Available online at www.sciencedirect.com
SCIENCE DIRECT®
Developmental Biology 270 (2004) 499–512
DEVELOPMENTAL BIOLOGY
www.elsevier.com/locate/ydbio

Genomes & Developmental Control

Identification of *C. elegans* sensory ray genes using whole-genome expression profiling

Douglas S. Portman[a,b,*] and Scott W. Emmons[a]

[a] *Department of Molecular Genetics, Albert Einstein College of Medicine, Bronx, NY 10461, USA*
[b] *Center for Aging and Developmental Biology and Department of Biomedical Genetics, University of Rochester School of Medicine and Dentistry, Rochester, NY 14642 USA*

Received for publication 2 January 2004, revised 26 February 2004, accepted 26 February 2004

Available online 2 April 2004

Abstract

The three cells that comprise each *C. elegans* sensory ray (two sensory neurons and a structural cell) descend from a single neuroblast precursor cell. The *atonal* ortholog *lin-32* and the E/*daughterless* ortholog *hlh-2* act to confer neural competence during ray development, but additional regulatory factors that control specific aspects of cell fate are largely unknown. Here, we use full-genome DNA microarrays to compare gene expression profiles in adult males of two mutant strains to identify new components of the regulatory network that controls ray development and function. This approach identified a large set of candidate ray genes. Using reporter genes, we confirmed ray expression for 13 of these, including a β-tubulin, a TWK-family channel, a putative chemoreceptor and four novel genes (the *cwp* genes) with a potential role in sensory signaling through the *C. elegans* polycystins *lov-1* and *pkd-2*. Additionally, we have found several ray-expressed transcription factors, including the Zn-finger factor *egl-46* and the bHLH gene *hlh-10*. The expression of many of these genes requires *lin-32* function, though this requirement may not reflect direct activation by *lin-32*. Our strategy provides a complementary foundation for modeling the genetic network that controls the development of a simple sensory organ.

Keywords: *C. elegans*; Male tail; Sensory rays; DNA microarrays; Proneural; *lin-32*; Genetic network; Neural subtype; Polycystins

Introduction

The nematode *C. elegans*, with its well-characterized, simple nervous system and amenability to a variety of experimental techniques, provides an excellent opportunity to address the genetic control of neural development (Thomas and Lockery, 1999). A better appreciation of this control is important both for characterizing conserved metazoan developmental mechanisms and for understanding the human pathology that can arise from defects in cell-fate specification and differentiation. In *C. elegans*, forward-genetic approaches to the study of neural development and function have successfully identified a variety of factors important for the generation of particular neural classes and for the establishment of subtype-specific differentiated characteristics (Ruvkun, 1997). However, this strategy has some inherent limitations. In particular, redundant or pleiotropic developmental mechanisms may not be revealed by such an approach; additionally, mutations that have inconspicuous phenotypes (especially subtle behavioral deficits) can be difficult or impossible to identify in genetic screens.

Here, we report the use of DNA microarrays in a strategy complementary to a forward-genetic approach to better characterize genetic networks operating in the development and function the sensory rays of *C. elegans* males. The rays are tail sensilla, likely both mechano- and chemosensory, used by males during copulation (Emmons and Sternberg, 1997) (Fig. 1A). Each of the 18 rays is composed of two distinct neurons, A-type and B-type (or RnA and RnB), as well as an associated structural cell (RnSt) (Fig. 1C). The two ray neurons differ morphologically and are thought to have distinct sensory modalities, whereas the ray structural

* Corresponding author. Center for Aging and Developmental Biology and Department of Biomedical Genetics, University of Rochester School of Medicine and Dentistry, Box 645, 601 Elmwood Avenue, Rochester, NY 14642. Fax: +1-585-506-1947.
E-mail address: douglas.portman@rochester.edu (D.S. Portman).

0012-1606/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.ydbio.2004.02.020

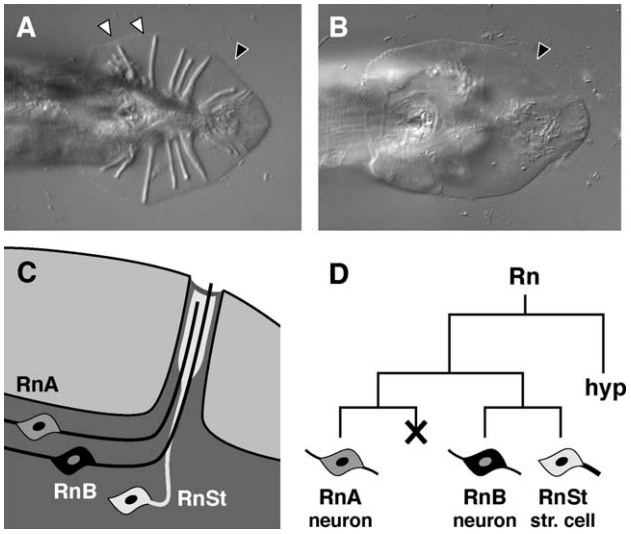

Fig. 1. Ray structure and development. (A) Wild-type adult male tail. Two of the eighteen rays are indicated with white arrowheads; the acellular cuticular fan in which the rays are embedded is indicated with a black arrowhead. (B) The tail of a *hlh-2(bx108);lin-32(e1926)* double mutant adult male. The fan (black arrowhead) contains no rays. (C) Schematic diagram of ray components, for simplicity, shown for only one ray. Each ray is composed of an A-type neuron (RnA), a B-type neuron (RnB), and a structural cell (RnSt). n represents a number from 1 to 9 to denote the cells of an individual ray. (D) The ray sublineage generates the three cells of each ray clonally from a single progenitor, the ray precursor cell (Rn). This sublineage also gives rise to one hypodermal cell (hyp) and a cell that undergoes programmed cell death (Rn.aap, indicated by x).

cell ensheathes the ray neuron endings (Sulston et al., 1980). Diversity among the rays is superimposed on this common structure, such the that rays have a distinctive anteroposterior pattern defined by their morphologies and patterns of gene expression (Chow and Emmons, 1994; Chow et al., 1995; Lints and Emmons, 1999, 2002; Zhang and Emmons, 1995).

The three cells of each ray arise clonally through a common developmental program, the ray sublineage (Fig. 1D), that generates the three cells of each ray from a single ray precursor cell (Rn) (Sulston et al., 1980). The ray precursor cell is specified in third-stage male larvae by the activity of the basic helix-loop-helix transcription factor encoded by *lin-32*, the *atonal* ortholog in *C. elegans* (Zhao and Emmons, 1995). *hlh-2*, the *C. elegans* E/*daughterless* ortholog, can heterodimerize with the *lin-32* product and is also required for ray precursor cell specification (Krause et al., 1997; Portman and Emmons, 2000). The development of individual ray cell types also independently requires *lin-32* and *hlh-2* (Portman and Emmons, 2000), suggesting that these genes promote a state competent for ray differentiation. *lin-32* expression is activated in ray precursor cells by the Hox genes *mab-5* and *egl-5* (Yi et al., 2000; Zhao and Emmons, 1995). Its activity is potentiated by the DM domain gene *mab-3* (Yi et al., 2000), and another DM-domain gene, *mab-23*, acts later in ray development to pattern dopaminergic fates in the rays (Lints and Emmons, 2002). However, little is known about other regulatory factors in ray development, particularly those that may act downstream of or in parallel with *lin-32* in all rays to specify differences among the presumptive RnA, RnB, RnSt, and Rn.aap cells.

Because adult males have 18 rays, the RnA and RnB neurons are the most numerous neural subtypes in these animals. We therefore considered the possibility that microarray comparison of total RNAs between two mutant strains that differ widely in the number of rays (Fig. 1B) might be sensitive enough to identify new components of the ray developmental network. Previous expression-profiling studies in *C. elegans* have exploited differences in gene expression between mutant strains to identify genes specifically expressed in major tissues such as the germline and pharynx (Gaudet and Mango, 2002; Reinke et al., 2000, 2003). However, for less-abundant cell types, this approach can be problematic and was found not to be successful for one neural type, the touch-receptor neurons (Zhang et al., 2002). Alternatives to the mutant-comparison approach have also been successful: microarray analysis of RNA from touch neurons cultured in vitro was used to identify touch-cell-expressed genes (Zhang et al., 2002), and a biochemical mRNA-tagging and purification strategy has identified genes expressed in muscle (Roy et al., 2002). Because in vitro culture of postembryonically derived cells has not been possible (Christensen et al., 2002; Zhang et al., 2002), the first approach is not available for the study of the rays. The mRNA-tagging strategy, although promising, may be problematic with less-abundant classes of cells such as specific neuron subtypes. We therefore chose to determine whether a mutant-comparison approach using RNA from entire animals could be used to identify ray-expressed genes.

Using this strategy, we generated a large set of candidate ray genes. With a reporter gene assay, we directly confirmed ray expression for 13 of these, a set that includes differentiation genes, novel genes, and potential regulatory factors. We conclude that a whole-organism microarray strategy can be used to guide the identification of genes expressed in a particular *C. elegans* neuronal type. Moreover, the set of identified ray-expressed genes (and additional candidate ray genes) provides a powerful framework complementary to traditional genetic approaches for the molecular characterization of the mechanisms of development and function of this model sensory organ.

Materials and methods

Nematode strains and RNA preparation

We constructed the two strains used for microarray analysis, EM672 *hlh-2(bx108)* I; *dpy-28(y1ts)* III; *him-5(e1490)* V; *lin-32(e1926)* X and EM673 *dpy-28(y1ts)* III; *lin-22(mu2)* IV; *him-5(e1490)* V, using strains obtained from the laboratory of S.W.E. and the *Caenorhabditis* Genetics Center (Minneapolis, MN) using standard methods (Brenner,

1974). To prepare large numbers of synchronized adult males, cultures were grown at the permissive *dpy-28(y1ts)* temperature of 16°C on 150-mm NGM plates and fed with an egg yolk–*E. coli* NA22 mixture (Krause, 1995). Crowded plates were shifted to the restrictive temperature of 25°C for 24–36 h to allow manifestation of the *dpy-28* hermaphrodite-lethal phenotype and were fed with additional egg yolk–*E. coli* paste as necessary. Hypochlorite treatment was used to obtain synchronous cultures (Lewis and Fleming, 1995) that were grown to young adulthood. This procedure routinely resulted in populations containing 90–95% males.

RNA preparation and microarray hybridization

Animals were washed several times in M9 and total RNA was harvested using Trizol (Invitrogen) or RNeasy (Qiagen). Poly(A)$^+$ RNA was selected using Oligotex (Qiagen). For each genotype, we isolated RNA from large-scale cultures on three separate occasions. Seven microarrays (two 11,917-gene arrays, Reinke et al., 2000; and five 17,871-gene whole-genome arrays, Jiang et al., 2001) were used in our analysis. The synthesis of labeled cDNA, microarray hybridization, and quantitative scanning of arrays were generously performed by S. Kim, K. Duke, and J. Ryu (Stanford University) as described (Reinke et al., 2000).

Microarray data analysis

For each array spot, we calculated the ratio of the normalized EM673:EM672 channel intensities. Spots with normalized intensities below a threshold of 100 units in both channels were discarded as undetectable. However, if only one channel was below threshold, we set it to a background value of 100 to prevent the ratio from becoming artificially exaggerated. This allowed us to retain genes expressed in one sample but potentially absent or present at very low levels in the other. Data from the seven arrays were merged by individual probes; from the pooled data, we calculated an average normalized ratio value for each gene. Percentile rankings were generated as described in the text. The significance of the change in expression level of each gene was evaluated with the CyberT algorithm (Baldi and Long, 2001); however, because of the high variability in our data even among previously known ray genes, we did not give strong consideration to the resulting *P* values in the selection of genes for expression pattern analysis.

Wormbase data mining

To retrieve gene expression patterns from Wormbase (www.wormbase.org), we used the following two AQL queries: (1) SELECT s,l,e,c FROM s IN CLASS Sequence, l IN s->Locus_genomic_seq, e IN l->Expr_pattern, c IN e->Cell WHERE EXISTS c and (2) SELECT s,e,c FROM s IN CLASS Sequence WHERE NOT EXISTS s->Locus_genomic_seq, e IN s->Expr_pattern, c IN e->Cell WHERE EXISTS c. Using Perl scripts, the output from these queries was merged and manipulated to group all expression patterns both by cell type and by gene. In total, expression patterns from 485 unique genes were retrieved from the WS110 release of Wormbase (September 2003).

Reporter genes

Sequence information and gene predictions were obtained from Wormbase (www.wormbase.org) (Stein et al., 2001) and the Intronerator (www.cse.ucsc.edu/~kent/intronerator) (Kent and Zahler, 2000). Existing annotations were supplemented by analysis using SMART (Letunic et al., 2002) and SignalP (Nielsen et al., 1997). For each reporter, we PCR-amplified a genomic fragment containing upstream sequence (generally at least 3 kb or up to the next flanking gene) and at least several amino acids of coding sequence; some reporters contained nearly full-length coding sequence. This fragment was then fused to GFP or YFP coding sequence using a PCR overlap-extension strategy (Hobert, 1997, 2002). Primer sequences used to generate reporter constructs are available upon request. PCR products were co-injected with pBX1 into a *pha-1(e2123ts)*; *him-5(e1490)* strain (Granato et al., 1994); several transgenic lines were obtained for each construct. We believe these reporters reflect the expression of the endogenous genes reasonably accurately; however, as with all such constructs, they may not recapitulate expression with complete accuracy.

Promoter E-box frequencies

We used a Perl script (Wall et al., 2000) to construct a database of upstream sequences by collecting upstream regions for each predicted gene in the *C. elegans* genome (annotated *C. elegans* genome sequence was obtained from Wormbase). For a given gene, the upstream region was chosen according to the following rule: if the upstream gene was in the opposite orientation (i.e., the intergenic region contained both promoters), the region selected included half of the intergenic distance. If the upstream gene was in the same orientation, two-thirds of the intergenic region was selected. Complex structures (e.g., nested or partially overlapping genes) were excluded from our database. For each promoter region, we calculated the frequency of each E-box (CANNTG) occurrence.

Data availability

The complete microarray data set in both raw and annotated forms is available for download from our laboratory website (http://wormweb.urmc.rochester.edu). Raw data are also available through the Stanford Microarray Database (http://genome-www.stanford.edu/microarray) (Sherlock et al., 2001).

Results and discussion

Microarray strategy

To identify candidate ray-expressed genes, we constructed two strains that differ significantly in the numbers of rays present in males. EM673 males contain the wild-type set of 18 rays (Fig. 1A) as well as several ectopic rays that result from a mutation in *lin-22*, a *hairy/E(spl)*-family gene (Wrischnik and Kenyon, 1997). In contrast, EM672 males are essentially rayless as a result of mutations in the bHLH genes *lin-32* and *hlh-2* (Portman and Emmons, 2000) (Fig. 1B). *him-5(e1490)* and *dpy-28(y1ts)* mutations (Plenefisch et al., 1989) were also present in both strains to allow the preparation of highly male-enriched cultures. We isolated poly(A)$^+$ mRNA from synchronized young adult males of these strains on three separate occasions; fluorescently labeled cDNA derived from these samples was then used for seven microarray hybridizations. For each spot on the array, an intensity ratio measuring the average change between the normalized EM673 and EM672 hybridization signal intensities was calculated. To restrict our analysis to genes that were detected in enough experiments to provide meaningful results, we considered only those that gave detectable signals in at least four of the seven arrays, yielding a set of 14,304 data points. (These points represent 13,305 unique genes since some genes are represented by more than one probe.) For each of these, we generated an average EM673 to EM672 expression ratio. We also determined the significance of reproducibility of the change in signal intensities using CyberT (Baldi and Long, 2001).

Unexpectedly, we found that the set of genes with the highest average ratios was highly enriched for genes known or predicted to be expressed in nematode sperm. Of the 133 genes with the highest average expression ratios (the top first percentile), 62 (47%) were present in a previously described set of sperm-enriched genes (Reinke et al., 2000); moreover, 102 (77%) belonged to the sperm gene-enriched cluster (mountain 4) in a global analysis of *C. elegans* gene expression patterns (Kim et al., 2001). Mountain 4 genes are therefore 11.3-fold over-represented in this set, a highly significant enrichment ($P \ll 0.001$, hypergeometric probability) (Kim et al., 2001). This enrichment could arise from a difference in total sperm number between these strains that is secondary to other phenotypes (e.g., a buildup of sperm in *lin-22* mutants resulting from other defects), or could reflect a more direct role for *lin-32, hlh-2*, or *lin-22* in germline development. Consistent with the latter possibility, both *lin-32* and *hlh-2* are expressed in the distal tip cells of the male and hermaphrodite gonads, and *lin-32* mutants have reduced brood sizes (Krause et al., 1997; Zhao and Emmons, 1995; data not shown). These differences in sperm gene expression, although not necessarily directly relevant to ray development, could reveal interesting functions for these genes and merit further attention.

To focus specifically on somatically expressed genes whose expression differs between our two strains, we took advantage of previous microarray studies in *C. elegans* that identified germline-enriched genes (Reinke et al., 2000) and germline-enriched expression clusters (Kim et al., 2001). This allowed us to remove 3975 potential germline genes from consideration as candidate ray genes, leaving a data set of 9330 unique somatic genes. Although this step almost certainly removed some legitimate ray-expressed genes from our data set, it generated a group more highly enriched for genes relevant to ray development and function.

Using the remaining data, we assigned each point a percentile rank for each experiment in which it gave a measurable signal. For each probe on the full-genome array, we then averaged its rank in each experiment, re-ranked probes by this average, and generated overall percentile scores for each probe. For genes that were represented by more than one probe on the array, we generated an average of this percentile score weighted by the number of measurable points for each probe. The complete data set is available for search and download at http://wormweb.urmc.rochester.edu.

Effectiveness of the approach

Several factors limit the accuracy with which our microarray data alone can identify ray-expressed genes. First, since we are measuring expression levels in the context of whole-animal mRNA (isolated from 1031 somatic and approximately 500 germline cells per animal, Kimble and White, 1981), the noise inherent in array experiments is likely to be significant compared to the relatively low signals we expect from ray-expressed genes. Second, ray-expressed genes with significant expression outside the rays may show relatively low changes in overall expression level. Only one gene, *ram-5* (Yu et al., 2000), is known to be expressed exclusively in rays; all other known ray genes are found in at least several other cells in the body, reducing the expression ratios that we would expect to observe for these genes (assuming that this other expression is not reduced in EM672 males; see below). Third, the *lin-32*, *hlh-2*, and *lin-22* mutations present in our strains affect other cells in addition to the rays: *lin-32* is required for the specification of other neuronal fates (Baumeister et al., 1996; Chalfie and Au, 1989; Portman and Emmons, 2000; Ruvkun, 1997; Shaham and Bargmann, 2002), and both *lin-32* and *hlh-2* have essential embryonic functions (Krause et al., 1997; Zhao and Emmons, 1995; data not shown). Our microarrays may therefore also identify genes expressed in other *lin-32*- or *hlh-2*-dependent cells, or genes repressed by *lin-22*. For these reasons, we do not expect to be able to identify all ray-expressed genes, nor do we expect to identify only ray-expressed genes.

Given these issues, we wished to assess how effective our microarray data would be in identifying genuine ray-expressed genes. We therefore examined the scores generated by our experiments of eight genes with previously known ray expression (Table 1A). These eight genes are those whose expression in adult rays is known to be reasonably specific; we did not consider ubiquitously expressed genes, genes expressed widely in the nervous system, or ray-expressed genes whose predominant expression is in other tissues. We found that none of these eight genes were excluded by the removal of potential germline genes and all met the detection criteria of at least four measurable points among the seven arrays. With regard to overall percentile rank, six of these genes (*ram-5*, Yu et al., 2000; *cat-2*, Lints and Emmons, 1999; *dat-1*, Nass et al., 2002; *pkd-2*, Barr and Sternberg, 1999; *lov-1*, Barr and

Table 1
Microarray results

ORF name	Gene	Cat.	Probe	*n*	Log^2 (Ratio)	*P*	Pctl.	M.E.	Expr.
(A) Previously known ray genes									
B0432.5	*cat-2*	D	B0432.5	4	0.71	0.022	1.1	–	r, n, o
Y73F8A.1	*pkd-2*[a]	D	Y73F8A.B	5	1.33	0.072	1.7	+	r, n
T24C2.1	*ram-5*	D	T24C2.1	6	0.80	0.056	2.0	–	r
T23G5.5	*dat-1*	D	T23G5.5	4	0.87	0.140	2.2	–	r, n
ZK945.9	*lov-1*[a]	D	ZK945.10	5	0.44	0.019	4.7	–	r, n
ZK945.9	*lov-1*[a]	D	ZK945.9	4	0.68	0.290	7.5	–	r, n
F16B3.1	*egl-2*	D	F16B3.1	6	0.51	0.110	7.8	–	r, n, m
F33H1.5	*srd-1*	D	F33H1.5	4	0.35	0.180	10.1	–	r, n
Y73F8A.1	*pkd-2*[a]	D	Y73F8A.A	6	0.90	0.150	12.4	–	r, n
ZK1290.2	*tph-1*	D	ZK1290.2	4	– 0.15	0.041	48.0	–	r, n
(B) Genes tested for ray expression									
C09B9.3	novel	N	C09B9.3	4	1.69	0.001	0.0	–	nd
Y49G5A.1	novel	N	Y49G5A.1	7	1.80	0.000	0.0	–	n, h, i
C37H5.10	*cwp-1*	**N**	C37H5.10	7	**1.67**	0.000	0.0	+	r, n
AH6.3	novel	N	AH6.3	6	**1.06**	0.000	0.1	+	r, n
T10H4.3	*srw-22*	D	T10H4.3	4	0.89	0.007	0.2	–	m
C25F9.1	*srw-85*	D	C25F9.1	5	1.74	0.039	0.3	–	n
F14H3.1	*srj-10*	D	F14H3.1	4	0.74	0.000	0.5	–	n
T28A11.1	*str-64*	D	T28A11.1	5	0.74	0.004	0.6	–	n
C50E10.11	*sre* class	D	C50E10.11	4	0.75	0.021	0.7	–	n
F43G6.6	*spt-3*	T	F43G6.6	5	0.90	0.032	0.8	–	n, m
C37H5.11	*cwp-2*	**N**	C37H5.11	7	**1.03**	0.017	0.9	+	r, n
ZK1251.8	*twk-9*	**D**	ZK1251.8	5	**0.71**	0.007	0.9	–	r, n, cc
C37H5.4	*cwp-3*	**N**	C37H5.4	5	**0.89**	0.032	1.2	+	r, n
B0272.1	*tbb-4*	**D**	B0272.1	6	**0.89**	0.035	1.6	–	r, n
C28A5.4	*ceh-43*	T	C28A5.4	5	0.70	0.037	1.8	–	n
C17H12.9	Onecut-HD	T	C17H12.9	7	**0.81**	0.010	1.8	–	r, n
C03A7.3	*srw-140*	**D**	C03A7.3	4	**0.63**	0.061	1.9	–	r, n
T05G5.2	*hlh-4*	T	T05G5.2	4	0.60	0.011	1.9	–	n
ZK909.4	*ces-2*	**T**	ZK909.4	4	**1.16**	0.180	2.4	–	r, n
F23F12.9	bZIP	T	F23F12.9	5	0.87	0.140	2.9	–	n
ZK682.4	*hlh-10*	**T**	ZK682.4	4	**0.50**	0.016	3.0	–	r, n
Y53C12C.1	HD	T	Y53C12C.1	4	0.72	0.150	3.6	–	nd
K11D12.1	*cwp-4*[b]	N	K11D12.1	7	**0.51**	0.048	6.1	–	r, n
F28C6.1	AP2-like	T	F28C6.1	5	**0.28**	0.000	6.4	–	r, n
K11G9.4	*egl-46*	**T**	K11G9.4	5	**0.50**	0.130	7.1	–	r, n

Microarray statistics and expression patterns are shown for 10 previously known ray-expressed genes (A) as well as for the 25 candidate ray genes we selected from their array scores (B). Candidate genes are sorted by increasing percentile rank; those found to be expressed in rays are shown in bold type. ORF indicates the sequence-based gene name (Stein et al., 2001), Gene gives the gene name or brief description, and Cat. shows the category of the encoded protein (D, differentiation gene; N, novel protein; T, transcription factor). Probe indicates the name of the PCR probe at the relevant spot on the array. *n* gives the number of measurable data points obtained in the seven experiments, log^2 (Ratio) shows the average of the base-two logarithm of the expression ratios, and *P* indicates the significance of the consistency in the change in expression ratio (Baldi and Long, 2001). Pctl. shows the overall percentile ranking. M.E. indicates whether each gene is present in a previously described set of male-enriched genes (Jiang et al., 2001). Expr. indicates the cell types in which we observed expression of reporter construct(s): r, ray neurons or structural cells; n, other neurons; i, intestine; m, muscle; h, hypodermal cells; cc, coelomocytes; o, other; nd, none detected.

[a] *pkd-2* and *lov-1* are both represented by two probes on the microarray; the "weighted-average" percentiles for these two genes are 7.5 and 5.9, respectively, obtained by weighting each percentile score by its *n* value.

[b] K11D12.1 was chosen initially because of a Wormbase annotation error that classified it as a transcription factor (it meets our transcription factor cutoff of eighth percentile). EST and reporter gene evidence suggests that it instead encodes a novel gene; the gene immediately upstream, K11D12.12, is a zinc-finger transcription factor.

Sternberg, 1999; and *egl-2*, Weinshenker et al., 1999) scored in the top seventh percentile, and one additional gene (*srd-1*, Troemel et al., 1995) scored within the top 12th percentile. The remaining gene, *tph-1*, did not show a change in expression, possibly as a result of its expression outside the rays (Sze et al., 2000). That six of these eight genes scored in the seventh percentile represents an 18.8-fold enrichment of ray-expressed genes in the top seventh percentile of our data set (653 genes) ($P \ll 0.001$, hypergeometric probability) and indicates that our approach yielded a data set significantly enriched for ray-expressed genes.

A second way to validate our approach is to compare the frequency with which we can successfully identify new ray-expressed genes with the frequency at which ray-expressed genes would be expected to be obtained by chance. In other words, how valuable was the array data in guiding our selection of candidate ray genes? As discussed below, we chose 25 candidate ray genes based on our microarray data set and found 13 of them to be expressed in rays, a frequency of 52%. None of the genes we tested were found to be ubiquitously expressed. In addition, none of the 13 ray-expressed genes were completely ray specific, but for most of them, the rays are their predominant site of expression. To determine whether selecting candidate ray genes from our microarray data conferred an advantage over selecting similar genes from the entire genome, we examined gene expression patterns reported in the Wormbase database. Genes with expression reported in Wormbase are clearly not a random sample; most represent genes that have been pursued by *C. elegans* researchers because they are likely to have interesting developmental and/or physiological functions. In this sense, this set is similar to our set of candidate genes, which we selected based both on their microarray score and on the potential functional relevance of their products (see below).

We therefore examined the expression patterns of all 485 genes for which expression data is present in Wormbase. Since many *C. elegans* researchers do not examine adult males when determining gene expression patterns, we could not accurately determine the frequency of ray expression among these genes. Instead, we reasoned that frequencies of gene expression in non-sex-specific *C. elegans* sensory neurons would be similar to those in the rays. Clearly, this approach can give only a very rough estimate of specific gene expression frequencies; however, we believe it provides an informative contrast to our results. We analyzed the 485 genes whose expression patterns are reported in Wormbase and calculated their expression in each of 73 *C. elegans* sensory neurons. We found that rates of gene expression varied from 1 in 485 (0.2%, in the neuron IL1DL) to 44 of 485 (9.1%, in each of the ASI neurons). The mean rate of gene expression among these 73 cells was 3.3%. Though this cannot be considered anything more than an approximate estimate of specific gene expression frequencies in these cells, it demonstrates that our rate of 52% is significantly higher than what one would expect to obtain by testing genes without the benefit of our microarray data.

Finally, in a survey of the expression patterns of 22 putative olfactory receptors (including their patterns in adult males) (Troemel et al., 1995), one (*srd-1*) was found to be expressed in the rays, a rate of 4.5%. This rate is consistent with the values obtained from the Wormbase analysis and provides additional independent evidence that our array data did indeed provide significant value to our strategy.

Identification of ray-expressed genes

We used our microarray data to select for expression pattern analysis 25 genes whose expression in the male tail had not previously been studied. We used the scores of the previously known ray genes discussed above (Table 1A) to establish criteria for selection. Genes were selected among three general classes: transcription factor genes, novel genes, and differentiation genes (i.e., genes without obvious potential regulatory roles in development). Putative differentiation genes and novel genes were chosen among those found in the top second percentile. Transcription factors, which occur less frequently in the genome and often have low expression levels (making them more prone to noisy measurements), were selected from the top eighth percentile. We did not take into account the consistency of the change in expression ratio, as P values less than 0.05 were not consistently associated with the set of previously known ray genes (Table 1). Neither did we consider whether a gene was previously found to have higher expression in males (Jiang et al., 2001), as this parameter was not significantly different in the known ray genes either; the reasons for this are not clear since each of these genes is expressed in significantly more cells in males than in hermaphrodites. The 25 genes we tested represent a small subset of those that fulfill our selection criteria (179 unique genes score in the top second percentile), suggesting that our high-scoring data set contains a significant number of additional ray-expressed genes.

For each of the 25 genes we selected (Table 1B), we constructed at least one GFP or YFP reporter (Fig. 2). For each reporter, several transgenic lines were inspected to determine whether ray fluorescence was observed. Using this approach, we identified expression in rays for 13 genes (Table 1); 11 of these are described in further detail below. Of the genes for which we did not observe ray expression, we found reporters for nine to be expressed in other neurons, one (*srw-22*) to be detectable only in muscle, and two to have no detectable expression (Table 1). It is possible that regulatory elements required for ray expression are missing from some of these reporter constructs. Additionally, as discussed above, at least some of these could also be direct or indirect targets of *lin-32*, *hlh-2*, or *lin-22*

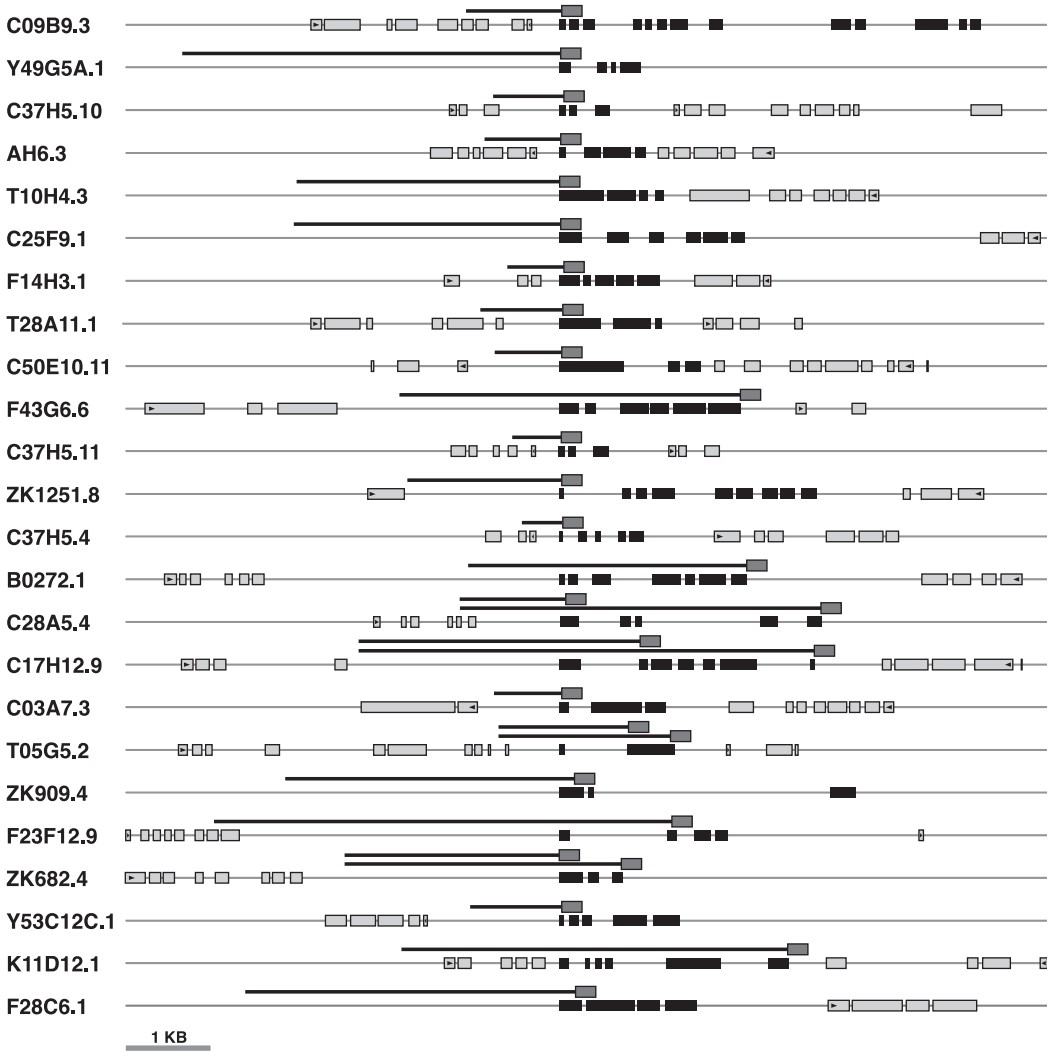

Fig. 2. Reporter constructs. The sequence amplified to create each reporter is shown above the predicted schematic exon–intron structure of each gene. Predicted exons of the tested gene are represented by black boxes. Exons of neighboring genes are shown with light-gray boxes; small triangles indicate the direction of transcription of neighboring genes. All representations are oriented so that the tested gene is in the left-to-right orientation. Dark black lines indicate the region amplified by PCR and fused to GFP or YFP (represented, not to scale, by the small rectangle that terminates each black line). No schematic for K11G9.4 (*egl-46*) is shown as these constructs were generously provided by J. Wu and M. Chalfie. The construct for K11D12.1 (*cwp-4*) included the immediately upstream gene, K11D12.12, as these two genes were predicted by Wormbase to be one large fused gene when our constructs were designed. Subsequent studies (data not shown) have demonstrated that K11D12.12 is a separate transcription unit and does not appear to be expressed in rays.

outside of the rays, though we have yet to obtain strong evidence in support of this possibility (see below).

tbb-4, a β-tubulin isoform

tbb-4 (B0272.1) encodes one of six β-tubulin isoforms in the *C. elegans* genome (Consortium, 1998; Stein et al., 2001). We constructed a full-length *tbb-4::GFP* fusion and found it to be expressed in all 18 ray A- and B-type neurons (Figs. 3A, B). In addition, this construct was expressed in amphid sensory neurons (Fig. 3C) as well as the PDE postdeirid sensory neuron (data not shown) in both males and hermaphrodites. These neurons all contain sensory cilia (White et al., 1986) (though the cilia of ray neurons may be atypical, Sulston et al., 1980), suggesting that the TBB-4 β-tubulin could be a component of ciliary microtubules. Consistent with this idea, the *tbb-4* promoter contains a potential binding site (GCTGCC AT GACAAC) for the RFX-type transcription factor DAF-19, which activates many targets required for ciliated neuron function (Haycraft et al., 2003; Schafer et al., 2003; Swoboda et al., 2000). Notably, the vertebrate β-tubulin isotype most similar to *tbb-4*, β_{IV}, has been shown to be enriched in both motile and nonmotile cilia (Renthal et al., 1993), suggesting that ciliary function could be an ancient conserved role for this β-tubulin isotype. Specific tubulin isotypes have previously been found in another class of neurons in *C. elegans*, the six mechanosensory neurons that express the MEC-12/MEC-7 α-/β-tubulin pair (Fukushige et al., 1999; Hamelin et al., 1992). No other tubulin

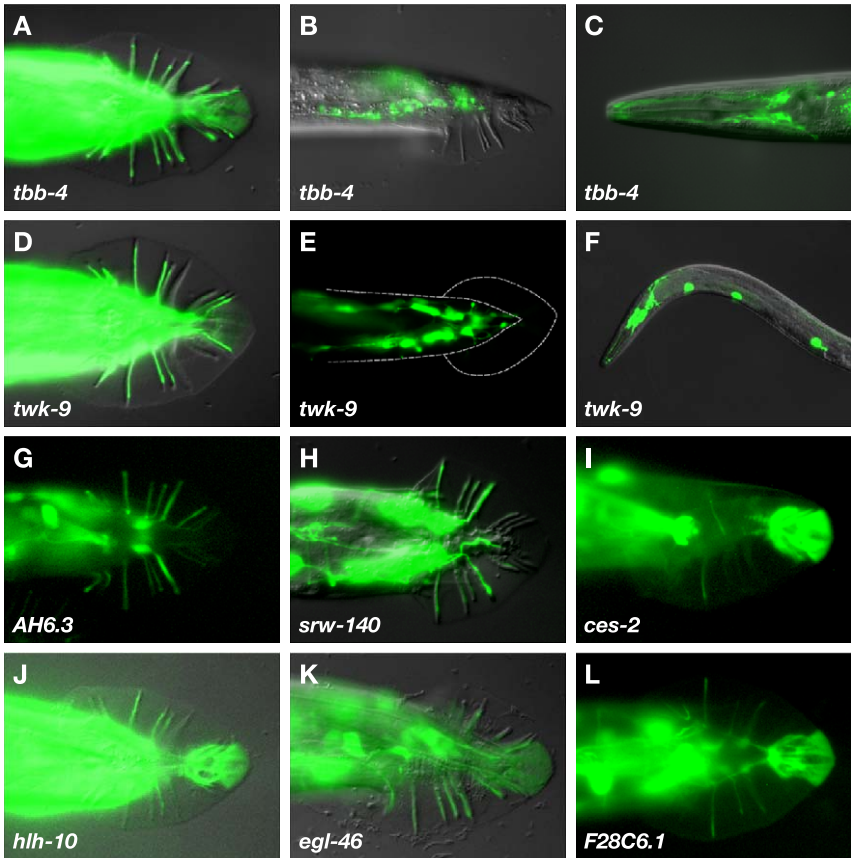

Fig. 3. Newly identified ray genes. Panels show ventral views of an epifluorescence signal (green) overlaid on a DIC image unless otherwise noted; anterior is to the left. (A) *tbb-4::YFP* expression in both neurons of each ray (overexposed to show expression in the rays). This YFP fusion contains nearly all of the *tbb-4* coding sequence and concentrates at ray tips. (B) Lateral view of a shorter exposure of *tbb-4::YFP* showing specific expression in ray cell bodies. (C) *tbb-4::YFP* expression in amphid sensory neurons (lateral view). (D) *twk-9::YFP* expression in ray neurons, overexposed to show ray expression. (E) A shorter exposure showing *twk-9::YFP* expression in ray cell bodies (epifluorescence only, dotted line indicates outline of body and fan). (F) *twk-9::YFP* expression in several head neurons and in coelomocytes (lateral view). (G) A YFP reporter for AH6.3, a gene encoding a novel protein, is expressed in all rays, both in the ray structural cells and in some ray neurons (epifluorescence only). (H) *srw-140::GFP* expression in ray neurons. (I) *ces-2::YFP* expression in ray 3 (ventrolateral view; epifluorescence only). (J) *hlh-10::YFP* expression in rays 2, 3, 4, and 5. (K) *egl-46::GFP* expression in all rays (dorsolateral view). (L) *F28C6.1::YFP* expression in rays 2, 3, 4, and 6 (epifluorescence only).

genes scored above the second percentile; however, two α-tubulins (*tba-8* and *tba-9*) had percentile ranks of 3.49 and 5.98, respectively. A reporter for *tba-8* showed no evidence of expression in the rays (data not shown); *tba-9* has not been tested.

twk-9, a two-pore domain K^+ channel

TWiK or K_{2P} channels are a recently described family of two-pore domain K^+ channels, members of which have been found in a variety of cell types in many organisms (Lesage and Lazdunski, 2000). The *C. elegans* genome contains approximately 45 TWiK channel genes; these are expressed primarily in neurons but are also found in other cell types (Kunkel et al., 2000; Salkoff et al., 1999, 2001). *twk-9* (ZK1251.8) was the only one of these genes to score above the second percentile. We observed expression of our *twk-9* reporter in rays (Figs. 3D, E) as well as several non-sex-specific cells including approximately 14 head and 2 pharyngeal neurons, phasmid neurons, and the coelomocytes (Fig. 3F). Expression of the *twk-9* reporter was not uniform in all rays, and expression in ray 6 was often undetectable; this may reflect genuine differences in endogenous *twk-9* expression or could be a property of the reporter itself. Most of the ray cells expressing the *twk-9* reporter appear to be A-type neurons based on co-labeling with a reporter for the B-neuron-expressed gene *pkd-2* (data not shown). The TWK-9 channel is most closely related to the vertebrate TASK family channels, pH-sensitive channels thought to contribute to background K^+ conductance in a variety of cell types (Lesage and Lazdunski, 2000). Other less closely related TWiK channels (such as TREK-1 and TRAAK) have been shown to be gated by membrane stretch (Lesage and Lazdunski, 2000). Though further experiments will be required to understand the contribution of *twk-9* to ray function, it provides a valuable molecular marker for developmental studies of A-type neuron fate.

A seven-transmembrane chemoreceptor

We tested several G-protein-coupled receptor genes and found ray expression for one, *srw-140* (C03A7.3) (Fig. 3H). *srw-140::YFP* expression is found in most rays, likely in the B-type neuron. The reporter is also expressed in many non-sex-specific neurons elsewhere in the body, including some nonsensory neurons (data not shown). *srw-140* is a divergent member of a large family (approximately 150 genes) of putative chemoreceptors; one possibility is that it may have a role in sensing chemical cues important for male-specific behaviors. Little is known about the *srw* family and, to our knowledge, no functional or expression data have been reported for any other member. In total, nine putative chemoreceptors scored above the second percentile. Of the six tested, *srw-140* was the only one for which we observed clear ray expression; most others were expressed non-sex-specifically in sensory neurons in the head (data not shown). We considered the possibility that some of these genes might have scored highly in our experiments because their neuronal expression might depend on *lin-32* function, however, our examination of the expression patterns of four of these genes in *lin-32* mutants did not reveal clear evidence in support of this possibility (data not shown).

Four ray-expressed transcription factors

Our data set also led us to find ray expression for several transcription factors. *ces-2* encodes a bZip-family transcription factor that is required for specification of programmed cell death in a pair of *C. elegans* pharyngeal cells (Metzstein et al., 1996); recently, *ces-2* has also been shown to be important for regulating *lin-48* in hindgut development, demonstrating that it has roles outside of cell-death specification (Wang and Chamberlin, 2002). We found that a *ces-2* reporter showed weak expression in ray 3 (Fig. 3I), suggesting a potential role for *ces-2* in neuronal subtype specification. Our characterization of *ces-2* mutant males did not reveal any ray defects (data not shown); however, the availability of molecular markers for ray 3 fate is limited.

hlh-10 (ZK682.4) encodes a bHLH protein of the MyoR/ABF-1 family (Ledent et al., 2002). Members of this family have roles in muscle and B-cell development in vertebrates (Lu et al., 1999, 2002; Massari et al., 1998). The HLH-10 protein, also called CeABF-1, has been shown to bind DNA and repress transcription in a mammalian transfection assay (Nguyen et al., 2001). We found expression of *hlh-10::YFP* in many ray neurons, often in rays 2–6 (Fig. 3J), though occasional expression was observed in all rays. ZK682.4 might have a role in patterning ray or ray neuron identities, or a more general role in the development of all rays.

The *egl-46* gene encodes a Zn-finger protein that belongs to a recently described class of factors called ‘‘nerfins’’ that include orthologs in *Drosophila* and vertebrates (Stivers et al., 2000; Wu et al., 2001). In *C. elegans*, *egl-46* function is required to repress touch-receptor cell fates, as touch-cell characteristics are expressed ectopically in *egl-46* mutants (Mitani et al., 1993; Wu et al., 2001). Recently, a role for *egl-46* has been described in the development of the male tail hook neuron HOB (Yu et al., 2003). We found that *egl-46* reporters (generously provided by J. Wu and M. Chalfie) were expressed in the B-type neuron of each ray (Fig. 3K). *egl-46* might act in these cells to promote aspects of ray neuron fate, to repress other fates, or both. *egl-46* function does not seem to be required for expression of the B-type neuron marker *pkd-2* in the ray neurons, though it does function to activate *pkd-2* in the HOB neuron (Yu et al., 2003; data not shown).

The fourth potential regulatory factor for which we detected ray expression, F28C6.1, encodes a member of the AP2 family of transcription factors. We detected expression of an F28C6.1 reporter gene most often in rays 2–6 (Fig. 3L) and non-sex-specifically in sensory and interneurons of the nerve ring (data not shown). To our knowledge, no mutations in F28C6.1 have been isolated. However, *tfap2a*, the zebrafish homolog of F28C6.1, has been shown to be required for the specification of hindbrain noradrenergic neurons (Holzschuh et al., 2003) and for the development of neural crest- and epidermal placode-derived sensory neurons (Knight et al., 2003), suggesting that a role in neural subtype specification could be a conserved, ancestral function for this gene family.

Identification of ray expression for each of these four genes represents the first step toward characterizing their potential roles in ray development and their relationship to *lin-32*. Because we identified these factors without regard to their function, it is possible that their expression during ray development would not have been identified through other means (e.g., forward-genetic or RNAi-based screens). In addition to these four transcription factors, we found that the consistently high scoring gene C17H12.9 (encoding a Onecut-family homeodomain protein) is expressed broadly throughout the adult nervous system of both males and hermaphrodites (data not shown). Since ray expression is a minor component of this gene’s overall expression, and because its expression pattern outside the rays does not change significantly in *lin-32* mutants (data not shown), C17H12.9 likely represents a false positive in our data set. Nevertheless, widespread expression of a Onecut-class gene in the nervous system is of interest; further studies should shed light on the potential role of this factor in neural development and/or function.

Four novel genes coexpressed with the C. elegans polycystins

The ray RnB neurons, as well as the male-specific CEM and HOB sensory neurons, use a polycystin signaling pathway to sense mechanosensory cues important for male mating behavior (Barr and Sternberg, 1999; Barr et al., 2001; Emmons and Somlo, 1999). *lov-1* and *pkd-2*, the *C.*

elegans orthologs of the human polycystin genes *PKD1* and *PKD2*, are required for multiple steps of male mating behavior, including response to the hermaphrodite and subsequent location of the hermaphrodite vulva (Barr and Sternberg, 1999; Barr et al., 2001; Kaletta et al., 2003). In humans, loss of *PKD1* or *PKD2* function results in autosomal dominant polycystic kidney disease (Igarashi and Somlo, 2002). The requirement for the polycystins in these seemingly disparate biological processes likely reflects their underlying conserved role in transducing mechanosensory stimuli in a variety of physiological contexts involving sensory cilia (McGrath et al., 2003; Nauli et al., 2003).

In *C. elegans*, *lov-1* and *pkd-2* are expressed in 21 male-specific neurons: the four CEM head sensory neurons, the hook neuron HOB, and the B-type neurons of all rays but ray 6 (Barr and Sternberg, 1999; Barr et al., 2001; Kaletta et al., 2003). Using our microarray data, we identified four novel genes that share the specific expression pattern characteristic of *lov-1* and *pkd-2* (Fig. 4); we have therefore named them *cwp* genes (coexpressed with polycystins). To our knowledge, no other *C. elegans* genes are known to be expressed in this pattern. All four predicted CWP proteins have signal peptides at their N-termini (Nielsen et al., 1997), indicating that they could be localized to an intracellular compartment or secreted into the extracellular space. These four genes also share an unusual physical arrangement: *cwp-1*, *-2*, and *-3* are located contiguously on the left arm of chromosome V (Fig. 4A), and *cwp-4* lies just 207 kb to the right. *cwp-1* and *-2* appear to have arisen from a recent tandem duplication as their protein products are 83% identical and they share some nucleotide sequence similarity outside the predicted coding region. *cwp-3* is transcribed in a head-to-head orientation with *cwp-2*, suggesting that these genes may share a common upstream regulatory region.

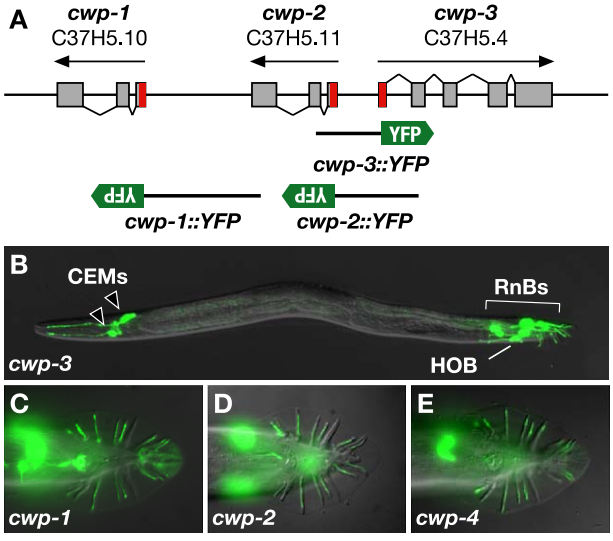

Fig. 4. (A) Schematic diagram of the arrangement of *cwp-1*, *-2*, and *-3* genes on the left arm of chromosome V. Exons are represented by rectangles, transcription units by arrows. Red shading indicates the region at the start of each gene that encodes a signal peptide. Reporter gene constructs are shown schematically below the genomic structure. (B) An adult male expressing *cwp-3::YFP*. The two pairs of CEM head sensory neurons, the HOB hook sensory neuron, and the region containing ray B neuron (RnB) cell bodies are indicated. (C–E) Adult male tails showing ray expression of *cwp-1::YFP*, *cwp-2::YFP*, and *cwp-4::YFP*, respectively. *cwp-4::YFP* expression can also be occasionally observed in ray 6 (as seen here in R6BL), similar to results observed with *pkd-2* reporters (Barr and Sternberg, 1999; Barr et al., 2001).

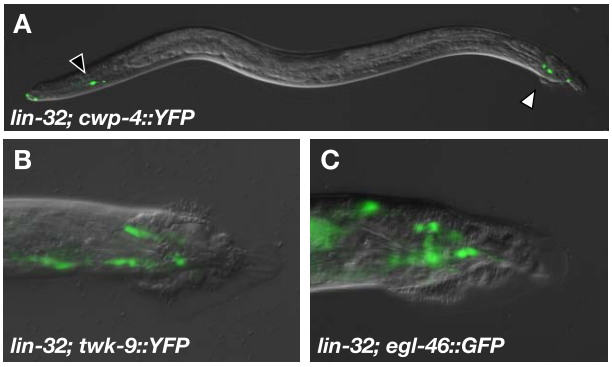

Fig. 5. *lin-32* function is required for ray expression of many ray genes. (A) *lin-32(gm239)* eliminates most expression of *cwp-4::YFP* in the male tail (HOB expression persists; open arrowhead). CEM expression (black arrowhead) is variably affected by *lin-32(gm239)*; cells that express the reporter often have defects in their position and/or in the structure of their processes. (B) Expression of *twk-9::YFP* in the rays requires *lin-32* function. YFP fluorescence in phasmid neurons suggests that this expression may not require *lin-32*. (Since *gm239* is not a null allele, it cannot be concluded that *lin-32* is absolutely not required.) (C) Strong expression of *egl-46::GFP* requires *lin-32* function. Some weak expression of *egl-46* persists, suggesting that its initial activation in the ray lineage might occur independently of *lin-32*.

Aside from their signal peptides, CWP-1, -2, and -3 do not contain any detectable sequence motifs. However, CWP-4, like the extracellular domain of LOV-1, contains a mucin or PTS domain, a serine/threonine/proline-rich region that is a potential target for O-linked glycosylation (Dekker et al., 2002). *cwp-4* is also the only gene of these four for which a reporter gene shows expression outside the set of *lov-1* and *pkd-2*-expressing cells: our *cwp-4::YFP* strain shows expression in several non-sex-specific head neurons in addition to the cells described above, though this expression is much weaker and more variable than that seen in the polycystin-expressing cells (data not shown).

As a TRP-family protein, PKD-2 is likely to act as a cation channel, and its activity may be regulated by the receptor-like LOV-1 protein in response to signals required for mating behavior (Barr and Sternberg, 1999; Barr et al., 2001). The expression patterns of the four *cwp* genes and the mucin domain of CWP-4 suggest a possible relationship between the *cwp* genes and the *lov-1*/*pkd-2* sensory pathway, though further genetic and biochemical studies will be necessary to test this hypothesis. In addition, these genes provide a valuable entry point for understanding the transcriptional specification of RnB-specific characteristics.

Table 2
E-box frequencies

Set	*n*	CANNTG	CACGTG
Detectable somatic genes[a]	9101	7.65 ± 10.02	0.32 ± 0.77
Top second percentile	179	5.36	0.32
Ray-expressed genes	33	7.39	0.21

n indicates the number of upstream regions searched for E-box sequences in each set. Values under "CANNTG" and "CACGTG" indicate the average number of matches to these sequences found per promoter; the range shown for "all genes" represents the standard deviation.

[a] This set represents all genes in our "universe" of eligible genes (those that have at least four detectable array measurements and are not excluded by our removal of germline-associated genes) that also have retrievable upstream regions. This number is slightly lower than the total number of eligible genes (9330) since our algorithm does not retrieve upstream regions from loci with complex structures (see Materials and methods).

E-boxes and lin-32-dependent regulatory mechanisms

As expected from the strategy we employed here (profiling mRNAs from young adult males), all of the ray-expressed genes we identified are expressed in mature, differentiated ray cells. We were interested to know whether this expression in adult males required *lin-32* function, and how direct the involvement of the LIN-32 protein in activating these genes might be. First, we selected several of the ray-expressed genes we identified and determined their expression patterns in *lin-32(gm239)* mutant males, in which almost all rays fail to develop. We found that the expression of the genes we tested (*cwp-4*, *tbb-4*, *twk-21*, and *egl-46*) in the male tail was strongly reduced or eliminated in *lin-32(gm239)* animals (Fig. 5 and data not shown). This is not unexpected, as the neuroblast identity of ray precursor cells requires *lin-32* function (Zhao and Emmons, 1995) and other differentiated ray markers have been shown to require *lin-32* for their expression (Portman and Emmons, 2000). Although this suggests that many ray genes are likely to be at least indirect targets of *lin-32*, it remains unclear from these data whether they might be direct targets as well.

Because bHLH proteins of the *atonal* class are known to bind to target sequences that contain E-box motifs (CANNTG) (Massari and Murre, 2000), we also searched the promoters of our high-scoring and ray-expressed genes for the occurrence of this motif and, more specifically, for CACGTG, which represents a high-affinity site for "Class B" bHLH transcription factors such as LIN-32 (Atchley and Fitch, 1997; Fisher and Caudy, 1998; Portman and Emmons, 2000). We found no evidence for a significant change in E-box frequencies in the putative regulatory regions of genes in the top second percentile or in those of ray-expressed genes (Table 2). Though this does not rule out the prospect that *lin-32* could act directly, it does support the possibility that *lin-32* acts largely through intermediate factors that then directly establish and/or maintain the patterns of transcription that determine the specific characteristics of each neuronal subtype (Fig. 6). Because *lin-32* expression is absent (or decreased substantially) in adult males (Portman and Emmons, 2000; Zhao and Emmons, 1995), *lin-32* is unlikely to have a prominent role in the maintenance of gene expression in adult rays. However, this alone does not rule out the possibility that *lin-32* could act directly in the initial activation of some (or all) ray genes, as *lin-32* reporter expression can be observed in early postmitotic ray cells (Portman and Emmons, 2000). The identification of more direct targets of the LIN-32:HLH-2 heterodimer might be possible through similar expression-profiling experiments using mRNA isolated from males undergoing the early stages of ray development.

Studying cell-fate specification with gene-expression profiling

Profiling gene expression in neural subtypes in a living animal presents a unique set of challenges. Although the total number of ray cells in wild-type males is relatively small compared to the total number of cells in the body (18 ray cells of each type vs. approximately 1500 somatic and

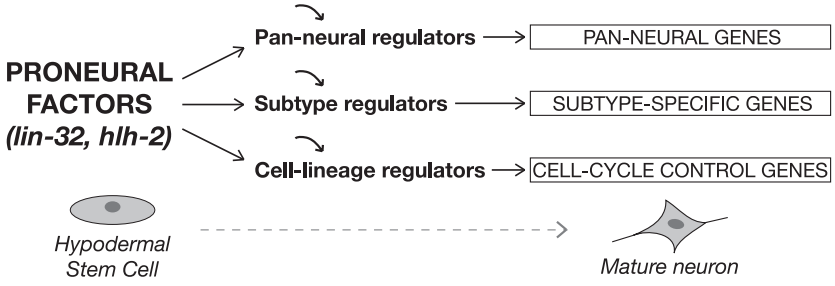

Fig. 6. Genetic control of ray cell specification. Multiple lines of evidence suggest that *lin-32* and its partner *hlh-2* act at the top of a hierarchy and that their function is required for multiple independent aspects of ray development (Portman and Emmons, 2000; Zhao and Emmons, 1995). We propose that *lin-32* activates multiple targets that control various aspects of cell-fate specification and cell lineage control. In particular, *lin-32* may activate a set of regulators of neural subtype as well as a pathway that implements pan-neural characteristics. The activation of downstream terminal differentiation genes may therefore require *lin-32* function only indirectly. Curved arrows indicate the likelihood that other independent inputs are also required to activate these intermediate factors with the correct temporal, spatial, and sexual specificity. In the case of the rays, these inputs are likely to include the Hox genes *mab-5* and *egl-5*, the DM domain genes *mab-3* and *mab-23*, and TGF-β and Wnt signals. Though we do not address cell lineage control here, it is likely that *lin-32* at least indirectly influences the competence of cells to divide and the timing of these divisions since ray progenitors become prematurely postmitotic in *lin-32* mutants.

germline cells, Kimble and White, 1981), our microarray measurements were sensitive enough to detect these differences in the context of mRNA prepared from whole animals. The success of this approach is encouraging, as it has allowed the identification of many ray-expressed genes that might have been difficult or impossible to identify using other means. Because many high-scoring genes in our data set have not yet been tested, further analysis may yield a significant number of additional ray-expressed genes. It is also possible that this general strategy may be successful for other repeated neuronal types in *C. elegans*, such as ventral-cord motor neurons. Nevertheless, the accuracy with which we can identify ray-expressed genes using this method leaves room for improvement. The use of alternative strategies, such as mRNA-tagging (Roy et al., 2002) or laser capture of small numbers of cells, may allow us to circumvent the specificity problems associated with the comparison of total RNA between two mutant strains.

The 13 genes we have identified here, particularly the *cwp* genes and new ray transcription factors, provide novel avenues to characterize ray development and function. Of special interest to us is the relationship between the early and intermediate activities of *lin-32* and *hlh-2* in the ray sublineage and the downstream or parallel functions of regulatory factors that may function specifically in each ray cell type. In addition, the novel *cwp* genes may yield insight into the mechanism and role of polycystin signaling in male mating behavior. With the combined use of forward-genetic, genomic, reverse-genetic, and biochemical approaches, it becomes a feasible goal to understand both the genetic network that directs ray development and the nature of the conserved neuronal specification pathways that are employed throughout metazoan organisms.

Acknowledgments

We are especially grateful to S. Kim, K. Duke, and J. Ryu for making microarray technology available to us, to R. Azevedo for discussion and statistical advice, and to K. Lee for technical assistance. We thank J. Hodgkin for his description in the *Worm Breeder's Gazette* of the use of *dpy-28* to obtain male-enriched cultures and J. Wu and M. Chalfie for providing reporter constructs for *egl-46*. We also thank H. Robertson and C. Bargmann for information about chemoreceptor genes in *C. elegans*, C. Laiosa for help with strain constructions, A. Fire and colleagues for GFP and YFP clones, O. Hobert for suggestions about the PCR fusion strategy, and T. Stiernagle and the *Caenorhabditis* Genetics Center, which is supported by the NIH National Center for Research Resources, for nematode strains. We thank the members of our laboratories for advice during this project and comments on the manuscript. This research was supported by NIH Grants R01 NS038906 (S.W.E.) and F32 GM018131 (D.S.P.). S.W.E. is the Siegfried Ullmann Professor of Molecular Genetics.

References

Atchley, W.R., Fitch, W.M., 1997. A natural classification of the basic helix-loop-helix class of transcription factors. Proc. Natl. Acad. Sci. U. S. A. 94, 5172–5176.

Baldi, P., Long, A.D., 2001. A Bayesian framework for the analysis of microarray expression data: regularized *t*-test and statistical inferences of gene changes. Bioinformatics 17, 509–519.

Barr, M.M., Sternberg, P.W., 1999. A polycystic kidney-disease gene homologue required for male mating behaviour in *C. elegans*. Nature 401, 386–389.

Barr, M.M., DeModena, J., Braun, D., Nguyen, C.Q., Hall, D.H., Sternberg, P.W., 2001. The *Caenorhabditis elegans* autosomal dominant polycystic kidney disease gene homologs *lov-1* and *pkd-2* act in the same pathway. Curr. Biol. 11, 1341–1346.

Baumeister, R., Liu, Y., Ruvkun, G., 1996. Lineage-specific regulators couple cell lineage asymmetry to the transcription of the *Caenorhabditis elegans* POU gene *unc*-86 during neurogenesis. Genes Dev. 10, 1395–1410.

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

Chalfie, M., Au, M., 1989. Genetic control of differentiation of the *Caenorhabditis elegans* touch receptor neurons. Science 243, 1027–1033.

Chow, K.L., Emmons, S.W., 1994. HOM-C/Hox genes and four interacting loci determine the morphogenetic properties of single cells in the nematode male tail. Development 120, 2579–2592.

Chow, K.L., Hall, D.H., Emmons, S.W., 1995. The *mab-21* gene of *Caenorhabditis elegans* encodes a novel protein required for choice of alternate cell fates. Development 121, 3615–3626.

Christensen, M., Estevez, A., Yin, X., Fox, R., Morrison, R., McDonnell, M., Gleason, C., Miller III, D.M., Strange, K., 2002. A primary culture system for functional analysis of *C. elegans* neurons and muscle cells. Neuron 33, 503–514.

Consortium, 1998. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science 282, 2012–2018.

Dekker, J., Rossen, J.W., Buller, H.A., Einerhand, A.W., 2002. The MUC family: an obituary. Trends Biochem. Sci. 27, 126–131.

Emmons, S.W., Somlo, S., 1999. Signal transduction. Mating, channels and kidney cysts. Nature 401, 339–340.

Emmons, S.W., Sternberg, P.W., 1997. Male development and mating behavior. In: Riddle, D.L., Blumenthal, T., Meyer, B.J., Priess, J.R. (Eds.), *C. elegans* II. Cold Spring Harbor Laboratory Press, Plainview, NY, pp. 295–334.

Fisher, A., Caudy, M., 1998. The function of hairy-related bHLH repressor proteins in cell fate decisions. BioEssays 20, 298–306.

Fukushige, T., Siddiqui, Z.K., Chou, M., Culotti, J.G., Gogonea, C.B., Siddiqui, S.S., Hamelin, M., 1999. MEC-12, an alpha-tubulin required for touch sensitivity in *C. elegans*. J. Cell Sci. 112 (Pt. 3), 395–403.

Gaudet, J., Mango, S.E., 2002. Regulation of organogenesis by the *Caenorhabditis elegans* FoxA protein PHA-4. Science 295, 821–825.

Granato, M., Schnabel, H., Schnabel, R., 1994. *pha-1*, a selectable marker for gene transfer in *C. elegans*. Nucleic Acids Res. 22, 1762–1763.

Hamelin, M., Scott, I.M., Way, J.C., Culotti, J.G., 1992. The *mec-7* beta-tubulin gene of *Caenorhabditis elegans* is expressed primarily in the touch receptor neurons. EMBO J. 11, 2885–2893.

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

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

Holzschuh, J., Barrallo-Gimeno, A., Ettl, A.K., Durr, K., Knapik, E.W., Driever, W., 2003. Noradrenergic neurons in the zebrafish hindbrain are induced by retinoic acid and require *tfap2a* for expression of the neurotransmitter phenotype. Development 130, 5741–5754.

Horton, R.M., 1997. In vitro recombination and mutagenesis of DNA. In: White, B.A. (Ed.), PCR Cloning Protocols, vol. 67. Humana Press, Totowa, NJ, pp. 141–149.

Igarashi, P., Somlo, S., 2002. Genetics and pathogenesis of polycystic kidney disease. J. Am. Soc. Nephrol. 13, 2384–2398.

Jiang, M., Ryu, J., Kiraly, M., Duke, K., Reinke, V., Kim, S.K., 2001. Genome-wide analysis of developmental and sex-regulated gene expression profiles in *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. U. S. A. 98, 218–223.

Kaletta, T., Van der Craen, M., Van Geel, A., Dewulf, N., Bogaert, T., Branden, M., King, K.V., Buechner, M., Barstead, R., Hyink, D., Li, H.P., Geng, L., Burrow, C., Wilson, P., 2003. Towards understanding the polycystins. Nephron Exp. Nephrol. 93, e9–e17.

Kent, W.J., Zahler, A.M., 2000. The intronerator: exploring introns and alternative splicing in *Caenorhabditis elegans*. Nucleic Acids Res. 28, 91–93.

Kim, S.K., Lund, J., Kiraly, M., Duke, K., Jiang, M., Stuart, J.M., Eizinger, A., Wylie, B.N., Davidson, G.S., 2001. A gene expression map for *Caenorhabditis elegans*. Science 293, 2087–2092.

Kimble, J.E., White, J.G., 1981. On the control of germ cell development in *Caenorhabditis elegans*. Dev. Biol. 81, 208–219.

Knight, R.D., Nair, S., Nelson, S.S., Afshar, A., Javidan, Y., Geisler, R., Rauch, G.J., Schilling, T.F., 2003. *lockjaw* encodes a zebrafish *tfap2a* required for early neural crest development. Development 130, 5755–5768.

Krause, M., 1995. Techniques for analyzing transcription and translation. In: Epstein, H.F., Shakes, D.C. (Eds.), *Caenorhabditis elegans*: Modern Biological Analysis of an Organism, vol. 48. Academic Press, San Diego, pp. 513–529.

Krause, M., Park, M., Zhang, J.M., Yuan, J., Harfe, B., Xu, S.Q., Greenwald, I., Cole, M., Paterson, B., Fire, A., 1997. A *C. elegans E/Daughterless* bHLH protein marks neuronal but not striated muscle development. Development 124, 2179–2189.

Kunkel, M.T., Johnstone, D.B., Thomas, J.H., Salkoff, L., 2000. Mutants of a temperature-sensitive two-P domain potassium channel. J. Neurosci. 20, 7517–7524.

Ledent, V., Paquet, O., Vervoort, M., 2002. Phylogenetic analysis of the human basic helix-loop-helix proteins. Genome Biol. 3 (RESEARCH0030).

Lesage, F., Lazdunski, M., 2000. Molecular and functional properties of two-pore-domain potassium channels. Am. J. Physiol. Renal Physiol. 279, F793–F801.

Letunic, I., Goodstadt, L., Dickens, N.J., Doerks, T., Schultz, J., Mott, R., Ciccarelli, F., Copley, R.R., Ponting, C.P., Bork, P., 2002. Recent improvements to the SMART domain-based sequence annotation resource. Nucleic Acids Res. 30, 242–244.

Lewis, J.A., Fleming, J.T., 1995. Basic culture methods. In: Epstein, H.F., Shakes, D.C. (Eds.), *Caenorhabditis elegans*: Modern Biological Analysis of an Organism, vol. 48. Academic Press, San Diego, pp. 3–29.

Lints, R., Emmons, S.W., 1999. Patterning of dopaminergic neurotransmitter identity among *Caenorhabditis elegans* ray sensory neurons by a TGF-beta family signaling pathway and a Hox gene. Development 126, 5819–5831.

Lints, R., Emmons, S.W., 2002. Regulation of sex-specific differentiation and mating behavior in *C. elegans* by a new member of the DM domain transcription factor family. Genes Dev. 16, 2390–2402.

Lu, J., Webb, R., Richardson, J.A., Olson, E.N., 1999. MyoR: a muscle-restricted basic helix-loop-helix transcription factor that antagonizes the actions of MyoD. Proc. Natl. Acad. Sci. U. S. A. 96, 552–557.

Lu, J.R., Bassel-Duby, R., Hawkins, A., Chang, P., Valdez, R., Wu, H., Gan, L., Shelton, J.M., Richardson, J.A., Olson, E.N., 2002. Control of facial muscle development by MyoR and capsulin. Science 298, 2378–2381.

Massari, M.E., Murre, C., 2000. Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. Mol. Cell. Biol. 20, 429–440.

Massari, M.E., Rivera, R.R., Voland, J.R., Quong, M.W., Breit, T.M., van Dongen, J.J., de Smit, O., Murre, C., 1998. Characterization of ABF-1, a novel basic helix-loop-helix transcription factor expressed in activated B lymphocytes. Mol. Cell. Biol. 18, 3130–3139.

McGrath, J., Somlo, S., Makova, S., Tian, X., Brueckner, M., 2003. Two populations of node monocilia initiate left–right asymmetry in the mouse. Cell 114, 61–73.

Metzstein, M.M., Hengartner, M.O., Tsung, N., Ellis, R.E., Horvitz, H.R., 1996. Transcriptional regulator of programmed cell death encoded by *Caenorhabditis elegans* gene *ces-2*. Nature 382, 545–547.

Mitani, S., Du, H., Hall, D.H., Driscoll, M., Chalfie, M., 1993. Combinatorial control of touch receptor neuron expression in *Caenorhabditis elegans*. Development 119, 773–783.

Nass, R., Hall, D.H., Miller III, D.M., Blakely, R.D., 2002. Neurotoxin-induced degeneration of dopamine neurons in *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. U. S. A. 99, 3264–3269.

Nauli, S.M., Alenghat, F.J., Luo, Y., Williams, E., Vassilev, P., Li, X., Elia, A.E., Lu, W., Brown, E.M., Quinn, S.J., Ingber, D.E., Zhou, J., 2003. Polycystins 1 and 2 mediate mechanosensation in the primary cilium of kidney cells. Nat. Genet. 33, 129–137.

Nguyen, L., Round, J., O'Connell, R., Geurts, P., Funes-Duran, M., Wong, J., Jongeward, G., Vierra, C.A., 2001. Isolation and characterization of the activated B-cell factor 1 homolog in *Caenorhabditis elegans*. Nucleic Acids Res. 29, 4423–4432.

Nielsen, H., Engelbrecht, J., Brunak, S., von Heijne, G., 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. 10, 1–6.

Plenefisch, J.D., DeLong, L., Meyer, B.J., 1989. Genes that implement the hermaphrodite mode of dosage compensation in *Caenorhabditis elegans*. Genetics 121, 57–76.

Portman, D.S., Emmons, S.W., 2000. The basic helix-loop-helix transcription factors LIN-32 and HLH-2 function together in multiple steps of a *C. elegans* neuronal sublineage. Development 127, 5415–5426.

Reinke, V., Smith, H.E., Nance, J., Wang, J., Van Doren, C., Begley, R., Jones, S.J., Davis, E.B., Scherer, S., Ward, S., Kim, S.K., 2000. A global profile of germline gene expression in *C. elegans*. Mol. Cell 6, 605–616.

Reinke, V., Gil, I.S., Ward, S., Kazmer, K., 2003. Genome-wide germline-enriched and sex-biased expression profiles in *Caenorhabditis elegans*. Development 131, 311–323.

Renthal, R., Schneider, B.G., Miller, M.M., Luduena, R.F., 1993. Beta IV is the major beta-tubulin isotype in bovine cilia. Cell Motil. Cytoskeleton 25, 19–29.

Roy, P.J., Stuart, J.M., Lund, J., Kim, S.K., 2002. Chromosomal clustering of muscle-expressed genes in *Caenorhabditis elegans*. Nature 418, 975–979.

Ruvkun, G., 1997. Patterning the nervous system. In: Riddle, D.L., Blumenthal, T., Meyer, B.J., Priess, J.R. (Eds.), *C. elegans* II. Cold Spring Harbor Laboratory Press, Plainview, NY, pp. 543–581.

Salkoff, A., Kunkel, M.T., Wang, Z.-W., Butler, A., Yuan, A., Nonet, M., Wei, A., 1999. The impact of the *Caenorhabditis elegans* genome project on potassium channel biology. In: Kurachi, Y., Jan, L.Y., Lazdunski, M. (Eds.), Potassium Ion Channels: Molecular Structure, Function, and Diseases, vol. 46. Academic Press, San Diego, pp. 9–27.

Salkoff, L., Butler, A., Fawcett, G., Kunkel, M., McArdle, C., Paz-y-Mino, G., Nonet, M., Walton, N., Wang, Z.W., Yuan, A., Wei, A., 2001. Evolution tunes the excitability of individual neurons. Neuroscience 103, 853–859.

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

Shaham, S., Bargmann, C.I., 2002. Control of neuronal subtype identity by the *C. elegans* ARID protein CFI-1. Genes Dev. 16, 972–983.

Sherlock, G., Hernandez-Boussard, T., Kasarskis, A., Binkley, G., Matese, J.C., Dwight, S.S., Kaloper, M., Weng, S., Jin, H., Ball, C.A., Eisen, M.B., Spellman, P.T., Brown, P.O., Botstein, D., Cherry, J.M., 2001. The Stanford Microarray Database. Nucleic Acids Res. 29, 152–155.

Stein, L., Sternberg, P., Durbin, R., Thierry-Mieg, J., Spieth, J., 2001. WormBase: network access to the genome and biology of *Caenorhabditis elegans*. Nucleic Acids Res. 29, 82–86.

Stivers, C., Brody, T., Kuzin, A., Odenwald, W.F., 2000. *Nerfin*-1 and-2, novel *Drosophila* Zn-finger transcription factor genes expressed in the developing nervous system. Mech. Dev. 97, 205–210.

Sulston, J.E., Albertson, D.G., Thomson, J.N., 1980. The *Caenorhabditis elegans* male: postembryonic development of nongonadal structures. Dev. Biol. 78, 542–576.

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

Sze, J.Y., Victor, M., Loer, C., Shi, Y., Ruvkun, G., 2000. Food and metabolic signalling defects in a *Caenorhabditis elegans* serotonin-synthesis mutant. Nature 403, 560–564.

Thomas, J.H., Lockery, S., 1999. Neurobiology. In: Hope, I.A. (Ed.), *C. elegans*: A Practical Approach. Oxford Univ. Press, Oxford, pp. 143–179.

Troemel, E.R., Chou, J.H., Dwyer, N.D., Colbert, H.A., Bargmann, C.I., 1995. Divergent seven transmembrane receptors are candidate chemosensory receptors in *C. elegans*. Cell 83, 207–218.

Wall, L., Christiansen, T., Orwant, J., 2000. Programming Perl. O'Reilly & Associates, Sebastopol, CA.

Wang, X., Chamberlin, H.M., 2002. Multiple regulatory changes contribute to the evolution of the *Caenorhabditis lin-48 ovo* gene. Genes Dev. 16, 2345–2349.

Weinshenker, D., Wei, A., Salkoff, L., Thomas, J.H., 1999. Block of an *ether-a-go-go*-like K(+) channel by imipramine rescues *egl-2* excitation defects in *Caenorhabditis elegans*. J. Neurosci. 19, 9831–9840.

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

Wrischnik, L.A., Kenyon, C.J., 1997. The role of *lin-22*, a *hairy/enhancer of split* homolog, in patterning the peripheral nervous system of *C. elegans*. Development 124, 2875–2888.

Wu, J., Duggan, A., Chalfie, M., 2001. Inhibition of touch cell fate by *egl-44* and *egl-46* in *C. elegans*. Genes Dev. 15, 789–802.

Yi, W., Ross, J.M., Zarkower, D., 2000. *mab-3* is a direct *tra-1* target gene regulating diverse aspects of *C. elegans* male sexual development and behavior. Development 127, 4469–4480.

Yu, R.Y., Nguyen, C.Q., Hall, D.H., Chow, K.L., 2000. Expression of *ram-5* in the structural cell is required for sensory ray morphogenesis in *Caenorhabditis elegans* male tail. EMBO J. 19, 3542–3555.

Yu, H., Pretot, R.F., Burglin, T.R., 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., Emmons, S.W., 1995. Specification of sense-organ identity by a *Caenorhabditis elegans Pax-6* homologue. Nature 377, 55–59.

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

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