-12(sa204)* worms were used as a wild-type background with regard to *daf-19*.

Adult hermaphrodites were transformed using standard protocols (Mello et al., 1991). Constructs were injected typically at 10-100 ng/μl along with coinjection markers such as pRF4 (contains the dominant marker *rol-6(su1006)*) or pBLH98 (contains the wild-type *lin-15* gene to rescue *lin-15(n765ts)*).

Microscopy and imaging

GFP expression patterns were analyzed in stable transgenic lines at 1000× magnification by conventional fluorescence microscopy (Zeiss Axioplan 2). Expression patterns were examined in at least two independent transgenic lines at most developmental stages of the worm. Neuronal cell anatomies and identities followed published descriptions (Ward et al., 1975; White et al., 1986).

For the analysis of XBX-2::GFP movement properties, worms were mounted on agarose pads and anesthetized with 10 mM levamisol. Adult worms were analyzed with a Leica confocal imaging spectrophotometer TCS SP unit mounted on a Leica DMIRBE inverted microscope, and the obtained images were processed using

Leica Confocal Software 2.5. Images were taken with a 63× objective and a 488 nm GFP filter. At least 40 stacked images were converted into an AVI file with a rate of two frames per second.

RNAi feeding experiments and fluorescent dye filling assays

RNA-mediated interference (RNAi) was performed according to standard methods (Timmons et al., 2001). PCR fragments for genes of interest were generated from N2 genomic DNA and cloned into the double T7 promoter-containing vector L4440 (gift from A. Fire). All constructs were transformed into HT115(DE3) bacterial cells and plated onto NGM plates with antibiotics and IPTG. L4-stage hermaphrodites were transferred to plates with induced bacteria and F_3 progenies were analyzed for possible phenotypes.

Fluorescent dye-filling assays were performed essentially as described previously (Starich et al., 1995) using the fluorescent dye DiI. Worm strains N2 and CB3323 were used as positive and negative controls, respectively. Stained adult hermaphrodites were analyzed at 1000× magnification by conventional fluorescence microscopy (Zeiss Axioplan 2).

Genetic characterization of *xbx* gene mutants

All deletion alleles analyzed in this study were generated by the *C. elegans* Gene Knockout Consortium (http://celeganskoconsortium.omrf.org/) using publicly available methodology (http://www.mutantfactory.ouhsc.edu/protocols.asp).

The original mutated strains RB819 *xbx-4(ok635)* IV and RB957 *xbx-6(ok852)* V were outcrossed three times with N2 and the following worm strains: JT7146 *egl-4(n478) unc-33(e204)* IV and DR108 *dpy-11(e224) unc-42(e270)* V, respectively. Outcrossed worms resulted in homozygous mutant strains containing the *xbx-4(ok635)* IV and *xbx-6(ok852)* V deletion alleles. These strains were then used as the basis for further analysis.

Results

Computational search for the X-box promoter motif

The discovery of X-boxes in promoter regions of certain ciliary genes (e.g. Swoboda et al., 2000; Haycraft et al., 2001; Haycraft et al., 2003; Fan et al., 2004) prompted us to analyze the whole *C. elegans* genome for the presence of these motifs. In order to perform this analysis of all *C. elegans* promoters we implemented an in-house searching algorithm – the X-box searcher, which searches for all possible matches to a defined motif sequence (Fig. 1A). For the initial search we used a 'relaxed' consensus (RYYNYY WW RRNRAC), that fits published mammalian X-box sequences (Emery et al., 1996b) and the first emerging *C. elegans* X-boxes from previously known ciliary genes (Swoboda et al., 2000). Using the X-box searcher with the relaxed consensus generates the highest number of output matches (1927) that were equally spread within 1000 bp of promoter regions (Fig. 1B,C). Through ongoing work with *xbx* gene candidates (see below) we obtained a 'refined' X-box consensus (GTHNYY AT

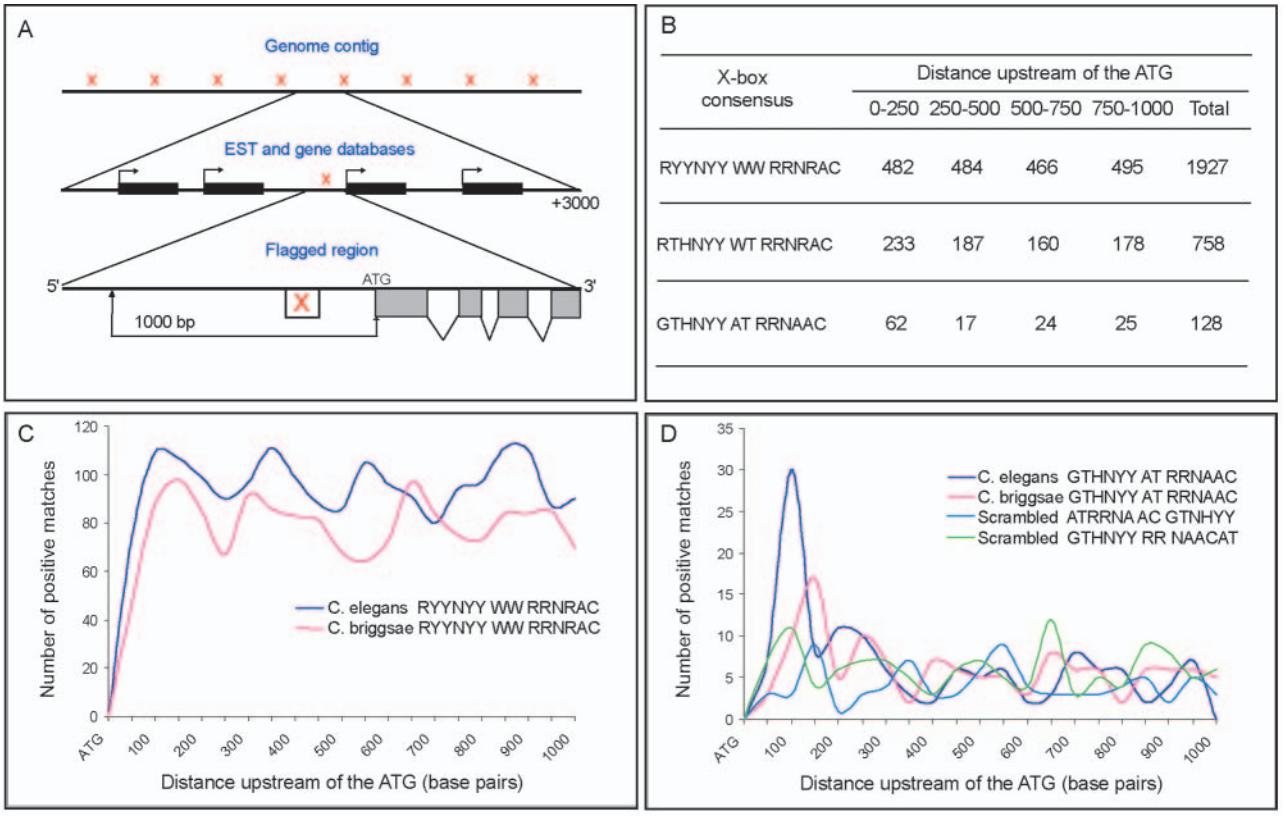


Fig. 1. Computational search for the X-box motif. (A) Schematic of the X-box search algorithm. The algorithm finds all matches for a defined motif consensus and cross-matches them against a list of predicted genes or available ESTs. The search space upstream of predicted genes was set to 1000 bp. (B) The number of matches obtained with different consensus sequences (top to bottom: relaxed, average, refined) searching the *C. elegans* genome. (C) The 'relaxed' X-box consensus (RYYNYY WW RRNRAC), used for initial searches, generates the largest number of matches that spread equally within promoter regions. (D) The 'refined' X-box consensus (GTHNYY AT RRNAAC), obtained on the basis of in vivo expression analysis of *xbx* genes, shows a significant concentration of matches in the region of around 100 bp upstream of the ATG.

RRNAAC), which corresponds to most of the experimentally proven *xbx* genes. We further re-analyzed the *C. elegans* genome using the refined consensus and found a significant reduction in the number of matches (128). Unlike the relaxed consensus, most of the refined matches show clustering around the region of 100 bp upstream of the ATG. This clustering was not observed when the X-box search sequence was scrambled (Fig. 1B,D). The refined *xbx* gene list includes the majority of experimentally proven *xbx* genes, but not all of them. At the same time, the list of genes obtained with the relaxed X-box consensus contains too much 'search noise'. According to our expression analysis (see below) we estimate that up to 90% of the relaxed matches can be found in promoter regions of non-ciliary genes or genes that are not specifically expressed in ciliated sensory neurons (CSN) (see Table S2 in supplementary material). Therefore, we additionally tried several different X-box sequences and found an 'average' consensus (RTHNYY WT RRNRAC). This consensus permitting ambiguities at three positions as compared to the refined consensus (A or G at position 1, A or T at position 7 and A or G at position 12) makes it possible to find all known X-boxes and reduces the search noise by about 2.5 times (758 matches) (Fig. 1B; see Table S1 in supplementary material). This consensus can be used for further analyses in *C. elegans*, as well as for X-box search efforts in other organisms.

The functional repertoire of the average candidate *xbx* gene list contains different molecular groups, including possible components of the ciliary structure and transport machinery (6%), transcription factors (5%), receptors (11%) and ion channels (1%). About 37% of the genes have no identified function and the rest of the list (40%) is composed of genes of various molecular identities (Fig. 2; see Table S1 in supplementary material).

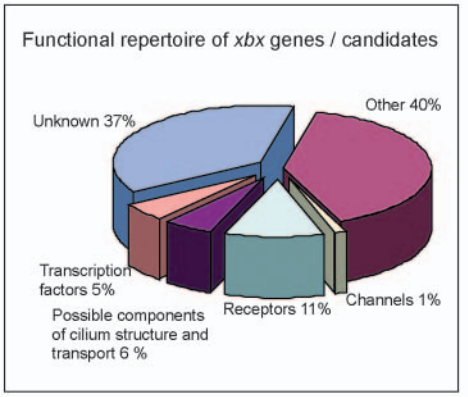
Our primary goal was to find new *xbx* and ciliary genes with high efficiency. Our X-box search might not be exhaustive, because the search algorithm is based on an originally small set of experimentally proven X-box sequences. To determine whether additional *xbx* gene candidates could be found in the *C. elegans* genome, we tried a different search approach, which is position-specific score matrix (PSSM) based (J.H.T., unpublished). Using the output data from the average list as a training set for the PSSM searcher we achieved almost the same list of *xbx* genes (data not shown), strengthening our overall strategy. However, we additionally obtained several prominent candidates that had a slightly different X-box motif than the average consensus and therefore were not found with the X-box searcher: for example *dlc-1* (GTTATT AT AAC*T*AC, which encodes a dynein light chain), C01B12.4 (GTTTCC AT AGC*T*AC, which encodes a predicted seven transmembrane receptor of the rhodopsin family).

Expression patterns of orthologous genes are often conserved. Because many orthologous transcription factors are also functionally conserved, one possible model to account for homologous gene expression patterns is conservation of specific binding sites within regulatory elements of orthologous genes (Ruvinsky et al., 2003). The nematodes *C. briggsae* and *C. elegans* are closely related species with very similar overall genome organizations (Stein et al., 2003). To find possible conservations of X-box regulatory elements between those two organisms we applied the *C. elegans* X-box search strategy to the *C. briggsae* genome. The number of obtained matches was slightly less than with *C. elegans*, probably because of the draft quality of the *C. briggsae* genome. Nevertheless, the profile of X-box distribution within promoter regions was similar to that of *C. elegans* (Fig. 1D).

In summary, using two different X-box promoter motif search approaches, X-box consensus sequences with varying degrees of refinement, together with cross-species comparisons and gene expression analysis (see below), we were able to identify a large number of bona fide *xbx* genes, a significant part of which we expect to also be ciliary- or CSN-specific genes.

Expression analysis of the *xbx* gene candidates

Our computational search has revealed a large, heterogeneous group of X-box matches (see Table S1 in supplementary



Possible components of cilium structure and transport	
Y110A7A.20	Homolog of *Brachydano rerio* IFT 20
C02H7.1	Homolog of human MIP-T3
K08D12.2	Homolog of human XRP-2
arl-3	GTP-binding ADP-ribosylation factor-like protein
Y37E3.5	GTP-binding ADP-ribosylation factor-like protein
C34B2.5	TPR repeat containing protein
C38B6.6	TPR repeat containing protein
C48B6.8	Contains WD40 domain
R06F6.8	Contains WD40 domain
C04C3.5	Contains coiled-coil domains

Transcription factors	
C06E2.1	Contains Zn-finger C2H2 domain
T22C8.3	Contains Zn-finger C2H2 domain
K05F1.5	Contains Zn-finger C2H2 domain
K10B3.5	Contains Zn-finger C2H2 domain
C01H6.3	Contains Zn-finger C2H2 domain
T07G12.11	Contains Zn-finger C2H2 domain
Y58A7A.3	Contains Zn-finger C6HC domain
C26B2.4	Nuclear hormone receptor
ZK488.4	Nuclear hormone receptor

Receptors	
str-1	7-transmembrane receptor
str-13	7-transmembrane receptor
str-44	7-transmembrane receptor
str-144	7-transmembrane receptor
srg-2	Serpentine receptor, class G
srh-74	Serpentine receptor, class H
sru-12	Serpentine receptor, class U
srx-54	Serpentine receptor, class X

Fig. 2. Functional repertoire of *xbx* genes. Functional repertoire of *xbx* genes and some representative members from different molecular groups, including possible components of cilium structure and transport machinery, transcription factors and receptors. The diagram represents data from the 'average' consensus list of *xbx* genes/candidates (758 members) (see Table S1 in supplementary material), including experimentally proven genes.

material). To determine the specificity of our computational search with regard to ciliary and CSN structure and function we initiated expression analyses for some of the genes. For this purpose we isolated a group of candidates with different nucleotide compositions of the X-box motifs, different positions within promoter regions and different proposed molecular functions of matching genes.

Our previous model associated the expression of *xbx* genes with CSN (Swoboda et al., 2000). We predicted that DAF-19 function, a particular X-box composition and its position within the promoter should be required for the expression of this group of genes. Therefore, expression patterns of selected genes were analyzed in both wild-type and *daf-19* mutant backgrounds. For some genes we also analyzed the actual X-box sequence by replacing it with nonspecific nucleotides within the respective expression construct.

Based on our obtained expression data (Table 1), we subdivide *xbx* genes into groups: (1) genes that are strongly regulated by DAF-19 and required for the general development of cilia; (2) genes that are partially regulated by DAF-19 and are probably required for more specific ciliary and/or CSN functions.

Genes of the first group are characterized by expression in many, most or all CSN. All of them are DAF-19 dependent and for some genes tested (through mutagenesis) we also demonstrated a dependence on the actual X-box sequence (Table 1).

Most genes from this group encode known participants of

Table 1. Expression analysis of *xbx* genes

C. elegans gene	X-box sequence (distance from ATG)	*C. briggsae* ortholog	X-box sequence (distance from ATG)	Expression patterns of *C. elegans* genes	Expression properties	References
Group 1						
*che-2** (F38G1.1)	GTTGTC AT GGTGAC (–130)	CBG13647	GTATCC AT GGCAAC (–182)	Many, most, all CSN	DD, XD	Fujiwara et al., 1999; Swoboda et al., 2000
*che-13** (F59C6.7)	GTTGCT AT AGCAAC (–75)	CBG02227	GTTTCC TT GACAAC (–85)	Many, most, all CSN	DD	Haycraft et.al., 2003
*osm-1** (T27B1.1)	GCTACC AT GGCAAC (–86)	CBG16355	GTTGCC AT GGACAC (–79)	Many, most, all CSN	DD, XD	Signor et al., 1999; Swoboda et al., 2000
*osm-5** (Y41G9A.1)	GTTACT AT GGCAAC (–116)	CBG02013	GTTGCC AG GGAAAC (–91)	Many, most, all CSN	DD, XD	Haycraft et.al., 2001
*osm-6** (R31.3)	GTTACC AT AGTAAC (–100)	CBG23329	X-box not found	Many, most, all CSN	DD, XD	Collet et al., 1998; Swoboda et al., 2000
*bbs-1** (Y105E8A.5)	GTTCCC AT AGCAAC (–99)	CBG08744	GTTGTT AT GGTAAC (–310)	Many, most, all CSN	DD	Ansley et al., 2003; current work
bbs-2 (F20D12.3)	GTATCC AT GGCAAC (–94)	CBG17712	ATATCC AT GGCAAC (–82)	Many, most, all CSN	DD, XD	Ansley et. al., 2003; current work
bbs-5 (R01H10.6)	GTCTCC AT GGCAAC (–66)	CBG23799	GTTACT AT GGCAAC (–69)	Many, most, all CSN	DD	Li et al., 2004
*bbs-7** (Y75B8A.12)	GTTGCC AT AGTAAC (–108)	CBG23043	GTTGCC AT GGTTAC (–138)	Many, most, all CSN	DD	Ansley et. al., 2003; current work
*bbs-8** (T25F10.5)	GTACCC AT GGCAAC (–84)	CBG19013	GTCTCT AT GGCAAC (–73)	Many, most, all CSN	DD	Ansley et. al., 2003; current work
*xbx-1** (F02D8.3)	GTTTCC AT GGTAAC (–79)	CBG11597	GTTTCC AT GGTTAC (–93)	Many, most, all CSN	DD, XD	Schafer et.al., 2003; current work
xbx-2 (D1009.5)	GTTGCC AT GACAAC (–78)	CBG00241	GTTTCC AT GGCTAC (–83)	Many, most, all CSN	DD, XD	Current work
Group 2						
xbx-3 (M04D8.6)	GTTGTC TT GGCAAC (–98)	CBG09908	GTTTCC AA GGAGAC (–128)	Amphids, phasmids	DD	Current work
xbx-4‡ (C23H5.3)	GTTGCC AT GACAAC (–82)	CBG10549	GTTGCC CT GGTGAC (–155)	Some CSN	DD	Current work
xbx-5 (T24A11.2)	GTCTCC AT GACAAC (–122)	CBG09228	GTCTCC AT GGCAAC (–142)	Some CSN	DD	Current work
xbx-6‡ (F40F9.1)	GTTTCC AT GGAAAC (–152)	CBG19349	GTATCC AT GGAAAC (–121)	Body wall muscles, pharyngeal muscles,ventral nerve cord, phasmids	DD (phasmids only)	Current work
xbx-7 (R148.1)	GTCACC AT AGGAAC (–70)	CBG22495	X-box not found	Labial neurons, some amphid neurons, phasmids	DD	Current work
nud-1 (F53A2.4)	GTATCC AT GAAAAC (–263)	CBG24281	X-box not found	Amphids, phasmids, vulva	DD (CSN only)	Dawe et al., 2001; current work
*che-11** (C27A7.4)	ATCTCC AT GGCAAC (–86)	CBG23392	GTATCC AT AGCAAC (–120)	Many, most, all CSN	DD (amphids and phasmids only)	Qin et al., 2001; current work
odr-4† (Y102E9.1)	ATCGTC AT CGTAAC (–164)	CBG16563	ATCGCC AT GGTTAC (–261)	10 amphid and 2 phasmid neurons	DD, XD	Dwyer et. al., 1998; current work
tub-1† (F10B5.4)	ATCTCC AT GACAAC (–183)	CBG00741	ATCACC AT GGCAAC (–232)	Many, most, all CSN	DD	Current work
nhr-44 (T19A5.4)	GTCTTC AT GGCAAC (–76)	CBG19141	X-box not found	ASK, other head neurons, other cell types	DD (in ASK)	Current work
Remaining genes analyzed						
F55D12.1	GTTACC AT AGTAAC (–234)	CBG08264	GTTGTC AT GACGAC (–252)	Glia, seam cells, vulva	DI	Current work
gpa-9‡ (F56H9.4)	GTTACC AT GGAAAC (–238)	Putative ortholog not identified		ASJ, PHB, PVQ, pharyngeal muscle, spermatheca	DI	Jansen et al., 1999; current work
zag-1‡ (F28F9.1)	ATTGTC TA GGTAAC (–128)	CBG10736	X-box not found	Head and tail neurons	DI	Wacker et al., 2003; current work
F17A2.3	ACCGCC AA AGAAAC (–83)	Putative ortholog not identified		Distal tip cells, ASJ, ASI	DI	Current work
aqp-2 (C01G6.1)	ACCACC TT GAAAAC (–115)	Putative ortholog not identified		Excretory system, intestine, body wall muscle	DI	Current work

All genes analyzed were grouped according to their expression patterns and dependence on DAF-19. Upon mutation, genes marked with * cause abnormalities in general ciliary structure and function (e.g. Dyf+sensory phenotypes). Upon mutation, genes marked with † cause abnormalities in specific ciliary functions (e.g. Odr, but not Dyf phenotype). Upon mutation, genes marked with ‡, even though tested, have not yet shown abnormalities in general ciliary structure or specific ciliary functions. For genes not marked, mutants are presently not available. We tested gene expression for DAF-19 and X-box dependence, as indicated: DD, expression is DAF-19 dependent; XD, expression is X-box dependent; DI, expression is DAF-19 independent. CSN, ciliated sensory neurons. The expression patterns of some *C. elegans* genes were previously described, as referenced. The DAF-19 or X-box dependence for the genes *che-2*, *che-13*, *osm-1*, *osm-5*, *osm-6*, *bbs-5* and *xbx-1* was previously described, as referenced.

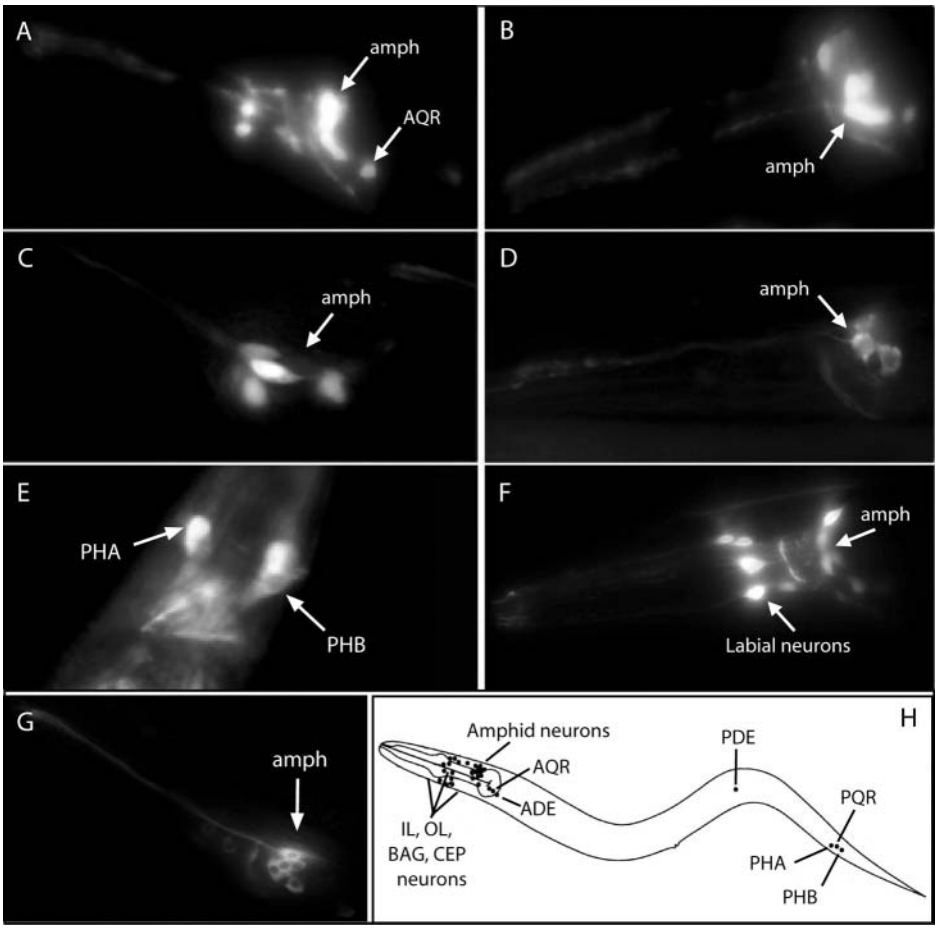


Fig. 3. Newly discovered X-box genes. (A) *xbx-2::gfp* is expressed in most or all CSN. (B) *xbx-3::gfp* is expressed in amphid and phasmid (not shown) neurons. (C) *xbx-4::gfp* is expressed in a subset of amphid neurons. (D) *xbx-5::gfp* is expressed in a subset of amphid neurons. (E) *xbx-6::gfp* is expressed in many different types of cells, including the phasmid neurons PHA and PHB. (F) *xbx-7::gfp* is expressed in a subset of amphid and interlabial neurons. (G) *tub-1::gfp* is expressed in most or all CSN. (H) Schematic diagram of CSN positioning in the *C. elegans* hermaphrodite (after Collet et al., 1998).

intraflagellar transport (IFT): *osm-1, osm-5, osm-6, che-2, che-13, xbx-1*. The mechanism of IFT was originally described in the biflagellate alga *Chlamydomonas reinhardtii* (Kozminski et al., 1993). It is characterized by the movement of IFT particles along ciliary/flagellar axonemal microtubules by means of kinesin and dynein motor molecules. *C. elegans xbx* genes encoding IFT proteins are well described. IFT components such as OSM-1, OSM-5, OSM-6, and CHE-13 are associated with the heterotrimeric motor protein kinesin-II. They are essential for anterograde transport (Signor et al., 1999; Haycraft et al., 2001; Haycraft et al., 2003). The gene *xbx-1* encodes a dynein light intermediate chain (DLIC) that is important for retrograde transport within cilia (Schafer et al., 2003).

Herein we report a new IFT gene, *xbx-2* (D1009.5). The *xbx-2::gfp* promoter fusion was strongly expressed in most of the ciliated sensory organs of the worm – amphids, phasmids, inner and outer labial quadrants (Fig. 3A). The XBX-2 protein contains a Tctex-1 domain that belongs to the family of dynein light chain (DLC) proteins. These molecules are essential for dynein assembly and participate in specific motor-cargo interactions (DiBella et al., 2001; Tai et al., 2001). To analyze the possible role of XBX-2 in the IFT process, we generated transgenic worms expressing XBX-2::GFP protein. Using time-lapse confocal microscopy, we observed movement of XBX-2::GFP particles along the ciliary axoneme in both retrograde (Fig. 4) and anterograde directions (see also Movie 1 in supplementary material).

A major part of the first group was also composed of *C. elegans* orthologs of human Bardet–Biedl syndrome genes: *bbs-1*, *bbs-2*, *bbs-5*, *bbs-7* and *bbs-8*. Recent data suggest that *bbs* genes are probably involved in the development of the basal body during cilium formation (Ansley et al., 2003; Li et al., 2004) and are required for assembly and proper function of some IFT components (Blacque et al., 2004; Fan et al., 2004). We expressed *bbs* gene members both in wild-type and in *daf-19* mutant backgrounds (in parallel with X-box mutagenesis experiments) and found that this group of genes strongly requires both DAF-19 function and proper X-box composition (Table 1). These results confirm that genes encoding general cilium structure molecules are strongly regulated by DAF-19.

The second group of *xbx* genes includes many novel genes, which are expressed in various subsets of CSN. For example, *xbx-3::gfp* was strongly expressed in all amphid and phasmid neurons, but not in other ciliated sensilla (Fig. 3B). In *daf-19* mutant worms, expression was restricted to one amphid neuron and abolished in phasmids, with occasional ectopic expression in other tissues. Expression of the novel gene *xbx-4* was also observed in some amphid and phasmid neurons (Fig. 3C), but in *daf-19* mutants it was completely abolished in both organs. The *xbx-5::gfp* promoter fusion was characterized by faint, punctate expression in phasmids and some amphid neurons (Fig. 3D). In a *daf-19* mutant background expression was absent in amphids but still visible in phasmids. The predicted XBX-5 protein contains seven transmembrane domains and can be considered as a possible receptor. The *xbx-6::gfp* construct was abundantly expressed in many cell types: pharyngeal muscles, numerous neurons in the head and tail regions, the ventral nerve cord and body wall muscles. Because of high overall expression levels, we were not able to identify individual CSN in the head region, but we observed expression in phasmids (Fig. 3E), which was strictly DAF-19 dependent. The expression in other cells was unchanged in *daf-19* mutants. The *xbx-6* gene encodes an *N*-methyl-D-aspartate receptor-associated protein. The novel gene *xbx-7* was expressed in phasmids, some amphid neurons

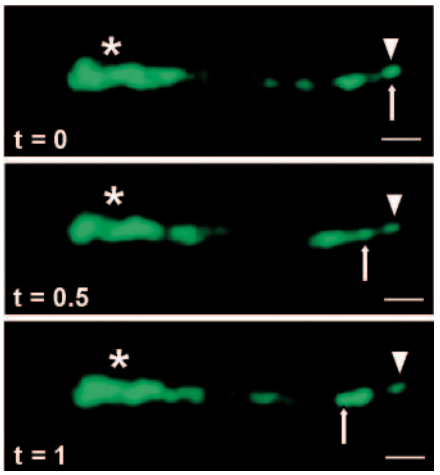


Fig. 4. An example of retrograde movement for XBX-2::GFP particles in a phasmid cilium (see also Movie 1 in supplementary material). The ciliary transition zone is marked with an asterisk. The arrowhead indicates the initial position of the moving particle at time zero (t=0). The arrow indicates the position of the moving particle at different time points (t=0.5 and 1 seconds). Scale bars: 1 μm.

and in interlabial neurons (Fig. 3F), while in *daf-19* mutants expression was significantly reduced. The gene *che-11* encodes a large protein, which is orthologous to *Chlamydomonas* IFT140 (Qin et al., 2001). The expression level of *che-11::gfp* was significantly reduced in phasmids and amphids in *daf-19* mutants (Table 2), while in the labial neurons it was almost unchanged. The nuclear hormone receptor gene *nhr-44* was expressed in head neurons, including the ciliated sensory neuron ASK, and also in other cell types, and its overall expression properties fit those of other *xbx* genes (Table 1).

Within the second group of *xbx* genes we especially note the genes *odr-4*, *nud-1* and *tub-1*, the molecular identities of which were previously described.

The gene *odr-4* has been shown to be important for the localization of some seven transmembrane domain odorant receptors to cilia (Dwyer et al., 1998). The expression of an ODR-4::GFP translational fusion was significantly reduced both in *daf-19* mutants and after mutation of the X-box sequence (data not shown). This indicates that not only transport mechanisms within cilia (IFT), but also to cilia (ODR-4) are under DAF-19 control.

The *C. elegans* ortholog of the *NudC* gene of the fungus *Aspergillus nidulans*, *nud-1*, was identified as a candidate *xbx* gene during our search. NUD-1 is an important component in microtubule-dependent nuclear positioning, which is required for proper growth, development and cellular function in both lower and higher eukaryotes. Sustained expression of *nud-1::gfp* in CSN was previously described (Dawe et al., 2001). We introduced this construct into a *daf-19* mutant background and found its expression drastically reduced (Table 2).

The gene *tub-1* is the worm ortholog of murine *tubby* which, when mutated, leads to neuronal deficits and late-onset obesity (Carroll et al., 2004). *tub-1* mutant worms exhibit functional defects in CSN and show a mild elevation of lipid accumulation (H.Y.M., unpublished data). The translational GFP fusion for *tub-1* was expressed only in the cytoplasm in all CSN (Fig. 1G). The expression level of this construct was significantly reduced in a *daf-19* mutant background (Table 2). We analyzed the 5′-UTR region of the human TUB gene and found two X-box-like sequences that perfectly match the *C. elegans* consensus: GTTGCC AT GGAAAC (–296) and GTTGCT AT AGTAAC (–339). Intriguingly, microtubule-associated protein 1A (MAP1A) can modify hearing defects in *tubby* mice (Ikeda et al., 2002), and its expression is regulated by RFX molecules (Nakayama et al., 2003). These observations suggest that the regulation of *tubby* pathways by RFX transcription factors can be conserved in evolution. Based on the data from the second group of *xbx* genes, we conclude that DAF-19 only partially regulates the expression of certain genes. These genes may only be required for specialized functions in CSN or during ciliogenesis.

The remaining genes analyzed were mostly expressed in many other different cell types and only in very few cases was expression observed predominantly in CSN (Table 1; see Table S2 in supplementary material). We checked some genes from this group (*zag-1*, *aqp-2*, *gpa-9*, F55D12.1, F17A2.3) in a *daf-19* mutant background and found that expression patterns were unchanged in the absence of DAF-19 function. Most of these X-box matches differ from the refined consensus or are located further upstream of the ATG (Table 1).

Table 2. Expression properties of novel *xbx* genes in wild type and in *daf-19(m86)* backgrounds*

	Number of amphid neurons expressing GFP						Average number of expressing amphid neurons	Number of phasmid neurons expressing GFP			Average number of expressing phasmid neurons
Genotype	0	1-5	6-10	11-15	16-20	21-24		0	1-2	3-4	
che-11::gfp											
Wild type	–	–	–	–	99	1	19	–	1	99	4
daf-19(m86)	61	39	–	–	–	–	1	87	12	1	1
nud-1::gfp											
Wild type	–	12	88	–	–	–	7	1	74	25	2
daf-19(m86)	100	–	–	–	–	–	0	100	–	–	0
tub-1::gfp											
Wild type	–	–	–	9	88	3	18	–	–	100	4
daf-19(m86)	–	–	91	9	–	–	8	19	75	6	2

*Data are given as percentage expression in different numbers of amphid and phasmid neurons. The average number of expressing neurons is specified. Two independent transgenic lines were analyzed for each gene.

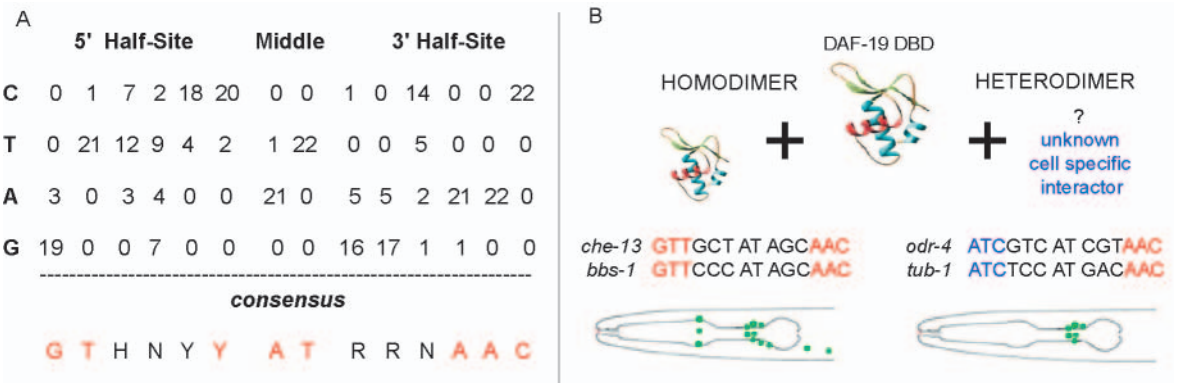


Fig. 5. Properties of the X-box motif in *C. elegans*. (A) The X-box consensus sequence obtained from in vivo expression analysis. Nucleotides marked in red are strongly conserved and important for the proper function of the motif. (B) Proposed scheme for the difference in expression patterns observed for *xbx* genes. Depending on the nucleotide composition of the X-box, DAF-19 can bind to the motif either in homodimer (driving expression in many, most or all CSN) or in heterodimer form, interacting together with some cell-specific factor (driving expression in a subset of CSN).

An additional confirmation and filtering mechanism used was a cross-species comparison, where candidates were classified as likely or unlikely *xbx* genes depending on the conservation of the X-box motif sequence between *C. elegans* and *C. briggsae* (Table 1). We found that most of the genes from the first group have putative *C. briggsae* orthologs with nearly identical X-box sequences and positions in their promoters. In the second group, the frequency of X-box occurrences in *C. briggsae* orthologs was reduced. *C. elegans xbx* gene candidates that were expressed predominantly in cell types other than CSN typically have no X-box-like sequences in promoters of their *C. briggsae* orthologs (Table 1; see Table S2 in supplementary material). Thus, the actual X-box sequence and position conservation between *C. elegans* and *C. briggsae* (and maybe other organisms) can be used as an additional measure to deduce possible molecular roles of *xbx* genes in cilia or CSN.

The expression data from 22 GFP fusions that show dependence on DAF-19 function and actual X-box sequence (Table 1) allowed us to derive an in vivo refined consensus for *C. elegans xbx* genes (Fig. 5A). This consensus can be characterized as a 14 nt imperfect palindromic sequence (GTHNYY AT RRNAAC) consisting of conserved nucleotides at 5′ (GT) and 3′ (AC) ends of the two half-sites separated by a conserved AT spacer. Most of the X-boxes matching this consensus are located in immediate proximity to the gene start (in the range of 50-250 bp upstream of the ATG). Lack of DAF-19 function or changes in the actual X-box sequence lead to drastic reduction or variation of gene expression patterns, suggesting a crucial role of the given motif sequence in the regulation of target genes (see also Discussion).

Functional analysis of *xbx* genes

Numerous mutations have been generated in *C. elegans* that affect cilia and their formation (Perkins et al., 1986; Starich et al., 1995). They cause various mutant phenotypes, including Dyf (fluorescent dye filling defective), Osm (avoidance of high osmotic strength defective), Mat (male mating defective), Che and Odr (chemo- and odorant sensation defective).

xbx genes from group 1 show Dyf and to various extents also different sensory phenotypes (Table 1) (Collet et al., 1998; Signor et al., 1999; Fujiwara et al., 1999; Haycraft et al., 2001; Haycraft et al., 2003; Schafer et al., 2003; Blacque et al., 2004; Li et al., 2004). Mutations that reduce dye filling of amphid and phasmid neurons are indicative of general defects in cilium structure and are often accompanied by various sensory mutant phenotypes (Starich et al., 1995).

The availability of a large group of *xbx* gene candidates obtained in our computational search prompted us to try a screening approach for new members of general cilium structure and function using genetic interference mediated by double stranded RNA (RNAi). We analyzed already known *dyf* genes (*che-13*, *osm-5*) together with novel candidates (*bbs-2*, *bbs-7*, *xbx-1*, *xbx-2*) with regard to the Dyf phenotype. It was known that RNAi is less efficient in neuronal types of cells (Simmer et al., 2002). To increase possible effects of interference, in parallel to wild-type worms we also tested *rrf-3* mutants, which are sensitive to RNAi in diverse tissues, especially in neurons (Simmer et al., 2002). We found that RNAi of the ciliary genes tested did not result in strong Dyf phenotypes in wild type or *rrf-3* mutants (data not shown). Therefore, RNAi cannot be implemented as a quick and easy screening technique with regard to the Dyf phenotype in *C. elegans*.

Unlike those in group 1, the roles of many genes in group 2 are largely unknown. To begin the functional investigation of the second group we analyzed two genes, *xbx-4* and *xbx-6*, mutants of which were available. The *xbx-4(ok635)* deletion extends over 951 bp starting in the promoter region and ending in the first intron, completely eliminating the beginning of the gene. The *xbx-6(ok852)* deletion extends over 1720 bp starting in the promoter region and covering five of the six exons of the gene. Since the expression of these genes is associated with CSN, we first focused our efforts on phenotypes related to defects in cilium structure or sensory abnormalities. The following results were obtained: both analyzed deletion alleles demonstrated wild-type responses with regard to fluorescent dye filling, high osmotic strength avoidance and in odorant sensation assays using three different odors (data not shown).

These *xbx-4* and *xbx-6* results suggest that in some instances members of the *xbx* gene family may have specialized molecular functions and therefore mutants have more

specialized sets of sensory phenotypes, although we cannot formally exclude genetic redundancy with other (*xbx*) genes expressed in CSN. For example, two other members of group 2, the genes *odr-4* and *tub-1*, when mutated, also do not produce general structural defects of cilia, but more specialized functional ciliary abnormalities, such as selective defects in odorant sensation (Dwyer et al., 1998) (H.Y.M., unpublished).

In conclusion, our data support the sorting of *xbx* genes into different groups, where members of group 1 are typically required for more general aspects of cilia formation, while genes from group 2 are typically required for more specialized functions within cilia and/or CSN.

Discussion

Efficiency of ciliary gene searching

To find groups of genes that belong to general and to specialized ciliary gene classes we used a promoter motif, the X-box, which participates in the regulation of certain ciliary genes in the nematode *Caenorhabditis elegans*. To make our search as efficient as possible we repeated, expanded and refined it with new sets of parameters (in a bootstrap-like fashion). As a result, we extracted from the *C. elegans* genome a list of 758 *xbx* gene candidates (Table S1 in supplementary material). We predict around 150 candidates from our list to be ciliary/ciliated sensory neuron (CSN)-specific genes, since they meet the following combination of criteria, found through experimental work:

(i) Out of more than 30 *xbx* genes tested for GFP expression patterns, about 60% of them, having an X-box motif fitting the refined consensus, were expressed in CSN, and most of them only in this group of neurons. When tested by mutational analyses expression in CSN was nearly always DAF-19 and X-box dependent.

(ii) Most of these experimentally confirmed *xbx* genes have the X-box promoter motif fairly close upstream of the ATG (up to –250), a pattern found previously for the first few cilium-specific *xbx* genes analyzed (Swoboda et al., 2000). By using various randomly scrambled X-box sequences for searches we found that the concentration of hits closely upstream of the ATG disappeared and became uniform throughout the search space (Fig. 1D). We also analyzed available expression data for genes, where the X-box match was located further upstream of the ATG and found that these genes are typically not expressed in a ciliary- or CSN-specific manner (Table S2 in supplementary material). By using the X-box as an anchor it is possible that gene predictions for some of our candidates could be re-evaluated. Therefore, some X-box matches located further upstream can still be considered as functional promoter motifs (Table S1 in supplementary material).

(iii) Finally, cross-species comparisons with the closely related nematode *C. briggsae* show strong conservation of the X-box sequence and position within promoters of many of the *C. elegans* experimentally confirmed *xbx* genes.

We note that our search is not completely unbiased, because the search algorithm is based on an originally small set of experimentally proven *xbx* genes. Therefore, as further *xbx* genes will be shown to be ciliary- or CSN-specific genes, their X-boxes will be included into the search parameters (bootstrap mechanism).

Other genome search approaches, in part utilizing X-box matches as a parameter, were used in different organisms to find general components of cilia formation (Li et al., 2004; Avidor-Reiss et al., 2004), yielding a set of conserved ciliary genes and gene candidates. We compared information from flagellar and basal body genes (Li et al., 2004) with our list of X-boxes and found an overlap of 15 X-box matches. At the same time, we found eight additional *xbx* gene candidates (Table S1 in supplementary material). Despite overlaps, many of the X-box matches were different between the respective searches, possibly because of different experimental parameters. Another important comparison was obtained from a recent study of olfactory neuron-specific genes (Colosimo et al., 2004), where we found 56 genes in common (Table S1 in supplementary material). All approaches together with further filtering mechanisms will give a complete, exhaustive list of genes important for structure and function of cilia and CSN.

Properties of the X-box motif in *C. elegans*

cis-Regulatory elements are information processing devices hardwired into the genomic DNA sequence, the function of which is to regulate gene expression (Davidson, 2001). Frequently, they are organized into modules that include many sites for DNA binding proteins (Howard and Davidson, 2004). In *C. elegans*, only some cell-type-defining transcription factors (CEH-10/TTX-3, MEC-3) target single binding sites in promoters of regulated genes (Zhang et al., 2002; Wenick and Hobert, 2004). Typically, also DAF-19 targets contain only a single X-box motif in their promoters. Nevertheless, different *xbx* genes show different expression properties, suggesting the presence of specific DAF-19 co-regulators. We propose that the information about particular gene expression profiles could already be stored at the level of the X-box sequence. For example, all genes for cilium structure and IFT from group 1 have perfect matches to the refined X-box motif consensus. Each gene of this group is expressed in most or all CSN and is strongly dependent on DAF-19 function. Whereas a perfect match to the 'refined consensus' does not automatically predict expression in all CSN, variations in the X-box motif sequence, especially in the more variable half-site (GTHNYY), predict different gene expression properties (*xbx-3*, *che-11*, *odr-4*, *tub-1*). These genes are either expressed only in a subset of CSN (*xbx-3*, *odr-4*) or are only partially dependent on DAF-19 function (*che-11*, *tub-1*).

Structural experiments have shown that each half-site of a symmetric X-box interacts with both DBDs of the RFX homodimer (Gajiwala et al., 2000). Nevertheless, binding of RFX is not ultimately dependent on dimerization and monomers can bind to a single 'high-affinity' half-site (RGYAAC) (Siegrist et al., 1993; Emery et al., 1996b).

We hypothesize that DAF-19 is a crucial transcription factor of genes required for general cilium formation. In this case, DAF-19 recognizes symmetric X-boxes as a homodimer and strongly activates their expression in most or all CSN (Fig. 5B). If a target gene contains a more asymmetric X-box sequence, it can be recognized by heterodimers of DAF-19 together with other, as yet unidentified factors, leading then to specific expression patterns in subsets of CSN (Fig. 5B). Hypothetically, X-box motif distances from the ATG could also contribute to target gene expression variability.

DAF-19 regulates the development of a 'ciliary module'

The development of sensory neurons in *C. elegans* is a complicated process that includes many stages and interactions of different transcription factors (Melkman and Sengupta, 2004). The gene *daf-19* acts at late stages of sensory neuron development when the respective cell fates have already been determined and specification and subsequent differentiation occurs. Several different types of genes are required to produce functional cilia in the worm: genes for their molecular structure, genes implicated in ciliary transport machineries and genes involved in signal reception and transduction (Jansen et al., 1999; Troemel, 1999; Rosenbaum and Witman, 2002; Melkman and Sengupta, 2004).

Our previous model associated DAF-19 regulation with only a certain group of genes functioning in cilium morphogenesis and architecture (Swoboda et al., 2000). The data obtained in our current research suggest that the repertoire of DAF-19-dependent genes is much broader (Fig. 2). For the first time we have found that DAF-19 can also regulate genes of ciliary function. For example, expression of the gene *odr-4* requires both a correct X-box sequence and the presence of DAF-19. The ODR-4 protein has been shown to be an important factor for localizing a subset of seven transmembrane domain odorant receptors to cilia (Dwyer et al., 1998). Moreover, we have shown that DAF-19 can directly regulate expression levels of some putative receptor proteins. For example, the genes *xbx-5* and *xbx-6* encode a seven transmembrane domain protein and an *N*-methyl-D-aspartate receptor-associated protein, respectively.

In addition to the group of genes with signal reception properties, we found X-boxes in promoters of proposed transcription factors (Fig. 2). Most of these factors contain a C2H2-type zinc-finger domain. The important role of this type of transcription factors for the development of cell-specific properties in CSN was already described for the gene *che-1* (Uchida et al., 2003). Thus, we predict the presence of cilium-specific developmental cascades directed by DAF-19-dependent transcription factors. These cascades may be required for specialized ciliary functions as well as being necessary for the possible parallel regulation of CSN specification and their final functional differentiation. For example, it has been demonstrated that DAF-19 can affect the expression of indirect targets in the HOB-specific pathway through some unknown factor(s) (the male-specific ciliated HOB neuron is necessary for sensation of the hermaphrodite vulva during mating) (Yu et al., 2003). Our own data also suggest that DAF-19 could be required not only for general cilia formation, but in some instances for the development of cell-specific properties as well. For example, we observed that the gene *nhr-44* was expressed in a DAF-19-dependent manner in the ciliated sensory neuron ASK. This gene belongs to the nuclear hormone receptor family, which includes many ligand-regulated transcriptional modulators involved in many developmental processes (Miyabayashi et al., 1999).

Another example is the gene *nud-1*. We propose that certain microtubule-associated molecules (like NUD-1) acting during early developmental stages could later be recruited by DAF-19 for the purposes of cilia development. The important role of *nud-1* in nuclear migration during embryogenesis in *C. elegans* was previously described (Dawe et al., 2001). It has also been shown that the mammalian ortholog of NUD-1, NudC, associates with the dynein motor complex during neuronal migration (Aumais et al., 2001). Therefore, we suggest a possible role for NUD-1 as a component of IFT during ciliogenesis. In addition to *nud-1*, we extracted two further X-box-containing genes from our list of candidates, *spd-5* (Hamill et al., 2002) and *dlc-1*, which are also involved in nuclear migration during early embryonic development and might later be recruited by DAF-19 for the development of CSN.

Considering the above we propose a model where DAF-19 regulates the development of a 'ciliary module' during the differentiation of sensory neurons in *C. elegans* (Fig. 6). According to this model, DAF-19 is a key factor for the general development of cilia. At the same time, together with other factors, it can drive the expression of genes required for specialized functions in cilia.

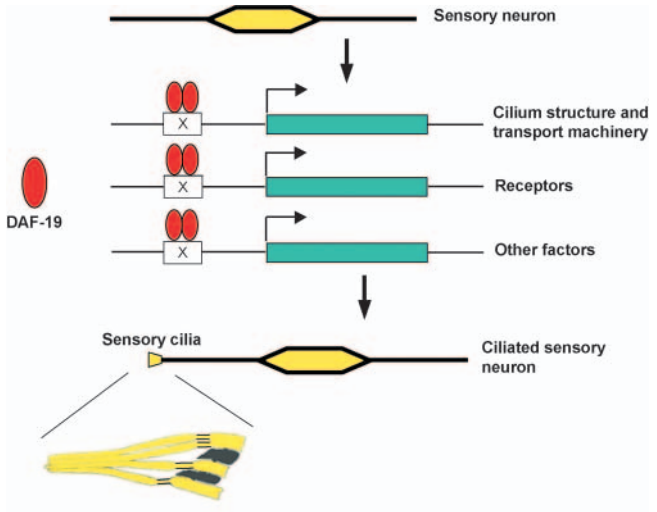


Fig. 6. Development of a 'ciliary module' in *C. elegans*. DAF-19 regulates the development of the module, which includes genes for the cilium structure and transport machinery, receptors and other factors. The activation of this module leads to the formation of functional ciliated endings during specification of sensory neurons in the worm.

xbx genes and cilia-dependent diseases

In mammals cilia are near ubiquitous organelles that project from the surfaces of many different cell types to carry out motility and sensory functions (Afzelius, 2004). Cilia have been implicated directly in many developmental processes such as generation of left-right asymmetry, heart development, maintenance of the renal epithelium, respiratory function, and physiological balance of the cerebrospinal fluid. Defects in cilia function and structure lead to a wide range of developmental problems and diseases (Pazour and Rosenbaum, 2002; Afzelius, 2004), among them: Bardet–Biedl syndrome, polycystic kidney disease (PKD), X-linked retinitis pigmentosa 2, nephronophthisis, maturity-onset obesity, etc. Many *C. elegans* genes that are orthologous to genes of human cilia-dependent diseases are X-box-containing genes. Therefore, we believe that the understanding of ciliogenesis in *C. elegans* will have a significant impact on the understanding and treatment of cilia-based pathologies in humans.

We thank the following people for technical assistance, helpful discussions, unpublished information, *C. elegans* strains, cosmid clones, cDNA clones, RNAi clones and GFP vectors: J. Ahringer, M. Barr, G. Caldwell, A. Coulson, A. Fire, C. Haycraft, Y. Kohara, G. Merrihew, E. Onischenko, S. Tat, H. Tian, B. Yoder and especially members of our and the T. Bürglin laboratories. We thank G. Gahmon and A. Laurençon for critical reading of the manuscript. We thank the C. elegans Genome Sequencing Consortium for providing genome sequence information and the Caenorhabditis Genetics Center, which is funded by the National Institutes of Health, for providing some of the *C. elegans* strains used in this study. We thank the C. elegans Knockout Consortium for providing gene deletion mutants. Work by H.Y.M. was supported by a postdoctoral fellowship from the Human Frontier Science Promotion Organization (Strasbourg, France). M.R.L. acknowledges scholar awards from the Canadian Institutes of Health Research (CIHR) and Michael Smith Foundation for Health Research (MSFHR), and funding provided by the Heart and Stroke Foundation of B. C. and Yukon and the Natural Sciences and Engineering Research Council of Canada (NSERC). This work was supported in part by Public Health Service Grants from NIH to J.H.T. and G.R., and by a grant to P.S. from the Swedish Foundation for Strategic Research (SSF) in Stockholm, Sweden.

Supplementary material

Supplementary material for this article is available at
http://dev.biologists.org/cgi/content/full/132/8/1923/DC1

Note added in proof

The *C. elegans* gene names *xbx-2* and *dylt-2* describe the same gene, *D1009.5*.

References

Afzelius, B. A. (2004). Cilia-related diseases. *J. Pathol.* **204**, 470-477.

Ansley, S. J., Badano, J. L., Blacque, O. E., Hill, J., Hoskins, B. E., Leitch, C. C., Kim, J. C., Ross, A. J., Eichers, E. R., Teslovich, T. M., Mah, A. K., Johnsen, R. C., Cavender, J. C., Lewis, R. A., Leroux, M. R., Beales, P. L. and Katsanis, N. (2003). Basal body dysfunction is a likely cause of pleiotropic Bardet–Biedl syndrome. *Nature* **425**, 628-633.

Aumais, J. P., Tunstead, J. R., McNeil, R. S., Schaar, B. T., McConnell, S. K., Lin, S., Clark, G. D. and Yu-Lee, L. (2001). NudC associates with Lis1 and the dynein motor at the leading pole of neurons. *J. Neurosci.* **21**, 1-7.

Avidor-Reiss, T., Maer, A. M., Koundakjian, E., Polyanovsky, A., Keil, T., Subramaniam, S. and Zuker, C. S. (2004). Decoding cilia function: defining specialized genes required for compartmentalized cilia biogenesis. *Cell* **117**, 527–539.

Blacque, O. E., Reardon, M. J., Li, C., McCarthy, J., Mahjoub, M. R., Ansley, S. J., Badano, J. L., Mah, A. K., Beales, P. L., Davidson, W. S., Johnsen, R. C., Audeh, M., Plasterk, R. H. A., Baillie, D. L., Katsanis, N., Quarmby, L. M., Wicks, S. R. and Leroux, M. R. (2004). Loss of *C. elegans* BBS-7 and BBS-8 protein function results in cilia defects and compromised intraflagellar transport. *Genes Dev.* **18**, 1630-1642.

Bonnafe, E., Touka, M., AitLounis, A., Baas, D., Barras, E., Ucla, C., Moreau, A., Flamant, F., Dubruille, R., Couble, P., Collignon, J., Durand, B. and Reith, W. (2004). The transcription factor RFX3 directs nodal cilium development and left-right asymmetry specification. *Mol. Cell. Biol.* **10**, 4417-4427.

Brenner, S. (1974). The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71-94.

Carroll, K., Gomez, C. and Shapiro, L. (2004). Tubby proteins: the plot thickens. *Mol. Cell. Biol.* **5**, 55-63.

Collet, J., Spike, C. A., Lundquist, E. A., Shaw, J. E. and Herman, R. K. (1998). Analysis of *osm-6*, a gene that affects sensory cilium structure and sensory neuron function in *Caenorhabditis elegans*. *Genetics* **148**, 187-200.

Colosimo, M. E., Brown, A., Mukhopadhyay, S., Gabel, C., Lanjuin, A. E., Samuel, A. D. and Sengupta, P. (2004). Identification of thermosensory and olfactory neuron-specific genes via expression profiling of single neuron types. *Curr. Biol.* **14**, 2245-2251.

DiBella, L. M., Benashski, S. E., Tedford, H. W., Harrison, A., Patel-King, R. S. and King, S. M. (2001). The Tctex1/Tctex2 class of dynein light chains. *J. Biol. Chem.* **276**, 14366-14373.

Davidson, E. H. (2001). *Genomic Regulatory Systems*. San Diego: Academic Press.

Dawe, A. L., Caldwell, K. A., Harris, P. M., Morris, N. R. and Caldwell, G. A. (2001). Evolutionarily conserved nuclear migration genes required for early embryonic development in *Caenorhabditis elegans. Dev. Genes. Evol.* **211**, 434-441.

Dubruille, R., Laurencon, A., Vandaele, C., Shishido, E., Coulon-Bublex, M., Swoboda, P., Couble, P., Kernan, M. and Durand, B. (2002). Drosophila Regulatory factor X is necessary for ciliated sensory neuron differentiation. *Development* **129**, 5487-5498.

Dutcher, S. K. (1995). Flagellar assembly in two hundred and fifty easy-to-follow steps. *Trends Genet.* **11**, 398-404.

Dwyer, N. D., Troemel, E. R., Sengupta, P. and Bargmann, C. I. (1998). Odorant receptor localization to olfactory cilia is mediated by ODR-4, a novel membrane-associated protein. *Cell* **93**, 455-466.

Emery, P., Durand, B., Mach, B. and Reith, W. (1996a). RFX proteins, a novel family of DNA binding proteins conserved in the eukaryotic kingdom. *Nucl. Acids Res.* **24**, 803-807.

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

Fan, Y., Esmail, M. A., Ansley, S. J., Blacque, O. E., Boroevich, K., Ross, A. J., Moore, S. J., Badano, J. L., May-Simera, H., Compton, D. S., Green, J. S., Lewis, R. A., Haelst, M. M., Parfrey, P. S., Baillie, D. L., Beales, P. L., Katsanis, N., Davidson, W. S. and Leroux, M. R. (2004). Mutations in a member of the Ras superfamily of small GTP-binding proteins causes Bardet-Biedl syndrome. *Nat. Genet.* **36**, 989-993.

Fujiwara, M., Ishihara, T. and Katsura, I. (1999). A novel WD40 protein, CHE-2, acts cell-autonomously in the formation of *C. elegans* sensory cilia. *Development* **126**, 4839-4848.

Gajiwala K. S., Chen, H., Cornille, F., Roques, B. P., Reith, W., Mach, B. and Burley, S. K. (2000). Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. *Nature* **403**, 916-921.

Hamill, D. R., Severson, A. F., Carter, J. C. and Bowerman, B. (2002). Centrosome maturation and mitotic spindle assembly in *C. elegans* require SPD-5, a protein with multiple coiled-coil domains. *Dev. Cell* **3**, 673-684.

Han, Y., Kwok, B. H. and Kernan, M. J. (2003). Intraflagellar transport is required in *Drosophila* to differentiate sensory cilia but not sperm. *Curr. Biol.* **13**, 1679-1686.

Haycraft, C. J., Swoboda, P., Taulman, P. D., Thomas, J. H. and Yoder, B. K. (2001). The *C. elegans* homolog of the murine cystic kidney disease gene Tg737 functions in a ciliogenic pathway and is disrupted in osm-5 mutant worms. *Development* **128**, 1493-1505.

Haycraft, C. J., Schafer, J. C., Zhang, Q., Taulman, P. D. and Yoder, B. K. (2003). Identification of CHE-13, a novel intraflagellar transport protein required for cilia formation. *Exp. Cell Res.* **284**, 251-263.

Howard, M. L. and Davidson, E. H. (2004). cis-Regulatory control circuits in development. *Dev. Biol.* **271**, 109-118.

Huang, M., Zhou, Z. and Elledge, S. J. (1998). The DNA replication and damage checkpoint pathways induce transcription by inhibition of the Crt1 repressor. *Cell* **94**, 595-605.

Ikeda, A., Zheng, Q. Y., Zuberi, A. R., Johnson, K. R., Naggert, J. K. and Nishina, P. M. (2002). Microtubule-associated protein 1A is a modifier of tubby hearing (*moth1*). *Nat. Genet.* **30**, 401-405.

Jansen, G., Thijssen, K. L., Werner, P., Horst, M., Hazendonk, E. and Plasterk, R. H. A. (1999). The complete family of genes encoding G proteins of *Caenorhabditis elegans. Nat. Genet.* **21**, 414-419.

Kozminski, K. G., Johnson, K. A., Forscher, P. and Rosenbaum, J. L. (1993). A motility in the eukaryotic flagellum unrelated to flagellar beating. *Proc. Natl. Acad. Sci. USA* **90**, 5519-5523.

Li, J. B., Gerdes, J. M., Haycraft, C. J., Fan, Y., Teslovich, T. M., May-Simera, H., Li, H., Blacque, O. E., Li, L., Leitch, C. C., Lewis, R. A., Green, J. S., Parfrey, P. S., Leroux, M. R., Davidson, W. S., Beales, P. L., Guay-Woodford, L. M., Yoder, B. K., Stormo, G. D., Katsanis, N. and Dutcher, S. K. (2004). Comparative genomics identifies a flagellar and basal body proteome that includes the *BBS5* human disease gene. *Cell* **117**, 541-552.

Maijgren, S., Sur, I., Nilsson, M. and Toftgard, R. (2004). Involvement of RFX proteins in transcriptional activation from a Ras-responsive enhancer element. *Arch. Dermatol. Res.* **295**, 482-489.

Mello, C. C., Kramer, J. M., Stinchcomb, D. and Ambros, V. (1991).

Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. *EMBO J.* **10**, 3959-3970.

Melkman, T. and Sengupta, P. (2004). The worm's sense of smell. Development of functional diversity in the chemosensory system of *Caenorhabditis elegans*. *Dev. Biol.* **265**, 302-319.

Miyabayashi, T., Palfreyman, M. T., Sluder, A. I., Slack, F. and Sengupta, P. (1999). Expression and function of members of a divergent nuclear receptor family in *Caenorhabditis elegans*. *Dev. Biol.* **215**, 314-331.

Nakayama, A., Murakami, H., Maeyama, N., Yamashiro, N., Sakakibara, A. and Takahashi, M. (2003). Role for RFX transcription factors in non-neuronal cell-specific inactivation of the microtubule-associated protein MAP1A promoter. *J. Biol. Chem.* **278**, 233-240.

Otsuki, K., Hayashi, Y., Kato, M., Yoshida, H. and Yamaguchi, M. (2004). Characterization of dRFX2, a novel RFX family protein in Drosophila. *Nucl. Acids Res.* **32**, 5636-5648.

Pazour, C. J. and Rosenbaum, J. L. (2002). Intraflagellar transport and cilia-dependent diseases. *Trends Cell Biol.* **12**, 551-555.

Perkins, L. S., Hedgecock, E. M., Thomson, J. N. and Culotti, J. G. (1986). Mutant sensory cilia in the nematode *Caenorhabditis elegans. Dev. Biol*. **117**, 456-487.

Qin, H., Rosenbaum, J. L. and Barr, M. M. (2001). An autosomal recessive polycystic kidney disease gene homolog is involved in intraflagellar transport in *C. elegans* ciliated sensory neurons. *Curr. Biol.* **11**, 457-461.

Reith, W. and Mach, B. (2001). The Bare lymphocyte syndrome and the regulation of MHC expression. *Annu. Rev. Immunol.* **19**, 331-373.

Rosenbaum, J. L. and Witman, G. B. (2002). Intraflagellar transport. *Nature Reviews: Mol. Cell. Biol.* **3**, 813-825.

Ruvinsky, I. and 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. and 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.

Siegrist, C. A., Durand, B., Emery, P., David, E., Hearing, P., Mach, B. and Reith, W. (1993). RFX1 is identical to EF-C and functions as a transactivator of the hepatitis B virus enhancer. *Mol. Cell. Biol.* **13**, 6375-6384.

Signor, D., Wedaman, K. P., Orozco, J. T., Dwyer, N. D., Bargmann, C. I., Rose, L. S. and Scholey, J. M. (1999). Role of a class DHC1b dynein in retrograde transport of IFT motors and IFT raft particles along cilia, but not dendrites, in chemosensory neurons of living *Caenorhabditis elegans*. *J. Cell. Biol.* **147**, 519-530.

Simmer, F., Tijsterman, M., Parrish, S., Koushika, S. P., Nonet, M. L., Fire, A., Ahringer, J. and Plasterk, R. (2002). Loss of the putative RNA-directed RNA polymerase RRF-3 makes *C. elegans* hypersensitive to RNAi. *Curr. Biol.* **12**, 1317-1319.

Starich, T. A., Herman, K. R., Kari, C. K., Yeh, W., Schackwitz, W. S., Schuyler, M. W., Collet, J., Thomas, J. H. and Riddle, D. L. (1995). Mutations affecting the chemosensory neurons of *Caenorhabditis elegans. Genetics* **139**, 171-188.

Stein, L. D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M. R., Chen, N., Chinwalla, A., Clarke, L., Clee, C., Coghlan, A. et al. (2003). The genome sequence of *Caenorhabditis briggsae:* a platform for comparative genomics. *PLoS Biol.* **1**, 166-192.

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

Tai, A. W., Chuang, J. and Sung, C. (2001). Cytoplasmic dynein regulation by subunit heterogeneity and its role in apical transport. *J. Cell. Biol.* **153**, 1499-1509.

Timmons, L., Court, D. L. and Fire, A. (2001). Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in *Caenorhabditis elegans*. *Gene* **263**, 103-112.

Troemel, E. R. (1999). Chemosensory signaling in *C. elegans*. *BioEssays* **21**, 1011-1020.

Uchida, O., Nakano, H., Koga, M. and Ohshima, Y. (2003). The *C. elegans che-1* gene encodes a zinc finger transcription factor required for specification of the ASE chemosensory neurons. *Development* **130**, 1215-1224.

Wacker, I., Schwarz, V., Hedgecock, E. M. and Hutter, H. (2003). *zag-1*, a Zn-finger homeodomain transcription factor controlling neuronal differentiation and axon outgrowth in *C. elegans*. *Development* **130**, 3795-3805.

Ward, S., Thomson, N., White, J. G. and Brenner, S. (1975). Electron microscopical reconstruction of the anterior sensory anatomy of the nematode *Caenorhabditis elegans*. *J. Comp. Neurol*. **160**, 313-337.

Wenick, A. S. and 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.

White, J. G., Southgate, E., Thomson, J. N. and Brenner, S. (1986). The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **314**, 1-340.

Wu, S. Y. and McLeod, M. (1995). The sak1 gene of *Schizosaccharomyces pombe* encodes an RFX family DNA-binding protein that positively regulates cyclic AMP-dependent protein kinase-mediated exit from the mitotic cell cycle. *Mol. Cell. Biol.* **15**, 1479-1488.

Yu, H., Pretot, R. F., Bürglin, T. R. and Sternberg, P. W. (2003). Distinct roles of transcription factors EGL-46 and DAF-19 in specifying the functionality of a polycystin-expressing sensory neuron necessary for *C. elegans* male vulva location behavior. *Development* **130**, 5217-5227.

Zhang, Y., Ma, C., Delohery, T., Nasipak, B., Foat, B. C., Bounoutas, A., Bussemaker, H. J., Kim, S. K. and Chalfie, M. (2002). Identification of genes expressed in *C. elegans* touch receptor neurons. *Nature* **418**, 331-335.

Table S1. List of 758 *xbx* genes/candidates, obtained with the 'average' consensus of the X-box

Gene identity	Gene name	X-box match	(I)	(II)	(III)	Distance from ATG	Molecular function or similarity
C06A6.7		GTAACTATGAAAAC				7	Unknown
C07G1.2		GTTACTATGAGAAC				7	Unknown
C45B11.4	*srh-74*	GTTTTTATGATAAC				7	Serpentine Receptor, class H
F02D8.1		GTTACCATGACAAC				7	Unknown
F41H10.2		GTAACCATGGAAAC				7	Contains similarity to Homo sapiens TTF2 protein
F59B10.2		GTATCCATGACAAC				7	Unknown
T28F3.6		GTTGCTATGGCAAC			+	7	Unknown
T12A2.15b		GTAGTTATAAAAAC				51	Contains C2 domain
C18E9.10		GTTTCTATGATAAC				53	Membrane protein involved in ER to Golgi transport
F13H10.6		GTACTCATGAAAAC				53	Unknown
Y110A7A.20		GTCTCTATAGCAAC				61	Component of IFT complex B
W10C8.1		GTTTCCATGAGAAC				65	Src homology domain 3
R01H10.6	*bbs-5*	GTCTCCATGGCAAC	+	+	+	66	Orthologous to human BBS5
Y37F4.2		GTTGTCATGGAAAC				66	Unknown
C26B2.4		GTAACTATGGAAAC			+	70	Nuclear hormone receptor
R148.1	*xbx-7*	GTCACCATAGGAAC			+	70	Unknown
F46F2.5		GTAACTATGGCAAC				71	Unknown
F59C6.7	*che-13*	GTTGCTATAGCAAC	+	+	+	75	Component of IFT complex B
C04C3.5	*dyf-3*	GTTTCTATGGGAAC	+	+		75	Contains coiled-coil domains
T19A5.4	*nhr-44*	GTCTTCATGGCAAC				77	Nuclear hormone receptor
C26B2.8		GTACCCATGAGAAC				78	Nuclear hormone receptor
D1009.5	*xbx-2*	GTTGCCATGACAAC	+	+		78	Dynein light chain (DLC)
F02D8.3	*xbx-1*	GTTTCCATGGTAAC	+		+	80	Dynein light intermediate chain (DLIC)
F28H6.3		GTCTTTATAGAAAC				81	Aminotransferase class 1
C48B6.8		GTTTCCATGACAAC			+	82	Contains WD40 domain
C23H5.3	*xbx-4*	GTTGCCATGACAAC				83	Unknown
K08D12.2		GTTTCCATAGCAAC	+			85	Human XRP-2
T25F10.5	*bbs-8*	GTACCCATGGCAAC	+		+	85	Orthologous to human BBS8
Y37F4.4		GTTTCCATGACAAC				85	Unknown
T27B1.1	*osm-1*	GCTACCATGGCAAC	+	+	+	86	Component of IFT complex B
ZC132.3		GTTGCTATGGAAAC			+	86	Unknown
C06E2.1		GTTTCTATGGAAAC				93	Contains Zn finger domain, C2H2 type
F20D12.3	*bbs-2*	GTATCCATGGCAAC	+	+	+	94	Orthologous to human BBS2
R31.3	*osm-6*	GTTACCATAGTAAC	+	+		100	Component of IFT complex B
Y105E8A.5	*bbs-1*	GTTCCCATAGCAAC	+	+	+	100	Orthologous to human BBS1
Y75B8A.12	*bbs-7*	GTTGCCATAGTAAC	+		+	108	Orthologous to human BBS7
Y4C6B.6		GTCTTTATAAAAAC				109	Homolog of the human gene GLCM
F33H1.1	*daf-19*	GTTTCCATGGAAAC			+	110	Member of the RFX-type transcription factors
T02G5.3		GTAGCCATAGCAAC				110	Unknown
Y37E3.5		GTAACTATGGCAAC				116	GTP-binding ADP-ribosylation factor-like protein ARL3
Y41G9A.1	*osm-5*	GTTACTATGGCAAC	+	+		116	Component of IFT complex B
T24A11.2	*xbx-5*	GTCTCCATGACAAC				122	7-transmembrane receptor
C15C8.1		GTAACCATAGCAAC				124	Unknown
F19H8.3	*arl-3*	GTCTCTATGGTAAC		+		152	Orthologous to human Arl-3
F40F9.1	*xbx-6*	GTTTCCATGGAAAC				152	N-methyl-D-aspartate receptor glutamate-binding subunit
Y45F10D.15		GTTTCCATGGAAAC				158	Unknown
F59E11.8		GCTGCCATAACAAC				159	Contains Zn finger domain
H01G02.2		GTCTCCATGACAAC			+	161	Protein kinase 7
F30F8.7		GTTTTCATAAAAAC				171	Unknown
Y58A7A.3		GTTGTCATGACAAC				173	Contains Zn finger domain, C6HC type
F14D7.8		GTTTCTATAGCAAC				178	Unknown
Y73B3A.18b		GTTATTATGGGAAC				178	Unknown
F56A3.4	*spd-5*	GTTTCCATAGTAAC				182	Member of Spindle Defective group

C14F5.1		GTTGCTATAACAAC			188	BCL2/adenovirus E1B 19-kDa protein-interacting protein 3
C02H7.1		GTCTCCATGACAAC	+	+	196	Ortholog of human MIP-T3
E03H12.2		GTTTCTATGAAAAC			204	C-type lectin
C04C11.1		GTTTTTATAGTAAC			205	SCP extracellular protein
F13H8.3		GTTGCCATGAAAAC			212	Predicted inosine-uridine preferring nucleoside hydrolase
F49E12.8		GTATCCATGGAAAC			224	Similar to H. Sapiens testis specific ZFP91 protein
T10C6.2		GTAACTATAAAAAC			233	7-transmembrane receptor
F55D12.1		GTTACCATAGTAAC			234	Unknown
K05F1.5		GTATCTATGGCAAC			237	Contains Zn finger domain, C2H2 type
F28B3.1		GTCGTTATAATAAC			238	Contains AMP binding domain
F56H9.4	*gpa-9*	GTTACCATGGAAAC			238	G alpha protein
F41E7.9		GTTACCATAACAAC			241	Unknown
C15F1.7		GTTTCTATAGAAAC			251	Cu/Zn superoxide dismutase
C35D10.3		GTATTTATAGCAAC			258	Unknown
F53A2.4	*nud-1*	GTATCCATGGGAAC			264	Orthologous to the Aspergillus nidulans nudC
T28H11.4	*pes-1*	GTTTTCATAAGAAC			271	Transcription factor of the Forkhead/HNF3 family
ZK265.6		GTTTTTATGAGAAC			273	Unknown
ZC404.9		GTTCCTATAAAAAC			287	Mitogen-activated protein kinase(MAP4K)
B0511.13		GTCTTCATAATAAC			303	Calcineurin-like phosphoesterase
Y94A7B.2		GTTTTCATAATAAC			312	Unknown
C24B9.8	*str-13*	GTTTTCATGGAAAC			329	7-transmembrane receptor
C26E6.11		GTAATTATAATAAC	+		363	Unknown
H35N03.1	*exp-1*	GTATTTATGAAAAC			366	GABA receptor
H41C03.3		GTTTCCATGACAAC			410	Branching enzyme
C09E8.1b		GTTGCCATGATAAC		+	411	Sodium-neurotransmitter symporter
F59E12.8		GTATTTATAAAAAC			420	Glutamate-gated kainate-type ion channel receptor subunit
K08C7.7		GTTCTTATGGTAAC		+	451	Contains an F-box domain
W03F9.4		GTTTTTATGAGAAC			472	Carnitine O-acyltransferase CPTI
C04C3.3		GTTCCCATAGAAAC		+	491	Pyruvate dehydrogenase E1, beta subunit
C08A9.6		GTTATCATAAAAAC			510	Unknown
R09H10.6		GTATTTATAACAAC			517	Unknown
T28A11.6		GTTGTTATAGTAAC			522	Unknown
F35D2.5a	*syd-1*	GTCACTATAACAAC			532	Contains PDZ domain
M163.5		GTTGTCATGGCAAC			549	Unknown
F53F4.13		GTTTTTATGGCAAC			553	Unknown
F08F3.2a	*acl-6*	GTATTTATGGTAAC			566	Acyltransferase-like protein
F54E7.7	*rcn-1*	GTTATTATAAGAAC			586	Regulator of Calcineurin
C17G10.9b		GTTTCTATGAAAAC			593	RNA polymerase I-associated factor - PAF67
Y39B6A.20	*asp-1*	GTTTTTATAAAAAC			648	Homolog of cathepsin D aspartic protease
T02E9.1		GTATTCATAATAAC			650	7 transmembrane receptor
C42D4.5	*str-1*	GTATTCATAGAAAC			663	7-transmembrane receptor
F28C6.1		GTAACCATAAAAAC			666	Transcription factor AP-2
T06E4.7		GTTATTATGAAAAC			678	7 transmembrane receptor
Y79H2A.6	*arx-3*	GTCTCCATGGGAAC			704	Actin-related protein Arp2/3 complex
R07B5.9		GTTCTTATAGGAAC			717	Unknown
F07A5.4		GTTATTATAATAAC			721	Unknown
F46F2.5		GTCTCCATGAAAAC			727	Unknown
Y38H6C.14		GTTTCCATAACAAC			730	Unknown
K11D2.1		GTTCTTATAAAAAC			740	Contains RCC1 domain
F48G7.8		GTTTTTATAAAAAC			745	Secreted surface protein
T01G1.2		GTTCTCATGAAAAC			749	Unknown
C48B6.3		GTTTTTATGAAAAC			752	Unknown
C48B6.6	*smg-1*	GTTTTCATAAAAAC			771	Suppressor with Morphological effect on Genitalia
F46H5.7c		GTTATCATAGAAAC			774	Unknown
F56A3.3b	*npp-6*	GTTACTATGGAAAC			778	Nuclear Pore complex Protein
B0310.6		GTTACTATAAAAAC			792	Unknown
F14D7.11		GTTCTCATAAAAAC			792	Unknown

F10C2.6	*drs-2*	GTCACTATGAAAAC				808	Aspartyl(D) tRNA Synthetase
F43G6.11b		GTATTTATAAGAAC				818	Histone deacetylase complex, catalytic component HDA1
F41C6.4		GTATCTATAATAAC				831	Unknown
Y39A1B.3	*dpy-28*	GTTACCATAACAAC				833	Required for negative regulation of X-chromosomal genes
F22D6.7		GTCCCCATAATAAC				843	Unknown
C34C6.3		GTAACTATGAAAAC				848	Unknown
R05H10.5		GTTTCCATGGAAAC				850	Glutathione peroxidase
W03D8.1		GTACCTATAAAAAC				853	Unknown
Y40B10A.5		GTTTTTATGAAAAC				901	Unknown
T07A9.9b		GTTTCTATGAAAAC				906	GTP-binding protein CRFG/NOG1 (ODN superfamily)
F44G3.10		GTATTTATAAAAAC				909	Encodes a claudin homolog
W02B12.2	*rsp-2*	GTAGCCATGACAAC				918	SR Protein (splicing factor)
C45B2.1		GTAATTATGGTAAC				956	Unknown
C15B12.7b	*cdf-1*	GTAACTATAAAAAC				963	Cation Diffusion Facilitator family
T20B12.5		GTCCCCATAATAAC				967	Unknown
F22F1.1	*hil-3*	GTATTTATAATAAC				974	Histone H1 Like
M04C9.5		GTTACCATAGAAAC			+	980	MAPK related serine/threonine protein kinase
F42G2.4b		GTTTCCATGACAAC				981	Contains an F-box domain
E03H4.4		GTCATTATAGTAAC				998	Unknown
F16H11.5	*nhr-45*	ATAACTATGACAAC				7	Nuclear hormone receptor
C49C8.6		ATTGCTATGAAAAC				7	Unknown
F32B4.4a		ATTTCTATGGAAAC				7	Unknown
F57B7.2		GTTCTTTTAGAAAC				15	Defense-related protein containing SCP domain
F54F11.2		GTCATTTTGGCAAC			+	16	Encodes a neprilysin
F53G2.8		ATCCCTATAAAAAC				19	Unknown
F41B4.4b	*grl-6*	GTTTTTTTAAAAAC				28	Glutamate-gated kainate-type ion channel receptor subunit
Y39A1A.16		ATTGTTATAACAAC				29	Contains an F-box domain
T08G3.5	*srh-141*	GTTACTTTAATAAC				32	Serpentine Receptor, class H
C04H5.1		ATTTCCATAAAAAC				34	Unknown
Y46H3C.3	*srw-90*	ATTATCATAAAAAC				39	Serpentine Receptor, class W
E03A3.4	*his-70*	GTTTTTTTGAGAAC				40	Encodes an H3 histone
C17C3.3		ATATTTATAAAAAC				41	Acyl-CoA thioesterase
Y37D8A.13	*unc-71*	ATCATCATAATAAC				42	Member of the metzincin superfamily of proteases
F38B6.6		GTATTTTTGAGAAC				47	Contains TPR repeats
F01D5.5		ATCACTATAAAAAC				49	Secreted surface protein
Y1A5A.1		GTATTCTTGAAAAC				52	Adaptor protein Enigma and related PDZ-LIM proteins
F59B1.6		ATTTCTATGGTAAC				59	Unknown
F35C11.4		GTCACCTTGACAAC			+	62	Unknown
F21D5.2		ATTTCCATGGAAAC				63	OTU (ovarian tumor)-like cysteine protease
C42D4.4	*str-44*	GTTCTTTTAGTAAC				64	7-transmembrane receptor
Y38H6C.13		GTAATTTTAACAAC				76	Unknown
C44E4.1a		GTTGTCTTAAAAAC				79	Zn-binding protein Push
C11H1.7		ATTTTCATGAAAAC				79	Unknown
F14H12.6		GTTCCTTTGATAAC				84	Unknown
K03E6.4		GTTCCCTTGGCAAC	+			85	Unknown
C27A7.4	*che-11*	ATCTCCATGGCAAC	+	+	+	86	Component of IFT complex A
F53A2.2		ATTCCCATGACAAC				90	Unknown
B0395.3		ATCACTATAGTAAC			+	90	Orthologous to the human choline acetyltransferase isoform R
T06E6.5		ATAATTATGACAAC				93	Contains an F-box domain
T24C2.4		GTTTCCTTGGCAAC				97	Contains an F-box domain
ZK488.4		GTCCTTTTAGAAAC				97	Nuclear hormone receptor
M04D8.6	*xbx-3*	GTTGTCTTGGCAAC			+	98	Unknown
F47C12.6		ATCTCCATGAGAAC				101	Unknown
K01C8.6		GTCTCTTTAGTAAC				102	Mitochondrial ribosomal protein L10
Y102A5C.22	*srx-54*	GTTGTTTTGAAAAC				108	Serpentine Receptor, class X
T23F2.3		ATTACCATGGAAAC				114	Stress responsive protein
K01C8.3b	*tdc-1*	ATAACCATAAAAAC				116	Aromatic-L-amino-acid/L-histidine decarboxylase

F35C8.5		ATTTTTATGATAAC				118	C-4 sterol methyl oxidase
F13H8.12		ATCATCATGACAAC				119	Contains Major sperm protein domain
C50H11.9	*str-244*	ATATTTATAAAAAC				119	7-transmembrane olfactory receptor
F38E11.12	*cng-3*	GTTTTCTTGATAAC			+	125	Cyclic Nucleotide Gated channel
C18F10.5	*srg-2*	ATTGTTATGATAAC				127	Serpentine Receptor, class G
C09H5.5	*str-144*	GTTTTCTTAAGAAC				127	7-transmembrane receptor
F38G1.1	*che-2*	GTTGTCATGGTGAC	+	+	+	131	Component of IFT complex B
F29B9.8		GTTTCCATAGCGAC				137	Unknown
Y5F2A.2		ATATTTATAAAAAC				134	Unknown
F32B6.9		GTCTCCTTGACAAC			+	135	Best vitelliform macular dystrophy-associated protein
T07G12.11		ATCATCATGATAAC				138	Contains Zn finger domain, C2H2 type
Y73F8A.11		GTCTCCTTAGTAAC				139	Best vitelliform macular dystrophy-associated protein
C06A12.5		GTTTCTTTAGAAAC				139	Predicted esterase
Y45F10B.5	*sru-12*	GTTATTTTGGAAAC				140	Serpentine Receptor, class U
Y102A5C.26		GTTTTCTTGAAAAC				140	7-transmembrane receptor
C42D4.13		ATTCTTATGGTAAC				144	Unknown
T13F2.3b	*pis-1*	GTTATTTTGAAAAC				147	Ortholog of mammalian Pax transcription activation domain
F48A9.3	*try-6*	GTATCCTTAAAAAC				148	Trypsin
F55D12.3		ATTATCATAACAAC				149	Nuclear hormone receptor
C03G6.3	*srj-14*	GTTCTCTTAAAAAC				150	Serpentine Receptor, class J
Y54G11A.9		GTATCTTTAATAAC				152	Mitochondrial Fe-S cluster biosynthesis protein ISA2
F40B5.3		GTTTTTTTAATAAC				152	Encodes a neprilysin
T10C6.8		ATTTCCATAAAAAC				159	Contains an F-box domain
F35D2.4		ATCACCATGACAAC				160	Unknown
C15C7.6		GTCATTTTGGCAAC				162	Unknown
R07E5.17		ATTACCATGGAAAC				164	Unknown
T25D10.4		GTATTTTTAAGAAC				165	Unknown
T22C8.3		GTTGTCATAGTGAC				168	Contains Zn finger domain, C2H2 type
Y82E9BR.14b		ATTTTCATGAAAAC				168	Glycolipid transfer protein
R13H4.1		ATTTCCATGACAAC	+		+	169	Contains similarity to Homo sapiens Nephrocystin 4
C45G9.11		GTTTTCTTAGAAAC				171	Unknown
C34B2.5		GTTATTTTAAAAAC				179	Contains TPR repeats
C35A11.2		ATCTTTATAAGAAC				179	Unknown
F32A6.5	*sto-2*	GTTTCCTTAAAAAC			+	183	STOmatin family
F10B5.4	*tub-1*	ATCTCCATGACAAC				183	Homolog of mammalian tubby
K10B3.5		ATATTTATGAAAAC				184	Unknown
Y37D8A.17		ATTTCCATAGAAAC				192	Lipocalin-interacting membrane receptor (LIMR)
C29E4.3	*ran-2*	GTCTCTTTGAAAAC				193	RanGAP (GTPase activating protein) homolog
Y46E12BR.1		ATAGTTATGAGAAC				193	Unknown
F35E12.8		GTATCCTTGAAAAC				193	Unknown
F09E10.5		ATTTTTATGAAAAC				193	Unknown
R10H10.3		ATATTTATAAAAAC				196	C-type lectin
F49E12.4	*ubc-24*	GTAGCTTTAAGAAC				197	Encodes a predicted conjugating enzyme (UBCs/E2s)
C01H6.3		ATTCTCATAGTAAC				201	Contains Zn finger domain, C2H2 type
T02D1.8		GTATTTTTAACAAC				202	Unknown
F23B12.4		GTCCCTTTGAGAAC				203	Secreted surface protein
H28G03.2a		ATCTCTATAAAAAC				203	Unknown
ZK1098.7		ATAATTATAAAAAC				205	Unknown
F11A5.6		GTAACCTTGAAAAC				207	Unknown
B0336.3		GTATTTTTAATAAC				212	Contains the RNA recognition motif
Y5H2B.1		GTTTTTTTAAGAAC				215	Predicted glycosyltransferase
C27C12.3		GTCATTTTGGGAAC				223	Unknown
B0547.4	*srd-8*	ATTTTCATAAAAAC				225	Serpentine Receptor, class D
C56E10.4	*nhr-137*	GTTTTTTTAAAAAC				225	Nuclear hormone receptor
C18A11.6		GTAACCTTGAAAAC				226	Unknown
C50C10.8	*sru-33*	GTTTTCTTGAAAAC				229	Serpentine Receptor, class U
F11A5.6		GTAACCTTGAAAAC				230	Unknown

F29D11.2		GTTTCTTTGGAAAC	+	236	Unknown
ZC204.10		ATAATCATAACAAC		241	Contains an F-box domain
C55B7.4a		ATCTTTATGAAAAC		246	Acyl CoA DeHydrogenase
C25A11.4d	*ajm-1*	ATAATCATGGTAAC		247	Apical junction molecule class
T15B7.10		GTATCTTTAAAAAC		259	Unknown
C16D9.6		ATCACCATAGTAAC		263	Galactosyltransferases
C44C1.6		GTCGTTTTGAAAAC		265	Unknown
Y40H7A.10		GTTCTTTTAAAAAC		268	Cysteine proteinase Cathepsin L
C34F11.5		ATTTTTATGAGAAC		272	Protein tyrosine kinase
T23F11.3		GTAGTTTTGAAAAC		273	CDK5 kinase activator p35/Nck5a
T21B4.3		ATTTTCATAGCAAC		274	Unknown
F15C11.2b		ATCACCATGGAAAC		276	Ubiquitin-like protein
T22C1.1		ATTTTCATAAAAAC		278	Contains N-recognin-type Zn-finger
C33F10.5b		GTTTCCTTGACAAC		287	Neural cell adhesion molecule L1
W07G4.6	*srr-1*	GTATTTTTGAAAAC		287	Serpentine Receptor, class R
F38B7.7	*srx-86*	ATTTTTATGGCAAC		288	Serpentine Receptor, class X
T09A5.9		ATTATTATAAAAAC		289	Protein phosphatase 1, regulatory subunit
W04G5.5		ATCGCTATAAAAAC		291	Predicted peptide:N-glycanase
F33H12.5	*sri-36*	ATTTTTATGAGAAC		293	Serpentine Receptor, class I
F25H5.6		ATTTTTATAATAAC		294	Mitochondrial/chloroplast ribosomal protein L54/L37
F55D10.5		GTCATTTTAGCAAC		295	Ligand-gated ion channel
R07B1.5		ATAATCATGAAAAC		296	Unknown
F59A7.6		ATTCCCATGAAAAC		298	Pseudogene
C49F5.3		GTTCTTTTAAGAAC		298	Unknown
F59A7.4	*hil-6*	GTCGTTTTAGTAAC		299	Histone H1 Like
F13C5.2		GTACTCTTGAAAAC	+	304	Transcription initiation factor TFIID, subunit BDF1
C26C6.1		ATATTTATAAGAAC		308	Ortholog of the vertebrate Polybromo protein
F20C5.1a	*pme-3*	ATAACTATAATAAC		308	Poly(ADP-ribose) Metabolism Enzyme
F37A8.5		GTCGCCTTGAAAAC		309	Predicted Yippee-type zinc-binding protein
F23F12.10	*srb-11*	ATTTCCATGAAAAC		312	Serpentine Receptor, class B
T28A11.14		GTTTTCTTAATAAC		312	Unknown
W04G3.9		ATTTCTATAAAAAC		314	Unknown
C55A6.6		ATTGCCATGAAAAC		316	Predicted short chain-type dehydrogenase
Y49C4A.9		ATCTCCATGGCAAC		317	Cytochrome P450 CYP2 subfamily
ZC395.3	*toc-1*	ATTGCCATGGAAAC		318	Transporter Of divalent Cations
B0280.5		GTAATTTTGAAAAC		321	Encodes a protein with a chitin binding peritrophin-A domain
K10D2.4		ATCTCTATGACAAC		323	Unknown
C25B8.6	*nhr-120*	ATAACTATGATAAC	+	323	Nuclear hormone receptor
H06H21.1	*srw-94*	ATATTTATAATAAC		324	Serpentine Receptor, class W
C28D4.4		ATCGCTATAGAAAC		327	Unknown
ZK380.1	*tbx-32*	GTTATTTTAATAAC		331	T-box transcription factor
Y45F10A.3		ATCGCCATAAAAAC		342	Kynurenine formamidase
F45H11.2		ATAGCTATGAGAAC		344	Mouse NEDd8 related
Y71H2AM.18		GTTTTTTTAGAAAC		344	ATP-dependent RNA helicase
F32E10.8		GTCATTTTGAAAAC		347	Low-density lipoprotein receptor
C41G11.4c		ATTTCCATAGAAAC		359	7-transmembrane receptor
Y39C12A.1		GTTTTCTTGAAAAC		365	Unknown
F57A10.3	*haf-3*	ATTCTCATGGGAAC		368	HAlF transporter (PGP related)
H06A10.2	*col-185*	ATATTTATAGAAAC		374	Collagen
F53B2.4	*sru-13*	GTTATTTTGGAAAC		376	Serpentine Receptor, class U
T27E9.1c	*tag-61*	ATTTTCATAAAAAC		383	Mitochondrial ADP/ATP carrier proteins
Y37E11AR.5		GTTACCTTGGAAAC		387	UDP-glucuronosyl and UDP-glucosyl transferase
H34P18.1		ATATTTATAGCAAC		388	Unknown
F59E10.1	*orc-2*	ATTATCATGGAAAC	+	393	Origin recognition complex, subunit 2
F14F3.3		ATAACTATGAAAAC		399	Predicted membrane protein
K07A1.14		ATATTTATGATAAC		403	Unknown
C33B4.4		ATTCTCATGAAAAC		403	Unknown

F25H5.1b		ATTTTCATGAAAAC		404	LIM domain
ZK381.5b		ATAACTATAGAAAC		407	LIM domain
T07H8.5	*srg-31*	GTAATTTTGGAAAC		410	Serpentine Receptor, class G
Y19D10B.5		GTTTTTTTAGGAAC		411	Predicted receptor
T21C9.11		ATCTCTATGGAAAC		411	Unknown
F22B7.1		ATAATTATAAAAAC		413	Unknown
F47H4.10	*skr-5*	ATTATTATAGAAAC	+	415	SKp1 Related (ubiquitin ligase complex component)
C28C12.5	*spp-8*	ATTTCCATAAAAAC		418	SaPosin-like Protein family
Y51A2A.1		GTATTTTTAAAAAC		424	C-type lectin
T21E12.4	*dhc-1*	ATCATTATAAAAAC		432	Dynein Heavy Chain
K09C4.5		ATAATTATGAAAAC		432	Unknown
F40G9.3	*ubc-20*	ATTCCTATAAAAAC		442	Ubiquitin-protein ligase
R11.3		GTAGTTTTAAAAAC		442	Unknown
C37H5.5		GTATTTTTGATAAC		452	Protein involved in the nuclear export of pre-ribosomes
B0240.2		GTACTTTTGGTAAC		452	Unknown
C17B7.2		ATTTTTATAAAAAC	+	454	Predicted cysteine rich protein found only in C.elegans
F02E11.3		GTATTTTTAGAAAC		460	Unknown
Y46E12A.3		ATCATTATAGGAAC		460	Glutaredoxin-related protein
Y5H2B.6		GTAGTTTTAGAAAC		466	Cytochrome P450 CYP2 subfamily
C30A5.9		ATTGCTATGGAAAC		480	Unknown
K10D3.3	*taf-11.2*	ATATTTATGAAAAC		482	TAF (TBP-associated transcription factor) family
F53C3.2		ATTCTCATGAGAAC		483	Contains an F-box domain
W02D7.3		ATTCTCATAAAAAC		484	Unknown
T12B5.11		GTATTTTTGAAAAC		486	Contains an F-box domain
C01B10.10		ATAATTATGAAAAC		488	Unknown
T28F4.2		GTTTTCTTGAAAAC		490	Non voltage-gated ion channels (DEG/ENaC family)
F14D12.6a		GTTTCTTTGACAAC		491	7 transmembrane receptor
T04A8.10		GTTTTCTTGGAAAC		497	Unknown
Y71H2AM.16		ATAATTATAGAAAC		507	Lysosomal & prostatic acid phosphatases
H04M03.12		GTTTTTTTGAAAAC		507	Unknown
T21B4.6	*srh-67*	GTCTTCTTGAAAAC		508	Serpentine Receptor, class H
Y51H4A.1		ATTTTTATAAAAAC		509	Unknown
H09I01.1		ATTTTTATAAAAAC		510	Unknown
T22E5.1		GTATTTTTGGCAAC		514	Unknown
M162.6		GTCATTTTAAGAAC		518	C-type lectin
C17H12.3		ATTTTTATAAAAAC		524	Protein tyrosine phosphatase
K01D12.8		ATCTTTATGAAAAC		526	Unknown
W03F11.5b		GTCATTTTGAAAAC		529	Unknown
Y106G6H.1		ATTGTCATGGAAAC		538	Unknown
W03F8.10		ATCGTTATAATAAC		540	Unknown
K11G12.5		ATTTTTTTAGAAAC		542	Mitochondrial oxoglutarate/malate carrier proteins
C05E4.6	*str-134*	ATTTTTTTGAAAAC		543	7-transmembrane olfactory receptor
K07F5.5		GTTTTTTTGAAAAC		547	Unknown
T26H2.1		ATCTCTATGGAAAC		547	Contains an F-box domain
Y41D4B.9	*nhr-122*	GTACTTTTGAAAAC		551	Nuclear Hormone Receptor
R03E1.2		GTTTTCTTAATAAC		551	ATPase membrane sector associated protein
T04D1.4		GTCTTTTTGGAAAC	+	554	Contains a glutamine/asparagine (Q/N)-rich ('prion') domain
ZK550.3		ATTTTCATGAAAAC		556	Unknown
C50C10.1	*sru-37*	ATTTTCATGAAAAC		556	Serpentine Receptor, class U
Y54G2A.16		ATCTTCATAAAAAC		557	Unknown
F47A4.5		GTACTTTTAACAAC		560	Ca2+-independent phospholipase A2
D2062.5		GTTTTTTTGGAAAC		562	Unknown
Y39A3A.1		ATTGCTATGGAAAC		576	Unknown
T23F2.3		ATTCCCATGGGAAC		581	Stress responsive protein
T01E8.3	*plc-3*	ATTGTTATAGAAAC		589	Phospholipase C
F28G4.1		ATATCCATGACAAC		601	Cytochrome P450 CYP4/CYP19/CYP26 subfamilies
W05H5.3		ATTGCCATAAAAAC		606	Na+/Pi symporter

C33G3.4		GTATCCTTAACAAC		611	Orthologous to the human gene beta-mannosidase
W09G12.9		ATTATTATAATAAC		614	Unknown
C28D4.10		GTAACTTTGAAAAC		614	Unknown
R06F6.8a		ATCACTATGACAAC		616	Contains WD40 domain
K09D9.8	*srh-8*	GTTTTTTTGAAAAC		616	Serpentine Receptor, class H
F13E9.10		ATTATCATGAAAAC		628	Predicted alpha-helical protein
Y50D7A.8		ATTCTCATGACAAC		630	Unknown
K05F1.9		ATTTTTATAAAAAC		633	VAMP-associated protein involved in inositol metabolism
B0495.8a		ATCCCTATGAAAAC		649	Spliceosome subunit
K09E3.6		ATTCCCATAGAAAC		652	Unknown
Y9C9A.2	*str-164*	ATATTTATAGTAAC		653	7-transmembrane olfactory receptor
T13F2.4		ATAATTATAAAAAC		662	Unknown
T19A6.4		ATTTTTATAAAAAC		669	Unknown
F39G3.1		GTCCTTTTAAGAAC		672	UDP-glucuronosyl and UDP-glucosyl transferase
Y55B1AL.3		ATTTCTATGAAAAC		673	DNA polymerase theta/eta, DEAD-box superfamily
F10C2.6	*drs-2*	ATAATTATGATAAC		690	Aspartyl(D) tRNA Synthetase
F09F3.10		GTATCTTTAAAAAC		692	Hormone receptors
Y87G2A.1		ATTGTCATGGCAAC	+	699	Unknown
C46A5.8		ATTACTATAGCAAC		706	Unknown
Y11D7A.13		GTAACTTTGAAAAC		708	Unknown
T10H4.9	*srx-51*	ATATTTATGAAAAC		708	Serpentine Receptor, class X
W08G11.1		GTATTTTTGGCAAC		713	Unknown
Y55F3AM.13		ATTTTTATGAAAAC		719	Unknown
F48C11.3	*nlp-3*	ATTTTCATAATAAC		721	Neuropeptide-Like Protein
C05E11.5	*amt-4*	GTTATTTTAGCAAC		736	Ammonium Transporter homolog
F20H11.6		GTTTTCTTGAAAAC		737	Unknown
F08B4.6	*hst-1*	GTCTTTTTAAAAAC		737	Heparan sulfate N-deacetylase/N-sulfotransferase
F35G12.1		ATCATTATGAAAAC		738	Orthologous to the human gene malonyl-coa decarboxylase
Y1B5A.1		GTTTTCTTAATAAC		744	Unknown
C18H7.6		ATTTTTATAAAAAC		746	Unknown
F49F1.7		ATCATTATGACAAC		746	Secreted surface protein
B0304.6	*sra-33*	GTTTTTTTGGAAAC		749	Serpentine Receptor, class A
B0336.8	*lgg-3*	GTTTTCTTAGAAAC		751	LC3, GABARAP and GATE-16 family
W08D2.3		ATTATCATAGAAAC		751	Unknown
C10F3.2	*dhs-16*	GTAGTTTTAACAAC		757	Dehydrogenases, Short chain
Y66A7A.8	*tbx-33*	ATTTTTATGAAAAC		760	T-box transcription factor
W07A12.4		GTTTTCTTGGGAAC		761	Topoisomerase TOP1-interacting protein BTBD1
F21D12.2		ATTCCCATAAAAAC		762	Unknown
C27D6.9	*srb-2*	ATTTTCATAAGAAC		762	Serpentine Receptor, class B
B0511.3		ATTTTCATGAAAAC		763	Contains an F-box domain
F49A5.5		ATACTCATGAAAAC		764	C-type lectin
F53C11.3		ATATTTATGGAAAC	+	765	Reductases with broad range of substrate specificities
C16E9.2a		GTTATTTTAACAAC		765	Unknown
ZK792.6	*let-60*	GTTTTTTTAAAAAC		768	Member of the GTP-binding RAS protooncogene family
Y39A1A.12		ATATTTATAAAAAC		770	Homolog of origin recognition complex 1
M01A12.3		ATTTTTATGGAAAC		772	Unknown
F43C11.7		ATCACCATAGCAAC		785	Predicted E3 ubiquitin ligase
B0294.3		GTTGCCATGGCGAC		788	Contains an F-box domain
K02E7.12		GTTTTCTTGGAAAC		798	Unknown
C01C10.3	*acl-12*	GTCTTTTTAAAAAC		799	Lysophosphatidic acid acyltransferase LPAAT
C18F10.5	*srg-2*	GTTTTCTTGAAAAC		802	Serpentine Receptor, class G
C18H2.5		ATTACCATGAAAAC		802	Contains BRCT, WSN domains and ankyrin repeats
C02D5.3		ATCTTTATAAGAAC		803	Glutathione S-transferase
F14D2.10		GTTTTCTTAACAAC		807	Alternative splicing factor SRp55/B52/SRp75 (RRM superfamily)
F23F12.11	*srb-10*	ATTTCCATGAAAAC		808	Serpentine Receptor, class B
F22D3.5		ATTGTCATGATAAC		809	Unknown
E02H4.4		GTTTCCTTGACAAC		810	Unknown

W06H8.3		ATATTTATGAAAAC		811	Pseudogene
Y75B8A.7		GTCTCTTTAAAAAC		813	U3 small nucleolar ribonucleoprotein (snoRNP) subunit
C14C11.3		GTTATTTTGGCAAC		817	Unknown
Y38H6C.18		GTCCTTTTGGGAAC		817	Unknown
Y51H7C.2		GTTTTTTTGAGAAC		820	Defense-related protein containing SCP domain
F37C12.16	*srb-9*	GTTTTTTTAAGAAC		822	Serpentine Receptor, class B
W03G11.3		ATTTCCATAAAAAC		825	Unknown
F40H6.2		ATTGCCATGGTAAC		826	Unknown
K08D8.1		ATTCCCATAACAAC		831	Unknown
Y51F10.5		ATATTTATAAAAAC		838	Unknown
T22C1.1		ATCCCCATAACAAC		838	Contains N-recognin-type Zn-finger
K02B12.4		GTACTTTTGATAAC		839	Unknown
C03D6.8	*rpl-24.2*	GTACTTTTAAAAAC		841	Ribosomal Protein, Large subunit
C05B10.1	*lip-1*	ATTATCATGAAAAC		843	Lateral-signal-Induced Phosphatase
K08B4.4		GTTTTTTTGAAAAC		846	UDP-glucuronosyl and UDP-glucosyl transferase
F59B2.12		ATTACTATAAGAAC		848	Unknown
F28C6.5		ATAACCATGAAAAC		852	Unknown
F55A4.1		GTTTTTTTAAAAAC		862	Synaptobrevin/VAMP-like protein SEC22
F46C5.7		GTCACTTTAACAAC		867	Unknown
C07A9.3a	*tlk-1*	ATTTCTATAGCAAC		868	Tousled-Like Kinase
Y55F3AM.14		GTTTTTTTGACAAC		868	Contains Zn-finger domain
F12F6.6	*sec-24.1*	ATATTTATAGGAAC		868	Yeast SEC homolog
K10D2.3		ATCTCTATGACAAC		869	S-M checkpoint control protein CID1
T22B3.3		GTTATTTTGGGAAC		870	VAMP-associated protein involved in inositol metabolism
C07G3.10		ATAATTATAACAAC		872	Unknown
F44G4.2		GTTATTTTGATAAC		880	Unknown
C03B1.6		GTAATTTTGGAAAC		880	Unknown
C03D6.1		ATACCTATAGAAAC		886	Translation initiation factor 2C (eIF-2C)
F54D11.1		ATTTTTATAAAAAC	+	890	SAM-dependent methyltransferases
T22E5.3		GTTGCTTTGAGAAC		893	Unknown
Y66D12A.10		ATTTCTATAAAAAC		898	Unknown
Y7A9A.1		ATTTTCATGGCAAC		898	Gamma-glutamyltransferase
R08A2.7		GTTTTTTTAGAAAC		902	Unknown
R11D1.2		ATCTCCATAGCAAC		904	Predicted transposase
F28B4.2		GTACTTTTGAGAAC		905	Guanine-nucleotide releasing factor
T10C6.5		ATTTTTATGACAAC		907	Unknown
F11D5.6		ATCACTATAAAAAC		912	Unknown
C07H6.1	*lig-4*	ATAGTTATAAAAAC		914	ATP-dependent DNA ligase IV
F07F6.8		ATAACTATAATAAC		916	Unknown
C03A7.10		ATAATTATAATAAC		916	Integral membrane O-acyltransferase
F37C12.17	*srb-7*	GTTCTTTTAAAAAC		917	Serpentine Receptor, class B
H05O09.2		GTTTTTTTAGAAAC		917	Unknown
C23F12.4		GTAGCTTTGAAAAC		917	Unknown
Y69A2AR.24		ATTTTTATGACAAC		925	Unknown
Y69E1A.4		ATTTTTATAGAAAC		926	Serine/threonine specific protein phosphatase PP1
Y111B2A.9a		ATCCTCATAAAAAC		929	Unknown
F53C11.8		GTCCTCTTGACAAC		947	Conserved WD40 repeat-containing protein AN11
B0222.5		ATTTCCATAGAAAC		953	Serine proteinase inhibitor (KU family)
F10G8.4		GTATTTTTAAAAAC		955	Protein tyrosine phosphatase
Y46H3D.8		GTTGTTTTAAAAAC		963	Secreted surface protein
T22F7.1		GTTATCTTGGTAAC		966	Synaptic vesicle transporter SVOP
K09F5.4		ATTTCCATAGCAAC		981	Unknown
F42G4.6		ATCATCATGGAAAC		988	Unknown
Y39C12A.8	*dnj-26*	ATTTCCATAACAAC		988	DNaJ domain (prokaryotic heat shock protein)
R09B3.4	*ubc-12*	ATATTTATGATAAC		989	Ubiquitin Conjugating enzyme
Y25C1A.12	*srg-24*	GTATTTTTAAAAAC		991	Serpentine Receptor, class G
F39H12.2		GTTTTTTTAAAAAC	+	993	Unknown

F40F8.10	*rps-9*	GTCGTTTTAGGAAC		997	Ribosomal Protein, Small subunit
W01B6.8		ATATTTTTAAAAAC		29	Unknown
ZK973.1		ATTTTTTTGATAAC		30	Transcription accessory protein TEX, contains S1 domain
C11H1.4	*prx-1*	ATTTTTTTGAAAAC		30	Encodes a predicted peroxin
ZK1320.12b	*taf-8*	ATATTTTTAAAAAC		33	TBP-associated transcription factor
F07D3.2	*flp-6*	ATTATTTTGAAAAC		34	FMRFamide-like peptide 6
T04B8.2		ATAGTTTTGAGAAC		45	Contains an F-box domain
T19C9.3	*srh-252*	ATCATTTTGAAAAC		45	Serpentine Receptor, class H
C44E4.1a		ATTCTCTTAAAAAC		46	Zn-binding protein Push
T17A3.12		ATTCTCTTAACAAC		48	Unknown
Y49E10.19		ATTTTCTTGGAAAC		48	Actin binding protein Anillin
F25E5.3		ATTATCTTGAAAAC		55	Trypsin
C29H12.1	*rrt-2*	ATCTCCTTGGAAAC		61	Arginyl-tRNA synthetase
Y82E9BR.13		ATTTTTTTAAAAAC		63	Unknown
E03H12.4		ATTTCCTTGGAAAC		69	Unknown
F20B10.3		ATTTTTTTAAAAAC		71	Unknown
F45G2.5	*bli-5*	ATCTTTTTAATAAC		73	Serine proteinase inhibitor (KU family)
F32B5.1		ATTTTTTTAAAAAC		76	Creatine kinase
F59B2.6	*zif-1*	ATATTTTTAAAAAC		91	Zinc Finger-interacting protein
F28C6.2		ATAACTTTAAAAAC		94	Transcription factor AP-2
F23H11.1		ATATTCTTGATAAC		95	PHD Zn-finger protein
K04F1.12		ATCTCCTTGGAAAC		98	Predicted receptor
C56G2.6	*let-767*	ATAATCTTGAAAAC		98	17 beta-hydroxysteroid dehydrogenase type 3
R13F6.2		ATAATTTTAGAAAC		101	C-type lectin
C12D8.10b	*akt-1*	ATTTTCTTAAAAAC	+	101	Serine/threonine protein kinase
W02B12.8		ATTTTCTTAGAAAC		102	CDC42 Rho GTPase-activating protein
E03H12.9		ATATCCTTGGCAAC		107	Unknown
K07F5.16		ATTTTTTTGAAAAC		108	Unknown
W05F2.5		ATCGCTTTAGGAAC		109	Contains an F-box domain
F28B12.1		ATTTTTTTAAAAAC		110	Unknown
T08G5.7		ATTTTTTTAAAAAC		111	C2H2-type Zn-finger
C25B8.1b	*kqt-1*	ATTGTCTTAGGAAC	+	113	Potassium channel, KvQLT family
F14F11.1e		ATTCTTTTAAAAAC		115	Voltage-gated K+ channel KCNB/KCNC
T21G5.2		ATTATTTTAAAAAC		116	Unknown
T28B8.5		ATATCCTTGGGAAC		118	Non voltage-gated ion channels (DEG/ENaC family)
F59B2.6	*zif-1*	ATTTTTTTAAAAAC		130	Zinc Finger-interacting protein
C10C6.3		ATATTCTTGAAAAC		130	Unknown
T04A8.5		ATAACTTTAAAAAC	+	131	Glutamine phosphoribosylpyrophosphate amidotransferase
F01F1.5	*dpf-4*	ATTGTTTTGGAAAC		131	Dipeptidyl Peptidase Four (IV) family
T08A9.12	*spp-2*	ATTATCTTGAAAAC		134	Orthologous to the human gene interferon gamma receptor 2
F40G12.1		ATTTTCTTGGAAAC		135	Sre G protein-coupled chemoreceptor
F17H10.2		GTTATTTTGGAGAC		140	Unknown
ZC239.17		ATTTTTTTAAAAAC		141	Polymerase delta-interacting protein PDIP1
Y47D3A.16		ATCTCTTTGGAAAC		148	Ribosomal protein S6 kinase
ZK546.11	*gst-30*	ATCATTTTGGCAAC		149	Glutathione S-transferase
M04D8.7		ATTTTCTTGAAAAC		149	Unknown
F20B6.4		ATTTTTTTGATAAC		149	Unknown
R04A9.3		ATTTCTTTAAAAAC		153	Unknown
F32D8.2		ATATTTTTAACAAC		162	Unknown
Y61B8A.1	*srh-116*	ATATTTTTGAGAAC		162	Serpentine Receptor, class H
R10D12.7		ATAATTTTGATAAC		169	Stress responsive protein
F40C5.3		ATCACTTTGAGAAC		181	Ground-Like domain
Y53G8B.1		ATTTTTTTGATAAC		182	Glutathione S-transferase
R05D3.2		ATTTTTTTTAAAAAC		184	Lipocalin-interacting membrane receptor (LIMR)
M03E7.5		ATTATTTTAAAAAC		195	Golgi SNAP receptor complex member
Y57A10C.5		ATTTTCTTAAAAAC		206	Sre G protein-coupled chemoreceptor
T21C12.2	*hpd-1*	ATCCCTTTAAAAAC		207	Encodes a 4-hydroxyphenylpyruvate dioxygenase

Y54F10BM.11		ATAACCTTAAAAAC			208	Unknown
F09C6.8		ATTTTTTTGGAAAC			209	Hormone receptor
C47A10.1	*pgp-9*	ATTACTTTAAAAAC			209	Multidrug/pheromone exporter, ABC superfamily
F53B6.1	*tsp-15*	ATTACTTTAAAAAC			211	Tetraspanin family integral membrane protein
T13H10.2		ATTTTTTTAAAAAC			213	Unknown
F35D11.8		ATATTCTTGACAAC			215	C-type lectin
C41C4.1		ATCTTTTTAATAAC			215	Chondroitin 6-sulfotransferase and related sulfotransferases
H12D21.10		ATCACCTTGAGAAC			215	Unknown
F25A2.1		ATTCCCTTGATAAC			218	Predicted lipase
D2045.6		ATTTTTTTGAAAAC		+	219	Cullin
C15H11.7	*pas-1*	ATATTCTTGATAAC			222	Encodes a type 6 alpha subunit of the 26S proteasome's
ZC504.4a	*mig-15*	ATAATCTTGGAAAC			226	Nck-interacting kinase (NIK)
Y51H7BR.4		ATATTTTTAAAAAC			228	Unknown
C05E4.2	*str-20*	ATCTTTTTGACAAC			234	7-transmembrane olfactory receptor
F40D4.3	*srh-159*	ATATTTTTGGCAAC			240	Serpentine Receptor, class H
Y41G9A.5		ATTTTTTTAGAAAC			243	Unknown
R107.7	*gst-1*	ATCATCTTAACAAC	+		247	Glutathione S-transferase
T08G3.12	*srh-138*	ATATTTTTGGTAAC			249	Serpentine Receptor, class H
T27A1.5		ATTTTTTTAAAAAC			252	Amino acid transporters
T01C1.2		ATTCTTTTGAAAAC			254	Unknown
F09C3.1	*pes-7*	ATTTTTTTAAAAAC			255	Ortholog of the vertebrate IQGAP protein
C02E7.3	*srh-20*	ATTACCTTAATAAC			258	Serpentine Receptor, class H
C38D4.5		ATTATCTTAACAAC			263	Predicted Rho GTPase-activating protein
T08G5.8		ATTTTTTTGAAAAC			275	Unknown
R02F11.3		ATTTTTTTAAAAAC			285	Mitochondrial Fe2+ transporter MMT1
C49C8.2		ATTTTCTTGATAAC			288	Unknown
K07F5.5		ATAATTTTAAAAAC			293	Unknown
Y6E2A.8		ATTGTCTTGAAAAC			295	Unknown
C08F1.5		ATTGTTTTGACAAC			297	Contains BTB/POZ domain
F52D2.4	*gei-12*	ATAGTCTTGAAAAC			299	Encodes a novel protein that affects embryonic viability
ZK742.4		ATTCCTTTAGTAAC			310	NADH:flavin oxidoreductase/12-oxophytodienoate reductase
F01G12.2b	*sur-7*	ATTTTTTTGGCAAC			310	Suppressor of activated let-60 Ras
C01G12.3		ATATTTTTAAAAAC			311	Unknown
F46F5.9		ATTATTTTAATAAC			312	Unknown
Y105E8A.1		ATTTTTTTAAAAAC			316	Unknown
F57B10.8		ATTATTTTAGAAAC			319	TBP-binding protein, activator of basal transcription
C05B5.1		ATTTTTTTAACAAC			319	Unknown
C53D5.4		ATAATTTTAAAAAC			321	Unknown
Y57A10A.9		ATATTTTTAAAAAC			321	Unknown
E04F6.2		ATTTTTTTGGCAAC	+		323	Unknown
Y53F4B.16		ATTTTTTTGGAAAC			327	Unknown
F25B4.5		ATTTCCTTGATAAC			328	mRNA processing protein
B0244.11		ATAATTTTGAAAAC			331	Unknown
C08B6.4		ATAGCCTTGGCAAC			333	Unknown
Y39C12A.1		ATCGCCTTGAGAAC			339	Unknown
H28G03.2a		ATTTCCTTGAAAAC			341	Unknown
T08B6.9		ATAATTTTAGGAAC			348	Splicing factor RNPS1, SR protein superfamily
K07A3.2	*ptr-12*	ATTTTTTTAGAAAC			349	Patched Related family
R13H9.4	*msp-53*	ATCATCTTGAGAAC			358	Contains Major sperm protein domain
R13H9.2	*msp-57*	ATCATCTTAAAAAC			361	Contains Major sperm protein domain
F36H12.7	*msp-19*	ATCATCTTGAGAAC			361	Contains Major sperm protein domain
F42C5.4		ATATTTTTAAAAAC			361	Similar to ARD GTP-binding proteins
C24H11.1		ATAGCTTTGATAAC			362	Serine/threonine specific protein phosphatase PP1
Y41E3.15		ATCGTTTTAAAAAC			364	7-transmembrane receptor
Y39C12A.1		ATTTTCTTGAAAAC			366	Unknown
K09A11.3		ATTTTTTTAAAAAC			379	Cytochrome P450 CYP2 subfamily
T10D4.7		ATATTCTTGAAAAC			381	Extracellular protein with cysteine rich structures

F09C6.10		ATTGTTTTGGGAAC		382	Unknown
F43G9.6		ATCCTTTTGAAAAC		384	Fertilization defective (abnormal sperm)
C02A12.6		ATTTCCTTGAAAAC		387	Unknown
F53B6.4		ATTTTTTTAAAAAC		388	Contains Major Sperm Protein domain
Y49E10.9		ATAATTTTAACAAC		396	Transporter, ABC superfamily
Y45F10D.11		GTAATTTTAATAAC		398	Unknown
Y45G5AM.5		GTAATTTTGGAAAC		398	Unknown
B0024.2	*col-150*	ATAATTTTAGAAAC		404	Collagens (type IV and type XIII)
T05C1.4c		ATCACCTTGGAAAC		405	CAMTA family of calmodulin-binding transcriptional activators
F35F11.3		ATTTTTTTGATAAC		405	Contains an F-box domain
T04C12.1		ATATTTTTGAAAAC		406	Unknown
T22C8.3		ATAATTTTGGGAAC		414	Zn-finger domain
AH9.1		ATATTTTTGAAAAC		414	7 transmembrane receptor
K08F4.3		ATAATTTTAAAAAC		416	Translocon-associated complex TRAP, beta subunit
C31H5.7		GTTCTTTTAACAAC		418	Unknown
F42H10.2		ATTTTCTTAGTAAC		418	Unknown
Y119D3B.14		ATTTTTTTAAAAAC		420	Elongation factor G
F14D7.9		ATTGTTTTAACAAC		420	Unknown
T26E4.2		ATCGCCTTAAGAAC		423	Secreted surface protein
F52A8.5		ATTGTTTTGGAAAC		428	Orthologous to the human glycine cleavage system protein
T27C4.3	*str-30*	ATTTTTTTGGAAAC		429	7-transmembrane olfactory receptor
F23B12.8	*bmk-1*	ATTTTTTTAAAAAC		432	BiMC related Kinase
R13H8.1a	*daf-16*	ATCATTTTGATAAC		434	Transcription factor of the HNF-3/forkhead family
ZK455.8b		ATTTTTTTAACAAC		439	Synaptic vesicle transporter SVOP and related transporters
ZK262.3		ATAGTTTTAAGAAC		443	Predicted lipase
F59D6.5	*srw-8*	ATATTTTTAAAAAC		447	Serpentine Receptor, class W
F19B10.6		ATATTTTTGAAAAC		450	Unknown
Y46G5A.20		ATATTTTTAAGAAC		452	Unknown
C01H6.4		ATCGTTTTGAAAAC		461	Flavin-containing monooxygenase
K02F2.5		ATCCCTTTGGGAAC	+	469	Unknown
C32D5.6		ATTACTTTGAAAAC		470	Unknown
F08H9.1	*coh-3*	ATATTTTTAGAAAC		471	Cohesin family
F58E10.2	*end-1*	ATAATTTTAATAAC		473	GATA-like transcription factor
R13F6.6a	*zak-1*	ATAATTTTAAAAAC		477	Mammalian ZAK kinase homolog
F09E5.12		ATTTTCTTGAAAAC		478	Unknown
C14A4.14		ATTATCTTGAAAAC		481	Mitochondrial 28S ribosomal protein S22
C30A5.9		ATTTTTTTAAAAAC		481	Unknown
T09B9.5		ATTTTTTTAAAAAC		482	Unknown
M162.3		ATTTCCTTAGTAAC		483	7-transmembrane receptor
T06C12.7	*nhr-84*	ATATTTTTAAGAAC		490	Nuclear hormone receptor
C18A11.7a	*dim-1*	ATAATTATGATAAC		490	Disorganized Muscle
C35D6.4		ATTTTCTTAGTAAC		491	CCCH-type Zn-finger protein
C36B1.7		ATTACTTTAAAAAC		492	Dihydrofolate reductase
F36A2.12		ATTGTCTTGGTAAC		492	Unknown
F16B4.5b		ATCACTTTGAAAAC		494	Unknown
C24H12.10		ATTTCTTTGAAAAC		503	Contains an F-box domain
VZK822L.2		ATTATTTTGAAAAC		504	Unknown
ZK512.1		ATTTTCTTGGAAAC		513	Unknown
C05D9.7		ATATTTTTAATAAC		524	Unknown
C40A11.9		ATCATCTTAAAAAC		527	Transposon-encoded protein
F52A8.4		ATTTTTTTAAAAAC		530	Unknown
C44H9.1		ATAATTTTAAAAAC		531	UDP-glucuronosyl and UDP-glucosyl transferase
R07C12.4		ATTTTTTTAAAAAC		534	7-transmembrane receptor
K05D4.8	*srz-13*	ATTTTTTTAAAAAC		534	Serpentine Receptor, class Z
F17C8.1	*acy-1*	ATTACTTTGATAAC	+	537	Adenylyl Cyclase
F08F8.8		ATTTCTTTGAAAAC		540	SNARE protein GS28
F38B6.3		ATTTTCTTAAAAAC		550	Unknown

C54F6.5		ATTCTTTTGAAAAC			551	Unknown
F30F8.5		ATCACTTTGGCAAC			553	Unknown
W01A8.5		ATTTTTTTAAAAAC			554	Similar to Homo sapiens sperm associated antigen 1
K02B12.4		ATTACTTTAAAAAC			555	Unknown
C16C8.10		ATCATCTTGAGAAC			562	Unknown
C29F4.1	*col-125*	ATCATTTTAGCAAC		+	564	Collagens (type IV and type XIII)
Y57G11C.42		ATTTTCTTGAAAAC			566	Unknown
F35F11.3		ATAATTTTGATAAC			574	Contains F-box domain
K07F5.5		ATAATTTTAAAAAC			579	Unknown
Y51H4A.13		ATCTCCTTAAAAAC			598	Unknown
F33C8.1b		ATAACCTTGAGAAC			606	Attractin and platelet-activating factor acetylhydrolase
K09D9.13	*srw-89*	ATTTTTTTAAAAAC			607	Serpentine Receptor, class W
EGAP2.2		ATTCTTTTAAAAAC			611	Unknown
T09D3.2	*srg-28*	ATCGCCTTGAGAAC			615	Serpentine Receptor, class G
B0334.7	*srh-44*	ATTTTTTTGGTAAC			617	Serpentine Receptor, class H
F56F10.2		ATTTTTTTAAAAAC		+	617	Unknown
H35N09.2		ATAGCTTTAGTAAC			618	Unknown
C03D6.3	*cel-1*	ATAATTTTGAAAAC			621	mRNA capping enzyme, guanylyltransferase (alpha) subunit
K10C9.7		ATTTTTTTGGAAAC			623	Unknown
F44D12.6		ATACTCTTGGAAAC			625	Unknown
T22D1.5		ATATTTTTAAAAAC			629	Protein phosphatase 2 regulatory subunit
VC5.3a	*npa-1*	GTCATTTTAGCAAC			629	Nematode Polyprotein Allergen related
C05B5.1		ATTTTTTTGGAAAC			630	Unknown
K08B4.4		ATAATTTTAAAAAC			632	UDP-glucuronosyl and UDP-glucosyl transferase
M03A8.1	*dhs-28*	ATTGTTTTAAAAAC		+	637	DeHydrogenases, Short chain
K01A2.11c	*cbn-1*	ATTTTCTTGGAAAC			641	Calcium Binding protein homolog
T09D3.2	*srg-28*	ATTATCTTGAAAAC			641	Serpentine Receptor, class G
F15A4.13		ATTTTCTTAGCAAC			644	Predicted transposase
F54D11.2		ATTTTTTTAAAAAC			645	Predicted alpha/beta hydrolase
ZK285.2		GTTTTTTTAAAAAC			647	Unknown
M02D8.6		ATAATTTTAATAAC			648	Unknown
H12I19.3	*srz-30*	ATAGTCTTAAAAAC			650	Serpentine Receptor, class Z
W04G3.7		ATATCCTTAGTAAC			655	Unknown
F56E10.1		ATTTTTTTAATAAC			658	Unknown
K04A8.7		ATTTTTTTGAAAAC			660	Reverse transcriptase
EEED8.6		ATTCTTTTAAAAAC	+	+	661	Zinc carboxypeptidase
Y7A5A.8		ATTTTTTTAACAAC			661	Unknown
B0478.3		ATACTTTTAAAAAC			664	Unknown
F07A11.4		ATTTCCTTGATAAC			667	Ubiquitin C-terminal hydrolase
H12I13.1		ATTTTTTTAAAAAC			670	Serine/threonine kinase (haspin family)
D1007.14	*pqn-24*	ATTTTTTTGAAAAC			675	Prion-like-(Q/N-rich)-domain-bearing protein
F55B12.9	*srx-129*	ATATTTTTGATAAC			677	Serpentine Receptor, class X
C48E7.6		ATAACTTTAATAAC		+	678	Proteoglycan
F11G11.8		ATAATTTTGAGAAC			680	Unknown
Y48E1A.1		ATCACTTTGAAAAC			683	RNA polymerase I, large subunit
C38H2.3		ATAATTTTAAAAAC			683	Unknown
T22H2.3	*sri-11*	ATTATTTTGACAAC			688	Serpentine Receptor, class I
K04A8.7		ATTTTTTTGAAAAC			690	Reverse transcriptase
B0285.5	*hse-5*	ATTTTTTTAAAAAC			696	Heparan Sulfate-glucuronic acid-5-Epimerase
K11G12.6		ATAACTTTAGAAAC			701	Predicted membrane protein
ZK822.1		ATTACTTTGGAAAC			703	Unknown
C16C4.8		ATACTTTTAAAAAC			705	Contains BTB/POZ domain
K04A8.7		ATTTTTTTGAAAAC			712	Reverse transcriptase
T24C4.1		ATCCTCTTAATAAC			713	Ubiquinol cytochrome c reductase, subunit QCR2
Y51F10.1		ATTTTTTTAGGAAC			714	Unknown
C07E3.4		ATTGCCTTAGAAAC			719	Protein tyrosine phosphatase
C06G8.4	*srd-7*	ATATTTTTAGTAAC			719	Serpentine Receptor, class D

R03H10.6		ATTTTTTTGAAAAC			721	Single-stranded DNA-binding replication protein A (RPA)
B0478.1a	*jnk-1*	ATCCTTTTGAGAAC		+	729	Jun-N-terminal kinase
ZK6.4		ATTATTTTAGAAAC			733	Nuclear hormone receptor
ZC239.16		ATTTTTTTAAAAAC			736	Polymerase delta-interacting protein PDIP1
F28D1.6		ATCATTTTGAAAAC			736	Unknown
F55A12.3	*ppk-1*	ATATTTTTAAAAAC			737	Phosphatidylinositol-4-phosphate 5-kinase
D1007.13		ATTGTTTTAAAAAC			741	Unknown
F43G6.9		ATATTCTTGAAAAC		+	746	Unknown
W08G11.1		ATTCCCTTGGTAAC			747	Unknown
C06A6.7		ATTATTTTAAAAAC			749	Unknown
F07B7.6		ATCATCTTAAAAAC			753	Transposon-encoded proteins with TYA
C41D11.3		ATCGTTTTGGCAAC			757	Unknown
T01B7.3	*rab-21*	ATTTCCTTGACAAC			762	GTPase Rab21, small G protein superfamily
F37C4.7		ATTATCTTGAAAAC			766	Unknown
K09D9.9		ATTTTTTTAGAAAC			782	Unknown
R05D8.6	*srg-58*	ATAATTTTAGAAAC			783	Serpentine Receptor, class G
C46E10.10	*srh-100*	ATTTCTTTAGAAAC			784	Serpentine Receptor, class H
F36H12.7	*msp-19*	ATATTTTTGATAAC			787	Contains Major sperm protein domain
F39G3.7	*prx-6*	ATTGTCTTAAAAAC			789	Peroxisome assembly factor
F21C3.3		ATTGCTTTAAAAAC			794	Zinc-binding protein of the histidine triad (HIT) family
F13G3.2	*srd-53*	ATAATTTTGAAAAC			796	Serpentine Receptor, class D
C24A1.4		ATTTTTTTGAAAAC			800	Unknown
ZK899.5		ATTTTTTTGGAAAC			801	Unknown
R13H9.4	*msp-53*	ATATTTTTGATAAC			809	Contains Major sperm protein domain
F02E8.5		ATTCTTTTAAAAAC			815	WD40 repeat protein TipD
Y54F10AM.2b	*feh-1*	ATTTTTTTGAGAAC			816	Mammalian FE65 Homolog
T19C3.5		ATTTTTTTGGAAAC			830	BPI/LBP/CETP family protein
K04A8.7		ATTATCTTGAAAAC			837	Reverse transcriptase
F49H12.2		ATCCCCTTAGAAAC			839	Unknown
Y48E1B.11		ATTTTTTTGAAAAC			840	Unknown
K08B12.1		ATTTTTTTAAAAAC			845	Predicted lipase
Y27F2A.5		ATTTTTTTGAAAAC			849	Predicted olfactory G-protein coupled receptor
W09D6.2	*str-261*	ATTCTTTTAAAAAC			850	7-transmembrane olfactory receptor
Y39A1A.23	*hpr-9*	ATCCCTTTAAAAAC			852	Homolog of S.Pombe Rad
T22F7.3		ATATTTTTAGTAAC			853	Serine proteinase inhibitor (KU family)
T01B6.1		ATCTTTTTGAAAAC			859	Unknown
C08F8.7	*rap-3*	ATAATTTTGGTAAC			863	RAP homolog (vertebrate Rap GTPase family)
Y40H7A.9		ATATCTTTAACAAC			867	7-transmembrane olfactory receptor
F59A1.8		ATTCTCTTAAAAAC			870	Unknown
B0412.1a	*dac-1*	ATTTTCTTAAAAAC		+	872	Dachshund transcription factor homolog
Y59A8B.11		ATAATCTTGATAAC			872	Unknown
K01C8.7		ATCCTTTTGATAAC			874	Mitochondrial FAD carrier protein
T07A9.12a		ATAGTTTTAACAAC			875	Predicted membrane protein
K10D3.2	*unc-14*	ATAGTTTTAAAAAC		+	876	Uncoordinated
F19H6.1		ATCTCCTTGGCAAC			879	NIMA (never in mitosis)-related G2-specific protein kinase
Y92C3A.1	*cdh-12*	ATTCCCTTGGCAAC			880	Cadherin family
F53B3.6		ATTTTTTTAAAAAC			880	Unknown
C54D2.5c	*cca-1*	ATATTTTTGGAAAC	+		885	Encodes a low-voltage-activated (T-type) calcium channel
B0495.7		ATAGTTTTAAAAAC			886	Aminopeptidases of the M20 family
K10H10.4		ATACTTTTAAAAAC			889	Unknown
R05G6.1		ATAGTTTTGAAAAC			897	Unknown
F21H7.7	*srh-104*	ATTTTTTTGAAAAC			902	Serpentine Receptor, class H
H21P03.3b	*sms-1*	ATTGTTTTGAAAAC			903	Sphingomyelin Synthase
F09E10.5		ATAACTTTAAAAAC			903	Unknown
ZK973.6		ATTTTTTTAAAAAC			906	Unknown
F58A3.2d	*egl-15*	ATTATTTTGAAAAC			912	Egg Laying defective
C02F5.10		ATAGTTTTAAAAAC			913	Unknown

R08F11.2		ATTTTTTTAAAAAC	913	7-transmembrane receptor
Y24D9B.1		ATCCCCTTAAAAAC	914	Unknown
Y51H4A.13		ATCTCCTTAAAAAC	915	Unknown
B0546.5		ATATTTTTAACAAC	917	Unknown
C53B4.4a		ATATTTTTGAAAAC	919	Predicted protein kinase
T16H12.8		ATTTTCTTGAAAAC	926	7-transmembrane receptor
T23F2.5		ATTTTTTTGAAAAC	931	Stress responsive protein
T05G11.6	*srw-55*	ATTTCTTTAAAAAC	933	Serpentine Receptor, class W
F49H12.6a	*acl-4*	ATAACTTTGAAAAC	933	Predicted phosphate acyltransferase, contains PlsC domain
T18H9.6	*mdt-27*	ATCCCTTTAATAAC	937	Mediator
F53A2.2		ATATTTTTGAAAAC	938	Unknown
C02B8.7		ATCTTTTTAAAAAC	939	Unknown
T07A9.11	*rps-24*	ATTTTCTTAGAAAC	944	Ribosomal Protein, Small subunit
ZC443.3		ATTTTTTTGATAAC	945	Unknown
T08H4.1		ATTTCTTTAAAAAC	950	Predicted guanine nucleotide exchange factor
C27C12.2		ATTGTTTTGAGAAC	955	Contains Zn-finger domain
C01B12.1	*sqt-2*	ATAGTTTTGGCAAC	956	Collagens (type IV and type XIII)
F52D2.3		ATTTTCATAAGGAC	963	Unknown
Y39D8C.1	*abt-4*	ATTTTTTTAAAAAC	963	ABC Transporter family
C02D4.1		ATCATCTTGAGAAC	988	Unknown
K01D12.15		ATTTCTTTAAAAAC	992	Unknown
C06E4.3		ATAATTTTAAAAAC	996	Reductases with broad range of substrate specificities

The table includes 758 matches, obtained with the 'average' consensus (RTHNYY WT RRNRAC).
All matches are sorted according to their nucleotide composition and position upstream of the ATG.
X-box matches that correspond to the 'Refined' consensus are highlighted with red.
X-box matches that differ by one nucleotide from the 'Refined' consensus are highlighted with yellow.
X-box matches that differ by two nucleotides from the 'Refined' consensus are highlighted with green.
X-box matches that differ by three nucleotides from the 'Refined' consensus were not found with these search parameters.
X-box matches that have been found in other searches for candidate ciliary genes are indicated in column (I) for Li et al. (Li et al., 2004), column (II) for Avidor-Reiss et al. (Avidor-Reiss et al., 2004) and column (III) for Colosimo et al. (Colosimo et al., 2004).

Table S2. Expression patterns of some of the *C. elegans xbx* gene candidates, where the X-box match is not conserved in *C. briggsae*

Gene name	X-box distance from ATG	Expression patterns of *C. elegans* genes
prx-1	30	Intestinal cells
flp-6	34	Nerve ring and other neuronal processes
unc-71	42	Excretory gland cells, some head neurons, sphincter muscles
rrt-2	61	Anterior intestine, pharynx, body wall muscles
let-767	98	Intestine
akt-1	101	Widely expressed
W02B12.8	101	Pharyngeal neurons and muscles
F29B9.8	136	Head muscles, coelomocytes, some neurons in the head
C45G9.12	152	IL, OL or BAG
mec-8	172	Intestine, nerve ring, head neurons, tail neurons
ran-2	193	Intestine
mig-15	226	Ventral and dorsal nerve cords, body wall muscles
pes-7	255	Ventral nerve cord of the elongating embryo, all major ganglia, vulva, spermathecal valves
pes-1	271	Embryonic expression
hil-6	299	All nuclei of the soma and the germline
gei-12	299	All somatic cells
sur-7	310	Most tissues with notable exception of intestinal cells
exp-1	366	Intestine and anal depressor muscles, RID, ADE and SABD
dhc-1	432	Embryonic expression
daf-16	434	Expressed in most if not all cells
coh-3	471	Expressed in the germline
dim-1	490	Body wall muscles
syd-1	532	Embryonic expression, nerve processes throughout the nervous system
acy-1	537	Head ganglia, ventral nerve cord, tail ganglia, vulva
rcn-1	586	Hypodermal seam cells, various neurons, vulval epithelial and muscle cells, marginal pharyngeal cells, male tail
asp-1	648	Intestinal cells of the late embryo and early larvae
str-1	663	AWB
hse-5	696	Hypodermal tissues
nlp-3	721	Neurons in the head, pharynx, vulva, and in the intestine
jnk-1	729	Most or all neurons and their processes
let-60	768	Vulva, muscles, hypodermis, intestine, ventral nerve cord
prx-6	789	Embryo, L1, L2, L4 and egg-laying adults
feh-1	816	Neuromuscular structures of the pharynx
lip-1	843	Most somatic cells
unc-14	876	Expressed in almost all neurons at all stages
cca-1	885	Pharyngeal MC motor neuron, head neurons, ventral nerve cord, anterior-most body wall muscles, and cells in the tail region
egl-15	912	Hypodermal and intestinal cells, vulval muscles
cdf-1	963	Expressed in vulval muscles, intestinal cells, and in vulval precursor cells
hil-3	974	All nuclei of the soma and the germline

Expression patterns of genes marked in red were determined in our current work using GFP fusions.

Expression patterns of genes not marked in color were extracted from WormBase (http://www.wormbase.org).

All matches are sorted according to their position upstream of the ATG.