Functional transcriptomics of a migrating cell in *Caenorhabditis elegans*

Erich M. Schwarz[a,b], Mihoko Kato[a,b], and Paul W. Sternberg[a,1]

[a]Division of Biology and Howard Hughes Medical Institute, California Institute of Technology, Pasadena, CA, 91125, U.S.A.
[b]These authors contributed equally to this work.

[1]To whom correspondence should be addressed. E-mail: *pws@caltech.edu*.

Biological Sciences: Genetics

Supporting Information

SI Materials and Methods

Culture and Strains

C. elegans were grown at 20° C. by standard methods (1). Wild-type N2 mixed-stage hermaphrodites (primarily larvae, with a few eggs and adults) were used for control RNA extractions. The strain PS4730 (*syIs128[unc-119(+)(100 μg/ml) + lag-2::YFP(50 μg/ml)]*II; *unc-119(ed4)*III; *him-5(e1490)*V) was used for LC dissections and for RNAi assays of LC migration. The following strains were used for assaying gene activity in LCs: CG308, *pha-1(e2123); him-5 (e1490) rgEx132 [pgar-3A::yfp]*; CG398, *pha-1(e2123); him-5 (e1490) rgEx132 [pacr-15:yfp pBX1]*; HC183, *qtIs20[nob-1::YFP, unc-119] unc-119(e2498)*; KP987, *lin-15(n765) nuIs1[pJM24(lin-15+), pV6 (glr-1::gfp), C06E1XP (glr-1+)]*; PS5149, *syIs78[ajm-1::GFP + unc-119(+)]; ayIs7[hlh-8::GFP + dpy-20(+)]; him-5(e1490)*; PS5309, *unc-119(ed4); syEx842[ins-18::NLS::YFP-unc-119(+)]*; PS5326, *unc-119(ed4); syEx855[ins-18::NLS::YFP-unc-119(+)]*; and IK934, *njEx386[full length ttx-4::GFP ges-1p::NLS-GFP],* obtained from Ikue Mori (Nagoya).

RNAi, Phenotypic Scoring, and Transgenes

To avoid embryonic lethality (e.g., in worms from which *nhr-67*[RNAi] L4-stage LCs were being harvested), feeding RNAi was carried out on *lag-2*::YFP worms (expressing YFP in the LC) from hatching as L1 larvae to the mid-L4 stage near the end of LC migration (2). Worms were scored for LC defects by Nomarski and fluorescence microscopy (100x); Nomarski microscopy allowed more detailed and sensitive screening than the dissection microscopy previously used by Kato and Sternberg (2). It is possible that RNAi against a single *msp* gene (*msp-3*) inactivated multiple *msp* genes due to the high nucleotide similarity of genes in the *msp* family, but we know of no reason to predict such multigene effects for other RNAi targets. YFP reporters were fused to promoter regions up to 6 kb in size, microinjected, and assayed transgenically. In all other respects, RNAi feeding and YFP fusions were done as in Kato and Sternberg (2).

LC Dissection

We harvested individual YFP-labeled LCs for RNA-seq by cutting them free of the gonad with a laser microbeam, cutting open the abdomen, and pipetting individual LCs (Fig. 1B). LCs were dissected from L3 and L4 male larvae in two stages. First, we mounted worms for laser microsurgery (3), and used the laser microbeam to sever their gonads from the LCs, since unsevered LCs proved essentially impossible to dissect. We then remounted the worms for abdominal dissection on a patch-clamp rig (3), but used a wide-mouthed, unpolished patch-clamp tube as a pipette to suck up individual LCs and transfer them to a prelubricated PCR tube. Tubes containing individual LCs were snap-frozen with liquid nitrogen and kept at -70° C. until amplification.

RT-PCR

For amplification, cells were thawed, lysed gently, and subjected to 3'-tailed RT-PCR by the method of Dulac and Axel (4). RT-PCRs were tested for basic quality by gel electrophoresis and secondary PCR against adjacent exons of *rpl-17*, cleaned and concentrated in TE with Qiagen MinElute PCR purification kits, quantitated by Nanodrop UV spectrophotometry, and stored at -20° C. We also harvested, diluted, and amplified aliquots of control RNA from mass starved wild-type hermaphroditess, primarily larvae, which contained a diverse set of non-LC cell types (5). Aliquots of N2 RNA were serially diluted in RNAse-free TE and independently amplified by RT-PCR. Equal aliquots of five parallel RT-PCRs for each cell type or for N2 larvae were pooled, and pools were sequenced as single-end 37-nt reads with an Illumina GA II or 50-nt reads with a HiSeq 2000. LC sets had five aliquots apiece. For N2 larvae, two sets of five aliquots were sequenced independently, their reads being merged after sequencing to provide a single N2 larval data set. As spike controls (Table S5), we used poly(A)$^+$-tailed versions of four spiked RNAs previously used by Mortazavi et al. (6): Apetala 2 (also termed AP2), λ 9786 (λ232), λ 11300 (λ11), and VATG (VATG3).

RNA-seq Analysis

RNA-seq reads were mapped to the WS190 sequence of the *C. elegans* genome with bowtie (7) given the arguments "*-k 11 -v 2 --best -m 10 --strata*"; RPKM counts for WS190 gene models were computed with ERANGE 3.1 (6), both for all reads (including those mapping to multiple genome sites) and for all non-multireads (which mapped either to unique genome sites or to splice junctions). For most analyses we used the simple RPKM values taken from all reads; but for the highly similar multicopy *msp* gene family, we also examined non-multireads to ensure unambiguous LC-enriched expression of individual genes such as *msp-3* (Table S4). We set an upper limit on gene models of 1.5 kb, since by gel electrophoresis we saw little evidence of 3'-RT-PCRs extending beyond this upper bound. Reads which failed to map to the WS190 *C. elegans* genome were successively mapped with bowtie to human cDNA and the AL1 primer of Dulac and Axel (4), using the same parameters for bowtie as with WS190. Human cDNA sequences were downloaded from ENSEMBL (8), at *ftp://ftp.ensembl.org/pub/release-63/fasta/homo_sapiens/cdna/Homo_sapiens.GRCh37.63.cdna.all.fa.gz*.

Genomic DNA and RNA

Genomic DNA and whole RNA was extracted from wild-type N2 *C. elegans* by previously described methods (9). Genomic DNA was used as a positive control template, distinguishable by product size from cDNA, for secondary PCRs testing single-cell RT-PCRs for success or

failure (in this study, this required oligonucleotides RPL17TEST01 and RPL17TEST012; see Detailed Protocol below).

The whole animal RNA used here was obtained and characterized in a previous study (9): it was quantitated and tested for quality with a Bioanalyzer, and successfully used for full-length RNA-seq of *C. elegans*. This RNA came from starved mixed-stage hermaphrodites; although we did not quantitate the source population, it appeared to consist mostly of L3 and L4 stage larvae, with some young adults and L2-stage larvae, and a few L1-stage larvae and eggs. We could therefore, in this study, reliably use diluted aliquots of the RNA as a background control for RT-PCRs and RNA-seq. We serially diluted aliquots of the whole RNA from ~1.4 μg/μl to ~$1.4 \cdot 10^{-3}$ μg/μl in RNAse-free T_{10} (10 mM Tris pH 8.0). 1.0-μl aliquots of the 10^3-diluted RNA were then (like dissected LCs) subjected to RT-PCR, purified, quantitated, and pooled for RNA-seq to provide larval expression data.

Gene Nomenclature

Gene names and many gene annotations were extracted from the WS220 release of WormBase (10) with TableMaker/xace from ACeDB 4.9.39. In a few cases, where a gene had an RNAi phenotype in the LC but only had a sequence (cosmid.number) name in WS220 (Tables 1, S4), we invented new names for them; these names were submitted to J. Hodgkin (Oxford) via WormBase and officially authorized before publication.

Gene Annotations

Gene annotations downloaded from WS220 included protein domains from PFAM (11) and NCBI orthology annotations (KOGs, TWOGs, etc.; (12)), as well as protein features such as signal, transmembrane, coiled-coil or low-complexity sequences (predicted respectively by the programs SignalP (13), TMHMM (14), Ncoils (15), and SEG (16). Annotations also included RNAi or mutant phenotypes annotated for a gene in WS220, with most phenotypes coming from genome-wide RNAi screens (17).

Other gene annotations were obtained from specialized databases complementary to WS220 or derived from it. GO annotations for genes (18) were downloaded from *ftp://ftp.sanger.ac.uk/pub2/wormbase/releases/WS220/ONTOLOGY/gene_association.WS220.wb.ce*. Updated orthology groups (KOGs, NOGs, etc.) from eggNOG v. 2.0 (19) were downloaded from *http://eggnog.embl.de/download/orthgroups.mapping.txt.gz*. Lists of transcription factor genes were obtained from J. Thomas (pers. comm.), the Walhout laboratory (20), or the Gupta laboratory (21). Strict orthologies between human and *C. elegans* genes were determined with OrthoMCL 1.3 (22) for the protein-coding genes of human (GRCh37 set, via ENSEMBL release 56; (8)), mouse (NCBIM37 set, via ENSEMBL 56), *C. elegans* (WormBase WS210), and *C. briggsae* (WS210). A subset of human genes known to cause disease when mutated were generated from the "morbidity" index of OMIM (23), via Ensmart of ENSEMBL release 60, and mapped onto OrthoMCL groups. A set of 931 genes preferentially expressed in male sperm, hermaphrodite sperm, or both sperm types was generated from the microarray analyses of Reinke et al. (24); a smaller set of 22 genes with well-characterized roles in spermatogenesis were extracted from recent literature (25-27).

Further data for specific genes pertinent to the LC transcriptome were derived from WormBase or GenBank. Of 46 paralogous *twk* genes in the *C. elegans* genome (28), the subset expressed in LCs were identified by their gene names, KOG domains, and PFAM domains in Table S4. Conserved genes of unknown function (29) expressed in the LC were identified

through their NCBI, eggNOG, and PFAM annotations. Cytoplasmic isoforms of MOSPD1 in mammals were identified in accessions AL546517 and CX166508 of GenBank by TBlastN searches of dbEST (30).

k-partitioning RNA-seq Data

Expression data were log-transformed and k-partitioned with Cluster 3.0 (31), with Euclidian distances for gene clustering, no clustering of expression types, no normalization of gene or cell-type data before k-partitioning, and 10,000 repetitions (32). Statistically optimal k-partitions for transformed data sets were determined with the Gap statistic (33) as implemented in the Perl module Statistics::Gap, using vector space, direct k-way clustering, the H2 criterion function, and 10,000 replicates per k-value: this gave an optimal k of 3 for log-transformed data and 2 for log-normalized data, which were combined to give 6 partitions overall (Fig. S1). For log-transformed data, this successfully resolves housekeeping genes, but fails to detect and group obvious sets of genes specific to individual cell types. We therefore examined k-partitionings of k=3 through k=10, and settled on k=5 for clustering log-transformed data to obtain the division shown in Fig. 2A. Transformed and k-partitioned gene data were plotted with TreeView 1.1.5r2 (34).

GO Term Analysis

The specificity of a given gene for a given cell type can be usefully considered a scalar variable, rather than a Boolean classification (35). For our GO term analyses below, we used scalar measurements of LC-enrichment. However, for our initial decisions about which genes to study with RNAi, we instead used a working definition of LC-enrichment: a gene had to have an absolute expression level of at least 0.5 RPKM, and its maximum observed expression in a wild-type LC (either L3 or L4 stage) had to be ≥20x greater than its observed expression in whole N2 larvae; this gave us a set of 963 genes to choose from for RNAi and YFP analyses (Tables S3B, S4). We later added a small number of genes because their expression profile was striking (MSPs), they were homologs of an interesting RNAi positive (SMC subunits), or because GO term analysis indicated they were particularly important (potassium channels).

Statistically overrepresented GO terms associated with LC-enriched genes were identified with FUNC (36). Gene Ontology (GO) terms statistically overrepresented among genes with high LC-enrichment or among genes strongly upregulated in the L4 stage, were determined using the Wilcoxon-rank test in FUNC (func_wilcoxon), with genes classified by floating-point scores (36). GO terms overrepresented among genes belonging to k-partitions were determined using the hypergeometric test in FUNC (func_hyper), with genes classified by Boolean categories (36). Both tests were run with the GO term tables from *http://archive.geneontology.org/full/2010-11-01/go_201011-termdb-tables.tar.gz*, and with 10,000 repetitions. For either floating-point and Boolean scores, we considered only the 8,011 genes for which we detected expression in pooled wild-type LCs (at either the L3 or the L4 stage). A gene's floating point score was generated from the logarithm ($\log_{10}$) of the ratios of that gene's expression levels in one cell type versus another: for instance, $\log_{10}$(maximum wild-type LC RPKM/N2 larval RPKM); $\log_{10}$(wild-type L4 RPKM/L3 RPKM); $\log_{10}$(wild-type L4 RPKM/*nhr-67*[RNAi] L4 RPKM); etc. We used $\log_{10}$ of gene expression ratios, rather than the original ratios, to avoid overweighting GO terms with outlier genes whose expression was exceptionally high. To allow ratios and logarithms to be computed for any gene, even if it had no observed expression in a given cell type, we defined the minimum expression level of each gene

as 0.01 rather than 0.00 RPKM, and the minimum expression ratio of each gene as 10^{-3} rather than 0. A gene whose observed maximum wild-type LC expression was 1.00 RPKM, and observed N2 larval expression was 0.00 RPKM, would be considered to have a maximum LC/N2 ratio of 100 (1.00/0.01), and a $\log_{10}$(maximum LC/N2) score of 2.0.

All Wilcoxon-rank and hypergeometric test results were filtered with FUNC's statistical refinement script before studying them further. GO terms were considered significant only if FUNC's statistical refinement script passed them at a p-value threshold of 0.01 or less, and if the ontology they belonged to (molecular_function [GO:0003674], biological_process [GO:0008150], or cellular_component [GO:0005575]) had a global statistical significance of 0.05 or lower. Where refinement at p=0.01 gave many terms, we sometimes produced a smaller subset by refinement at p=0.001.

While statistics defined which GO terms might be significant, making biological sense of the results and summarizing them for human readers required some judgement. Many terms that passed FUNC's refined statistical tests were unremarkable. Others were probably nontrivial, but were not easy to decipher. For instance, "locomotion" was observed among overrepresented GO terms. This is likely to reflect the uncoordinated phenotypes of genes elicited in genome-wide mass RNAi screens, which are used to infer by their phenotype that a given gene is involved in locomotion. That linker cells express many such genes was notable, but to make real sense of this was not easy, since the genes could be specialized for muscle function, neuronal function, or something else entirely. For this reason, we have focused in our summary of the GO term analysis on genes with a well-defined, narrow function that can be ascribed to a single migrating cell. We have also passed over, without comment, terms that are statistically significant but very broad.

Cross-comparison to known sperm genes and testing MSP/MSD gene expression for reproducibility

Reinke et al. identified a set of 970 genes with preferential activity in male sperm, hermaphrodite sperm, or both sperm types (24). We linked 932 of these genes to current gene names in WormBase WS220 (others may have been designated as pseudogenes, or have undergone splits or fusions of their gene models since 2004). Comparing these 932 genes known to be active in sperm to the 10,064 genes that we detected in wild-type LCs, we found 569 LC genes with Reinke annotations (male_sperm, herm_sperm, or shared_sperm; Table S4). Among the smaller set of 8,011 genes expressed in pooled wild-type LCs, we found 389 such sperm genes.

In pooled LCs, we observed strong upregulation for MSP and MSD genes from the wild-type L3 to the wild-type L4 stage, but much weaker upregulation to *nhr-67*(RNAi) L4 (Fig. 2C). For MSP genes, mean expression levels were 0.04 and 9.76 RPKM in pooled L3-stage and *nhr-67*(RNAi) L4-stage LCs (with ranges of 0.00-0.80 and 0.00-215 RPKM, respectively), but rose to 148 RPKM in pooled wild-type L4-stage LCs (range 1.01-1,372 RPKM). For MSD genes, mean expression values were 2.14 and 2.84 RPKM in L3-stage and *nhr-67*(RNAi) L4-stage LCs (with ranges of 0.00-36.5 and 0.00-34.7 RPKM), but rose to 54.9 RPKM in wild-type L4-stage LCs (range 0.00-775 RPKM).

To statistically evaluate these data, we scored the 8,011 genes and their subset of 389 sperm-associated genes by their ratio of gene expression in wild-type pooled L4 versus L3-stage LCs, and performed a Wilcoxon test of their ranking with the Perl module Statistics::Test::WilcoxonRankSum (version 0.0.6). We found that the 389 sperm genes were

overrepresented among genes upregulated between L3- and L4-stage LCs with high significance ($p < 10^{-6}$; Table S13). This significance remained if we considered only 345 genes which did not encode a PFAM domain for major sperm protein (PF00635) in WormBase WS220 ($p < 10^{-6}$), or even smaller gene sets annotated by Reinke et al. as enriched in male or hermaphrodite sperm only (24 and 37 genes, respectively; $p = 1.2 \cdot 10^{-5}$ and $7.3 \cdot 10^{-5}$).

To determine whether upregulation of MSP/MSD genes was reproducible or due to sporadic RNA-seq signals (e.g., from a freak contamination event during dissection), we also examined the expression levels of ~60 LC genes encoding MSP domains in RNA-seq data from individual, rather than pooled, LCs. Expression of these ~60 genes was reproducibly strong in five individual wild-type L4-stage LCs (median = 0.50 RPKM), while being substantially lower among five individual *nhr-67*(RNAi) L4-stage LCs (median = 0.00 RPKM); the median expression levels of *msp-3* in these two sets of individual LCs were 84.1 and 1.14 RPKM, respectively.

Statistical analysis of differential gene expression

We computed statistical significance scores for differential gene expression with DESeq 2.10 (37). As input data, we used counts of reads mapped to genes from the RNA-seq data individually amplified and sequenced LCs, and from two RNA-seq analyses of whole larval RNA (Table S2) which we had previously pooled as a single larval data set. This gave us five biological replicates of each LC type and two technical replicates of whole larvae. We compared L3- to L4-stage LCs, wild-type to *nhr-67*(RNAi) L4-stage LCs, and wild-type LCs (both L3 and L4) to whole larvae. For the LC/larval comparison, we grouped all ten L3- and L4-stage LCs into a single LC condition. For each comparison, we first estimated dispersions of gene activity with *estimateDispersions(cds1, method = "per-condition", sharingMode = "maximum", fitType = "local")*, and then ran a negative binomial test for differential gene expression with *nbinomTest(cds1, "[condition 1]", "[condition 2]")*. DESeq generates both raw p-values (uncorrected for multiple hypothesis testing with thousands of genes) and false discovery rates (p-values corrected for multiple hypothesis testing); both statistical values for each differential gene analysis are given in Table S4. At least two LC-expressed genes that we showed experimentally to be required for normal LC migration (*hlh-19* and *srsx-18*) scored as insignificant by these criteria, probably because each gene was detected in only one out of 10 wild-type LCs, and thus resembled statistical noise. Nevertheless, DESeq did indicate which genes had the most robust signal of differential gene expression, given the limits of single-cell RNA-seq.

Supporting Figures

Figure S1. k-partitioning of LC Genes With Six Partitions Defined by the Gap Statistic (Methods; (33)). As in Fig. 2A, expression data from 8,011 genes expressed in pooled wild-type LCs ("LC genes") were compared between LC types (L3, L4, *nhr-67*[RNAi] L4), and mass larvae (Table S1, S3). Yellow and blue denote high and low expression levels, respectively. LC genes were split with a statistically optimal k=3 for log-transformed data; in parallel, they were split with an optimal k=2 for log-normalized data; finally, the k=2 split was superimposed on the k=3 split of the log-transformed data, to give the six partitions shown here. Initial k=3-derived and subsequent k=2-derived splits are shown with black and grey dividing lines, respectively. This method is statistically less arbitrary than the simpler k=5 partitioning used in Fig. 2A, but gives groupings of genes that are less intuitively clear.

Figure S2. LC Genes Also Expressed in Sperm are More Strongly Upregulated in Wild-type L4-Stage LCs Than in *nhr-67*(RNAi) L4-Stage LCs. RNA-seq data from 8,011 genes expressed in pooled wild-type LCs are shown, with the ratios of observed gene activity for wild-type L4-stage/wild-type L3-stage LCs (x-axis) versus *nhr-67*(RNAi) L4-stage/wild-type L3-stage LCs (y-axis). The following subsets of genes are highlighted: *msp-3*; 43 other "sperm/MSP" genes, which were previously found to be expressed in sperm (24) and which also encode proteins with MSP domains (PFAM domain PF00635; Table S4); 345 "sperm/non-MSP" genes, which are sperm-expressed but which do not encode MSP domains; 15 "MSP/non-sperm" genes, which encode MSP domains but which were not previously observed to be sperm-expressed; and the overall background of LC genes. Most LC genes with previously known sperm expression, regardless of whether they encode MSP domains or not, are strongly upregulated from the L3 to the wild-type L4 stages, but show much less upregulation from the L3 to the *nhr-67*(RNAi) L4 stages. This skew is statistically discernable as well (Table S13).

Figure S3. Activity of *msp-3* Transgene (A,B) and RNAi (C,D) in Linker Cell. (A) Nomarski image and **(B)** fluorescence image of a migrating L4-stage LC (arrow), followed by the gonad (row of cells to the left). A YFP transgene driven by 3.3-kb of the *msp-3* promoter region (Table S10) expresses in the LC (arrow), but not in the gonad trailing behind it. Cells between the linker cell and the body wall are ventral cord neurons or other cells in that plane. **(C)** A wild-type mid-L4-stage LC expressing cytoplasmic YFP is typically elongated and pointed towards its direction of migration (overlay of Nomarski and fluorescence images). **(D)** In a similarly staged LC subjected to *msp-3*(RNAi), the LC's migration is delayed, and its shape is abnormally rounded.

Supporting Tables

Table S1. Expressed Genes Detected in Pooled LCs and Larvae.

Condition:	Reads sequenced:	Mapped to *C. elegans*:	Mapped to human cDNA:	Mapped to AL1:	Unmapped reads:	Genes:
In *C. elegans*	n/a	n/a	n/a	n/a	n/a	20,252
L3-stage LC pool	14,291,807	5,496,633 (38.5%)	906,562 (6.3%)	4,874,927 (34.1%)	3,007,274 (21.1%)	5,740 [28%]
L4-stage LC pool	10,303,633	4,765,657 (46.3%)	1,485,637 (14.4%)	1,484,930 (14.4%)	2,564,303 (24.9%)	6,603 [33%]
nhr-67(RNAi) L4-stage LC pool	13,571,986	4,276,676 (31.5%)	2,023,122 (14.9%)	2,197,942 (16.2%)	5,072,207 (37.4%)	5,290 [26%]
L3- and L4-stage LC pool	24,595,440	10,262,290 (41.7%)	2,392,199 (9.7%)	6,359,857 (25.9%)	5,571,577 (22.7%)	8,011 [40%]
Mixed larvae	26,268,281	20,040,697 (76.3%)	14,243 (0.05%)	4,698,805 (17.9%)	1,509,867 (5.7%)	13,146 [65%]
Any pooled source	64,435,707	34,579,663 (53.7%)	4,429,654 (6.9%)	13,256,604 (20.6%)	12,153,651 (18.8%)	13,939 [69%]

Reads of 38 nt were mapped with bowtie to the WS190 release of the *C. elegans* genome, the human cDNA sequences from ENSEMBL 63, or the PCR primer AL1 (Methods). For each cell type or combination of cell types, percentages of reads are against the total reads sequenced. Manual dissection of LCs clearly introduced human poly(A)$^+$ RNA to our RT-PCRs, since RNA-seq of undissected whole larvae gave much lower frequencies of reads mapping to human cDNA. However, the predominant contaminant was the PCR primer AL1, found in all four read sets. We do not know the source of reads which we failed to map to human cDNA or AL1, but they might arise from other contaminants in the *C. elegans* cultures, such as the bacteria used to grow worms. Genes are from the WS190 annotation of *C. elegans*; there were 20,252 in the entire genome, of which 69% showed detectable expression in at least one of our RNA-seq data sets. Of the genes for which we detected expression at all, 94% were detected in whole larvae, which was reasonable given the wide variety of larval cell types (5).

Table S2. Expressed Genes Detected in Individual LCs and Larval RNA Samples.

Condition:	Reads sequenced:	Mapped to *C. elegans*:	Genes:
In *C. elegans*	n/a	n/a	20,252
L3.indiv1	15,113,212	2,866,508 (19.0%)	2,983 (14.7%)
L3.indiv2	1,348,957	626,707 (46.5%)	4,133 (20.4%)
L3.indiv3	20,094,930	10,339,482 (51.5%)	4,238 (20.9%)
L3.indiv4	15,355,272	7,102,435 (46.3%)	3,433 (17.0%)
L3.indiv5	18,612,313	5,618,173 (30.2%)	2,868 (14.2%)
L4.indiv1	19,692,491	11,485,407 (58.3%)	7,267 (35.9%)
L4.indiv2	21,182,740	12,441,710 (58.7%)	6,384 (31.5%)
L4.indiv3	19,729,469	424,071 (2.2%)	3,278 (16.2%)
L4.indiv4	1,232,463	420,408 (34.1%)	4,436 (21.9%)
L4.indiv5	19,979,116	8,146,424 (40.8%)	4,053 (20.0%)
nhr.indiv1	13,328,842	297,067 (2.2%)	2,744 (13.6%)
nhr.indiv2	15,429,375	7,763,475 (50.3%)	4,855 (24.0%)
nhr.indiv3	11,453,227	6,231,720 (54.4%)	4,046 (20.0%)
nhr.indiv4	4,599,820	767,189 (16.7%)	2,871 (14.2%)
nhr.indiv5	17,915,437	5,819,769 (32.5%)	3,760 (18.6%)
Mixed larvae 1	16,155,370	11,250,288 (69.6%)	12,856 (63.5%)
Mixed larvae 2	10,112,911	8,790,409 (86.9%)	9,485 (46.8%)

Reads, mapping to the *C. elegans* genome, and gene counts were computed as in Table S1. The original reads for individual LCs were 50 nt long, but were truncated in silico to 38 nt before mapping to allow consistency between analyses. Data from the two larval samples were merged for most analyses of pooled larval RNA-seq, but were analyzed separately here. Note that gene expression levels for each cell were normalized by being computed as RPKM (6), so that comparisons of gene expression levels between different individual cells or their mean values (Table S6) were feasible.

Table S3. Subsets of Genes With Expression Detected in Wild-type LCs.

S3A. Subsets of genes with expression detected in pooled versus individual LCs.

Condition:	**Genes:**
Any wild-type LCs, pooled or individual, L3- or L4-stage	10,064 [100%]
Pooled L3- or L4-stage pooled LCs	8,011 [79.6%]
Individual L3- or L4-stage LCs (in aggregate)	9,968 [99.0%]
Both pooled and individual L3- or L4-stage LCs (detected both ways)	7,915 [78.6%]
Only pooled L3- or L4-stage LCs, not individual ones	96 [1.0%]
Only individual L3- or L4-stage LCs (in aggregate), not pooled ones	2,053 [20.4%]

S3B. Subsets of genes with expression detected in pooled LCs.

Condition:	**Genes:**
Wild-type L3- or L4-stage LCs	8,011 [100%]
≥20x greater RPKM in wild-type LCs (defining larger RPKM from either L3- or L4-stage LCs) than in whole N2 larvae	1,097 [13.7%]
≥20x greater RPKM in LCs than in whole larvae, along with RPKM ≥ 0.5	963 [12.0%]
≥20-fold differences of RPKM (either greater or lower) between L3- and L4-stage LCs	3,963 [49.5%]
≥20-fold greater RPKM in wild-type L4-stage LCs than in *nhr-67*(RNAi) L4-stage LCs	2,478 [30.9%]
≥20-fold greater RPKM in wild-type than in *nhr-67*(RNAi) L4-stage LCs, and also ≥20-fold greater RPKM in L4- than in L3-stage LCs	1,473 [18.4%]
≥20-fold lower RPKM in wild-type L4-stage LC than in *nhr-67*(RNAi) L4-stage LCs	903 [11.3%]
≥20-fold lower RPKM in wild-type versus *nhr-67*(RNAi) L4-stage LCs, and also ≥20-fold lower RPKM in L4-stage than in L3-stage LCs	504 [6.3%]
Parallel ≥20-fold changes of RPKM for L4-stage LCs versus either L3-stage or *nhr-67*(RNAi) L4-stage LCs (1,473 + 504 genes)	1,977 [24.7%]
Transcription factor genes expressed in L3- or L4-stage LCs	426 [5.3%]
Transcription factor genes expressed in L3- or L4-stage LCs, with ≥20-fold differences of RPKM between wild-type versus *nhr-67*(RNAi) L4	180 [2.2%]

S3C. Subsets of genes with expression detected in individual LCs, with comparisons between types of LCs being done between the means of individuals from each type.

Condition:	**Genes:**
Wild-type L3- or L4-stage LCs	9,968 [100%]
≥20x greater RPKM in wild-type LCs (defining larger RPKM from either L3- or L4-stage LCs) than in whole N2 larvae	1,840 [18.5%]
≥20-fold differences of RPKM (either greater or lower) between L3- and L4-stage LCs	3,984 [40.0%]

≥20-fold greater RPKM in wild-type L4-stage LCs than in *nhr-67*(RNAi) L4-stage LCs	2,949 [30.0%]
≥20-fold greater RPKM in L4-stage LCs than in L3-stage LCs	3,003 [30.1%]
≥20-fold greater RPKM in wild-type than in *nhr-67*(RNAi) L4-stage LCs, and also ≥20-fold greater RPKM in L4- than in L3-stage LCs	1,929 [19.4%]
≥20-fold lower RPKM in wild-type L4-stage LC than in *nhr-67*(RNAi) L4-stage LCs	736 [7.4%]
≥20-fold lower RPKM in L4-stage LCs than in L3-stage LCs	981 [9.8%]
≥20-fold lower RPKM in wild-type versus *nhr-67*(RNAi) L4-stage LCs, and also ≥20-fold lower RPKM in L4-stage than in L3-stage LCs	277 [2.8%]
Parallel ≥20-fold changes of RPKM for L4-stage LCs versus either L3-stage or *nhr-67*(RNAi) L4-stage LCs (1,929 + 277 genes)	2,206 [22.1%]
Transcription factor genes expressed in L3- or L4-stage LCs	538 [5.4%]
Transcription factor genes expressed in L3- or L4-stage LCs, with ≥20-fold differences of RPKM between wild-type versus *nhr-67*(RNAi) L4	207 [2.1%]

Table S4. Names, Expression Levels, and Annotations for the 10,064 Genes Whose Expression Was Detected in Wild-type L3- or L4-stage LCs. [Data are in the Excel file "Dataset S1".] These include both the 8,011 genes detected in pooled L3 and L4 LCs, and an additional 1,957 genes which were detected only in RNA-seq data from individually amplified L3 and L4 LCs. (Conversely, only 96 genes detected in pooled LCs were not also detected in individual LCs.) Columns denote the following data: "Gene" gives the full gene identifier (WormBase name, sequence name, and CGC name) of a gene. "L3_pool", "L4_pool", "nhr_pool", and "larvae_pool" denote the RPKM for a given gene observed in pooled L3 LCs, L4-stage LCs, *nhr-67*(RNAi) L4-stage LCs, and whole wild-type (N2) larvae. "L3_pool_nm" through "larvae_pool_nm" denote the same data, but with RPKMs calculated only from non-multireads (unique or spliced reads, which are important for ascertaining locus-specific expression of genes like *msp-3* that belong to highly duplicated gene families). "max_wtLC_pool/larvae_pool" gives the ratio of the highest observed expression in pooled wild-type L3-stage or L4-stage LCs to expression in whole larvae, with a minimum RPKM of 0.01 when no expression was observed; "max_wtLC_pool_nm/larvae_pool_nm" denotes the same ratio for non-multiread RPKMs; and "max_wtLC_indiv_mean/larvae_pool" denotes the same ratio, but with maximum LC expression taken from the mean RPKMs per gene of individually amplified L3 or L4 LCs. "L3.indiv_Mean", "L4.indiv_Mean", and "nhr.indiv_Mean" give the mean RPKMs per gene for each set of RNA-seq data from five individual L3, L4, or *nhr-67*(RNAi) LCs. "L3.vs.L4 p-value" and "L3.vs.L4 FDR" give the p-value (uncorrected for multiple hypothesis testing) and the false discovery rate (corrected for such testing) for differential expression of genes, as determined by DESeq (37) with RNA-seq data from individually amplified wild-type L3- versus L4-stage LCs (Table S2). "L4.vs.nhr p-value", "L4.vs.nhr FDR", "LC.vs.larvae p-value", and "LC.vs.larvae FDR" give the same values, for comparisons of individual wild-type versus *nhr-67*(RNAi) L4-stage LCs and of wild-type L3- and L4-stage LCs versus two technical replicates of whole larval RNA. "Protein size(s)" lists the sizes of protein products. "Protein feature" lists predicted features such as signal, transmembrane, coiled-coil or low-complexity sequences, predicted respectively by the programs SignalP (13), TMHMM (14), Ncoils (15), and SEG (16). "Reinke_class" lists 569 genes detected in wild-type LCs (from a set of 931 such genes in the *C. elegans* genome) that are preferentially expressed in male sperm, hermaphrodite sperm, or both sperm types, as determined by the microarray analyses of Reinke et al. (24). "Known_sperm_genes" indicates 22 genes with well-characterized roles in *C. elegans* spermatogenesis (25-27). "TF" indicates whether a gene's product was predicted to be a transcription factor by J. Thomas (pers. comm.), the Walhout laboratory (20), or the Gupta laboratory (21). "Disease ortholog" denotes orthology to a human disease gene. "Strict human ortholog" denotes strict orthology to a human gene. "NCBI KOG" and "Bork KOG" list orthology annotations by NCBI (12) and eggNOG 2.0 (19), respectively. "PFAM domain" lists any such domains annotated in WormBase WS220. "RNAi_screen" lists whether a gene was screened by RNAi, with "nonwt_RNAi" and "WT_RNAi" indicating that a phenotype either was or was not observed. "WBPhenotype" lists any RNAi or mutant phenotypes annotated for a gene in WS220, with most phenotypes coming from mass RNAi screens. "NOT WBPheno" indicates that a gene was annotated as negative for such phenotypes in WS220. "k-cluster group" indicates which of the five k-cluster groups shown in Fig. 2A a gene belonged to. "LC-enriched GO term" lists any genes annotated for a GO term mentioned in the main text as being both overrepresented in LC-enriched genes (the full term list is in Table S11) and being particularly noteworthy (e.g., "axon"). "L4-enriched GO term" does the same for

GO terms listed in Table S12C as being associated with genes upregulated in L4 LCs by NHR-67. "L3.indiv1" through "nhr.indiv5" give expression data in RPKM for the individually amplified LCs themselves; "L3.indiv1_nm" through "nhr.indiv5_nm" give these values for non-multireads only. Detailed annotations for LC genes in Table S4 are derived from the WS220 release of WormBase, which contains references for their phenotypic data.

Table S5. RPKM Counts for Spikes in RT-PCRs.

Spike type	RPKM of a given pooled RT-PCR type				Estimated. spiked txs. per RT-PCR:
	L3-stage LC	L4-stage LC	*nhr-67*(RNAi) L4-stage LC	Whole larvae	
VATG	10,834.44	937.56	3,648.17	2.73	5,000
λ 11300	457.16	73.04	338.72	0.82	500
λ 9786	5.63	0.40	6.97	0.00	50
Apetala 2	0.00	0.00	0.00	0.00	5

Table S6. Correlation of Measured Gene Expression Levels Between Cells and Pools.

S6A. The correlations (r^2) of gene expression levels between individual, mean individual, or pooled LCs versus mean individual or pooled LCs or whole larvae. These comparisons used all 10,064 genes observed in either pooled or individual wild-type L3- or L4-stage LCs. Since the pools were made from aliquots of the RT-PCR products from single cells, this comparison shows technical rather than biological replicability: RNA-seq with a pool of single RT-PCRs gives a result quite similar to the data obtained by averaging RNA-seq data from RT-PCRs sequenced individually. It also shows that the agreement between different cells varied, with one notably divergent L4 cell (L4.indiv3) in our data set. "nhr" denotes "*nhr-67*(RNAi) L4-stage"; "im" denotes "individual mean"; "pl" denotes "pool".

	L3 pl	L3.im	L4 pl	L4.im	nhr pl	nhr.im	Larval pl
L3 pool	1.00	0.93	0.81	0.66	0.72	0.63	0.27
L3.indiv mean	0.93	1.00	0.67	0.74	0.67	0.72	0.20
L3.indiv1	0.53	0.65	0.41	0.39	0.32	0.37	0.11
L3.indiv2	0.78	0.80	0.57	0.66	0.53	0.55	0.17
L3.indiv3	0.82	0.78	0.59	0.59	0.71	0.70	0.25
L3.indiv4	0.77	0.79	0.58	0.66	0.68	0.67	0.18
L3.indiv5	0.59	0.72	0.39	0.50	0.35	0.42	0.06
L4 pool	0.81	0.67	1.00	0.56	0.68	0.48	0.35
L4.indiv mean	0.66	0.74	0.56	1.00	0.56	0.67	0.15
L4.indiv1	0.71	0.77	0.66	0.74	0.55	0.58	0.14
L4.indiv2	0.70	0.75	0.59	0.79	0.58	0.65	0.12
L4.indiv3	0.16	0.18	0.10	0.51	0.11	0.15	0.02
L4.indiv4	0.46	0.54	0.46	0.66	0.37	0.47	0.09
L4.indiv5	0.43	0.46	0.37	0.55	0.46	0.57	0.18
nhr pool	0.72	0.67	0.68	0.56	1.00	0.78	0.33
nhr.indiv mean	0.63	0.72	0.48	0.67	0.78	1.00	0.17
nhr.indiv1	0.25	0.35	0.20	0.29	0.28	0.53	0.05
nhr.indiv2	0.73	0.76	0.61	0.71	0.72	0.75	0.17
nhr.indiv3	0.49	0.60	0.37	0.53	0.53	0.62	0.10
nhr.indiv4	0.40	0.41	0.27	0.42	0.54	0.71	0.13
nhr.indiv5	0.27	0.31	0.19	0.31	0.49	0.60	0.11

Table S6 (continued). Correlation of Measured Gene Expression Levels Between Cells and Pools.

S6B. The correlations (r^2) of gene expression levels in individual LCs versus one another. These comparisons used all 10,064 genes observed in either pooled or individual wild-type L3- or L4-stage LCs. Correlations of individual cells to one another are lower than to their mean, but higher than to whole larvae.

	L3.indiv1	L3.indiv2	L3.indiv3	L3.indiv4	L3.indiv5
L3.indiv1	1.00	0.44	0.32	0.30	0.47
L3.indiv2	0.44	1.00	0.60	0.60	0.48
L3.indiv3	0.32	0.60	1.00	0.73	0.38
L3.indiv4	0.30	0.60	0.73	1.00	0.44
L3.indiv5	0.47	0.48	0.38	0.44	1.00
L4.indiv1	0.57	0.72	0.49	0.58	0.55
L4.indiv2	0.42	0.72	0.55	0.61	0.52
L4.indiv3	0.08	0.18	0.14	0.19	0.12
L4.indiv4	0.31	0.59	0.38	0.42	0.37
L4.indiv5	0.18	0.31	0.52	0.49	0.29
nhr.indiv1	0.34	0.20	0.24	0.23	0.30
nhr.indiv2	0.41	0.67	0.63	0.69	0.49
nhr.indiv3	0.38	0.56	0.39	0.46	0.48
nhr.indiv4	0.12	0.29	0.61	0.51	0.14
nhr.indiv5	0.12	0.20	0.37	0.32	0.17
	L4.indiv1	L4.indiv2	L4.indiv3	L4.indiv4	L4.indiv5
L3.indiv1	0.57	0.42	0.08	0.31	0.18
L3.indiv2	0.72	0.72	0.18	0.59	0.31
L3.indiv3	0.49	0.55	0.14	0.38	0.52
L3.indiv4	0.58	0.61	0.19	0.42	0.49
L3.indiv5	0.55	0.52	0.12	0.37	0.29
L4.indiv1	1.00	0.80	0.17	0.68	0.30
L4.indiv2	0.80	1.00	0.25	0.58	0.34
L4.indiv3	0.17	0.25	1.00	0.14	0.08
L4.indiv4	0.68	0.58	0.14	1.00	0.31
L4.indiv5	0.30	0.34	0.08	0.31	1.00
nhr.indiv1	0.32	0.28	0.04	0.23	0.25
nhr.indiv2	0.75	0.79	0.19	0.53	0.38
nhr.indiv3	0.61	0.59	0.13	0.45	0.27
nhr.indiv4	0.25	0.37	0.09	0.23	0.50
nhr.indiv5	0.18	0.23	0.08	0.19	0.38

	nhr.indiv1	nhr.indiv2	nhr.indiv3	nhr.indiv4	nhr.indiv5
L3.indiv1	0.34	0.41	0.38	0.12	0.12
L3.indiv2	0.20	0.67	0.56	0.29	0.20
L3.indiv3	0.24	0.63	0.39	0.61	0.37
L3.indiv4	0.23	0.69	0.46	0.51	0.32
L3.indiv5	0.30	0.49	0.48	0.14	0.17
L4.indiv1	0.32	0.75	0.61	0.25	0.18
L4.indiv2	0.28	0.79	0.59	0.37	0.23
L4.indiv3	0.04	0.19	0.13	0.09	0.08
L4.indiv4	0.23	0.53	0.45	0.23	0.19
L4.indiv5	0.25	0.38	0.27	0.50	0.38
nhr.indiv1	1.00	0.29	0.39	0.16	0.21
nhr.indiv2	0.29	1.00	0.62	0.44	0.27
nhr.indiv3	0.39	0.62	1.00	0.21	0.19
nhr.indiv4	0.16	0.44	0.21	1.00	0.45
nhr.indiv5	0.21	0.27	0.19	0.45	1.00

Table S6 (continued). Correlation of Measured Gene Expression Levels Between Cells and Pools.

S6C. The correlations (r^2) of gene expression levels in individual LCs to the mean of the other four cells of their type ("Mean4"), which provides a measurement of biological rather than technical reproducibility. These comparisons used all 10,064 genes observed in either pooled or individual wild-type L3- or L4-stage LCs. Correlation of individual LCs to our background control of whole larval RNA was substantially lower.

	Mean4	Larval pool
L3.indiv1	0.48	0.11
L3.indiv2	0.71	0.17
L3.indiv3	0.65	0.25
L3.indiv4	0.67	0.18
L3.indiv5	0.57	0.06
L4.indiv1	0.62	0.14
L4.indiv2	0.68	0.12
L4.indiv3	0.19	0.02
L4.indiv4	0.55	0.09
L4.indiv5	0.30	0.18
nhr.indiv1	0.34	0.05
nhr.indiv2	0.61	0.17
nhr.indiv3	0.48	0.10
nhr.indiv4	0.45	0.13
nhr.indiv5	0.43	0.11

Table S7. The Full Set of GO Terms Overrepresented Among Each of Five k-partitioned Gene Clusters Shown in Fig. 2A.

Cluster 1: NHR-67-inhibited, broad expression (2,233 genes)			
root_node_name	**node_name**	**node_id**	**p-value**
molecular_function	protein dimerization activity	GO:0046983	0.000699563
cellular_component	mitochondrion	GO:0005739	0.000247493
cellular_component	organelle envelope	GO:0031967	7.83E-05
biological_process	ameboidal cell migration	GO:0001667	0.000667211
biological_process	receptor-mediated endocytosis	GO:0006898	9.20E-05
biological_process	multicellular organismal development	GO:0007275	0.000719621
biological_process	axon guidance	GO:0007411	0.000340993
biological_process	tissue development	GO:0009888	0.000854911
biological_process	gonad morphogenesis	GO:0035262	0.000683918
biological_process	positive regulation of growth rate	GO:0040010	0.000135525
biological_process	regulation of transport	GO:0051049	0.000667211
biological_process	establishment of localization in cell	GO:0051649	0.000187531
biological_process	cellular response to biotic stimulus	GO:0071216	0.000575511
biological_process	cellular response to protein stimulus	GO:0071445	0.000575511
Cluster 2: NHR-67-activated, broad expression (1,013 genes)			
root_node_name	**node_name**	**node_id**	**p-value**
biological_process	cellular amino acid catabolic process	GO:0009063	0.000700349
biological_process	water-soluble vitamin biosynthetic process	GO:0042364	0.000148556
Cluster 3: Ubiquitous (1,164 genes)			
root_node_name	**node_name**	**node_id**	**p-value**
molecular_function	structural constituent of ribosome	GO:0003735	3.43E-36
molecular_function	translation elongation factor activity	GO:0003746	2.02E-05
molecular_function	cytochrome-c oxidase activity	GO:0004129	6.67E-06
molecular_function	threonine-type endopeptidase activity	GO:0004298	2.39E-11
molecular_function	GTP binding	GO:0005525	0.000164665
molecular_function	sugar binding	GO:0005529	0.000130867
molecular_function	NADH dehydrogenase (ubiquinone) activity	GO:0008137	2.02E-05
molecular_function	hydrogen ion transmembrane transporter activity	GO:0015078	2.61E-06
molecular_function	rRNA binding	GO:0019843	0.000165167
molecular_function	ribosome binding	GO:0043022	0.000944749
molecular_function	hydrogen ion transporting ATP synthase activity, rotational mechanism	GO:0046933	0.000488813
molecular_function	proton-transporting ATPase activity, rotational	GO:0046961	3.38E-05

	mechanism		
molecular_function	unfolded protein binding	GO:0051082	4.90E-06
cellular_component	mitochondrion	GO:0005739	1.73E-05
cellular_component	proteasome core complex	GO:0005839	4.32E-06
cellular_component	ribosome	GO:0005840	7.06E-33
cellular_component	striated muscle thin filament	GO:0005865	0.000151494
cellular_component	small ribosomal subunit	GO:0015935	7.44E-05
cellular_component	prefoldin complex	GO:0016272	0.000151494
cellular_component	vacuolar proton-transporting V-type ATPase complex	GO:0016471	0.000320163
cellular_component	proteasome core complex, alpha-subunit complex	GO:0019773	4.32E-06
cellular_component	organelle membrane	GO:0031090	0.000723281
cellular_component	proton-transporting two-sector ATPase complex, catalytic domain	GO:0033178	1.63E-05
cellular_component	cytoplasmic part	GO:0044444	6.22E-05
cellular_component	proton-transporting ATP synthase complex, coupling factor F(o)	GO:0045263	0.000517661
biological_process	reproduction	GO:0000003	3.82E-20
biological_process	nematode larval development	GO:0002119	2.42E-37
biological_process	translation	GO:0006412	1.45E-26
biological_process	translational elongation	GO:0006414	4.56E-10
biological_process	protein folding	GO:0006457	1.59E-05
biological_process	receptor-mediated endocytosis	GO:0006898	4.28E-05
biological_process	determination of adult lifespan	GO:0008340	3.65E-15
biological_process	embryo development ending in birth or egg hatching	GO:0009792	3.23E-45
biological_process	body morphogenesis	GO:0010171	1.42E-05
biological_process	ATP synthesis coupled proton transport	GO:0015986	5.81E-10
biological_process	molting cycle, collagen and cuticulin-based cuticle	GO:0018996	1.56E-06
biological_process	electron transport chain	GO:0022900	0.000669231
biological_process	pronuclear migration	GO:0035046	2.45E-05
biological_process	growth	GO:0040007	9.91E-15
biological_process	positive regulation of growth rate	GO:0040010	1.70E-13
biological_process	locomotion	GO:0040011	1.69E-08
biological_process	regulation of multicellular organism growth	GO:0040014	0.000301983
biological_process	hermaphrodite genitalia development	GO:0040035	3.39E-07
biological_process	cellular respiration	GO:0045333	7.00E-05
biological_process	negative regulation of growth	GO:0045926	0.00081114
biological_process	protein polymerization	GO:0051258	1.27E-05
biological_process	proteolysis involved in cellular protein catabolic process	GO:0051603	0.000106085
Cluster 4: L3-specific (1,400 genes)			
root_node_name	**node_name**	**node_id**	**p-value**
molecular_function	sequence-specific DNA binding transcription factor activity	GO:0003700	0.000538886
molecular_function	monooxygenase activity	GO:0004497	0.000114898

molecular_function	heme binding	GO:0020037	0.000539324
molecular_function	sequence-specific DNA binding	GO:0043565	9.07E-05
cellular_component	nucleosome	GO:0000786	1.37E-07
biological_process	nucleosome assembly	GO:0006334	7.07E-07
biological_process	RNA biosynthetic process	GO:0032774	0.000822246
biological_process	ncRNA processing	GO:0034470	0.00057598
biological_process	kinetochore organization	GO:0051383	0.000788237
Cluster 5: NHR-67-activated, L4-specific (2,201 genes)			
root_node_name	**node_name**	**node_id**	**p-value**
molecular_function	protein serine/threonine kinase activity	GO:0004674	2.40E-10
molecular_function	protein tyrosine kinase activity	GO:0004713	4.07E-08
molecular_function	protein tyrosine phosphatase activity	GO:0004725	4.73E-06
molecular_function	cation channel activity	GO:0005261	0.000785088
molecular_function	potassium channel activity	GO:0005267	5.72E-05
molecular_function	gated channel activity	GO:0022836	9.15E-05
cellular_component	integral to membrane	GO:0016021	2.62E-24
biological_process	protein phosphorylation	GO:0006468	3.85E-11
biological_process	protein dephosphorylation	GO:0006470	4.84E-06
biological_process	G-protein coupled receptor protein signaling pathway	GO:0007186	0.000147227
biological_process	metal ion transport	GO:0030001	0.00041068
biological_process	lipid glycosylation	GO:0030259	0.000541829

Table S8. Genes Selected for RNAi and Scored for LC Migration Defects. Results are listed as either "nonwt_RNAi" or "WT_RNAi", for genes that did or did not show a phenotype.

Gene	RNAi_screen
WBGene00020142|T01C8.1|aak-2	WT_RNAi
WBGene00017967|F32A5.1|ada-2	WT_RNAi
WBGene00000074|C04A11.4|adm-2	WT_RNAi
WBGene00000205|M01B12.3|arx-7	nonwt_RNAi
WBGene00000219|F21F8.7|asp-6	WT_RNAi
WBGene00000221|T04C10.4|atf-5	WT_RNAi
WBGene00015096|B0261.8	WT_RNAi
WBGene00009945|F52H3.3|bath-38	WT_RNAi
WBGene00007237|C01G10.10	WT_RNAi
WBGene00015361|C02H6.1	WT_RNAi
WBGene00015510|C06A6.5	WT_RNAi
WBGene00007496|C09G9.7|npax-4	WT_RNAi
WBGene00015815|C16A11.2	WT_RNAi
WBGene00016046|C24B5.4	WT_RNAi
WBGene00016162|C27D6.4|crh-2	nonwt_RNAi
WBGene00007941|C34F6.7	WT_RNAi
WBGene00016439|C35D10.1	WT_RNAi
WBGene00007994|C37E2.2	WT_RNAi
WBGene00008065|C43C3.2|npr-18	WT_RNAi
WBGene00016749|C48E7.1	WT_RNAi
WBGene00008260|C52G5.2	WT_RNAi
WBGene00016922|C54F6.4	WT_RNAi
WBGene00000437|R13A5.5|ceh-13	WT_RNAi
WBGene00000462|T26C11.5|ceh-41	WT_RNAi
WBGene00011122|R07H5.2|cpt-2	WT_RNAi
WBGene00017028|D1044.2	WT_RNAi
WBGene00008366|D1046.5|tpra-1	nonwt_RNAi
WBGene00000908|F11A1.3|daf-12	nonwt_RNAi
WBGene00000914|F33H1.1|daf-19	WT_RNAi
WBGene00017381|F11D5.3|ddr-2	WT_RNAi
WBGene00012116|T28B8.5|del-4	WT_RNAi
WBGene00000959|F46H6.2|dgk-2	WT_RNAi
WBGene00000961|T21B6.1|dgn-1	WT_RNAi
WBGene00018943|F56C3.6|dgn-2	WT_RNAi
WBGene00008438|DH11.5	nonwt_RNAi
WBGene00012832|Y43F8C.10|dmd-3	WT_RNAi
WBGene00010484|K01H12.1|dph-3	WT_RNAi
WBGene00001078|F22B7.10|dpy-19	WT_RNAi
WBGene00001086|R13G10.1|dpy-27	WT_RNAi
WBGene00001131|F15D3.1|dys-1	WT_RNAi
WBGene00008483|E04D5.4	WT_RNAi

WBGene00001148|H30A04.1|eat-20	WT_RNAi
WBGene00001186|F55A8.1|egl-18	WT_RNAi
WBGene00001196|M01D7.7|egl-30	WT_RNAi
WBGene00001173|F55A8.2|egl-4	WT_RNAi
WBGene00001210|K11G9.4|egl-46	WT_RNAi
WBGene00001174|C08C3.1|egl-5	nonwt_RNAi
WBGene00007772|C27C12.2|egrh-1	WT_RNAi
WBGene00001253|F52C12.5|elt-6	WT_RNAi
WBGene00001309|M01D7.6|emr-1	WT_RNAi
WBGene00008564|F08A8.1|acox-1	WT_RNAi
WBGene00017265|F08F3.9	WT_RNAi
WBGene00008621|F09C8.1	WT_RNAi
WBGene00017317|F09G2.9	WT_RNAi
WBGene00017501|F15E11.15	nonwt_RNAi
WBGene00017534|F17A9.2	WT_RNAi
WBGene00017535|F17A9.3	WT_RNAi
WBGene00008951|F19C6.2	WT_RNAi
WBGene00017950|F31E3.2	WT_RNAi
WBGene00009306|F32A7.5	WT_RNAi
WBGene00018378|F43C11.1	WT_RNAi
WBGene00009688|F44E5.1	WT_RNAi
WBGene00018623|F48G7.12	WT_RNAi
WBGene00009888|F49E2.5	WT_RNAi
WBGene00018928|F56B3.2	WT_RNAi
WBGene00016138|C26E6.2|flh-2	WT_RNAi
WBGene00001449|F07D3.2|flp-6	WT_RNAi
WBGene00010453|K01B6.1|fozi-1	WT_RNAi
WBGene00014177|ZK1010.3|frg-1	WT_RNAi
WBGene00001519|Y40H4A.1|gar-3	WT_RNAi
WBGene00021697|Y48G9A.3|gcn-1	WT_RNAi
WBGene00001570|F58A4.11|gei-13	nonwt_RNAi
WBGene00001580|F28D1.10|gex-3	WT_RNAi
WBGene00021980|Y58A7A.6|glb-32	nonwt_RNAi
WBGene00001636|Y75B8A.9|gly-11	WT_RNAi
WBGene00010593|K06A4.3|gsnl-1	WT_RNAi
WBGene00001753|R03D7.6|gst-5	WT_RNAi
WBGene00010416|H25P06.1	WT_RNAi
WBGene00006452|R148.6|heh-1	WT_RNAi
WBGene00001860|F28B3.7|him-1	nonwt_RNAi
WBGene00009002|F21C3.3|hint-1	WT_RNAi
WBGene00001957|F48D6.3|hlh-13	WT_RNAi
WBGene00001962|F57C12.3|hlh-19	nonwt_RNAi
WBGene00001949|M05B5.5|hlh-2	nonwt_RNAi
WBGene00001953|C02B8.4|hlh-8	nonwt_RNAi
WBGene00001995|K08H2.6|hpl-1	WT_RNAi
WBGene00002012|F38E11.1|hsp-12.3	nonwt_RNAi

WBGene00002026|C12C8.1|hsp-70	WT_RNAi
WBGene00002071|T04G9.3|ile-2	WT_RNAi
WBGene00019030|F58B6.2|inft-1	nonwt_RNAi
WBGene00002101|T28B8.2|ins-18	WT_RNAi
WBGene00002175|Y105C5B.21|jac-1	WT_RNAi
WBGene00019344|K02G10.1	WT_RNAi
WBGene00010704|K09A11.1	WT_RNAi
WBGene00019609|K10B3.1	WT_RNAi
WBGene00010767|K11B4.2	WT_RNAi
WBGene00019673|K12C11.1	WT_RNAi
WBGene00002216|T09A5.2|klp-3	WT_RNAi
WBGene00002267|C44F1.3|lec-4	WT_RNAi
WBGene00020605|T20B12.9|lgc-50	WT_RNAi
WBGene00010716|K09C8.4|lge-1	WT_RNAi
WBGene00006407|F25H5.1|lim-9	nonwt_RNAi
WBGene00003015|W03C9.4|lin-29	nonwt_RNAi
WBGene00003089|K02C4.4|ltd-1	WT_RNAi
WBGene00019737|M02F4.3	WT_RNAi
WBGene00019751|M03D4.4	WT_RNAi
WBGene00003235|C52E4.2|mif-2	WT_RNAi
WBGene00003238|Y34D9B.1|mig-1	WT_RNAi
WBGene00003367|M106.1|mix-1	nonwt_RNAi
WBGene00003426|F36H12.7|msp-19	nonwt_RNAi
WBGene00003424|F26G1.7|msp-3	nonwt_RNAi
WBGene00003435|C33F10.9|msp-40	nonwt_RNAi
WBGene00003438|F58A6.8|msp-45	nonwt_RNAi
WBGene00003442|C34F11.6|msp-49	nonwt_RNAi
WBGene00021839|Y54E10BL.5|nduf-5	WT_RNAi
WBGene00017742|F23F1.1|nfyc-1	WT_RNAi
WBGene00003696|T01G6.4|nhr-106	WT_RNAi
WBGene00003700|Y46H3D.5|nhr-110	WT_RNAi
WBGene00003718|F44C8.5|nhr-128	WT_RNAi
WBGene00017510|F16B4.9|nhr-178	WT_RNAi
WBGene00017367|F10G2.9|nhr-263	WT_RNAi
WBGene00003661|K11E4.5|nhr-71	WT_RNAi
WBGene00003684|C12D5.8|nhr-94	WT_RNAi
WBGene00003687|H27C11.1|nhr-97	nonwt_RNAi
WBGene00003738|F54F3.1|nid-1	WT_RNAi
WBGene00003779|Y75B8A.2|nob-1	nonwt_RNAi
WBGene00003968|T14F9.4|peb-1	WT_RNAi
WBGene00004010|Y48A6C.5|pha-1	nonwt_RNAi
WBGene00004024|Y75B8A.1|php-3	WT_RNAi
WBGene00004032|F57F5.5|pkc-1	WT_RNAi
WBGene00004076|W09B6.1|pod-2	WT_RNAi
WBGene00004134|F59B10.1|pqn-47	WT_RNAi
WBGene00004166|Y43H11AL.3|pqn-85	nonwt_RNAi

WBGene00004192|ZK809.7|prx-2	WT_RNAi
WBGene00004213|C48D5.2|ptp-1	WT_RNAi
WBGene00004224|F55F8.1|ptr-10	WT_RNAi
WBGene00004219|C45B2.7|ptr-4	WT_RNAi
WBGene00004259|D2085.1|pyr-1	WT_RNAi
WBGene00019842|R02F11.4	WT_RNAi
WBGene00019867|R04B3.2	WT_RNAi
WBGene00019904|R05G9.2	WT_RNAi
WBGene00019945|R08C7.1	WT_RNAi
WBGene00020125|R173.3	WT_RNAi
WBGene00016354|C33F10.5|rig-6	WT_RNAi
WBGene00017183|F02E11.5|scl-15	WT_RNAi
WBGene00004780|C09B7.1|ser-7	WT_RNAi
WBGene00004873|Y47D3A.26|smc-3	nonwt_RNAi
WBGene00004874|F35G12.8|smc-4	nonwt_RNAi
WBGene00004921|F31E8.2|snt-1	WT_RNAi
WBGene00004962|F53G12.6|spe-8	WT_RNAi
WBGene00007918|C34C6.5|sphk-1	nonwt_RNAi
WBGene00004999|K09F5.3|spp-14	WT_RNAi
WBGene00005003|F27C8.4|spp-18	WT_RNAi
WBGene00009743|F45H11.1|sptf-1	WT_RNAi
WBGene00011926|T22C8.5|sptf-2	WT_RNAi
WBGene00005113|R186.2|srd-35	WT_RNAi
WBGene00005517|F28D9.2|sri-5	WT_RNAi
WBGene00015419|C04E6.2|srsx-18	nonwt_RNAi
WBGene00005719|F53F1.7|srv-8	WT_RNAi
WBGene00005936|K01B6.2|srx-45	WT_RNAi
WBGene00044068|ZK867.1|syd-9	nonwt_RNAi
WBGene00011373|T02D1.7	WT_RNAi
WBGene00011772|T14G8.4	WT_RNAi
WBGene00011798|T16G1.4	WT_RNAi
WBGene00011824|T18D3.7	WT_RNAi
WBGene00020615|T20D4.9	WT_RNAi
WBGene00020630|T20F7.1	nonwt_RNAi
WBGene00011873|T20G5.12	WT_RNAi
WBGene00020808|T25F10.6	WT_RNAi
WBGene00006520|D2013.10|tag-175	WT_RNAi
WBGene00016423|C34H3.1|tag-275	WT_RNAi
WBGene00044324|ZK652.3|tag-277	WT_RNAi
WBGene00016871|C52B11.2|tag-303	WT_RNAi
WBGene00006562|K01C8.3|tdc-1	WT_RNAi
WBGene00006498|R13F6.4|ten-1	nonwt_RNAi
WBGene00006575|F13B10.1|tir-1	WT_RNAi
WBGene00017298|F09E10.8|toca-1	nonwt_RNAi
WBGene00006668|R04F11.4|twk-13	nonwt_RNAi
WBGene00006672|C24A3.6|twk-18	nonwt_RNAi

WBGene00006683|Y47D3B.5|twk-31	WT_RNAi
WBGene00006831|C52E12.2|unc-104	WT_RNAi
WBGene00006792|T06H11.1|unc-58	WT_RNAi
WBGene00006804|Y37D8A.13|unc-71	WT_RNAi
WBGene00006823|C06A5.7|unc-94	nonwt_RNAi
WBGene00006875|T22C8.8|vab-9	nonwt_RNAi
WBGene00012193|W02A11.2|vps-25	WT_RNAi
WBGene00012162|VZC374L.1	WT_RNAi
WBGene00020917|W01B11.6	WT_RNAi
WBGene00012366|W09G3.2|maea-1	nonwt_RNAi
WBGene00012549|Y37D8A.8	WT_RNAi
WBGene00012658|Y39A1A.21	WT_RNAi
WBGene00021546|Y43H11AL.2	WT_RNAi
WBGene00012871|Y45F10B.3	WT_RNAi
WBGene00012996|Y48C3A.16	WT_RNAi
WBGene00013042|Y49E10.23|cccp-1	nonwt_RNAi
WBGene00021736|Y50D4A.2|wrb-1	nonwt_RNAi
WBGene00013178|Y53H1A.2	WT_RNAi
WBGene00013239|Y56A3A.22	WT_RNAi
WBGene00022280|Y74C10AL.2	WT_RNAi
WBGene00010178|F57A8.2|yif-1	WT_RNAi
WBGene00022518|ZC123.3	nonwt_RNAi
WBGene00006975|F54F2.2|zfp-1	nonwt_RNAi
WBGene00018704|F52E4.8|ztf-13	WT_RNAi
WBGene00006993|Y79H2A.11|zyg-8	nonwt_RNAi

Table S9. Genes With Either RNAi Phenotypes or YFP Transgenes Tested for LC Expression.

S9A. Genes tested with both RNAi and YFP.

Gene	Orthology/description				LC exp.	Spec.	Shape	Path	Delay	Det.	Str.	n=	% abn
hlh-8	Helix-loop-helix (TWIST) TF	TF	D	O	no	-	--		-			23	35
nob-1	Homeobox protein abdominal-B	TF			no		-	---	---			59	51
syd-9	Zinc-finger	TF			no	-	---		---			22	50
arx-7	Arp2/3 complex, subunit ARPC5			O	yes		--	---	---		--	62	60
gar-3	Muscarinic acetylcholine receptor			O	yes								
him-1	Cohesin subunit SMC1		D		yes (Ab)		---		---	---		26	69
hsp-12.3	Alpha-crystallin				no		---		---			25	52
ins-18	Insulin				no								
msp-3	Major sperm protein (MSP)				yes		---	-	---			28	36
pkc-1	Protein kinase C		D		yes								
R02F11.4	Centrosomal protein of 97 kDa (CEP97)			O	no								
smc-4	Condensin subunit SMC4			O	yes		---		---	---		28	82
sphk-1	Sphingosine kinase			O	yes			---	--		---	44	32
srsx-18	7-transmembrane receptor				yes			-	---			79	41
srx-45	7-transmembrane receptor				yes								
T20D4.9	Neprilysin; peptidase family M13				no								
yif-1	Yip1 interacting factor; cons. membrane protein			O	yes								
zyg-8	Doublecortin				no		---	-	--			42	45

S9B. Genes tested with RNAi.

Gene	Orthology/description				LC exp.	Spec.	Shape	Path	Delay	Det.	Str.	n=	% abn
crh-2	CREB/ATF family transcription factor	TF		O			-	--	---		--	33	42
daf-12	Nuclear Hormone Receptor	TF	D	O			---		---		---	21	48
egl-5	zerknullt-like homeobox protein	TF					--		---	---	---	13	69
gei-13	Nondescript; strongly expressed in L3 LCs	TF					---	---	---			34	56
hlh-19	Helix-loop-helix TF	TF				--	-	--	---		-	84	45
hlh-2	Helix-loop-helix (E/Da) TF	TF		O		---						11	55
lin-29	Zinc finger protein 362	TF		O		---	---	--	---			17	94
nhr-97	Nuclear factor 4 hormone receptor	TF					--	-	---			71	44
pha-1	Nematode-specific TF	TF					---	---	--			20	45
T20F7.1	Large (738-residue) nematode-specific TF	TF					---	---	---			35	54
ZC123.3	Zinc-finger homeodomain protein 2	TF	D				-		---		--	9	56
zfp-1	PHD finger protein AF10	TF	D	O			--		---		---	38	50
cccp-1	Conserved coiled-coil protein						--		---	-		17	35
DH11.5	Nondescript; strongly expressed in L4 LCs						-	-	---		---	18	56
F15E11.15	BED Zn-finger; strongly expressed in LCs								---			33	36
glb-32	Globin					-	---	---	---	-	--	103	43
inft-1	Rho GTPase effector BNI1 and rel. formins			O			---		--			22	59
lim-9	Four and a half lim domains protein			O			--	---	--			52	44
maea-1	Macrophage-erythroblast attacher protein			O		--	--	---	--			51	51
mix-1	Condensin subunit SMC2			O			-		---			11	55
msp-19	Major sperm protein (MSP)						-	-	---			13	46
msp-40	Major sperm protein (MSP)						--	-	---		-	10	70
msp-45	Major sperm protein (MSP)						--		-			8	25
msp-49	Major sperm protein (MSP)								---			10	60
pqn-85	SCC2/Nipped-B		D	O		-	--	-	---	-	--	82	45
smc-3	Cohesin subunit SMC3		D	O			---		---	---		12	83
ten-1	Teneurin-1						---		---		--	24	67
toca-1	Cdc42-interacting protein CIP4						--	---	---			62	39
tpra-1	Conserved membrane protein			O			--	--	---	-	---	71	45
twk-13	Tandem pore domain K+ channel					-	--	---	---			21	62
twk-18	Tandem pore domain K+ channel							---	-			23	39
unc-94	Tropomodulin and leiomodulin			O			--	--	---		--	73	43
vab-9	Brain cell membrane protein 1 (BCMP1)					--	-	---	---			51	31
wrb-1	Tryptophan-rich basic protein						---		---			28	39

S9C. Genes tested with YFP.

Gene	Orthology/description				LC exp.	Spec.	Shape	Path	Delay	Det.	Str.	n=	% abn
acr-15	Acetylcholine receptor				no	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.		
acr-16	Acetylcholine receptor				yes	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.		
glr-1	AMPA glutamate receptor subunit GluR2		D		no	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.		
glr-2	AMPA glutamate receptor subunit GluR2		D		yes	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.		

"LC exp." denotes genes tested for LC expression with YFP transgenes (except *him-1*, determined with antibodies). "TF" denotes known or predicted transcription factors, "D" denotes human disease orthologs, and "O" denotes strict orthologies to human genes. Phenotypes for RNAi are listed as in Fig. 3. Increasingly severe phenotypes are listed from "-" to "---", with "n/d" for genes not tested with RNAi. "n=" and "% abn." denote the number of animals scored for a gene's RNAi phenotype and the percentage showing at least one phenotype.

Table S10. Details of Genes Tested for LC Expression by YFP Transgenes.

Gene	LC expr.	Assay	5'- forward primer-3' or other source	5'- (YFP seq) + reverse primer-3'
acr-15	no	p::YFP	CG398	
acr-16	yes	p::YFP	GCTCAACAACTACTTTTTTCCTCC	ACGGACATGAGAATCAGGGAAAGA
arx-7	yes	p::YFP	GATTTTCACCAACCAACCAACCCA	TTTTTCCTATCTGAAACAAGAGAT
gar-3	yes	p::YFP	CG308	
glr-1	no	p::YFP	KP987	
glr-2	yes	p::YFP	CCTGATGATGCTATTGTGTTTTGC	CGCTACAAAAGCTCCTGAAATTAAACC
him-1	yes	antibody		
hlh-8	no	p::YFP	PS5149	
hsp-12.3	no	p::YFP	GTCAACATGTTTACGTGGTTTCTG	ATCAGAAAATAGTTTCAAAAAGTCCC
ins-18	no	p::YFP	PS5309 PS5326	
msp-3	yes	p::YFP	CACTTCTCAATATACTAACCCTAC	GGCGACTTGTTGGAAGAGTGTGAA
nob-1	no	p::YFP	HC183	
pkc-1	yes	p::YFP	IK934	
R02F11.4	no	p::YFP	GCCTTCTTTCATTCCACACGCTTT	AGTGCCTGAAAGAAAATTGCATCA
smc-4	yes	p::YFP	CAAAGCTTCAGCGTCAAAAAATGA	CAACCCTAAAATACGTTGATTTTT
sphk-1	yes	p::YFP	GCTCCAGCAACTTCTACGTTAAGT	GTGATGTTCGGTGCAAGTGCACTT
srsx-18	yes	p::YFP	CCACCTTTTTTGGAAGCTGTTTAT	TGTCTGAAAATATAGATGGTAATG
srx-45	yes	p::YFP	GGAGTAAAGCTAGTGGGGATTTTG	ATCTTCATTTTTAGTGAGAAAACA
syd-9	no	SYD-9::YFP		
T20D4.9	no	p::YFP	GTAGTAGAGCAAACGGAAAATGGA	TCTCAGTTAGTATTAAGAACAAAA
yif-1	yes	p::YFP	GAACAAGTGGATAGGAGTCCTGAG	CTTCAGCTGGAAAATAAAATTGGT
zyg-8	no	p::YFP	CTCGAACAACATACTCGGCAAAAT	CGTTCGAGTAGAGATAGTCTAACA

YFPseq is: GAAAAGTTCTTCTCCTTTACTCAT. Strains and their genotypes are listed in Methods.

Table S11. Gene Functions Preferentially Expressed in Pooled LCs.

root_node_name	node_name	node_id	p-value
molecular_function	sequence-specific DNA binding transcription factor activity	GO:0003700	6.47E-05
molecular_function	receptor activity	GO:0004872	0.000634083
molecular_function	protein binding	GO:0005515	5.60E-05
molecular_function	cytoskeletal protein binding	GO:0008092	0.000757682
molecular_function	hydrogen ion transmembrane transporter activity	GO:0015078	0.000365059
molecular_function	sequence-specific DNA binding	GO:0043565	3.35E-08
molecular_function	metal ion binding	GO:0046872	0.000365329
cellular_component	axon	GO:0030424	0.00011527
biological_process	nucleosome assembly	GO:0006334	0.000488117
biological_process	regulation of transcription, DNA-dependent	GO:0006355	0.000103383
biological_process	intracellular protein transport	GO:0006886	0.0001177
biological_process	cell adhesion	GO:0007155	1.11E-05
biological_process	G-protein coupled receptor protein signaling pathway	GO:0007186	0.000129775
biological_process	small GTPase mediated signal transduction	GO:0007264	0.000246439
biological_process	synaptic transmission, cholinergic	GO:0007271	0.000837103
biological_process	response to heat	GO:0009408	0.000106112
biological_process	regulation of signal transduction	GO:0009966	0.000299928
biological_process	regulation of cellular component size	GO:0032535	0.000453619
biological_process	locomotion	GO:0040011	3.61E-05
biological_process	regulation of locomotion	GO:0040012	0.000302899
biological_process	regulation of multicellular organism growth	GO:0040014	0.000153664
biological_process	regulation of pharyngeal pumping	GO:0043051	0.000348343
biological_process	post-translational protein modification	GO:0043687	0.000627103
biological_process	regulation of protein metabolic process	GO:0051246	0.000500838
biological_process	regulation of biological quality	GO:0065008	0.000456184
biological_process	cellular component assembly at cellular level	GO:0071844	0.000477604

This gives the full set of GO terms overrepresented among LC-enriched genes. LC-enriched genes are defined here as those with high ratios of expression in LCs (either L3- or L4-stage, whichever was stronger) to expression in whole larvae.

Table S12. Gene Functions Upregulated From Pooled L3- to L4-stage LCs in an NHR-67-dependent Manner.

S12A. The full set of GO terms overrepresented among genes with high ratios of expression in pooled L4-stage LCs to expression in pooled L3-stage LCs.

root_node_name	node_name	node_id	p-value
molecular_function	dopamine beta-monooxygenase activity	GO:0004500	0.00487488
molecular_function	ribonuclease H activity	GO:0004523	0.00408664
molecular_function	protein serine/threonine kinase activity	GO:0004674	4.61E-05
molecular_function	protein tyrosine kinase activity	GO:0004713	0.000453022
molecular_function	protein tyrosine phosphatase activity	GO:0004725	1.58E-05
molecular_function	structural molecule activity	GO:0005198	0.00171799
molecular_function	potassium channel activity	GO:0005267	0.000933981
molecular_function	cyclin-dependent protein kinase regulator activity	GO:0016538	0.00512191
cellular_component	integral to membrane	GO:0016021	5.38E-11
cellular_component	contractile fiber part	GO:0044449	0.00690771
biological_process	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	GO:0000184	0.00414124
biological_process	protein phosphorylation	GO:0006468	8.60E-06
biological_process	protein dephosphorylation	GO:0006470	0.000110879
biological_process	methionine metabolic process	GO:0006555	0.00620301
biological_process	water-soluble vitamin metabolic process	GO:0006767	0.00821988
biological_process	regulation of muscle contraction	GO:0006937	0.0013925
biological_process	G-protein coupled receptor protein signaling pathway	GO:0007186	0.00713486
biological_process	cellular amino acid catabolic process	GO:0009063	0.0033618
biological_process	nucleotide biosynthetic process	GO:0009165	0.00815794
biological_process	post-embryonic body morphogenesis	GO:0040032	0.00195914
biological_process	phosphoinositide-mediated signaling	GO:0048015	0.00190175

Table S12 (continued). Gene Functions Upregulated From L3- to L4-stage LCs in an NHR-67-dependent Manner.

S12B. The full set of GO terms overrepresented among genes with high ratios of expression in pooled wild-type L4-stage LCs to expression in pooled *nhr-67*(RNAi) L4-stage LCs.

root_node_name	node_name	node_id	p-value
molecular_function	DNA ligase (ATP) activity	GO:0003910	0.00203034
molecular_function	protein serine/threonine kinase activity	GO:0004674	1.34E-05
molecular_function	protein tyrosine kinase activity	GO:0004713	0.000120011
molecular_function	protein tyrosine phosphatase activity	GO:0004725	0.000308518
molecular_function	nicotinic acetylcholine-activated cation-selective channel activity	GO:0004889	0.00777197
molecular_function	structural molecule activity	GO:0005198	0.000621201
molecular_function	structural constituent of cytoskeleton	GO:0005200	0.00550409
molecular_function	cation channel activity	GO:0005261	0.00201471
molecular_function	potassium channel activity	GO:0005267	0.00153337
molecular_function	calcium ion binding	GO:0005509	0.00100914
molecular_function	protein binding	GO:0005515	0.000519604
molecular_function	voltage-gated cation channel activity	GO:0022843	0.0047468
molecular_function	neurotransmitter binding	GO:0042165	0.00895835
cellular_component	integral to membrane	GO:0016021	1.44E-06
cellular_component	sarcomere	GO:0030017	0.00143301
cellular_component	A band	GO:0031672	0.00573005
cellular_component	postsynaptic membrane	GO:0045211	0.0067511
cellular_component	striated muscle dense body	GO:0055120	0.000416344
biological_process	protein phosphorylation	GO:0006468	3.27E-06
biological_process	protein dephosphorylation	GO:0006470	7.63E-05
biological_process	glutamine biosynthetic process	GO:0006542	0.00958687
biological_process	Mo-molybdopterin cofactor biosynthetic process	GO:0006777	0.00410362
biological_process	sodium ion transport	GO:0006814	0.00857127
biological_process	regulation of muscle contraction	GO:0006937	0.00182022
biological_process	homophilic cell adhesion	GO:0007156	0.00451715
biological_process	signal transduction	GO:0007165	0.00314681
biological_process	G-protein coupled receptor protein signaling pathway	GO:0007186	0.00468049
biological_process	vitelline membrane formation	GO:0030704	0.00975122
biological_process	directional locomotion	GO:0033058	0.00914112
biological_process	positive regulation of locomotion	GO:0040017	0.00337218
biological_process	negative regulation of cell cycle	GO:0045786	0.00805532
biological_process	positive regulation of smooth muscle contraction	GO:0045987	0.00403835
biological_process	regulation of behavior	GO:0050795	0.00665497

Table S12 (continued). Gene Functions Upregulated From L3- to L4-stage LCs in an NHR-67-dependent Manner.

S12C. The overlap of GO terms from Tables S12A and S12B, representing genes upregulated from pooled L3- to L4-stage LCs in an NHR-67-dependent manner. The largest (least significant) p-value from Tables S12A-S12B is shown for each term.

root_node_name	node_name	node_id	p-value
biological_process	G-protein coupled receptor protein signaling pathway	GO:0007186	0.00713486
biological_process	protein dephosphorylation	GO:0006470	0.000110879
biological_process	protein phosphorylation	GO:0006468	8.60E-06
biological_process	regulation of muscle contraction	GO:0006937	0.00182022
cellular_component	integral to membrane	GO:0016021	1.44E-06
molecular_function	potassium channel activity	GO:0005267	0.00153337
molecular_function	protein serine/threonine kinase activity	GO:0004674	4.61E-05
molecular_function	protein tyrosine kinase activity	GO:0004713	0.000453022
molecular_function	protein tyrosine phosphatase activity	GO:0004725	0.000308518
molecular_function	structural molecule activity	GO:0005198	0.00171799

Table S13. Overrepresentation of Sperm Genes Among LC Genes With High Wild-Type L4/L3 Expression Ratios Versus *nhr-67*(RNAi) L4/L3 Ratios.

Genes		L4/L3 expression ranking		*nhr-67*/L3 expression ranking	
Class of genes	**No. of genes**	**p-value**	**z-score**	**p-value**	**z-score**
Sperm	389	$<10^{-6}$	22	$<10^{-6}$	5.9
Shared sperm	328	$<10^{-6}$	21	$<10^{-6}$	6.2
Male sperm	24	$1.2 \cdot 10^{-5}$	4.4	0.94	0.1
Herm. sperm	37	$7.3 \cdot 10^{-5}$	4.0	0.67	0.4
Sperm/MSP	44	$<10^{-6}$	11	$<10^{-6}$	6.2
Sperm/non-MSP	345	$<10^{-6}$	20	$7.6 \cdot 10^{-5}$	4.0
MSP/non-sperm	15	$1.1 \cdot 10^{-2}$	2.6	0.32	1.0
LC-enriched	1,097	0.74	0.3	$<10^{-6}$	-5.7
Housekeeping	1,164	$<10^{-6}$	-9.0	$<10^{-6}$	7.0

The 8,011 genes detected in wild-type pooled LCs were ranked by ratios of their observed gene activity, for either wild-type L4-stage to wild-type L3-stage LCs or *nhr-67*(RNAi) L4-stage to wild-type L3-stage LCs. Subsets of the pooled LC genes were then selected by various criteria and given a Wilcoxon rank-sum test to see if they were overrepresented among either high-ranking or low-ranking genes by either the L4/L3 or the *nhr-67*/L3 ranking. The degree of their bias was assayed by their z-score (obtained by the normal approximation in the Wilcoxon test). Genes with positive z-scores are skewed towards higher ranks; negative z-scores, the reverse. The subsets given here include: all of the LC genes which were previously observed to have sperm-enriched expression by Reinke et al. (24); the subset of those genes which were enriched in both male and hermaphrodite sperm ("Shared sperm"); the subset enriched in male sperm; the subset enriched in hermaphrodite sperm; subdivisions of the sperm-expressed LC genes shown in Figure S2 ("sperm/MSP", "sperm/non-MSP", and "MSP/non-sperm"), "LC-enriched" (defined as in Figure 2), and "Housekeeping" (defined as in Figure 2). Generally speaking, sperm-associated genes are more upregulated than expected by chance from L3- to wild-type L4-stage LCs, but much less strongly so from L3- to *nhr-67*(RNAi) L4-stage LCs. That this is not a universal pattern is shown by the rank distributions of LC-enriched genes (which show no bias for L4/L3, while being sharply downranked in *nhr-67*/L3) and of housekeeping genes (which have lower ranks than expected randomly in L4/L3, but higher ranks than expected in *nhr-67*/L3).

Supplemental Method: Detailed Protocol For Single-Cell RT-PCR

General notes

Although our method is closely similar to that of Dulac and Axel (4), we have had several requests for our detailed protocol, and thus include it here.

There are two materials in this protocol which readers may be tempted to replace with cheaper substitutes. These reagents are Platinum Taq DNA polymerase, and filter-protected pipet tips. However, for this procedure to succeed reliably, Platinum Taq or an equivalent polymerase, rather than a cheaper generic Taq polymerase, must be used in the RT-PCR. We have had at least one coworker disregard this advice and end up wasting time and scarce cells. We suspect this is because ordinary Taq is more susceptible than Platinum Taq to primer-dimer artifacts, which are a very high risk in single-celled RT-PCRs with low annealing temperatures and high initial ratios of primer to substrate.

We also urge that filter-protected pipet tips be used throughout the procedure, not merely during the RT-PCR but during all preparatory steps in which reagents for the RT-PCR are made, because unfiltered pipet tips allow PCRs to be taken over by aerosol-borne products of previous PCRs (38). The reader should remember that this is a procedure aimed at getting very rare transcripts out of single cells, and that these transcripts must be amplified out of a sea of contaminants and artifacts. Moreover, it is far more expensive to perform next-generation sequencing on junk than to use filter tips the first time.

We serially dilute oligonucleotide solutions to 100 μM and then 10 μM in T_{10} (10 mM Tris pH 3.0), store them at -20˚ C., and routinely thaw only the 10 μM stock for use. We prefer T_{10} to H_20 because distilled water, unless specifically deionized, can be acidic rather than neutral.

We found that dNTPs from New England Biolabs performed reliably in making stock solutions for RT-PCRs.

The dNTPs in the reverse transcription step are at a very low concentration. We found that successful RT-PCRs declined if an aliquot of dNTPs was used even twice. We thus make many small aliquots of dilute dNTPs, freeze them all, only thaw one aliquot of dilute dNTPs for a given reverse transcription step, and then throw it away.

More recent single-cell RT-PCR protocols have been developed which may well have superior sensitivity or reliability to the protocol given here, and which readers should consider using (39, 40). We consider the main advantages of this protocol to be its simplicity and robustness.

Set up reagents (before Day 1):

1. Make RNAse-free H_20, TE, and T_{10}. H_20 should be glass-distilled and autoclaved. TE should be made by standard means (41) and autoclaved. T_{10} is simply TE without EDTA, i.e., 10 mM Tris pH 8.0.

Each solution should be made RNAse-free after autoclaving by filtering twice through two 0.2-μm Nalgene CN nitrocellulose filter units (not tissue-culture filters, and not 0.45-μm filters; VWR, catalog number 28199-076). Solutions can then be stored at 4˚ C. or room temperature. In our experience, solutions prepared this way are RNAse-free enough to allow rare RNAs of up to ~10 kb in size to remain intact (42). There are two different reasons why this method may work. The first is that the nitrocellulose filter, unlike filters used for tissue culture medium, is likely to absorb trace proteins; in particular, it is likely to absorb trace RNAses which have survived autoclaving. The second is that at least some "RNAse" contamination may be due to the

blooming of bacterial spores that survive autoclaving and later come to life in a solution; double-filtering is likely to purge solutions of such spores. In any event, double-filtering is more convenient than DEPC treatment, and can be easily adapted to any filterable aqueous solution.

2. Order reagents for RT-PCR. For this protocol, we used the following reagents from these specific suppliers:

Platinum Taq Polymerase from Invitrogen. 500 reactions, 5 U/ul, catalog number 10966-034.

10x PCR buffer, from Invitrogen (included with Platinum Taq Polymerase).

50 mM $MgCl_2$, from Invitrogen (included with Platinum Taq Polymerase).

AMV reverse transcriptase (AMV-RT) from New England Biolabs. 500 U, catalog number M0227 T.

M-MuLV reverse transcriptase (MuLV-RT) from New England Biolabs. 200 U/µl, catalog number M0253S.

MuLV-RT reaction buffer from New England Biolabs (included with MuLV-RT).

Anti-RNAse from Ambion, 2500 U (15-30 U/µl), catalog number AM2690.

Stop RNase Inhibitor from 5 PRIME, 2000 U (1x 500 µl; 4 U/µl), catalog number 2500200.

Terminal Transferase (TdT) from New England Biolabs. 500 U, catalog number M0315 S.

2.5 mM $CoCl_2$ from New England Biolabs (included with Terminal Transferase).

10x terminal transferase buffer (NEBuffer 4) from New England Biolabs (included with Terminal Transferase).

Oligo$(dT)_{12-18}$ primer from Invitrogen, catalog number SKU# 18418-012.

dNTP solution set (250 µl apiece of 100 mM dATP, dCTP, cGTP, and dTTP) from New England Biolabs; catalog number N0446S.

Pre-lubricated, DNAse- and RNAse-free 1.7-ml microfuge tubes from Sorenson Bioscience; catalog number 1170.

AeroGuard Aerosol-Barrier Pipet Tips from National Scientific Supply Company, 0.1-10.0 µl, pack of 960, catalog number UN0011-MRS.

AeroGuard Aerosol-Barrier Pipet Tips from VWR. 1-30 µl, pack of 960, catalog number 53502-902; 1-140 µl, pack of 960, catalog number 53502-901; 200-1000 µl, pack of 1000, catalog number 53502-905.

Bovine Serum Albumin (nuclease-free) from Roche Applied Science. 20 mg/ml, catalog number 10711454001.

Igepal CA-630 (Nonidet-P40 substitute) from USB. 500 ml, catalog number 19628.

Triton X-100, molecular biology grade, Sigma, cat. no. T8787.

Qiagen MinElute PCR Purification Kit, catalog number 28004; or Qiagen Qiaquick PCR Purification Kit, catalog number 28104. While we used MinElute for this study, Qiaquick may be better at removing oligonucleotides, and may thus be preferable.

3. Order AL1 primer. Its sequence is: 5'-ATT-GGA-TCC-AGG-CCG-CTC-TGG-ACA-AAA-TAT-GAA-TTC-TTT-TTT-TTT-TTT-TTT-TTT-TTT-TTT-3'.

4. Order 5' and 3' primers for two adjacent exons of a control positive gene. In our protocol, this gene was *rpl-17* from *C. elegans*, and was targeted with the following primers: RPL17TEST01, 5'-CAC-CTC-GTT-ATT-GAG-CAC-ATC-AAC-GTT-CAA-3'; and RPL17TEST02, 5'-GGC-

ATC-ATC-GGT-TGG-CTT-GGA-GAC-AA-3'. These primers are expected to amplify a product of ~155 nt from cDNA (e.g., valid RT-PCR products) but one of ~390 nt from genomic DNA. BlastN showed no close matches of these primers to any other amplicons in the genome. One can thus use them with genomic DNA as a positive control for amplification, while distinguishing its product from amplified cDNA.

5. Make a stock 100 µM AL1 primer solution in RNAse-free T_{10}. Note that 2.7 µl of 100 µM AL1 equals 5 µg of AL1. Make stock 100 µM and 10 µM primer solutions for RPL17TEST01 and RPL17TEST02. Store all oligonucleotide solutions at -20˚ C.

6. Make stock primer mix, for reverse transcription step:

5.0 µl 100 mM dATP
5.0 µl 100 mM dCTP
5.0 µl 100 mM dGTP
5.0 µl 100 mM dTTP
19.6 µl of 0.5 µg/µl oligo$(dT)_{12\text{-}18}$ primer
0.4 µl H_20 [with 2.0-µl pipettor]

Dispense into 1.5 µl aliquots, and store at -20˚ C. between uses.

The amounts of dNTPs in this mixture, after being diluted 25-fold for the reaction itself, are deliberately low. This is in order to bias the reverse transcription towards making equally short copies (0.5-1.5 kb) of all poly$(A)^+$ mRNAs, regardless of their actual length. A practical implication of this is that reusing the same aliquot is unreliable; even a small amount of degradation of the dNTPs turns a limited reverse transcription step into a failed reverse transcription step.

7. Make 2x tailing buffer, with added cobalt:

20.0 µl of 10x terminal transferase buffer (NEBuffer 4, New England Biolabs)
20.0 µl 2.5 mM $CoCl_2$ (New England Biolabs)
1.5 µl 100 mM dATP (New England Biolabs)
47.5 µl H_20

Store at -20˚ C. between uses. 2x tailing buffer should give a final concentration of 10 U/4.5 µl when mixed with terminal transferase (TdT) enzyme. For New England Biolabs TdT (20 U/µl), mix 0.5 µl enzyme with 4.0 µl 2x tailing buffer.

8. Make 1x RNAse-free dissection buffer (1x XM2; (3)), with the following composition: 145 mM NaCl; 5 mM KCl; 1 mM CaCl2; 5 mM MgCl2; 10 mM HEPES; 20 mM D-glucose; pH 7.2.

9. Program PCR machine. Given pre-dissected, frozen cells, one can run both reactions on a single reserved machine on a single day. The first half of this PCR has a theoretical run time of 4 hours and 35 minutes, but (in our experience) a real run time of 5 hours. The second half has a theoretical run time of 3 hours and 45 minutes, but in practice is most conveniently run overnight.

Set up the following SSPCR program:

1. 94˚ C., 30".
2. 94˚ C., 30".
3. 42˚ C., 2'.

4. 72.0˚ C., 6' + 10"/cycle.
5. Go back to step 2, 24x.
6. 42˚ C., 1 hr. [Enzyme boost step, without overt mixing.]
7. 94˚ C., 30".
8. 94˚ C., 30".
9. 42˚ C., 2'.
10. 72˚ C., 6'.
11. Go back to step 8, 24x.
12. 4˚ C., 18 hr.
13. Stop.

Dissections (Day 1):

1. Set up reagents for snap-freezing dissected samples: a small container of liquid nitrogen (a thermos flask is ideal, but an ice bucket will do if necessary; and an ice bucket with well-ground dry ice. Do not expect cells to remain intact if kept on wet ice at 4˚ C. after dissection.

2. Label pre-lubricated RNAse-free microfuge tubes into which to put dissected cells. Cells will be deposited into a small volume (~1 µl) of RNAse-free dissection buffer.

3. Mix 1x RNAse-free dissection buffer (1x XM2) with RNAse inhibitors and nuclease-free BSA for the dissections:

48.5 µl XM2
0.5 µl Anti-RNAse inhibitor (15-30 U/µl)
0.5 µl Stop RNase Inhibitor (30 U/µl)
0.5 µl BSA

The BSA is meant to nonspecifically bind sticky sites and prevent rare cellular RNAs from being lost. Do not use DTT, even though some protocols recommend, it, because it clogs dissection needles.

4. Dissect cells (including one mock dissection as a negative control, if possible) into tubes labelled 1 to around 15, keeping track in one's notebook of which tubes were actually used for what cells. Use a pre-lubricated microfuge tube for each dissection, to help ensure that tiny aliquots of cellular material (e.g., rare RNA) do not stick to the tube walls or otherwise get lost.

The detailed dissection protocol closely follows that of Lockery and Goodman (3), but with the following differences. Cells that are tightly bound to their neighbors should be severed from them before the dissection itself, by using laser microsurgery (43) to ablate neighboring cells. Pipets are left blunt (and thus able to take up a dissected cell) rather than polished into being patch pipettes. A second dissection near the head is not necessary if the cell in question is associated with the gonad (as the linker cell is), since the initial abdominal cut should free the gonad immediately. Finally, a nontoxic cyanoacrylate glue, such as PeriAcrly®90 (GluStitch Inc., catalog number P-ACRYL5(V)CE), is used to glue down worms.

5. Spin down and snap-freeze each newly dissected cell. After each dissection, spin the cell(s) down briefly. Then immediately snap-freeze the cell's tube by immersion in liquid N_2. Move the frozen cells to -80˚ C. (on well-ground dry-ice in a ice bucket), and store them at -80˚ C. until amplification. Typically, it will be most convenient to do dissections on one day, then start

RT-PCR on a later day. Cells can be safely kept at -80˚ C. for up to a few weeks, or even shipped on dry ice from one location to another (which can be useful when the dissection and the amplification must be done in different laboratories).

Reverse-transcription and RT-PCR (Day 2, earlier)

The first steps of RT-PCR can be done most efficiently if one has water baths set to 37˚ C. and 65˚ C., along with dedicated floaters, so that one can readily pick up 1-2 floaters of tubes rather than transfer samples tube by tube. Also, a floater of tubes can be moved from one water bath to another of a given temperature at once, whereas hand-transferring samples tube by tube smears out step times. Alternatives to water baths or floaters are acceptable if they soak the entire tube in the given temperature and allow rapid changes of heat.

For experiments using ~10 cells or cell equivalents, a small glass dish or Tupperware tub with distilled water in it, on one's desktop, makes the best (most convenient) room-temperature water bath. It can be used at one's bench and then carried over to other water baths before transferring floaters from it.

Unless one has thin fingers and fine muscular control, it is probably a good idea to carry out most of the steps below in 1.5-ml tubes, even though the final PCR will be carried out in 0.5-ml tubes with thin walls. We routinely use 1.5-ml tubes up to the very end of setting up the reaction, transfer reactions to 0.5-ml tubes for the PCR itself, and then transfer them back after the PCR to 1.5-ml tubes for long-term storage at -20 or -80˚ C.

1. Plan appropriate reactions. In addition to amplification of dissected cells, there should be several controls.

First of all, there should be a known dilute RNA sample from the organism of interest, which one can reliably use as a positive control for amplification with reverse transcriptase (+control RNA, +RT). In our case, we used diluted whole RNA that we had previously extracted from starved mass wild-type (N2) *C. elegans* larvae, which we had already shown to work well in standard RNA-seq (6). Since this RNA contained transcripts from many different cell types, any gene strongly expressed both in our cell type of interest (linker cells) and in the mass larval RNA sample was likely to be housekeeping, and could be classified as such.

There should be another control with reverse transcriptase but no substrate (either a mock dissection, or just H_20), and a third control with no reverse transcriptase. These provide a check against spurious amplification or contaminants in one's reaction mix.

2. Set up reagents. Before starting, get a wet ice bucket; make sure the 37˚ C. bath is well watered; get the 65˚ C. bath set up; and have a "bath" of room-temperature water for the oligo-T annealing step. Have labelled, pre-lubricated, RNAse-free microfuge tubes for the reactions. Have some standard way to label PCRs reliably (e.g., the system described below).

3. Set up a new 1/25 dilution of new stock primer mix:

1.0 µl stock primer mix

2x 12.0 µl H_20

For greater accuracy, one can pipet 12.0 µl twice with a 20-ul pipettor, rather than pipet 24.0 µl once with a 200-ul pipettor. The stock primer mix should from an aliquot freshly thawed, used once, and then discarded after this experiment.

4. Set up cDNA-lysis buffer. For ≤10 samples, set up 50 µl:
40.5 µl H_20
5.0 µl MuLV-RT reaction buffer, 10x
2.5 µl 10% v/v Igepal CA-630 [This is a substitute for NP40.]
[Vortex the tube containing the above components, and microfuge the tube, before adding the following components.]
0.5 µl Anti-RNAse inhibitor (15-30 U/µl)
0.5 µl Stop RNase Inhibitor (30 U/µl)
1.0 µl 1/25 dilution of new stock primer mix
Gently mix the full solution by inverting its tube or flicking its tube with a finger; then microfuge the tube's contents down.

5. Make 1:1 v:v mix of AMV- and MuLV-RT. Viscosity of these solutions at -20˚ C. makes aliquotting inefficient. To compensate for this, make a slight excess of mix. E.g., for 5 reactions, mix 1.5 µl AMV-RT + 1.5 µl MuLV-RT.

6. Keep the cDNA-lysis buffer and the AMV-RT/MuLV-RT mix on ice until use.

7. Put all vials containing frozen single cells in room-temperature water bath, and let them thaw.

8. Aliquot 4.0 µl of pre-chilled cDNA-lysis buffer to each tube with ~1.0 µl substrate -- whether that substrate is blank H_20, dilute control RNA, or cells. Briefly spin down all tubes; flick each tube gently with one's finger to mix contents; spin the tubes again. Warm them in a floater rack in a room-temperature water bath for a few seconds.

9. Lyse the cells by moving them, in their floater rack, to a 65 C water bath for 1'.

10. Anneal oligo-T, by immediately moving the tubes in their rack to a room-temperature water bath for 1'. Then put the tubes on ice.

11. Microfuge the tubes for 2' at 4˚ C.

12. Add 0.5 µl AMV-RT/MuLV-RT mix apiece to all PCR reactions except the (highly recommended) non-RT control. Put tubes on ice. Flick and spin to get well mixed.

13. Incubate at 37˚ C. for exactly 15'.

14. Incubate the reaction tubes in a water bath for 10' at 65˚ C.

15. Put tubes on ice. Microfuge for 2' at 4˚ C.

16. Mix tailing/transferase solution. For 5 reactions mix:
2x 12.0 µl 2x tailing buffer
3.0 µlTerminal transferase, NEB, 20 U/µl
For 15 reactions mix:
72.0 µl 2x tailing buffer

9.0 µl Terminal transferase, NEB, 20 U/µl
Again, it is more accurate to aliquot 2x12.0 µl than 24.0 µl.

17. Add 4.5 µl of tailing/transferase solution to each reaction tube. Flick and spin to get well-mixed.

18. Incubate each tube at 37˚ C. for 15'.

19. Prepare 5% Triton-X100. While the previous step is going on, mix 5.0 µl Triton-X100 and 95.0 µl H_20. Since Triton-X100 is viscous and does not mix easily with water at room temperature, this requires successive pre-heating of aliquots of H_20 and Triton-X100, and then of the 5% mixture, to get the former to pipet accurately and the latter to mix quickly. Preheating of microfuge vials containing small amounts of H_20 and Triton-X100 can be done in a 65˚ C. water bath.

20. Set up labeled thin-walled tubes for PCR.

21. To inactivate the TdT, incubate reactions for 20' at 65˚ C. (75˚ C. may be a better temperature for inactivation, but the protocol we adapted used 65˚ C.)

22. Microfuge for 2' at 4˚ C.

23. Set up PCR reactions. Do all PCR mixes in pre-lubricated, RNAse-free microfuge tubes in order to prevent loss of reaction materials. Use all 10 µl of each finished reverse transcription reaction, to add to 90 µl of PCR mix and make a 100 µl final reaction volume.

24. Make a fresh PCR mix. For N reactions, premix 1.1xN aliquots to prevent shortfalls due to imprecise micropipetting.

Material:	1x rxn.	5 rxns.:	15 rxns, in 2 mixes:
H_20	65.0 µl	4x 89.5 µl	5x 117 µl [**twice**]
10x PCR buffer	10.0 µl	55.0 µl	82.5 µl ["]
50 mM MgCl2	5.0 µl	27.5 µl	41 µl [etc.]
20 mg/ml BSA	0.5 µl	2.8 µl	4.1 µl
----------mix here before proceeding----------			
100 mM dATP	1.0 µl	5.5 µl	8.3 µl
100 mM dCTP	1.0 µl	5.5 µl	8.3 µl
100 mM dGTP	1.0 µl	5.5 µl	8.3 µl
100 mM dTTP	1.0 µl	5.5 µl	8.3 µl
5% Triton X-100	1.0 µl	5.5 µl	8.3 µl
Platinum Taq**	2.0 µl	11.0 µl	16.5 µl
100 µM AL1****	2.7 µl	14.9 µl	11.1 + 11.2 µl

*The recipe here gives 5.5x the needed amounts for a single reaction, since losses of up to 10% through pipetting can cause shortfalls.

**In other words, make two mixes of 8.25x the reaction mixture needed for a single reaction, each in an sterile microfuge tube.

***Dulac's original protocol required 2.0 µl for the first step and 1.0 µl for the boost. This is 5x as much Taq as Invitrogen's manual says needed (0.4 µl per 100-µl reaction; at most, up to 1.0 µl). However, this protocol uses Dulac's enzyme amounts.

****This is 5 µg of AL1 primer.

Flick and microfuge them to get them well-mixed.

25. Aliquot 90 µl of PCR mix to each reaction tube, with a fresh pipet tip for each aliquot. Flick and microfuge them to get them well-mixed.

26. Transfer the reactions from 1.5-ml tubes to to thin-walled 0.5-ml PCR tubes. Microfuge the reactions down briefly to prevent bubbling.

27. Run the first half of the SSPCR program for 5 hours (check it at 4 hr., 45 min.; it should be done in exactly 5 hr.) before hot boost at 42˚ C.

Second round of RT-PCR (Day 2, later):

1. As soon as the first half of SSPCR ends, add a boost of 1.0 µl Platinum Taq/reaction. Keep the reaction at 42˚ C. to prevent primer-dimer. Do this one tube at a time, quickly, while the other tubes remain on the 42˚ C. block, and return each tube to the block immediately after pipetting. Do not try to mix the enzyme into solution; simply pipet it into the top of the tube, let it sink to the bottom of the tube, and let heat convection mix it in. After all tubes have been given an enzyme boost, seal the PCR machine and prompt the machine to go on to step 7 of SSPCR.

2. Run the rest of SSPCR. This will typically run overnight.

Initial results of RT-PCR (Day 3):

1. Store the bulk of reactions at -20˚ C. until further work; do long-term storage at -80˚ C. For convenient storage and handling, it will probably be best to pipet the completed reactions back into 1.5-ml microfuge tubes.

2. Run 10.0 µl of each reaction per lane on a 1.5% EtBr-stained gel, between flanking lanes of 100-bp molecular weight markers.

3. Photograph and evaluate the gel. Any successful RT-PCR will give a smear of DNA from 0.5-1.5 kb in size, which should not appear in negative controls. Ideally, negative controls should be blank. However, if bands appear in the negative controls, this is not necessarily a sign of failure; there may be cryptic poly(A)$^+$ RNAs in at least some reagents used that were amplified in the controls. What is really crucial is that broad smears should appear in positive controls and valid cell amplifications, and that secondary PCRs should give cDNA-sized products only from valid amplifications.

Secondary PCRs (Day 3 or later):

1. Test successful RT-PCRs for the presence or absence at least one diagnostic gene, with gene-specific oligonucleotide pairs. These should span an intron and amplify 300-500 bp, with the 5'-most one being targetted ≤500 bp from the predicted/known end of the transcript, with blank and genomic DNA as control substrates. We have found that *rpl-17* in *C. elegans* is a

highly robust positive control that fits these criteria. Cell-specific genes, particularly if they are expressed at low levels, may not show up reliably in secondary PCR of even a valid reaction, because individual cells have highly variable patterns of amplification.

2. Purify successful PCRs with the Qiagen MinElute or the Qiagen Qiaquick PCR Purification Kit. Quantitate them with Qubit. For both purification and quantitation, follow manufacturer's instructions. We find Qubit quantitation to be more reliable than other methods such as Nanodrop.

3. Carry out next-generation sequencing, either on individual cleaned RT-PCRs, or on a quantitatively balanced pool of them (e.g., 1.0 µg apiece for each reaction from a given cell type, pooled into a single solution for sequencing). The details of sequencing will depend on which sequencing facility one is using. Generally, though, Illumina libraries should be constructed and sequenced as if they were meant for standard RNA-seq (6): i.e., the inserts should be ~300 nt in size (i.e., they should have ~200 nt of real insert and 2x50 nt of flanking adapters), and the reads should be single-end. We used 38-nt reads in this study, but 50-nt reads have recently become more common and are probably preferable.

Labelling PCRs reliably

The above protocol entails many different PCRs, which rapidly fill one's freezer boxes and are easy to confuse if not tracked stringently. To avoid this problem, we use the following labeling method, adopted from early *Drosophila* genetics (where it was used to identify newly discovered alleles). We label PCRs by the code "XX [lowercase a-l] YY.ZZ" which denotes the last two digits of the year (XX), the month (a = January, l = December), the day (YY), and the reaction number for that day (ZZ). Thus, "12f20.01" would denote PCR reaction number 1 on June 20th, 2012. This code ensures that the exact date and identity of any PCR tube is identifiable simply by reading its lid, and lowers the risk of confusing PCRs.

Supplemental References

1. Lewis JA & Fleming JT (1995) Basic culture methods. *Methods Cell Biol.* 48:3-29.
2. Kato M & Sternberg PW (2009) The *C. elegans tailless*/Tlx homolog *nhr-67* regulates a stage-specific program of linker cell migration in male gonadogenesis. *Development* 136(23):3907-3915.
3. Lockery SR & Goodman MB (1998) Tight-seal whole-cell patch clamping of *Caenorhabditis elegans* neurons. *Methods Enzymol.* 293:201-217.
4. Dulac C & Axel R (1995) A novel family of genes encoding putative pheromone receptors in mammals. *Cell* 83(2):195-206.
5. Hall D, Altun, ZF (2008) *C. elegans Atlas* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York).
6. Mortazavi A, Williams BA, McCue K, Schaeffer L, & Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5(7):621-628.
7. Langmead B, Trapnell C, Pop M, & Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10(3):R25.
8. Flicek P*, et al.* (2011) Ensembl 2011. *Nucleic Acids Res.* 39(Database issue):D800-806.
9. Mortazavi A*, et al.* (2010) Scaffolding a *Caenorhabditis* nematode genome with RNA-seq. *Genome Res.* 20(12):1740-1747.
10. Yook K*, et al.* (2012) WormBase 2012: more genomes, more data, new website. *Nucleic Acids Res.* 40(D1):D735-D741.
11. Finn RD*, et al.* (2010) The Pfam protein families database. *Nucleic Acids Res.* 38(Database issue):D211-222.
12. Koonin EV*, et al.* (2004) A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol* 5(2):R7.
13. Bendtsen JD, Nielsen H, von Heijne G, & Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* 340(4):783-795.
14. Krogh A, Larsson B, von Heijne G, & Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305(3):567-580.
15. Lupas A (1996) Prediction and analysis of coiled-coil structures. *Methods Enzymol.* 266:513-525.
16. Wootton JC (1994) Non-globular domains in protein sequences: automated segmentation using complexity measures. *Comput. Chem.* 18(3):269-285.
17. Boutros M & Ahringer J (2008) The art and design of genetic screens: RNA interference. *Nat Rev Genet* 9(7):554-566.
18. Ashburner M*, et al.* (2000) Gene ontology: tool for the unification of biology. *Nat. Genet.* 25(1):25-29.
19. Muller J*, et al.* (2010) eggNOG v2.0: extending the evolutionary genealogy of genes with enhanced non-supervised orthologous groups, species and functional annotations. *Nucleic Acids Res.* 38(Database issue):D190-195.
20. Reece-Hoyes JS*, et al.* (2005) A compendium of *Caenorhabditis elegans* regulatory transcription factors: a resource for mapping transcription regulatory networks. *Genome Biol* 6(13):R110.

21. Haerty W, Artieri C, Khezri N, Singh RS, & Gupta BP (2008) Comparative analysis of function and interaction of transcription factors in nematodes: extensive conservation of orthology coupled to rapid sequence evolution. *BMC Genomics* 9:399.
22. Li L, Stoeckert CJ, Jr., & Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13(9):2178-2189.
23. Amberger J, Bocchini C, & Hamosh A (2011) A new face and new challenges for Online Mendelian Inheritance in Man (OMIM®). *Hum. Mutat.* 32(5):564-567.
24. Reinke V, Gil IS, Ward S, & Kazmer K (2004) Genome-wide germline-enriched and sex-biased expression profiles in *Caenorhabditis elegans*. *Development* 131(2):311-323.
25. Nishimura H & L'Hernault SW (2010) Spermatogenesis-defective (spe) mutants of the nematode *Caenorhabditis elegans* provide clues to solve the puzzle of male germline functions during reproduction. *Developmental dynamics : an official publication of the American Association of Anatomists* 239(5):1502-1514.
26. Kulkarni M, Shakes DC, Guevel K, & Smith HE (2012) SPE-44 implements sperm cell fate. *PLoS Genet* 8(4):e1002678.
27. Gleason EJ*, et al.* (2012) Developmental genetics of secretory vesicle acidification during *Caenorhabditis elegans* spermatogenesis. *Genetics* 191(2):477-491.
28. Salkoff L*, et al.* (2005) Potassium channels in *C. elegans*. *WormBook* 2005 Dec 30:1-15.
29. Galperin MY & Koonin EV (2004) 'Conserved hypothetical' proteins: prioritization of targets for experimental study. *Nucleic Acids Res.* 32(18):5452-5463.
30. Gertz EM, Yu YK, Agarwala R, Schäffer AA, & Altschul SF (2006) Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. *BMC Biol* 4:41.
31. de Hoon MJ, Imoto S, Nolan J, & Miyano S (2004) Open source clustering software. *Bioinformatics* 20(9):1453-1454.
32. Gollub J & Sherlock G (2006) Clustering microarray data. *Methods Enzymol.* 411:194-213.
33. Tibshirani R, Walther G, & Hastie T (2001) Estimating the number of clusters in a data set via the gap statistic. *J. R. Statist. Soc. B* 63:411-423.
34. Saldanha AJ (2004) Java Treeview--extensible visualization of microarray data. *Bioinformatics* 20(17):3246-3248.
35. Ben-Shaul Y, Bergman H, & Soreq H (2005) Identifying subtle interrelated changes in functional gene categories using continuous measures of gene expression. *Bioinformatics* 21(7):1129-1137.
36. Prüfer K*, et al.* (2007) FUNC: a package for detecting significant associations between gene sets and ontological annotations. *BMC Bioinformatics* 8:41.
37. Anders S & Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* 11(10):R106.
38. Kwok S & Higuchi R (1989) Avoiding false positives with PCR. *Nature* 339(6221):237-238.
39. Tang F*, et al.* (2010) RNA-Seq analysis to capture the transcriptome landscape of a single cell. *Nat Protoc* 5(3):516-535.
40. Islam S*, et al.* (2011) Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res.* 21(7):1160-1167.
41. Sambrook JR, D.W. (2001) *Molecular Cloning: a Laboratory Manual.* (CSHL Press, Cold Spring Harbor, NY) 3rd Ed.

42. Schwarz EM & Benzer S (1997) *Calx*, a Na-Ca exchanger gene of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U. S. A.* 94(19):10249-10254.
43. Fang-Yen C, Gabel CV, Samuel AD, Bargmann CI, & Avery L (2012) Laser microsurgery in *Caenorhabditis elegans*. *Methods Cell Biol.* 107:177-206.

Figure S1

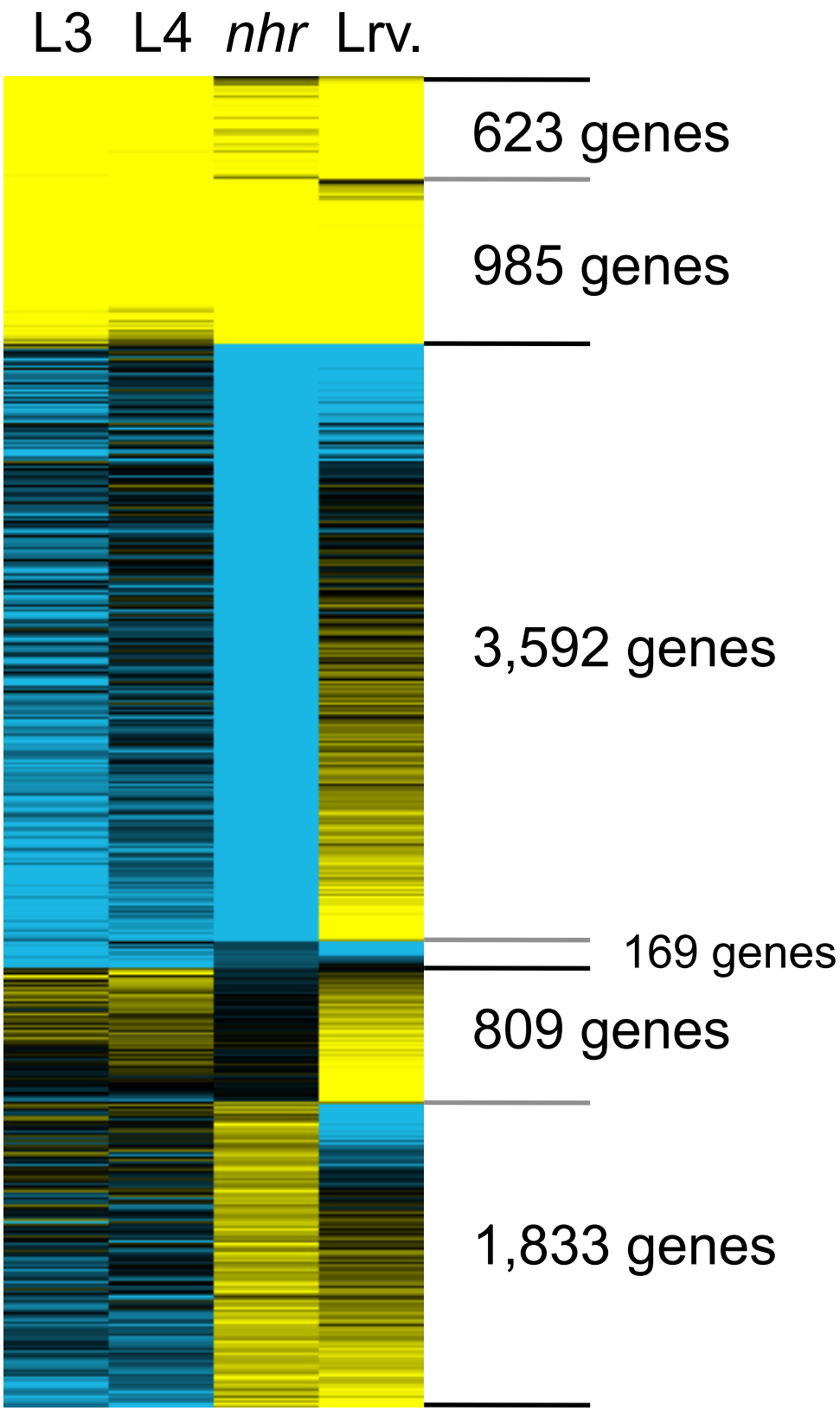
L3
L4
nhr
Lrv.
623 genes
985 genes
3,592 genes
169 genes
809 genes
1,833 genes

Figure S2

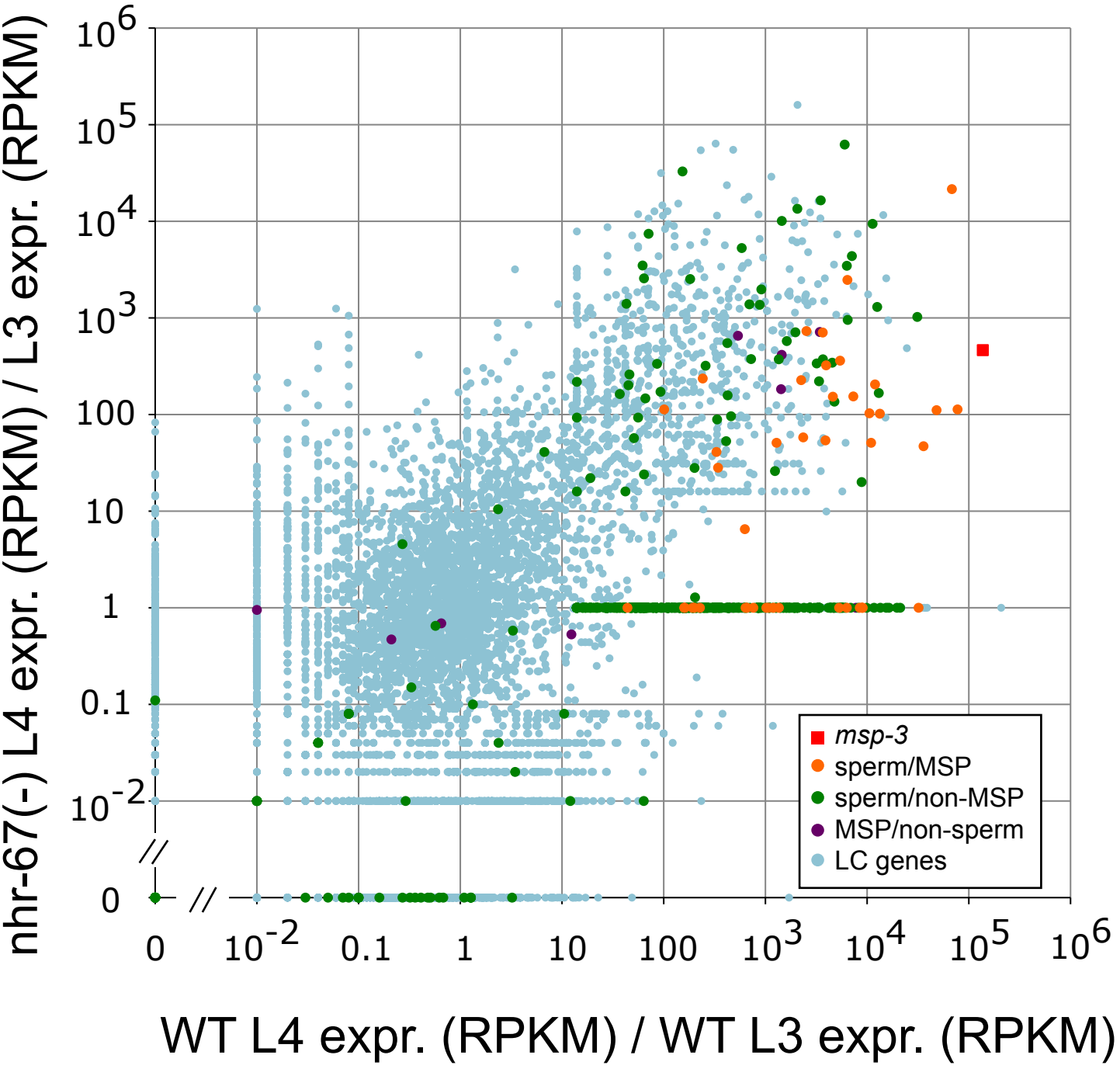
nhr-67(-) L4 expr. (RPKM) / L3 expr. (RPKM)
WT L4 expr. (RPKM) / WT L3 expr. (RPKM)
0
10^-2
0.1
1
10
100
10^3
10^4
10^5
10^6
msp-3
sperm/MSP
sperm/non-MSP
MSP/non-sperm
LC genes

Figure S3