Supplemental Results

SR1 Comparisons of gene sets enriched in certain tissues and cell types to published data sets

To validate the enriched gene lists we compared a select number of our data sets to similar profiles generated by other groups. For example, we identified 318 genes annotated as expressed in the excretory cell by combining GFP expression patterns from WormBase (WS200) and from the Genome BC *C. elegans* Gene Expression Consortium (Hunt-Newbury et al. 2007; Rogers et al. 2008). Our L2 excretory cell enriched gene list (531 genes) contains 61 of these 318 genes, which is significantly higher than expected for a random distribution (6.7X over-represented, $p < 7.1e-33$). Similarly, the list of gene enriched in L2 body-wall muscle generated by our study (1,152 genes) shares 146 genes (2X over-represented, $p < 8.426e-17$) with a comparable list (1,157 genes) obtained from L1 larval body-wall (Roy et al. 2002). A previously produced L2 stage intestine-specific data set (1,925 genes) significantly overlaps with our L2 intestine profile (195 out of 678 genes) (2.8X over-represented, $p < 4.352e-42$) (Pauli et al. 2005). The union of our embryonic and L2 stage intestine enriched data sets contains 1540 transcripts that overlaps with 103 of 153 (8.1 fold over-represented, $p < 2.9e-74$) genes previously identified as intestine-specific in SAGE (serial analysis of gene expression) data sets derived from embryonic and adult intestine (McGhee et al. 2007; McGhee et al. 2009). A SAGE data set from the young adult gonad (Wang et al. 2009) identified 1,063 genes enriched in the germ line in comparison to the all somatic cells. We generated a tiling array profile also from dissected yound adult gonads and identified 4,363 enriched genes in comparison to the soma. These germ line enriched SAGE and tiling array data sets significantly overlap, sharing 462 genes (1.8X over-representation, $p < 4.016e-49$). A comparison of the previously generated germ line SAGE list to our embryonic Z2/Z3 germ line precursor enriched gene, also shows significant enrichment (3.0X over-representation, $p < 1.051e-40$). The significant overlap between our cell-specific enriched gene lists and data sets generated by other groups reinforces the validity of each data set by confirming previous results.

SR2 Comparison of cell-types profiled at embryonic and larval stages.

We profiled eight cell types (all neurons, A-class motor neurons, GABA neurons, dopaminergic neurons, intestine, body muscle, coelomocytes, hypodermis) at both embryonic and larval stages (Table 1, Supplemental Table 1). We compared the embryonic and corresponding postembryonic data sets for each tissue to identify a "core" set of genes that is detected in both data sets as well as the subset of these genes that is enriched in both members of each pair. The core lists are generated by taking the intersection of expressed or enriched genes for each cell type from embryonic and larval stages (Supplemental Files #3-10). For instance, we have profiled dopaminergic neurons at embryonic and L3/L4 larval stages. As shown in Fig. 4B, several genes required for dopamine synthesis and dopaminergic neuron function (*ast-1, dat-1, cat-2, cat-4*) are enriched in dopaminergic neurons in both embryonic and larval stages. Other genes with known dopaminergic function, *bas-1* (aromatic amino acid decarboxylase) (Hare and Loer 2004) and *asic-1* (DEG/ENaC cation channel subunit) (Voglis and Tavernarakis 2008) are also included in this core list of dopaminergic neuron-enriched transcripts. Thus, genes with potentially important for the function of a particular cell-type can be identified by testing candidates included in the shared gene lists (http://www.vanderbilt.edu/wormmap/Core_enriched_genes/).

SR3 Genes encoding membrane transporter proteins are highly enriched in the excretory cell.

Osmoregulation and excretion are fundamental biological processes that all animals share. In a typical multicellular organism, specialized cell types are assembled into an excretory organ that collects and removes metabolic wastes or functions to maintain ionic balance in changing aqueous environments. In *C. elegans*, these complex physiological tasks are accomplished with a simple excretory system composed of only 4 types of cells; the pore cell, duct cell, gland cells and excretory canal. The largest of these cells, the excretory canal, assumes a unique H-shaped architecture in which elongated tubular processes emanate from the cell soma beneath the posterior bulb of the pharynx, bifurcate to the right and left sides and then separate again to extend in both anterior and posterior directions along the entire length of the animal (**Supplemental Fig. S25A**)(Nelson et al. 1983; Altun 2002-2010). The excretory cell cytoplasm is

contained within a cylindrical membrane-bound domain that is penetrated from the interior or basal side by elaborate networks of canals. These “canaliculi” converge on an internal, fluid filled “tunnel” that connects with the duct and pore cells on the ventral side of the head region. Disruption of any one of these cell types, duct, pore or excretory canal, disables osmoregulatory capacity as evidenced by a swollen, lethal phenotype in hypotonic solutions (Nelson et al. 1983). We used the mRNA-tagging strategy to generate a tiling array profile of the excretory cell in L2 larvae, a developmental stage of both active excretory cell growth and essential osmoregulatory function (**Table 1**). This data set identified 531 transcripts that are enriched (≥ 2-fold, FDR ≤ 5%) in the excretory cell in comparison to the average L2 larval stage cell (see Methods). GFP reporter genes generated from three genes that are highly enriched in this data set illuminate the elongated anatomy of this unique cell type (**Supplemental Fig. S25C-E**). As would be expected for a cell type with high osmoregulatory activity, molecular function gene ontology (GO) terms for membrane transporter related activities are over-represented in the excretory cell data set (FDR < 0.01, hypergeometric distribution, **Supplemental Fig. S25B**) and thus are indicative of excretory cell specific profile. In addition to detecting genes that code for physiological functions, the enriched profile also includes 17 transcription factors with potential roles in excretory cell differentiation (http://edgedb.umassmed.edu, (Reece-Hoyes et al. 2005)). Indeed, the POU domain transcription factor, CEH-6, is highly enriched (4-fold) and has been previously shown to control excretory cell morphogenesis and gene expression (Burglin and Ruvkun 2001; Mah et al. 2007; Armstrong and Chamberlin 2010; Mah et al. 2010). All 17 (100%) of the known CEH-6-regulated genes are included in the excretory cell profile (see also Supplemental Fig. S25). Another 79 genes from this list have a perfect match to the CEH-6 binding site octamer, ATTTGCAT, within 1 kb upstream of the translational start site and are thus candidates for additional CEH-6 target genes. Two other members of the homeodomain family in this list, *ceh-26* (3.6-fold) and *ceh-37* (3.7-fold), are known to be expressed in the excretory cell (Lanjuin et al. 2003; Reece-Hoyes et al. 2005) but downstream targets have not been identified. Our finding that multiple transcription factors are expressed in the excretory cell is consistent with the earlier suggestion that excretory cell differentiation likely depends on the gene regulatory roles of multiple transcription factors functioning in parallel pathways

(Burglin and Ruvkun 2001; Mah et al. 2007; Armstrong and Chamberlin 2010; Mah et al. 2010). For example, in addition to detecting all of the known *ceh-6* targets, our data set also includes 9/16 (56%) of vacuolar ATPase proton pump subunit genes that are coordinately regulated by the nuclear hormone receptor, *nhr-31* (Hahn-Windgassen and Van Gilst 2009).

SR4 Comparison between mSTAD TARs and modMine TARs

TARs predicted from the same tiling array data sets, but with different computational methods were compared to the integrated transcript model, which is based on RNA sequencing (Gerstein et al., *in press*). On the level of individual tiling probes, we assessed precision and sensitivity relative to probes mapped to exons of the integrated transcript model (see Supplemental Fig. S6). For modMine TARs derived from cell-type data sets, we observed an average sensitivity of 33% at an average precision of 90%. On average, TARs predicted by mSTAD achieved a sensitivity of 56% and a precision of 87%; for the subset of mSTAD TARs for which expression was confirmed by a statistical test, precision increased to 90% at a reduced sensitivity of 47%. Thus expressed TARs detect ~14% more exon probes than modMine TARs at approximately the same precision. All mSTAD TARs detect 23% more exon probes at slightly lower precision. For individual genes, in particular ones with intermediate expression levels, this improved TAR accuracy is also reflected by more accurate detection of exon boundaries (see Supplemental Fig. S7 for example cases). The superior accuracy of mSTAD can be ascribed to a more complex modeling approach, which specifically accounts for introns, as well as a state-of-the-art inference method (Hidden Markov Support Vector Machines, see (Zeller et al. 2008) and references therein) (see Supplemental File #11 modMINE comparison).

SR5 Comparison of SOM clusters to germ line and sex-specific profiles

Of the 644 genes in cell-type SOM cluster C2, 283 are enriched in the gonad (1.9-fold over-representation, $p < 3.2e-31$), 119 overlap with spermatogenesis genes (4.0-fold over-representation, $p < 1.2e-40$), but only 27 overlap with soma-enriched genes (0.4-fold under-representation, $p < 3.5e-10$). Of the 1255 genes in cluster cell-type SOM C3, 476 are enriched in the hermaphrodite gonad (1.6-fold over-representation, $p < 2.1e-32$), 185 are enriched in the male profile (1.8-fold over-representation, $p < 3.7e-14$), but only 43 are enriched in the somatic

cell profile (0.3-fold under-representation, $p < 8.9e-24$). In addition, 185/1255 genes in C3 have been previously identified in sperm (Reinke et al. 2004).

Supplemental Protocols

SP1 Construction of cell-specific 3XFLAG::PAB-1 plasmids.

A Gateway (Invitrogen) compatible mRNA-tagging vector, *pSV41 (Pgateway::3XFLAG::PAB-1 + unc-119 minigene)* was constructed to provide a convenient method for inserting cell-specific promoters and for generating transgenic lines by bombardment. The *unc-119* minigene plasmid, MM051 (Maduro and Pilgrim 1995), was digested with HindIII, blunted with T4 DNA polymerase, digested with BamHI and the resulting fragment subcloned into plasmid pSV15 which contains the 3XFLAG::PAB-1 insert (Von Stetina et al. 2007), using BamHI and EcoRV restriction sites. The resulting plasmid was digested with KpnI then treated with T4 DNA polymerase for blunt end ligation with the Gateway vector conversion fragment A (Invitrogen). The resulting plasmid (pSV41) contains the *unc-119* minigene in the opposite orientation *vs*. promoter sequences inserted between the attR1 and attR2 sites upstream of the 3XFLAG::PAB-1 coding region.

SP2 Constructs generated using Gateway LR recombination with pSV41 plasmid

Cell-specific promoter fragments were generated from genomic DNA for *unc-122* (coelomocytes), *dpy-7* (hypodermis), *glr-1* (*glr-1*-expressing neurons) and subcloned into pCR8/GW-TOPO (Invitrogen). PCR amplicons and primer pairs were: *unc-122* (800 bp, *unc-122*_5prime/*unc-122*_3p); *dpy-7* (354 bp, *dpy-7*_5p/*dpy-7*_3p); *glr-1* (5.3 bp, *glr-1*_5p/*glr-1*_3p). The 716 bp *dat-1* (dopaminergic neurons) promoter was PCR amplified from plasmid pRN200, (a gift from R. Blakely), using primers *dat-1*p1 and *dat-1*p2 and cloned into pCR8/GW-TOPO. The *hlh-17* (CEP sheath cell) promoter was generated by amplifying the 4 kb promoter sequence upstream of the first ATG start (McMiller and Johnson 2005) with primers containing flanking attB recombination sites. This fragment was subcloned into pDONOR221 (Invitrogen) by a BP recombination reaction. LR recombination reactions were performed using pSV41 as the destination vector to create the following expression plasmids: pJW7 (P*glr-1*::3XFLAG::PAB-1), pJW5 (P*unc-122*::3XFLAG::PAB-1), pJW8 (P*dpy-7*::3XFLAG::PAB-1), pKW63 (P*dat-1*::3XFLAG::PAB-1), pMK107L (P*hlh-17*::3XFLAG::PAB-1) (**Supplemental Table S7**).

SP3 Generating cell-specific::3XFLAG::PAB-1 strains by microparticle bombardment.

Microparticle bombardment was used as previously described (Fox et al. 2005) to generate transgenic lines from plasmids containing the *unc-119* minigene. Additional modifications were used for plasmids pJW5, pJW7, pJW8 and pKW63, which were linearized by digesting with a unique ApaI restriction site upstream of the *unc-119* + minigene cassette. The reaction was then ethanol-precipitated and re-suspended in ddH2O. 8-10 μg of linearized plasmid was used to coat gold beads for bombardment. Animals were bombarded at 1800 psi, allowed to recover for 1 hr and washed to 7 x 100 mm NGM plates seeded with $OP50^{-1}$ bacteria. Plates were allowed to starve for 2 weeks at 23-25 °C and viable animals showing wildtype movement were picked for selfing. Transgenic lines were screened by anti-FLAG immunostaining (Von Stetina et al. 2007) to confirm specific expression (see **Supplemental Fig. S1**).

SP4 Other constructs generated for cell-specific profiling

The excretory cell-specific promoter P*clh-4* was amplified from genomic DNA using primers *clh-4* F and *clh-4* R. The 4 kb PCR product was then cloned into TOPO-2.1 (Invitrogen) to generate pDM1. pDM1 was used as a template to construct a Gateway donor vector by PCR amplification of the *clh-4* promoter using *clh-4* primers flanked with attB1 and attB2 sites. The promoter fragment was subcloned into pDONOR221 by performing a BP recombination reaction to create pDM2. pDM2 was combined with destination vector pSV41 in a LR recombination reaction creating the expression vector pJW6 (P*clh-4*::3XFLAG::PAB-1). The P*clh-4*::3XFLAG::PAB-1 cassette was then PCR amplified and 6ul of PCR product was coinjected with pRF4 [*rol-6 (su1006)*] at 25 ng into wild type animals. The transgenic line was integrated by gamma irradiation and outcrossed five times. The 861 bp putative promoter of *ttr-39* was amplified via PCR with primers pC04G21_5 and pC04G21_3 and inserted into pENTR D-TOPO (Invitrogen) via TOPO TA reaction. *Pttr-39* was then inserted upstream of 3XFLAG::PAB-1 via Gateway LR reaction with pSV41 resulting in the expression vector pSA2. The P*ttr-39*::3XFLAG::PAB-1 cassette was then amplified via PCR from pSA2 with primers pC04G21_5 and PAB1UTR_3 (5' CAATAGCAGCCAAATGCA 3'). The PCR reaction (12 μl) was co-injected with *dpy-5* rescuing plasmid pCes361 (25ng) into *dpy-5(e907)* animals. Gamma irradiation of the transgenic line yielded NC1645 *dpy-5(e907); wdIs31[Pttr-39::3xFLAG::PAB-1 dpy-5(+)]* IV. The integrant was

outcrossed five times prior to microarray profiling. Expression of the epitope-tagged PAB-1 for both the excretory cell and D-class motor neuron cell-specific lines was confirmed by immunostaining (Roy et al. 2002) with monoclonal mouse anti-FLAG antibodies (Sigma).

SP5 Isolation of cell-specific RNA by the mRNA tagging method

The mRNA-tagging method was used to isolate RNA from 12 different cell types in either larvae or young adults. Methods for obtaining RNA from L2 stage larvae were as previously described (Von Stetina et al. 2007). The following modifications were used for L4 stage larvae and young adults. Gravid adults were obtained from 20 x 150 mm culture plates (8P media, 8X peptone NGM) and treated with hypochlorite to release embryos. Arrested L1 larvae were isolated after hatching overnight at 20 °C in M9 buffer and transferred to Na^{22} seeded 8P plates for growth at 20 °C for 22-25hrs and then transferred to 23 °C for an additional 24-26hrs to reach mid-L4 stage larvae as shown by the appearance of a tree-shaped vulva (~80%). To obtain Young Adults (YAs), the arrested L1 larvae were grown on Na^{22}-seeded 8P plates at 20 °C for ~72hrs to reach early YA, as evidenced by a mature (everted) vulva in ~80% of animals. Synchronized L4 and YA animals were resuspended in 3 ml homogenization buffer and passed through a French press four times at 6,000 psi to obtain lysate as opposed to three times for L2 larvae. Mock IPs were performed to obtain reference data sets of non-specifically bound RNA for synchronized populations of L2, L4 and YA animals (Von Stetina et al. 2007). At least 3 independent RNA samples were prepared for each cell type and for each of the reference data sets.

SP6 Preparation of embryonic cells and primary cell culture

Methods used for generating preparations of embryonic cells and for primary cell culture have been previously described and are summarized here (Christensen et al. 2002). Embryos were obtained by hypochlorite treatment of synchronized populations of adult hermaphrodites and digested with chitnase to remove the egg shell. The resultant single-cell suspension of embryonic cells (in egg buffer) (Christensen et al. 2002) was either submitted directly for FACS-isolation of GFP-labeled cells (germ line precursors, BAG neurons) (see below) or was resuspended in L-15 cell culture medium, supplemented with 10% FBS and penicillin/streptomycin and plated at a density of 1×10^{-6} ml-1 on 1-well chamber slides (Nunc)

coated with poly-L-lysine (Sigma) (all other cell types, see **Table 1**). Primary cultures were maintained overnight at 23-25 °C.

SP7 Isolation of fluorescently-labeled embryonic cells by FACS

FACS was used to isolate 13 specific embryonic cell types each labeled with either GFP or DsRed2. Cells derived from freshly dissociated embryos were passed through a 5 μm filter (Durapore - Millipore) to remove debris. Primary cultures were examined 24 hr after plating to confirm expression of fluorescent markers (GFP and/or DsRed2). Cultured cells were resuspended in egg buffer and prepared for FACS as previously described (Fox et al. 2005). Dead cells were labeled by staining with propidium iodide (~1 μg/mL of cells) for GFP markers or with 7-AAD (Invitrogen) (~1-2 μg/mL of cells) for DsRed2 labels. Viable cells were isolated using either a FACStar Plus or FACSAria flow cytometer (75 μm nozzle, ~10,000-15,000 events/sec) (Becton Dickinson, San Jose, CA). FACS gates were empirically adjusted to achieve >80% purity for target cells (Table S1). The fraction of target cells (80-97%) for each cell type was determined by direct inspection in the fluorescence microscope 24 hr after plating on 4-well chamber slides coated with peanut lectin (Sigma) (Fox et al. 2005). AVA and AVE neurons were obtained by isolating cells expressing both GFP and DsRed2 markers (Spencer et al., manuscript in preparation). Yields of target cells ranged from ~5,000 to ~100,000 for each FACS run. At least 3 independent samples were collected for each target cell type. Reference samples for freshly dissociated embryos (early embryos or EE) and cells obtained from primary cultures (late embryos, LE) were obtained by isolating all viable cells from the wildtype (N2) strain (Fox et al. 2005).

SP8 RNA extraction from embryonic cells isolated by FACS

Cells collected for RNA isolation were sorted directly into Trizol LS (Invitrogen). The sample was extracted with chloroform, RNA precipitated with isopropanol, washed 2X with 75% EtOH and resuspended in RNAase-free H_2O. A DNA-free RNA purification kit (Zymo Research) was used to DNAase-treat and purify RNA according to the manufacturer's instructions. RNA quality and yield was determined using a Bioanalyzer (Agilent). Total amounts of RNA for each sample ranged from 600 pg to ~20 ng.

SP9 RNA amplification

The WT-Ovation Pico kit (NuGEN Technologies, Inc) was used to amplify RNA (0.6 ng to 10 ng starting material). 3 μg from each reaction was used to generate double stranded cDNA with the WT-Ovation Exon module (NuGEN Technologies, Inc). 4-5 μg of ds-cDNA was fragmented and labeled using the FL-Ovation Biotin V2 module (NuGEN Technologies, Inc).

SP10 Microarray hybridization

The *C. elegans* 1.0R tiling array (Affymetrix) contains > 3 million perfect match (PM)/mismatch (MM) probe pairs representing the *C. elegans* non-repetitive genome. Probes are 25 nt in length and tiled at an average distance of 25 nt as measured from the centers of adjacent probes. Double-stranded cDNA targets were used for hybridization because all probe sequences match a single DNA strand whereas individual transcripts can be derived from either the plus or minus strands. At least 3 independent replicates were obtained for each cell type. *Interse* Pearson correlation coefficients were calculated between replicates to ensure consistent sample preparation and hybridization.

SP11 RT-PCR to detect novel RNA

Single-stranded cDNA previously generated for microarray analysis was used as template for PCR-based validation of novel TARs. The –RT L2-intestine sample used the same RNA input for amplification, but reverse transcriptase was omitted and dH_2O was added to maintain constant volume. Primers (**Supplemental Table S8**) were designed to generate small amplicons of 75-150 bp using Batch-Primer3 (You et al. 2008). PCR conditions are as follows: 4 ng ss cDNA, 500 nM each primer, 1.5 μM MgCl2, 2.5 U GoTaq polymerase (Promega), and 200 nM dNTPs in a 50 μl reaction. The reactions were run in a MJ Research Minicycler with the following program: 94 °C 30 sec, 35 cycles of 94 °C sec, 58 °C 30 sec, 72 °C 30 sec. PCR products were electrophoresed on a 2% agarose gel and stained with ethidium bromide (Sigma). The products were visualized with a Bio-rad Gel Doc.

SP12 Quantitative PCR validation of novel TAR differential expression

Quantitative PCR (qPCR) was performed on ss cDNA used for microarray analysis. Primers (**Supplemental Table S9**) were designed to generate amplicons of 75 to 150 bp using Batch-Primer3 (You et al. 2008). Ssofast Eva green reaction mix was used with a 2-step 98 °C 2 sec, 60

°C 5 sec reaction and melting curve on a CFX96 Real Time Thermal Cycler (Bio-rad). Data were normalized to an internal 26S rRNA control using the Pfaffl method of determining relative expression (Pfaffl 2001).

SP13 Analysis of the Z2/Z3 tiling array data set and immunostaining to detect protein expression in Z2/Z3 in embryos.

Inspection of the P*pie-1::*PGL-1::GFP transgenic line used to isolate Z2/Z3 cells from the embryo (**Supplemental Table S1**) revealed ectopic expression in hypodermal cells in older embryos (unpublished data, **Supplemental Fig. S14**.) A comparison of the list of Z2/Z3-enriched transcripts to genes enriched in embryonic hypodermal cells detected 335 shared genes (2.7-fold higher than expected, p <1.9e-70, by hypergeometric test). These 335 shared genes were manually removed from the list of Z2/Z3-enriched genes to produce the "Z2/Z3 core" data set. Notably, overlap of Z2/Z3-enriched genes with embryonic muscle and neurons was lower than expected (0.4-fold and 0.1-fold, respectively) for a random distribution, indicating that the Z2/Z3 dataset is otherwise specific.

Embryos were fixed using methanol/acetone (Strome and Wood 1983). Antibody dilutions were 1:30,000 rabbit anti-PGL-1 (Kawasaki et al. 1998), 1:5000 rat anti-PGL-3 (Kawasaki et al. 2004), 1:1000 guinea pig anti-HTP-3 (MacQueen et al. 2005), 1:10,000 rabbit anti-REC-8 (SDI Q0802); 1:300 Alexa Fluor 488 goat anti-rabbit and anti-guinea pig IgG, and 1:300 Alexa Fluor 594 goat anti-rabbit and anti-rat IgG (Molecular Probes, Eugene, OR). Images were acquired, using identical exposure settings and times across stages, with a Volocity spinning disk confocal system (Perkin-Elmer/Improvision, Norwalk, CT) fitted on a Nikon Eclipse TE2000-E inverted microscope.

SP14 Mapping tiling probes to the *C. elegans* genome and its annotation

Perfect match (PM) 25mer tiling probe sequences were mapped to the *C. elegans* genome sequence (release WS200) (Rogers et al. 2008) using vmatch to detect all (direct and inverse) matches of length ≥17 with at most one mismatch or indel (Abouelhoda 2004). Only probes that perfectly aligned to a single genomic location were retained thereby discarding the most highly repetitive probes. Repeat information was kept for probes with multiple imperfect alignments as a filter for subsequent analyses. These included 70,189 PM tiling probes with

exact matches and an additional 113,054 probes with inexact matches leaving a total of 2,758,587 non-repetitive probes according to the above criteria.

SP15 Normalization of tiling array data

Prior to transcript identification, array data were first preprocessed with a background normalization technique that removed the mean array background of a sliding window neighborhood of 51x51 features (Borevitz et al. 2003; Zeller et al. 2009). In a second normalization step, array data were quantile normalized (Bolstad et al. 2003) followed by transcript normalization to reduce probe sequence bias as the third step (Zeller et al. 2008). Ridge regression models underlying transcript normalization were fitted on one replicate of each reference sample (**Table 1**, **Supplemental Table S1**); one regression model was fitted for each developmental stage data set and for hermaphrodite gonads (EE through YA and YA gonad), respectively) (**Supplemental Table S1**) and normalization of data from L4 males and L4 hermaphrodite soma, respectively) was carried out with the model fitted on L4 data. Adopting a fivefold cross-validation scheme, we evaluated how transcript normalization improved the separation between exon and background signal in comparison to DNA reference normalization (Huber et al. 2006). Evaluations were based on a global threshold above which probes are called exonic and background otherwise.

SP16 *De novo* transcript identification using mSTAD

For *de novo* identification of transcriptionally active regions (TARs) we adopted mSTAD, a previously proposed machine-learning based method (Laubinger et al. 2008; Zeller et al. 2008). For the analysis of cell type samples, a separate mSTAD model was optimized for each developmental stage by training on corresponding reference hybridization data and annotation information belonging to chromosomal chunks, each of which contained one annotated gene with half the intergenic space surrounding it (see **Supplemental Table S1**). The fitted models were used for transcript identification from all samples belonging to the same developmental stage (*e.g.*, the mSTAD model trained on EE-ref was used for transcript identification in EE-ref, EE BAG neurons, and EE germline precursors). Array data from developmental stages was analyzed with another set of models, each of which was trained on the same array sample for

which it identified transcripts (see **Supplemental Table S1**). One more model was trained for hermaphrodite gonads and TARs for L4 males and L4 hermaphrodite soma were identified with the L4 mSTAD model (see **Supplemental Table S1**).

SP17 Accuracy of TAR predictions assessed relative to the integrated transcript model

For comparison between the TARs from modMine (based on WS190) and mSTAD (based on WS200) with the integrated transcript model (Hillier et al. 2009) Gerstein et al., *in press*; WS170 based), we first mapped the probe sequences to all three genome versions using GenomeMapper (Schneeberger et al. 2009). http://www.1001genomes.org/downloads/genomemapper.html). All probe sequences that previously mapped uniquely to WS200 also mapped uniquely to WS170 and WS190. On the basis of the integrated transcript model, we labeled each probe as either exonic or non-exonic and used this labeling as a gold standard for the following evaluation. Probe-level sensitivity and precision were calculated for modMine and mSTAD TAR predictions derived from each data set available (**Supplemental Fig. S6**). Shown is the evaluation of mSTAD TAR predictions for all cell type and developmental stage data in two versions: original TARs and TARs that were confirmed to be expressed by a statistical test (see below). Additionally, we included the evaluation of unfiltered modMine TARs for the cell type data sets where predictions were available.

SP18 Determining overlap between TARs and known genes to identify novel TARs

TARs were compared to the following features annotated in WS199: Protein-coding genes (and their corresponding exon features), pseudogenes (and pseudogenic exons) and non-protein coding genes. We called TARs "unannotated" if they overlapped less than 20 nt with exons (of coding genes and pseudogenes) or with non-coding genes (**Fig. 2D**, **Supplemental Fig. S8**). Moreover, we determined the overlap between TARs and genes of the integrated transcript model to obtain "novel" TARs that neither overlapped with annotated features nor with exons of gold standard gene models by ≥20 nt (**Fig. 2E**, **F**).

Non-redundant (nr) TARs resulted from the union of positions inclusive to TARs obtained in any individual sample. Similarly, nr expressed TARs, nr differentially expressed TARs, nr unannotated TARs and nr novel TARs were obtained as the position-wise union of expressed, differentially expressed, unannotated and novel TARs, respectively (**Fig. 2F**). For each position within nr expressed TARs, we counted the number of individual samples in which an expressed TAR was detected. Partitioning expressed nrTARs according to overlap with known transcripts resulted in the histograms shown (**Fig. 4C**, **Supplemental Fig. S17**).

SP19 Probe set definition and estimation of expression for annotated genes and TARs

For each protein-coding gene model annotated in WS199 (Rogers et al. 2008), we constructed a probe set containing all PM tiling probes that could be perfectly aligned to corresponding constitutive exons. Repetitive probes (see above definition) were removed from gene probe sets and probe set information was converted into CDF. Similarly, for each TAR a probe set was generated from all probes it comprised. Subsequently, expression was estimated for genes and TARs with a minimal probe set size of three using RMA, which involves quantile normalization and summarization with median polish (RMA's default array-background normalization was omitted (Bolstad et al. 2003; Irizarry et al. 2003; Gautier et al. 2004).

SP20 Testing genes and TARs for expression above background

To establish whether expression of a particular gene was significantly higher than the array background intensity, we compared its hybridization signal to an empirical null model. For each gene probe set (or TAR probe set derived by mSTAD) we constructed a background probe set from an equally sized random sample of probes mapped to annotated intergenic regions. This sampling process was repeated until $\geq 10^6$ background probe sets had been collected. For a given biological sample, we established the null model from the median of the PM intensities of the background samples pooling replicate data. The empirical p-value of a gene's expression was estimated as the proportion of background probe sets with the same or higher median intensity than the median PM intensity of the gene probe set. Expression p-values were adjusted for multiple testing using the false discovery rate (FDR) method by Benjamini &

Hochberg (as implemented in the R function p.adjust(x, method="fdr") (Benjamini and Hochberg 1995).

SP21 Determining differentially expressed genes and TARs

Differentially expressed genes were identified using a linear model and an empirical Bayes moderated t-statistic (Smyth 2004) implemented in the Bioconductor package Limma (Smyth 2005). In contrast to annotated gene models for which coding regions are explicitly defined, TAR boundaries were obtained directly from segmentation of a single set of triplicate array data and thus varied between samples. When comparing TARs from different samples, we performed the test for differential expression with respect to the TARs identified in the sample tested for upregulation (**Supplemental Fig. S9B**, **C**, **Supplemental Fig. S12C**, upper panels), because on average, the accuracy of TAR boundaries increased with hybridization intensity (see **Fig. 2C**). Hence, instead of testing the TARs detected in sample A for downregulation versus sample B, we performed the reciprocal test assessing whether the TARs derived for sample B were upregulated in comparison to sample A (**Supplemental Fig. S9B**, **C**, **Supplemental Fig. S12C**, lower panels).

SP22 Additional correction for multiple testing of (differentially) expressed genes

Since FDRs for genes expressed above background or differentially expressed between samples were calculated for each individual sample (comparison), we applied an additional stringent Bonferroni-style correction for multiple testing (C.E. Bonferroni, 1935). We divided individual FDR estimates by the number of samples and sample comparisons, respectively, obtaining an adjusted FDR of 1.3×10^{-4} for expression above background and of 7.4×10^{-4} for differential expression (Table 2, see also **Supplemental Fig. S12**).

SP23 Entropy-based detection of selecitvely enriched genes

Gene-expression entropy was calculated based on the fold change relative to the corresponding reference sample. Fold-changes < 1.0e-5 were set to a pseudocount of 1.0e-5 before they were rescaled to the interval [0, 1] by dividing by the sum of fold changes across tissues and cell types for each gene. Afterwards, expression entropy was calculated as described (Schug et al. 2005). Selectively enriched genes were extracted from the set of enriched genes in a given tissue (FDR ≤ 0.05 and FC ≥ 2.0), if additionally (i) their fold change vs. reference was among the

upper 40% of the positive FC range observed for this gene across all tissues and (ii) their entropy was among the lower 40% of the distribution observed for all genes (i.e., $H \leq 3.03$).

SP24 Fold change histograms for differentially expressed genes

To generate the histograms of expression differences (**Supplemental Fig. S12A**), we first calculated the fold change between expression in a given cell type to the corresponding reference for all genes for which differential expression was detected in at least one comparison (FDR ≤ 0.05). We next determined the maximal fold change across cell types and depending on its direction tabulated the gene either as upregulated or downregulated (relative to reference).

SP25 Revealing developmental and cell-type specific expression patterns with self-organizing maps

Self-organizing maps (SOMs)(Kohonen 1982) were constructed using the Matlab SOM toolbox version 2.0 (Vesanto et al. 2000). As an input for SOM training, we selected the subset of genes detected as differentially expressed in the respective samples, applied log2 transformation and normalized by subtracting the mean expression across conditions for each individual gene (yielding mean-centered log expression). We chose SOM topologies with a hexagonal neighborhood consisting of 30 x 15 and 60 x 60 units for developmental and cell-type data sets, respectively. SOM training proceeded in 100 and 300 epochs with Gaussian neighborhood radius shrinking linearly from 5 to 1 and from 15 to 3 for developmental and cell type data sets, respectively.

Some regions were identified by *k*-means clustering as implemented in the Matlab SOM toolbox. We varied *k*, the pre-chosen number of clusters, from 1 to 15 and 1 to 20 for developmental and cell type data sets, respectively. To obtain a robust clustering, we only retained cluster information that was consistent in 75% of 50 - 100 replicates each of which resulted from the best out of five independent *k*-means runs with randomly initialized cluster centroids. We selected $k = 8$ and $k = 14$ for the developmental and the cell type data set, respectively, based on biological interpretation and silhouette coefficients, a means of assessing which SOM units lie tightly within clusters or which are in between clusters

(Rousseeuw 1987). In addition, we used silhouette coefficients to select the top half of SOM units close to cluster centroids, which resulted in the clusterings shown (**Fig. 6**, **Fig. 7**).

SP26 Analysis of the Z2/Z3 tiling array data set and immunostaining to detect protein expression in embryonic Z2/Z3 germ line precursor cells.

Inspection of the P*pie-1::*GFP::PGL-1 transgenic line used to isolate Z2/Z3 cells from the embryo (**Supplemental Table S1**) (see also **Supplemental Fig. S14**) revealed ectopic expression in hypodermal cells in older embryos (unpublished data). A comparison of the list of Z2/Z3-enriched transcripts to genes enriched in embryonic hypodermal cells detected 335 shared genes (2.7-fold higher than expected, $p<1.9e-70$, by hypergeometric test). These 335 shared genes were manually removed from the list of enriched Z2/Z3 genes to produce the "Z2/Z3 core" data set. Notably, overlap of Z2/Z3 enriched list with embryonic muscle and neurons was lower than expected (0.4-fold and 0.1-fold, respectively) for a random distribution thereby indicating that the Z2/Z3 dataset was otherwise specific.

Embryos were fixed using methanol/acetone (Strome and Wood 1983). Antibody dilutions were 1:30,000 rabbit anti-PGL-1 (Kawasaki et al. 1998), 1:1000 guinea pig anti-HTP-3 (MacQueen et al. 2005), 1:250 rabbit anti-REC-8; 1:300 Alexa Fluor 488 goat anti-rabbit IgG, 1:300 Alexa Fluor 488 goat anti-guinea pig IgG, and 1:300 Alexa Fluor 594 goat anti-rabbit IgG (Molecular Probes, Eugene, OR). Images were acquired with a Volocity spinning disk confocal system (Perkin-Elmer/Improvision, Norwalk, CT) fitted on a Nikon Eclipse TE2000-E inverted microscope.

SP27 Hypergeometric tests

To test for significant overlap between separate lists of genes, we applied the hypergeometric test as implemented by Jim Lund (http://elegans.uky.edu/MA/progs/overlap_stats.html). The number of genes in the whole genome was set at 18,451 for the number of genes represented on the *C. elegans* tiling array.

Microscopy. Isolated embryonic cells were imaged using differential interference contrast (DIC) and epifluorescence optics with a Zeiss Axiovert inverted microscope equipped with an ORCA ER (Hamamatsu) high-resolution, cooled CCD camera. Intact animals were imaged with a Zeiss

Axioplan compound microscope equipped with an ORCA ER camera and a Leica TCS SP5 confocal microscope.

Supplemental Tables

Supplemental Table S1. Detailed description of samples used for expression profiling.

Sample Cell types	Stage	Genotype	Description	FACS Cell Purity	DCC #	GEO #	RNA
emb-0hr-ref	embryo 0hr	N2	all viable freshly dissociated (0 hr) embryonic cells	100%	3172	GSE25350	total RNA
emb-BAG	embryo 0hr	nIs242[gcy-33::GFP];lin-15(n765)	embryonic BAG neurons	>82%	2499	GSE23769	total RNA
emb-GLP	embryo 0hr	bnIs1 (pie-1p8::GFP::PGL-1+unc119)	embryonic germ-line precursor cells	>95%	661	GSE23285	total RNA
emb-reference	embryo 24hr	N2	all viable cultured (24 hr) embryonic cells	100%	456	GSE23246	total RNA
emb-AVA	embryo 24hr	OH4326[otEx239(rig-3::GFP) pha-1(e2123)III] VH804[hdIs32(glr-1::DsRed2)	embryonic AVA neurons	>80%	459	GSE23249	total RNA
emb-GABA	embryo 24hr	CZ1200[juIs76(unc-25::GFP) II; lin-15(n765ts)X]	embryonic GABAergic motor neurons	>87%	468	GSE23257	total RNA
emb-bwm-v2	embryo 24hr	ccIs4251 [I;dpy-20(e1282) IV]	embryonic body wall muscle	>97%	470	GSE23260	total RNA
emb-coelomocytes	embryo 24hr	wyIs58 (opt-3::GFP::RAB-3; unc-122::RFP)	embryonic coelomocytes	nd	458	GSE23248	total RNA
emb-dop	embryo 24hr	dat-1::GFP (pRN2003)	embryonic dopaminergic motor neurons	>86%	467	GSE23257	total RNA
emb-intestine	embryo 24hr	wIs84	embryonic intestine	>91%	457	GSE23247	total RNA
emb-panneural	embryo 24hr	evIs111	embryonic neurons	>90%	455	GSE23245	total RNA
emb-A-class	embryo 24hr	wdIs5[unc-4::GFP; dpy-20(e1282)]	embryonic A-class motor neurons	>88%	654	GSE23278	total RNA
emb-hypodermis	embryo 24hr	/+; raIs/[rol-6(SU1006)+pdpy-7::GFP]	embryonic hypodermal cells	>85%	662	GSE23286	total RNA
emb-AVE	embryo 24hr	KM173 (opt-3::GFP[pRF4]); hdIs32 (glr-1::DsRed2)	embryonic AVE neurons	>88%	3173	GSE25351	total RNA
emb-PhM	embryo 24hr	ccIs9753[myo-2::GFP]	embryonic pharyngeal muscle	>91%	2548	GSE23770	total RNA
L2-glr	L2	unc-119 (ed1); [unc-119 (+); glr-1::3XFLAG::PAB-1]	L2 glutamate receptor neurons	na	658	GSE23282	poly A+ / total RNA
L2-A-class	L2	unc-119 (ed1); wdEx257 [unc-119 (+); unc-4::3XFLAG::PAB-1]	L2 A-class motor neurons	na	469	GSE23259	poly A+ / total RNA
L2-GABA_neurons	L2	dpy-5 (e907); wdIs31 [dpy-5 (+); pC04G2.1::3XFLAG::PAB-1]	L2 GABA neurons	na	466	GSE23256	poly A+ / total RNA

Sample	Stage	Genotype	Description	FACS Cell Purity	DCC #	GEO #	RNA
L2-bwm	L2	gaIs146 [(myo-3p::FLAG::PAB-1) + (sur-5::GFP)]	L2 body wall muscle	na	465	GSE23255	poly A+ / total RNA
L2-excretory_cell	L2	wdIs47 [clh-4::3XFLAG::PAB-1 + rol-6 (su1006)]	L2 excretory cell	na	464	GSE23254	poly A+ / total RNA
L2-intestine	L2	gaIs148 [(ges-1p::FLAG::PAB-1) +(sur-5::GFP)]	L2 intestine	na	463	GSE23253	poly A+ / total RNA
L2-panneural	L2	gaIs153 [(F25B3.3::FLAG::PAB-1) + (sur-5::GFP)]	L2 neurons	na	462	GSE23252	poly A+ / total RNA
L2-coelomocytes	L2	unc-119(ed1); wdEx638 [unc-119(+); unc-122::3XFLAG::PAB-1]	L2 coelomocytes	na	657	GSE23281	poly A+ / total RNA
L2-reference	L2	N2	mock-IP from L2 stage animals	na	461	GSE23251	poly A+ / total RNA
L3-L4-PVD_OLL	L3-L4	unc-119 (ed1); wdEx460 [unc-119 (+); ser-2prom3B::3XFLAG::PAB-1]	L3-L4 PVD and OLL neurons	na	460	GSE23250	poly A+ / total RNA
L3-L4-dop	L3-L4	unc-119 (ed1); wdEx637 [unc-119 (+); dat-1::3XFLAG::PAB-1]	L3-L4 dopaminergic neurons	na	655	GSE23279	poly A+ / total RNA
L3-L4-reference	L3-L4	N2	mock-IP from L3-L4 stage animals	na	659	GSE23283	poly A+ / total RNA
L3-L4-hypodermis	L3-L4	unc-119(ed1); wdEx626[unc-119+; dpy-7::3xFLAG::PAB-1]	L3-L4 hypodermis	na	2454	GSE23287	poly A+ / total RNA
YA-CEPsh	YA	unc-119 (?); nsIs191 [unc-119 (+); hlh-17::3XFLAG::PAB-1]	Young adult CEP sheath cells	na	660	GSE23284	poly A+ / total RNA
YA-ref	YA	N2	Mock-IP from young adult stage animals	na	656	GSE23280	poly A+ / total RNA
Gonad	YA	N2	Dissected gonad from YA hermaphrodite	na	481	GSE23269	total RNA
Whole Animal							
N2EE	early embryo	N2	Early embryos	na	476	GSE23265	total RNA
N2LE	late embryo	N2	Late embryos	na	479	GSE23268	total RNA
L1	L1	N2	L1 animals	na	484	GSE23270	total RNA
L2	L2	N2	L2 animals	na	472	GSE23261	total RNA
L3	L3	N2	L3 animals	na	474	GSE23263	total RNA
L4	L4	N2	L4 animals	na	473	GSE23262	total RNA
YA	YA	N2	Young adult animals	na	475	GSE23264	total RNA
soma-only	L4	glp-1(q224)	L4 somatic cells only	na	485	GSE23271	total RNA
male	L4	dpy-28(y1) III; him-8(e1489) IV	L4 males	na	478	GSE23267	total RNA

Supplemental Table S2. Genome coverage of non-redundant TARs filtered for expression over background.

Feature class	Samples	# nt in TARs	% of genome
nt covered by expressed nrTARs (expr. FDR≤0.05)	cells & stages	40,568,743	40.5% (100,286,002)
	cells	37,797,653	37.7% (100,286,002)
	stages	24,388,451	24.3% (100,286,002)

nt covered by unannotated expressed nrTARs (expr. FDR≤0.05)	cells & stages	11,959,806	29.5% (40,568,743)
	cells	10,924,840	28.9% (37,797,653)
	stages	3,807,275	15.6% (24,388,451)
nt covered by novel expressed nrTARs (expr. FDR≤0.05)	cells & stages	10,534,028	26.0% (40,568,743)
	cells	9,600,877	25.4% (37,797,653)
	stages	3,206,075	13.1% (24,388,451)

Supplemental Table S3. Non-redundant TARs overlapping with genic features from WormBase and integrated transcript model.

Feature class	Samples	#	% of WS199 annotation
WS199 exons of coding genes overlapping with nrTARs	cells & stages	119,521 exons	87.1% (137,193)
	cells	116,929 exons	
	stages	100,658 exons	
WS199 coding genes with exons overlapping with nrTARs	cells & stages	18,183 genes	91.3% (19,912)
	cells	18,049 genes	
	stages	15,400 genes	
WS199 genic regions overlapping with nrTARs	cells & stages	18,211 genes	91.5% (19,912)
	cells	18,074 genes	
	stages	15,466 genes	
ITM* exons (unique) overlapping with nrTARs	cells & stages	138,433 exons	87.8% (157,612)
	cells	135,654 exons	86.1% (157,612)
	stages	116,799 exons	74.1% (157,612)
ITM* genes with exons overlapping with nrTARs	cells & stages	19,325 genes	88.8% (21,774)
	cells	19,173 genes	88.1% (21,774)
	stages	16,152 genes	74.2% (21,774)

* ITM (Integrated Transcript Model)

Supplemental Table S4. Overlap between non-coding RNA features annotated in WormBase with non-redundant TARs (reported only if overlap ≥ 20 nt)

Feature class	Samples	# detected	% of WS199
all noncoding RNAs overlapping with nrTARs[a]	cells & stages	1,299 genes	9.8% (13,264)
	cells	1,188 genes	
	stages	495 genes	
ncRNAs overlapping with nrTARs[b]	cells & stages	1,033 genes	8.1% (12,830)
	cells	947 genes	
	stages	332 genes	
miRNAs overlapping with nrTARs	cells & stages	67 genes	42.7% (157)
	cells	62 genes	
	stages	25 genes	
snRNAs overlapping with nrTARs	cells & stages	43 genes	45.7% (94)
	cells	37 genes	
	stages	24 genes	
snoRNAs overlapping with nrTARs	cells & stages	122 genes	87.8% (139)
	cells	109 genes	
	stages	93 genes	

tRNAs overlapping with nrTARs	cells & stages	14 genes	63.6% (22)
	cells	13 genes	
	stages	9 genes	
rRNAs overlapping with nrTARs	cells & stages	19 genes	90.5% (21)
	cells	19 genes	
	stages	11 genes	
exons of pseudogenes overlapping with nrTARs	cells & stages	2,259 exons	52.2% (4,346)
	cells	2,165 exons	
	stages	1,249 exons	
pseudogenes with exons overlapping with nrTARs	cells & stages	882 pseudogenes	58.2% (1,516)
	tissues	850 pseudogenes	
	stages	462 pseudogenes	

[a] all features annotated as non-coding RNAs in the *C. elegans* WS199 annotation, including all sub-categories detailed in the table: unspecified ncRNAs, miRNAs, snRNAs, snoRNAs, tRNAs, rRNAs, scRNA, snlRNA and pseudogenes (overlap data for 1 scRNA and 4 snlRNAs not shown).
[b] features annotated as ncRNAs without further specification ("ncRNA" in the source-column and "gene" in the type-column of the *C. elegans* WS199 annotation in gff3 file format)

Supplemental Table S5. Z2/Z3-expressed genes involved in proliferation and later germline events.

Proliferation		Meiosis I/II		Oogenesis	
Core Z2/Z3	**Z2/Z3 & Hyp**	**Core Z2/Z3**	**Z2/Z3 & Hyp**	**Core Z2/Z3**	**Z2/Z3 & Hyp**
aak-1	*deps-1**	*coh-3*	*apc-10*	*cpg-1**	*cbd-1*
*asb-1**	*gld-3*	*cpb-3*	*him-3*	*cpg-4*	*cpg-2**
cdc-25.1	*glh-2*	*htp-1*	*htp-3**	*egg-1*	*daz-1**
gld-1	*glp-1*	*htp-2*	*prom-1*	*egg-2*	*mex-5*
*glh-1**	*ima-2**	*ksr-2*	*rec-8*	*egg-3*	*mex-6*
glh-4	*pgl-1**	*lab-1*	*sun-1*	*meg-1*	*moe-3*
*iff-1**		*mei-2*	*xnd-1/gak-1*	*mex-1*	*oma-2*
pgl-3		*mesp-1*		*pos-1*	*spn-4*
pie-1		*syp-2*		*puf-5*	*zif-1*
*prg-1**		*syp-3*			
*ppw-2**		*zim-3*			
puf-8					

Genes from the core Z2/Z3 list (Core Z2/Z3) and genes shared between the Z2/Z3-enriched and hypodermis-enriched lists (Z2/Z3 & Hyp) with established roles in proliferation, meiosis or oogenesis were evaluated for transcription *in vivo*. Genes highlighted with asterisks show evidence of zygotic transcription in Z2/Z3 by *in situ* hybridization (NextDB; http://nematode.lab.nig.ac.jp).

Supplemental Table S6. Analysis of in situ hybridization signal for 100 Z2/Z3-enriched and 100 randomly-selected genes.

Z2/Z3-ENRICHED GENES

Chr	No. genes	No. genes in NEXTDB	Positive in adult germ line	Positive in Z2/Z3 in L1s	Positive in Z2/Z3 in embryos
I	24	18	13	4	5
II	36	18	13	0	1
III	22	18	14	4	1
IV	28	18	12	0	0
V	30	19	8	3	1
X	12	9	3	0	0
All	152	100	63	11 [a,c]	8 [b,c]

RANDOM GENES

Chr	No. genes	No. genes in NEXTDB	Positive in adult germ line	Positive in Z2/Z3 in L1s	Positive in Z2/Z3 in embryos
I	29	18	10	0	0
II	40	18	8	1	0
III	27	18	12	1	0
IV	37	18	10	1	0
V	65	13	5	0	0
X	30	15	4	0	0
All	228	100	49	3 [d]	0

[a] *glh-1, ppw-2, hil-5, iff-1, hil-4, asb-1, ucr-2.3, cpg-1, ife-3, rpl-11.1, prg-1*
[b] *glh-1, ppw-2, hil-5, iff-1, hil-4, lsl-1, pas-5, hmg-12*
[c] Z2/Z3 signal for *hil-5* and *hmg-12* RNA may be due to persistence of maternal load. Z2/Z3 signal for other genes was preceded by a stage at which signal was very low or undetectable.
[d] *top-2, cra-1, mel-46*

Supplemental Table S7. Primers used to amplify promoters for generating mRNA-tagging plasmids (see Methods in main text).

Promoter	Length	Primers	Primer Sequence
Pglr-1	5298 bp	glr-1_5p	CTGTAGCCGGTATGCACTG
		glr-1_3p	GTGAATGTGTCAGATTGGG

Pttr-39	861 bp	C04G2.1_5	ATTATTATTTCTATCGGCT
		C04G2.1_3	ATGATTTTTTGTTTTAAC
Pclh-4	4033 bp	clh-4 F	CGACAAAATTCAGGCGAGAAAGC
		clh-4 R	CCACATTGGTGGTGCTATGAATTCAGC
Punc-122	800 bp	unc122_5prime	GTAATGTTTTCCCGCTGATA
		unc122_3prime	ATTGTGAGCCCAATGAAGTA
Pdat-1	716 bp	dat-1p1	CCATGAAATGGAACTTGAATCC
		dat-1p2	GGCTAAAAATTTGTTGAGATTCG
Pdpy-7	354 bp	dpy7_5p	ACAATCTATTTGTAATCTCATTCC
		dpy7_3p	GGAACAAAATGTAAGAATATTC

Supplemental Table S8. Primers used for reverse-transcriptase PCR validation of novel TARs (see Methods in main text)

Name	length	Tm	GC%	sequence	Amplicon length
TAR_E-pan_77592_F	20	60.23	55	TTCCTCTGGAACTGGACAGG	104
TAR_E-pan_77592_R	20	59.35	55	CCCTGAGCTTTCCACGTAGT	
TAR_E-pan_77593_F	20	59.66	45	CACCCCAAAAATACCTGGAA	131
TAR_E-pan_77593_R	20	59.95	40	TTGATTGCGATGAAAAGCAG	
TAR_E-bwm_63930_F	20	59.94	45	ATCATCCCAAACGCTTTCAC	123
TAR_E-bwm_63930_R	20	58.88	50	TTTCCACTATGCAGCTGACC	
TAR_E-coel_06773_F	20	59.8	45	AAGAGGGTCCAACCGAATTT	121
TAR_E-coel_06773_R	20	59.96	50	CCGGGACTGTGCAAGATAAT	
TAR_L2-int_19313_F	20	59.65	45	GCCGAGATTGAGGAAAAATG	112
TAR_L2-int_19313_R	21	58.78	48	CCGGTACTTATTCGTTTGCTC	
TAR_E-AVE_07153_F	20	59.8	45	AAGAGGGTCCAACCGAATTT	121
TAR_E-AVE_07153_R	20	59.96	50	CCGGGACTGTGCAAGATAAT	
TAR_E-AVE_52539_F	20	59.14	50	GGCTGGTTCTGAAGTCCAAT	137
TAR_E-AVE_52539_R	20	59.88	45	GTGTTGCAGGTTGGGTTTTT	

Supplemental Table S9. Primers used for quantitative real-time PCR validation of differential expression of novel TARs (see Methods in main text).

Name	Oligo length	Tm	%GC	Sequence	Amplicon length
TAR_L3-L4-dop_35596_F	21	60.13	42.86	TTGAACCCGAAAAAGTGTCTG	97
TAR_L3-L4-dop_35596_R	21	59.27	47.62	TGGAGTCAAGGATTCTGAAGG	
TAR_L3-L4-hypo_34173_F	21	60.13	42.86	TTGAACCCGAAAAAGTGTCTG	97
TAR_L3-L4-hypo_34173_R	21	59.27	47.62	TGGAGTCAAGGATTCTGAAGG	
TAR_L3-L4-hypo_36011_F	20	60.86	45	CACATTGAGCGGGAAATGAT	130

TAR_L3-L4-hypo_36011_R	21	58.46	47.62	TTCTCTTCGGAGATGTTCCTC	
TAR_YA-CEPsh_52288_F	20	60.31	40	ACGTTCCAATCGGAATTCAA	125
TAR_YA-CEPsh_52288_R	20	59.14	45	AGACCACCAGCATGTTCAAA	
TAR_L2-exc-cell_23646_F	20	60.05	40	TCAAATGTGCCCAATGAGAA	115
TAR_L2-exc-cell_23646_R	20	58.49	45	GACCGATTCATGGAAGTTCA	
TAR_L2-exc-cell_40020_F	20	60.28	45	TTTGTGTGTGGCAAGAGGAA	128
TAR_L2-exc-cell_40020_R	20	60.58	50	TGGTCGTACCCCAAATATCG	
TAR_L2-glr_26400_F	20	60.06	50	AGTGTCAACAGCTGCAATCG	134
TAR_L2-glr_26400_R	20	59.99	60	CCAGTCCTCTGCCTGTCTTC	
TAR_L2-A-class_72252_F	20	59.66	55	GCTTCTGGTCCATCCAAGAC	131
TAR_L2-A-class_72252_R	20	60.48	55	GGCACCAGGATAATCTCACG	
TAR_E-hypo_29647_F	24	58.90	45.83	CACTGGTGTAGAAGAACAAGAGGT	94
TAR_E-hypo_29647_R	20	59.17	45	AGGTCGTGCATTTTTCCTTC	

Supplemental Figures

Supplemental Fig. S1. New strains constructed for mRNA-tagging stained for 3XFLAG to confirm specific expression of transgenes.

Strains are listed in **Table 1** and **Supplemental Table S1**.
(A) Dopaminergic neurons
(B) Hypodermis
(C) *Pglr-1*+ neurons
(D) CEP sheath cells
(E) Coelomocytes
(F) Excretory cell

Supplemental Fig. S2. Marker gene expression in the tissue and cell type panel profiled.

Lines visualize tiling array-based expression estimates of genes with promoters were used as markers for isolation of specific tissues or cell types (see key, **Supplemental Table S1**, **Supplemental Fig. S1**). Red arrowheads indicate the tissue- or cell-type sample marked with the corresponding gene (for example*, Pgcy-33* was used as a marker for BAG neurons). Note that LE AVA and LE AVE neuron samples were acquired using two markers in combination. For isolation of RNA from PVD and OLL neurons, a modified *ser-2* promoter with expression specific to these neurons was used.

Supplemental Fig. S3. Principal component analysis of expression estimates shows agreement in clustering between cell type and developmental stage data.

Principal component 1 identifies the striking difference between the gonad and all other profiles. Germ line contribution to other stages (EE, LE, L4, L4 males) separates those profiles

from other data sets (*e.g.,* early larval stages L1, L2, L3) along the X-axis. Principal component 2 separates data sets on the Y-axis based on embryonic or postembryonic stages. All embryonic stages are colored red and postembryonic stages are blue.

Supplemental Fig. S4. Correlation between expression estimates for the tissue data and the developmental stage data.

Matrix tabulates Pearson correlation coefficients for pairwise comparisons between gene expression values obtained by pooling tissue/cell data sets belonging to the same developmental stage (rows) and expression values from developmental stage data sets (columns). For this analysis, all cell type data sets belonging to the same stage were treated as replicates to obtain one average expression value per gene and stage. Correlation strength is color-coded (see color bar). For example, pooled tissue data for L2 correlates best (r = 86%) with the L2 sample in the whole animal developmental stage data sets.

Supplemental Fig. S5. Genes expressed in cell type samples

Bar height corresponds to the number of genes detected as expressed above background. The fraction of genes with higher expression in a given cell type than in the corresponding reference sample is indicated (see key).

Supplemental Fig. S6. Comparison of tiling array-based TARs to the integrated transcript model at tiling probe resolution.

Cross-validation accuracy of modMINE TARs (blue squares), mSTAD TARs for cell types (green circles), expressed mSTAD TARs for cell types (orange diamonds), mSTAD TARs for developmental stages (brown triangles), expressed mSTAD TARs for developmental stages (green triangles) was assessed relative to integrated transcript models. At probe level, sensitivity is defined as the percentage of probes included in both tiling array TARs as well as exons of integrated transcript models, among all probes in exons of integrated transcript models. Precision is defined as the percentage of probes in both tiling array TARs as well as exons of integrated transcript models, among those included in tiling array-based TARs.

Supplemental Fig. S7. Screenshot from genome browser.

This customized generic genome browser (Stein et al. 2002) shows mSTAD TAR predictions for two example regions in comparison to modMine TARs (Gerstein et al., *in press*). Annotated gene models (WormBase WS199) are shown on top ("Gene models"). Expression coloring of mSTAD TARs has been replaced by gray.
(A) Chromosome I positions 948,000 - 958,400
(B) Chromosome II positions 1,210,000 - 1,255,000

Supplemental Fig. S8. Overlap of TARs with annotated non-coding RNAs.

(A) Overlap between non-redundant TARs (nrTARs), the portion detected as expressed and annotated miRNAs. 57.3% of annotated miRNAs do not overlap with nr TARs.

(B) Overlap between non-redundant TARs (nrTARs), the portion detected as expressed and annotated snoRNAs. Only 12.2% of annotated snoRNAs do not overlap with nr TARs.
(C) Overlap between non-redundant TARs (nrTARs), the portion detected as expressed and annotated pseudogenes. 41.8% of annotated pseudogenes do not overlap with nr TARs.

Supplemental Fig. S9. TAR predictions

(A) Predicted transcriptionally active regions (TARs) per tissue/cell type for which expression could be confirmed by a statistical test (expressed) (see Methods in main text for details). EE (Early Embryo), LE (Late Embryo), L2 larva, L3/L4 larva, YA (Young Adult).
(B) TARs detected as differentially expressed between tissue samples (labels as in A) and reference samples.
(C) Novel TARs detected as differentially expressed between tissue samples (labels as in A) and reference samples.

Supplemental Fig. S10. A subset of non-redundant TARs is differentially expressed between cell types and corresponding refererences

Unannotated and novel TARs and their overlap with TARs expressed above background and differentially expressed between samples. Unannotated TARs are defined as TARs without significant overlap (≥ 20bp) with exons of annotated coding genes, pseudogenes and non-coding RNAs. Novel TARs are defined as the subset of unannotated TARs without significant overlap (≥ 20bp) with exons in the integrated transcript model (see main text for details). Expressed TARs were determined using a statistical test and differentiall expressed TARs were identified in comparisons between cell types or developmental stages using a linear model approach (see main text).

Supplemental Fig. S11. Overlap between TARs described here and predicted long ncRNA

Displayed are overlaps between predicted long ncRNA (Liu et al. DOI:10.1101/gr.110189.110) and nrTARs as well as the subset of expressed and novel TARs, showing that, *e.g.*, the majority (2,538 Kb) of predicted long ncRNAs overlaps with the novel nrTARs from this study.

Supplemental Fig. S12. Expression fold changes of differentially expressed genes and TARs.

(A) Histogram depicting numbers of gene models binned according to maximal relative expression (fold change) in specific cell types vs. corresponding reference samples derived from all cells (FDR ≤ 0.05) (see supplemental protocol SP24).
(B) Histogram counting gene models differentially expressed between cell types and corresponding reference samples (FDR ≤ 0.05). Expression fold change is color-coded (see key between B and C).
(C) Histogram showing novel TARs that are differentially expressed between cell types and corresponding reference samples (FDR ≤ 0.05). Expression fold change is color-coded (see key between B and C).

Supplemental Fig. S13. Overlap between genes detected as expressed over the array background and genes found to be differentially expressed at two different levels of confidence.

(A) Venn diagram showing the overlap between the union of genes detected as expressed in any data set and the union of genes detected as differentially expressed across all pair-wise comparisons performed at a false discovery rate of 5%.
(B) Venn diagram showing the overlap between expressed genes and differentially expressed ones as in (A) except that a more stringent FDR cutoff was required (0.14% and 0.11%, respectively) corresponding to a Bonferroni-style correction which accounts for a possible accumulation of type-I errors (see Methods in main text).

Supplemental Fig. S14. Overlap between sets of genes specifically enriched in certain tissues or cell types

Percentage of genes found in the overlap between two sets of specifically enriched genes is shown as a gray-scale heatmap (see key). For tissues sampled at two time points, red box outlines indicate overlap between corresponding samples.

Supplemental Fig. S15. Seven-transmembrane proteins are overrepresented among genes that are selectively enriched in neuronal samples.

Pie chart shows the subset of most highly over-represented protein domains for 1,242 genes selectively enriched in any of the neuronal samples, but not enriched in any other tissue or cell type profiled here.

Supplemental Fig. S16. Detection of non-redundant differentially expressed genes as a function of data set size.

(A) Differentially expressed genes were detected from comparisons of each of the 25 tissue-derived data sets to its corresponding reference. Cumulative totals were determined for 100 separate experiments of randomized sample order with a fixed FDR < 5% for each comparison (black) as well as with an FDR divided by the total number of sample comparisons as a stringent correction for multiple testing (red). The set of non-redundant genes was obtained as the union of genes differentially expressed in pairwise comparisons. Plotted are the mean (solid line) and the 5th and 95th percentile (error bars) of 100 experiments with randomized sample order.
(B) Random subsets of 7 developmental stage data sets were considered. Differentially expressed genes were obtained from all pairwise comparisons between samples in a given subsets. Shown are mean, 5th and 95th percentile of 100 permutation experiments as in (A).

Supplemental Fig. S17. Coverage of the genome by transcripts expressed in developmental stage data.

Nucleotides in non-redundant TARs (for 7 samples corresponding to developmental stages) were binned according to the number of samples for which a TAR was detected at the given position. Bars pointing up correspond to expressed TARs overlapping with exons of coding genes (in WS199) and those of the integrated transcript model (also mapped to WS199). Bars pointing down correspond to nucleotides in expressed novel TARs (see main text for definition)

broken down by their location relative to annotated protein coding gene models (see legend). Positions were classified as proximal if within 500 bp around any annotated gene and otherwise as distal.

Supplemental Fig. S18. Online resource for differential gene expression visualization.

The adapted screenshot shows that *ceh-6*, a POU domain transcription factor, is highly expressed in excretory cells relative to other cell types. Two of its target genes, *clh-4* and *pgp-3*, show peaked expression in L2 excretory cells as well (see Supplemental Results SR3 for details).

Supplemental Fig. S19. GO analysis of genes enriched in germ line precursor cells.

GO term analysis was performed on genes enriched in the primordial germ cells Z2/Z3. The top ten GO Biological Process IDs are shown with the corresponding number of genes annotated with the GO term in parentheses (FDR ≤ 0.01).

Supplemental Fig. S20. Additional SOM clusters found in the developmental stage data set

Mean-centered log2-expression values of genes corresponding to additional clusters not shown in **Fig. 6** are plotted for the 25% of best-fitting genes (as gray lines).
(A) CS2 is enriched in genes with highest expression in the embryo.
(B) CS4 contains genes with fluctuating expression between successive embryonic and larval stages.
(C) Genes in CS6 show progressively stronger expression with each developmental stage with a plateau in L3/L4 larvae.
(D) CS7 features a cluster in which transcripts show a modest increase during embryonic to L4 development (see main text for discussion).

Supplemental Fig. S21. SOM component planes for cell-type data.

(A) SOM component planes visualizing the correspondence between SOM units and mean-centered log2-expression values for one cell type sample each are shown (see labels above plots and color bar).
(B) *k*-means clusters of SOM units (k=14, see Methods in main text for details).
(C) SOM clusters after filtering by silhouette coefficients to identify core regions (see Methods for details).

Supplemental Fig. S22. SOM regions and cluster-specific gene expression patterns for cell-type data.

Top left shows core regions of cell type SOM each representing a cohort of genes with similar cell type expression pattern. Expression patterns of the top 20% best-matching genes are shown as box plots for each cluster (if not already contained in **Fig. 7**, main text). Cell-types are as indicated at bottom with gray shading indicating reference samples and colored shading highlighting cell types with elevated expression in a particular cluster (see main text for discussion).

Supplemental Fig. S23. All motifs detected in developmental stage and tissue data sets

Complete table of motifs detected by FIRE as over- and under-represented in (A) developmental profile clusters and (B) cell-type profile clusters (see **Fig. 8** and Methods for additional details).

Supplemental Fig. S24. Correlations between regulatory motifs detected in stage and cell-type expression clusters.

To identify functional modules of motifs, FIRE determines whether pairs of motifs are informative for expression of a gene in a cluster. A heat map shows whether each motif pair has a positive, negative or neutral interaction. Significant interactions are shown with colored borders (promoter-promoter = blue, UTR-UTR = pink, promoter-UTR = green). A correlation between positions of motifs is indicated with a cross.
(A) Interactions between motifs found in stage *k*-means clusters. The first five motifs show significant interactions, which define a functional module.
(B) Interactions between motifs found in cell-type *k*-means clusters. The first two motifs show a significant interaction to define a functional module.

Supplemental Fig. S25. The excretory cell expresses many transport-related genes.

(A) The excretory cell body is located ventral to the terminal bulb of the pharynx, and extends canals anteriorly and posteriorly along either side of the body. These canals collect ions and fluid for osmoregulation.
(B) Pie chart showing that the top ten GO molecular function categories enriched in the excretory cell profile correspond to transporter proteins (FDR < 0.01).
(C) - (E) GFP-reporters selected from excretory cell enriched genes demonstrate robust expression in the excretory canal (C) *cnx-1*, 2.2X, (D) *ral-1*, 3.4X, (E) *srv-1*, 4.5X.

SUPPLEMENTAL REFERENCES

Abouelhoda, M. 2004. Replacing suffix trees with enhanced suffix arrays. *Journal of Discrete Algorithms* **2**(1): 53-86.

Altun, Z.F., Herndon, L.A., Crocker, C., Lints, R. and Hall, D.H. (ed.s). 2002-2010. WormAtlas.

Armstrong, K.R. and Chamberlin, H.M. 2010. Coordinate regulation of gene expression in the C. elegans excretory cell by the POU domain protein CEH-6. *Mol Genet Genomics* **283**(1): 73-87.

Benjamini, Y. and Hochberg, Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B (Methodological)* **57**(1): 289-300.

Bolstad, B.M., Irizarry, R.A., Astrand, M., and Speed, T.P. 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics (Oxford, England)* **19**(2): 185-193.

Borevitz, J.O., Liang, D., Plouffe, D., Chang, H.-s., Zhu, T., Weigel, D., Berry, C.C., Winzeler, E., and Chory, J. 2003. Large-scale identification of single-feature polymorphisms in complex genomes. *Genome research* **13**(3): 513-523.

Burglin, T.R. and Ruvkun, G. 2001. Regulation of ectodermal and excretory function by the C. elegans POU homeobox gene ceh-6. *Development* **128**(5): 779-790.

Christensen, M., Estevez, A., Yin, X., Fox, R., Morrison, R., McDonnell, M., Gleason, C., Miller, D.M., 3rd, and Strange, K. 2002. A primary culture system for functional analysis of C. elegans neurons and muscle cells. *Neuron* **33**(4): 503-514.

Fox, R.M., Von Stetina, S.E., Barlow, S.J., Shaffer, C., Olszewski, K.L., Moore, J.H., Dupuy, D., Vidal, M., and Miller, D.M., 3rd. 2005. A gene expression fingerprint of C. elegans embryonic motor neurons. *BMC Genomics* **6**(1): 42.

Gautier, L., Cope, L., Bolstad, B.M., and Irizarry, R.a. 2004. affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics (Oxford, England)* **20**(3): 307-315.

Hahn-Windgassen, A. and Van Gilst, M.R. 2009. The Caenorhabditis elegans HNF4alpha Homolog, NHR-31, mediates excretory tube growth and function through coordinate regulation of the vacuolar ATPase. *PLoS Genet* **5**(7): e1000553.

Hare, E.E. and Loer, C.M. 2004. Function and evolution of the serotonin-synthetic bas-1 gene and other aromatic amino acid decarboxylase genes in Caenorhabditis. *BMC Evol Biol* **4**: 24.

Hillier, L.W., Reinke, V., Green, P., Hirst, M., Marra, M.A., and Waterston, R.H. 2009. Massively parallel sequencing of the poly-adenylated transcriptome of C. elegans. *Genome Res*.

Huber, W., Toedling, J., and Steinmetz, L.M. 2006. Transcript mapping with high-density oligonucleotide tiling arrays. *Bioinformatics (Oxford, England)* **22**(16): 1963-1970.

Hunt-Newbury, R., Viveiros, R., Johnsen, R., Mah, A., Anastas, D., Fang, L., Halfnight, E., Lee, D., Lin, J., Lorch, A. et al. 2007. High-throughput in vivo analysis of gene expression in Caenorhabditis elegans. *PLoS Biol* **5**(9): e237.

Irizarry, R.A., Bolstad, B.M., Collin, F., Cope, L.M., Hobbs, B., and Speed, T.P. 2003. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* **31**(4): e15.

Kawasaki, I., Amiri, A., Fan, Y., Meyer, N., Dunkelbarger, S., Motohashi, T., Karashima, T., Bossinger, O., and Strome, S. 2004. The PGL family proteins associate with germ granules and function redundantly in Caenorhabditis elegans germline development. *Genetics* **167**(2): 645-661.

Kawasaki, I., Shim, Y.H., Kirchner, J., Kaminker, J., Wood, W.B., and Strome, S. 1998. PGL-1, a predicted RNA-binding component of germ granules, is essential for fertility in C. elegans. *Cell* **94**(5): 635-645.

Kohonen, T. 1982. Self-organized formation of topologically correct feature maps. *Biological Cybernetics* **43**(1): 59-69.

Lanjuin, A., VanHoven, M.K., Bargmann, C.I., Thompson, J.K., and Sengupta, P. 2003. Otx/otd homeobox genes specify distinct sensory neuron identities in C. elegans. *Dev Cell* **5**(4): 621-633.

Laubinger, S., Zeller, G., Henz, S.R., Sachsenberg, T., Widmer, C.K., Naouar, N., Vuylsteke, M., Scholkopf, B., Rätsch, G., and Weigel, D. 2008. At-TAX: a whole genome tiling array resource for developmental expression analysis and transcript identification in Arabidopsis thaliana. *Genome Biol* **9**(7): R112.

MacQueen, A.J., Phillips, C.M., Bhalla, N., Weiser, P., Villeneuve, A.M., and Dernburg, A.F. 2005. Chromosome sites play dual roles to establish homologous synapsis during meiosis in C. elegans. *Cell* **123**(6): 1037-1050.

Maduro, M. and Pilgrim, D. 1995. Identification and cloning of unc-119, a gene expressed in the Caenorhabditis elegans nervous system. *Genetics* **141**(3): 977-988.

Mah, A.K., Armstrong, K.R., Chew, D.S., Chu, J.S., Tu, D.K., Johnsen, R.C., Chen, N., Chamberlin, H.M., and Baillie, D.L. 2007. Transcriptional regulation of AQP-8, a Caenorhabditis elegans aquaporin exclusively expressed in the excretory system, by the POU homeobox transcription factor CEH-6. *J Biol Chem* **282**(38): 28074-28086.

Mah, A.K., Tu, D.K., Johnsen, R.C., Chu, J.S., Chen, N., and Baillie, D.L. 2010. Characterization of the octamer, a cis-regulatory element that modulates excretory cell gene-expression in Caenorhabditis elegans. *BMC Mol Biol* **11**: 19.

McGhee, J.D., Fukushige, T., Krause, M.W., Minnema, S.E., Goszczynski, B., Gaudet, J., Kohara, Y., Bossinger, O., Zhao, Y., Khattra, J. et al. 2009. ELT-2 is the predominant transcription factor controlling differentiation and function of the C. elegans intestine, from embryo to adult. *Dev Biol* **327**(2): 551-565.

McGhee, J.D., Sleumer, M.C., Bilenky, M., Wong, K., McKay, S.J., Goszczynski, B., Tian, H., Krich, N.D., Khattra, J., Holt, R.A. et al. 2007. The ELT-2 GATA-factor and the global regulation of transcription in the C. elegans intestine. *Dev Biol* **302**(2): 627-645.

McMiller, T.L. and Johnson, C.M. 2005. Molecular characterization of HLH-17, a C. elegans bHLH protein required for normal larval development. *Gene* **356**: 1-10.

Nelson, F.K., Albert, P.S., and Riddle, D.L. 1983. Fine structure of the Caenorhabditis elegans secretory-excretory system. *J Ultrastruct Res* **82**(2): 156-171.

Pauli, F., Liu, Y., Kim, Y.A., Chen, P.J., and Kim, S.K. 2005. Chromosomal clustering and GATA transcriptional regulation of intestine-expressed genes in C. elegans. *Development*.

Pfaffl, M.W. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* **29**(9): e45.

Reece-Hoyes, J.S., Deplancke, B., Shingles, J., Grove, C.A., Hope, I.A., and Walhout, A.J.M. 2005. A compendium of Caenorhabditis elegans regulatory transcription factors: a resource for mapping transcription regulatory networks. *Genome biology* **6**(13): R110-R110.

Rogers, A., Antoshechkin, I., Bieri, T., Blasiar, D., Bastiani, C., Canaran, P., Chan, J., Chen, W.J., Davis, P., Fernandes, J. et al. 2008. WormBase 2007. *Nucleic acids research* **36**(Database issue): D612-617.

Rousseeuw, P. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* **20**(1): 53-65.

Roy, P.J., Stuart, J.M., Lund, J., and Kim, S.K. 2002. Chromosomal clustering of muscle-expressed genes in Caenorhabditis elegans. *Nature* **418**(6901): 975-979.

Schneeberger, K., Hagmann, J.r., Ossowski, S., Warthmann, N., Gesing, S., Kohlbacher, O., and Weigel, D. 2009. Simultaneous alignment of short reads against multiple genomes. *Genome biology* **10**(9): R98-R98.

Schug, J., Schuller, W.P., Kappen, C., Salbaum, J.M., Bucan, M., and Stoeckert, C.J., Jr. 2005. Promoter features related to tissue specificity as measured by Shannon entropy. *Genome Biol* **6**(4): R33.

Smyth, G.K. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical applications in genetics and molecular biology* **3**(1): Article3.

Smyth, G.K. 2005. Limma: linear models for microarray data. In *Bioinformatics and Computational Biology Solutions using R and Bioconductor*, (ed. R. Gentleman, Carey, V.J., Huber, W., Irizarry, R.A., and Dudoit, S.), pp. 397-420. Springer, New York.

Stein, L.D., Mungall, C., Shu, S., Caudy, M., Mangone, M., Day, A., Nickerson, E., Stajich, J.E., Harris, T.W., Arva, A. et al. 2002. The generic genome browser: a building block for a model organism system database. *Genome Res* **12**(10): 1599-1610.

Strome, S. and Wood, W.B. 1983. Generation of asymmetry and segregation of germ-line granules in early C. elegans embryos. *Cell* **35**(1): 15-25.

Vesanto, J., Himberg, J., Alhoniemi, E., and Parhankangas, J. 2000. SOM Toolbox for Matlab 5.

Voglis, G. and Tavernarakis, N. 2008. A synaptic DEG/ENaC ion channel mediates learning in C. elegans by facilitating dopamine signalling. *Embo J* **27**(24): 3288-3299.

Von Stetina, S.E., Watson, J.D., Fox, R.M., Olszewski, K.L., Spencer, W.C., Roy, P.J., and Miller, D.M., 3rd. 2007. Cell-specific microarray profiling experiments reveal a comprehensive picture of gene expression in the C. elegans nervous system. *Genome Biol* **8**(7): R135.

Wang, X., Zhao, Y., Wong, K., Ehlers, P., Kohara, Y., Jones, S.J., Marra, M.A., Holt, R.A., Moerman, D.G., and Hansen, D. 2009. Identification of genes expressed in the hermaphrodite germ line of C. elegans using SAGE. *BMC Genomics* **10**: 213.

You, F.M., Huo, N., Gu, Y.Q., Luo, M.C., Ma, Y., Hane, D., Lazo, G.R., Dvorak, J., and Anderson, O.D. 2008. BatchPrimer3: a high throughput web application for PCR and sequencing primer design. *BMC Bioinformatics* **9**: 253.

Zeller, G., Henz, S.R., Laubinger, S., Weigel, D., and Rätsch, G. 2008. Transcript normalization and segmentation of tiling array data. *Pac Symp Biocomput*: 527-538.

Zeller, G., Henz, S.R., Widmer, C.K., Sachsenberg, T., Rätsch, G., Weigel, D., and Laubinger, S. 2009. Stress-induced changes in the Arabidopsis thaliana transcriptome analyzed using whole-genome tiling arrays. *Plant J*.