Transcriptomic Description of an Endogenous Female State in *C. elegans*

David Angeles-Albores[1,†] Daniel H.W. Leighton[2,†] Tiffany Tsou[1]
Tiffany H. Khaw[1] Igor Antoshechkin[3] Paul W. Sternberg[1,*]

October 29, 2016

† These authors contributed equally to this work
1 Department of Biology and Biological Engineering, and Howard Hughes Medical Institute, Caltech, Pasadena, CA, 91125, USA
2 Department of Human Genetics, Department of Biological Chemistry, and Howard Hughes Medical Institute, University of California, Los Angeles, Los Angeles, CA 90095, USA
3 Department of Biology and Biological Engineering, Caltech, Pasadena, CA, 91125, USA
***** Corresponding author. Contact:pws@caltech.edu

Abstract

Understanding genome and gene function in a whole organism requires us to fully comprehend the life cycle and the physiology of the organism in question. Although *C. elegans* is traditionally though of as a hermaphrodite, XX animals exhaust their sperm and become (endogenous) females after 3 days of egg-laying. The molecular physiology of this state has not been studied as intensely as other parts of the life cycle, in spite of documented changes in behavior and metabolism that occur at this stage. To study the female state of *C. elegans*, we designed an experiment to measure the transcriptomes of 1st day adult females; endogenous, 6th day adult females; and at the same time points, mutant feminized worms. At these time points, we were able to separate the effects of biological aging from the transition into the female state. We find that spermless young adult animals partially phenocopy 6 day old wild-type animals that have depleted their sperm after egg-laying, and that spermless animals also exhibit fewer differentially expressed genes as they age throughout these 6 days. Our results indicate that sperm loss is responsible for some of the naturally occuring transcriptomic changes that occur during the life cycle of these animals. These changes are enriched in transcription factors canonically associated with neuronal development and differentiation. Our data provide a high-quality picture of the changes that happen in global gene expression throughout the period of early aging in the worm.

Introduction

In order to understand gene and genome function, we must understand the important components of or-

ganisms' life cycles and physiological states. Here, we investigate what we argue is a distinct state in the *C. elegans* life cycle, the endogenous female state. *C. elegans* is best known as hermaphrodite, but after 3–4 days of egg-laying, the animals become sperm-depleted, which marks a transition into an endogenous female state. Transition into this state is marked by increased male-mating success [8], which may be due to an increased attractiveness to males [19]. This increased attractiveness acts at least partially through production of volatile chemical cues [14]. Despite these observations, up until this point, a physiologic description of the physiology of this state is lacking. We address this problem by using transcriptomic studies to measure genome-wide expression changes that occur as the worm enters into an endogenous female stage.

Since hermaphrodites are reproductive adults, the transition into the female state might also be accompanied by signs of aging. Comparing a hermaphrodite worm with an endogenous (post-hermaphroditic) female would not enable us to separate biological aging from any development that may occur after a nematode depletes its own sperm. In order to decouple the effects of aging and sperm-loss, we deviced a two factor experiment. First, we examined wild type XX animals at the beginning of adulthood (before worms contained embryos, referred to as 1st day adults) and after sperm depletion (6 days after the last molt, which we term 6th day adults). Second, we examined feminized XX animals that fail to produce sperm but are fully fertile if supplied sperm by mating with males (see Fig. 1). We selected a *fog-2* mutant as a spermless worm suitable for analysis. *fog-2* is involved in germ-cell sex determination in the hermaphrodite worm and is required for sperm production [5, 25].

C. elegans defective in sperm formation will never transition to a hermaphroditic stage. As time moves forward, these spermless worms only exhibit changes related to biological aging. We also 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 do not lay a single one. The different life histories could affect gene expression and aging.

Here, we show that we can detect a transcriptional signature associated both with loss of hermaphroditic sperm and entrance into the endogenous female state. We can also detect changes associated specifically with the biological aging. Loss of sperm leads to increases in the expression levels of transcription factors that are canonically associated with development and cellular differentiation and enriched in neuronal functions. Biological aging also causes transcriptomic changes consisting of 5,592 genes in *C. elegans*. 4,552 of these changes are common between wild-type and *fog-2* worms, indicating they do not depend on life history or genotype. To facilitate exploration of the data, we have generated a website with interactive plots: https://wormlabcaltech.github.io/Angeles_Leighton_2016/.

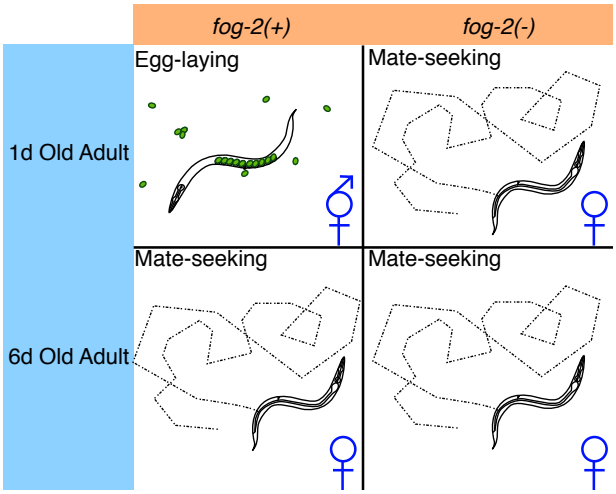

Figure 1. Experimental design to identify genes associated with life-history.

Materials and Methods

Strains

Strains were grown at 20°C on NGM plates containing *E. coli* OP50. We used the laboratory *C. elegans* strain N2 as our wild-type strain [28]. We also used the N2 mutant strain JK574, which contains the *fog-2(q71)* allele, for our experiments.

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 [14]. 1,000–5,000 worms from each replicate were rinsed into a microcentrifuge tube in S basal (5.85g/L NaCl, 1g/L K_2HPO_4, 6g/L KH_2PO_4), and then spun down at 14,000rpm for 30s. The supernatant was removed and 1mL of TRIzol was added. Worms were lysed by vortexing for 30 s at room temperature and then 20 min at 4°. The TRIzol lysate was then spun down at 14,000rpm for 10 min at 4°C to allow removal of insoluble materials. Thereafter the Ambion TRIzol protocol was followed to finish the RNA extraction (MAN0001271 Rev. Date: 13 Dec 2012). 3 biological replicates were obtained for each genotype and each time point.

RNA-Seq

RNA integrity was assessed using RNA 6000 Pico Kit for Bioanalyzer (Agilent Technologies #5067–1513) and mRNA was isolated using NEBNext Poly(A) mRNA Magnetic Isolation Module (New England Biolabs, NEB, #E7490). RNA-Seq libraries were constructed using NEBNext Ultra RNA Library Prep Kit for Illumina (NEB #E7530) following manufacturer's instructions. Briefly, mRNA isolated from $\sim 1\mu$g of total RNA was fragmented to the average size of 200 nt by incubating at 94 °C for 15 min in first strand buffer, cDNA was synthesized using random primers and ProtoScript II Reverse Transcriptase followed by second strand synthesis using Second Strand Synthesis Enzyme Mix (NEB). Resulting DNA fragments were end-repaired, dA tailed and ligated to NEBNext hairpin adaptors (NEB #E7335). After ligation, adaptors were converted to the 'Y' shape by treating with USER enzyme and DNA fragments were size selected using Agencourt AMPure XP beads (Beckman Coulter #A63880) to generate fragment sizes between 250 and 350 bp. Adaptor-ligated DNA was PCR amplified followed by AMPure XP bead clean up. Libraries were quantified with Qubit dsDNA HS Kit (ThermoFisher Scientific #Q32854) and the size distribution was confirmed with High Sensitivity DNA Kit for Bioanalyzer (Agilent Technologies #5067–4626). Libraries were sequenced on Illumina HiSeq2500 in single read mode with the read length of 50 nt following manufacturer's instructions. Base calls were performed with RTA 1.13.48.0 followed by conversion to FASTQ with bcl2fastq 1.8.4.

RNA interference

RNAi was performed as described in previous protocols [13] using a commercially available RNAi library [12]. RNAi bacterial strains were grown in LB plus 100μg/mL ampicillin overnight. Fresh RNAi cultures were then plated onto NGM agar plates containing 25μg/mL carbenicillin and 1mM IPTG. N2 or *fog-2(q71)* hermaphrodites grown on *E. coli* OP50 were bleached onto NGM plates without *E. coli*, and allowed to hatch and develop into L1 larvae. Starved L1s were transferred to recently seeded RNAi plates. All assays were performed on the offspring of these L1s (F1 generation). Worms grown on every RNAi 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-treated worms were grown at 20° C. We extracted plasmid DNA from all RNAi strains that showed a statistically significant phenotype to verify the sequence identity of the hits.

Ovulation assay

We performed a slightly modified ovulation assay based on previously described protocols [31]. 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 (6cm diameter plates, NGM agar seeded with 15μL 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 [15]. Young adult N2 hermaphrodites were selected the day of the assay and placed on assay plates (same as the ovulation assay plates), twenty worms per assay, and allowed to incubate at room temperature overnight. Assays for *fog-2(q71)* were performed on the same worms as used in the ovulation assay. 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°C. 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 [3] with 200 bootstraps. The commands used for read-alignment are in the S.I.. Differential expression analysis was performed using Sleuth [22]. The following General Linear Model (GLM) was fit:

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

where y_i are the TPM counts for the ith gene; $\beta_{0,i}$ is the intercept for the ith gene, and $\beta_{X,i}$ is the regression coefficient for variable X for the ith gene; A is a binary age variable indicating 1st day adult (0) or 6th day 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 ith 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 [21]. Data analysis was performed using the Pandas, NumPy and SciPy libraries [11, 18, 29]. Graphics were created using the Matplotlib and Seaborn libraries [10, 30]. Interactive graphics were generated using Bokeh [2].

Tissue Enrichment Analysis was performed using WormBase's TEA tool [1] for Python.

Brood Size Analysis.

Brood size results were analyzed using Welch's t-test to identify RNAi treatments 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 on this dataset does not lead to statistical acceptance of any gene that was not accepted without a bootstrap; however, applying a boostrap does lead to statistical rejection of a large number of results.

Ovulation Assay Analysis.

ovulation assay 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 from the *Caenorhabditis* Genetics Center upon request. All of the data and scripts pertinent for this project except the raw reads can be found on our Github repository `https://github.com/WormLabCaltech/Angeles_Leighton_2016`. File S1 contains the Kallisto commands used to process the read alignments as well as the accession numbers for the raw-reads, whcih are available at GenBank. It also contains a detailed explanation for the format of files S2 and S3. File S2 contains the list of genes that were altered in aging regardless of genotype. File S3 contains the list of genes and their associations with the *fog-2* phenotype. File S4 contains the genes that changed differently in aging wild-type worms and in *fog-2* mutants. Raw reads will be deposited in the Sequence Read Archive and made available shortly.

Results

Transcriptomics

We obtained 16–19 million reads mappable to the *C. elegans* genome per biological replicate, which enabled us to identify 14,702 individual genes totalling 21,143 isoforms (see Figure 2a)

We used a linear generalized model (see RNA-Seq Analysis) with interactions to identify a transcriptomic profile associated with the *fog-2* genotype independently of age, as well as a transcriptomic profile of *C. elegans* aging common to both genotypes. We identified a transcriptomic aging signature consisting of 5,592 genes that were differentially expressed in 6th day adult animals of either genotype relative to 1st day adult animals. This constitutes more than

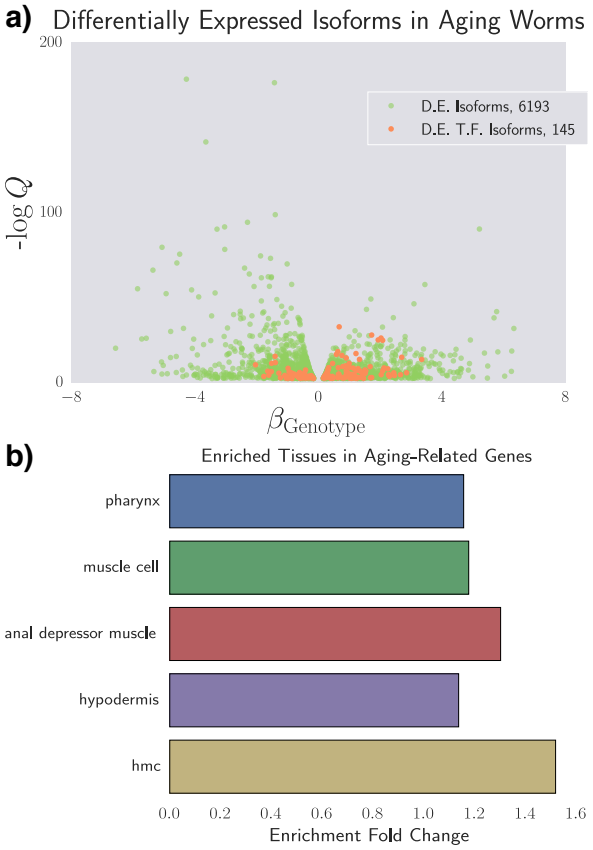

Figure 2. **a** We identified a common aging transcriptome between N2 and *fog-2* animals, which identified 6,193 differentially expressed isoforms totaling 5,592 genes. The volcano plot is randomly down-sampled 30% for ease of viewing. Each point represents an individual isoform. β_{Aging} is the regression coefficient in our model. Larger magnitudes of β indicate a larger log-fold change. The y-axis shows the negative logarithm of the q-values for each point. Green points are differentially expressed isoforms; orange points are differentially expressed isoforms of genes predicted to be transcription factors [24]. An interactive version of this graph can be found on our website. **b** Tissue Enrichment Analysis [1] showed that genes associated with muscle tissues and hypodermis are enriched in aging-related genes. Only statistically significantly enriched tissues are shown. Enrichment Fold Change is defined as $Observed/Expected$.

one quarter of the genes in *C. elegans*. Tissue Enrichment Analysis (TEA) [1] showed that muscle- and hypodermis-associated genes are particularly enriched in this dataset (see Figure 2b).

To verify the quality of our dataset, we generated a list of 1,056 golden standard genes expected to be altered in 6th day adult worms using previous literature reports including downstream genes of *daf-12*, *daf-16*, and aging and lifespan extension datasets [7,9,16,17,20]. We found 506 genes of 1,056 genes in our golden standard. This result was statistically significant with a p-value $< 10^{-38}$.

Next, we used a published compendium [24] to search for known or predicted transcription factors. We found 145 transcription factors in the set of genes with differential expression in aging nematodes. We expected these transcription factors to reflect the same tissue enrichment as the bulk dataset, but the results showed enrichment of hypodermal and neuronal tissues, not muscle (see Table 1). Many of these transcription factors have been associated with developmental processess, 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.

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 specification factor (see Jupyter Notebook, section on Tissue Enrichment Analysis). Similarly to the transcriptio factors associated with aging, 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 response either aging or germline feminization.

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 age differently between genotypes. However, the large fraction of genes that exhibit shared changes between both variables suggests that almost all genes that are associated with sperm loss through mutation of *fog-2* have an aberrant aging behavior.

This behavior can be most clearly observed by plotting the β regression coefficients for each variable against each other (see Fig. 3). This reveals a clear trend along the line $y = x$. However, our model (see RNA-Seq Analysis) is specifically built to 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, $\beta_{\text{aging::genotype}}$, cancels the aging (or genotype) term. The transcriptional change seen in 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 show significant expression changes compared to the wild type in response to sperm loss through mutation of *fog-2* do not change in this strain as these animals age. However, in wild-type animals, 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 that will later occur in wild-type animals 3 days post-adulthood.

Screens

To test whether the genes we identified might lead to phenotypes when they are perturbed, we performed a series of screens that would identify altered phenotypes associated with sperm loss: fertility, altered lawn-leaving or ovulation rate. We performed a brood size screen, selecting as targets genes that were upregulated in N2 but downregulated in *fog-2*, although we also included a gene that was visually determined to have a brood size phenotype (*zip-3*, which goes down with age). In total, we identified seven genes whose knockdown altered brood size (see Figure 4). Most of these genes were associated with decreased brood size. Of the six genes that showed decreased brood size, one had previously been described as having a sterile or lethal phenotype (*ard-1*; [27]). Knock-down of *exc-5* had not been de-

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, leading to many false positives). The best 5 results by Q-value are shown below.

Tissue	Expected	Observed	Q-value
P11	1	9	$< 10^{-5}$
ventral nerve cord	9	26	$< 10^{-5}$
dorsal nerve cord	5	19	$< 10^{-5}$
head muscle	3	13	$< 10^{-4}$
P7.p	1	6	$< 10^{-2}$

scribed to have a sterile phenotype before. The remaining four genes were identified in our screen had not been described previously to have a phenotype. *R09H10.5* and *WO7G4.5* had mild effects on brood size; whereas *zip-3*, and *R07E4.1* had strong effects on fertility. *ard-1*, *R09H10.5* and *WO7G4.5* are all expressed in the intestine. Although we cannot rule out that these genes are essential in development, there is a strong functional connection between the intestine and the germline through yolk production [6], and we hypothesize that these genes participate in communication between these tissues in adult worms. We also identified *ape-1* as a gene associated with increases in brood size. Given the small effect size, we cannot rule out that this is a false positive, even though we have applied a very conservative methodology to our statistical analysis.

Some of the genes in our genotype dataset could also be associated with ovulation rate in *fog-2*. We selected genes that showed upregulation in *fog-2* mutants in our RNA-seq dataset and tested the effect of RNAi treatment against these genes on lawn-leaving behaviour. We observed large variation in the ovulation rate for the control RNAi, and as a result only a single RNAi treatment was 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 experimental conditions for a single day/event. Therefore, it is likely the source of variation in our screen was systemic and of equal magnitude across replicates and RNAi strains. Our screen identified a single hit: *C23H4.4,* a predicted carboxyl ester lipase with unknown function.

We also performed lawn-leaving assays because previous research has reported differences in hermaphrodite wild-type worms and virgin *fog-2* leaving rates, and these differences are believed to be associated with differences in the status of the reproductive system [15]. 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, and 9 genes that 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 behavior 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 back-

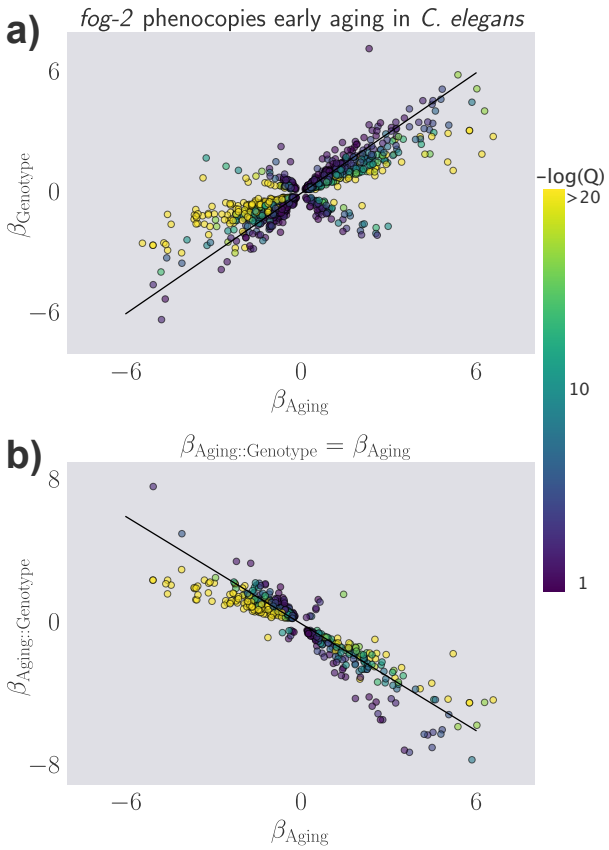

Figure 3. *fog-2* partially phenocopies early aging in *C. elegans*. The β in each axes is the regression coefficient from the GLM, and can be loosely interpreted as an estimator of the log-fold change. Feminization by loss of *fog-2* is associated with a transcriptomic phenotype involving 1,881 genes. 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.** We identified genes that change similarly in feminization as in aging. Black line is the line $y = x$, not a line of best fit. **b.** Many of the genes in **a** do not further change with age in a *fog-2* animal. The black line is the line $y = -x$, and is not a line of best fit. Color is proportional to $-\log Q$. An interactive version of these graphs can be found on our website.

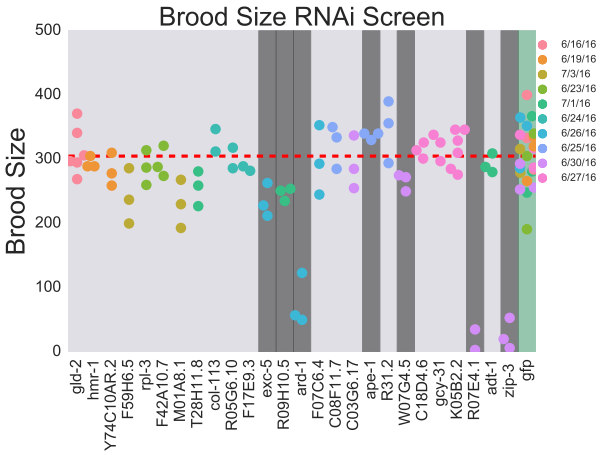

Figure 4. Brood size screen results. *gfp* control RNAi is shaded in green. Results that are statistically significantly different from *gfp* are shaded in dark grey. Dotted red line shows the mean of the pooled *gfp* controls. Points are colored by the date the assay was started on.

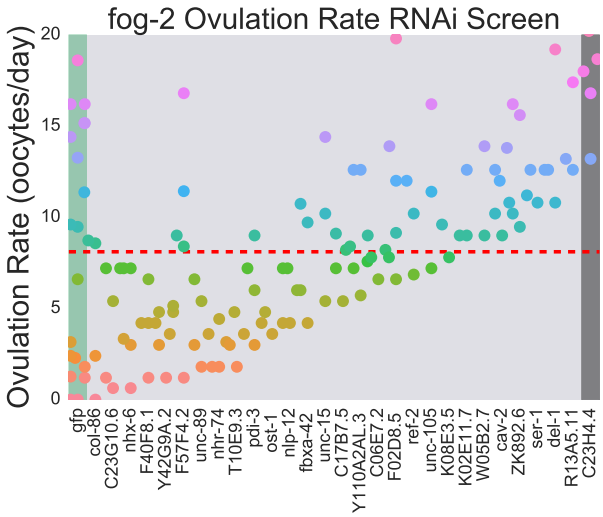

Figure 5. Ovulation Rate Assay. *gfp* control RNAi is shaded in green. Results that are statistically significantly different from *gfp* RNAi are shaded in dark grey. Dotted red line shows the mean of the pooled *gfp* RNAi controls. Points are colored proportionally to their y-coordinate

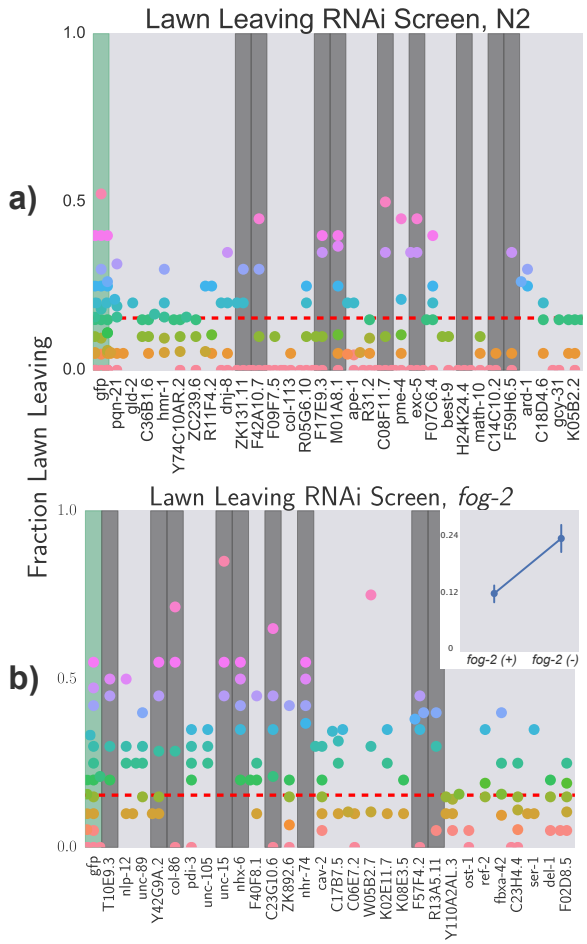

Figure 6. Lawn-leaving Screen Results. *gfp* control RNAi is shaded in green. Results that are statistically significantly different from *gfp* are shaded in dark grey. Dotted red line shows the mean of the pooled *gfp* controls. Points are colored proportionally to their y-coordinate. **a** N2 lawn-leaving 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 [15].

ground. All other hits stimulated lawn-leaving.

We also looked for correlations between lawn-leaving phenotypes and ovulation rate in RNAi treated *fog-2* worms. This was possible because we used the same worms for both assays. However, we did not find any correlation between the lawn-leaving rate of a worm and ovulation rate (data not shown).

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 the highest-quality description of the transcriptomic changes that occur in any aging population of *C. elegans* yet published. Although our data only capture $\sim 50\%$ of the expression changes reported in earlier aging transcriptome literature, this disagreement can be explained by a difference in methodology; earlier publications typically addressed the aging of fertile wild-type hermaphrodites only indirectly, or queried aging animals at a much later stage of their life cycle.

Measurement of a female state

We set out to study the self-fertilizing (hermaphroditic) to self-sterile (female) transition by comparing wild-type animals with *fog-2* mutants as they aged. In so doing, we were able to uncouple reproductive changes from the aging processes that self-sterile and fertile animals experience in common. A key prediction that emerges naturally from postulating the female state is that loss of the hermaphroditic stage should not impact the female-specific transcriptome. Moreover, as long as a female worm remains in this state, the transcriptome should reflect this permanence. Therefore, the spermless female state should have a transcriptomic

signature, regardless of what state preceded the female state (the L4 stage or a hermaphroditic stage) and regardless of the worm's age. Indeed, our results indicate the existence of a discrete female state. In fact, genes associated with the female were largely invariant through time. In a sense, the wild-type worms caught up to their *fog-2* counterparts.

Developmental Factors Change Between 1st to 6th Day of Adulthood in *C. elegans*

Our transcriptomes reveal a host of transcription factors that changed between the 1st and 6th day of adulthood in *C. elegans*. Many of these transcription factors have been associated with development via cellular differentiation and specification. For example, we identified the transcription factor *lin-32*, which has been associated with neuron development [4,23,33]; the Six5 ortholog *unc-39* has been associated with axonal pathfinding and neuron migration [32]; *cnd-1*, a homolog of the vertebrate NeuroD transcription factors, is expressed in the early embryo and is also involved in axon guidance [26]. Why these transcription factors turn on is a mystery, but we speculate that this hints at important neuronal changes that are taking place. Such changes might not be unexpected, given changes in behaviour and pheromone production that are known to occur as hermaphrodites age [14,19].

Our explorations have shown that the loss of *fog-2* partially phenocopies the endogenous female state in *C. elegans*. 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 production and behavioral pathways, although we have been unable to find these genes. 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 physiologic changes could reasonably be the same as the timescale of onset of embryonic transcription. All in all, our research supports the idea that wide-ranging transcriptomic effects of aging in various tissues can be observed well before onset of mortality, and that the *C. elegans* continues to develop at this stage of its life as it enters a new stage of its life cycle.

Acknowledgements

We thank the *Caenorhabditis* Genetics Center for providing worm strains. This work would not be possible without the central repository of *C. elegans* information generated by WormBase, without which mining the genetic data would not have been possible. D.H.W.L. was supported by a National Institutes of Health US Public Health Service Training Grant (T32GM07616). This research was supported by the Howard Hughes Medical Institute, for which P.W.S. is an investigator.

Author Contributions:

DA, DHWL and PWS designed all experiments. DHWL and THK collected RNA for library preparation. IA generated libraries and performed sequencing. DA performed all bioinformatics and statistical analyses. DA, TT and DHWL performed all screens. DA, DHWL and PWS wrote the paper.

References

1. David Angeles-Albores, Raymond Y. N. Lee, Juancarlos Chan, and Paul W. Sternberg. Tissue enrichment analysis for C. elegans genomics. *BMC Bioinformatics*, 17(1):366, 2016.

2. Bokeh Development Team. Bokeh: Python library for interactive visualization. 2014.

3. Nicolas L. Bray, Harold Pimentel, Páll Melsted, and Lior Pachter. Near-optimal probabilistic RNA-seq quantification. *Nature biotechnology*, 34(5):525–7, 2016.

4. Martin Chalfie and Macy Au. Genetic control of differentiation of the Caenorhabditis elegans touch receptor neurons. *Science (New York, N.Y.)*, 243(4894 Pt 1):1027–33, 1989.

5. Robert Clifford, M H Lee, Sudhir Nayak, Mitsue Ohmachi, Flav Giorgini, and Tim Schedl. FOG-2, a novel F-box containing protein, associates with the GLD-1 RNA binding protein and directs male sex determination in the C. elegans hermaphrodite germline. *Development (Cambridge, England)*, 127(24):5265–5276, 2000.

6. Ana S. DePina, Wendy B. Iser, Sung-Soo Park, Stuart Maudsley, Mark A. Wilson, and Catherine A. Wolkow. Regulation of Caenorhabditis elegans vitellogenesis by DAF-2/IIS through separable transcriptional and posttranscriptional mechanisms. *BMC physiology*, 11:11, 2011.

7. D. Mark Eckley, Salim Rahimi, Sandra Mantilla, Nikita V. Orlov, Christopher E. Coletta, Mark A. Wilson, Wendy B. Iser, John D. Delaney, Yongqing Zhang, William Wood, Kevin G. Becker, Catherine A. Wolkow, and Ilya G. Goldberg. Molecular characterization of the transition to mid-life in Caenorhabditis elegans. *Age*, 35(3):689–703, 2013.

8. L. René Garcia, Pinky Mehta, and Paul W. Sternberg. Regulation of distinct muscle behaviors controls the C. elegans male's copulatory spicules during mating. *Cell*, 107(6):777–788, 2001.

9. Julius Halaschek-Wiener, Jaswinder S. Khattra, Sheldon Mckay, Anatoli Pouzyrev, Jeff M. Stott, George S. Yang, Robert A. Holt, Steven J. M. Jones, Marco A. Marra, Angela R. Brooks-Wilson, and Donald L. Riddle. Analysis of long-lived C. elegans daf-2 mutants using serial analysis of gene expression. *Genome Research*, pages 603–615, 2005.

10. John D. Hunter. Matplotlib: A 2D graphics environment. *Computing in Science and Engineering*, 9(3):99–104, 2007.

11. Eric Jones, Travis Oliphant, Pearu Peterson, and Others. {SciPy}: Open source scientific tools for {Python}. *Computing in Science and Engineering*, 9:10–20, 2001.

12. Ravi S. Kamath, Andrew G. Fraser, Yan Dong, Gino Poulin, Richard Durbin, Monica Gotta, Alexander Kanapin, Nathalie Le Bot, Sergio Moreno, Marc Sohrmann, David P. Welchman, Peder Zipperlen, and Julie Ahringer. Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. *Nature*, 421(6920):231–7, 2003.

13. Ravi S. Kamath, Maruxa Martinez-Campos, Peder Zipperlen, Andrew G. Fraser, and Julie Ahringer. Effectiveness of specific RNA-mediated interference through ingested double-stranded RNA in Caenorhabditis elegans. *Genome biology*, 2(1):RESEARCH0002, 2001.

14. Daniel H. W. Leighton, Andrea Choe, Shannon Y. Wu, and Paul W. Sternberg. Communication between oocytes and somatic cells regulates volatile pheromone production in Caenorhabditis elegans. *Proceedings of the National Academy of Sciences*, 111(50):17905–17910, 2014.

15. Jonathan Lipton, Gunnar Kleeman, Rajarshi Ghosh, Robyn Lint, and Scott W. Emmons. Mate Searching in Caenorhabditis elegans: A Genetic Model for Sex Drive in a Simple Invertebrate. *Journal of Neuroscience*, 24(34):7427–7434, 2004.

16. James Lund, Patricia Tedesco, Kyle Duke, John Wang, Stuart K. Kim, and Thomas E. Johnson. Transcriptional profile of aging in C. elegans. *Current Biology*, 12(18):1566–1573, 2002.

17. Mark McCormick, Kan Chen, Priya Ramaswamy, and Cynthia Kenyon. New genes that extend Caenorhabditis elegans' lifespan in response to reproductive signals. *Aging Cell*, 11(2):192–202, 2012.

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

19. Natalia S. Morsci, Leonard A. Haas, and Maureen M. Barr. Sperm status regulates sexual attraction in Caenorhabditis elegans. *Genetics*, 189(4):1341–1346, 2011.

20. Coleen T. Murphy, Steven A. McCarroll, Cornelia I. Bargmann, Andrew Fraser, Ravi S. Kamath, Julie Ahringer, Hao Li, and Cynthia Kenyon. Genes that act downstream of DAF-16 to influence the lifespan of Caenorhabditis elegans. *Nature*, 424(6946):277–283, 2003.

21. F. Pérez and B.E. Granger. IPython: A System for Interactive Scientific Computing Python: An Open and General- Purpose Environment. *Computing in Science and Engineering*, 9(3):21–29, 2007.

22. Harold J Pimentel, Nicolas L. Bray, Suzette Puente, Páll Melsted, and Lior Pachter. Differential analysis of RNA-Seq incorporating quantification uncertainty. *bioRxiv*, page 058164, 2016.

23. D S. Portman and Scott W. Emmons. The basic helix-loop-helix transcription factors LIN-32 and HLH-2 function together in multiple steps of a C. elegans neuronal sublineage. *Development (Cambridge, England)*, 127(24):5415–5426, 2000.

24. John S. Reece-Hoyes, Bart Deplancke, Jane Shingles, Christian A. Grove, Ian A. Hope, and Albertha J. M. Walhout. A compendium of Caenorhabditis elegans regulatory transcription factors: a resource for mapping transcription regulatory networks. *Genome biology*, 6(13):R110, 2005.

25. Tim Schedl and Judith Kimble. fog-2, a germ-line-specific sex determination gene required for hermaphrodite spermatogenesis in Caenorhabditis elegans. *Genetics*, 119(1):43–61, 1988.

26. Caroline Schmitz, Parag Kinge, and Harald Hutter. Axon guidance genes identified in a large-scale RNAi screen using the RNAi-hypersensitive Caenorhabditis elegans strain nre-1(hd20) lin-15b(hd126). *Proceedings of the National Academy of Sciences of the United States of America*, 104(3):834–9, 2007.

27. Femke Simmer, Celine Moorman, Alexander M. Van Der Linden, Ewart Kuijk, Peter V E Van Den Berghe, Ravi S. Kamath, Andrew G. Fraser, Julie Ahringer, and Ronald H A Plasterk. Genome-wide RNAi of C. elegans using the hypersensitive rrf-3 strain reveals novel gene functions. *PLoS Biology*, 1(1):77–84, 2003.

28. J. E. Sulston and S. Brenner. The DNA of Caenorhabditis elegans. *Genetics*, 77(1):95–104, 1974.

29. Stéfan Van Der Walt, S. Chris Colbert, and Gaël Varoquaux. The NumPy array: A structure for efficient numerical computation. *Computing in Science and Engineering*, 13(2):22–30, 2011.

30. Michael Waskom, Samuel St-Jean, Constantine Evans, Jordi Warmenhoven, Kyle Meyer, Marcel Martin, Luc Rocher, Paul Hobson, Pete Bachant, Tamas Nagy, Daniel Wehner, Olga Botvinnik, Tobias Megies, Saulius Lukauskas, Drewokane, Erik Ziegler, Tal Yarkoni, Alistair Miles, Antony Lee, Luis Pedro Coelho, Yaroslav Halchenko, Tom Augspurger, Gregory Hitz, Jake Vanderplas, Clark Fitzgerald, John B. Cole, Gkunter, Santi Villalba, Stephan Hoyer, and Eric Quintero. seaborn: v0.7.0 (January 2016). 2016.

31. Ana White, Abegail Fearon, and Casonya M Johnson. HLH-29 regulates ovulation in C. elegans by targeting genes in the inositol triphosphate signaling pathway. *Biology open*, 1(3):261–8, 2012.

32. Judith L. Yanowitz, M. Afaq Shakir, Edward Hedgecock, Harald Hutter, Andrew Z.

Fire, and Erik A. Lundquist. UNC-39, the C. elegans homolog of the human myotonic dystrophy-associated homeodomain protein Six5, regulates cell motility and differentiation. *Developmental Biology*, 272(2):389–402, 2004.

33. C Zhao and S W Emmons. A transcription factor controlling development of peripheral sense organs in C. elegans., 1995.