

Sperm Loss Phenocopies Early Aging in *C. elegans*

David Angeles-Albores*, †, Daniel H Leighton*, ‡, Tiffany Tsou†, Tiffany K. Shaw†, Igor Antoshechkin§ and Paul W Sternberg†, 1

*Co-First Authors, †Department of Biology and Biological Engineering, and Howard Hughes Medical Institute, Caltech, Pasadena, CA, 91125, USA ,

‡Department of Human Genetics, Department of Biological Chemistry, and Howard Hughes Medical Institute, University of California, Los Angeles, Los Angeles, CA 90095, USA., §Department of Biology and Biological Engineering, Caltech, Pasadena, CA, 91125, USA

ABSTRACT

KEYWORDS

C. elegans

fog-2

aging

transcriptome

INTRODUCTION

Aging is a natural phenomenon that affects most, if not all, animals.

While there has been a long-standing interest in *C. elegans* longevity and aging past 20 days of life, which has led to a rich literature (), there has not been such intense study into the changes that occur during the period of early aging. We refer to early aging as the period of time that occurs between the first day of adulthood (post-L4 molt) and 72 hours later. In this period of time, *C. elegans* worms undergo a transition between a hermaphroditic, self-fertile stage and its final form as a female worm.

We attempted to elucidate the transcriptomic changes that occur in early aging in the worm. We reasoned that early aging would have two components that ought to be separable in principle. Since the final stage of the worm is as a female, there might be transcriptomic changes associated with sperm status. A second component of aging, on the other hand, should be decoupled from sperm-status. We designed a two factor experiment to test our hypothesis.

Wild-type worms naturally become female as they age. These worms should experience both components of early aging as time proceeds. On the other hand, *C. elegans* defective in sperm formation will never transition to a hermaphroditic stage. As time moves forward, these spermless worms should not exhibit changes related to sperm or fertility status, and should only exhibit the genotype-independent changes. Additionally, we reasoned that we might be able to identify gene expression changes due to different life histories: whereas hermaphrodites lay almost 300 eggs over three days, spermless females likely engage in more mate-seeking behavior. The different life histories could affect gene expression and aging.

As a model for a spermless female worm, we selected the *fog-2* mutant for analysis. *fog-2* is involved in germ-cell sex determination in the hermaphrodite worm and is required for spermatogenesis (Schedl and Kimble 1988; Clifford *et al.* 2000). This protein is only expressed in the germline, which makes it a good target to generate spermless mutants with, since our perturbation would be targeted exclusively within the gonad.

Here, we show that early aging can be divided into two additive components. First, a component that is associated with loss of hermaphroditic sperm and which leads to increases in the levels of transcription factors that are canonically associated with development and cellular differentiation, and enriched in neuronal functions. We also describe a second transcriptomic module consisting of 5,592 genes that is probably associated with biological age in *C. elegans* and that is invariant between worms that undergo a hermaphroditic egg-laying stage and worms that are never fertilized. Our results can be viewed online in interactive plots at the following address: https://wormlabcaltech.github.io/Angeles_Leighton_2016/.

MATERIALS AND METHODS

RNA extraction

Synchronized worms were grown to either young adulthood or the 6th day of adulthood prior to RNA extraction. Synchronization and aging were carried out according to protocols described previously (Leighton *et al.* 2014). 1000–5000 worms from each replicate were rinsed into a microcentrifuge tube in S basal (5.85g/L NaCl, 1g/L K₂HPO₄, 6g/L KH₂PO₄), and then spun at 14,000rpm for 30s. The supernatant was removed and 1mL of Trizol was added. Worms were lysed by vortexing for 30s at room temperature and then 20 minutes at 4°. The Trizol lysate was then spun at 14,000rpm for 10 minutes at 4° down to allow removal of insoluble materials. Thereafter the Ambion TRIzol protocol was followed to finish the RNA extraction (MAN0001271 Rev. Date: 13 Dec 2012).

Copyright © 2016 Angeles-Albores *et al.*

Manuscript compiled: Tuesday 4th October, 2016%

¹Contact: pws@caltech.edu, Address: Department of Biology and Biological Engineering, and Howard Hughes Medical Institute, Caltech, Pasadena, CA, 91106, USA

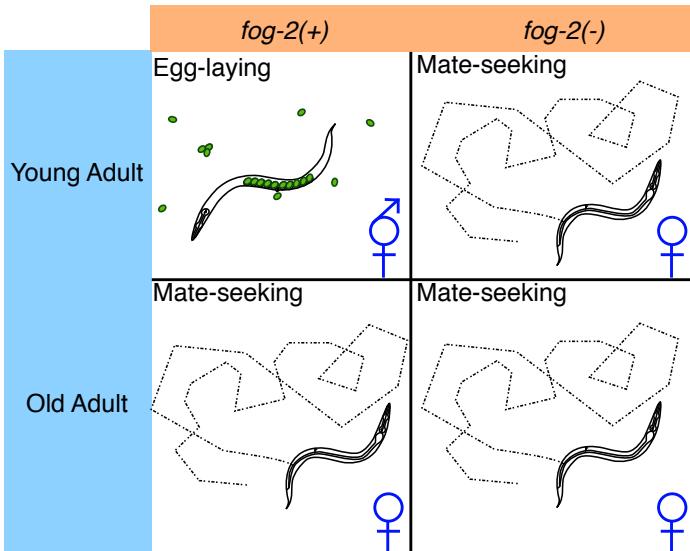


Figure 1 Experimental design to identify genes associated with life-history.

RNA-seq

RNA interference

RNAi was performed as described in previous protocols (Kamath *et al.* 2001) using a commercially available RNAi library (Kamath *et al.* 2003). RNAi bacterial strains were grown in LB plus 100 μ g/mL ampicillin overnight. Fresh RNAi cultures were then plated onto NGM agar plates containing 25 μ g/mL carbenicillin and 1mM IPTG. N2 or *fog-2(q71)* hermaphrodites grown on *E. coli* OP50 were bleached onto sterile plates, and starved L1s transferred to recently seeded RNAi plates. All assays were performed on the offspring of these L1s. Worms grown on every strain were monitored for gross abnormalities, such as sterility, lethality and larval arrest. Control worms in all assays were fed with an anti-GFP RNAi strain. All RNAi worms were grown at 20° C.

Oocyte dropping assay

We performed a slightly modified oocyte dropping assay based on previously described protocols (White *et al.* 2012). Feminized *fog-2(q71)* hermaphrodites were picked as virgin L4s the day before the assay to a fresh RNAi plate. The following day, the adult animals were placed on assay plates (NGM agar seeded with 15 μ L of *E. coli* OP50 four days earlier), twenty worms per assay, and allowed to incubate at room temperature. Laid oocytes were counted after two hours. Plates were then left at room temperature overnight to serve as lawn-leaving assays.

Lawn-leaving assay

Lawn-leaving assays were performed as described previously (Lipton 2004). Young adult N2 hermaphrodites were selected the day of the assay and placed on assay plates (same as the oocyte dropping plates), twenty worms per assay, and allowed to incubate at room temperature overnight. Assays for *fog-2(q71)* hermaphrodites were performed on the same worms as used in the oocyte dropping assays. The following morning, plates were scored for leaving, with any worm touching any part of the bacterial lawn with any part of its body deemed to be “on” the lawn, and all others deemed to be “off”.

Brood size counting

Worms were selected for this assay as L4 hermaphrodites to ensure that all progeny could be counted. For each replicate of each assay, a single worm was placed on a fresh RNAi plate and incubated at 20°. Every 1–2 days, the test worm was moved to a fresh RNAi plate, until it stopped laying eggs. Progeny were counted on each plate before they reached adulthood to ensure that only a single generation was counted.

Statistical Analysis

RNA-Seq Analysis. RNA-seq alignment was performed using Kallisto (Bray *et al.* 2015) with 200 bootstraps. The commands used for read-alignment are in file . Differential expression analysis was performed using Sleuth (Pimentel *et al.* 2016). The following Generalized Linear Model (GLM) was fit:

$$\log(y_i) = \beta_{0,i} + \beta_{G,i} G + \beta_{A,i} A + \beta_{A::G,i} A G, \quad (1)$$

where y_i are the TPM counts for the i th gene; $\beta_{0,i}$ is the intercept for the i th gene, and $\beta_{X,i}$ is the regression coefficient for variable X for the i th gene; A is a binary age variable indicating young adult (0) or old adult (1) and G is the genotype variable indicating wild-type (0) or *fog-2* (1); $\beta_{A::G,i}$ refers to the regression coefficient accounting for the interaction between the age and genotype variables in the i th gene. Genes were called significant if the FDR-adjusted q -value for any regression coefficient was less than 0.1. Our script for differential analysis is available on GitHub.

Regression coefficients and TPM counts were processed using Python 3.5 in a Jupyter Notebook (Pérez and Granger 2007). Data analysis was performed using the Pandas, NumPy and SciPy libraries (McKinney 2011; Van Der Walt *et al.* 2011; Oliphant 2007). Graphics were created using the Matplotlib and Seaborn libraries (Waskom *et al.* 2016; Hunter 2007). Interactive graphics were generated using Bokeh (Bokeh Development Team 2014).

Tissue Enrichment Analysis was performed using WormBase’s TEA tool (Angeles-Albores *et al.* 2016) for Python.

Brood Size Analysis. Brood size results were analyzed using Welch’s t-test to identify genes that were significantly different from a GFP RNAi control. RNAi control results were pooled over multiple days because we could not detect systematic day-day variation. We did not apply FDR or Bonferroni correction because, at a p-value threshold for significance of 0.05, we expected 1 false positive on average per screen.

Lawn-leaving Analysis. Lawn-leaving results were analyzed using a χ^2 test for categorical variables. Results were considered statistically significant if $p < 0.05$. No FDR or Bonferroni correction was applied because the size of the screen was too small, with 1 false positive expected on average per screen. However, the lawn-leaving results suffered from high variance, which can lead to false positive results. To safeguard against false positive discovery, we used a non-parametric bootstrap to estimate the true χ^2 value. Using a bootstrap does not lead to statistical acceptance of any gene that was not accepted without a bootstrap; however, applying a bootstrap does lead to statistical rejection of a large number of results.

Oocyte Dropping Analysis. Oocyte dropping results were analyzed using a non-parametric bootstrapped Mann-Whitney U-test because the GFP control variance was very large relative to the variance of the RNAi treatments. Results were considered statistically significant if $p < 0.05$.

Data Availability

Strains are available upon request. File S1 XXXX. File S2 contains XXXX. File S3 contains XXXX. Sequence data are available at GenBank and the accession numbers are listed in File S3. Gene expression data are available at GEO with the accession number: XXXXXX. Code used to generate the simulated data is provided in file XXXX.

RESULTS

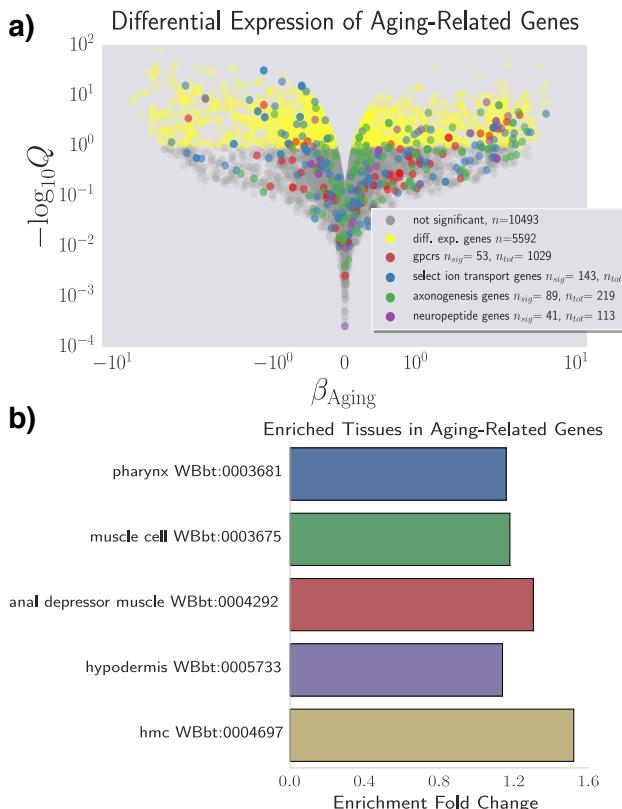


Figure 2 a We identified a common aging transcriptome between N2 and *fog-2* animals, which showed 5,592 altered genes. The volcano plot was randomly down-sampled 30%. **b** Tissue Enrichment Analysis showed that genes associated with muscle tissues and hypodermis are enriched in this dataset. An interactive version of this graph can be found on our [website](#).

Transcriptomics

We used a linear generalized model (see [Statistical Analysis](#)) with interactions to identify a transcriptomic profile associated with the *fog-2* genotype, a transcriptomic profile of *C. elegans* aging common to both genotypes. We identified an aging transcriptomic phenotypic consisting of 5,592 genes that were differentially expressed in 6 day old animals relative to 3 day old animals. This constitutes almost one quarter of the genes in *C. elegans*. Tissue Enrichment Analysis (TEA) showed that muscle-associated and hypodermis-associated genes are particularly enriched in this dataset (see Figure 2).

To verify the quality of our dataset, we generated a list of genes expected to be altered in 6 day old worms using previous literature reports including downstream genes of *daf-12*, *daf-16*, and aging and lifespan extension datasets (1). Out of 1056 genes in our golden

standards, we found 506 genes. This results was statistically significant with a p-value < 10⁻³⁸.

To better understand this dataset, we extracted all the transcription factors that were altered in the aging gene-set. We found 145 transcription factors. We expected these transcription factors to reflect the same tissue enrichment as the bulk dataset, but the results showed enrichment of hypodermic tissues and neuronal tissues, not muscle (see Table 1). Many of these transcription factors have been typically associated with developmental processes, and it is unclear why they would change expression in adult animals. Interactive volcano plots for each gene-set can be found in our [website](#), where all the transcription factors are shown as green and all significantly altered genes are plotted in gray (no downsampling).

Next, we extracted the regression coefficients associated with genotype change, we were also able to identify 1,881 genes associated with the *fog-2* genotype, including 60 transcription factors. Gonad-related tissues were enriched in this gene set, consistent with the function of *fog-2* as a sperm transcription factor. As before, tissue enrichment analysis of these transcription factors reveals that they are involved in neural development. Of the 1,881 genes that we identified in the *fog-2* transcriptome, 1,040 genes were also identified in our aging set. Moreover, of these 1,040 genes, 905 genes changed in the same direction in age and genotype.

We were surprised at the large fraction of genes that overlapped between these two categories. We built our model to explicitly avoid overlap between variables. Our original expectation had been that certain genes would show a common aging phenotype regardless of genotype; that *fog-2* would exhibit a specific set of changes; and that a small set of genes would be altered between genotypes throughout the course of aging. However, the large fraction of genes that exhibit shared changes between both variables suggests that almost all genes that are involved in sperm loss through mutation of *fog-2* have an aberrant aging behaviour.

This aberrant behaviour can be most clearly observed by plotting the β regression coefficients for each variable against each other. Doing so reveals a clear trend along the line $y = x$. However, our model is built to specifically disallow this. The only situation in which $\beta_{\text{aging}} = \beta_{\text{genotype}}$ is a valid statement in our model is when $\beta_{\text{aging:genotype}} \neq 0$. Therefore, we also plotted $\beta_{\text{aging:genotype}}$ against β_{aging} . This revealed a strong inverse relationship: The interaction term cancels the aging (or genotype) term. An old, *fog-2* worm would be represented as $\beta_{\text{aging}} + \beta_{\text{genotype}} + \beta_{\text{aging:genotype}} = \beta_{\text{genotype}}$. Therefore, genes that are associated with sperm loss through mutation in the *fog-2* gene do not change through early aging in these animals. However, in animals that are wild-type, these same genes will eventually change to the same levels as in the *fog-2* mutants. In other words, *fog-2* partially phenocopies the aging process in wild-type animals 3 days post-adulthood (see Fig. 3).

Screens

We designed three screens to test our gene sets. We selected gene targets by identifying genes that belonged exclusively to one of the three sets defined previously. We selected two assays that could be performed to test genes associated with genotype change. We reasoned that genes that are negatively correlated with the *fog-2* genotype could be associated with fertility, whereas genes that are positively correlated with the *fog-2* genotype could be associated with ovulation rate. Moreover, lawn leaving is known to be associated with sperm status (2), so we decided to perform a second screen to study lawn-leaving behaviour.

We performed a brood size screen, selecting as targets genes

Table 1 Transcription Factor TEA Results. We ran TEA on the transcription factors associated with aging. Raw results showed a large list of terms associated with embryonic tissues that are neuronal precursors (AB lineage), so we trimmed the results by removing any terms that were expected to show up less than once in our list (if a term is expected to show up 10^{-3} times in a list, 1 occurrence is enough to show enrichment in this list). The best 4 results by Q-value are shown below.

Tissue	Expected	Observed	Q-value
P11	1	9	0.000006
ventral nerve cord	9	26	0.000006
dorsal nerve cord	5	19	0.000006
head muscle	3	13	0.000040
P7.p	1	6	0.001382

that were upregulated in N2 but downregulated in *fog-2*, and we also included some genes from alternative categories if we visually noticed serious decreases in brood size. We identified 9 genes that altered brood size (see Figure 4). Of these 7 genes, XXX were previously known. XXX/7 were genes that were associated specifically with the *fog-2* phenotype. XXX/7 were associated with aging or life history.

We reasoned that some of the genes in our genotype dataset might also be associated with ovulation rate in *fog-2*. We selected genes that showed upregulation in *fog-2* animals and placed them on a lawn for two hours. We observed large variation in the ovulation rate for the control RNAi, and as a result no genes were associated with alterations in ovulation rate. However, the pooled variance across the screen was very similar to the control variance, suggesting that our failure to identify genes was not a result of poor conditions or experimental variance. Our screen identified a single hit: C23H4.4, a carboxyl ester lipase with unknown function.

We also performed lawn-leaving assays because previous research has reported differences in N2 and *fog-2* leaving rates (). We tested genes that increased in *fog-2* animals in a *fog-2* background, expecting that decreasing these genes should lead to a reversal of the lawn-leaving assay. Likewise, we tested genes that were decreased in *fog-2* animals in an N2 background, expecting that knockdown of these genes would cause lawn-leaving. We identified 9 genes that had an altered lawn-leaving profile in an N2 background. We also identified 9 genes that had an altered lawn-leaving profile in a *fog-2* background. Oddly, both screens identified genes that mainly stimulate lawn-leaving. Initially, we had expected that RNAi knockdown would allow us to identify genes that inhibit lawn-leaving behaviour in an N2 background; whereas RNAi knockdown in a *fog-2* background would identify genes that promote lawn-leaving. However, our screen only identified genes that suppress lawn-leaving in an N2 background. All other hits stimulated lawn-leaving.

DISCUSSION

Defining an Early Aging Phenotype

Our experimental design enables us to decouple the effects of egg-laying from aging. As a result of this, we identified a set of almost 4,000 genes that are altered similarly between wild-type and *fog-2* mutants. Due to the read depth of our transcriptomic data (20 million reads) and the number of samples measured (3 biological replicates for 4 different life stages/genotypes), this dataset constitutes a high-quality description of the genetic changes that occur

in natural *C. elegans* aging.

Developmental Factors Change During Early Aging

Our transcriptomic explorations revealed a host of transcription factors that changed during early aging in *C. elegans*. Many of these transcription factors have been associated with embryonic development via cellular differentiation and specification. For example, we identified the transcription factor *lin-32*, which has been associated with neuron development (); the Six5 ortholog *unc-39* has been associated with axonal pathfinding; *cnd-1*, a homolog of the vertebrate NeuroD transcription factors, is expressed in the early embryo and its expression is reported to disappear by the end of the first larval stage ()�.

Our explorations shown that the loss of the *fog-2* transcription factor phenocopies early aging in *C. elegans*. The reason for phenocopying is unclear at this point, but we speculate that such a result could happen if an animal could sense its own fertilization status. Indeed, prior work has established that *C. elegans* is capable of detecting its fertilization state, and that this state has consequences for pheromone production and mating behaviours. Given the enrichment of neuronal transcription factors that are associated with sperm loss in our dataset, we believe this dataset should contain some of the transcriptomic modules that are involved in these pheromone and behavioural pathways. Currently, we cannot judge how many of the changes induced by loss of hermaphroditic sperm are developmental (i.e., irreversible), and how many can be rescued by mating to a male. While an entertaining thought experiment, establishing whether these transcriptomic changes can be rescued by males is a daunting experimental task, given that the timescales for robust, whole-animal transcriptomic changes could reasonably be the same as the timescale of onset of embryonic transcription.

Interpretation of Screen Hits

LITERATURE CITED

- Angeles-Albores, D., R. Y. N. Lee, J. Chan, and P. W. Sternberg, 2016 Tissue enrichment analysis for *C. elegans* genomics. *BMC Bioinformatics* 17: 366.
- Bokeh Development Team, 2014 Bokeh: Python library for interactive visualization .
- Bray, N. L., H. Pimentel, P. Melsted, and L. Pachter, 2015 Near-optimal RNA-Seq quantification. *aRxiv* 34: 525–528.
- Clifford, R., M. H. Lee, S. Nayak, M. Ohmachi, F. Giorgini, and T. Schedl, 2000 FOG-2, a novel F-box containing protein, asso-

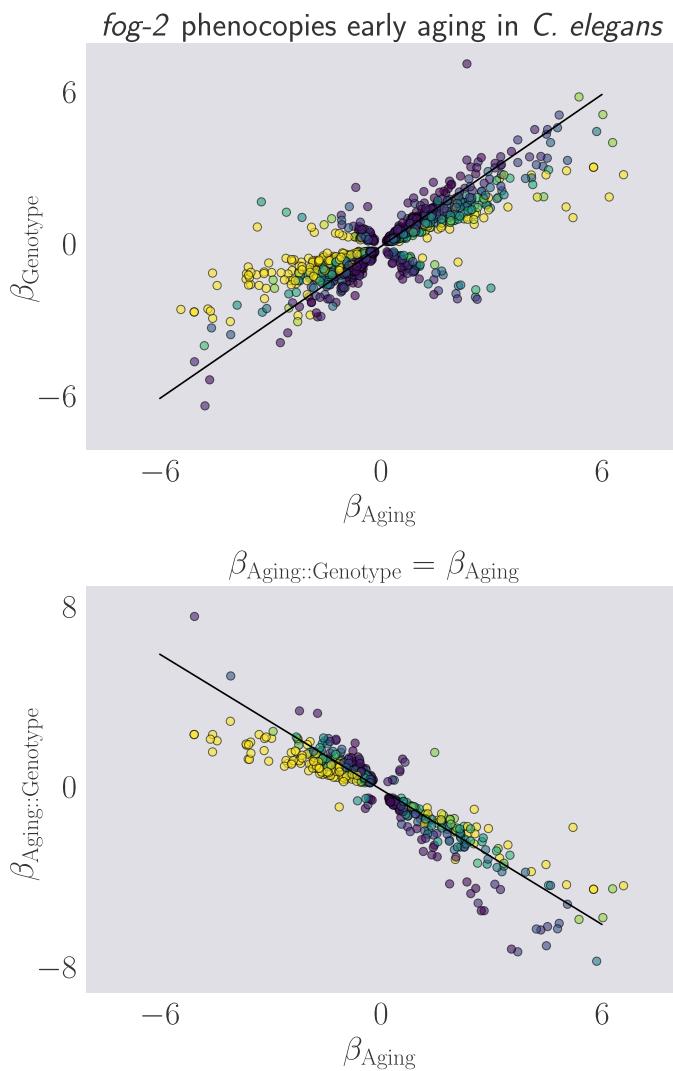


Figure 3 *fog-2* partially phenocopies early aging in *C. elegans*. Feminization by loss of *fog-2* is associated with a transcriptomic phenotype involving 1,881 genes. 1,040/1,881 of these genes are also altered in wild-type worms as they progress from young adulthood to old adulthood, and 905 change in the same direction. However, progression from young to old adulthood in a *fog-2* background results in no change in the expression level of these genes. **a** β_{Aging} vs β_{Genotype} shows that these values have a correlation close to unity. Black line is the line $y = x$, not a line of best fit. **b** $\beta_{\text{Aging}} \cdot \text{Genotype}$ vs β_{Aging} shows that there is an inverse correlation between these two variables, which means that whereas in a wild-type animal these genes would change expression level as it ages, in a *fog-2* animal these genes would not change. The black line is the line $y = -x$, and is not a line of best fit. For both graphs, color is proportional to $-\log Q$. Purple shows Q -values close to 0.1, whereas yellow shows values $< 10^{-20}$. An interactive version of these graphs can be found on our [website](#).

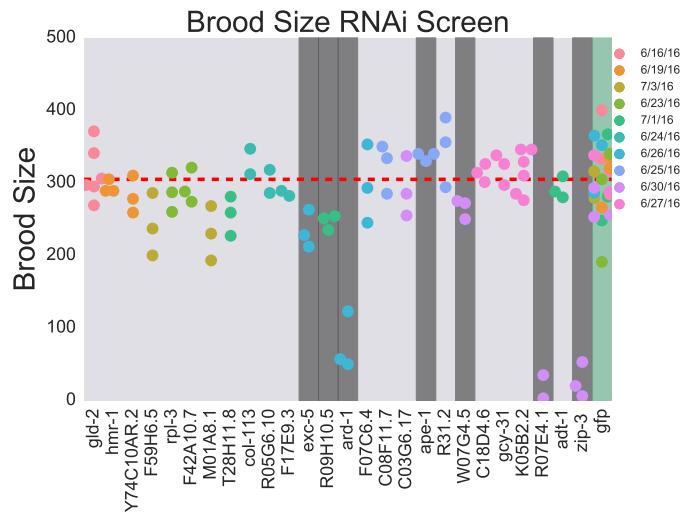


Figure 4 Brood size screen results. GFP control RNAi is shaded in green. Results that are statistically significantly different from GFP are shaded in dark grey. Dotted red line shows the mean of the pooled GFP controls. Points are colored by the date the assay was started on.

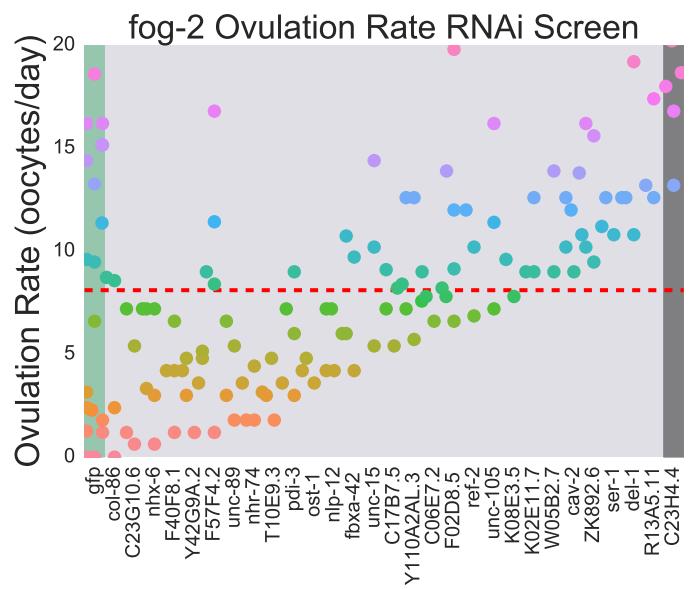


Figure 5 Ovulation Rate Assay. GFP control RNAi is shaded in green. Results that are statistically significantly different from GFP are shaded in dark grey. Dotted red line shows the mean of the pooled GFP controls. Points are colored proportionally to their y-coordinate

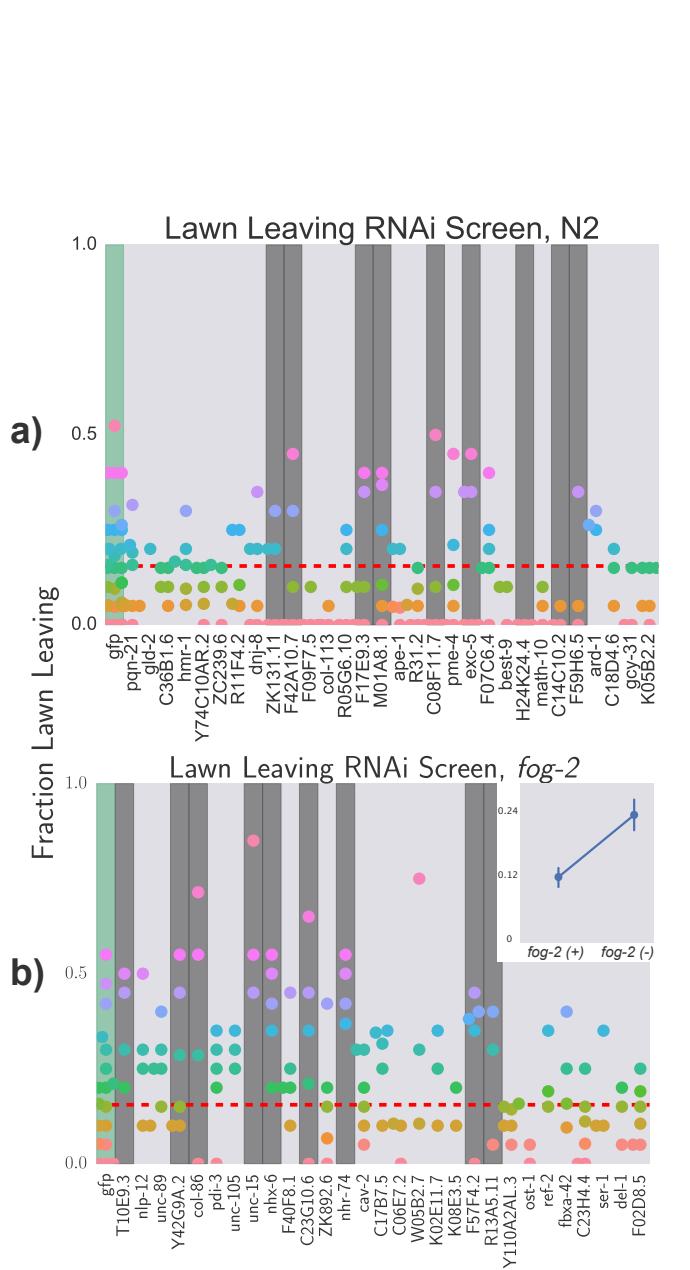


Figure 6 lawn-leaving Screen Results. GFP control RNAi is shaded in green. Results that are statistically significantly different from GFP are shaded in dark grey. Dotted red line shows the mean of the pooled GFP controls. Points are colored proportionally to their y-coordinate. **a** N2 lawn-leaving assay. **b** fog-2 lawn-leaving assay. Inset shows the screen-wide average lawn-leaving rate of N2 and fog-2, which shows that the screen-wide lawn-leaving rate of fog-2 worms is twice the rate of lawn-leaving of N2 worms, in agreement with previous literature ()�.

ciates with the GLD-1 RNA binding protein and directs male sex determination in the *C. elegans* hermaphrodite germline. *Development* (Cambridge, England) **127**: 5265–5276.

Hunter, J. D., 2007 Matplotlib: A 2D graphics environment. *Computing in Science and Engineering* **9**: 99–104.

Kamath, R. S., A. G. Fraser, Y. Dong, G. Poulin, R. Durbin, M. Gotta, A. Kanapin, N. Le Bot, S. Moreno, M. Sohrmann, D. P. Welchman, P. Zipperlen, and J. Ahringer, 2003 Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421**: 231–7.

Kamath, R. S., M. Martinez-Campos, P. Zipperlen, A. G. Fraser, and J. Ahringer, 2001 Effectiveness of specific RNA-mediated interference through ingested double-stranded RNA in *Caenorhabditis elegans*. *Genome biology* **2**: RESEARCH0002.

Leighton, D. H. W., A. Choe, S. Y. Wu, and P. W. Sternberg, 2014 Communication between oocytes and somatic cells regulates volatile pheromone production in *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences* **111**: 17905–17910.

Lipton, J., 2004 Mate Searching in *Caenorhabditis elegans*: A Genetic Model for Sex Drive in a Simple Invertebrate. *Journal of Neuroscience* **24**: 7427–7434.

McKinney, W., 2011 pandas: a Foundational Python Library for Data Analysis and Statistics. *Python for High Performance and Scientific Computing* pp. 1–9.

Oliphant, T. E., 2007 SciPy: Open source scientific tools for Python. *Computing in Science and Engineering* **9**: 10–20.

Pérez, F. and B. Granger, 2007 IPython: A System for Interactive Scientific Computing Python: An Open and General-Purpose Environment. *Computing in Science and Engineering* **9**: 21–29.

Pimentel, H. J., N. Bray, S. Puente, P. Melsted, and L. Pachter, 2016 Differential analysis of RNA-Seq incorporating quantification uncertainty. *bioRxiv* p. 058164.

Schedl, T. and J. Kimble, 1988 fog-2, a germ-line-specific sex determination gene required for hermaphrodite spermatogenesis in *Caenorhabditis elegans*. *Genetics* **119**: 43–61.

Van Der Walt, S., S. C. Colbert, and G. Varoquaux, 2011 The NumPy array: A structure for efficient numerical computation. *Computing in Science and Engineering* **13**: 22–30.

Waskom, M., S. St-Jean, C. Evans, J. Warmenhoven, K. Meyer, M. Martin, L. Rocher, P. Hobson, P. Bachant, T. Nagy, D. Wehner, O. Botvinnik, T. Megies, S. Lukauskas, Drewokane, E. Ziegler, T. Yarkoni, A. Miles, A. Lee, L. P. Coelho, Y. Halchenko, T. Augspurger, G. Hitz, J. Vanderplas, C. Fitzgerald, J. B. Cole, Gkunter, S. Villalba, S. Hoyer, and E. Quintero, 2016 seaborn: v0.7.0 (January 2016).

White, A., A. Fearon, and C. M. Johnson, 2012 HLH-29 regulates ovulation in *C. elegans* by targeting genes in the inositol triphosphate signaling pathway. *Biology open* **1**: 261–8.