

Genome-wide germline-enriched and sex-biased expression profiles in *Caenorhabditis elegans*

Valerie Reinke^{1,*}, Inigo San Gil¹, Samuel Ward² and Keith Kazmer¹

¹Department of Genetics, Yale University School of Medicine, New Haven, CT 06520, USA

²Department of Molecular and Cellular Biology, University of Arizona, Tucson, AZ 85721, USA

*Author for correspondence (e-mail: valerie.reinke@yale.edu)

Accepted 9 October 2003

Development 131, 311-323
Published by The Company of Biologists 2004
doi:10.1242/dev.00914

Summary

We performed a genome-wide analysis of gene expression in *C. elegans* to identify germline- and sex-regulated genes. Using mutants that cause defects in germ cell proliferation or gametogenesis, we identified sets of genes with germline-enriched expression in either hermaphrodites or males, or in both sexes. Additionally, we compared gene expression profiles between males and hermaphrodites lacking germline tissue to define genes with sex-biased expression in terminally differentiated somatic tissues. Cross-referencing hermaphrodite germline and somatic gene sets with *in situ* hybridization data demonstrates that the vast majority of these genes have appropriate spatial expression patterns. Additionally, we examined gene expression at

multiple times during wild-type germline development to define temporal expression profiles for these genes. Sex- and germline-regulated genes have a non-random distribution in the genome, with especially strong biases for and against the X chromosome. Comparison with data from large-scale RNAi screens demonstrates that genes expressed in the oogenic germline display visible phenotypes more frequently than expected.

Supplemental data available online

Key words: *C. elegans*, Germline, Sex determination, Microarrays, Gene expression, X chromosome

Introduction

Knowledge of spatial and temporal gene expression profiles facilitates functional annotation of the genome by providing likely site(s) of action for the corresponding gene products. Previously, we defined a set of 1416 genes with germline-enriched expression using DNA microarrays representing a subset of the predicted genes in the *C. elegans* genome (Reinke et al., 2000). This partial set of genes has been the focus of several subsequent reverse-genetic and yeast-two-hybrid screens designed to identify gene functions and interactions important for either germline or early embryonic processes (Colaiácovo et al., 2002; Piano et al., 2002; Walhout et al., 2002; Pellettieri et al., 2003). Additionally, the partial germline-enriched gene set facilitated the cloning of genetic mutants with known defects in germline development (MacQueen and Villeneuve, 2001) (S.W., unpublished), as well as providing a gene list to test candidates required for specific events during germline development (i.e. oocyte maturation) (Miller et al., 2003). We present a comprehensive analysis of temporal and mutant gene expression profiles in the hermaphrodite germline of *C. elegans*. Additionally, we examine sex-regulated gene expression differences between males and hermaphrodites, in both germline and somatic tissue.

In the bi-lobed gonad of the hermaphrodite nematode *C. elegans*, the cells destined to become gametes originate from a common pool of germline stem cells that exist within a niche at the distal end of the gonad. Upon leaving the niche, germ cells enter meiosis I and undergo chromosomal synapsis and recombination. During the fourth larval stage (L4), meiotic

germ cells within the proximal region of the gonad differentiate into spermatocytes. After the L4-to-adult molt, the germ cells in the proximal gonad instead differentiate into oocytes.

Regulation of gene expression at transcriptional and post-transcriptional levels is crucial for the proper specification, proliferation and differentiation of germ cells (reviewed by Seydoux and Strome, 1999; Goodwin and Evans, 1997; Kuwabara and Perry, 2001). However, the regulatory networks controlling genes involved in either germ cell fate decisions or terminal differentiation of gametes are only superficially understood. Additionally, global silencing mechanisms that control the expression of large regions of the genome overlay gene-specific regulation, as demonstrated by the selective silencing of repetitive transgenes in the germline (Kelly et al., 1997). The largest endogenous target of silencing identified to date is the X chromosome, which is silenced to differing degrees in the germline of males and hermaphrodites (Kelly et al., 2002; Fong et al., 2002). A thorough understanding of how both global and gene-specific regulation of gene expression contributes to the proper functioning of the germline requires a comprehensive knowledge of the genes expressed in that tissue.

Males appear among the hermaphrodite population at a low frequency, when X chromosome nondisjunction occurs during meiosis. Somatic tissues of males differ from hermaphrodites in several ways: males lack a vulva and uterus, have a single-lobed gonad, and do not produce vitellogenin in the intestine. Male-specific structures include both a broad, fan-shaped tail as well as sex-specific neurons that direct physical and behavioral aspects of mating (Emmons and Sternberg, 1997). In the male

germline, many of the initial steps in germ cell development are morphologically similar to hermaphrodites, although they make only sperm. Male spermatozoa are larger than hermaphrodite spermatozoa, and this size difference is one factor that promotes the preferential use of male sperm over self-sperm by a mated hermaphrodite (LaMunyon and Ward, 1998). Genome-wide gene expression studies that identify molecular similarities and differences between the two sexes for both somatic and germline tissues are crucial for investigations of the underlying genetic pathways that generate sex-specific structures.

In the past few years, many microarray-based expression studies have been performed on diverse aspects of *C. elegans* development, such as embryogenesis, aging, pharyngeal development, muscle development and dauer formation (Baugh et al., 2003; Lund et al., 2002; Gaudet and Mango, 2002; Roy et al., 2002; Wang and Kim, 2003). Additionally, several large-scale expression and functional analyses have been performed in *C. elegans*, including in situ hybridization of ESTs and genome-wide RNAi (Kohara, 2001; Kamath et al., 2003). Integration of the diverse data types from these global investigations can result in increased strength and specificity of functional predictions. For example, in situ data provides spatial patterns that are only inferred in whole-animal microarray studies, while RNAi studies provide valuable information about reduction-of-function phenotypes. Beyond compiling evidence for single genes, genome-wide functional approaches also facilitate observations on a different scale from single-gene studies. In particular, several of these global studies have demonstrated that genes are non-randomly arranged in the *C. elegans* genome with respect to both gene expression and function (reviewed by Reinke, 2002; Piano et al., 2002; Kamath et al., 2003).

We use DNA microarrays corresponding to 92% of the currently predicted genes in the *C. elegans* genome to examine expression profiles among mutant strains with defects in germline proliferation or gamete production, as well as between males and hermaphrodites. Together these experiments identify 5629 genes that show distinct germline- or sex-dependent expression profiles. In addition, we have pinpointed a small set of genes with expression likely to be specific in the male germline. Investigation of the kinetics of gene expression during wild-type hermaphrodite larval and adult development demonstrates that sets of genes with germline-enriched expression are highly temporally co-regulated, but that genes with sex-biased expression in the soma are less so. We also extend the previous observation that the chromosomal location of germline-enriched transcripts is non-random in the genome. Our studies show that genes expressed in the germline are under-represented on the X chromosome and, conversely, that genes with hermaphrodite soma-biased expression are enriched on the X chromosome. We also discuss previously unreported biases on autosomes. Comparison between germline- and sex-regulated genes and large-scale RNAi screens demonstrates that genes with expression in oogenic germlines are enriched for visible phenotypes relative to other gene expression sets.

Materials and methods

Sample preparation

Strains used: wild type is *C. elegans* variety Bristol strain N2; linkage

group I, *glp-4(bn2ts)* (Beanan and Strome, 1992); linkage group IV, *fem-1(hc17ts)* (Nelson et al., 1978); *fem-3(q23gf)* (Barton et al., 1987); and linkage group V, *him-5(e1490)* (Browerman and Meneely, 1994).

The wild-type reference RNA used is identical to that of Reinke et al. (Reinke et al., 2000), and by mass is approximately composed of 40% gravid adults, 30% larvae, 15% embryos and 15% post-reproductive adults. The reference sample was used solely to allow comparison between genotypes; we did not infer any biological meaning for the sample/reference ratios. The L4 and adult N2 and *glp-4* hermaphrodite samples and the *fem-1* and *fem-3* hermaphrodite samples are also identical to those described by Reinke et al. (Reinke et al., 2000). The *him-5* samples and *glp-4;him-5* samples were prepared by growing worms on 15-cm² plates and synchronizing as described previously (Reinke et al., 2000). Starved L1 larvae were plated on 15 cm² plates, raised at 25°C until the young adult stage, and then harvested and washed in S basal buffer. The worms were centrifuged on a sucrose cushion to remove debris and recovered in several washes of S-basal. The worms were then filtered through a 30 µm nylon mesh (Spectrum Laboratories, Rancho Dominguez, CA) into a petri dish containing S-basal for 10 minutes. Males passed through the mesh, while hermaphrodites were retained on top. The filtration was repeated a second time, and male purity was determined by examining the sex of 100 animals. A population had to consist of 90-95% males before it was used in a microarray experiment.

Timecourse samples were grown by taking synchronized L1 larvae, plating them on 15 cm plates and growing them at 25°C until the middle of L3. Samples were taken every three hours for 36 hours by harvesting approx. five plates and washing the worm pellet several times in S-basal until most bacteria was removed, with a final resuspension of the worm pellet in four volumes of Trizol (Gibco/BRL). At each collection, the developmental stage of the animal was verified by inspection of vulval and germline formation using Nomarski optics.

Total RNA and polyA purification was performed as described elsewhere (Reinke et al., 2000).

Microarrays

Microarrays were constructed essentially as described elsewhere (Reinke et al., 2000; Jiang et al., 2001). The set of 19,213 primer pairs corresponding to ~94% of the genes in the genome was used to PCR amplify gene fragments as described previously (Jiang et al., 2001). Out of these, 18,010 produced a single band of the correct size. Reverse transcription, labelling and hybridization to the arrays were performed as described elsewhere (Reinke et al., 2000).

Data analysis

The average mean log₂ ratio was calculated for each set of replicates comparing a staged sample of specific genotype to the common reference. Comparisons were made between genotypes by subtracting the mean log value of one ratio from another, and the significance of the difference was evaluated using Student's *t*-test for two populations. For the *fem-3(gf)* versus *fem-1(lf)* direct comparison, we performed the same analysis, except we used a Student's *t*-test for one population. We chose a combination of a twofold difference with a *t* value exceeding 99% confidence ($P < 0.01$), because these criteria allowed the inclusion of essentially all genes that had previously been identified as germline-enriched in a wt/*glp-4* hermaphrodite comparison (Reinke et al., 2000). Additionally, requiring a twofold difference reduced false positives, as the number of genes with twofold difference and a $P < 0.01$ only included ~100 genes more than with $P < 0.001$, and almost all genes showed germline expression by in situ hybridization (Table 1).

At our laboratory website (<http://wormgermline.yale.edu>), the expression profiles of genes can be searched as single-gene or multi-gene queries and the text files containing the gene sets described in Figs 1-3, and the order of genes in the cluster diagram of Fig. 5 can

be downloaded. Additionally, the gene sets are available in a zipped file (Data S1) at <http://dev.biologists.org/supplemental>. The raw data are available at Yale Microarray Database (YMD; http://ymd.med.yale.edu/ymd_prod/cgi-bin/ymd_public_data.cgi) and at the Gene Expression Omnibus (GEO; <http://www.ncbi.nlm.nih.gov/geo/>, Accession Numbers GSE715-GSE737).

The molecular function category of Gene Ontology (GO) was used to assign functional annotation to ~20% of the genes within the different expression subsets systematically. To consolidate related categories, these molecular annotations were combined under slightly broader categories according to GO Slim annotations used for *Drosophila*, and then graphed using pie charts. Gene clustering analysis was performed with a hierarchical clustering package based on the Pearson correlation coefficient, using the average linkage of \log_2 ratios (Eisen et al., 1998). The significance of the distribution of sex- and germline-regulated genes on different chromosomes relative to the expected number was evaluated using a hypergeometric probability test. To examine the distribution of RNAi phenotypes among the gene sets, available RNAi data were downloaded from Wormbase (www.wormbase.org) and compiled so that independent assays were combined. Any gene that displayed an RNAi phenotype of embryonic lethal (Emb) or sterile (Ste or Stp) in even one assay was counted. All other phenotypes affecting growth, body morphology or somatic development were coalesced into the post-embryonic category.

Results

Definition of primary sex- and germline-enriched gene sets

We wished to globally identify genes with germline- and sex-regulated expression profiles in *C. elegans* hermaphrodites and males. To generate large populations of hermaphrodite nematodes lacking a germline, we used the *glp-4(bn2)* strain, which bears a temperature-sensitive mutation that results in under-proliferation of germ cells and a largely empty somatic gonad at the restrictive temperature (Beanan and Strome, 1992). To generate large numbers of males, we used the *him-5(e1490)* mutation, which causes an increase in X chromosome nondisjunction and leads to a population consisting of ~30% males (Hodgkin et al., 1979). We increased the percentage of males to ~95% by using a filtration system to separate males from hermaphrodites (Materials and methods).

We collected poly A+ RNA from populations of wild type (N2) and *glp-4(bn2)* hermaphrodites that were synchronized to either the fourth larval stage (L4) or young adult stage. We also collected poly A+ RNA from *him-5(e1490)* and *glp-4(bn2);him-5(e1490)* purified adult male populations. Each sample was independently grown and harvested four times. Using DNA microarrays representing 18,010 of the 19,546 protein-encoding genes currently annotated in the *C. elegans* genome, we compared each of these samples with a common reference sample. The reference sample comprises hermaphrodites at all life stages: embryos, L1-L4 larvae, young adults and post-reproductive adults. For each gene in each microarray experiment, we calculated a \log_2 ratio of the staged sample relative to the reference sample. We averaged the \log_2 ratios for each set of four replicates. The common reference used in all hybridizations allows us to compare the average expression ratios among these four genotypes and define sex- and germline-regulated genes (Fig. 1). In each comparison, we required the fold difference in expression to exceed twofold and a confidence level of 99% ($P < 0.01$, Student's *t*-test) for

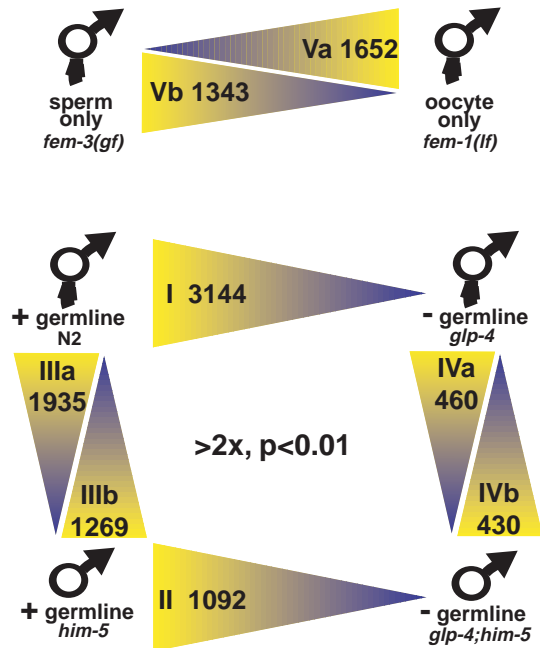


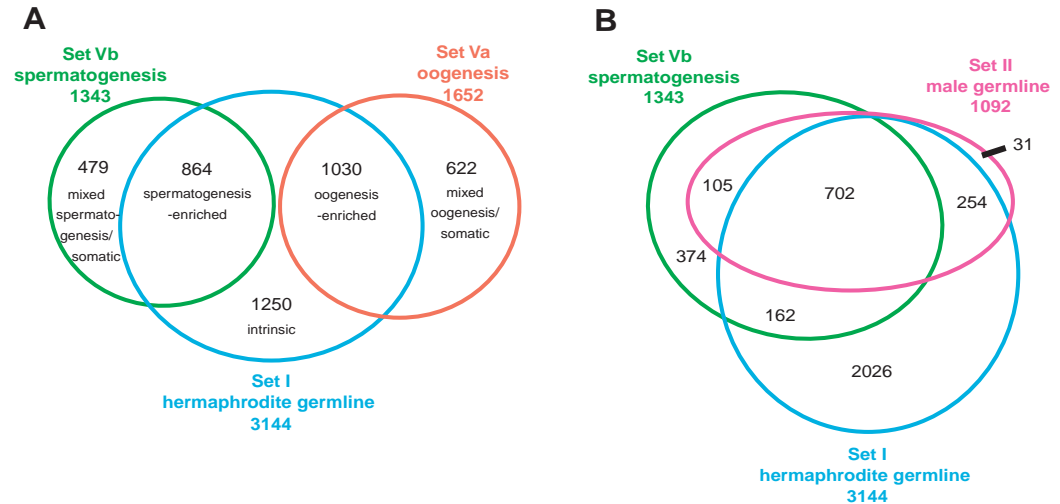
Fig. 1. Experimental design. Five sets of differentially expressed genes (I-V, triangles) were defined by indirect comparisons between males and hermaphrodites with and without a germline. For each set of genes, the wide, yellow end of the triangle abuts the sample with higher expression levels, while the blue tip of the triangle points to the sample with lower expression levels. The number of differentially expressed genes is listed within the triangle. For the N2 hermaphrodite and *glp-4* hermaphrodite comparison (set I), both L4 and adult stages were included. All samples were compared indirectly through use of a common reference, except *fem-1(lf)* was compared directly with *fem-3(gf)*.

inclusion in the defined gene set (see Materials and methods). These criteria were selected based on the expression profiles of genes that had been previously characterized as germline-enriched by microarray analysis (Reinke et al., 2000).

To identify genes with germline-enriched expression relative to somatic tissues, we compared either L4 or young adult wild-type hermaphrodites to *glp-4(bn2)* hermaphrodites of the corresponding stage. Hermaphrodites produce sperm as L4 larvae and switch to oogenesis as young adults, so we combined the L4 and adult gene sets to include genes expressed during both sperm and oocyte production. This set of comparisons defined 3144 genes with germline-enriched gene expression in hermaphrodites (set I, Fig. 1). Comparison of adult *him-5(e1490)* males to adult *glp-4(bn2);him-5(e1490)* males identified a total of 1092 genes with germline-enriched expression in males (set II, Fig. 1).

To define genes with differential expression between the two sexes, we compared wild-type (N2) hermaphrodite gene expression with that of *him-5(e1490)* males. This comparison identified 1935 genes with hermaphrodite-biased expression, and 1269 genes with male-biased expression (sets IIIa and IIIb). Finally, we compared *glp-4(bn2)* hermaphrodites and *glp-4(bn2);him-5(e1490)* males, both of which lack a germline, to define 460 genes with enriched expression in the hermaphrodite adult soma and conversely, 430 genes with enriched expression in male adult soma (sets IVa and IVb).

Fig. 2. Venn diagram analysis of genes with germline-enriched expression. Comparisons between major sets of genes define restricted subsets of germline-enriched expression. Numbers of genes within each subset are listed. (A) Combination of the spermatogenesis and oogenesis sets with the hermaphrodite germline-enriched set defines a subset of germline-intrinsic transcripts. (B) Combination of the male and hermaphrodite germline-enriched sets with the spermatogenesis set defines a subset of male germline-specific transcripts.



In an independent set of microarray experiments, we further characterized germline-enriched gene expression by determining which genes were differentially expressed during spermatogenesis and oogenesis in hermaphrodites. In three independent replicates, we directly compared gene expression levels in adult *fem-1(lf)* hermaphrodites, which make only oocytes, with adult *fem-3(gf)* hermaphrodites, which make only sperm (Nelson et al., 1978; Barton et al., 1987). This comparison identified 1652 genes with high levels of expression during oogenesis [high in *fem-1(lf)* relative to *fem-3(gf)*; set Va], and 1343 genes with high levels of expression during spermatogenesis [high in *fem-3(gf)* relative to *fem-1(lf)*; set Vb]. Genes with high expression during oogenesis would probably encode proteins required for oocyte differentiation as well as maternally provided factors necessary for proper development of the early embryo. Genes with high expression during spermatogenesis are likely to be involved in spermatocyte specification and differentiation.

Together the above experiments identify 5629 genes, ~29% of the protein-encoding genome, the expression of which is regulated by sexual identity and/or the presence of a germline. All of these data are available for either single or multiple gene queries at <http://wormgermline.yale.edu>, and the raw data are available at the Gene Expression Omnibus at NCBI and Yale Microarray Database (see Materials and methods).

Classification of hermaphrodite germline-enriched gene expression

We consider all genes with increased expression in wild type relative to *glp-4* (set I) and/or significantly different expression between *fem-1(lf)* and *fem-3(gf)* (set V) as the entire set of genes with hermaphrodite germline-enriched expression. To examine the relationship among these sets of genes, we performed a Venn diagram analysis (Fig. 2A). The combination of sets Va and Vb with set I define the total set of 4245 genes with germline-enriched expression. The intersection of Va or Vb with set I identifies those germline-enriched genes with increased expression in hermaphrodites producing only sperm or only oocytes. These gene sets are termed spermatogenesis-enriched and oogenesis-enriched, respectively. In our prior experiments, we categorized germline-enriched genes with no significant difference between spermatogenesis or oogenesis

(i.e. genes in set I that do not overlap Va or Vb) as ‘intrinsic’ (Reinke et al., 2000). Genes with germline-intrinsic expression are predicted to function in mitotic proliferation and early meiosis I in the distal germline, rather than in gametogenesis or embryogenesis. However, as the data presented below show, we cannot distinguish between the intrinsic and oogenesis sets by their temporal regulation, their in situ hybridization patterns, or the predicted functions of the encoded proteins. We therefore consider these two groups to have significant functional overlap, and thus some genes found in the oogenesis set are likely to have intrinsic functions, and vice versa.

A fraction of the *fem-3(gf)*-enriched or *fem-1(lf)*-enriched transcripts, encoded by 479 and 622 genes, respectively, did not meet the criteria for germline-enrichment (Fig. 2A). The *fem-3(gf)* and *fem-1(lf)* mutants make larger numbers of gametes than wild-type hermaphrodites; thus, many differentially expressed transcripts of low abundance are reliably detected in the *fem-3(gf)/fem-1(lf)* comparison, but not in the wild-type/*glp-4* comparison (Reinke et al., 2000). Additionally, *fem-1* is expressed in somatic tissues as well as the germline (Gaudet et al., 1996). Even though the *fem-1* mutation we used does not have a phenotypic effect in the soma of hermaphrodites, the expression of a subset of somatically expressed genes could still be affected in this background. Because some of the genes that appear differentially regulated in the *fem-3(gf)/fem-1(lf)* comparison may actually be somatically expressed, we term these subsets ‘mixed spermatogenesis/somatic’ and ‘mixed oogenesis/somatic’ (Fig. 2A). Additional evidence supporting this possibility is discussed below.

We examined the predicted molecular functions of the genes within the intrinsic, oogenesis and spermatogenesis gene sets using the existing gene ontology (GO) annotation for *C. elegans* because it is a structured annotation system and allows objective classification (Ashburner et al., 2000). Currently, ~20% of the genes in the nematode have an entry in the ‘molecular function’ category of GO (www.geneontology.org). We cross-referenced these molecular annotations with the intrinsic, oogenesis and spermatogenesis subsets to provide a preliminary annotation of these genes. To simplify inspection of the annotations, we combined related annotations into broader categories based on

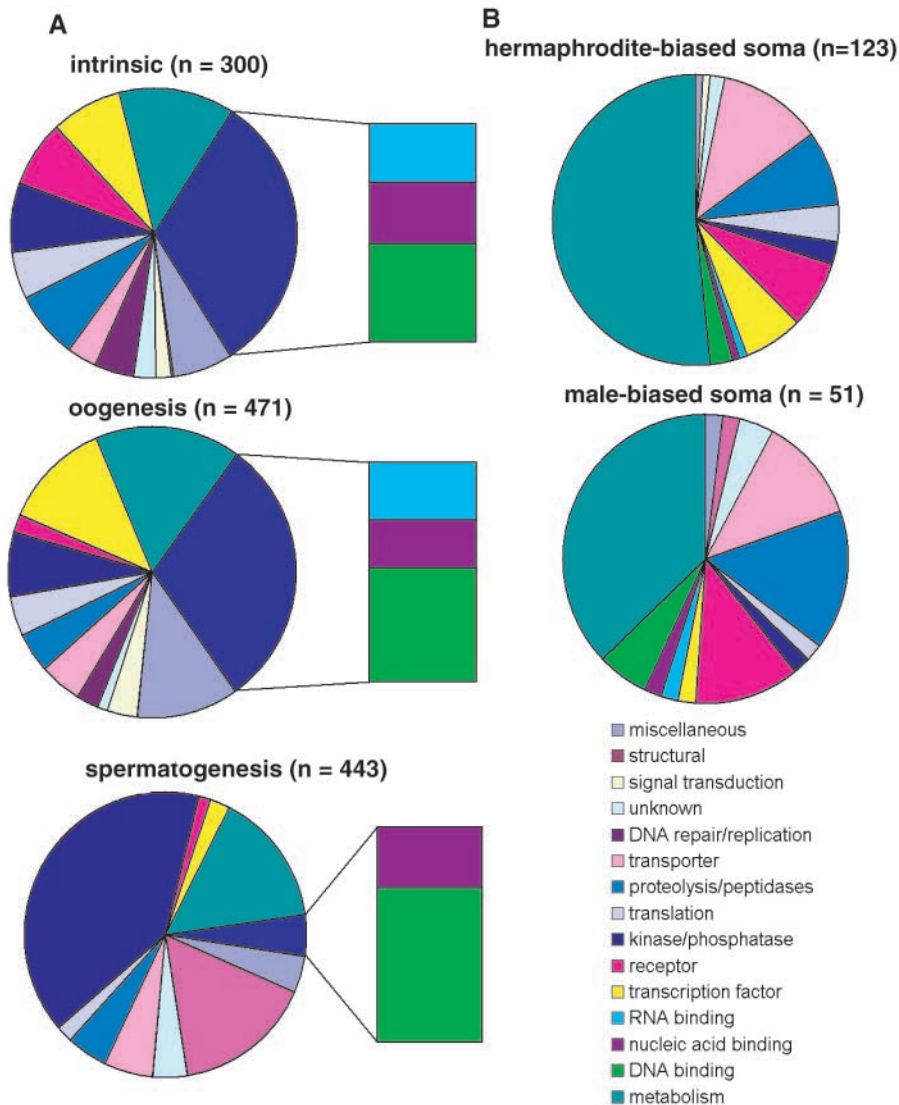


Fig. 3. Functional categories of germline- and sex-regulated genes. Pie charts show functional annotation of each of the major gene sets, based on the assigned molecular function using gene ontology (GO) annotation. Most genes with unknown function are not included in this analysis; the few included have a domain of unknown function that has been annotated by GO. (A) Functional categories of the germline-enriched gene sets. The sidebar divides genes encoding nucleic acid binding proteins into three categories: RNA binding, DNA binding or unspecified nucleic acid binding. (B) Functional categories of sex-biased, somatic-expressed gene sets.

the 'GO slim' ontology for *Drosophila* (www.flybase.org). The results are presented as pie charts in Fig. 3A. Most notably, the oogenesis and intrinsic gene sets have genes in the same categories at approximately the same ratios. For example, genes encoding predicted nucleic acid-binding proteins comprise 31% and 32% of the intrinsic and oogenesis gene sets, respectively. In both sets, 27% of the nucleic acid binding proteins are predicted to bind RNA specifically (aqua box in bar). By contrast, only 4% of genes in the spermatogenesis set encode predicted nucleic acid-binding proteins. Of these, fewer than 1% encode RNA-binding proteins, probably because mature sperm have extremely low levels of mRNA (Roberts et al., 1986). Instead, the spermatogenesis set has an enrichment of cytoplasmic signaling molecules such as protein kinases and protein phosphatases, as noted previously (Reinke et al., 2000).

Identification of genes expressed in the male germline

Most events in male germ cell development also occur in hermaphrodite germ cells, such as mitotic proliferation,

recombination, chromosome segregation and spermatogenesis. We therefore expected a large overlap between male and hermaphrodite germline-enriched gene sets. To determine which of the male germline-enriched genes (set II) were also enriched in the hermaphrodite germline, we performed a Venn diagram analysis with three datasets: the male and hermaphrodite germline-enriched gene sets and the spermatogenesis gene set (sets I, II and Vb; Fig. 2B). We found that 87% of genes with male germline-enriched expression are also enriched in the hermaphrodite germline. Of these 956 shared genes, 702 show significantly enriched expression during spermatogenesis, while 254 do not show spermatogenesis-enrichment and are therefore likely to be involved in other aspects of germline development shared between the two sexes, such as mitotic proliferation. Another 105 male germline-enriched genes overlap with the mixed spermatogenesis/somatic subset, and thus likely also represent shared spermatogenesis-enriched genes.

The remaining 31 genes with germline-enriched expression in males, but not hermaphrodites, are candidates for the molecular basis of two known male germline-specific characteristics: heterochromatization of the X chromosome (Kelly et al., 2002), and the ability of male sperm to out-compete hermaphrodite sperm for fertilization of oocytes (LaMunyon and Ward, 1998). Under the strictest definition, a truly male-specific germline gene should show no evidence of expression in hermaphrodites. We excluded any genes with hermaphrodite expression by any of several criteria: (1) a high *fem-3(gf)/fem-1(lf)* ratio, (2) significant fluctuation in either of two hermaphrodite timecourses in larvae or embryos (see below) (Baugh et al., 2003), (3) hermaphrodite expression by in situ hybridization (Kohara, 2001), or (4) an RNAi phenotype in hermaphrodites (e.g. Kamath et al., 2003). The remaining eight genes include several that encode novel proteins as well as one that contains a predicted MSP (major sperm protein) domain found in structural and signaling sperm proteins, and two that potentially bind DNA (Fig. 4).

Identification of genes with sex-biased expression in somatic tissues

We examined the gene sets that correspond to sex-biased, somatically expressed genes (sets IVa and IVb). Because we compared the somatic tissues of adult hermaphrodites and males, we primarily identified genes that correspond to terminal morphological differences between the two sexes. Categorization of the molecular functions of these genes is presented in Fig. 3B. Named genes among the hermaphrodite soma-biased gene set include *lin-2*, *sir-2.2*, *egl-13* and the genes encoding vitellogenin. Among the most highly male soma-biased genes are several that are similar to *Tpx1*, a vertebrate gene required for male germ cell interaction with surrounding somatic tissue in vertebrates (Giese et al., 2002). We also found 12 genes encoding neuropeptide-like proteins (*nlp-1-3*, *-12*, *-14*, *-25*, *-31* and *flp-3*, *-6*, *-8*, *-9*) with male-biased expression that could potentially be involved in mediating male-specific behaviors. Among the expected male-specific somatic proteins, *pkd-2* and *her-1* had significant male-biased expression. Additional known male-specific proteins such as *lov-1* and *mab-3* had mild enrichment but did not meet our statistical criteria, while others were not present on the array (e.g. *mab-23*).

Comparison with independent large-scale gene expression datasets

To validate our gene subsets for the hermaphrodite germline, we compared our results with other large-scale gene expression datasets. Our experiments measure the abundance of transcripts in one sample relative to another. Therefore, we will not identify germline-expressed genes with only mild or no enrichment in the germline, relative to somatic tissues. Additionally, the criteria we set for inclusion in a gene set are stringent, so genes with mild germline-enriched expression in our experiments are excluded. Comparisons with gene expression data generated by in situ hybridization or Affymetrix microarray analysis allow us to estimate both the number of genes we are missing, and the number of genes that are incorrectly included in our germline datasets.

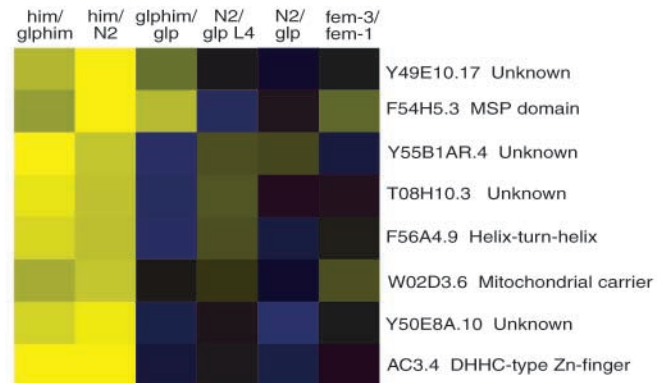


Fig. 4. Male germline-specific genes. Shown are relative levels of gene expression of several indirect comparisons between genotypes for eight genes with significantly male germline-enriched expression, but no evidence of hermaphrodite-enriched expression. him, *him-5* males; glp, *glp-4;him-5* males; N2, wild-type hermaphrodites; glp, *glp-4* hermaphrodites; fem-3, *fem-3(gf)* hermaphrodites; fem-1, *fem-1(lf)* hermaphrodites. All comparisons shown were made between adult animals, except for column 4. Yellow represents increased expression in the numerator of each comparison, while blue represents increased expression in the denominator of each comparison.

First, we examined whole-mount in situ expression patterns in hermaphrodites in NextDB (<http://nematode.lab.nig.ac.jp>) (Kohara, 2001) for randomly chosen genes in each subset (Table 1). Considering only genes for which expression was detected by in situ hybridization, this comparison showed that 98% of the genes in the intrinsic and oogenesis-enriched groups had detectable in situ staining solely or primarily in the germline. Additionally, 88% of spermatogenesis-enriched transcripts were detected either in the germline or spermatheca of the animal. Overall, the in situ data indicates that most of our germline datasets have a low false positive rate, as few genes that had germline-enriched expression by microarray were found solely in the soma by in situ hybridization. However, we note that the sensitivity of the large-scale in situ

Table 1. Overlap with independent gene expression data sets

Category	I Number examined	II % with associated ESTs (n)	III % with in situ pattern (n)	IV % germline only* (n)	V % germline+ somatic (n)	VI % somatic only (n)	VII % overlap with Baugh et al.† (n)
Hermaphrodites							
Intrinsic (n=1250)	252	65 (165)	52 (130)	98 (125)	1 (3)	1 (2)	71 (890)
Oogenesis-enriched (n=1030)	256	85 (218)	67 (172)	98 (169)	1 (2)	1 (1)	88 (904)
Mixed oogenesis/somatic (n=622)	137	76 (105)	54 (75)	91 (68)	4 (3)	5 (4)	72 (452)
Spermatogenesis enriched (n=864)	290	42 (123)	26 (75)	88 (66)	4 (3)	8 (6)	9 (79)
Mixed spermatogenesis/somatic (n=479)	250	37 (92)	22 (55)	63 (35)	2 (1)	34 (19)	16 (76)
Soma enriched (n=460)	145	70 (101)	35 (51)	20 (10)	2 (1)	78 (40)	14 (66)
Males							
Germline enriched (n=31)	31	29 (9)	16 (5)	60 (3)	20 (1)	20 (1)	19 (6)
Soma enriched (n=430)	127	45 (58)	18 (23)	26 (6)	13 (3)	53 (14)	15 (50)

n, number used to calculate percentages. For calculation of percentages in columns II and III, (N) was divided by column I and multiplied by 100%. For calculation of percentages in columns IV, V and VI, (n) was divided by (n) of column III and multiplied by 100%.

*Genes with staining at regions corresponding to spermathecae were considered germline (this pattern was common among spermatogenesis genes and almost absent among other groups).

†Baugh et al., 2003.

hybridization is unknown, so it is possible that some genes with detectable expression only in the germline are also expressed in the soma.

In contrast to all other germline-enriched gene sets, 34% of the genes in the mixed spermatogenesis/somatic gene subset (see Fig. 2A) show only somatic expression by in situ hybridization. This finding is consistent with the possibility discussed above that some of the genes in this subset might be those whose expression is affected by the *fem-1* mutation in somatic tissues.

In this analysis, we found that our gene expression subsets had an unequal distribution of associated ESTs and detectable in situ staining. Approximately 60% of all annotated genes have a corresponding EST. Within most of our hermaphrodite gene expression subsets, an average of 74% of the genes were represented by ESTs; of these, 73% had visible staining (54% of subset). However, only 40% of genes with spermatogenesis-enriched or mixed spermatogenesis/somatic expression had associated ESTs; of these, 60% displayed visible staining (24% of subset). Thus, genes expressed during spermatogenesis are less likely to be represented by ESTs than are genes expressed in the oogenic germline. We note that the low numbers seen for genes with male-biased expression are not surprising, because the EST and in situ projects focused on hermaphrodites.

The majority of genes with hermaphrodite soma-biased expression (78%) had staining in specific somatic tissues, such as the intestine, vulva and body wall muscle, as well as broad staining that was difficult to attribute to a specific tissue(s). About 20% of the examined transcripts from the hermaphrodite-biased soma-enriched subset stained the germline, indicating that we missed about 20% of genes with detectable expression in the germline with our experimental design and statistical criteria.

We also compared our gene subsets to gene expression profiles of early embryogenesis, collected using an Affymetrix microarray platform (Baugh et al., 2003). By examining embryos prior to the onset of zygotic transcription, this study identified a large set of transcripts that are maternally provided, which (by definition) must be expressed in the adult germline. As expected, genes with intrinsic or oogenesis-enriched expression had very high overlap with the maternally provided gene set, ranging from 71-88% (Table 1). Genes enriched during spermatogenesis, or in somatic tissues, had a lower overlap of 9-19%. Our comparisons to both the in situ dataset and the maternally provided dataset allows us to estimate that we are missing ~20% of genes expressed in the germline.

Temporal regulation during hermaphrodite development

In a second set of microarray experiments, we performed a temporal analysis of gene expression during wild-type hermaphrodite development. This analysis allows us to examine the normal kinetics of gene expression of many of the genes defined in our mutant sets. We collected 12 samples at 3-hour intervals, beginning in the middle of the third larval stage (L3) and extending through adulthood (Fig. 5A). During this time, germ cells initiate several events, including exiting mitosis and initiating meiosis, differentiating into sperm then oocytes, and launching embryogenesis. Formation of several somatic gonad structures such as the vulva, the spermatheca

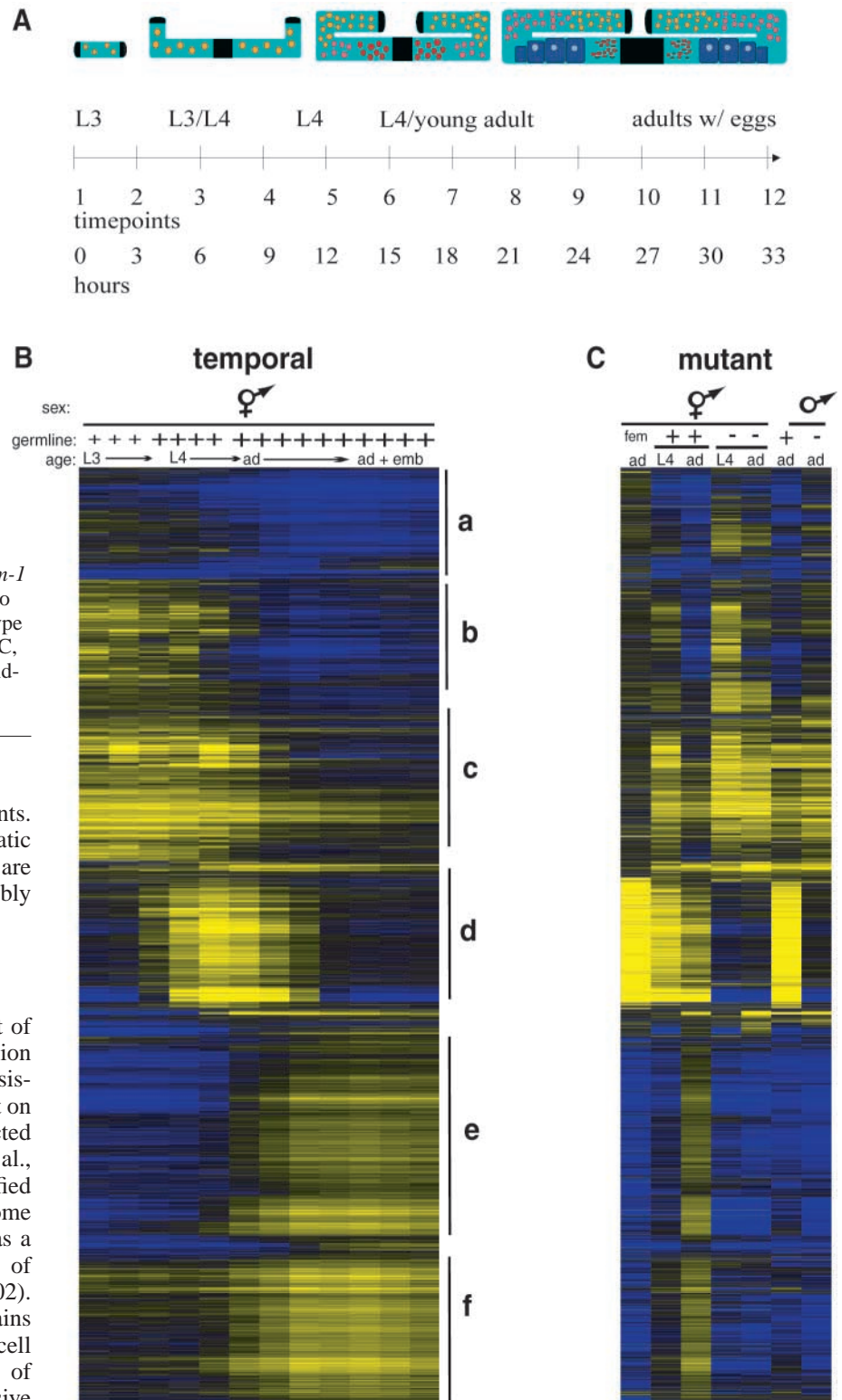
and the uterus also occurs during this time. We collected three series of staged hermaphrodites and performed 36 hybridizations against the same mixed stage reference sample used in the mutant hybridizations. For each gene, we averaged the three replicates at each time point, and used ANOVA to identify 5083 genes that showed a significant alteration in gene expression levels between two or more time points ($P < 0.01$). Of these, 2925 are germline- or sex-regulated genes.

We used hierarchical clustering to group all 5083 genes solely by the similarity in their temporal expression profiles, and defined six large clusters (A-F) with distinct patterns of expression over time, as defined by a correlation coefficient exceeding 0.80 within a cluster (Fig. 5B) (Eisen et al., 1998). So that these genes would be grouped only by their temporal expression profiles, we 'carried' the mutant expression data in the analysis and did not give it any weight in the clustering (Fig. 5C). We then asked whether specific sets of germline- or sex-regulated genes were over-represented in these clusters that were defined solely by temporal regulation. We found that 98% of the genes in the intrinsic and oogenesis-enriched subsets included in the analysis comprise ~84% of, and are evenly distributed among, clusters E and F. The remaining 16% are likely also expressed in the germline, based on the striking similarity in temporal expression. Clusters E and F have largely similar expression profiles (correlation coefficient of 0.74), with a few subtle differences. Genes in Cluster E first display very low levels of expression from mid-L3 to the end of L4, and then show an abrupt increase at the transition to young adulthood (time points 6 and 7), with high levels persisting through the rest of the time course, while genes in Cluster F have higher levels of expression at the earlier time points and show a gradual increase starting slightly earlier (timepoints 5 and 6).

In contrast to the intrinsic and oogenesis gene sets, 99% of spermatogenesis-enriched genes (as defined in Fig. 2A) are found in cluster D, which is characterized by sharp induction at the beginning of L4 (time point 3) and a sharp decline at the end of L4 (time points 6 and 7). The reciprocal relationship of the expression of the spermatogenesis-enriched and oogenesis-enriched/intrinsic transcripts at the onset of adulthood reflects the switch from spermatogenesis to oogenesis that occurs at this time. Strikingly, only 26% of genes in the mixed spermatogenesis/somatic gene set (see Fig. 2A) are present in cluster D, with the remaining 74% distributed among the genes in clusters A-C (Fig. 5B). The fact that most genes in the mixed spermatogenesis/somatic set differ in temporal regulation from the spermatogenesis-enriched set supports the possibility that the mixed spermatogenesis/somatic set contains many genes that are likely to be differentially expressed in somatic tissues in response to the effects of the *fem-1(lf)* mutation. By contrast, 86% of the genes in the mixed oogenesis/somatic subset are found in the E and F clusters with the intrinsic and oogenesis-enriched gene subsets, and thus are largely co-regulated with those sets.

In general, genes with sex-biased somatic expression were evenly distributed among clusters A-C, along with most remaining temporally regulated genes that showed no sex- or germline-regulated expression. Cluster A includes genes with a moderate level of expression at the L3 stage, which decreases at later stages, while Clusters B and C contain genes with a high level of expression in L3 and L4 larvae that decreases at

Fig. 5. Temporal analysis of wild-type larval and adult gene expression. (A) Diagram of germline development and timepoints taken for analysis. Black, somatic gonad; orange, proliferating germ cells; pink, meiotic germ cells; red, differentiating spermatocytes; blue, differentiating oocytes. (B) Temporal expression profiles for all genes with $P < 0.01$ (ANOVA) in timecourse, organized by similarity in expression using hierarchical clustering. Each row represents a gene, while each column represents an average of each set of comparisons. The \log_2 ratios have been replaced with color that demonstrates the relative level of expression. Six clusters, a-f, with different temporal patterns are defined by bars down right side of clustergram. (C) The mutant expression profiles of the different genotypes of temporally regulated genes. Genes are present in the same order as B. The mutant data had no weight in the clustering, but were included to demonstrate the mutant profile of each genotype. The 'fem' column (column 1) corresponds to the *fem-3/fem-1* direct comparison. The + and – notations refer to the presence or absence of the germline (wild type or *gfp-4* mutants, respectively). For both B and C, yellow indicates higher expression in staged wild-type or mutant sample; blue indicates higher expression in reference sample or *fem-1(lf)*.



the onset of adulthood to varying extents. Similarity in temporal regulation of somatic genes will help to determine which genes are likely to be co-expressed and possibly function together.

Biased chromosomal distribution of germline- and sex-regulated genes

Our previous study defining the partial set of genes with germline-enriched expression demonstrated that many spermatogenesis-enriched and intrinsic genes are not present on the X chromosome at the numbers expected given a random distribution (Reinke et al., 2000). Subsequent studies have identified transcriptional silencing of the X chromosome in the male and hermaphrodite germline as a potential force prohibiting X-linkage of germline-expressed genes (Kelly et al., 2002). In the male, the single X chromosome remains silent at all stages of male germ cell development, and displays a hallmark of heterochromatin formation, extensive methylation of lysine 9 on histone H3 (Nakayama et al., 2001; Rea et al., 2000). However, in oogenic hermaphrodites, the pair of X chromosomes display only transient and partial H3 lysine 9 methylation and is silenced only in mitotically dividing and meiotic germ cells through the pachytene stage of meiosis I before becoming transcriptionally active late in pachytene, just prior to diplotene and diakinesis (Kelly et al., 2002). The entire

genome is apparently silenced in meiotic metaphase I just prior to fertilization.

We examined the chromosomal distribution of our genome-wide set of germline-enriched and sex-biased somatic genes and, as before, found that genes in the spermatogenesis and intrinsic sets are greatly under-represented on the X

Fig. 6. Chromosomal distribution of germline- and sex-regulated genes. The ratio of genes in each gene set observed on each chromosome, relative to the expected number for the number of genes per chromosome on the microarray, was plotted. Any statistically significant deviation of observed from expected is marked by an asterisk ($P < 0.001$; hypergeometric probability test). See <http://wormgermline.yale.edu> for corresponding table with numbers of each gene set on each chromosome.

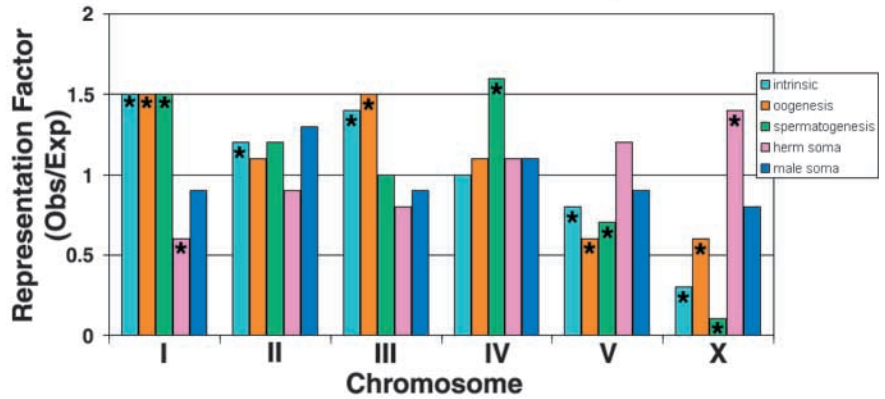
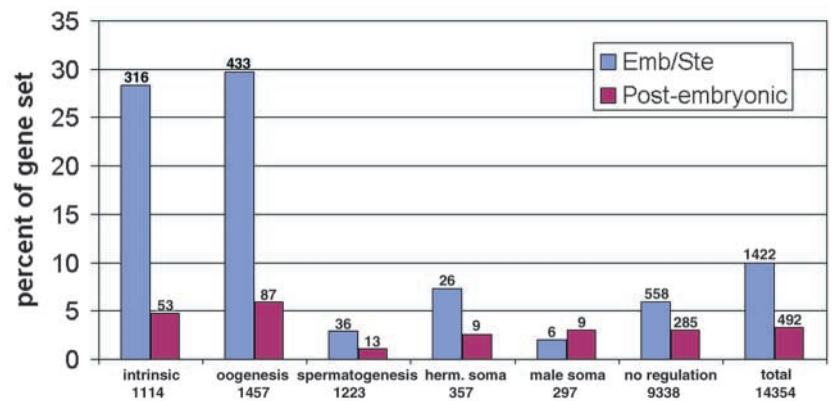


Fig. 7. RNAi phenotypes of germline- and sex-regulated genes. The fraction of genes in each dataset that display a phenotype when functionally depleted by RNAi are graphed. Above each bar in the graph is the number of genes with an RNAi phenotype. Below the name of each dataset is the total number of genes within that set. Genes with no significant regulation by sex or the presence of a germline are included in the 'no regulation' category, while all genes with RNAi assays performed and listed in Wormbase (www.wormbase.org) are counted in the 'total' category.



chromosome (Fig. 6). Approximately 196 genes in the spermatogenesis set were expected on the X chromosome given the number of X-linked genes on the microarray, and we only found 25 ($P < 0.001$). Similarly, 180 genes with intrinsic expression were expected on the X chromosome, but only 47 are present ($P < 0.001$). The increased numbers in the genome-wide data set allowed us to determine that genes with oogenesis-enriched expression are also significantly under-represented on the X chromosome, although to a lesser extent than the spermatogenesis-enriched and intrinsic genes; we expected 241 genes in the oogenesis subset to be X-linked, and found 138 ($P < 0.001$; see Discussion). Conversely, genes encoding hermaphrodite-biased somatically expressed transcripts are significantly enriched on the X chromosome. A recent report on X chromosome distribution of genes with male-biased expression in *Drosophila* demonstrated that genes expressed in both the soma and germline of the male were depleted from the X chromosome (Parisi et al., 2003). In contrast to these observations in *Drosophila*, male-biased somatically expressed genes in *C. elegans* are present at close to expected numbers on the X chromosome, and only genes expressed in the male germline of *C. elegans* are strongly depleted from the X chromosome, as described above (see Discussion).

In addition to the major effect on the X chromosome, we found that the number of germline-enriched genes on several autosomes was significantly different from a random distribution. For example, all germline-enriched gene subsets were significantly enriched on chromosome I and depleted

from chromosome V. Additionally, genes in the intrinsic and oogenesis groups are enriched on chromosome III, while genes in the spermatogenesis category are over-represented on chromosome IV. Several tightly clustered groups of genes encoding major sperm proteins are present on IV that contribute to this enrichment.

RNA-mediated interference phenotypes of germline- and sex-regulated genes

The vast majority of the genes in the *C. elegans* genome have been assayed for reduction-of-function phenotypes in hermaphrodites by large-scale RNA-mediated interference screens (Piano et al., 2002; Maeda et al., 2001; Kamath et al., 2003; Gonczy et al., 2000). The combined results of these large-scale screens have identified visible phenotypes for ~13% of assayed genes, with almost 10% showing either an embryonic lethal or sterile phenotype. One study focusing on a set of 751 genes expressed primarily in oogenic germlines found that 322 (42%) produced either an embryonic lethal or sterile phenotype upon functional depletion by RNAi, demonstrating that these phenotypes are enriched among germline-expressed genes (Piano et al., 2002).

We compared the genes with germline-enriched expression that we identified in our microarray analysis to the combined results from these large-scale RNAi screens, focusing on embryonic lethality and sterility (Fig. 7). Of the genes within the intrinsic and oogenesis gene sets that had been tested by RNAi, functional depletion of ~28% of the genes results in either or both of these phenotypes. This number is lower than

the 42% mentioned above, probably because of differing false negative rates among the various studies (Piano et al., 2002). Functional disruption of <3% of the genes in the spermatogenesis set causes embryonic lethality or sterility. This number is a significant underestimate of the genes whose function is required for spermatogenesis, as RNAi is inefficient at reproducing phenotypes of known mutants in spermatogenesis-expressed genes (S.W., unpublished).

Among genes with sex-biased expression in the soma, genes with hermaphrodite-biased expression gave rise to lethal or sterile phenotypes 7% of the time, whereas fewer than 2% of the genes with male-biased expression displayed these phenotypes when functionally depleted in hermaphrodites. Of the ~11,000 genes that are not significantly differentially regulated by either sex or germline status, 6% showed embryonic lethality or sterility. Embryonic lethal phenotypes can occur upon functional depletion of zygotically expressed genes that are not expressed in the germline. Similarly, a sterile phenotype can result upon depletion of critical genes expressed in the somatic gonad or somatic reproductive organs.

We also examined a diverse group of phenotypes that affect post-embryonic development without affecting fecundity [i.e. *Vpep* + *Gro*, as defined by Kamath et al. (Kamath et al., 2003)]. We found that genes in the intrinsic and oogenesis data sets display post-embryonic phenotypes more frequently than other gene sets, including the hermaphrodite-biased somatic set. Genes in the intrinsic and oogenesis sets that display post-embryonic phenotypes upon RNAi include multiple genes with known somatic functions, such as *egl-27*, *nhr-23*, *daf-18* PTEN, *lin-35* Rb and *ptp-2* (Ch'ng and Kenyon, 1999; Kostrouchova et al., 2001; Mihaylova et al., 1999; Lu and Horvitz, 1998; Gutch et al., 1998). This observation suggests that a subset of the genes expressed primarily in the germline also function in other tissues at other places and times. Because incomplete RNAi can result in partial depletion of a gene product, many of these post-embryonic phenotypes might be caused by decreased levels of a gene product that causes embryonic lethality or sterility when completely depleted.

Discussion

Using genome-wide DNA microarrays, we have assayed the currently annotated genes to identify global gene expression profiles for the germline and somatic tissues of both hermaphrodite and male *C. elegans*, and defined subsets of germline- and sex-regulated genes. Additionally, we have examined the temporal regulation of many of these genes as wild-type hermaphrodite larvae develop into reproductively mature adults. These profiles provide a comprehensive overview of mRNA abundance during germline development and define somatic terminal differences between sexes. The identification of germline- and sex-regulated genes will facilitate both single-gene and large-scale functional studies in the future. Additionally, this global overview shows striking biases in how these genes are distributed in the genome, the classes of proteins they encode and the phenotypes that are produced upon functional depletion.

Characterization of germline-enriched genes

The germline-enriched gene set has been resolved into three subsets, which roughly correspond to different germline

functions: spermatogenesis, oogenesis and intrinsic. This last group is defined by a lack of significant regulation during gametogenesis, and was originally thought to correspond to distal germline functions, such as stem cell proliferation and early meiosis I (Reinke et al., 2000). However, several observations made in this study point to the conclusion that genes in the intrinsic and oogenesis-enriched subsets are highly similar and not easily functionally distinguishable. First, these genes encode similar types and proportions of proteins with predicted or known functions (Fig. 3A). Second, the temporal regulation of these genes is essentially identical, with the vast majority of the genes showing an abrupt induction at the onset of young adulthood (clusters E and F; Fig. 5B). Third, the range and percentage of RNAi phenotypes is comparable between these two categories (Fig. 7). Fourth, many genes among the oogenesis gene set show detectable expression in the distal germline by *in situ* hybridization. Why then can we distinguish these two gene sets by microarray analysis? One possibility is based on our observation that the *fem-3* mutant makes excess sperm at the expense of more distal germ cells, reducing the number of cells in pachytene of meiosis I relative to *fem-1* mutants. Thus, many genes that are expressed in distal germ cells ('intrinsic') actually appear enriched in *fem-1* relative to *fem-3* and are classified as 'oogenesis-enriched'. From the above set of observations, we conclude that many genes with 'intrinsic' function are likely to be found in the oogenesis category, and vice versa.

The single distinguishing characteristic we have observed between the intrinsic and oogenesis sets is that genes in the oogenesis set are found on the X chromosome at a slightly higher frequency than genes in the intrinsic set, although they are still present at much lower levels than expected (Fig. 6). Genes expressed during early oogenesis are probably not as strongly excluded from the X chromosome because the X chromosome becomes re-activated as germ cells enter oogenesis (Kelly et al., 2002). This observation suggests that although the oogenesis category contains many genes with intrinsic function, it also contains genes that are specifically expressed at a later point in the germline, relative to the intrinsic category. Indeed, some autosomal and X-linked genes in the oogenesis category have detectable *in situ* hybridization signals only in the proximal germline (<http://nematode.lab.nig.ac.jp>) (Kelly et al., 2002).

In our analysis of *in situ* hybridization patterns presented in NextDB for the germline- and sex-regulated genes in hermaphrodites, we found that more genes in the spermatogenesis set fail to have a corresponding EST than do hermaphrodite-biased somatic genes or intrinsic/oogenesis genes. One possible explanation for this observation is that genes required for spermatogenesis are expressed only briefly during the development of the animal and are therefore likely to be under-represented in a cDNA library. However, hermaphrodite-biased, somatically expressed genes, many of which are also expressed in a restricted area and for a brief period of time (e.g. during vulval induction), have a corresponding EST as frequently as genes in the intrinsic or oogenesis sets. Interestingly, the fraction of ESTs that give visible *in situ* expression patterns is roughly equivalent between hermaphrodite-biased somatic and spermatogenesis gene sets (50-56%), but is considerably lower than for the intrinsic/oogenesis gene sets (71-79%). This observation

suggests that genes with highly restricted spatial or temporal expression tend not to be detected by in situ hybridization, or that genes with expression in the oogenic germline are more easily detectable than other tissues.

Comparison between male-biased and hermaphrodite-biased germline-enriched gene sets defined a small subset of genes whose expression appears largely restricted to the male germline. These genes are candidates for male germline-specific functions, such as heterochromatinization of the X chromosome and effective sperm competition. Interestingly, among the eight genes that show no apparent expression in hermaphrodites, one contains an MSP domain commonly found among structural sperm proteins, as well as additional domains of unknown function. Future functional studies using RNAi or deletion mutant analysis to investigate the role of these proteins in male germline development and function will shed light on how the male germline performs its unique roles.

Temporal analysis of germline- and sex-regulated gene expression

We analyzed high-resolution temporal gene expression profiles during the stages that encompass most of germline development: in L3 and L4 larvae, and in pre- and post-reproductively mature adults. When we clustered genes based solely on their temporal expression profiles, we found that subsets of the germline genes showed very similar co-regulation. Genes in the spermatogenesis set show expression profiles that correspond with spermatogenesis in the fourth larval stage, as expected (Reinke et al., 2000). Surprisingly, even with a fairly high-resolution sampling, we were able to distinguish only two other germline-enriched clusters (clusters E and F), each containing a mixture of genes in the intrinsic and oogenesis sets. Not only do the genes within each cluster show a strong degree of correlated expression (>0.85), but the correlation between clusters E and F is also high (0.74). These temporal gene expression profiles indicate that essentially all gene expression in the germline is tightly temporally controlled, not just the genes with spermatogenesis-enriched expression. In addition to providing profiles of germline-enriched gene expression, these data also provide temporal profiles for genes expressed during somatic events that occur during late larval and early adult stages, including formation of several structures of the somatic gonad.

The timecourse data also provide supporting evidence that some genes differentially regulated in the *fem-3(gf)/fem-1(lf)* comparison are actually somatically expressed, because a subset of genes with enrichment only in *fem-3(gf)* animals ('mixed spermatogenesis/somatic' subset, see Fig. 2B) do not cluster with all the other spermatogenesis genes in cluster D. For example, among this subset is the family of Mariner transposases that have been found to have sperm-enriched expression (Reinke et al., 2000; Kim et al., 2001). However, transposons are known to be silenced in the germline (Emmons and Yesner, 1984), so the expression of transposases, which excise and mobilize transposons, during spermatogenesis was surprising. The data presented in this report demonstrate that although these transposases show *fem-3*-enriched expression, their transcripts are not germline-enriched in wild type relative to *glp-4(lf)* animals and they do not cluster with the other spermatogenesis-enriched transcripts. Instead, these transposases either show no temporal regulation or are found in cluster A. Thus, it is very

likely that these transposases are not expressed during spermatogenesis, but show increased activity in somatic tissue upon loss of the activity of the *fem-1* gene product.

Chromosomal biases of germline- and sex-regulated genes

Our data show that the chromosomal locations of the genes with enriched expression in the germline are non-random. We found a very strong bias against genes in the spermatogenesis and intrinsic sets residing on the X chromosome, as seen before (Reinke et al., 2000). Additionally, with our expanded data set, we were also able to detect a significant reduction in the number of observed X-linked genes in the oogenesis category compared with the expected number. Our data also indicate that the location of hermaphrodite-biased, somatically expressed genes is reciprocal with the germline-enriched gene sets: whereas genes with germline-enriched expression are enriched on chromosome I and lacking on the X chromosome, hermaphrodite-biased, somatically expressed genes are enriched on the X chromosome and lacking on chromosome I. Notably, this reciprocal relationship is not limited to chromosomes I and X; a similar trend is also seen for chromosomes II, III and V.

Recently, several investigations of gene expression in different organisms have revealed sex chromosome biases for genes expressed in the germline. These biases differ between organisms: in mice, genes expressed in male spermatogonia are concentrated on the X chromosome, while in both *Drosophila* and *C. elegans*, genes expressed during spermatogenesis are found on the X chromosome much less frequently than expected (Wang et al., 2001; Parisi et al., 2003; Reinke et al., 2000) (this work). However, genes with male-biased expression in somatic tissues in *Drosophila* are also under-represented on the X chromosome, whereas male-biased somatically expressed genes in *C. elegans* are not. In *C. elegans*, the silencing of the X chromosome in the germline provides an excellent candidate for the mechanism behind under-representation of germline genes on the X chromosome (Kelly et al., 2002). By contrast, in *Drosophila*, several lines of evidence indicate that the X chromosome remains transcriptionally active during spermatogenesis (reviewed by McKee and Handel, 1993); therefore the forces preventing genes with male-biased expression from staying on the X chromosome must differ from those in *C. elegans*. Genes with male-biased expression in *Drosophila* show a strong tendency to move off the X chromosome, based on comparisons to the distantly related mosquito, *Anopheles gambiae* (Parisi et al., 2003). It will be interesting to perform a similar analysis of genes with spermatogenesis-enriched expression in *C. elegans*, once the genome sequence of the related nematode *C. briggsae* is fully assembled and chromosomes are assigned.

When we compared our gene sets with existing RNAi phenotypic data, we found that the intrinsic and oogenesis gene sets contain a high percentage of genes that result in either embryonic lethality or sterility when functionally depleted, as observed previously (Piano et al., 2002). Large-scale functional studies have demonstrated that RNAi of X-linked genes results in embryonic lethal or sterile phenotypes much less frequently than expected (Piano et al., 2002; Kamath et al., 2003). Much of this observation can therefore be attributed to the fact that genes with germline-enriched

expression are excluded from the X chromosome and that the X chromosome is silenced throughout much of the germline, as discussed above. However, even taking into account the reduced number of germline-enriched genes on the X chromosome, embryonic lethal and sterile phenotypes are still much rarer than expected (Piano et al., 2002). One possible explanation for this observation is that the brief window of expression of X-linked genes in the germline reduces the ability of RNAi to effectively functionally deplete them. Another possibility is that genes on the X chromosome are simply less likely to encode proteins required for viability and fecundity (Piano et al., 2002). This second possibility is supported by the fact that post-embryonic phenotypes are found more frequently among X-linked genes (Kamath et al., 2003). However, we did not see an enrichment of post-embryonic phenotypes among the hermaphrodite-biased, somatically expressed genes, which are enriched on the X chromosome (Figs 6 and 7).

Conclusion

Gene expression profiling in developmental model organisms such as *C. elegans* provides valuable information about how gene expression is regulated during metazoan development. Genome-scale expression studies complement forward and reverse genetic screens, because genes that do not give rise to a specific phenotype upon mutation or functional depletion can now be associated with a specific biological process. Additionally, combining data from multiple independent large-scale functional analyses, including expression profiling, strengthens functional predictions that can be made.

Our work has identified the vast majority of genes with expression in the germline of *C. elegans*. The identification and characterization of cis-regulatory elements in the noncoding regions surrounding these genes will allow us to better understand the gene-specific and global regulatory mechanisms that govern gene expression in the germline. In the future, computational analysis, in conjunction with experiments using genomic DNA microarrays to investigate the interactions of regulatory proteins with cis-regulatory elements, should shed light on these still-mysterious mechanisms.

The authors thank Kevin White, Kris Gunsalus, and members of the Reinke lab for critical reading of the manuscript. We also thank Jeremy Nance and Elizabeth Davis for preparation of the *fem-1* and *fem-3* RNA samples, and Stuart Kim and members of the Kim lab for assistance in preparation of some of the microarrays used in this work. We are especially grateful to Yuji Kohara and others who performed the large-scale *in situ* hybridization project (NextDB; <http://nematode.lab.nig.ac.jp>), whose data was extremely valuable in analyzing our own. This work was supported by grants GM065682 (V.R.) and GM25243 (S.W.) from the NIH.

References

- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T. et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25-29.
- Barton, M. K., Schedl, T. and Kimble, J. (1987). Gain-of-function mutations of *fem-3*, a sex-determination gene in *Caenorhabditis elegans*. *Genetics* **115**, 107-119.
- Baugh, L. R., Hill, A. A., Slonim, D. K., Brown, E. L. and Hunter, C. P. (2003). Composition and dynamics of the *Caenorhabditis elegans* early embryonic transcriptome. *Development* **130**, 889-900.
- Beanan, M. and Strome, S. (1992). Characterization of a germline proliferation mutation in *C. elegans*. *Development* **116**, 755-766.
- Broverman, S. A. and Meneely, P. M. (1994). Meiotic mutants that cause a polar decrease in recombination on the X chromosome in *Caenorhabditis elegans*. *Genetics* **136**, 119-127.
- Ch'ng, Q. and Kenyon, C. (1999). *egl-27* generates anteroposterior patterns of cell fusion in *C. elegans* by regulating Hox gene expression and Hox protein function. *Development* **126**, 3303-3312.
- Colaiaçovo, M. P., Stanfield, G. M., Reddy, K. C., Reinke, V., Kim, S. K. and Villeneuve, A. M. (2002). A targeted RNAi screen for genes involved in chromosome morphogenesis and nuclear organization in the *Caenorhabditis elegans* germline. *Genetics* **162**, 113-128.
- Eisen, M. B., Spellman, P. T., Brown, P. O. and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **95**, 14863-14868.
- Emmons, S. W. and Sternberg, P. S. (1997). Male development and mating behavior. In *C. elegans II* (ed. D. L. Riddle, T. Blumenthal, B. J. Meyer and J. R. Priess), pp. 295-334. Plainview, NY, Cold Spring Harbor Laboratory Press.
- Emmons, S. W. and Yesner, L. (1984). High-frequency excision of transposable element Tc1 in the nematode *Caenorhabditis elegans* is limited to somatic cells. *Cell* **36**, 599-605.
- Fong, Y., Bender, L., Wang, W. and Strome, S. (2002). Regulation of the different chromatin states of autosomes and X chromosomes in the germ line of *C. elegans*. *Science* **296**, 2235-2238.
- Gaudet, J. and Mango, S. E. (2002). Regulation of organogenesis by the *Caenorhabditis elegans* FoxA protein PHA-4. *Science* **295**, 821-825.
- Gaudet, J., VanderElst, I. and Spence, A. M. (1996). Post-transcriptional regulation of sex determination in *Caenorhabditis elegans*: widespread expression of the sex-determining gene *fem-1* in both sexes. *Mol. Biol. Cell* **7**, 1107-1121.
- Giese, A., Jude, R., Kuiper, H., Raudsepp, T., Piumi, F., Schambony, A., Guerin, G., Chowdhary, B. P., Distl, O., Topfer-Petersen, E. and Leeb, T. (2002). Molecular characterization of the equine testis-specific protein 1 (TPX1) and acidic epididymal glycoprotein 2 (AEG2) genes encoding members of the cysteine-rich secretory protein (CRISP) family. *Gene* **299**, 101-109.
- Gonczy, P., Echeverri, C., Oegema, K., Coulson, A., Jones, S. J., Copley, R. R., Duperon, J., Oegema, J., Brehm, M., Cassin, E. et al. (2000). Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. *Nature* **408**, 331-336.
- Goodwin, E. B. and Evans, T. C. (1997). Translational control of development in *C. elegans*. *Semin. Cell Dev. Biol.* **8**, 551-559.
- Gutch, M. J., Flint, A. J., Keller, J., Tonks, N. K. and Hengartner, M. O. (1998). The *Caenorhabditis elegans* SH2 domain-containing protein tyrosine phosphatase PTP-2 participates in signal transduction during oogenesis and vulval development. *Genes Dev.* **12**, 571-585.
- Hodgkin, J., Horvitz, H. R. and Brenner, S. (1979). Nondisjunction mutants of the nematode *Caenorhabditis elegans*. *Genetics* **91**, 67-94.
- Jiang, M., Ryu, J., Kiraly, M., Duke, K., Reinke, V. and Kim, S. K. (2001). Genome-wide analysis of developmental and sex-regulated gene expression profiles in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* **98**, 218-223.
- Kamath, R. S., Fraser, A. G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M. et al. (2003). Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421**, 231-237.
- Kelly, W. G., Schaner, C. E., Dernburg, A. F., Lee, M.-H., Kim, S. K., Villeneuve, A. M. and Reinke, V. (2002). X-chromosome silencing in the germline of *C. elegans*. *Development* **129**, 479-492.
- Kim, S. K., Lund, J., Kiraly, M., Duke, K., Jiang, M., Stuart, J. M., Eizinger, A., Wylie, B. N. and Davidson, G. S. (2001). A gene expression map for *Caenorhabditis elegans*. *Science* **293**, 2087-2092.
- Kohara, Y. (2001). Systematic analysis of gene expression of the *C. elegans* genome. *Tanpakushitsu Kakusan Koso* **46**, 2425-2431.
- Kostrouchova, M., Krause, M., Kostrouch, Z. and Rall, J. E. (2001). Nuclear hormone receptor CHR3 is a critical regulator of all four larval molts of the nematode *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* **98**, 7360-7365.
- Kuwabara, P. E. and Perry, M. D. (2001). It ain't over till it's ova: germline sex determination in *C. elegans*. *BioEssays* **23**, 596-604.
- LaMunyon, C. W. and Ward, S. (1998). Larger sperm outcompete smaller

- sperm in the nematode *Caenorhabditis elegans*. *Proc. R. Soc. Lond. B. Biol. Sci.* **265**, 1997-2002.
- Lu, X. and Horvitz, H. R.** (1998). *lin-35* and *lin-53*, two genes that antagonize a *C. elegans* Ras pathway, encode proteins similar to Rb and its binding protein RbAp48. *Cell* **95**, 981-991.
- Lund, J., Tedesco, P., Duke, K., Wang, J., Kim, S. K. and Johnson, T. E.** (2002). Transcriptional profile of aging in *C. elegans*. *Curr. Biol.* **12**, 1566-1573.
- MacQueen, A. J. and Villeneuve, A. M.** (2001). Nuclear reorganization and homologous chromosome pairing during meiotic prophase require *C. elegans* *chk-2*. *Genes Dev.* **15**, 1674-1687.
- Maeda, I., Kohara, Y., Yamamoto, M. and Sugimoto, A.** (2001). Large-scale analysis of gene function in *Caenorhabditis elegans* by high-throughput RNAi. *Curr. Biol.* **11**, 171-176.
- McKee, B. D. and Handel, M. A.** (1993). Sex chromosomes, recombination, and chromatin conformation. *Chromosoma* **102**, 71-80.
- Mihaylova, V. T., Borland, C. Z., Manjarrez, L., Stern, M. J. and Sun, H.** (1999). The PTEN tumor suppressor homolog in *Caenorhabditis elegans* regulates longevity and dauer formation in an insulin receptor-like signaling pathway. *Proc. Natl. Acad. Sci. USA* **96**, 7427-7432.
- Miller, M. A., Ruest, P. J., Kosinski, M., Hanks, S. K. and Greenstein, D.** (2003). An Eph receptor sperm-sensing control mechanism for oocyte meiotic maturation in *Caenorhabditis elegans*. *Genes Dev.* **17**, 187-200.
- Nakayama, J.-L., Rice, J. C., Strahl, B. E., Allis, C. D. and Grewal, S. I. S.** (2001). Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* **292**, 110-113.
- Nelson, G. A., Lew, K. K. and Ward, S.** (1978). Intersex, a temperature-sensitive mutant of the nematode *C. elegans*. *Dev. Biol.* **66**, 386-409.
- Parisi, M., Nuttall, R., Naiman, D., Bouffard, G., Malley, J., Andrews, J., Eastman, S. and Oliver, B.** (2003). Paucity of genes on the *Drosophila* X chromosome showing male-biased expression. *Science* **299**, 697-700.
- Pellettieri, J., Reinke, V., Kim, S. K. and Seydoux, G.** (2003). Coordinate activation of maternal protein degradation during the egg-to-embryo transition in *C. elegans*. *Dev. Cell* **5**, 451-462.
- Piano, F., Schetter, A. J., Morton, D. G., Gunsalus, K. C., Reinke, V., Kim, S. K. and Kemphues, K. J.** (2002). Gene clustering based on RNAi phenotypes of ovary-enriched genes in *C. elegans*. *Curr. Biol.* **12**, 1959-1964.
- Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z. W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C. P., Allis, C. D. and Jenuwein, T.** (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593-599.
- Reinke, V.** (2002). Functional exploration of the *C. elegans* genome using DNA microarrays. *Nat. Genet.* **32**, 541-546.
- Reinke, V., Smith, H. E., Nance, J., Wang, J., van Doren, C., Begley, R., Jones, S. J. M., Davis, E. B., Scherer, S., Ward, S. and Kim, S. K.** (2000). A global profile of germline gene expression in *C. elegans*. *Mol. Cell* **6**, 605-616.
- Roberts, T. M., Pavalko, F. M. and Ward, S.** (1986). Membrane and cytoplasmic proteins are transported in the same organelle complex during nematode spermatogenesis. *J. Cell Biol.* **102**, 1787-1796.
- Roy, P. J., Stuart, J. M., Lund, J. and Kim, S. K.** (2002). Chromosomal clustering of muscle-expressed genes in *Caenorhabditis elegans*. *Nature* **418**, 975-979.
- Seydoux, G. and Strome, S.** (1999). Launching the germline in *Caenorhabditis elegans*: regulation of gene expression in early germ cells. *Development* **126**, 3275-3283.
- Walhout, A. J. M., Reboul, J., Shtanko, O., Bertin, N., Vaglio, P., Ge, H., Lee, H., Doucette-Stamm, L., Gunsalus, K. C., Schetter, A. J. et al.** (2002). Integrating interactome, phenome, and transcriptome mapping data for the *C. elegans* germline. *Curr. Biol.* **12**, 1952-1958.
- Wang, J. and Kim, S. K.** (2003). Global analysis of dauer gene expression in *Caenorhabditis elegans*. *Development* **130**, 1621-1634.
- Wang, P. J., McCarrey, J. R., Yang, F. and Page, D. C.** (2001). An abundance of X-linked genes expressed in spermatogonia. *Nat. Genet.* **27**, 422-426.