Biophysical Journal Volume 109 December 2015 2436–2445

Article

A Transport Model for Estimating the Time Course of ERK Activation in the *C. elegans* Germline

Henry H. Mattingly,[1] Jessica J. Chen,[2] Swathi Arur,[2,*] and Stanislav Y. Shvartsman[1,*]
[1]Department of Chemical and Biological Engineering and Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey; and [2]The University of Texas Graduate School of Biomedical Sciences and Department of Genetics, University of Texas M. D. Anderson Cancer Center, Houston, Texas

ABSTRACT The *Caenorhabditis elegans* germline is a well-studied model system for investigating the control of cell fate by signaling pathways. Cell signals at the distal tip of the germline promote cell proliferation; just before the loop, signals couple cell maturation to organism-level nutrient status; at the proximal end of the germline, signals coordinate oocyte maturation and fertilization in the presence of sperm. The latter two events require dual phosphorylation and activation of ERK, the effector molecule of the Ras/MAPK cascade. In *C. elegans*, ERK is known as MPK-1. At this point, none of today's methods for real-time monitoring of dually phosphorylated MPK-1 are working in the germline. Consequently, quantitative understanding of the MPK-1-dependent processes during germline development is limited. Here, we make a step toward advancing this understanding using a model-based framework that reconstructs the time course of MPK-1 activation from a snapshot of a fixed germline. Our approach builds on a number of recent studies for estimating temporal dynamics from fixed organisms, but takes advantage of the anatomy of the germline to simplify the analysis. Our model predicts that the MPK-1 signal turns on ~30 h into germ cell progression and peaks ~7 h later.

INTRODUCTION

The *Caenorhabditis elegans* germline is a well-studied model system for genetic studies of cell signaling (1–3). The adult hermaphrodite germline consists of two U-shaped tubes that meet at a common uterus. Each tube is filled with germ cells spatially arranged according to their maturity (4). Stem cell divisions at the distal tip of each tube (distal to the common uterus) maintain the germ cell population as the more mature cells move away from the distal tip and transition into meiosis (5,6). After the loop, the bend in the U, a single-file line of oocytes prepare for fertilization and ovulation. Here, we focus on the anterior region of the germline, from the distal tip cell until the loop (Fig. 1), where ~1000 germ cells line the periphery of the gonad tube and are connected to a common cytoplasmic core called the rachis (7,8). Nuclei are separated by incomplete cell membranes that are open to the rachis, so the terms "nucleus" and "cell" in this region are essentially interchangeable. Near the loop, a large portion of healthy germ cells undergo apoptosis, which may leave more cytoplasmic material for the surviving cells that become oocytes (9). In an adult hermaphroditic nematode, the germline reaches steady state, maintaining a roughly constant number of germ cells and ovulation rate for most of the organism's reproductive life (7,10).

Within the rachis are spatial gradients of effector molecules that regulate germ cell development as they progress through the germline (1,2). Among the regulatory molecules in the rachis is the extracellular signal-regulated kinase (ERK), the terminal kinase of the Ras/mitogen-activated protein kinase (MAPK) signaling cascade. The *C. elegans* homolog of ERK is known as MPK-1 (2,11). In the pachytene region of the germline, DAF-2 insulin-like signaling activates the MAPK cascade when the organism is in a nutritionally replete environment, resulting in MPK-1 becoming dually phosphorylated (dpMPK-1) and catalytically active (12); activation of MPK-1 then drives meiotic progression and oocyte production. The ligand to the DAF-2 receptor that results in activation of MPK-1 in the germline is unknown as of this writing. In the proximal part of the germline, MPK-1 activation by signals from sperm couples oocyte maturation to sperm availability (10).

Active MPK-1 has numerous substrates that control multiple biological functions (13). Mutations that result in complete loss of active MPK-1 cause germ cells to arrest in the early pachytene stage of meiosis I. Reduction of MPK-1 activation results in multiple phenotypes, such as delayed progression of pachytene stage germ cells and the formation of excessively large oocytes (2). Alternatively, mutations that overactivate MPK-1 result in higher and ectopic activation of MPK-1 in the loop region, causing defects in oocyte growth and maturation (2). All of these mutations render the animal infertile.

Clearly, the amplitude of the MPK-1 signal in the germline is important for normal germ cell development. The duration of MPK-1 signaling has also been shown to be important in

Submitted June 16, 2015, and accepted for publication October 1, 2015.
*Correspondence: sarur@mdanderson.org or stas@princeton.edu
Editor: Leah Edelstein-Keshet.

0006-3495/15/12/2436/10
http://dx.doi.org/10.1016/j.bpj.2015.10.021

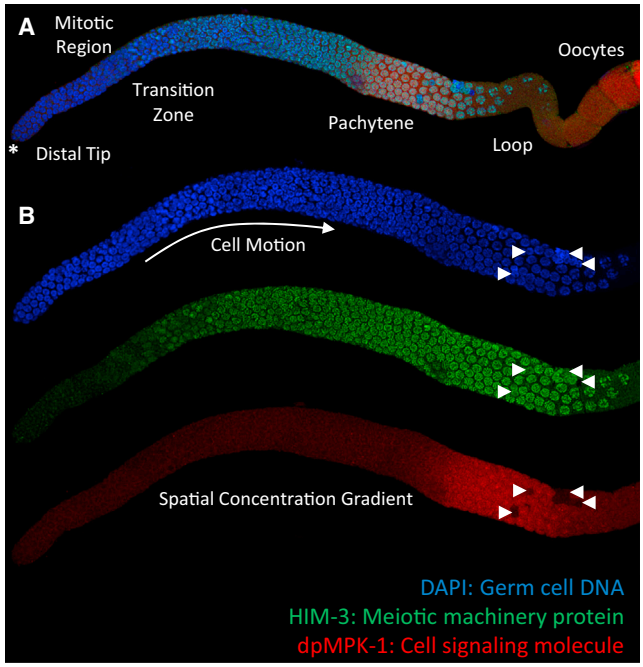


FIGURE 1 An adult, hermaphrodite *C. elegans* germline. (*A*) Shows the anterior region of the germline, from the distal tip (*left* in the image, marked by an *asterisk*) through the first few cellularized oocytes (*right* in the image). (*B*) Focuses on the germline from (*A*), but zoomed in only on the region from the distal tip to the loop. (*Blue*) Nuclei stained with DAPI. Mitotic cells in the mitotic region undergo divisions that maintain the germ cell population. As the cells divide and expand the mitotic region, more mature cells are pushed toward the loop. (*Green*) Germ cells differentiate in the transition zone, entering meiosis, marked by the HIM-3 protein. HIM-3 labels the synaptonemal complex axis of the meiotic nuclei (53). (*Red*) Activation of MPK-1, which occurs in the final two-thirds of the anterior germline. Near the loop, a large fraction of arriving germ cells undergo apoptosis (apoptotic cells in each color channel marked by *white arrowheads*). Surviving cells pass the loop and become oocytes. To see this figure in color, go online.

other cell decision-making contexts, such as the proliferation versus differentiation decision in PC12 cells (14–18), and may also be a relevant factor in the germline. However, because existing methods for real-time monitoring of MPK-1 signaling in living organisms are not yet working in the germline, quantitative information about the dynamics of this signal is not available as of this writing (19–28). In the early *Drosophila melanogaster* embryo, ERK activation dynamics were estimated from fixed samples by matching morphological features to developmental time (29–31). Here, we provide an alternative method to quantitatively estimate MPK-1/ERK activation dynamics in the *C. elegans* germline. Our approach builds on recent work in cultured cells estimating dynamic information from static data (32,33). However, these studies required molecular markers to order cells according to their progress in the cell cycle; in the germline, cells are naturally ordered in space. We take advantage of this anatomical feature and develop a model-based computational approach for the reconstruction of active MPK-1 dynamics in the distal germline (which we define here as the region from the distal tip cell to the loop). While multiple mathematical models for germ cell dynamics have been proposed (34–36), all of these simulate the stem cell programs in the distal region of the germline. To our knowledge, this is the first model aimed at the quantitative estimation of signaling dynamics in the germline from fixed samples and that can be extended to multiple signaling conditions and environmental contexts.

MATERIALS AND METHODS

Nematode fixation and staining

Wild-type N2 hermaphroditic animals were placed on NGM (normal growth media (37)) plates seeded with OP50 *Escherichia coli* bacteria at the fourth larval stage of development (L4) at 20°C for 24 h, and germlines were extruded at the end of 24 h. Dissections were performed as described in Arur et al. (13,38). Briefly, all dissections were performed under 5 min (immediately after adding levamisole to render the animals immobile) to achieve optimal dpMPK-1 staining. The dissected germlines were then fixed in 3% paraformaldehyde for 10 min at room temperature, followed by fixation in 100% methanol at −20°C for 1 h. The fixed germlines were then processed for immuno-fluorescence staining via blocking in 30% Normal Goat Serum (in 1× PBST with 0.1% TWEEN-20) for 1 h, followed by incubation with anti-MAP Kinase antibody used at 1:200 (Clone MAPK-YT; Sigma-Aldrich, St. Louis, MO) overnight in 30% Normal Goat Serum at 4°C. Secondary antibodies were donkey anti-mouse Alexa Fluor 594. The secondary antibodies were used at 1:400. All germlines were treated with the same staining solution in the same vial to mitigate antibody-staining variation, as described in Lee et al. (2).

dpMPK-1 staining was standardized using wild-type germlines and germlines from an *mpk-1* null allele, *ga117*, processed in the same tube (2,12). *ga117* was described in Lackner and Kim (39) and is an early stop in the *mpk-1* gene, resulting in no expression of the MPK-1 protein. The MAPK-YT antibody was rigorously tested for specificity by multiple laboratories against *mpk-1*(*ga117*) null mutant germlines (2,40). The staining conditions have been standardized such that the antibody does not detect any signal in the *mpk-1* null germline, but reproducibly detects the two peaks of MPK-1 activation in wild-type N2 germlines.

Imaging

An A1 confocal microscope with a Plan-Apo VC 60× Oil objective was used for imaging germlines (Nikon, Melville, NY). Z-stacks, with 0.5 μm spacing between slices, were taken of the distal part of each germline. A single z-stack did not fit the entire distal germline in the field of view. Multiple z-stacks taken along the length of a single germline were stitched together using a custom MATLAB program (The MathWorks, Natick, MA). The overlap position between a pair of overlapping z-stacks was found by finding the position with the highest the normalized cross correlation between the DAPI images in the stacks (MATLAB *normxcorr2*).

Image segmentation and extracting fluorescence data

Individual germ cells were detected in each DAPI image slice of the stitched z-stacks using the two-dimensional (2D) circular Hough transform. The built-in MATLAB function *imfindcircles* was used to implement this. This approach was used because the germ cells took on irregular shapes in 3D when pressed by the coverslip. However, within a slice they appeared circular. As a result, the structure imposed by the 2D circular Hough transform produced more accurate detection rates than other less structured

methods, such as watershed-based methods. Before applying the Hough transform, images were smoothed with a 3-pixel-wide Gaussian filter (*imfilter*), eroded (*imerode*) with a 4-pixel disk structuring element (*strel*), and reconstructed (*imreconstruct*), using the Gaussian-smoothed image as the mask. The purpose of these preprocessing steps was to blur out intracellular features, while leaving the circular cell shapes intact.

Within a slice, circles with projected area overlap above a threshold were merged by keeping the circle containing the highest average DAPI fluorescence intensity among those overlapping and removing the others. Because germ cells span multiple slices, a single germ cell would almost always be identified in multiple slices. Therefore, multiple detections of the same germ cell across slices had to be registered. This was done by starting from the bottom (or top) slice, detecting circles in that slice, and then assigning the mutually closest circle in the next slice to the same cell. This was repeated until the last slice, with the maximum allowed extent of a single cell being four slices.

To determine the effectiveness of the image segmentation code, three germlines were manually segmented in ImageJ (National Institutes of Health, Bethesda, MD) using the oval selection tool to circle cells in each slice. The result of this process was a list of cells, the slices they appear in, their 2D positions in each slice, and their radii in each slice. The automated detections then had to be matched to corresponding manual detections. For each manually detected cell, all automatically detected cells in the same slice range were considered. It was visually determined that sometimes the automated segmentation code found that cells extended one slice above or below the same cell in the manual segmentation, so the slice range for each manually detected cell was increased above and below by one slice.

Among the automatically detected cells in the same slice range as a given manually detected cell, the one with the mutually closest 2D position (averaged over slices) was found. A metric was then needed to decide whether this closest cell was actually the same cell as the manually detected one. Because we knew the two cells were in the same slice range, we only considered their projected areas. The metric used was the area of intersection of the 2D projections of the two cells divided by the area of their union. This metric is common in the object detection community and accounts for both the locations and scales of the objects being compared (41,42). If the metric was above a preset threshold (0.3), the automatically detected cell was considered to be the same cell as the manually detected one. Otherwise, the manually detected cell was considered to be undetected by the segmentation code. An automatically-detected cell that was matched to a manually-detected one could not be matched to any other manually detected ones. After cycling through all of the manually detected cells, any cells that were automatically detected but not assigned to a manually detected one were considered false detections. Among the three manually segmented germlines, on average the detector found 90% of the germ cells present, and of the ones detected, 97% were true positives. Images comparing the automated and manual detections are provided in the Section S1 in the Supporting Material.

With the three-dimensional (3D) positions and radii of each germ cell, fluorescence intensities could be extracted from z-stacks (Fig. 2). In some cases, the segmentation code could detect faint DAPI signal even when only the tip of a cell was present in a slice. The dpMPK-1 signal from the cells in these slices was very weak and not representative of the actual dpMPK-1 content of the cell. For this reason, the slice of maximum DAPI intensity was determined for each cell, and the average nuclear dpMPK-1 intensity was measured only from that slice. If the dpMPK-1 signal in a germline significantly decreased with slice depth due to photobleaching during imaging, then only the signal from cells in the first few slices were used to measure the dpMPK-1 spatial profile. Additionally, background fluorescence was subtracted from the dpMPK-1 profiles. In each germline, background fluorescence was calculated by averaging the

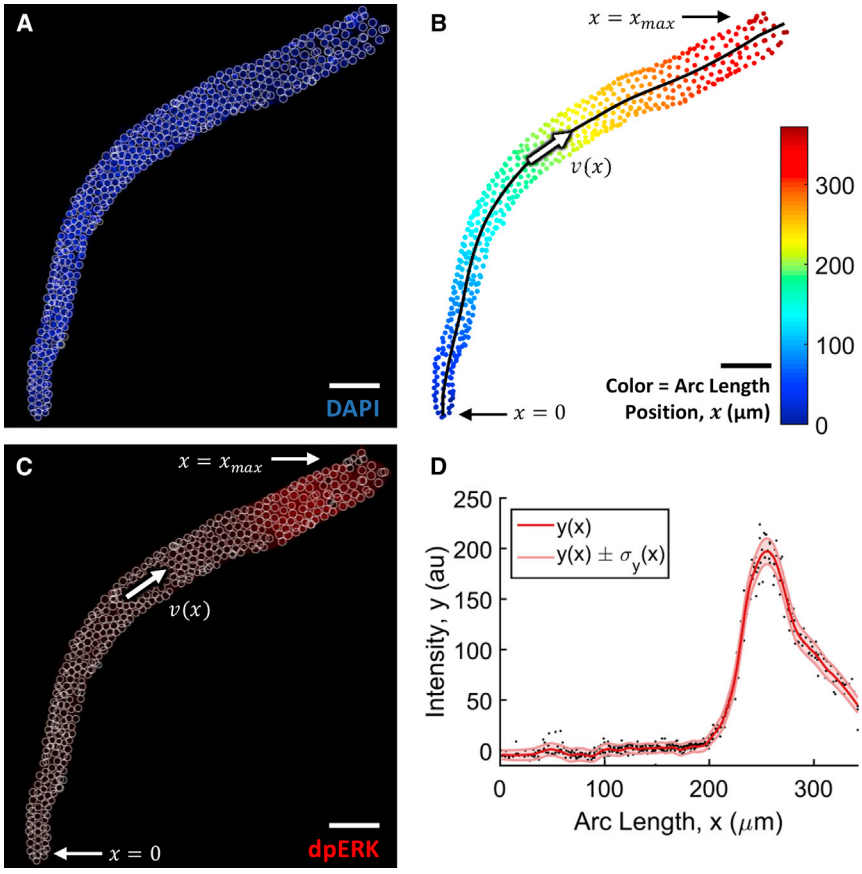


FIGURE 2 Workflow of the image analysis for a given germline. (*A*) Confocal fluorescence microscope image of a DAPI-stained germline, extruded from a wild-type, hermaphroditic adult nematode. (*White circles*) Detected nuclei. Only the distal part of the germline is shown. Scale bar is 21 μm. (*B*) Nuclei detected from z-stack of the germline in (*A*). (*Black*) The center line estimated for this germline. Nuclei are colored by their arc length position along the center line. At the distal tip $x = 0$, and at the beginning of the loop $x = x_{max}$ (in this germline $x_{max} \sim 350$ μm). Germ cells are assumed to move unidirectionally through the germline with spatially-dependent velocity $v(x)$. (*C*) (*Red color channel image*) Germline in (*A*), corresponding to the dpMPK-1 antibody staining. (Overlaid in *white*) Nuclei locations. (*D*) The nuclear mask in (*B*) was used to extract nuclear fluorescence intensity of dpMPK-1 from (*C*) for each germ cell. Plotted is dpMPK-1 fluorescence intensity versus arc length position. (*Black dots*) Measurements of dpMPK-1 intensity from each germ cell. (*Dark-red line*) Smoothed average across cells; (*shaded region*) average intensity ± 1 SD. To see this figure in color, go online.

dpMPK-1 intensity in the mitotic and early meiotic region, where there should be much less active MPK-1 than in the pachytene region. For each experiment, ~6–10 intact germlines were imaged with the exact same microscope settings, and each experiment was repeated at least three times. The data was internally compared for consistency.

Finding the center line of the gonad tube

The scatter plot of individual cell positions in a single germline viewed from a distance appears to be a one-dimensional curve. Diffusion Maps, a nonlinear dimensionality reduction technique (43), essentially takes advantage of this coarse perspective to order the germ cells according to their positions along the one-dimensional curve, which in this case corresponds to the center line of the gonad tube. More details about the Diffusion Maps algorithm are given in the Supporting Material. Diffusion Maps is not necessarily required to find the center line; we used it because it quickly and robustly sorts cells by arc length along the center line. A user needed to manually specify which end of the one-dimensional curve corresponded to the distal tip. Once the cells had been ordered along the center line and the distal tip had been chosen, the positions of germ cells within a local neighborhood of each other were averaged, producing a noisy estimate of the center line. Following the approach of Kafri et al. (32), the ordered set of points was converted to spherical coordinates, and each of the three ordered sets of coordinates was smoothed by taking moving averages using MATLAB's *smooth* function. Converting back to Cartesian coordinates resulted in a smooth estimate of the center line. Due to neighborhood averaging, the center line estimated this way did not extend all the way to the two ends of the distal germline. Each end of the center line was extended in a straight line in the direction of the tangent to the center line at that end. The center line was extended at each end until it intersected the plane containing the distal- (or proximal-) most cell with normal vector parallel to the direction in which the center line was being extended.

Estimating probability distribution functions

Germ cell positions were projected from 3D space onto the closest point on the center line of the gonad tube to calculate their arc length positions. Smooth probability density functions and cumulative distribution functions were estimated in MATLAB using Z. Botev's code *kde*, available on the MathWorks website at: http://www.mathworks.com/central/fileexchange/14034-kernel-density-estimator (44).

RESULTS

Transport model for the germline

In this section, we formulate a mathematical model of germ cell transport along the germline. The prediction of this model will be the average time it takes a germ cell to travel from the distal tip to any given position in the germline. The approach, much like models used to describe unidirectional traffic flow (45), will be to estimate the local velocity as a function of spatial position. If we were given the average velocity of germ cells at each position, we could immediately calculate the average transit times. Because we do not get this information from fixed germlines, we will use our model to express velocity in terms of quantities that are measurable from fixed samples.

We will only consider distal-to-proximal motion and neglect perpendicular motion because the length of the gonad tube is much longer than its diameter. Position in the germline will be measured by arc length: if one were to measure the total distance driven down a winding road, that distance would be the arc length of the road. Arc length position, x, will be measured as distance in microns from the distal tip, with $x = 0$ located at the distal tip. We aim to derive an equation $t(x)$ that describes the time it takes germ cells to travel from the distal tip to each arc length position along the germline.

In this analysis, we focus only on the distal germline. The approach described here for estimating time as a function of arc length is statistical in nature and therefore benefits from having as many cells as possible in the sample. In the proximal germline, where oocytes are in single-file, there are too few cells to make any meaningful estimates using this approach.

We will only consider adult hermaphroditic organisms (defined as 24 h past the fourth larval molt (L4)) with fully developed germlines because these germlines have reached steady state (7,10). Specifically, we assume that the number of germ cells in the germline and the rate of oocyte ovulation are constant over time (10). Young adults are in a state of germline expansion, increasing their germ cell numbers. By 24 h past the fourth larval molt, the germline switches to a homeostatic mode, maintaining the number of germ cells more or less constant (7,46). In normal laboratory conditions at 20°C, this state lasts until ~72–90 h past L4 (12,47). Time course analysis of hermaphroditic germlines assayed at 18, 20, 24, 36, and 48 h past the L4 stage of development for dpMPK-1 activation suggests that the concentration gradients inside the germline are also time-invariant in homeostatic adult germlines (12).

Next, we assume that germ cells move unidirectionally through the germline, from which it follows that germ cells are perfectly arranged according to their maturity. This is not strictly true, but there is clear directionality to germ cells' motion when looking at timescales on the order of hours, the timescale over which germ cells develop. This assumption is weakest in the mitotic region, where it is unclear whether there is a mechanism that would prevent some cells from moving unidirectionally down the germline, such as asymmetric cell divisions (7), trapping by distal tip cell processes (7,48), or diffusive motion created by random orientation of the axes of cell divisions (36). We acknowledge that our model does not account for effects like these. However, while an individual cell could get trapped in the mitotic region for an extended period of time before beginning unidirectional motion toward the loop, this would only add a time delay to the beginning of the time course of MPK-1 activation, without otherwise changing the dynamics.

The final assumption we make is that in a wild-type germline, germ cells undergo stereotypical dynamics, so that the population-level dynamics are representative of an individual cell. With this, we assume that the time it takes each cell to reach the loop is the same across cells.

The intuition behind the derivation is that the rate at which cells arrive at each position must be balanced by

the rate at which cells leave that position for there to be no accumulation of cells at any position over time. We can construct a balance on the number of cells arriving and leaving an arbitrary arc length position:

$$0 = -\frac{dJ}{dx} + S(x) - R(x). \quad (1)$$

Here, $J(x)$ describes the number of cells per time that pass through each position x and is known as the convective flux. In this formulation, cell flux and cell flow rate are equivalent. $S(x)$ is a spatially dependent source term that accounts for the production of cells by mitotic divisions, and $R(x)$ is a sink term that accounts for the removal of apoptotic cells from the germline before they reach the loop. The average rate at which cells are produced is $s \equiv \int_0^{x_{max}} S(x)dx$ cells per hour, and, likewise, the average rate at which cells are removed via apoptosis is $r \equiv \int_0^{x_{max}} R(x)dx$ cells per hour.

We assumed that all new cells are introduced into the germline at $x = 0$, reducing Eq. 1 to

$$0 = -\frac{dJ}{dx} - R(x), \quad (2)$$

with cell production included as a boundary condition on the flux at $x = 0$, $J(x = 0) = s$. The left-hand sides of Eqs. 1 and 2 correspond to the rate of accumulation of cells over time at each position x, which is equal to zero under the steady-state assumption.

The convective flux term can be expressed as $J(x) = C(x)v(x)$, where $C(x)$ is the concentration of cells at x, and $v(x)$ is the average velocity of cells at x. We want to solve for $v(x)$. The concentration of cells can be expressed as $C(x) = N_{tot}f_X(x)$, where N_{tot} is the average number of germ cells in the distal germline, and $f_X(x)$ is the probability density function of germ cell arc length positions x (Fig. 3). The concentration $C(x)$ is a smoothed histogram of germ cell positions, and $f_X(x)$ is the same but normalized so that the area under the curve equals 1. We estimated the function $f_X(x)$ empirically for each germline as follows: germ cells were automatically detected in a z-stack of an extruded germline using custom code written in MATLAB; the center line of the germline was constructed; arc length positions of cells in that germline were calculated; then $f_X(x)$ for that germline was estimated from the collection of cell positions using kernel density estimation, a standard technique for empirically estimating probability distributions (44).

Solving the model

In this section we solve the transport model for $v(x)$, the average local velocity, and $t(x)$, the average transit time to each position along the gonad tube. First, we substitute $J(x) = N_{tot}f_X(x)v(x)$ for the convective flux term of Eq. 2. Moving the convective flux term to the left-hand side and integrating once gives

$$N_{tot}f_X(x)v(x) = s - \int_0^x R(u)du, \quad (3)$$

where s, the rate of cell production, is the constant of integration that results from applying the boundary condition $J(x = 0) = s$. Solving for $v(x)$ gives

$$v(x) = \frac{s - \int_0^x R(u)du}{N_{tot}f_X(x)}, \quad (4)$$

We can draw an analogy here to unidirectional, steady-state fluid flow in a pipe. In the pipe flow problem, $v(x) = Q(x)/A(x)$—i.e., the average local velocity of fluid elements equals the local volumetric flow rate divided by the local cross-sectional area. The numerator of the expression on the right-hand side of Eq. 4 is the local cell flow rate (cells per unit time), analogous to the local volumetric flow rate in a pipe (volume per unit time). The cell flow rate at $x = 0$, where cells are produced by mitotic divisions, equals s, and the second term in the

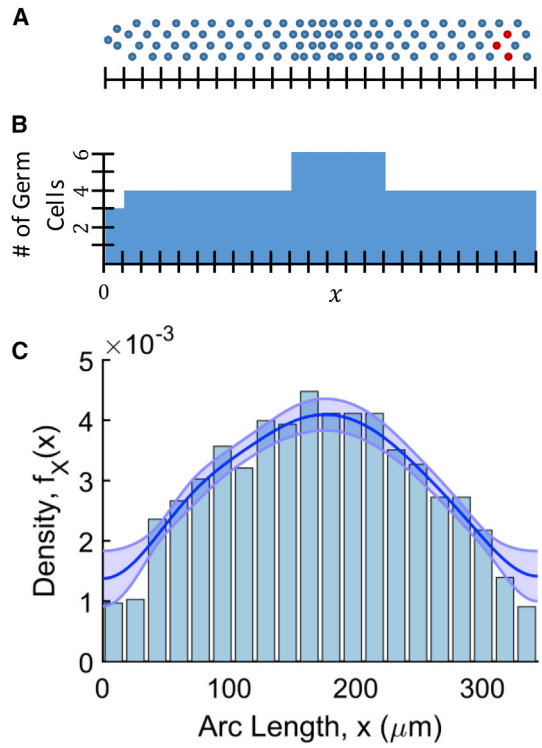


FIGURE 3 (*A*) Schematic of a germline with a nonuniform distribution of cells over arc length position x. (*Red*) Cells undergoing apoptosis. (*B*) Histogram of germ cell arc length positions from (*A*). Smoothing this histogram and normalizing by the total number of cells in the germline would give an estimate of $f_X(x)$, the probability density function over arc length positions. (*C*) A plot of $f_X(x)$, calculated from germ cell positions in a real germline and overlaid on the normalized histogram of cell positions in that germline. (*Shaded region*) Mean estimate of $f_X(x)$ ± the root-mean-squared error of the estimate (54). To see this figure in color, go online.

numerator reduces the local cell flow rate as cells undergo apoptosis. In the absence of cell death, the cell flow rate everywhere in the germline would be constant and equal to the production rate, s. The denominator of the expression, the local concentration of cells, is analogous to the local cross-sectional area in the pipe flow case.

The velocity $v(x)$ is the time rate of change of a cell's position x, or $v(x) = dx/dt$. Separating the variables x and t, giving $dt = dx/v(x)$, and then integrating, produces

$$t(x) = \int_0^x \frac{N_{\text{tot}} f_X(w)}{s - \int_0^w R(u)du} dw. \quad (5)$$

The variables u and w are dummy variables of integration. The lower bound of the outer integral on the right-hand side is zero because at $x = 0$, $t = 0$. The average time cells take to transit from the distal tip at $x = 0$ to the loop at $x = x_{\max}$ will be denoted τ.

Factoring out N_{tot}/s from Eq. 5 shows that there are essentially two lumped parameter combinations that govern the behavior of $t(x)$: the ratios N_{tot}/s and r/s. After factoring, the equation becomes

$$t(x) = \frac{N_{\text{tot}}}{s} \int_0^x \frac{f_X(w)}{1 - \frac{r}{s}\int_0^w \widehat{R}(u)du} dw, \quad (6)$$

where $\widehat{R}(x)$ is a dimensionless function with the same shape as $R(x)$ and, as mentioned in the previous section, r is the average rate at which cells undergo apoptosis. The first ratio, N_{tot}/s, is the total number of cells in the germline divided by the cell production rate and has units of time. In the absence of apoptosis, the average transit time from the distal tip to the loop would be equal to N_{tot}/s. The second ratio, r/s, is the fraction of germ cells that undergo apoptosis and is dimensionless. N_{tot}/s gives the scale of $t(x)$ but does not affect its shape, which is determined by the expression inside the integral.

In practice, $f_X(x)$ was estimated for each germline and $R(x)$ was calculated by pooling data from multiple germlines. Estimation of $R(x)$, which accounts for germ cell apoptosis and removal from the germline, required pooling data from multiple germlines and is described in the Supporting Material (Sections S3 and S4). The same values of the parameters N_{tot}, s, and r, were used for all germlines. In particular, N_{tot} and s have strong opposing effects on the model's predictions; as a result, using a value of N_{tot} from an individual germline and the average value for s would give biased results. During the error analysis (Section S5 in the Supporting Material), the parameters N_{tot}, s, and r were sampled uniformly within ranges consistent with the literature to propagate parameter uncertainties through the model, which will be discussed in the next section. The trapezoidal rule was used to approximate all integrals.

Exploring the feasible space of physiological parameters

The model contains one equation and four physiological parameters: N_{tot}, the average number of cells in a germline at any given time once adulthood is reached; s, the average rate of cell production; r, the average rate of cell removal by apoptosis; and τ, the average time cells take to travel from the distal tip to the loop. The steady-state assumption provides an equation relating these physiological parameters, meaning their values are no longer entirely independent of each other. Evaluating the indefinite integral in Eq. 6 at $x = x_{\max}$, where $t(x_{\max}) = \tau$ gives

$$\tau = \frac{N_{\text{tot}}}{s} \int_0^{x_{\max}} \frac{f_X(w)}{1 - \frac{r}{s}\int_0^w \widehat{R}(u)du} dw. \quad (7)$$

The density function $f_X(x)$ is measurable for each individual germline, and $\widehat{R}(x)$ can be estimated by pooling data from multiple germlines. This leaves only the physiological parameters as unknowns in the equation. Additionally, choosing values for three of the parameters constrains the value of the fourth.

Previous studies have estimated ranges for these physiological parameters from data on wild-type animals. The total number of cells in an adult hermaphroditic germline is ~800–1000 cells (7). The cell production rate by mitotic divisions, s, has been reported to be between 16 and 24 cells per hour (7,49). An engulfment marker shows that in wild-type animals, each gonad arm contains 2–4 apoptotic cells at a given time (50). However, the time rate of cell death, per se, has not been precisely measured. Under the steady-state assumption, the rate of cell removal via apoptosis, r, must equal the difference between the rate of cell production, s, and the rate of cell removal via ovulation. The latter has been estimated to be between 2 and 5 cells per hour per gonad arm (10). Therefore, a consistent estimate for the rate of cell death, r, lies between 11 and 22 cells per hour. Finally, pulse-chase experiments suggest that τ lies between 48 and 54 h (51).

Although there are only three independent physiological parameters after applying Eq. 7, we have information about all four parameters. To capitalize on this, we sampled the parameters N_{tot}, s, and r uniformly within the literature-reported ranges, then discarded all parameter combinations for which the value of τ, calculated using Eq. 7, was outside of the range consistent with the literature. This procedure carved out a sliver in the 3D parameter space consisting of the parameter combinations that are consistent with contemporary knowledge (52). This collection of consistent parameters is called the "feasible set".

Model predictions

With Eq. 6 and the feasible set of parameters, we were able to calculate, for each germline, a prediction of the transit time to each position along that germline, $t(x)$. These maps from position to time were used to transform the spatial profiles of dpMPK-1 fluorescence intensity to dynamic profiles for each germline. Essentially, dpMPK-1 intensity at each position in a germline was plotted against the time at which cells arrive at that position. The result was a picture of the time course of MPK-1 signaling that germ cells experience/produce as they move through that germline (Fig. 4). Doing this for multiple germlines allowed us to compare signaling dynamics across germlines.

Comparing the estimated time courses of dpMPK-1 from multiple germlines shows that they approximately collapse. It was not obvious a priori that this would be the case from the raw spatial profiles of dpMPK-1. Additionally, naively plotting dpMPK-1 intensity against normalized arc length for each germline does not cause the profiles to collapse (Fig. S5 in the Supporting Material). The collapse of these curves suggests that, despite varying geometries and dpMPK-1 gradients among germlines, germ cells see stereotypical dynamics of MPK-1 activation during meiotic progression.

The average time at which germ cells in different germlines first begin to be exposed to active MPK-1 (30.5 h, SD 1.5 h) and the time of peak dpMPK-1 concentration (mean 37.7 h, SD 1.8 h) are very similar across germlines. These averages were taken over all germlines and all parameter combinations in the feasible set. The uncertainties are dominated by the uncertainty in the rate of apoptosis, r. Future experiments that constrain the range of possible values for the parameter r will dramatically reduce the uncertainty of this model's estimates.

The model also makes quantitative predictions about the relationships between physiological parameters in an animal. From Eqs. 6 and 7, one can immediately see that because ratios of physiological parameters govern the behavior of $t(x)$, numerous values of the individual parameters N_{tot}, s, and r can give rise to the same dynamics in the germline. Additionally, looking at the dependence of $\tau = t(x_{max})$, the average transit time from the distal tip to the loop, on the lumped parameter ratios is also informative (Fig. 5). In a situation where the rate of ovulation decreases, germ cells will subsequently spend more time in the region distal to the loop. To maintain steady state, the animal's physiological parameters must change. According to the model, steady state can be achieved by: 1) decreasing the rate of cell production, s; 2) increasing the number of cells in the distal region N_{tot}; or 3) increasing the rate of apoptosis, r. The quantitative model predictions are therefore in line with qualitative intuitions.

Thus, during the first wave of MPK-1 activation in the germline, each of the developing germ cells is exposed to high levels of MPK-1 signaling for several hours, during which MPK-1 activation is translated into changes in a large

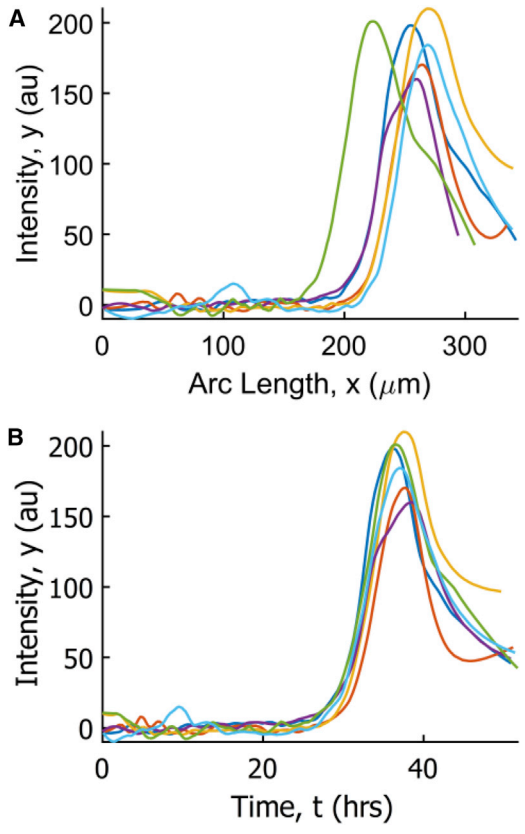


FIGURE 4 (*A*) Raw spatial profiles of dpMPK-1 fluorescence intensity, $y(x)$, in six germlines from a single experiment, imaged under identical conditions. (*B*) For each germline, the function $t(x)$ relating arc length positions to the transit times to reach those positions was estimated using Eq. 5. Plotted are the results of transforming the spatial dpMPK-1 profiles in (*A*) into average time courses of dpMPK-1 seen by traveling cells. To see this figure in color, go online.

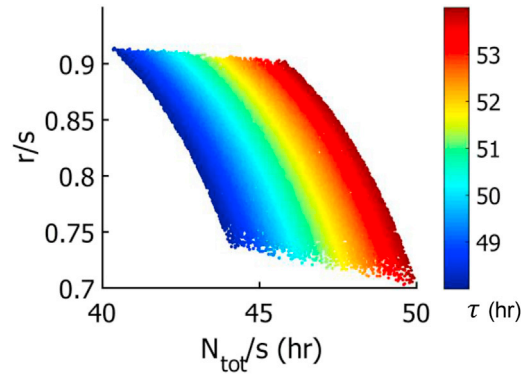


FIGURE 5 Dependence of τ, the average transit time from the distal tip to the loop, on the two lumped parameter ratios in the model. The first, N_{tot}/s, is the ratio of the number of cells in the germline to the cell production rate. The second, r/s, is the ratio of the rate of apoptosis to the cell production rate, or the fraction of cells that undergo apoptosis. To see this figure in color, go online.

number of processes, including apoptosis, meiotic progression, and membrane biogenesis.

DISCUSSION

Cell fate decision processes rely on the joint dynamics of multiple molecular players. Some of these species may be difficult to follow in real-time, calling for approaches for reconstructing dynamics based on snapshots from fixed samples. In the *Drosophila melanogaster* embryo, cell signaling and gene induction dynamics were estimated from fixed samples by matching embryo morphology to time (29–31). Additionally, a number of recent articles demonstrated that for a stationary process, in which cells progress unidirectionally through a sequence of states, average dynamics can be reconstructed from fixed cells. For instance, Kafri et al. (32) used data from a population of unsynchronized, fixed vertebrate cultured cells to reconstruct the average growth dynamics of an individual cell. In this work, each cell was labeled for DNA and Geminin (a protein degraded during mitosis), and this pair of signals provided a quantitative measure of cell cycle progression for each cell. The large numbers of cells available to them in each experiment enabled estimation of cell size distributions throughout the cell cycle and revealed a feedback mechanism that reduces the variability in cell sizes. A similar approach was used by Akopyan et al. (33) to reconstruct the dynamics of the mitotic entry network. In both of these studies, reconstruction of dynamics from snapshots relied on the introduction of some measure of progress, which was critical for temporal ordering of data from multiple individual cells. While this step is essential for a dataset extracted from cells that are independent of each other, it is not necessary for data from cells within the germline, where the arc length provides a natural ordering coordinate.

Assuming that the number and spatial arrangement of cells are invariant over time, we derived a transport model, in which cells move similar to cars on a packed highway, to relate arc length, rather than the cell state, to time. In doing so, we used literature estimates of the physiological parameters in the germline (7,10,49,51). Finally, we pooled data from multiple germlines to estimate the average local rates of cell death. This led to the average time course of ERK activation within a developing germ cell. Looking at this pulse, one can ask, what controls the onset of ERK activation, its amplitude, and duration? It is known that this phase of ERK signaling relies on a spatially uniform insulin receptor, but the relevant ligand and the mechanisms responsible for the intracellular control of the duration of ERK activation are still unknown. As the molecular candidates for controlling these processes are proposed, such as a localized extracellular ligand, our approach can be used to probe their quantitative contributions to multiple aspects of ERK signaling.

The estimates derived from our model-based approach of cell transit times and ERK activation dynamics remain to be verified by data from studies with live reporters of ERK activation. At this point, these estimates provide the only available insight into the dynamics of a key regulatory signal in one of the most advanced experimental systems for studies of cell fate decisions. As measurements of the physiological parameters in the model become more precise, the model may be updated and its uncertainties reduced. Applying this approach to germlines at other steady states under altered conditions is relatively straightforward. Changes in the rate of cell division, rate of cell death, total number of cells, or time cells take to reach the loop can cause germlines to operate at different steady states. These alternative steady states can be achieved in the lab setting by changing the environmental conditions (temperature, crowding, or presence of a noxious substance) or by introducing genetic perturbations. However, for analysis of mutant backgrounds, the assumptions that the germline is at steady state, that the dpERK gradient is time-invariant, and that the distribution of germ cells is not changing over time, must be verified. Additionally, at least three of the four physiological parameters for that mutant background must be measured to apply to the model to alternative steady states: the average rate of cell production, the average number of cells in the germline, the average rate of apoptosis, and the average time cells take to travel from the distal tip to the loop.

SUPPORTING MATERIAL

Supporting Materials and Methods and eight figures are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(15)01060-7.

AUTHOR CONTRIBUTIONS

J.J.C. extruded, fixed, immunostained, and mounted the germlines on glass slides for imaging; H.H.M. derived the model, imaged the germlines, wrote the MATLAB code, and performed the image and data analysis; and all authors contributed to designing the research and writing the article.

ACKNOWLEDGMENTS

H.H.M. thanks Yannis Kevrekidis and Carmeline Dsilva for helpful discussions about the implementation of the Diffusion Maps algorithm.

H.H.M. and S.Y.S. were supported by National Institutes of Health grant No. R01GM086537. S.A. is supported by National Institutes of Health grant No. GM 98200 and American Cancer Society grant No. RSG014-044-DDC. The Deans Scholarship at University of Texas Graduate School of Biomedical Sciences supports J.J.C.

SUPPORTING CITATIONS

References (55–57) appear in the Supporting Material.

REFERENCES

1. Hansen, D., and T. Schedl. 2006. The regulatory network controlling the proliferation-meiotic entry decision in the *Caenorhabditis elegans* germ line. *Curr. Top. Dev. Biol.* 76:185–215.

2. Lee, M. H., M. Ohmachi, …, T. Schedl. 2007. Multiple functions and dynamic activation of MPK-1 extracellular signal-regulated kinase signaling in *Caenorhabditis elegans* germline development. *Genetics.* 177:2039–2062.
3. Hubbard, E. J. A. 2007. *Caenorhabditis elegans* germ line: a model for stem cell biology. *Dev. Dyn.* 236:3343–3357.
4. Schedl, T. 1997. Developmental genetics of the germline. *In C. Elegans* II. D. L. Riddle, T. Blumenthal, B. J. Meyer, and J. R. Priess, editors. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 241–269.
5. Kimble, J., and H. Seidel. 2008. *C. elegans* Germline Stem Cells and their Niche. *In* StemBook. Harvard Stem Cell Institute, Cambridge, MA.
6. Crittenden, S. L., E. R. Troemel, …, J. Kimble. 1994. GLP-1 is localized to the mitotic region of the *C. elegans* germ line. *Development.* 120:2901–2911.
7. Crittenden, S. L., K. A. Leonhard, …, J. Kimble. 2006. Cellular analyses of the mitotic region in the *Caenorhabditis elegans* adult germ line. *Mol. Biol. Cell.* 17:3051–3061.
8. White, J. G. 1988. The anatomy. *In* The Nematode *Caenorhabditis elegans*. W. B. Wood, editor. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
9. Gumienny, T. L., E. Lambie, …, M. O. Hengartner. 1999. Genetic control of programmed cell death in the *Caenorhabditis elegans* hermaphrodite germline. *Development.* 126:1011–1022.
10. McCarter, J., B. Bartlett, …, T. Schedl. 1999. On the control of oocyte meiotic maturation and ovulation in *Caenorhabditis elegans*. *Dev. Biol.* 205:111–128.
11. Church, D. L., K. L. Guan, and E. J. Lambie. 1995. Three genes of the MAP kinase cascade, mek-2, mpk-1/sur-1 and let-60 ras, are required for meiotic cell cycle progression in *Caenorhabditis elegans*. *Development.* 121:2525–2535.
12. Lopez, A. L., 3rd, J. Chen, …, S. Arur. 2013. DAF-2 and ERK couple nutrient availability to meiotic progression during *Caenorhabditis elegans* oogenesis. *Dev. Cell.* 27:227–240.
13. Arur, S., M. Ohmachi, …, T. Schedl. 2009. Multiple ERK substrates execute single biological processes in *Caenorhabditis elegans* germline development. *Proc. Natl. Acad. Sci. USA.* 106:4776–4781.
14. Gotoh, Y., E. Nishida, …, H. Sakai. 1990. Microtubule-associated-protein (MAP) kinase activated by nerve growth factor and epidermal growth factor in PC12 cells. Identity with the mitogen-activated MAP kinase of fibroblastic cells. *Eur. J. Biochem.* 193:661–669.
15. Nguyen, T. T., J. C. Scimeca, …, E. van Obberghen. 1993. Co-regulation of the mitogen-activated protein kinase, extracellular signal-regulated kinase 1, and the 90-kDa ribosomal S6 kinase in PC12 cells. Distinct effects of the neurotrophic factor, nerve growth factor, and the mitogenic factor, epidermal growth factor. *J. Biol. Chem.* 268:9803–9810.
16. Marshall, C. J. 1995. Specificity of receptor tyrosine kinase signaling: transient versus sustained extracellular signal-regulated kinase activation. *Cell.* 80:179–185.
17. Ebisuya, M., K. Kondoh, and E. Nishida. 2005. The duration, magnitude and compartmentalization of ERK MAP kinase activity: mechanisms for providing signaling specificity. *J. Cell Sci.* 118:2997–3002.
18. Futran, A. S., A. J. Link, …, S. Y. Shvartsman. 2013. ERK as a model for systems biology of enzyme kinetics in cells. *Curr. Biol.* 23:R972–R979.
19. Green, H. M., and J. Alberola-Ila. 2005. Development of ERK activity sensor, an in vitro, FRET-based sensor of extracellular regulated kinase activity. *BMC Chem. Biol.* 5:1.
20. Burack, W. R., and A. S. Shaw. 2005. Live cell imaging of ERK and MEK: simple binding equilibrium explains the regulated nucleocytoplasmic distribution of ERK. *J. Biol. Chem.* 280:3832–3837.
21. Fujioka, A., K. Terai, …, M. Matsuda. 2006. Dynamics of the Ras/ERK MAPK cascade as monitored by fluorescent probes. *J. Biol. Chem.* 281:8917–8926.
22. Harvey, C. D., A. G. Ehrhardt, …, K. Svoboda. 2008. A genetically encoded fluorescent sensor of ERK activity. *Proc. Natl. Acad. Sci. USA.* 105:19264–19269.
23. Komatsu, N., K. Aoki, …, M. Matsuda. 2011. Development of an optimized backbone of FRET biosensors for kinases and GTPases. *Mol. Biol. Cell.* 22:4647–4656.
24. Aoki, K., Y. Kumagai, …, M. Matsuda. 2013. Stochastic ERK activation induced by noise and cell-to-cell propagation regulates cell density-dependent proliferation. *Mol. Cell.* 52:529–540.
25. Ahmed, S., K. G. Grant, …, J. M. Haugh. 2014. Data-driven modeling reconciles kinetics of ERK phosphorylation, localization, and activity states. *Mol. Syst. Biol.* 10:718.
26. Kumagai, Y., H. Naoki, …, M. Matsuda. 2015. Heterogeneity in ERK activity as visualized by in vivo FRET imaging of mammary tumor cells developed in MMTV-Neu mice. *Oncogene.* 34:1051–1057.
27. Regot, S., J. J. Hughey, …, M. W. Covert. 2014. High-sensitivity measurements of multiple kinase activities in live single cells. *Cell.* 157:1724–1734.
28. Hiratsuka, T., Y. Fujita, …, M. Matsuda. 2015. Intercellular propagation of extracellular signal-regulated kinase activation revealed by in vivo imaging of mouse skin. *eLife.* 4:e05178.
29. Lim, B., N. Samper, …, S.Y. Shvartsman. Kinetics of gene derepression by ERK signaling. *Proc. Natl. Acad. Sci. USA.* 110:10330–10335.
30. Lim, B., C. J. Dsilva, …, S. Y. Shvartsman. 2015. Dynamics of inductive ERK signaling in the *Drosophila* embryo. *Curr. Biol.* 25:1784–1790.
31. Dsilva, C. J., B. Lim, …, S. Y. Shvartsman. 2015. Temporal ordering and registration of images in studies of developmental dynamics. *Development.* 142:1717–1724.
32. Kafri, R., J. Levy, …, M. W. Kirschner. 2013. Dynamics extracted from fixed cells reveal feedback linking cell growth to cell cycle. *Nature.* 494:480–483.
33. Akopyan, K., H. Silva Cascales, …, A. Lindqvist. 2014. Assessing kinetics from fixed cells reveals activation of the mitotic entry network at the S/G2 transition. *Mol. Cell.* 53:843–853.
34. Beyer, A., R. Eberhard, …, J. Fisher. 2012. A dynamic physical model of cell migration, differentiation and apoptosis in *Caenorhabditis elegans*. *Adv. Exp. Med. Biol.* 736:211–233.
35. Setty, Y., D. Dalfó, …, H. Kugler. 2012. A model of stem cell population dynamics: in silico analysis and in vivo validation. *Development.* 139:47–56.
36. Hall, B. A., N. Piterman, …, J. Fisher. 2015. Emergent stem cell homeostasis in the *C. elegans* germline is revealed by hybrid modeling. *Biophys. J.* 109:428–438.
37. Brenner, S. 1974. The genetics of *Caenorhabditis elegans*. *Genetics.* 77:71–94.
38. Arur, S., M. Ohmachi, …, T. Schedl. 2011. MPK-1 ERK controls membrane organization in *C. elegans* oogenesis via a sex-determination module. *Dev. Cell.* 20:677–688.
39. Lackner, M. R., and S. K. Kim. 1998. Genetic analysis of the *Caenorhabditis elegans* MAP kinase gene mpk-1. *Genetics.* 150:103–117.
40. Miller, M. A., V. Q. Nguyen, …, D. Greenstein. 2001. A sperm cytoskeletal protein that signals oocyte meiotic maturation and ovulation. *Science.* 291:2144–2147.
41. Ferrari, V.-J., M. Marin-Jimenez, and A. Zisserman. 2008. Progressive search space reduction for human pose estimation. *Proc. CVPR IEEE.* 1–8.
42. Marin-Jimenez, M. J., A. Zisserman, …, V. Ferrari. 2014. Detecting people looking at each other in videos. *Int. J. Comput. Vis.* 106:282–296.
43. Coifman, R. R., and S. Lafon. 2006. Diffusion maps. *Appl. Comput. Harmon. Anal.* 21:5–30.
44. Botev, Z. I., J. F. Grotowski, and D. P. Kroese. 2010. Kernel density estimation via diffusion. *Ann. Stat.* 38:2916–2957.
45. Treiber, M., and A. Kesting. 2013. Traffic Flow Dynamics: Data, Models and Simulation. Springer, Berlin, Germany.

46. Gerhold, A. R., J. Ryan, …, P. S. Maddox. 2015. Investigating the regulation of stem and progenitor cell mitotic progression by in situ imaging. *Curr. Biol.* 25:1123–1134.

47. Hughes, S. E., K. Evason, C. Xiong, and K. Kornfeld. 2007. Genetic and Pharmacological Factors That Influence Reproductive Aging in Nematodes. *PLoS Genet.* 3:e25.

48. Wong, B. G., A. Paz, …, E. E. Hui. 2013. Live imaging reveals active infiltration of mitotic zone by its stem cell niche. *Integr. Biol. (Camb.).* 5:976–982.

49. Fox, P. M., and T. Schedl. 2015. Analysis of germline stem cell differentiation following loss of GLP-1 notch activity in *Caenorhabditis elegans*. *Genetics.* 201:167–184.

50. Park, D., H. Jia, …, H. M. Chamberlin. 2006. Pax2/5/8 proteins promote cell survival in *C. elegans*. *Development.* 133:4193–4202.

51. Jaramillo-Lambert, A., M. Ellefson, …, J. Engebrecht. 2007. Differential timing of S phases, X chromosome replication, and meiotic prophase in the *C. elegans* germ line. *Dev. Biol.* 308:206–221.

52. Frenklach, M., A. Packard, …, R. Feeley. 2004. Collaborative data processing in developing predictive models of complex reaction systems. *Int. J. Chem. Kinet.* 36:57–66.

53. Zetka, M. C., I. Kawasaki, …, F. Müller. 1999. Synapsis and chiasma formation in *Caenorhabditis elegans* require HIM-3, a meiotic chromosome core component that functions in chromosome segregation. *Genes Dev.* 13:2258–2270.

54. Hansen, B. 2009. Lecture Notes on Nonparametrics. University of Wisconsin-Madison, Madison, WI, Accessed October 2015. http://www.ssc.wisc.edu/~bhansen/718/NonParametrics1.pdf.

55. Coifman, R. R., Y. Shkolnisky, …, A. Singer. 2008. Graph Laplacian tomography from unknown random projections. *IEEE Trans. Image Process.* 17:1891–1899.

56. Ferguson, A. L., A. Z. Panagiotopoulos, …, I. G. Kevrekidis. 2010. Systematic determination of order parameters for chain dynamics using diffusion maps. *Proc. Natl. Acad. Sci. USA.* 107:13597–13602.

57. Lant, B., and W. B. Derry. 2013. Methods for detection and analysis of apoptosis signaling in the *C. elegans* germline. *Methods.* 61:174–182.

Biophysical Journal

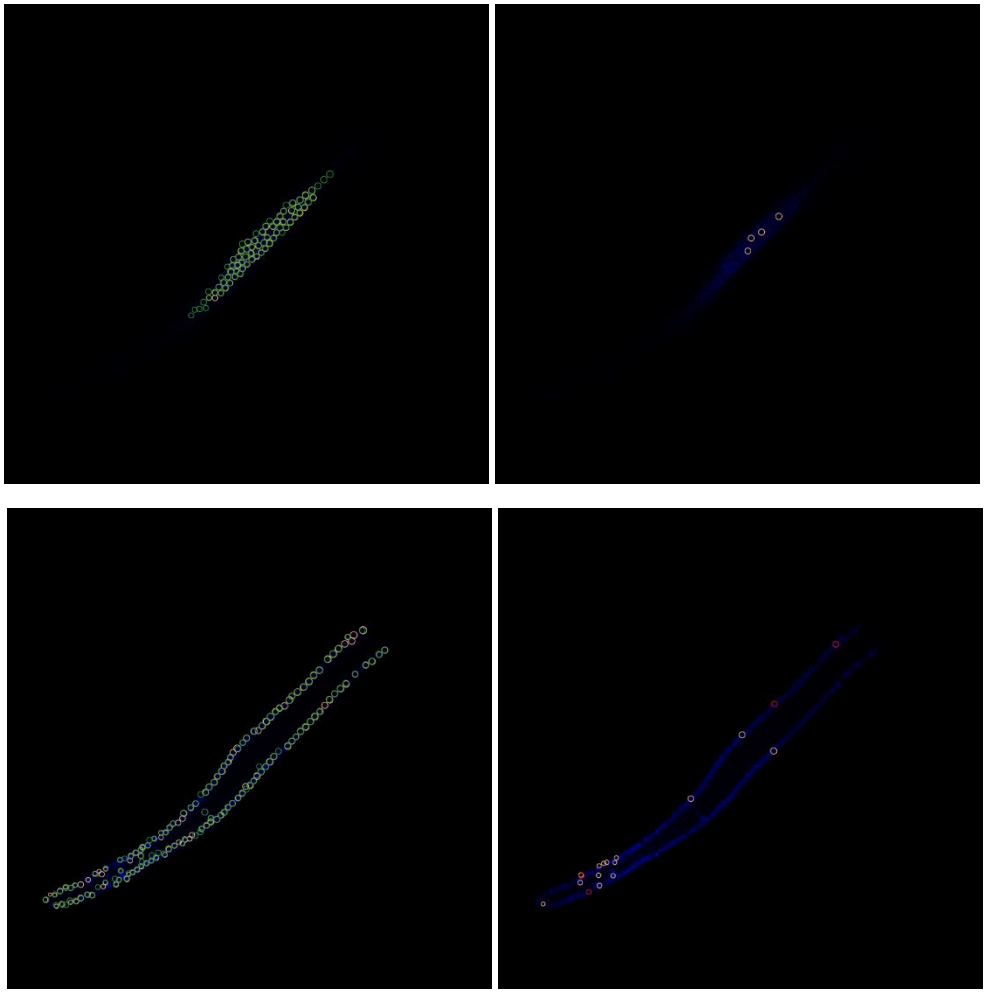
A Transport Model for Estimating the Time Course of ERK Activation in the *C. elegans* Germline

Henry H. Mattingly[1], Jessica J. Chen[2], Swathi Arur[2,*], Stanislav Y. Shvartsman[1,*]

[1]Department of Chemical and Biological Engineering and Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ, [2]UT Graduate School of Biomedical Sciences and Department of Genetics, University of Texas M.D. Anderson Cancer Center, Houston, TX

SUPPORTING MATERIAL

Section S1: Comparison of Manual and Automated Image Segmentation



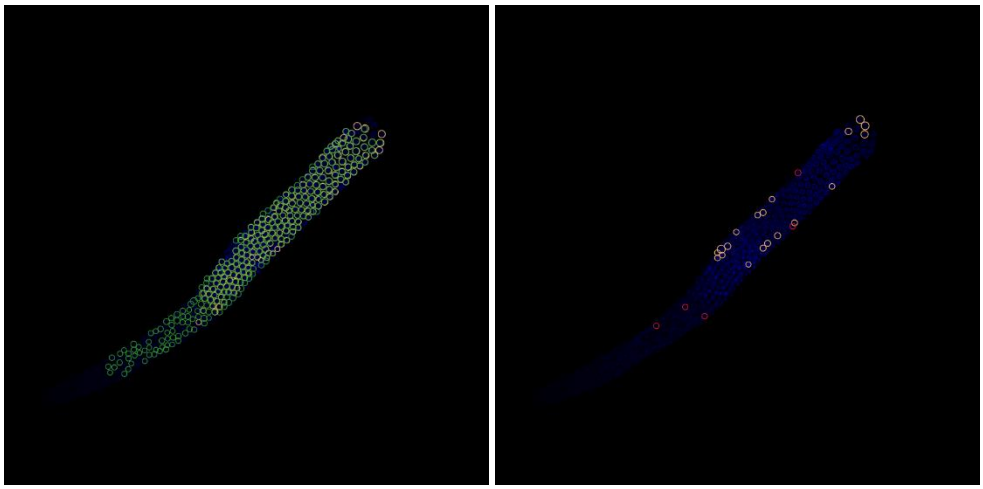


Figure S1: Detection of germ cells in z-stacks of germlines. Each row is an image of the same germline taken at a different depth. The distal tip is at the bottom-left of each image. The green circles in the left column are the cells that were correctly identified by the automatic segmentation algorithm, and the orange circles are the corresponding manually-segmented cells. Green circles without an orange partner correspond to cases in which the cell was first detected in that slice by the automatic segmentation algorithm, but first detected in the slice above or below that one by the manual segmentation. The red circles in the right column are the objects that were incorrectly identified as cells by the automatic segmentation, while the orange circles are the cells that were identified by the manual segmentation but not by the automatic one.

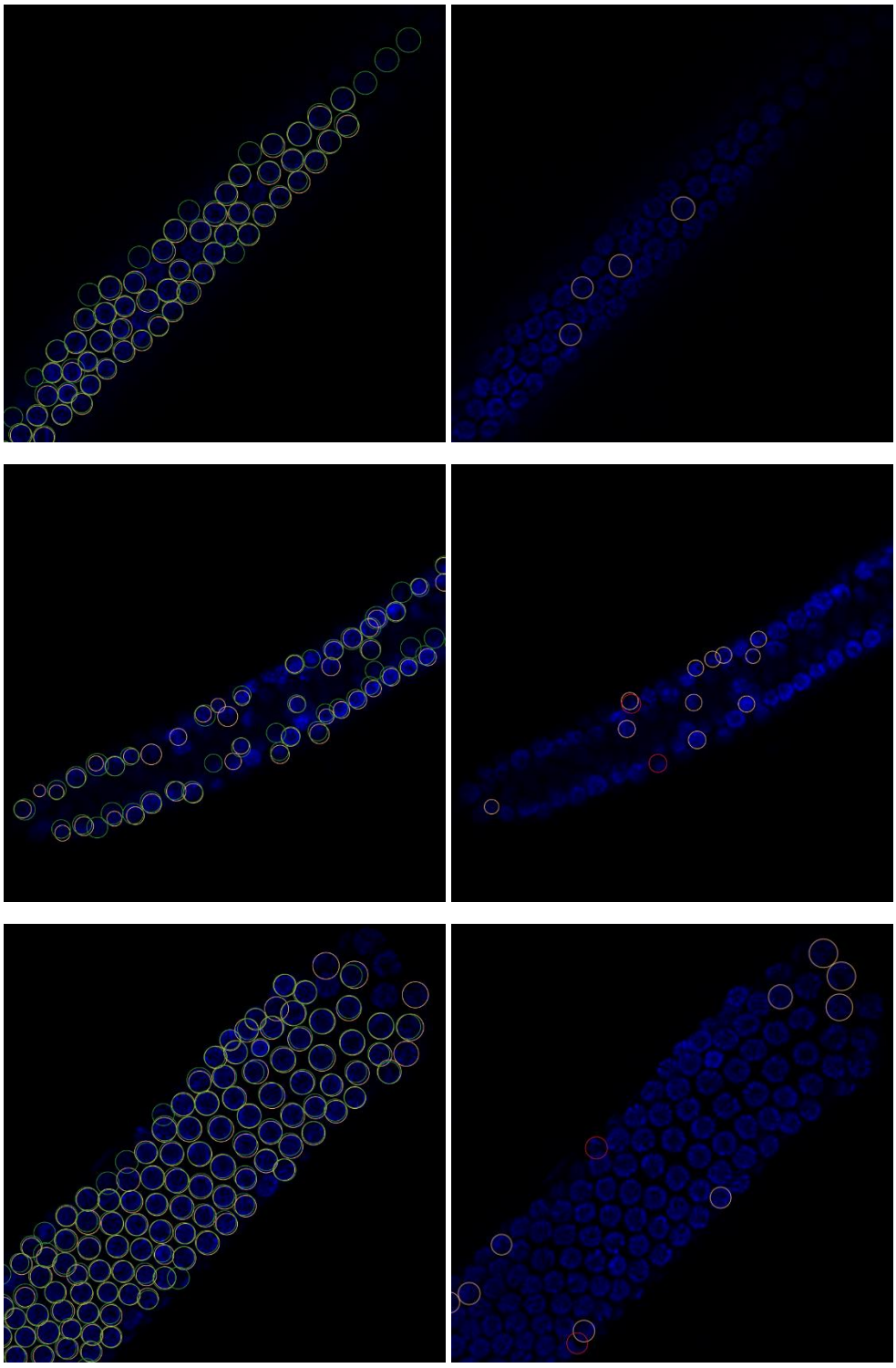


Figure S2: Zoomed-in regions of the images in Figure S1. Circle colors have the same meaning as in Figure S1. The second row is zoomed in on the distal tip. The third row is zoomed in on the pachytene region, just before the loop.

Section S2: Diffusion Maps Algorithm for Ordering Cells along the Center Line

In the Diffusion Maps algorithm (1), a random walk is constructed over a set of data points, with hopping probabilities between pairs of points determined by their pairwise distances and a kernel function. Here, the data points were the positions of the germ cells in a single germline. When the data points lie on a lower-dimensional manifold, the algorithm produces a robust ordering of the data along its principal nonlinear axis (or axes) on the manifold. In the case of the germline, the germ cells essentially lie on a 1D manifold, the center line of the germline. The algorithm was implemented using custom MATLAB code.

First, weights between data points (germ cell positions) were calculated by passing their pairwise distances through a Gaussian kernel function. Weights between data points are related to the probability of a random walker jumping from one of those data points to the other, with higher weights corresponding to higher hopping probabilities. The width of the Gaussian kernel determines the relevant scale of hopping. If the kernel width is much smaller than even the smallest distance between cells, then all weights between data points will be near zero, and a random walker cannot jump between any pair of data points. As the width is increased from zero, there is a scale at which the data appear 3D, then 2D, then 1D. At the 2D scale, a hopper can jump across the entire depth of the flattened germline in one jump, but not the diameter or length of the germline. At the 1D scale, the hopper can jump across the entire diameter of the germline in one jump, but not the length of the germline. If the kernel width is larger than the entire germline, all weights between data points will be close to one, and a random walker can jump between any two data points, no matter how far apart they are in space. At this scale, the data is essentially zero-dimensional from the point of view of the hopper, all collapsing to a single point. Previous work has developed an automated way of choosing the kernel width (2, 3). In practice, Diffusion Maps is not sensitive to the precise value of the kernel width, as long as it is in the correct dimensionality regime. We chose the kernel width for each germline so that the data "appeared" one-dimensional to a random walker.

The weights were then assembled into a symmetric matrix, with entry (i, j) containing the weight between germ cell i and germ cell j. The rows were normalized so that the sum of each row equaled one. This normalized matrix can be interpreted as a Markov transition matrix, with entry (i, j) containing the probability of a hopper located at data point i jumping to data point j in one time step. As the number of data points approaches infinity, the eigenvectors of this Markov matrix approach the eigenfunctions of the Laplace (diffusion) operator with Neumann (reflecting) boundary conditions (2). The first eigenvector of this matrix is a vector of ones, and contains no information. The first nontrivial eigenvector is one-to-one with and parameterizes the principal nonlinear axis of the data, the center line of the gonad tube. Element i of this eigenvector is associated with germ cell i; therefore, the monotonic ordering of the elements gives the ordering of the cells according to their positions along the center line. Diffusion Maps does not give the arc length positions of the cells, only their ordering.

Section S3: Estimating the Apoptosis Term, $\boldsymbol{R(x)}$

Evaluation of Equation 5 of the main text requires an expression for $R(x)$, which accounts for cell death in the germline. It is common to observe several germ cells undergoing apoptosis in a given fixed germline. Apoptotic germ cells are recognizable because their chromatin condenses and they undergo cellularization (4, 5). The former causes the cells to exhibit strong fluorescence when stained for DNA, while the latter causes the cells to exhibit essentially no fluorescence when stained for dpMPK-1 (Figure S1). From the time that a germ cell first shows symptoms of apoptosis to the time that the cell is removed from the germline by sheath cells is about 1 hour (4). In this time, a dying germ cell can only travel about one cell diameter before being cleared from the germline (6). As a result, the frequency at which cell corpses are observed at a given position is essentially the same as the frequency at which cells undergo apoptosis at that position.

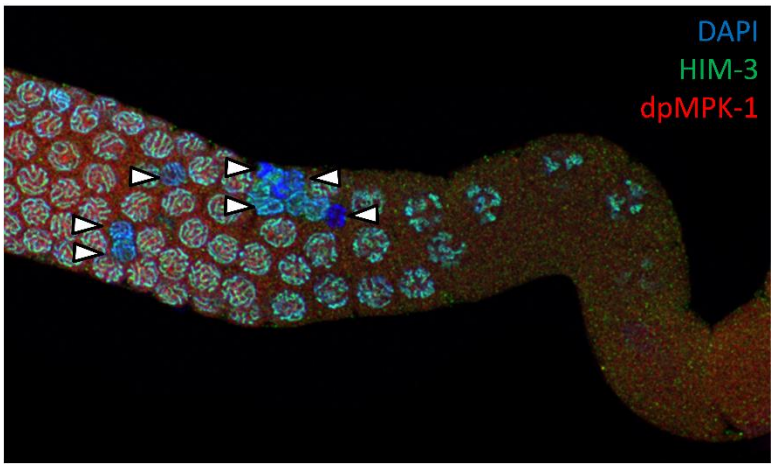


Figure S3: Image of the loop region of a *C. elegans* germline. Apoptotic cells (denoted by white arrowheads) are recognizable by their strong DAPI signal and lack of dpMPK-1 signal.

By this argument, we assumed that the death/clearance rate $R(x)$ is proportional to the number of cell corpses observed at x. We can decompose $R(x)$ into a shape function times a constant of proportionality that determines its scale. Under our assumptions, the shape function is given by the probability distribution of cell corpse locations (Figure S2). The constant of proportionality is r, the total rate of cell death, which can be estimated from data in the literature, as discussed in the main text. Note that the source term $S(x)$ could be estimated in a similar manner by looking at the relative frequency of cell divisions as a function of arc length across germlines, but this was not explored here.

However, in a given germline there are too few corpses to estimate the shape of $R(x)$ accurately. To address this, we pooled corpse counts from multiple fixed germlines to estimate the average shape of $R(x)$ over multiple germlines. The pooling process, itself, requires aligning spatial positions across different germlines. Here we will describe a method for registering arc length positions across different germlines. This is the only part of our approach that requires averaging across germlines.

Section S4: Registering Arc Length Positions across Germlines

Since there are so few cell corpses in a given germline, we need to estimate the shape of $R(x)$ from multiple germlines. Germlines come in different sizes, so pooling corpse positions across germlines requires that arc length positions in different germlines be registered or transformed to a common axis. Registering positions between two germlines is equivalent to determining an invertible function that maps positions in one germline to corresponding positions in the other. We assume that germ cells at "corresponding positions" are at the same developmental maturity, are the same age, and have spent the same amount of time in their respective germlines. Under the assumption that germ cells are arranged according to their maturity, this invertible function exists. The mapping will locally stretch or compress positions in one germline, like an accordion, to match the corresponding positions in the other. There will be a different mapping between each pair of germlines.

If x_1 refers to arc length positions in the first germline and x_2 refers to those in the second. The goal is to estimate an invertible function that maps x_1 to x_2, i.e. $x_2 = g(x_1)$ and $x_1 = g^{-1}(x_2)$. In the Derivation section, we introduced the probability density of germ cell arc length positions, $f_X(x)$, which quantifies the local "concentration" of germ cells in a particular germline. A related quantity is the cumulative distribution of arc length positions, $F_X(x)$, which quantifies the cumulative fraction of germ cells located at or before x.

Since x_1 and x_2 are related by an invertible function, their cumulative distributions must satisfy:

$$F_{X_1}(x_1) = F_{X_2}(x_2). \tag{S1}$$

Proof of Equation S1:

1. $F_{X_1}(x_1) = P(X_1 \leq x_1)$ (definition of a cumulative distribution function)
2. $= P(g(X_1) \leq g(x_1))$ (applying $g(\cdot)$ to both sides, and noting that $g(\cdot)$ is invertible, one-to-one and onto, and monotonically *increasing*)
3. $= P(X_2 \leq x_2)$ (using $X_2 = g(X_1)$ and $x_2 = g(x_1)$)
4. $= F_{X_2}(x_2)$. (definition of a cumulative distribution function)

Therefore, the invertible function we are seeking is $x_2 = g(x_1) = F_{X_2}^{-1}\left(F_{X_1}(x_1)\right)$. For each germline, $F_X(x)$ and its inverse are both measurable from data, meaning the mapping between any pair of germlines is measureable.

Using this approach, germ cell corpse positions from all germlines were transformed to their corresponding positions in a single germline. The particular germline used does not affect the result. The shape of $R(x)$ was estimated from these corpse positions the same way as $f_X(x)$, by kernel density estimation. This shape function for $R(x)$ was then transformed back to the arc length axis of each germline.

Note that the local source term $S(x)$ in Equation 1 of the main text can also be estimated in a similar fashion by identifying the positions of mitotically dividing cells in multiple germlines and registering the positions across germlines. Recent work suggests, though, that production throughout the mitotic region is roughly uniform (7).

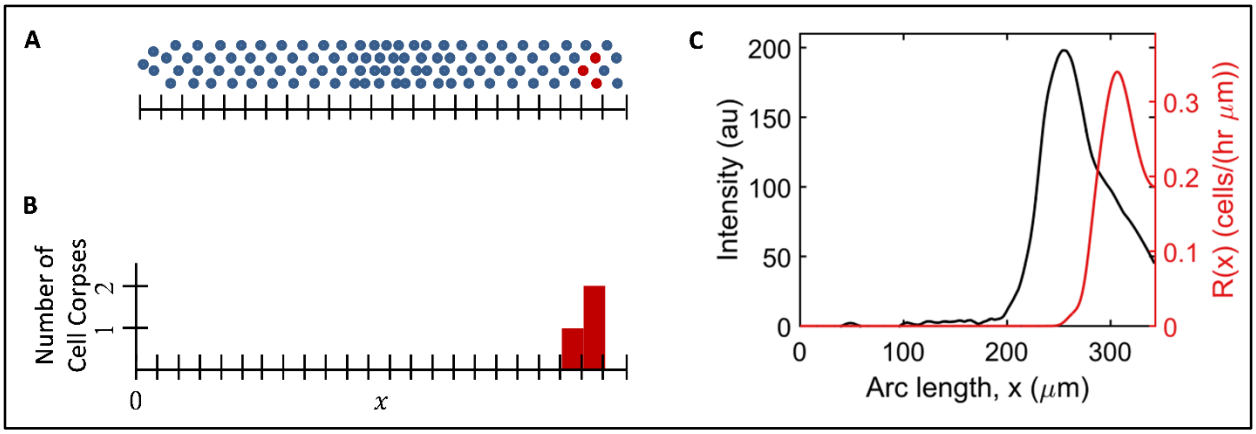


Figure S4: Estimation of $R(x)$. Assuming that the time rate at which cells undergo apoptosis at a given location is proportional to the frequency with which corpses are observed at that location, the shape of the spatially-dependent sink term $R(x)$ is the probability density function of cell corpse locations. **A)** Shows a schematic of a germline, where red cells are undergoing apoptosis. **B)** The histogram of cell corpse positions. With enough corpses, normalizing this histogram would give a good approximation of the shape of $R(x)$. Since there are not enough corpses in a single germline to estimate $R(x)$, corpse positions from multiple germlines must be aligned and pooled. **C)** Plotted in red is $R(x)$, estimated from 63 cell corpses pooled from 6 germlines and plotted against arc length position in a particular germline. The spatial dpMPK-1 profile from the same germline is shown in black. This shows that the peak rate of cell death occurs spatially (and temporally) after the peak of the dpMPK-1 pulse.

Section S5: Error Analysis

Uncertainty in $f_X(x)$

Assuming that we can measure germ cell arc length positions accurately, uncertainty in the probability density functions $f_X(x)$ for each germline can be approximated by their root mean squared error (RMSE) in the asymptotic limit of many samples. The mean squared error (MSE) of the estimate of $f_X(x)$ is the squared bias of the estimate (introduced by oversmoothing the true function) plus the variance of the estimate (introduced by estimating the function from a finite set of observations). When kernel density estimation is used to estimate probability density functions, the expression for the MSE is (8):

$$\delta f_X(x)^2 = \frac{h^4\left(f_X{}''(x)\right)^2}{4} + \frac{R f_X(x)}{nh}. \qquad \text{(S1)}$$

The first term in the sum is the squared bias of the estimate of $f_X(x)$, and the second term is the variance of the estimate. h is the bandwidth of the smoothing kernel used in the density estimation, and n is the number of observations (here, the number of cells in a germline). ${f_X}''(x)$ is the second derivative of the density, meaning that regions of the density function with larger curvature are more difficult to estimate accurately. This quantity was calculated by fitting the estimates of $f_X(x)$ with splines (MATLAB *csapi*) and taking the second derivative of the spline (MATLAB *fnder*). Finally, R is a property of the kernel function used in density estimation; for a kernel function $g(u)$, $R = \int_{-\infty}^{\infty} g(u)^2 du$. Here, an approximately Gaussian kernel was used, for which $R = 1/2\sqrt{\pi}$. Technically, density estimation was done via solving a diffusion equation, which acts much like a Gaussian smoothing kernel, but with better estimates of the density near the boundaries of the domain. This expression for the MSE should overestimate the error of the estimate near the boundaries.

Uncertainty in $t(x)$

Uncertainty in the estimates of $t(x)$ propagate from: uncertainty in $f_X(x)$, uncertainty in the values of the parameters in the model, and uncertainty in the shape of the apoptosis function $R(x)$.

Uncertainty in the model parameters was accounted for by uniformly sampling the literature ranges for N_{tot}, s, and the rate of ovulation (used to calculate r). Sampling was performed using a Latin Hypercube design (MATLAB's *lhsdesign*) to generate 100,000 samples. The value of τ produced by each parameter combination was calculated; if the value of τ was outside of the literature reported range for τ (48-54 hrs (9)), then the parameter set was discarded. After this pruning, 31,552 parameter sets remained. This collection of acceptable parameter sets was sampled from during the next step.

To estimate the effect of uncertainty in the shape of the apoptosis function $R(x)$, 5,000 samples were bootstrapped per germline from the collection of corpse observations. The shape of $R(x)$ was calculated for each randomly-sampled set of corpses. Then, for each sample, a parameter set was drawn at random from the collection of acceptable parameter sets, with replacement. Finally, $t(x)$ was calculated for that set of corpses, that parameter set, and that germline. The result was 5,000 estimates of $t(x)$ for each germline, the distribution of which accounted for uncertainty in the model parameters and the shape of $R(x)$. We denote the standard deviation of this distribution, as a function of x, $\delta t_{boot}(x)$.

The total uncertainty in $t(x)$, for each germline, is given by:

$$\delta t(x)^2 = \delta t_{boot}(x)^2 + \left(\frac{dt}{df_X}\right)^2 \delta f_X(x)^2, \tag{S2}$$

where $\frac{dt}{df_X}$ is, from Equation 5 in the main text, $\frac{dt}{df_X} = \int_0^x \frac{N_{tot}}{s - \int_0^w R(u)du} dw$, and $\delta f_X(x)$ is the RMSE of $f_X(x)$. This quantity was calculated separately for each germline.

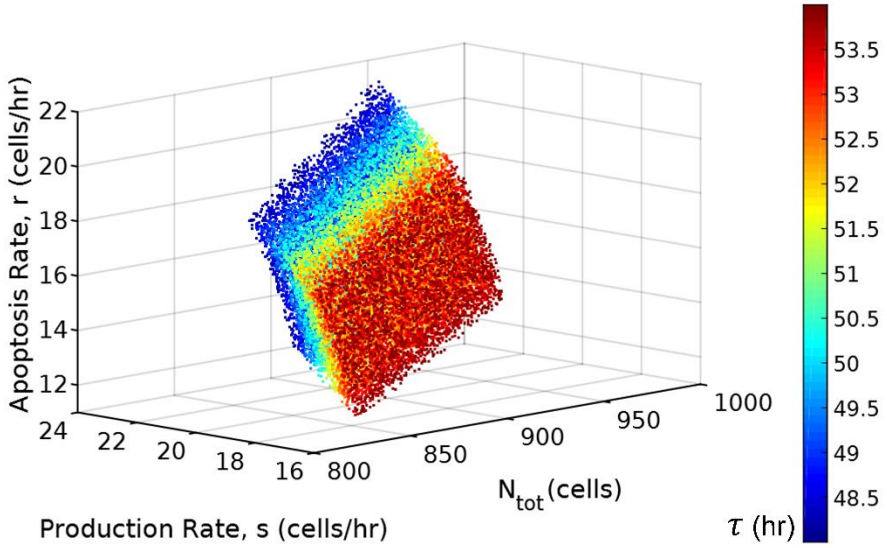


Figure S5: The collection of acceptable samples of N_{tot}, s, and r. Colors correspond to the value of τ, in hours, calculated from that parameter set, using Equation 7 of the main text. Axes limits are the ranges consistent with the literature. Using knowledge from the literature of all four parameters significantly reduces the volume of acceptable parameter combinations. The resulting region is called the feasible set (10).

Uncertainty in MPK-1 activation dynamics

Uncertainty in the dynamics of dpMPK-1 estimated from fixed samples arose from measurement uncertainty of the antibody staining and propagation of uncertainty from the time estimates. The measurement uncertainty was taken to be the standard deviation of the nuclear dpMPK-1 intensity measurements around the smoothed dynamics for that germline. If $y(t)$ is the fluorescence intensity of dpMPK-1 with respect to time and $\delta y_{measure}$ is the measurement uncertainty, then the total uncertainty in $y(t)$ is:

$$\delta y(t)^2 = \delta y(t)^2_{measure} + \left(\frac{dy}{dt}\right)^2 \delta t(x)^2, \qquad \text{(S3)}$$

where dy/dt is the derivative of $y(t)$ with respect to t, and δt is the uncertainty in t. This derivative was calculated by fitting $y(t)$ with splines (MATLAB *csapi*) and taking the derivative (MATLAB *fnder*). This calculation was performed for each germline.

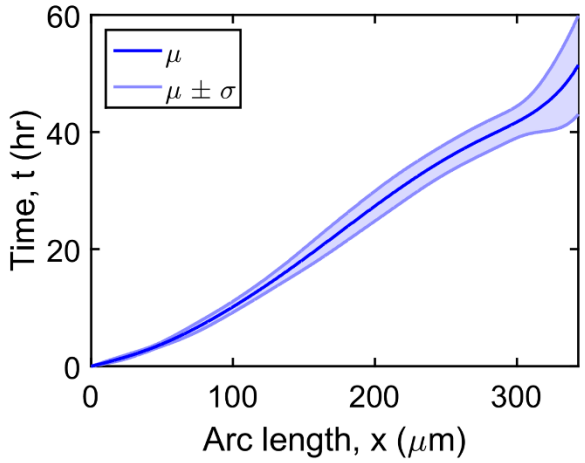


Figure S6: Sample $t(x)$ from a germline, plus and minus one standard deviation $\delta t(x)$, which accounts for uncertainties propagated from errors in estimating $f_X(x)$, uncertainty in the parameter values, and uncertainty in the shape of $R(x)$.

Section S6: Additional Figures

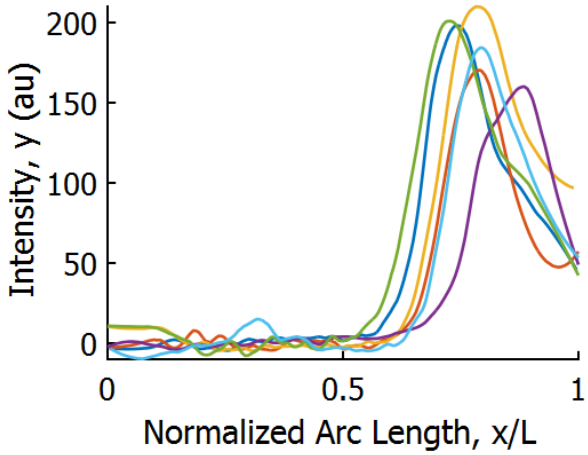


Figure S7: dpMPK-1 fluorescence intensity versus normalized arc length (arc length divided by the total distance from the distal tip to the loop) for multiple germlines. The images were acquired in the same experiment, at a set microscope condition. Plotting this way does not cause the spatial dpMPK-1 profiles from multiple germlines to collapse.

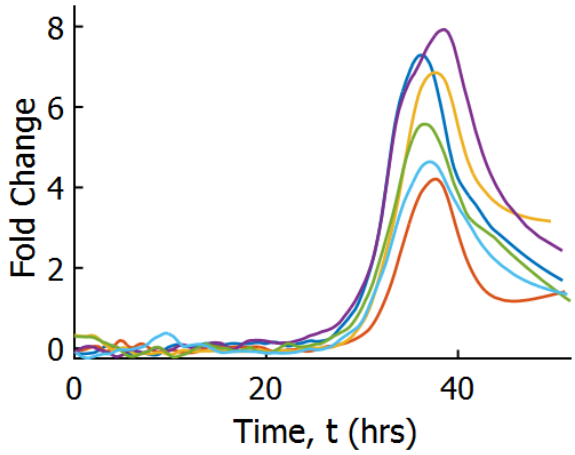


Figure S8: The fold change in dpMPK-1 fluorescence intensity, relative to background levels, as a function of time. Background fluorescence was calculated by averaging the dpMPK-1 intensity in the mitotic and early meiotic region, where there should be much less active MPK-1 than in the pachytene region.

SUPPORTING REFERENCES

1. Coifman, R. R., and S. Lafon. 2006. Diffusion maps. Appl Comput Harmon A 21:5-30.
2. Coifman, R. R., Y. Shkolnisky, F. J. Sigworth, and A. Singer. 2008. Graph Laplacian tomography from unknown random projections. IEEE T Image Process 17:1891-1899.
3. Ferguson, A. L., A. Z. Panagiotopoulos, P. G. Debenedetti, and I. G. Kevrekidis. 2010. Systematic determination of order parameters for chain dynamics using diffusion maps. P Natl Acad Sci USA 107:13597-13602.
4. Gumienny, T. L., E. Lambie, E. Hartwieg, H. R. Horvitz, and M. O. Hengartner. 1999. Genetic control of programmed cell death in the Caenorhabditis elegans hermaphrodite germline. Development 126:1011-1022.
5. Lant, B., and W. B. Derry. 2013. Methods for detection and analysis of apoptosis signaling in the C. elegans germline. Methods 61:174-182.
6. Crittenden, S. L., K. A. Leonhard, D. T. Byrd, and J. Kimble. 2006. Cellular analyses of the mitotic region in the Caenorhabditis elegans adult germ line. Mol Biol Cell 17:3051-3061.
7. Fox, P. M., and T. Schedl. 2015. Analysis of germline stem cell differentiation following loss of GLP-1 notch activity in Caenorhabditis elegans. Genetics 201:167-184.
8. Hansen, B. 2009. Lecture Notes on Nonparametrics. University of Wisconsin. Web. Oct. 2015.
9. Jaramillo-Lambert, A., M. Ellefson, A. M. Villeneuve, and J. Engebrecht. 2007. Differential timing of S phases, X chromosome replication, and meiotic prophase in the C-elegans germ line. Dev Biol 308:206-221.
10. Frenklach, M., A. Packard, P. Seiler, and R. Feeley. 2004. Collaborative data processing in developing predictive models of complex reaction systems. Int J Chem Kinet 36:57-66.