Supplemental Materials

Molecular Biology of the Cell

Lockhead et al.

Supplementary Material

The tubulin repertoire of *C. elegans* sensory neurons and its context-dependent role in process outgrowth

Supplementary Table S1: List of strains used in this study.

Strain Name	Genotype	Source/citation	Comments
N2	wild type	Brenner, 1974	
GN607	*uIs30[mec-17p::gfp]I*	This study	TRN::GFP
TU2769	*uIs31[mec-17p::gfp]III*	O'Hagan et al, 2005	TRN::GFP
TU4008	*uIs115[mec-17p::rfp]IV;uIs116[egl-5::gfp]?*	C. Zheng, M. Chalfie; Zheng et al, 2015b	
TU253	*mec-4(u253)X*	CGC; Hong et al, 2000	
GN627	*tba-1(ok1135)I;uIs31[mec-17p::gfp]III*	This study	
GN626	*tba-2(tm6948)I;uIs31[mec-17p::gfp]III*	This study	
GN672	*tba-1(ok1135)I;mec-12(pg75)uIs31[mec-17p::gfp]III*	This study	
GN673	*tba-2(tm6948)I; mec-12(pg75)uIs31[mec-17p::gfp]III*	This study	
GN625	*uIs30[mec-17p::gfp]I;tbb-1(gk207)III*	This study	
GN662	*uIs30[mec-17p::gfp]I;tbb-2(gk130)III*	This study	Grows slow
GN633	*uIs30[mec-17p::gfp]I;tbb-1(gk207)III;mec-7(u142)X*	This study	
GN666	*uIs30[mec-17p::gfp]I;tbb-2(gk130)III;mec-7(u142)X*	This study	Grows slow
TU2845	*uIs31[mec-17p::gfp]III;mec-7(u142)X*	O'Hagan et al, 2005	
GN613	*uIs30[mec-17p::gfp]I;mec-12(e1607)III*	This study	
GN665	*mec-12(pg75)uIs31[mec-17p::gfp]III*	This study	
GN663	*uIs30[mec-17p::gfp]I;mec-12(e1607)III;mec-7(u142)X*	This study	
TU2589	*uIs25[mec-18p::gfp]*	Zhang *et al.*, 2002	PLM RNA-seq
DA1267	*lin-15(n765)X;adEx1267[lin-15(+) gcy-8p::gfp]*	Yu *et al.*, 1997	AFD RNA-seq
DA1262	*lin-15(n765)X;adEx1262[lin-15(+) gcy-5p::gfp]*	Yu *et al.*, 1997	ASER RNA-seq
GN596	*uIs31[mec-17p::gfp]III;pmk-3(ok169)IV*	This study	
GN675	*pg77[TBA-1::TagRFP-T + loxP]I;uIs31[mec-17p::gfp)III*	This study	Broad strong RFP expression
GN683	*pg79[TBB-1::TagRFP-T + loxP]uIs31[mec-17p::gfp)III*	This study	Broad strong RFP expression
GN688	*pg81[TBA-2::TagRFP-T + loxP]I;uIs31[mec-17p::gfp)III*	This study	Broad weak RFP expression
PT3155	*pha-1(e2123) III; him-5(e1490)V; myEx882 [tba-1p::gfp]*	This study	
PT3156	*pha-1(e2123) III; him-5(e1490)V; myEx883 [tba-2p::gfp]*	This study	
PT3147	*pha-1(e2123) III; him-5(e1490)V; myEx897[tbb-2p::gfp]*	This study	

References: Brenner S, *Genetics* 77:71, 1974; O'Hagan, Chalfie & Goodman, *Nat Neurosci* 8:43, 2005; Yu, et al., *PNAS* 94:3384, 1997; Zhang, et al., *Nature* 418:331-5, 2002; Zheng et al, *Neuron* 88:514, 2015.

Supplementary Table S2: Numbers of filtered and mapped reads in RNA-seq data sets.

Input data (post-filter)	Total reads	Aligned 0 times	Aligned exactly 1 time	Aligned 2+ times	Overall alignment rate
PLM	20,140,971	9,927,990 (49.29%)	4,446,728 (22.08%)	5,766,253 (28.63%)	50.71%
AFD	27,057,771	19,369,009 (71.58%)	3,709,162 (13.71%)	3,979,600 (14.71%)	28.42%
ASER	19,316,855	13,303,583 (68.87%)	2,791,906 (14.45%)	3,221,366 (16.68%)	31.13%
Larvae	23,369,056	5,513,350 (23.59%)	7,422,251 (31.76%)	10,433,455 (44.65%)	76.41%

"Total reads" denotes the complete set of quality-filtered reads for a given data set that were mapped with RSEM to a *C. elegans* gene index (from WormBase release WS245) in order to compute gene expression values. With RSEM, a read can be mapped to the gene index either 0 times (i.e., it can fail to map at all); it can map exactly 1 time (i.e., it can map to a unique site in the gene index); or it can map 2+ times. For each RNA-seq data set, the numbers and percentages of reads with each status are given, as is the overall percentage of reads that mapped to the gene index.

Supplementary Table S3: Numbers of genes with above-background expression in PLM, AFD, ASER, and larval RNA-seq data.

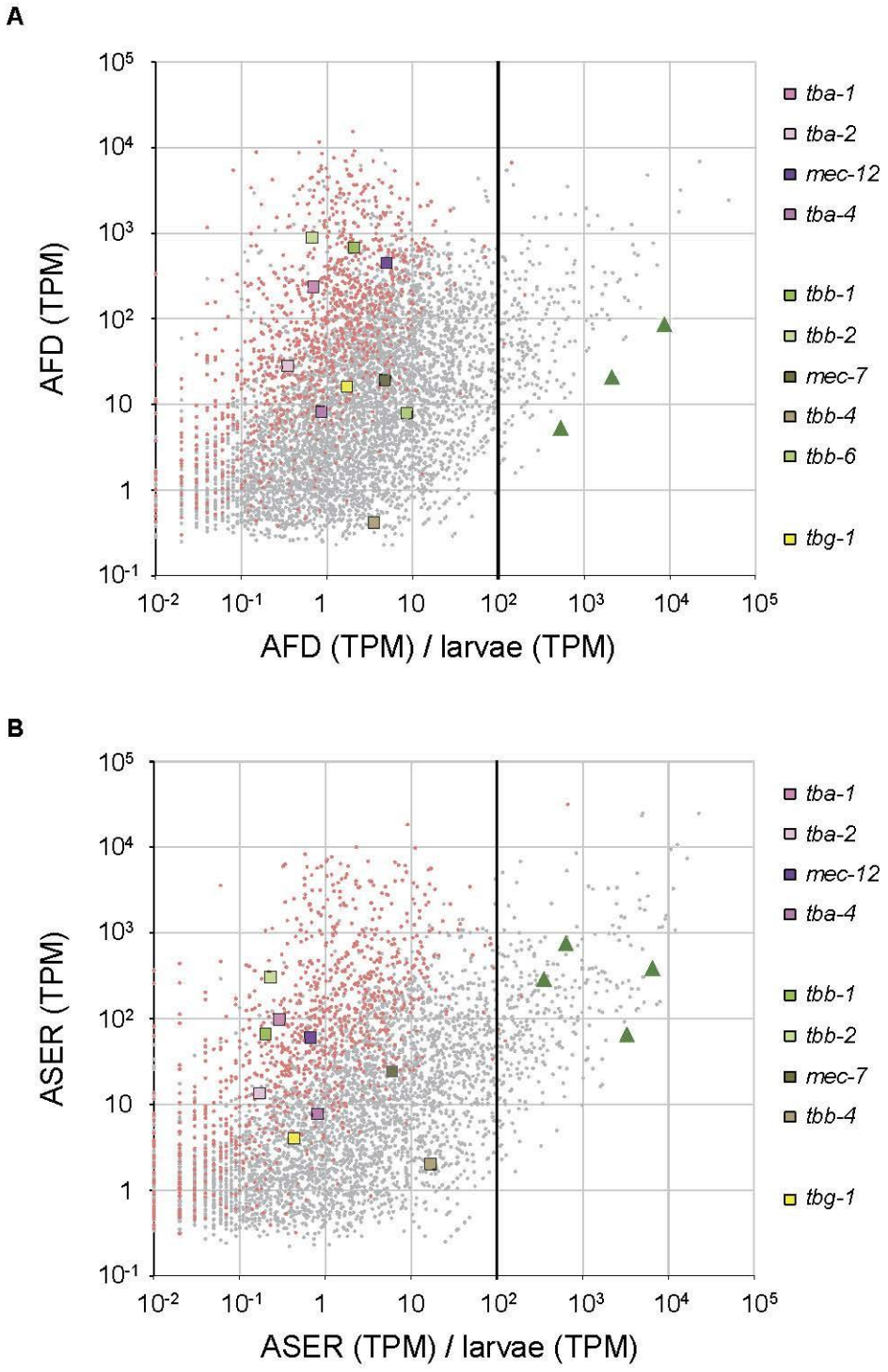
Data set	Protein-coding	ncRNA	Housekeeping	Tubulins
PLM	5,086	13	1,060	8
AFD	5,824	32	1,083	10
ASER	4,601	15	1,029	9
Larvae	9,918	6	1,153	11

Protein-coding and ncRNA-coding genes were counted as expressed if, in a given RSEM analysis of a given data set, the gene had a minimum expression value (minTPM) of ≥ 0.1 TPM in a 99% credibility interval (i.e., ≥ 0.1 minTPM). Housekeeping genes were from the gene set originally defined by Schwarz, et al., *PNAS* 109:16246, 2012. Tubulins were genes in the WS245 release of WormBase annotated as encoding a 'Tubulin' domain from PFAM-A (accession number PF00091.20). All gene annotations (including those used here, such as coding status) are given in Supplementary Table S4.

Supplementary Table S4: Traits of *C. elegans* genes expressed in PLM, other sensory neurons (AFD and ASER), and whole larvae.
See the Excel spreadsheet *PLM_AFD_ASER_larvae_data.xlsx*. Its data columns are as follows:

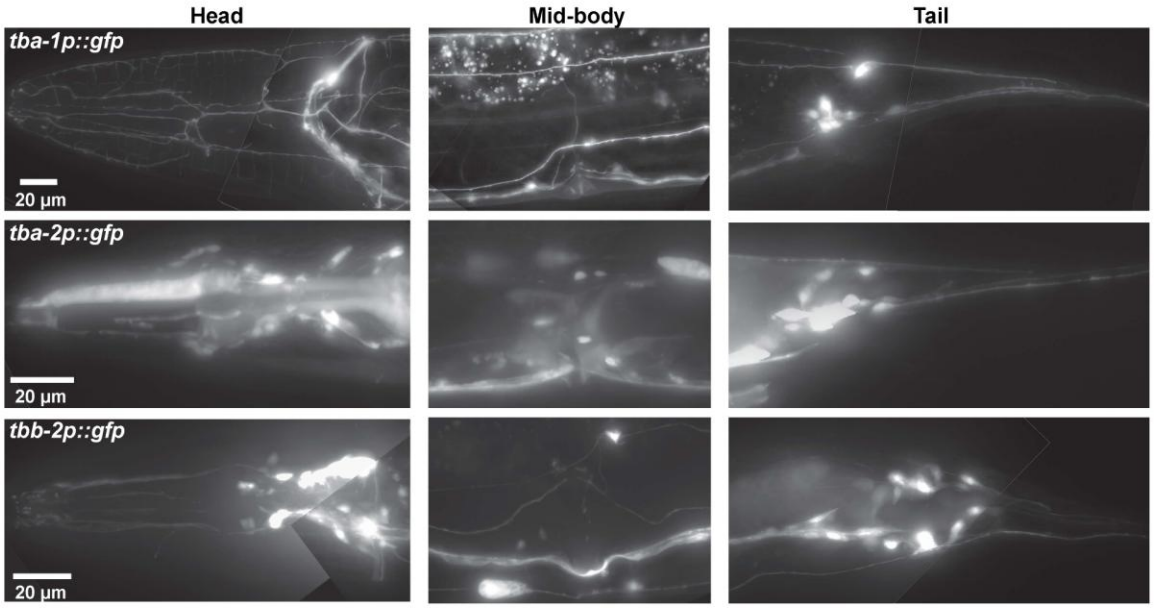
Gene: a given predicted protein-coding or ncRNA-coding gene in the *C. elegans* genome, from WormBase release WS245. All further data columns are pertinent to that particular gene.
PLM_TPM: the expression level for a given gene in PLM neurons, determined with RNA-seq reads from a pooled set of three individual PLM single-cell RT-PCR products (Supplementary Table S2), measured in transcripts per million (TPM).
PLM/larvae: the ratio of gene expression (measured in TPM) between PLM neurons and whole *C. elegans* larvae. Genes in this table have been ranked by descending values of PLM/larvae, as a general measure of their PLM-specificity.
AFD_TPM: the expression level for a given gene in AFD neurons, determined with RNA-seq reads from a pooled set of seven individual AFD single-cell RT-PCR products (Supplementary Table S2), measured in TPM.
AFD/larvae: the ratio of gene expression (measured in TPM) between AFD neurons and whole *C. elegans* larvae.
ASER_TPM: the expression level for a given gene in ASER neurons, determined with RNA-seq reads from a pooled set of seven individual ASER single-cell RT-PCR products (Supplementary Table S2), measured in TPM.
ASER/larvae: the ratio of gene expression (measured in TPM) between ASER neurons and whole *C. elegans* larvae.
Larvae_TPM: the expression level for a given gene in whole larvae, generated from a pooled set of all larval RNA-seq reads (Supplementary Table S2), measured in TPM.
Coding: the nature of a given gene's coding potential, as annotated in WormBase WS245. Most genes are either solely protein-coding or solely ncRNA-coding, and are noted as such in this data column. For 301 genes in *C. elegans*, WS245 predicts both protein-coding and non-protein-coding transcripts; in this table, such genes are denoted with "protein; ncRNA". However, for purposes of gene analysis, we assume that any gene with dual predicted nature is solely protein-coding.
Prot_size: this shows the full range of sizes for all protein products from a gene's predicted isoforms.
Max_prot_size: the size of the largest predicted protein product.
Housekeeping: a set of genes that were previously observed, by single-cell RNA-seq, to be constitutively active both in whole *C. elegans* larvae and in three different developmental stages/genotypes of migrating *C. elegans* linker cells (Schwarz *et al.*, 2012).
TF: genes annotated as encoding transcription factors by one or more of three different censuses by Gupta, Thomas, or Walhout, as previously compiled by Schwarz et al. (Schwarz *et al.*, 2012).
PFAM-A: for protein-coding genes, predicted domains from the annotated (PFAM-A) subdivision of PFAM 27 (Finn *et al.*, 2014), with an E-value of $\leq 10^{-5}$.
eggNOG: for protein-coding genes, predicted orthology groups from the eggNOG 3.0 database (Powell *et al.*, 2012).
Phobius: this denotes predictions of signal and transmembrane sequences made with Phobius (Käll *et al.*, 2004). 'SigP' indicates a predicted signal sequence, and 'TM' indicates one or more transmembrane-spanning helices, with N helices indicated with '(Nx)'. Varying predictions from different isoforms are listed

NCoils: this shows coiled-coil domains, predicted by ncoils (Lupas, 1996). As with Psegs, the relative and absolute fractions of each protein's coiled-coil residues are shown.
Psegs: this shows what fraction of a protein is low-complexity sequence, as detected by pseg (Wootton, 1994). Both the proportion of such sequence (ranging from 0.01 to 1.00) and the exact ratio of low-complexity residues to total residues are given. Proteins with no predicted low-complexity residues are blank.
[*RNA-seq read set*]_minTPM: for the gene in question, and for a given RNA-seq data set, this denotes the minimum estimate of that gene's activity as measured in that RNA-seq data set and computed by RSEM with a 99% credibility interval in Transcripts Per Million (minTPM). All cell types are as with "[*RNA-seq read set*]_TPM" above. Generating a statistically robust minimum estimate of gene activity allowed background noise to be distinguished from authentic but low expression levels above background.
[*RNA-seq read set*]_reads: for the gene in question, and for a given RNA-seq data set, this denotes a posterior mean estimate of the number of RNA-seq reads mapping to that gene as computed by RSEM, and with decimal fractions rounded off. All cell types are as with "**[*RNA-seq read set*] TPM**" above.

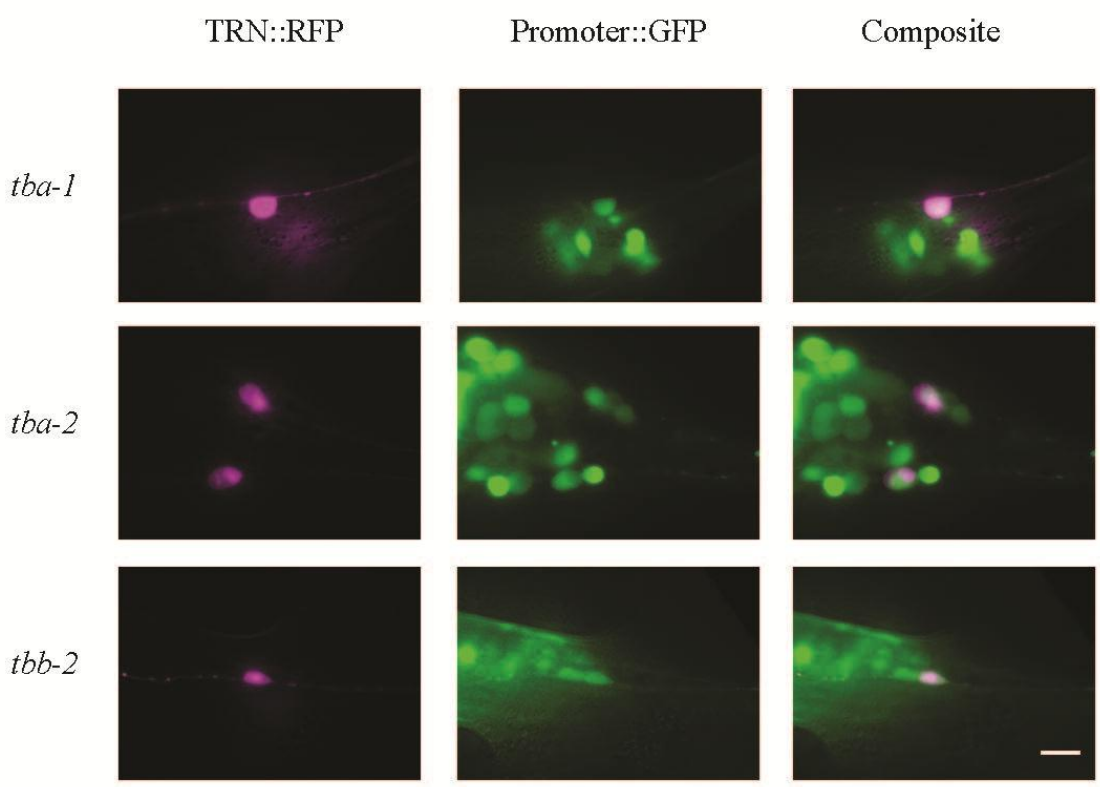


Supplemental Figure S1: Tubulin expression varies between AFD, and ASER neurons. **A.** Expression data are shown for 5,856 *C. elegans* genes (5,824 protein-coding and 32 ncRNA-coding) that exhibited above-background expression in our AFD RNA-seq data (i.e., that had a minimum expression level, in a 99% credibility interval, of ≥ 0.1 TPM; detailed data are given in Supplementary Table S4). The x-axis shows AFD-specificity, computed as the ratio of

AFD gene expression (in TPM) to larval gene expression (in TPM) for each gene. The y-axis shows the absolute magnitude of AFD gene activity in TPM. Both axes are in logarithmic scale; the AFD/larval expression ratio of 10^2 is highlighted. Among most genes (labeled "AFD" and shown as gray dots) are highlighted ten genes encoding tubulins (shown as individually color-coded squares), three genes encoding guanylate cyclase receptors (*gcy-8*, *gcy-23*, and *gcy-18*, shown as green triangles) that are required for thermosensation (Takeishi *et al.*, 2016) and 1,083 genes encoding housekeeping functions (labeled "Hkeep" and shown as red dots). The housekeeping genes were identified in a previous single-cell RNA-seq analysis (Schwarz *et al.*, 2012). **B.** Expression data are shown for 4,616 *C. elegans* genes (4,601 protein-coding and 15 ncRNA-coding) that exhibited above-background expression in our ASER RNA-seq data (i.e., that had a minimum expression level, in a 99% credibility interval, of ≥ 0.1 TPM; detailed data are given in Supplementary Table S4). The x-axis shows ASER-specificity, computed as the ratio of ASER gene expression (in TPM) to larval gene expression (in TPM) for each gene. The y-axis shows the absolute magnitude of ASER gene activity in TPM. Both axes are in logarithmic scale; the ASER/larval expression ratio of 10^2 is highlighted. Among most genes (labeled "ASER" and shown as gray dots) are highlighted nine genes encoding tubulins (shown as individually color-coded squares), two genes encoding guanylate cyclase receptors (*gcy-19* and *gcy-22*) or neighboring a guanylate cyclase gene (M02G9.4 and M02G9.3; adjacent to *gcy-5*) that have been previously found to be specifically expressed in ASER (collectively shown as green triangles) and 1,029 genes encoding housekeeping functions (labeled "Hkeep" and shown as red dots). In the cases of *gcy-19* and *gcy-22*, our expression data match previously published observations (Ortiz *et al.*, 2006). In the case of *gcy-5*, we do not observe previously observed ASER-specific expression (Yu *et al.*, 1997) in our RNA-seq data, but do see strongly ASER-specific expression for two genes (M02G9.4 and M02G9.3) that flank a common promoter region shared by *gcy-5* and M02G9.4. The housekeeping genes were identified in a previous single-cell RNA-seq analysis (Schwarz *et al.*, 2012).



Supplemental Figure S2: Tubulin promoter expression patterns. All three markers show extensive expression in neurons, including ventral nerve cord and several tail neurons. *tba-1p::gfp* is expressed in extensively branched FLP neurons in the head, and is visible in commisures and neuronal processes that span the body. In *tba-2p::GFP* transgenic animals, there is some pharyngeal and some vulval muscle expression in addition to neuronal expression. *tbb-2p::gfp* is expressed in numerous ciliated sensory neurons in the head and several neurons in the tail. Some pictures are composed of multiple micrographs stitched together in Photoshop.



Supplemental Figure S3: *tba-1*, *tba-2*, and *tbb-2* promoter-driven GFP expressed in TRNs. PLMs marked with *uIs115* transgene (TRN::RFP). Images of promoter::GFP shown for the *tba-1*, *tba-2*, and *tbb-2* promoters above. Expression of promoter::GFP detected in all six TRNs for all three promoters examined. Anterior is left, dorsal top. Scale bar: 10 µm.