Stochastic Methods for Inferring States of Cell Migration

R. J. Allen[1†], C. Welch[1†], Neha Pankow[1], Klaus M. Hahn[1,2] and Timothy C. Elston[1,2*]*

[1] *Department of Pharmacology, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, [2] Computational Medicine Program, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States*

Cell migration refers to the ability of cells to translocate across a substrate or through a matrix. To achieve net movement requires spatiotemporal regulation of the actin cytoskeleton. Computational approaches are necessary to identify and quantify the regulatory mechanisms that generate directed cell movement. To address this need, we developed computational tools, based on stochastic modeling, to analyze time series data for the position of randomly migrating cells. Our approach allows parameters that characterize cell movement to be efficiently estimated from cell track data. We applied our methods to analyze the random migration of Mouse Embryonic Fibroblasts (MEFS) and HeLa cells. Our analysis revealed that MEFs exist in two distinct states of migration characterized by differences in cell speed and persistence, whereas HeLa cells only exhibit a single state. Further analysis revealed that the Rho-family GTPase RhoG plays a role in determining the properties of the two migratory states of MEFs. An important feature of our computational approach is that it provides a method for predicting the current migration state of an individual cell from time series data. Finally, we applied our computational methods to HeLa cells expressing a Rac1 biosensor. The Rac1 biosensor is known to perturb movement when expressed at overly high concentrations; at these expression levels the HeLa cells showed two migratory states, which correlated with differences in the spatial distribution of active Rac1.

Keywords: cell migration, stochastic modeling, RHOG, Rac1, biosenor, migration states

INTRODUCTION

The ability of cells to move is essential to many biological processes, such as tissue development, the immune response and wound healing (Franca-Koh et al., 2007; Petrie et al., 2009; Cain and Ridley, 2012). Anomalous cell migration plays a role in diseases, such as cancer and atherosclerosis (Cain and Ridley, 2012; Hall, 2009; Lemarié et al., 2010; Finney et al., 2017). During cell migration, intracellular signaling networks tightly control the spatiotemporal dynamics of the cytoskeleton. In particular, the Rho family of small GTPases has been implicated in membrane protrusion, adhesion, contraction and de-adhesion, all steps necessary for cell migration (Rottner et al., 1999; Jaffe and Hall, 2005; Goley and Welch, 2006; Ridley, 2006; Iden and Collard, 2008; Ladwein and Rottner, 2008). Rac1, the family member studied here, produces cell protrusions by interacting with effector proteins that modulate actin polymerization, including formins and Paks. A prevailing hypothesis is that Rac1 induces localized actin polymerization to trap random, thermal driven outward movements of the cell edge (Ridley, 2015; Marston et al., 2019; Schaks et al., 2019).

During random cell migration, in which cells do not experience directional environmental cues, cells move in a persistent manner, but with significant variability in their direction and speed. Therefore, methods for quantifying cell movement that take into account the stochastic nature of this phenomenon are needed. Previous studies have analyzed cell migration in terms of quantitative metrics such as the mean squared deviation in cell position, which can be linked to both speed and persistence (Othmer et al., 1988; Dimilla et al., 1992; Rosello et al., 2004; Dieterich et al., 2008). Additionally, it has been suggested that fractional diffusion models are required to accurately describe cell movement (Dieterich et al., 2008). We refer the reader to a recent review which describes these approaches and others (Svensson et al., 2018). We used stochastic modeling to develop tools for quantifying cell migration such that it can be characterized in terms of biologically relevant parameters. In our approach, the motion of cells is assumed to follow a 2D random walk with persistence. A related method that takes into account the probability of turning and contains a parameter related to persistence also has been applied to analyze random cell migration (Arrieumerlou and Meyer, 2005). An important distinction of our approach is that our model allows for the possibility of multiple states of migration, distinguished by differences in speed and persistence. This feature allowed us to determine that Mouse Embryonic Fibroblasts (MEFS) exist in two distinct states during random migration. Knock down of the Rho-GTPase RhoG suggests that this protein plays an important role in establishing the two states. We next demonstrated how our method allows the migration state of a cell to be predicted from time series data. Finally, we applied our method to examine the activation of Rac1, a GTPase known to be important in producing localized protrusions. Interestingly, we found that overexpressed, biosensor induced two states of migration in HeLa cells that correlated with different numbers of active Rac1 foci.

RESULTS

Preliminary Analysis

To develop our methods, we collected data sets that consisted of time series for the x and y coordinates of the cell centroids of randomly migrating MEF cells (**Figures 1A,B**). We chose this cell type because it shows persistent migration in the absence of directional cues. As an initial analysis of the data, we computed the average persistence of cell movement defined as $P = < \cos(\theta) >$, where θ is the change in the direction of cell movement between measurements (**Figure 1C**) and the angular bracket denotes averaging over cell tracks. If θ is uniformly distributed, then the motion of the cell lacks persistence and $P = 0$. This behavior would be consistent with a pure random walk (diffusive motion). For values of P greater than zero, the movement of the cell shows persistence, with a value of 1 indicating motion in a straight line. Combining the cell tracks for individual cells, produced a value of $P = 0.43$. This value is consistent with cells that show persistent motion. We also generated histograms from the Δx and Δy displacements and empirically calculated
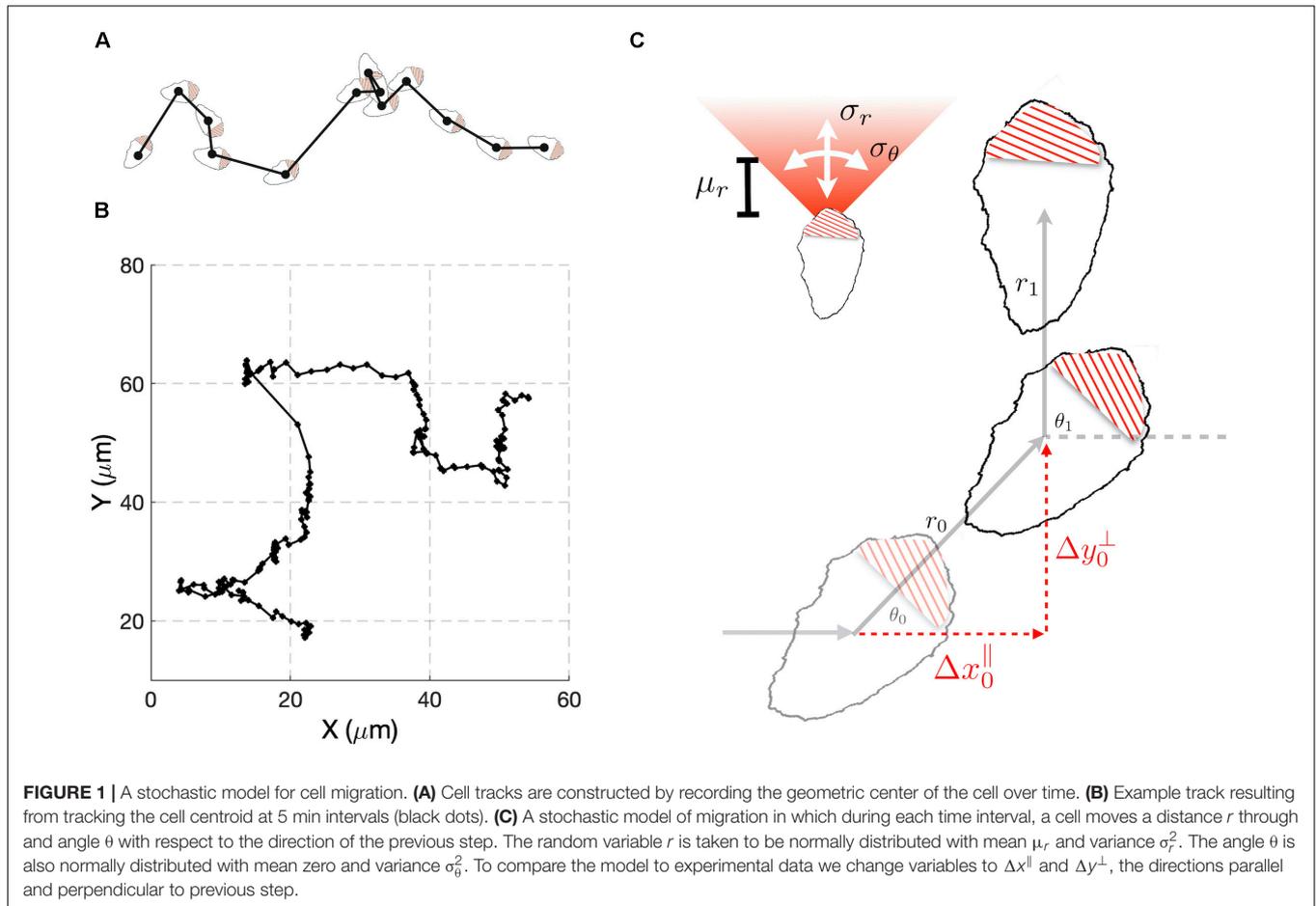
cumulative density functions (**Supplementary Figure S1**, top left panel). These distributions were found to show slight deviations from a Gaussian distribution.

A Stochastic Model for Cell Migration

Our preliminary cell track analysis led us to model cell movement as a 2D random walk with persistence (**Figure 1C**). In our model, for each time interval i, the distance, r_i, traveled by a cell and the angle, θ_i, through which the cell moves are considered random variables. The random variable r_i is taken to have a Gaussian distribution characterized by mean μ_R, and variance σ_R^2. We allowed for negative values of r_i to account for the scenario in which a cell maintains its direction of polarization, but its centroid moves in a rearward direction. The directional angle θ_i, is also taken to have a Gaussian distribution with variance σ_s^2, and centered on the value of the previous angle θ_{i-1}. Small values of σ_s^2 correspond to highly persistent migration. For large values of σ_s^2 the new direction becomes uniformly distributed on the interval $[-\pi, \pi]$ and the model represents a purely diffusive process.

It is not possible to tell from cell track data alone if changes in θ_i of magnitude greater than $\pi/2$ resulted from large deviations in orientation or negative r_i. Thus, the probability distribution for these variables cannot be constructed unambiguously from the cell track data. To overcome this difficulty, we performed a change of variables from (r_i, θ_i) to $(\Delta x_i^{\parallel}, \Delta y_i^{\perp})$, where these new variables correspond to changes in the centroid's position during the ith time interval that are parallel and perpendicular to the direction of the previous step (**Figure 1C**). An important feature of the model is that analytical expressions for the probability density functions (pdfs) of $\Delta x_i^{\parallel}$ and $\Delta y_i^{\perp}$ can be found (**Supplementary Information**), allowing estimation of model parameters from experimental data to be performed in a computationally efficient manner, relative to the alternative of estimating probability density functions via repeated simulation of the stochastic model (**Figure 1C**). These co-ordinates explicitly handle the degeneracy in θ_i described above, because in these co-ordinates all possibilities that could have led to a given observation are considered. If cells show persistent motion, $\Delta x_i^{\parallel}$ has a positive mean value. Also, if there are no external cues in the experiments to define a preferred direction of motion, $\Delta y_i^{\perp}$ is symmetric about zero. Therefore, the distribution for $\Delta x_i^{\parallel}$ is more informative, and we use it to compare experimental results with the model's behavior. It is possible to simultaneously fit the $\Delta x_i^{\parallel}$ and $\Delta y_i^{\perp}$ distributions, but this comes at an increased computational cost. As a consistency check, after performing parameter estimation, we verify that the model accurately captures the $\Delta y_i^{\perp}$ distribution. If the model failed this consistency test, we could repeat the parameter estimation using both distributions. However, this was not required for any of the cases considered here.

We used a Monte Carlo method based on the Metropolis algorithm to perform parameter estimation. This was followed by local optimization algorithms to identify parameters associated with the global minimum error between the model and data (**Supplementary Information**). To test the accuracy and

FIGURE 1 | A stochastic model for cell migration. **(A)** Cell tracks are constructed by recording the geometric center of the cell over time. **(B)** Example track resulting from tracking the cell centroid at 5 min intervals (black dots). **(C)** A stochastic model of migration in which during each time interval, a cell moves a distance r through and angle θ with respect to the direction of the previous step. The random variable r is taken to be normally distributed with mean μ_r and variance σ_r^2. The angle θ is also normally distributed with mean zero and variance σ_θ^2. To compare the model to experimental data we change variables to $\Delta x^{\parallel}$ and $\Delta y^{\perp}$, the directions parallel and perpendicular to previous step.
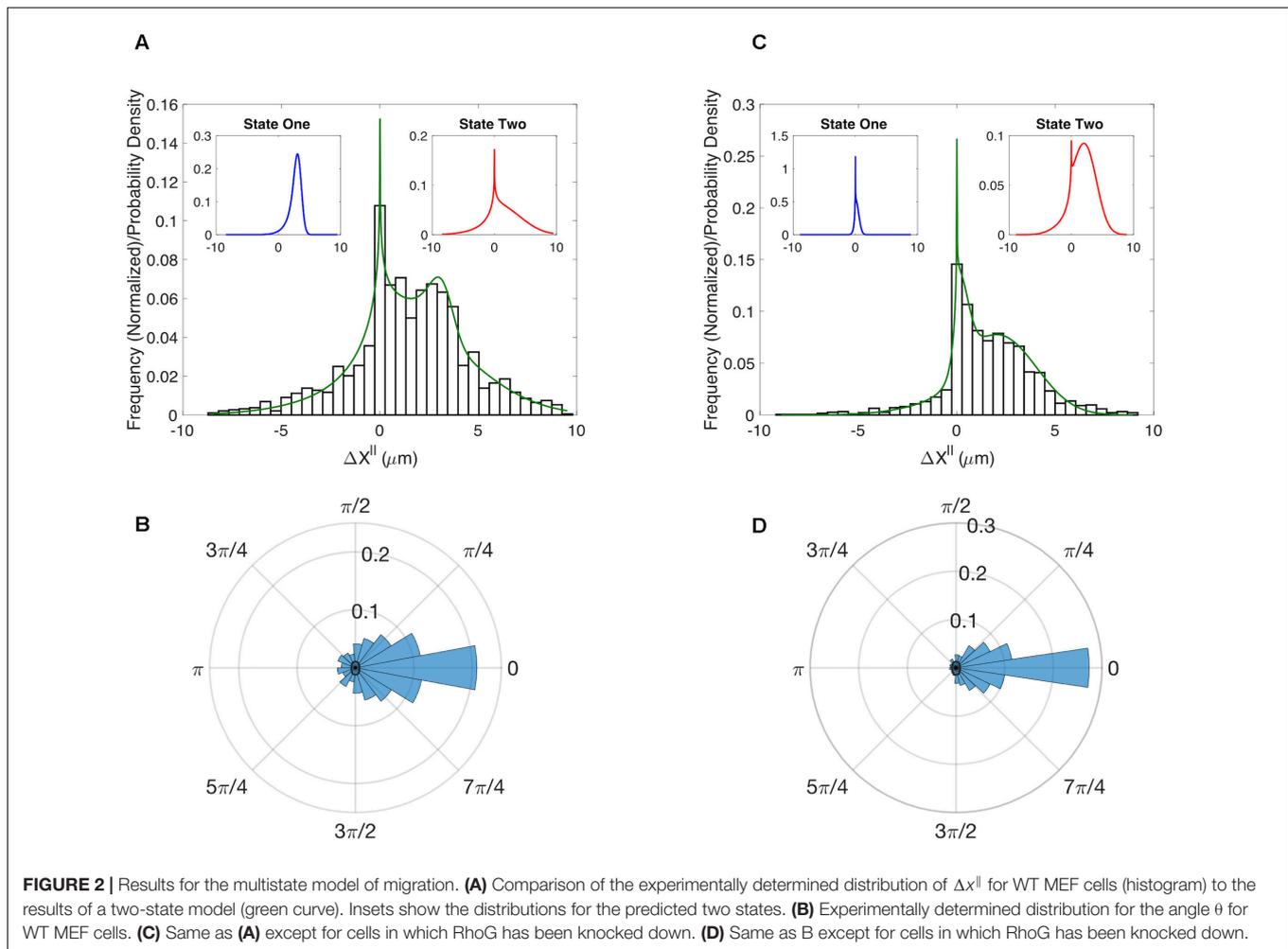
efficiency of this method, we benchmarked our approach using data generated from computational simulations of the stochastic model (**Supplementary Figure S2**). Having validated our computational methods, we next fit the model to the experimentally measured distributions. The model did not generate a good fit to experimental data for MEF cells (**Supplementary Figure S3**, dashed curve). In particular, we found that the model could not capture the second mode observed in the $\Delta x_i^{\parallel}$.

A Multistate Model for Cell Migration

Further inspection of the MEF cell tracks suggested that individual cells might exist in different modes of migration, distinguished by differences in speed and persistence. We therefore expanded our model to allow for different states of migration. That is, we hypothesized that at any given time a migrating cell is in one of n states denoted by S_i, with $i \in \{1 \dots n\}$. Each state is characterized by the parameters μ_r^i, σ_r^i, and σ_θ^i. The additional parameters, α^i, denoting the fraction of time spent in state i, are required to fully specify the model. Since $\sum \alpha^i = 1$, in the two-state case the total number of parameters is seven. Note that if a two-state model is fit to data consisting of only a single state, then we expect our Monte Carlo method to produce parameter sets in which α^1 takes on values of 0 or

1, or $\mu_r^1 = \mu_r^2$, $\sigma_r^1 = \sigma_r^2$, and $\sigma_\theta^1 = \sigma_\theta^2$. The extended model is essentially a mixture model, which is itself a reduced hidden Markov model under the assumption that the probabilities of transitioning between states are independent and identically distributed. We again used simulated data to validate the accuracy and efficiency of our Monte Carlo method when multiple states are considered (**Supplementary Figure S4**).

The multi-state model produced a good fit to the MEF $\Delta x_i^{\parallel}$ distribution (**Figure 2A**). To assess the accuracy of our parameter estimates we used confidence-interval profiling (Raue et al., 2009). To determine acceptable values for the sum of the squared errors (SSE) we boot-strapped the original datasets to assess plausible differences in our observed distributions should we repeat the experiments (**Supplementary Information**). The results of this analysis provide a measure of the confidence that should be placed on each estimated parameter value (**Supplementary Figures S5A,B**). Of particular interest is the parameter α which represents the fraction of time in each state. The best fits were achieved with α = 0.12. We confirmed that the model also captured the distributions for $\Delta y_i^{\perp}$ (**Supplementary Figure S6A**). The results of our analysis suggest that randomly migrating MEFs exist in one of two states. About 12% of time these cells are in a state with a well-defined characteristic step of ~3 μm (State 1 – blue distribution in **Figure 2A** left

FIGURE 2 | Results for the multistate model of migration. **(A)** Comparison of the experimentally determined distribution of $\Delta x^{\parallel}$ for WT MEF cells (histogram) to the results of a two-state model (green curve). Insets show the distributions for the predicted two states. **(B)** Experimentally determined distribution for the angle θ for WT MEF cells. **(C)** Same as **(A)** except for cells in which RhoG has been knocked down. **(D)** Same as B except for cells in which RhoG has been knocked down.

inset) and an angular distribution with $\sigma_{\theta}^{1} = 0.7$. In the second state, the step size is highly variable (State 2 – red distribution **Figure 2A** right inset) and the motion is less persistent $\sigma_{\theta}^{2} = 1.3$. For completeness, we also show the distribution for the angle θ (**Figure 2B**).

RhoG's Role in Migration

It has long been appreciated that the canonical Rho-GTPases RhoA, Rac1, and Cdc42 play important roles in cell migration. However, the role of RhoG in migration is less well studied. To determine if RhoG plays a role in the random migration of MEFs, we generated time series data for cells in which this protein was knocked down. While the angular distributions for the WT and knockdown do not show clear differences (**Figures 2B,D**), the $\Delta x_{\parallel}$ distributions indicate RhoG does effect migration (**Figures 2A,C**). Moreover, by fitting our simple two-state model we can quantify this effect and ascertain that the persistent state 1 in the MEF control has been converted to a state in which the cells do not show significant movement ($\mu_{r}^{1} = 3.2\,\mu$m for the WT to $\mu_{r}^{1} = 0.25\,\mu$m for the KD). State 2 seems to be preserved by the KD in the sense that the confidence intervals defining state 2 parameters are overlapping in the two

cases (**Supplementary Figure S5**). A putative mechanism for how RhoG activation influences cell migration via recruitment of the DOCK180/ELMO complex (Katoh and Negishi, 2003; Katoh et al., 2006), which acts as GEF for Rac1. However, whether this is the key pathway in this process, and how it is organized spatio-temporally, is a direction of future research.
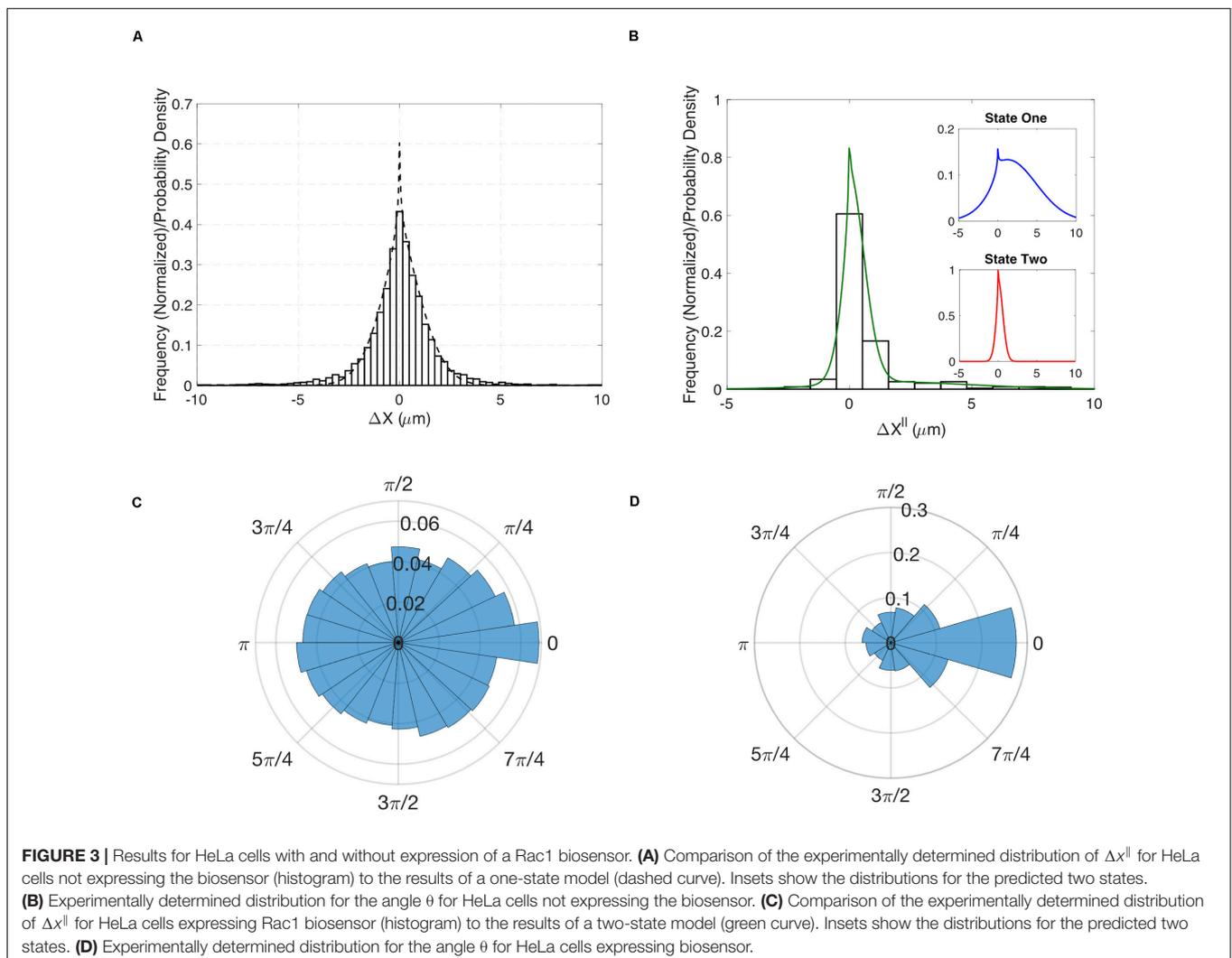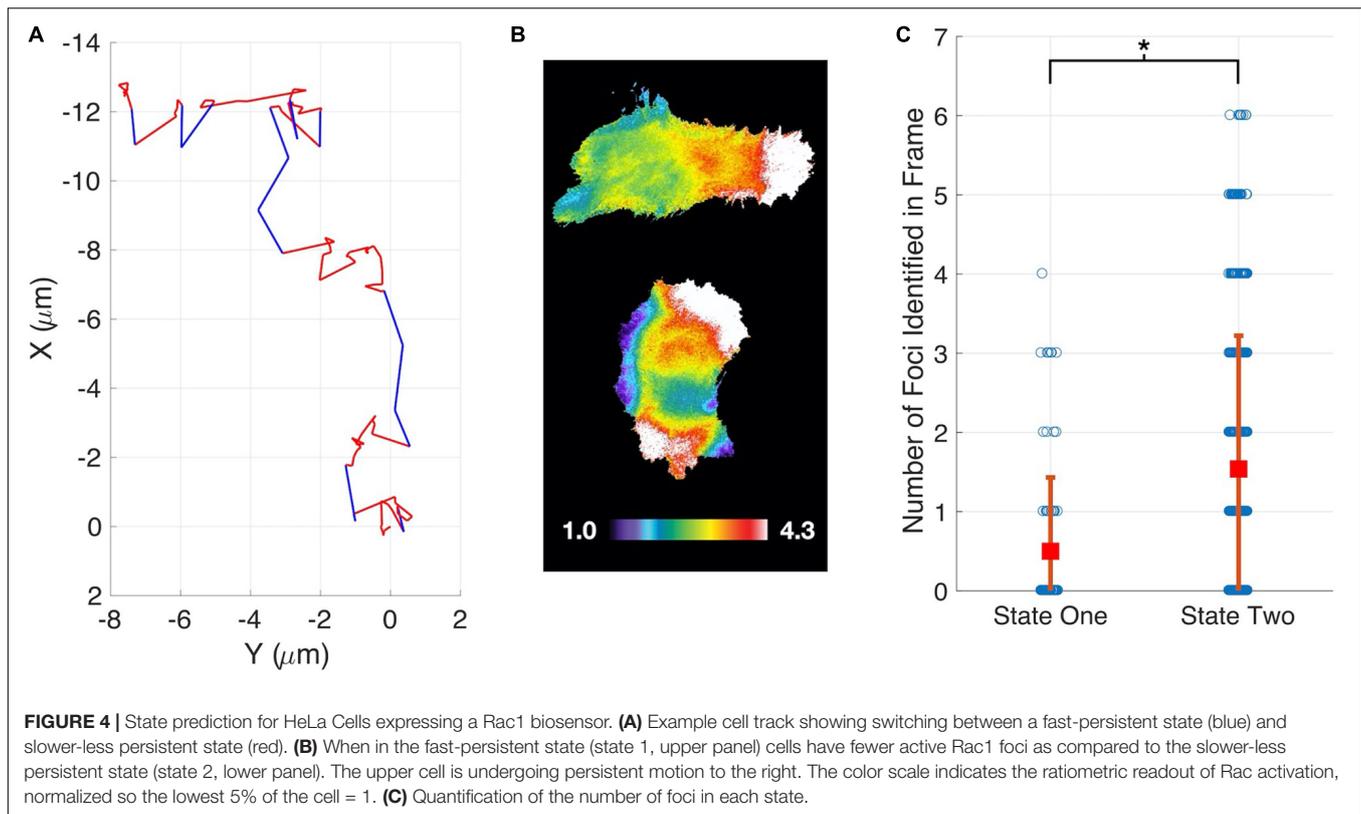
Inferring States From Time Series Data

We next sought to develop computational tools that could be used to determine if the predicted states of migration correspond to subpopulations of cells with distinct phenotypes or if individual cells could transition between states. To test if individual cells change their migration state, we developed a method to infer migration states from individual cell tracks. Our approach uses a Bayesian prediction method based on the probability that a sequence of k successive steps arises from one of the identified states (see Methods for details). Before applying our state prediction method on the experimental data, we first validated the approach using synthetic data. To generate this data, we performed computational simulations of the stochastic model using the parameters estimated from the experimental data for MEF cells. With these values our state-prediction

algorithm correctly identified the states more than 90% of the time, validating the approach (**Supplementary Figure S7**).

Having demonstrated our method's ability to infer cell migration states from simulated track data and demonstrate a role for RhoG, we examined whether the different migration states could be correlated with molecular changes within cells. The Hahn lab has used HeLa cells to develop new biosensors and optogenetic probes. It is well established that these molecular tools must be used at controlled concentrations, below levels where they perturb cell movement (Kraynov et al., 2000; Machacek et al., 2009). Controls in earlier studies have shown that HeLa cells exhibit altered motility when the Rac1 biosensor is expressed at high levels. We decided to investigate if our stochastic modeling approach could quantify the effects of biosensor overexpression. We compared WT HeLa cells without biosensor expression to cells with high levels of Rac1 biosensor. Our analysis revealed that WT cells showed little directed motion and a single migration state was sufficient to capture the distributions of steps sizes (**Figures 3A,B**). In contrast, cells with highest levels of biosensor exhibited two states of migration

(**Figure 3C**). In particular, two states were needed to capture the long tail of the distribution (see **Supplementary Figure S9** for comparison of one-and two state results). In state 1 the cell moves in persistent manner, whereas in state 2 the cell is mostly stationary. To test if the predicted two states are correlated with differences in cell signaling, we ran our state prediction algorithm on the track data. Interestingly, our analysis predicted that individual cells randomly switch between the two states (**Figure 4A**), and qualitative observations indicated that the slow state showed multiple disperse Rac activation events at the edge of the cell, while the fast state showed a single Rac activation at the leading edge (see **Supplementary Movies M1, M2**). To quantify this observation, we identified and counted the number of foci of active Rac1 in each image and grouped these counts by the predicted state (see **Supplementary Information** for details), reasoning that random movement would require more cell protrusions distributed around the cell perimeter. Rac activation is known to be sufficient to generate cell protrusions (Wu et al., 2009; Wang et al., 2016). Cells predicted be in state 1, which corresponds to the fast-persistent state, had fewer Rac1 foci than

FIGURE 3 | Results for HeLa cells with and without expression of a Rac1 biosensor. **(A)** Comparison of the experimentally determined distribution of $\Delta x^{\parallel}$ for HeLa cells not expressing the biosensor (histogram) to the results of a one-state model (dashed curve). Insets show the distributions for the predicted two states. **(B)** Experimentally determined distribution for the angle θ for HeLa cells not expressing the biosensor. **(C)** Comparison of the experimentally determined distribution of $\Delta x^{\parallel}$ for HeLa cells expressing Rac1 biosensor (histogram) to the results of a two-state model (green curve). Insets show the distributions for the predicted two states. **(D)** Experimentally determined distribution for the angle θ for HeLa cells expressing biosensor.

FIGURE 4 | State prediction for HeLa Cells expressing a Rac1 biosensor. **(A)** Example cell track showing switching between a fast-persistent state (blue) and slower-less persistent state (red). **(B)** When in the fast-persistent state (state 1, upper panel) cells have fewer active Rac1 foci as compared to the slower-less persistent state (state 2, lower panel). The upper cell is undergoing persistent motion to the right. The color scale indicates the ratiometric readout of Rac activation, normalized so the lowest 5% of the cell = 1. **(C)** Quantification of the number of foci in each state.

those predicted to be in state 2, which show little net movement (**Figure 4B**). This observation is consistent with highly motile cells typically showing strong polarity.

DISCUSSION

We developed novel computational methods for analyzing the movement of randomly migrating cells. Our approach combines stochastic modeling with statistical inference methods to detect and quantify migratory phenotypes. Migrating cells have a biochemical, morphological, and structural orientation that persists as these cells move. Our model captures this 'memory' by conditioning the cell's movement during the current time interval on its previous direction of motion. An important feature of our model is that analytic expressions for the probability densities for cell displacements parallel and perpendicular to the previous direction of motion can be found. This feature allows us to generate the probability density function for a given set of parameters rather than generating an approximation to this PDF via stochastic simulation of our migration model (**Figure 1C**). In most use cases the analytical PDF is computationally more efficient due to the high number of repeats required to estimate the PDF with sufficent accuracy. We have validated all our approaches using simulated data, and then applied the methodology to study randomly migrating MEF and HeLa cells.

Our modeling approach allows for multiple states of migration. This feature allowed us to demonstrate that migrating

cells randomly transition between modes of movement. Crucial to the detection of these states is the quantification of parameter values and the associated confidence in those estimates. This process allowed us to be confident in the existence of two states of migration for MEF and HeLa cells over-expressing a Rac1 biosensor.

The identification of multiple states of migration for MEF cells led us to assess the role of RhoG in establishing these states. To do this we used siRNA to reduce RhoG expression. This perturbation suggests that RhoG plays a role in directed migration, because reducing RhoG eliminated net movement in the first predicted state and shortened the range of step sizes taken in the second state. We next developed a Bayesian approach to predict the current migration state of a cell from time series of the cell's position. Using this method, we demonstrated that individual HeLa cells expressing a Rac1 biosensor switched between migratory states. Importantly, we were able to correlate these two states with differences in the distribution of Rac1 activity.

We believe that our methods provide useful tools for quantifying and characterizing cell migration. Our stochastic model characterizes cell migration using parameters with straightforward biological interpretations. Hence, application of this model can lead to biological insights not apparent in the data from visual inspection or simple quantitative measures. In this case, our analysis suggests a role of RhoG in allowing cells to change direction, which may play a role in the ability of randomly migrating cells to search their environment.

MATERIALS AND METHODS

Computational Methods

The full code and analysis for this paper is available (Allen, 2020).

Coordinate Transformation

We modeled cell migration as a stochastic sequence of steps characterized by the step size r_i and directional angle θ_i (**Figure 1**). Since we assume r_i and θ_i to be realizations of independent random variables R and Θ the probability the cell moves (r, θ) is defined by

$$f(r, \theta | \theta_{i-1}) = g_R(r).g_\Theta(\theta | \theta_{i-1}) \qquad (1)$$

where $g_R(r)$ is the probability density function (pdf) for the step magnitude, which we take to have the normal distribution $\mathcal{N}(\mu_r, \sigma_r^2 \Delta t)$, and $g_\Theta(\theta | \theta_{i-1})$ is the pdf generating the new orientation conditioned on the previous angle, which we take to have the normal distribution $\mathcal{N}(\theta_{i-1}, \sigma_\theta^2 \Delta t)$. The experimental data is collected in Cartesian coordinates (X, Y). In principle we could transform the data into the coordinates R and Θ. However this transformation cannot be completed uniquely, because there is no way to distinguish a backward step in which the cell maintains its direction of polarity ($\theta_i = \theta_{i-1}$) from one in which the front and back of the cell have reversed ($\theta_i = \theta_{i-1} + (2k + 1)\pi$). Furthermore, the value of θ_i cannot be determined if $r_i = 0$. For these reasons, we transform the model to the coordinates $(\Delta x_i^\parallel, \Delta y_i^\perp)$, where these new variables correspond to changes in the centroid's position during the ith time interval that are parallel and perpendicular to the direction of the previous step.

To compare with the model the data needs to be manipulated to generate histograms for steps in the $x^\parallel$ and $y^\perp$ directions. For each sequential triplet of coordinates $\{(x_{i-1}, y_{i-1}), (x_i, y_i), (x_{i+1}, y_{i+1})\}$, we rotate the steps as a rigid body about (x_{i-1}, y_{i-1}) by a four quadrant inverse tangent based on $\tan^{-1} \frac{y_i - y_{i-1}}{x_i - x_{i-1}}$. The result of this is that all steps are pre-oriented in a positive x-direction and initiated at $(0, 0)$, and can be plotted as histograms of step distance in the x and y direction: $(\Delta x_i^\parallel, \Delta y_i^\perp) = (x_{i+1}' - x_i', y_{i+1}' - y_i')$.

The pdf for $\Delta x^\parallel$ is:

$$f_X(\Delta x^\parallel) = \int_{-1}^{1} \frac{f_H(\Delta x^\parallel / h, h)}{h} dh \qquad (2)$$

where $h = \cos(\theta)$, and,

$$f_H(\Delta x^\parallel / h, h) = g_r(\Delta x^\parallel / h) \sum_k \frac{1}{(1 - h^2)^{\frac{1}{2}}} \times$$

$$(g_\Theta(\arccos(h) + 2\pi k) + g_\Theta(-\arccos(h) + 2\pi k)). \qquad (3)$$

The expression for $f_Y(\Delta y^\perp)$ is similar, however now with $g_\Theta(\arccos(h) + \pi/2 + 2\pi k) + g_\Theta(-\arccos(h) + \pi/2 + 2\pi k)$ in the summation term. A derivation of these results is presented in the **Supplementary Information**.

Parameter Estimation

Parameters were estimated by simulated annealing, which is a Monte Carlo method based on the Metropolis algorithm (24, 25).

Initial choices of parameters generate an analytical solution (Eq. 2), which is scored against the experimental data ($\Delta x_i^\parallel$) by the sum of least squared differences. At each step of the algorithm the parameters are updated by a small addition of Gaussian noise, if this update scores better than the current score then these parameters are accepted. If the score is higher, the parameter set is accepted with probability $e^{-\frac{\Delta s}{T}}$, where Δs is the difference between the current and previous scores and T is the current temperature. Over the course of the fitting T, the temperature is reduced. This fixes the parameter choices into a local minimum. Here we choose a geometric cooling regime. Due to the stochastic nature of the simulation, and that there could be many local minima, it is necessary to run this fitting procedure multiple times. The best fit of this routine was then further refined using MATLABs fmincon routine, which was also used to assess the sensitivity of our fit to altering parameter values via confidence-interval profiling (**Supplementary Figure S4**, **Supplementary Material** for details).

The histograms were amalgamated from multiple cell tracks. For the case of two states, the pdf for Δx becomes

$$f_X(\Delta x) = \alpha f_X^1(\Delta x) + (1 - \alpha) f_X^2(\Delta x)$$

where α is the fraction of time spent in state 1 and the distributions $f_X^1(\Delta x)$ and $f_X^2(\Delta x)$ are parameterized by $(\mu_r^1, \sigma_r^1, \sigma_\theta^1)$ and $(\mu_r^2, \sigma_r^2, \sigma_\theta^2)$, respectively.

Parameter sets were identified by multiple simulated annealing runs, followed by local-optimization routines.

Validation of Methods

To validate the pdfs and the parameter estimation algorithm, we simulated cell tracks using the stochastic model (**Figure 1**). Cell tracks were generated using two states, each with distinct parameter sets. At each step a state was chosen at random with probability 0.5. As above, the simulated cell tracks were used to construct the distributions for $\Delta x_i^\parallel$ and $\Delta y_i^\perp$. We assumed model parameters were not known and used the Monte Carlo method to fit Eq. 2, modified to two states (see below) to the simulated data for $\Delta x_i^\parallel$. The Monte Carlo method quickly converged on the correct parameter values (**Supplementary Figure S4**), validating the analytical solution to the model and our fitting procedure. In theory we also could fit the pdf for $\Delta y_i^\perp$. However, the pdf for $\Delta y_i^\perp$ is symmetric, because there is no preferred direction of migration and therefore less informative than the distribution for $\Delta x_i^\parallel$. We found that we could maintain the accuracy of our parameter estimation while improving the computational cost by only considering the $\Delta x_i^\parallel$ distribution. As a consistency check, we always verify that the estimated parameters accurately reproduce the pdfs for $\Delta y_i^\perp$ (**Supplementary Figure S6**).

State Prediction

To identify which state a cell is in at a given time, we used Bayes' theorem to invert the problem. That is, we calculate the probability that a cell is in state S_i given the experimental data. Note that in calculating this probability, we also get the false positive rate or p-value. To make a reliable prediction of

S_i may require an n-step window, where n is odd, such that, $\{x_{i-n/2}, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_{i+n/2}\}$. Then:

$$P(s_i|X) = \frac{P(X|s_i)P(s_i)}{P(X|s_i)P(s_i) + P(X|s_i^c)P(s_i^c)}$$

where $P(X|s_i)$ is calculated from the model, and we take $P(s_i) = \alpha$. Windows of length one, three and five were tested. For the case presented here, we found that the window of length one produced results similar to the other two window lengths.

Foci Identification

Ratiometric images of the FRET based Rac1 biosensor were analyzed for localized regions of higher Rac1 activity near the periphery of the cell. We call these regions "foci". We used custom application of the image processing toolbox in MATLAB to identify foci, which we define as contiguous regions within the cell that were simultaneously: (1) 60% above the average intensity of the cell, (2) greater than 100 pixels in area, and (3) contained at least one point within 5 pixels of the cell edge. The length of time (or number of frames) that a cell could be followed for varied. So, to not overweight any one cell, the number of image frames analyzed, n, was selected to maximize $n \times m$ where m is the number of cells with at least n images.

Experimental Methods
Cell Culture and Transient Transfections

HeLa cells were maintained in Dulbecco's modified Eagle's Medium (DMEM) (Cellgro) supplemented with 10% fetal bovine serum (FBS) (HyClone), 100 U/mL penicillin and 100 μg/mL streptomycin (Cellgro) and 2 mM L-glutamine (Invitrogen) at 37°C and 5% CO₂. All cDNA constructs were transfected into cells using FuGene6 (Roche) according to the manufacturer's instructions. IA32 Mouse Embryonic Fibroblast (MEF) cells were maintained in Dulbecco's modified Eagle's Medium (DMEM) (Cellgro) supplemented with 10% fetal bovine serum (FBS) (HyClone) and 1× GlutaMAX (Thermo Fisher Scientific).

IA32MEFs were transfected with either RhoG siRNA (CAGGTTTACCTAAGAGGCCAA) or Allstars Negative Control siRNA (Qiagen, United States). 7.5 μL, 10 μM siRNA was added to 250 μL serum-free DMEM. 3 μL lipofectamine RNAimax was added to another 250 μL serum-free DMEM. After 5 min, the two solutions were mixed and incubated for 20 min, followed by dropwise addition to a 35 mm dish. Medium was changed after 24 h and cells were split as required for use in experiments 48–72 h post-transfection, when knock-down efficiency was maximal. Control siRNA cells were incubated with 5 μM CFDA green for 20 min in serum-free DMEM. CFDA-labeled control cells were mixed with unlabeled RhoG siRNA cells immediately prior to the experiment.

Live Cell Imaging

For live cell imaging, cells were plated on fibronectin-coated coverslips (10 μg/ml fibronectin) 4 h before imaging, then transferred to Ham's F12-K imaging medium supplemented with 2% FBS and 15 mM HEPES. Live cell imaging was performed in a closed heated chamber (20/20 Bionomic).

For biosensor imaging, photobleach-corrected time-lapse image stacks were acquired for 18 h at 5 min intervals and processed as previously described (Pertz et al., 2006; Machacek et al., 2009; Hodgson et al., 2010). The following filter sets were used (Chroma Technology Corp.): CFP: D436/20, D470/40; FRET: D436/20, ET535/30; YFP: D500/20, ET535/30. Cells were illuminated with a 100 W Hg arc lamp through a 1.0 neutral density filter.

For RhoG siRNA experiments cell tracks were generated through 10× DIC imaging of cells plated as above, but using Ham's F12K medium supplemented with 5% FBS. Images were acquired for at least 70 frames at 10 min intervals in a closed, heated chamber. This length of track was objectively identified as optimal by maximizing the total number of analyzed frames in the entire data set.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

RA, TE, and KH devised the research plan and wrote the manuscript. RA performed the mathematical and computational calculations. CW and NP collected the data. All authors contributed to the article and approved the submitted version.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphys.2020.00822/full#supplementary-material

FIGURE S1 | Cumulative distribution functions (CDFs) for *x* and *y* step sizes compared to the CDFs for normal distributions.

FIGURE S2 | Validation of analytical approach. Simulated data (histogram) is generated by simulating data (blue bars) stochastically with parameters $(\mu_r, \sigma_r, \sigma_\theta, \sigma_0) = (0.1, 0.1, 0.5, \pi/5)$. Comparison with the analytical PDF (red) gives good agreement.

FIGURE S3 | Fitting the model to MEF WT cells indicates that a one state model of migration cannot adequately capture key features of the data. Observed data (open bars), model fit (green).

FIGURE S4 | Analytics and Fitting Validation. Simulated data (histogram) is generated by applying the model with parameters $(\mu_r^1, \sigma_r^1, \sigma_\theta^1) = (1, 0.3, 3)$, $(\mu_r^2, \sigma_r^2, \sigma_\theta^2) = (5, 1, 0.5)$ and $\alpha = 0.5$. Assuming the underlying parameters were unknown, we used simulated annealing to fit the analytical solution (green line). This validates our analytical solution and fitting

procedure (we estimate $(\hat{\mu}_r^1, \hat{\sigma}_r^1, \hat{\sigma}_\theta^1) = (1.02, 0.30, 2.72)$, $(\hat{\mu}_r^2, \hat{\sigma}_r^2, \hat{\sigma}_\theta^2) = (5.08, 0.93, 0.51)$, and $(\alpha = 0.50)$.

FIGURE S5 | Parameter values and associated confidence intervals. For each condition, and each parameter, the best-fit is the minimum of the blue curve. Acceptable parameter values are those which can lead to a value below the threshold (red dots) when the other parameters are re-fit, highlighted by the red-dash lined. Note, that two states were clearly identified in both cases.

FIGURE S6 | Model prediction versus the ΔY distribution from the cases shown in **Figure 2**. **(A)** MEF WT **(B)** MEF RhoG KD. In **(A)** and **(B)**: Observed data (open bars), model fit (green) and individual pdfs for state one and state two (insets).

FIGURE S7 | Testing the accuracy of predicting states using HeLa Rac1 model parameters. Data simulated stochastically, where we know the state a given step was generated from, is compared with our prediction of the state using Bayes Theorem. Overall, we are correct more than 90% of the time. However, steps truly in state one get occasionally mischaracterized are overlapping leading to an accuracy of around 70% in this case.

FIGURE S8 | Bootstrapped Analysis of the Difference in Foci Between States. To ascertain if the difference we observed in the mean foci count in state one versus state two was meaningful or could have been observed by chance, we performed a boostrap analysis. For a given iteration of the bootstrap, we randomly assigned state one or state two to each image frame with probability α and $1 - \alpha$, respectively. Then, for each iteration, calculated Δ mean foci as the difference in the mean number of foci in state one versus state two. Blue bars, bootstrapped distribution (50,000 iterations), dotted red line observed Δ mean foci.

FIGURE S9 | Comparison of 1-state (dashed curves) and 2-state (green curves) model fits for all of the experimental results.

FIGURE S10 | Overview of algorithm to estimate model parameters and confidence intervals.

FILES M1 | M1 (slow_state_hela_rac1.mp4) and M2 (fast_state_hela_rac1.mp4). The predicted "fast state" of migration was correlated with fewer Rac1 foci, most frequently a single focus at the leading edge, and in the slow state multiple brief Rac1 activation foci appear.

REFERENCES

Allen, R. J. (2020). *Cell Tracking Project*. Available at: https://github.com/rallen81/CellTracking (accessed January 17, 2018).

Allen, R. J., Welch, C., Pankow, N., Hahn, K., and Elston, T. C. (2018). Stochastic methods for inferring states of cell migration. *bioRxiv [Preprint]*. doi: 10.1101/249656

Arrieumerlou, C., and Meyer, T. (2005). A local coupling model and compass parameter for eukaryotic chemotaxis. *Dev. Cell* 8, 215–227. doi: 10.1016/j.devcel.2004.12.007

Cain, R. J., and Ridley, A. J. (2012). Phosphoinositide 3-kinases in cell migration. *Biol. Cell* 101, 13–29. doi: 10.1042/bc20080079

Dieterich, P., Klages, R., Preuss, R., and Schwab, A. (2008). Anomalous dynamics of cell migration. *Proc. Natl. Acad. Sci. U.S.A.* 105, 459–463. doi: 10.1073/pnas.0707603105

Dimilla, P. A., Quinn, J. A., Albelda, S. M., and Lauffenburger, D. A. (1992). Measurement of individual cell migration parameters for human tissue cells. *AIChE J.* 38, 1092–1104. doi: 10.1002/aic.690380712

Finney, A. C., Stokes, K. Y., Pattillo, C. B., and Orr, A. W. (2017). Integrin signaling in atherosclerosis. *Cell Mol. Life Sci.* 74, 2263–2282. doi: 10.1007/s00018-017-2490-4

Franca-Koh, J., Kamimura, Y., and Devreotes, P. N. (2007). Leading-edge research: PtdIns(3,4,5)P3 and directed migration. *Nat. Cell Biol.* 15–17. doi: 10.1038/ncb0107-15

Goley, E. D., and Welch, M. D. (2006). The ARP2/3 complex: an actin nucleator comes of age. *Nat. Rev. Mol. Cell Biol.* 7, 713–726. doi: 10.1038/nrm2026

Hall, A. (2009). The cytoskeleton and cancer. *Cancer Metastasis Rev.* 28, 5–14.

Hodgson, L., Shen, F., and Hahn, K. (2010). Biosensors for characterizing the dynamics of rho family GTPases in living cells. *Curr. Protoc. Cell Biol.* 14, Unit 14.11.1–26. doi: 10.1002/0471143030.cb1411s46

Iden, S., and Collard, J. G. (2008). Crosstalk between small GTPases and polarity proteins in cell polarization. *Nat. Rev. Mol. Cell Biol.* 9, 846–859. doi: 10.1038/nrm2521

Jaffe, A. B., and Hall, A. (2005). "Rho GTPases: biochemistry and biology. *Annu. Rev. Cell Dev. Biol.* 21, 247–269.

Katoh, H., Hiramoto, K., and Negishi, M. (2006). Activation of Rac1 by RhoG regulates cell migration. *J. Cell Sci.* 119, 56–65. doi: 10.1242/jcs.02720

Katoh, H., and Negishi, M. (2003). RhoG activates Rac1 by direct interaction with the Dock180-binding protein Elmo. *Nature* 424, 461–464. doi: 10.1038/nature01817

Kraynov, V. S., Chamberlain, C., Bokoch, G. M., Schwartz, M. A., Slabaugh, S., and Hahn, K. M. (2000). Localized Rac activation dynamics visualized in living cells. *Science* 290, 333–337. doi: 10.1126/science.290.5490.333

Ladwein, M., and Rottner, K. (2008). On the Rho'd: the regulation of membrane protrusions by Rho-GTPases. *FEBS Lett.* 582, 2066–2074. doi: 10.1016/j.febslet.2008.04.033

Lemarié, C. A., Tharaux, P. L., and Lehoux, S. (2010). Extracellular matrix alterations in hypertensive vascular remodeling. *J. Mol. Cell. Cardiol.* 48, 433–439. doi: 10.1016/j.yjmcc.2009.09.018

Machacek, M., Hodgson, L., Welch, C., Elliott, H., Pertz, O., Nalbant, P., et al. (2009). Coordination of Rho GTPase activities during cell protrusion. *Nature* 461, 99–103. doi: 10.1038/nature08242

Marston, D. J., Anderson, K. L., Swift, M. F., Rougie, M., Page, C., Hahn, K. M., et al. (2019). High Rac1 activity is functionally translated into cytosolic structures with unique nanoscale cytoskeletal architecture. *Proc. Natl. Acad. Sci. U. S. A.* 116, 1267–1272. doi: 10.1073/pnas.1808830116

Othmer, H. G., Dunbar, S. R., and Alt, W. (1988). Models of dispersal in biological systems. *J. Math. Biol.* 26, 263–298. doi: 10.1007/bf00277392

Pertz, O., Hodgson, L., Klemke, R. L., and Hahn, K. M. (2006). Spatiotemporal dynamics of RhoA activity in migrating cells. *Nature* 440, 1069–1072. doi: 10.1038/nature04665

Petrie, R. J., Doyle, A. D., and Yamada, K. M. (2009). Random versus directionally persistent cell migration. *Nat. Rev. Mol. Cell Biol.* 10, 538–549. doi: 10.1038/nrm2729

Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., et al. (2009). Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics* 25, 1923–1929. doi: 10.1093/bioinformatics/btp358

Ridley, A. J. (2006). Rho GTPases and actin dynamics in membrane protrusions and vesicle trafficking. *Trends Cell Biol.* 16, 522–529. doi: 10.1016/j.tcb.2006.08.006

Ridley, A. J. (2015). Rho GTPase signalling in cell migration. *Curr. Opin. Cell Biol.* 36, 103–112. doi: 10.1016/j.ceb.2015.08.005

Rosello, C., Ballet, P., Planus, E., and Tracqui, P. (2004). Model driven quantification of individual and collective cell migration. *Acta Biotheor.* 52, 343–363. doi: 10.1023/b:acbi.0000046602.58202.5e

Rottner, K., Hall, A., and Small, J. V. (1999). Interplay between rac and rho in the control of substrate dynamics. *Curr. Biol.* 9, 640–648.

Schaks, M., Giannone, G., and Rottner, K. (2019). Actin dynamics in cell migration. *Essays. Biochem.* 63, 483–495.

Svensson, C. M., Medyukhina, A., Belyaev, I., Al Zaben, N., and Figge, M. T. (2018). Untangling cell tracks: quantifying cell migration by time lapse image data analysis. *Cytometry* 93, 357–370. doi: 10.1002/cyto.a.23249

Wang, H., Vilela, M., Winkler, A., Tarnawski, M., Schlichting, I., Yumerefendi, H., et al. (2016). LOVTRAP: an optogenetic system for photoinduced protein dissociation. *Nat. Meth.* 13, 755–758. doi: 10.1038/nmeth.3926

Wu, Y. I., Frey, D., Lungu, O. I., Jaehrig, A., Schlichting, I., Kuhlman, B., et al. (2009). A genetically encoded photoactivatable Rac controls the motility of living cells. *Nature* 461, 104–108. doi: 10.1038/nature08241