

# The *C. elegans* female state: Decoupling the transcriptomic effects of aging and sperm-status

David Angeles-Albores<sup>1, †</sup>      Daniel H.W. Leighton<sup>1, 2, †</sup>      Tiffany Tsou<sup>1</sup>  
Tiffany H. Khaw<sup>1</sup>      Igor Antoshechkin<sup>3</sup>      Paul W. Sternberg<sup>1,\*</sup>

March 9, 2017

† 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** Current: 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 thought 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 as intensely studied as other parts of the life cycle, despite documented changes in behavior and metabolism that occur at this stage. To study the female state of *C. elegans*, we measured the transcriptomes of 1st day adult hermaphrodites; endogenous, 6th day adult females; and at the same time points, mutant *fog-2(lf)* worms that have a feminized germline phenotype. At these time points, we could separate the effects of biological aging from the transition into the female state. *fog-2(lf)* mutants partially phenocopy 6 day adult wild-type animals and exhibit fewer differentially expressed genes as they age throughout these 6 days. Therefore, *fog-2* is epistatic to age as assessed by this transcriptomic phenotype, which indicates that both factors act on sperm status to mediate entry into the female state. 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.

## 1 Introduction

2 Transcriptome analysis by RNA-seq [1] has allowed  
3 for indepth analysis of gene expression changes be-

4 tween life stages and environmental conditions in  
5 many species [2, 3]. *Caenorhabditis elegans*, a ge-  
6 netic model nematode with extremely well defined  
7 and largely invariant development [4,5], has been sub-

jected to extensive transcriptomic analysis across all stages of larval development [6–8] and many stages of embryonic development [7]. Although RNA-seq was used to develop transcriptional profiles of the mammalian aging process soon after its invention [9], few such studies have been conducted in *C. elegans* past the entrance into adulthood.

A distinct challenge to the study of aging transcriptomes in *C. elegans* is the hermaphroditic lifestyle of wild-type individuals of this species. Young adult hermaphrodites are capable of self-fertilization [10, 11], and the resulting embryos will contribute RNA to whole-organism RNA extractions. Most previous attempts to study the *C. elegans* aging transcriptome have addressed the aging process only indirectly, or relied on the use of genetically or chemically sterilized animals to avoid this problem [7, 12–17]. In addition, most of these studies obtained transcriptomes using microarrays, which are less accurate than RNA-seq, especially for low-expressed genes [18].

Here, we investigate what we argue is a distinct state in the *C. elegans* life cycle, the endogenous female state. Although *C. elegans* hermaphrodites emerge into adulthood already replete with sperm, after about 3 days of egg-laying the animals become sperm-depleted and can only reproduce by mating. This marks a transition into what we define as the endogenous female state. This state is behaviorally distinguished by increased male-mating success [19], which may be due to an increased attractiveness to males [20]. This increased attractiveness acts at least partially through production of volatile chemical cues [21]. These behavioral changes are also coincident with functional deterioration of the germline [22], muscle [23], intestine [24] and nervous system [25], changes traditionally attributed to the aging process [26].

To decouple the effects of aging and sperm-loss, we devised a two factor experiment. 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 used *fog-2(lf)* mu-

tants to obtain feminized animals. *fog-2* is involved in germ-cell sex determination in the hermaphrodite worm and is required for sperm production [27, 28].

*C. elegans* defective in sperm formation will never transition into or out of 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.

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 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 causes transcriptomic changes consisting of 5,592 genes in *C. elegans*. 4,552 of these changes occur in both genotypes we studied, indicating they do not depend on life history or genotype. To facilitate exploration of the data, we have generated a website where we have deposited additional graphics, as well as all of the code used to generate these analyses: [https://wormlabcaltech.github.io/Angeles\\_Leighton\\_2016/](https://wormlabcaltech.github.io/Angeles_Leighton_2016/).

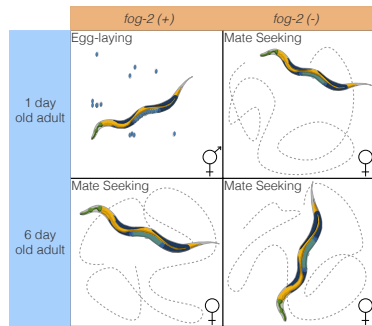
## Materials and Methods

### 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 [10]. 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



**Figure 1.** Experimental design to identify genes associated with sperm loss and with aging. Studying the wild-type worm alone would measure time- and sperm-related changes at the same time, without allowing us to separate these changes. Studying the wild-type worm and a *fog-2(lf)* mutant would enable us to measure sperm-related changes but not time-related changes. By mixing both designs, we can measure and separate both modules.

94 out according to protocols described previously [21].  
 95 1,000–5,000 worms from each replicate were rinsed  
 96 into a microcentrifuge tube in S basal (5.85g/L NaCl,  
 97 1g/L K<sub>2</sub>HPO<sub>4</sub>, 6g/L KH<sub>2</sub>PO<sub>4</sub>), and then spun down  
 98 at 14,000rpm for 30s. The supernatant was removed  
 99 and 1mL of TRIzol was added. Worms were lysed  
 100 by vortexing for 30 s at room temperature and then  
 101 20 min at 4°. The TRIzol lysate was then spun down  
 102 at 14,000rpm for 10 min at 4°C to allow removal of  
 103 insoluble materials. Thereafter the Ambion TRIzol  
 104 protocol was followed to finish the RNA extraction  
 105 (MAN0001271 Rev. Date: 13 Dec 2012). 3 biological  
 106 replicates were obtained for each genotype and  
 107 each time point.

## 108 RNA-Seq

109 RNA integrity was assessed using RNA 6000 Pico Kit  
 110 for Bioanalyzer (Agilent Technologies #5067–1513)  
 111 and mRNA was isolated using NEBNext Poly(A)  
 112 mRNA Magnetic Isolation Module (New England Bi-  
 113 olabs, NEB, #E7490). RNA-Seq libraries were con-  
 114 structed using NEBNext Ultra RNA Library Prep  
 115 Kit for Illumina (NEB #E7530) following manufac-  
 116 turer’s instructions. Briefly, mRNA isolated from

~ 1μg of total RNA was fragmented to the average 117  
 size of 200nt by incubating at 94°C for 15 min in first 118  
 strand buffer, cDNA was synthesized using random 119  
 primers and ProtoScript II Reverse Transcriptase fol- 120  
 lowed by second strand synthesis using Second Strand 121  
 Synthesis Enzyme Mix (NEB). Resulting DNA frag- 122  
 ments were end-repaired, dA tailed and ligated to 123  
 NEBNext hairpin adaptors (NEB #E7335). After 124  
 ligation, adaptors were converted to the ‘Y’ shape 125  
 by treating with USER enzyme and DNA fragments 126  
 were size selected using Agencourt AMPure XP beads 127  
 (Beckman Coulter #A63880) to generate fragment 128  
 sizes between 250 and 350 bp. Adaptor-ligated DNA 129  
 was PCR amplified followed by AMPure XP bead 130  
 clean up. Libraries were quantified with Qubit ds- 131  
 DNA HS Kit (ThermoFisher Scientific #Q32854) and 132  
 the size distribution was confirmed with High Sensi- 133  
 tivity DNA Kit for Bioanalyzer (Agilent Technologies 134  
 #5067–4626). Libraries were sequenced on Illumina 135  
 HiSeq2500 in single read mode with the read length 136  
 of 50nt following manufacturer’s instructions. Base 137  
 calls were performed with RTA 1.13.48.0 followed by 138  
 conversion to FASTQ with bcl2fastq 1.8.4. 139

## 140 Statistical Analysis

### 141 RNA-Seq Analysis

142 RNA-Seq alignment was performed using  
 143 Kallisto [29] with 200 bootstraps. The com-  
 144 mands used for read-alignment are in the S.I. file 1.  
 145 Differential expression analysis was performed using  
 146 Sleuth [30]. The following General Linear Model  
 147 (GLM) was fit:

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

where  $y_i$  are the TPM counts for the  $i$ th gene;  $\beta_{0,i}$  148  
 is the intercept for the  $i$ th gene, and  $\beta_{X,i}$  is the re- 149  
 gression coefficient for variable  $X$  for the  $i$ th gene; 150  
 $A$  is a binary age variable indicating 1st day adult 151  
 (0) or 6th day adult (1) and  $G$  is the genotype vari- 152  
 able indicating wild-type (0) or *fog-2(lf)* (1);  $\beta_{A::G,i}$  153  
 refers to the regression coefficient accounting for the 154  
 interaction between the age and genotype variables 155

156 in the  $i$ th gene. Genes were called significant if the  
157 FDR-adjusted q-value for any regression coefficient  
158 was less than 0.1. Our script for differential analysis  
159 is available on GitHub.

160 Regression coefficients and TPM counts were pro-  
161 cessed using Python 3.5 in a Jupyter Notebook [31].  
162 Data analysis was performed using the Pandas,  
163 NumPy and SciPy libraries [32–34]. Graphics  
164 were created using the Matplotlib and Seaborn li-  
165 braries [35, 36]. Interactive graphics were generated  
166 using Bokeh [37].

167 Tissue, Phenotype and Gene Ontology Enrichment  
168 Analyses (TEA, PEA and GEA, respectively) were  
169 performed using the WormBase Enrichment Suite for  
170 Python [38, 39].

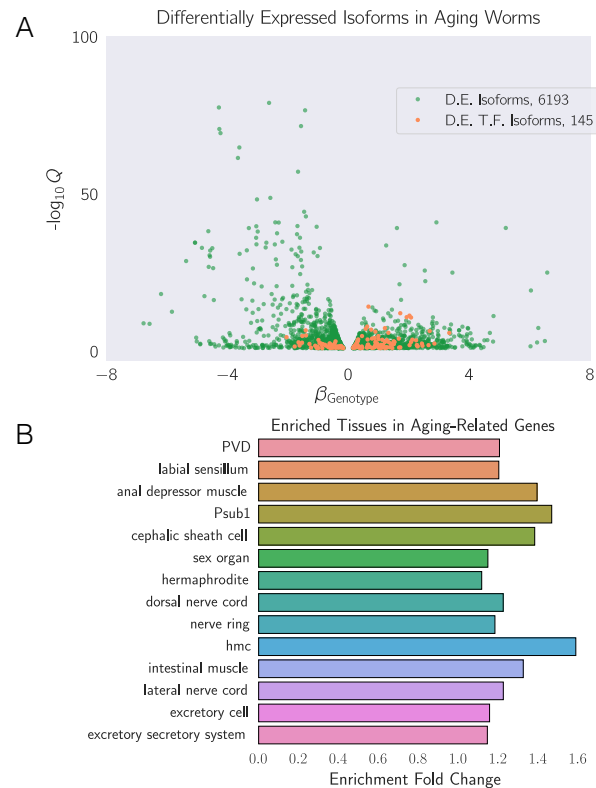
## 171 Data Availability

172 Strains are available from the *Caenorhabditis* Genet-  
173 ics Center. All of the data and scripts pertinent  
174 for this project except the raw reads can be found  
175 on our Github repository [https://github.com/](https://github.com/WormLabCaltech/Angeles_Leighton_2016)  
176 [WormLabCaltech/Angeles\\_Leighton\\_2016](https://github.com/WormLabCaltech/Angeles_Leighton_2016). File S1  
177 contains the list of genes that were altered in aging re-  
178 gardless of genotype. File S2 contains the list of genes  
179 and their associations with the *fog-2(lf)* phenotype.  
180 File S3 contains genes associated with the female  
181 state. Raw reads were deposited to the Sequence  
182 Read Archive under the accession code SUB2457229.

## 183 Results and Discussion

### 184 Decoupling time-dependent effects 185 from sperm-status via general linear 186 models

187 In order to decouple time-dependent effects from  
188 changes associated with loss of hermaphroditic  
189 sperm, we measured wild-type and *fog-2(lf)* adults at  
190 the 1st day adult stage (before visible embryos were  
191 present) and 6th day adult stage, when all wild-type  
192 hermaphrodites had laid all their eggs (see Fig 1), but  
193 mortality was still low (< 10%) [41]. We obtained 16–  
194 19 million reads mappable to the *C. elegans* genome  
195 per biological replicate, which enabled us to identify



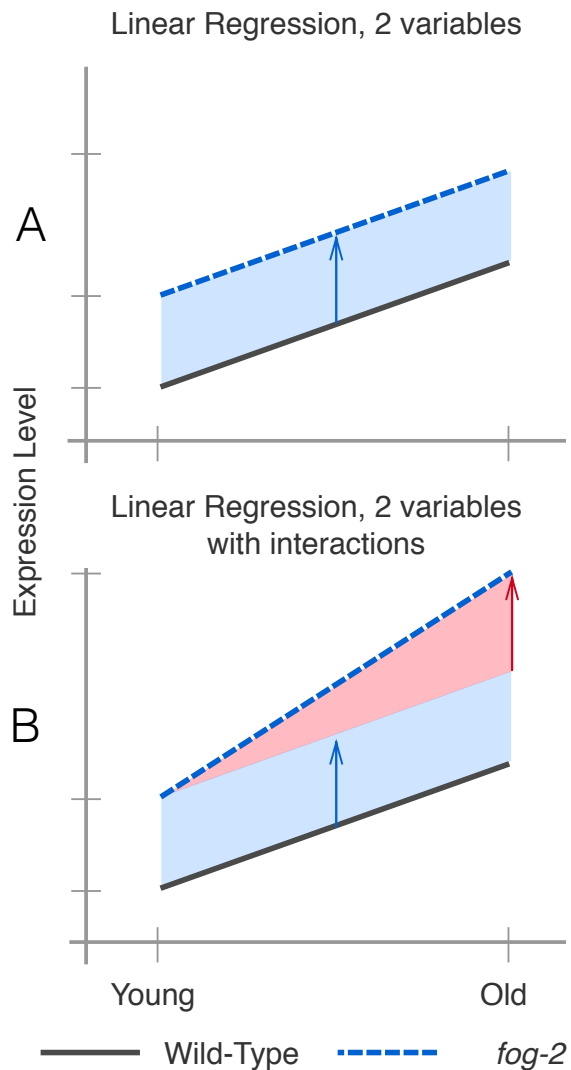
**Figure 2.** **A** We identified a common aging transcriptome between N2 and *fog-2(lf)* animals, consisting of 6,193 differentially expressed isoforms totaling 5,592 genes. The volcano plot is randomly down-sampled 30% for ease of viewing. Each point represents an individual isoform.  $\beta_{\text{Aging}}$  is the regression coefficient. Larger magnitudes of  $\beta$  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 predicted transcription factor genes [40]. An interactive version of this graph can be found on our [website](#). **B** Tissue Enrichment Analysis [38] showed that genes associated with muscle tissues and the nervous system are enriched in aging-related genes. Only statistically significantly enriched tissues are shown. Enrichment Fold Change is defined as *Observed/Expected*. hmc stands for head mesodermal cell.

196 14,702 individual genes totalling 21,143 isoforms (see  
197 Figure 2a).

198 One way to analyze the data from this two-factor  
199 design is by pairwise comparison of the distinct  
200 states. However, such an analysis would not make  
201 full use of all the statistical power afforded by this ex-  
202 periment. Another method that makes full use of the  
203 information in our experiment is to perform a linear  
204 regression in 3 dimensions (2 independent variables,  
205 age and genotype, and 1 output). A linear regression  
206 with 1 parameter (age, for example) would fit a line  
207 between expression data for young and old animals.  
208 When a second parameter is added to the linear re-  
209 gression, said parameter can be visualized as altering  
210 the y-intercept, but not the slope, of the first line in  
211 question (see Fig. 3a).

212 Although a simple linear model is oftentimes use-  
213 ful, sometimes it is not appropriate to assume that  
214 the two variables under study are entirely independ-  
215 ent. For example, in our case, three out of the four  
216 timepoint-and-genotype combinations we studied did  
217 not have sperm, and sperm-status is associated with  
218 both the *fog-2(lf)* self-sterile phenotype and with bi-  
219 ological age of the wild-type animal. One way to statis-  
220 tically model such correlation between variables is to  
221 add an interaction term to the linear regression. This  
222 interaction term allows extra flexibility in describing  
223 how changes occur between conditions. For exam-  
224 ple, suppose a given theoretical gene *X* has expression  
225 levels that increase in a *fog-2*-dependent manner, but  
226 also increases in an age-dependent manner. However,  
227 aged *fog-2(lf)* animals do not have expression levels  
228 of *X* that would be expected from adding the effect of  
229 the two perturbations; instead, the expression levels  
230 of *X* in this animal are considerably above what is  
231 expected. In this case, we could add a positive inter-  
232 action coefficient to the model to explain the effect  
233 of genotype on the y-intercept as well as the slope  
234 (see Fig. 3b). When the two perturbations are loss-  
235 of-function mutations, such interactions are epistatic  
236 interactions.

237 For these reasons, we used a linear generalized  
238 model (see [Statistical Analysis](#)) with interactions to  
239 identify a transcriptomic profile associated with the  
240 *fog-2(lf)* genotype independently of age, as well as  
241 a transcriptomic profile of *C. elegans* aging common



**Figure 3.** **A.** A linear regression with two variables, age and genotype. The expression level of a gene increases by the same amount as worms age regardless of genotype. However, *fog-2(lf)* has more mRNA than the wild-type at all stages (blue arrow). **B.** A linear regression with two variables and an interaction term. In this example, the expression level of this hypothetical gene is different between wild-type worms and *fog-2(lf)* (blue arrow). Although the expression level of this gene increases with age, the slope is different between wild-type and *fog-2(lf)*. The difference in the slope can be accounted for through an interaction coefficient (red arrow).

242 to both genotypes. The change associated with each  
243 variable is referred as  $\beta$ ; this number, although re-  
244 lated to the natural logarithm of the fold change, is  
245 not equal to it. However, it is true that larger mag-  
246 nitudes of  $\beta$  indicate greater change. Thus, for each  
247 gene we performed a linear regression, and we eval-  
248 uated the whether the  $\beta$  values associated with each  
249 coefficient were significantly different from 0 via a  
250 Wald test corrected for multiple hypothesis testing.  
251 A coefficient was considered to be significantly differ-  
252 ent from 0 if the q-value associated with it was less  
253 than 0.1.

## 254 A quarter of all genes change expres- 255 sion between the 1st day of adulthood 256 and the 6th day of adulthood in *C. el-* 257 *egans*.

258 We identified a transcriptomic signature consisting  
259 of 5,592 genes that were differentially expressed in  
260 6th day adult animals of either genotype relative to  
261 1st day adult animals (see SI file 2). This constitutes  
262 more than one quarter of the genes in *C. elegans*. Tis-  
263 sue Enrichment Analysis (TEA) [38] showed that ner-  
264 vous tissues including the ‘nerve ring’, ‘dorsal nerve  
265 cord’, ‘PVD’ and ‘labial sensillum’ were enriched in  
266 genes that become differentially expressed through  
267 aging. Likewise, certain muscle groups (‘anal depres-  
268 sor muscle’, ‘intestinal muscle’) were enriched. (see  
269 Figure 2b). Gene Enrichment Analysis (GEA) [39]  
270 revealed that genes that were differentially expressed  
271 during the course of aging were enriched in terms  
272 involving respiration (‘respiratory chain’, ‘oxoacid  
273 metabolic process’); translation (‘cytosolic large ribo-  
274 somal subunit’); and nucleotide metabolism (‘purine  
275 nucleotide’, ‘nucleoside phosphate’ and ‘ribose phos-  
276 phate’ metabolic process). Phenotype Enrichment  
277 Analysis (PEA) [39] showed enrichment of pheno-  
278 types that affect the *C. elegans* gonad, including ‘go-  
279 nad vesiculated’, ‘gonad small’, ‘oocytes lack nucleus’  
280 and ‘rachis narrow’.

281 To verify the quality of our dataset, we generated a  
282 list of 1,056 golden standard genes expected to be al-  
283 tered in 6th day adult worms using previous literature  
284 reports including downstream genes of *daf-12*, *daf-16*,

285 and aging and lifespan extension datasets [12–16].  
286 Out of 1,056 standard genes, we found 506 genes in  
287 our time-responsive dataset. This result was statisti-  
288 cally significant with a p-value  $< 10^{-38}$ .

289 Next, we used a published compendium [40] to  
290 search for known or predicted transcription factors.  
291 We found 145 transcription factors in the set of  
292 genes with differential expression in aging nema-  
293 todes. We subjected this list of transcription fac-  
294 tors to TEA to understand their expression patterns.  
295 6 of these transcription factors were expressed in  
296 the ‘hermaphrodite specific neuron’ (HSN), a neuron  
297 physiologically relevant for egg-laying (*hlh-14*, *sem-4*,  
298 *ceh-20*, *egl-46*, *ceh-13*, *hlh-3*), which represented a sta-  
299 tistically significant 2-fold enrichment of this tissue  
300 ( $q < 10^{-1}$ ). The term ‘head muscle’ was also over-  
301 represented at twice the expected level ( $q < 10^{-1}$ ,  
302 13 genes). Many of these transcription factors have  
303 been associated with developmental processes, and  
304 it is unclear why they would change expression in  
305 adult animals.

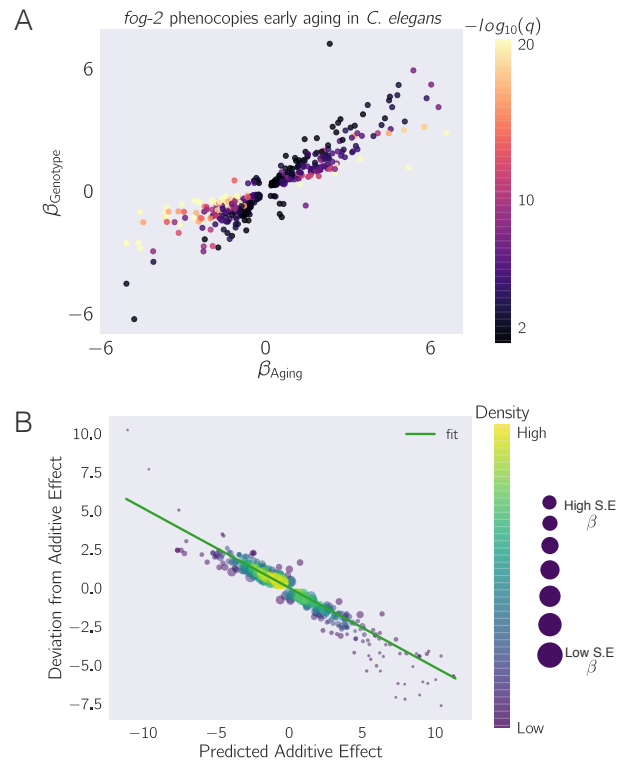
## 306 The whole-organism *fog-2(lf)* tran- 307 sriptome in *C. elegans*.

308 We identified 1,881 genes associated with the *fog-2(lf)*  
309 genotype, including 60 transcription factors (see SI  
310 file 3). TEA showed that the terms ‘AB’, ‘midbody’,  
311 ‘uterine muscle’, ‘cephalic sheath cell’, ‘anal depres-  
312 sor muscle’ and ‘PVD’ were enriched in this gene set.  
313 The terms ‘AB’ and ‘midbody’ likely reflect the im-  
314 pact of *fog-2(lf)* on the germline. Phenotype enrich-  
315 ment showed that only a single phenotype, ‘spindle  
316 orientation variant’ was enriched in the *fog-2(lf)* tran-  
317 scriptome ( $q < 10^{-1}$ , 38 genes, 2-fold enrichment).  
318 Most genes annotated as ‘spindle orientation vari-  
319 ant’ were slightly upregulated, and therefore are un-  
320 likely to uniquely reflect reduced germline prolifera-  
321 tion. GO term enrichment was very similar to the ag-  
322 ing gene set and reflected enrichment in annotations  
323 pertaining to translation and respiration. Unlike the  
324 aging gene set, the *fog-2(lf)* transcriptome was sig-  
325 nificantly enriched in ‘myofibril’ and ‘G-protein cou-  
326 pled receptor binding’ ( $q < 10^{-1}$ ). Enrichment of the  
327 term ‘G-protein coupled receptor binding’ was due to  
328 14 genes: *cam-1*, *mom-2*, *dsh-1*, *spp-10*, *flp-6*, *flp-7*,

329 *flp-9*, *flp-13*, *flp-14*, *flp-18*, *K02A11.4*, *nlp-12*, *nlp-13*,  
330 and *nlp-40*. *dsh-1*, *mom-2* and *cam-1* are members  
331 of the Wnt signaling pathway. Most of these genes'  
332 expression levels were up-regulated, suggesting in-  
333 creased G-protein binding activity in *fog-2(lf)* mu-  
334 tants.

### 335 The *fog-2(lf)* transcriptome overlaps 336 significantly with the aging transcrip- 337 tome

338 Of the 1,881 genes that we identified in the *fog-2(lf)*  
339 transcriptome, 1,040 genes were also identified in our  
340 aging set. Moreover, of these 1,040 genes, 905 genes  
341 changed in the same direction in response either aging  
342 or germline feminization. The overlap between these  
343 transcriptomes suggests an interplay between sperm-  
344 status and age. The nature of the interplay should be  
345 captured by the interaction coefficients in our model.  
346 There are four possibilities. First, the *fog-2(lf)* worms  
347 may have a fast-aging phenotype, in which case the  
348 interaction coefficients should match the sign of the  
349 aging coefficient. Second, the *fog-2(lf)* worms may  
350 have a slow-aging phenotype, in which case the inter-  
351 action coefficients should have an interaction coeffi-  
352 cient that is of opposite sign, but not greater in mag-  
353 nitude than the aging coefficient (if a gene increases  
354 in aging in a wild-type worm, it should still increase  
355 in a *fog-2(lf)* worm, albeit less). Third, the *fog-2(lf)*  
356 worms exhibit a rejuvenation phenotype. If this is  
357 the case, then these genes should have an interaction  
358 coefficient that is of opposite sign and greater magni-  
359 tude than their aging coefficient, such that the change  
360 of these genes in *fog-2(lf)* mutant worms is reversed  
361 relative to the wild-type. Finally, if these genes are  
362 indicative of a female state, then these genes should  
363 not change with age in *fog-2(lf)* animals, since these  
364 animals do not exit this state during the course of  
365 the experiment. Moreover, because wild-type worms  
366 become female as they age, a further requirement for  
367 a transcriptomic signature of the female state is that  
368 aging coefficients for genes in this signature should  
369 have genotype coefficients of equal sign and magni-  
370 tude. In other words, entrance into the female state  
371 should be not be path-dependent.

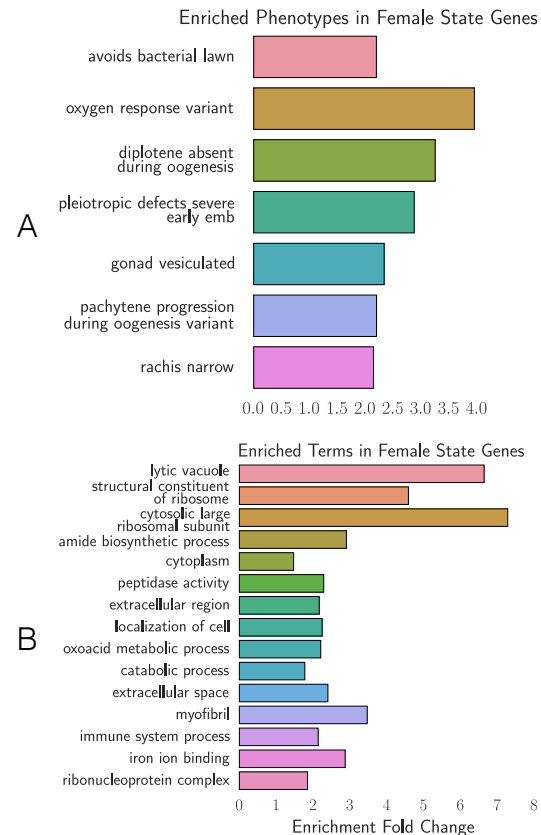


**Figure 4.** *fog-2(lf)* partially phenocopies early aging in *C. elegans*. The  $\beta$  in each axes is the regression coefficient from the GLM, and can be loosely interpreted as an estimator of the log-fold change. Feminization by loss of *fog-2(lf)* is associated with a transcriptomic phenotype 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(lf)* background results in no change in the expression level of these genes. **A** We identified genes that change similarly during feminization and aging. The correlation between feminization and aging is almost 1:1. **B** Epistasis plot of aging versus feminization. Epistasis plots indicate whether two genes (or perturbations) act on the same pathway. When two effects act on the same pathway, this is reflected by a slope of  $-0.5$ . The measured slope was  $-0.51 \pm 0.01$ .

372 To evaluate which of these possibilities was most  
 373 likely, we selected the 1,040 genes that had ag-  
 374 ing, genotype and interaction coefficients significantly  
 375 different from zero and we plotted their temporal  
 376 coefficients against their genotype coefficients (see  
 377 Fig. 4a). We observed that the aging coefficients  
 378 were strongly predictive of the genotype coefficients.  
 379 Most of these genes fell near the line  $y = x$ , sug-  
 380 gesting that these genes define a female state. As a  
 381 further test that these genes actually define a female  
 382 state, we generated an epistasis plot using this gene  
 383 set. We have previously used epistasis plots to mea-  
 384 sure transcriptome-wide epistasis between genes in a  
 385 pathway [42]. Briefly, an epistasis plot plots the ex-  
 386 pected expression of a double perturbation under an  
 387 additive model (null model) on the x-axis, and the  
 388 deviation from this null model in the y-axis. In other  
 389 words, we calculated the x-coordinates for each point  
 390 by adding  $\beta_{\text{Genotype}} + \beta_{\text{Aging}}$ , and the y-coordinates  
 391 are equal to  $\beta_{\text{Interaction}}$  for each isoform. Previously  
 392 we have shown that if two genes act in a linear path-  
 393 way, an epistasis plot will generate a line with slope  
 394 equal to  $-0.5$ . When we generated an epistasis plot  
 395 and found the line of best fit, we observed a slope of  
 396  $-0.51 \pm 0.01$ , which suggests that the *fog-2* gene and  
 397 time are acting to generate a single transcriptomic  
 398 phenotype along a single pathway. Overall, we iden-  
 399 tified 405 genes that increased in the same direction  
 400 through age or mutation of the *fog-2(lf)* gene and that  
 401 had an interaction coefficient of opposite sign to the  
 402 aging or genotype coefficient (see SI file 4). Taken to-  
 403 gether, this information suggests that these 405 genes  
 404 define a female state in *C. elegans*.

## 405 Analysis of the Female State Transcrip- 406 tome

407 To better understand the changes that happen after  
 408 sperm loss, we performed tissue enrichment, pheno-  
 409 type enrichment and gene ontology enrichment anal-  
 410 yses on the set of 405 genes that we associated with  
 411 the female state. TEA showed no tissue enrichment  
 412 using this gene-set. GEA showed that this gene list  
 413 was enriched in constituents of the ribosomal sub-  
 414 units almost four times above background ( $q < 10^{-5}$ ,  
 415 17 genes). The enrichment of ribosomal constituents



**Figure 5.** Phenotype and GO enrichment of genes involved in the female state. **A.** Phenotype Enrichment Analysis. **B.** Gene Ontology Enrichment Analysis. Most of the terms enriched in PEA reflect the abundance of ribosomal subunits present in this gene set.



416 in this gene set in turn drives the enriched pheno- 456  
417 types: ‘avoids bacterial lawn’, ‘diplotene absent dur- 457  
418 ing oogenesis’, ‘gonad vesiculated’, ‘pachytene pro- 458  
419 gression during oogenesis variant’, and ‘rachis nar- 459  
420 row’. The expression of most of these ribosomal sub- 460  
421 units is down-regulated in aged animals or in *fog-2(lf)* 461  
422 mutants. 462

## 423 Discussion 463

### 424 Defining an Early Aging Phenotype 464

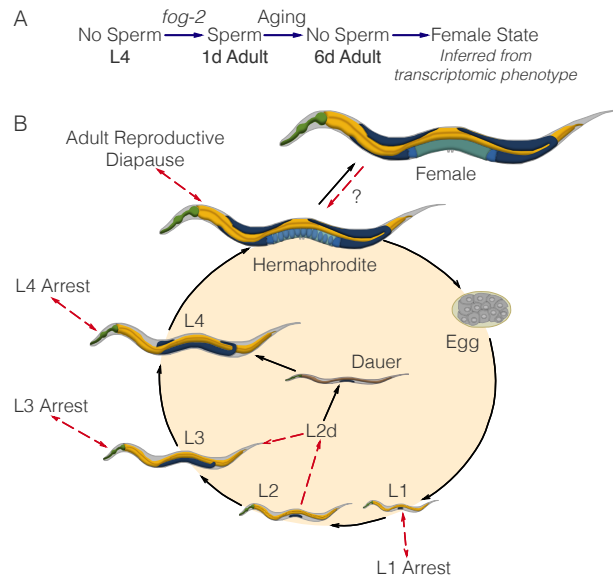
425 Our experimental design enables us to decouple the 465  
426 effects of egg-laying from aging. As a result, we 466  
427 identified a set of almost 4,000 genes that are al- 467  
428 tered similarly between wild-type and *fog-2(lf)* mu- 468  
429 tants. Due to the read depth of our transcrip-  
430 tomic data (20 million reads) and the number of  
431 samples measured (3 biological replicates for 4 dif-  
432 ferent life stages/genotypes), this dataset consti-  
433 tutes a high-quality description of the transcriptomic  
434 changes that occur in aging populations of *C. ele-*  
435 *gans*. Although our data only capture  $\sim 50\%$  of  
436 the expression changes reported in earlier aging tran-  
437 scriptome literature, this disagreement can be ex-  
438 plained by a difference in methodology; earlier pub-  
439 lications typically addressed the aging of fertile wild-  
440 type hermaphrodites only indirectly, or queried aging  
441 animals at a much later stage of their life cycle.

### 442 Measurement of a female state is en- 469 443 abled by linear models 470

444 We set out to study the self-fertilizing 471  
445 (hermaphroditic) to self-sterile (female) transi- 472  
446 tion by comparing wild-type animals with *fog-2(lf)* 473  
447 mutants as they aged. Our computational approach 474  
448 enabled us to separate between two biological pro- 475  
449 cesses that are correlated within samples. Because 476  
450 of this intra-sample correlation, identifying this 477  
451 state via pairwise comparisons would not have been 478  
452 straightforward. Although it is a favored method 479  
453 amongst biologists, such pairwise comparisons suffer 480  
454 from a number of drawbacks. First, pairwise compar- 481  
455 isons are unable to draw on the full statistical power 482

456 available to an experiment because they discard 457  
458 almost all information except the samples being 459  
460 compared. Second, pairwise comparisons require a 461  
462 researcher to define *a priori* which comparisons are 463  
464 informative. For experiments with many variables, 465  
466 the number of pairwise combinations is explosively 467  
468 large. Indeed, even for this two-factor experiment, 469  
470 there are 6 possible pairwise comparisons. On the 471  
472 other hand, by specifying a linear regression model, 473  
474 each gene can be summarized with three variables, 475  
476 each of which can be analyzed and understood 477  
478 without the need to resort to further pairwise 479  
480 combinations. 481

469 Our explorations have shown that the loss of 470  
471 *fog-2(lf)* partially phenocopies the transcriptional 472  
473 events that occur naturally as *C. elegans* ages from 474  
475 the 1st day of adulthood to the 6th day of adult- 476  
477 hood. Moreover, epistasis analysis of these pertur- 478  
479 bations suggest that they act on the same path- 480  
481 way, namely sperm generation and depletion (see 482  
483 Fig. 6). Sperm generation promotes a non-female 484  
485 states, whereas sperm depletion causes entry into the 486  
487 female state. Given the enrichment of neuronal tran- 488  
489 scription factors that are associated with sperm loss 489  
490 in our dataset, we believe this dataset should contain 490  
491 some of the transcriptomic modules that are involved 491  
492 in these pheromone production and behavioral path- 492  
493 ways, although we have been unable to find these 493  
494 genes. Currently, we cannot judge how many of the 494  
495 changes induced by loss of hermaphroditic sperm are 495  
496 developmental (i.e., irreversible), and how many can 496  
497 be rescued by mating to a male. While an entertain- 497  
498 ing thought experiment, establishing whether these 498  
499 transcriptomic changes can be rescued by males is a 499  
500 daunting experimental task, given that the timescales 500  
501 for physiologic changes could reasonably be the same 501  
502 as the timescale of onset of embryonic transcription. 502  
503 All in all, our research supports the idea that wide- 503  
504 ranging transcriptomic effects of aging in various tis- 504  
505 sues can be observed well before onset of mortality, 505  
506 and that *C. elegans* continues to develop as it enters 506  
507 a new state of its life cycle. 507



**Figure 6. A.** A substrate model showing how *fog-2* promotes sperm generation, whereas aging promotes sperm depletion, leading to entry to the female state. Such a model can explain why *fog-2* and aging appear epistatic to each other. **B.** The complete *C. elegans* life cycle. Recognized stages of *C. elegans* are marked by black arrows. States are marked by red arrows to emphasize that at the end of a state, the worm returns to the developmental timepoint it was at before entering the state. The L2d state is an exception. It is the only stage that does not return to the same developmental timepoint; rather, the L2d state is a permissive state that allows entry into either dauer or the L3 stage. We have presented evidence of a female state in *C. elegans*. At this point, it is unclear whether the difference between hermaphrodites and females is reversible by males. Therefore, it remains unclear whether it is a stage or a true state.

## The *C. elegans* life cycle, life stages and life states

*C. elegans* has a complicated life cycle, with two alternative developmental pathways that have multiple stages (larval development and dauer development), followed by reproductive adulthood. In addition to its developmental stages, researchers have recognized that *C. elegans* has numerous life states that it can enter into when given instructive environmental cues. One such state is the L1 arrest state, where development ceases entirely upon starvation [43]. More recently, researchers have described additional diapause states that the worm can access at the L3, L4 and young adult stages under conditions of low food [44–46]. Not all states of *C. elegans* are arrested, however (see Fig. 6). For example, the L2d state is induced by crowded and nutrient poor conditions [47]. While within this state, the worm is capable of entry into either dauer or the L3 larval stage, depending on environmental conditions. Thus, the L2d state is a permissive state, and marks the point at which the nematode development is committed to a single developmental pathway.

Identification of the *C. elegans* life states has often been performed by morphological studies (as in the course of L4 arrest or L2d) or via timecourses (L1 arrest). However, not all states may be visually identifiable, or even if they are, the morphological changes may be very subtle, making positive identification difficult. However, the detailed information afforded by a transcriptome should in theory provide sufficient information to definitively identify a state, since transcriptomic information underlies morphology. Moreover, transcriptomics can provide an informative description into the physiology of complex metazoan life state's via measurements of global gene expression. By identifying differentially expressed genes and using ontology enrichment analyses to identify gene functions, sites of expression or phenotypes that are enriched in a given gene set, researchers can obtain a clearer picture of the changes that occur in the worm in a less biased manner than by identifying gross morphological changes. RNA-seq is emerging as a powerful technology that has been used successfully in the past as a qualitative tool for

543 target acquisition. More recent work has successfully  
544 used RNA-seq to establish genetic interactions be-  
545 tween genes [48, 49]. In this work, we have shown  
546 that whole-organism RNA-seq data can also be ana-  
547 lyzed via a similar formalism to successfully identify  
548 internal states in a multi-cellular organism.

## 549 Acknowledgments

550 We thank the *Caenorhabditis* Genetics Center for  
551 providing worm strains. This work would not be  
552 possible without the central repository of *C. elegans*  
553 information generated by WormBase, without which  
554 mining the genetic data would not have been possi-  
555 ble. DHWL was supported by a National Institutes  
556 of Health US Public Health Service Training Grant  
557 (T32GM07616). This research was supported by the  
558 Howard Hughes Medical Institute, for which PWS is  
559 an investigator.

## 560 Author Contributions:

561 DA, DHWL and PWS designed all experiments.  
562 DHWL and THK collected RNA for library prepara-  
563 tion. IA generated libraries and performed sequenc-  
564 ing. DA performed all bioinformatics and statistical  
565 analyses. DA, TT and DHWL performed all screens.  
566 DA, DHWL and PWS wrote the paper.

## References

1. Mortazavi A, Williams BA, McCue K, Schaefer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* 5(7):621–628.
2. Gerstein MB *et al.* (2014) Comparative analysis of the transcriptome across distant species. *Nature* 512:445–448.
3. Blaxter M, Kumar S, Kaur G, Koutsovoulos G, Elsworth B (2012) Genomics and transcriptomics across the diversity of the Nematoda. *Parasite Immunology* 34(2-3):108–120.
4. Sulston JE, Horvitz HR (1977) Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Developmental Biology* 56(1):110–156.
5. Sulston JE, Schierenberg E, White JG, Thomson JN (1983) The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Developmental Biology* 100(1):64–119.
6. Hillier LW *et al.* (2009) Massively parallel sequencing of the polyadenylated transcriptome of *C. elegans*. *Genome Research* 19(4):657–666.
7. Boeck ME *et al.* (2016) The time-resolved transcriptome of *C. elegans*. *Genome Research* pp. 1–10.
8. Murray JI *et al.* (2012) Multidimensional regulation of gene expression in the *C. elegans* embryo. Multidimensional regulation of gene expression in the *C. elegans* embryo. pp. 1282–1294.
9. Magalhães JD, Finch C, Janssens G (2010) Next-generation sequencing in aging research: emerging applications, problems, pitfalls and possible solutions. *Ageing research reviews* 9(3):315–323.
10. Sulston JE, Brenner S (1974) The DNA of *Caenorhabditis elegans*. *Genetics* 77(1):95–104.
11. Corsi AK, Wightman B, Chalfie M (2015) A transparent window into biology: A primer on *Caenorhabditis elegans*. *Genetics* 200(2):387–407.
12. Murphy CT *et al.* (2003) Genes that act downstream of DAF-16 to influence the lifespan of *Caenorhabditis elegans*. *Nature* 424(6946):277–283.
13. Halaschek-Wiener J *et al.* (2005) Analysis of long-lived *C. elegans* *daf-2* mutants using serial analysis of gene expression. *Genome Research* pp. 603–615.

- 
14. Lund J *et al.* (2002) Transcriptional profile of aging in *C. elegans*. *Current Biology* 12(18):1566–1573.
  15. McCormick M, Chen K, Ramaswamy P, Kenyon C (2012) New genes that extend *Caenorhabditis elegans*' lifespan in response to reproductive signals. *Aging Cell* 11(2):192–202.
  16. Eckley DM *et al.* (2013) Molecular characterization of the transition to mid-life in *Caenorhabditis elegans*. *Age* 35(3):689–703.
  17. Rangaraju S *et al.* (2015) Suppression of transcriptional drift extends *C. elegans* lifespan by postponing the onset of mortality. *eLife* 4(December2015):1–39.
  18. Wang C *et al.* (2014) The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. *Nature biotechnology* 32(9):926–32.
  19. Garcia LR, LeBoeuf B, Koo P (2007) Diversity in mating behavior of hermaphroditic and male-female *Caenorhabditis* nematodes. *Genetics* 175(4):1761–1771.
  20. Morsci NS, Haas LA, Barr MM (2011) Sperm status regulates sexual attraction in *Caenorhabditis elegans*. *Genetics* 189(4):1341–1346.
  21. Leighton DHW, Choe A, Wu SY, Sternberg PW (2014) Communication between oocytes and somatic cells regulates volatile pheromone production in *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences* 111(50):17905–17910.
  22. Andux S, Ellis RE (2008) Apoptosis maintains oocyte quality in aging *Caenorhabditis elegans* females. *PLoS Genetics* 4(12).
  23. Herndon La *et al.* (2002) Stochastic and genetic factors influence tissue-specific decline in ageing *C. elegans*. *Nature* 419(6909):808–814.
  24. McGee MD *et al.* (2011) Loss of intestinal nuclei and intestinal integrity in aging *C. elegans*. *Aging Cell* 10(4):699–710.
  25. Liu J *et al.* (2013) Functional aging in the nervous system contributes to age-dependent motor activity decline in *C. elegans*. *Cell Metabolism* 18(3):392–402.
  26. Golden TR, Melov S (2007) Gene expression changes associated with aging in *C. elegans*. *WormBook : the online review of C. elegans biology* pp. 1–12.
  27. Schedl T, Kimble J (1988) *fog-2*, a germ-line-specific sex determination gene required for hermaphrodite spermatogenesis in *Caenorhabditis elegans*. *Genetics* 119(1):43–61.
  28. Clifford R *et al.* (2000) 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.
  29. Bray NL, Pimentel HJ, Melsted P, Pachter L (2016) Near-optimal probabilistic RNA-seq quantification. *Nature biotechnology* 34(5):525–7.
  30. Pimentel HJ, Bray NL, Puente S, Melsted P, Pachter L (2016) Differential analysis of RNA-Seq incorporating quantification uncertainty. *bioRxiv* p. 058164.
  31. Pérez F, Granger B (2007) IPython: A System for Interactive Scientific Computing Python: An Open and General- Purpose Environment. *Computing in Science and Engineering* 9(3):21–29.
  32. McKinney W (2011) pandas: a Foundational Python Library for Data Analysis and Statistics. *Python for High Performance and Scientific Computing* pp. 1–9.
-

- 
33. Van Der Walt S, Colbert SC, Varoquaux G (2011) The NumPy array: A structure for efficient numerical computation. *Computing in Science and Engineering* 13(2):22–30.
  34. Oliphant TE (2007) SciPy: Open source scientific tools for Python. *Computing in Science and Engineering* 9:10–20.
  35. Waskom M *et al.* (2016) seaborn: v0.7.0 (January 2016).
  36. Hunter JD (2007) Matplotlib: A 2D graphics environment. *Computing in Science and Engineering* 9(3):99–104.
  37. Bokeh Development Team (2014) Bokeh: Python library for interactive visualization.
  38. Angeles-Albores D, N. Lee RY, Chan J, Sternberg PW (2016) Tissue enrichment analysis for *C. elegans* genomics. *BMC Bioinformatics* 17(1):366.
  39. Angeles-Albores D, Lee RY, Chan J, Sternberg PW (2017) Phenotype and gene ontology enrichment as guides for disease modeling in *C. elegans*. *bioRxiv*.
  40. Reece-Hoyes JS *et al.* (2005) A compendium of *Caenorhabditis elegans* regulatory transcription factors: a resource for mapping transcription regulatory networks. *Genome biology* 6(13):R110.
  41. Stroustrup N *et al.* (2013) The *Caenorhabditis elegans* Lifespan Machine. *Nature methods* 10(7):665–70.
  42. Angeles-Albores D, Puckett Robinson C, Williams BA, Sternberg PW (2017) Genetic Analysis of a Metazoan Pathway using Transcriptomic Phenotypes. *bioRxiv*.
  43. Johnson TE, Mitchell DH, Kline S, Kemal R, Foy J (1984) Arresting development arrests aging in the nematode *Caenorhabditis elegans*. *Mechanisms of Ageing and Development* 28(1):23–40.
  44. Angelo G, Gilst MRV (2009) Cells and Extends Reproductive. *Science* 326(November):954–958.
  45. Seidel HS, Kimble J (2011) The oogenic germline starvation response in *c. elegans*. *PLoS ONE* 6(12).
  46. Schindler AJ, Baugh LR, Sherwood DR (2014) Identification of Late Larval Stage Developmental Checkpoints in *Caenorhabditis elegans* Regulated by Insulin/IGF and Steroid Hormone Signaling Pathways. *PLoS Genetics* 10(6):13–16.
  47. Golden JW, Riddle DL (1984) The *Caenorhabditis elegans* dauer larva: Developmental effects of pheromone, food, and temperature. *Developmental Biology* 102(2):368–378.
  48. Dixit A *et al.* (2016) Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* 167(7):1853–1866.e17.
  49. Adamson B *et al.* (2016) A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell* 167(7):1867–1882.e21.
-