



**HAL**  
open science

## Inferring geometrical dynamics of cell nucleus translocation

Sirine Amiri, Yirui Zhang, Andonis Gerardos, Cécile Sykes, Pierre Ronceray

► **To cite this version:**

Sirine Amiri, Yirui Zhang, Andonis Gerardos, Cécile Sykes, Pierre Ronceray. Inferring geometrical dynamics of cell nucleus translocation. *Physical Review Research*, 2024, 6 (4), pp.043030. 10.1103/physrevresearch.6.043030 . hal-04778056

**HAL Id: hal-04778056**

**<https://hal.science/hal-04778056v1>**

Submitted on 12 Nov 2024






**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

## Inferring geometrical dynamics of cell nucleus translocation

Sirine Amiri <sup>1,\*</sup> Yirui Zhang <sup>2,\*</sup> Andonis Gerardos <sup>2</sup> Cécile Sykes <sup>1,3,†</sup> and Pierre Ronceray <sup>2,‡</sup>

<sup>1</sup>Laboratoire de Physique de l'ENS, École Normale Supérieure, PSL University, CNRS, Sorbonne Université, Université de Paris, 75005 Paris, France

<sup>2</sup>Aix Marseille Université, CNRS, CINAM, Turing Center for Living Systems, 13009 Marseille, France

<sup>3</sup>Laboratoire Physico-Chimie Curie, Institut Curie, 75005 Paris, France



(Received 20 December 2023; accepted 9 July 2024; published 11 October 2024)

The ability of eukaryotic cells to squeeze through constrictions is limited by the stiffness of their large and rigid nucleus. However, migrating cells are often able to overcome this limitation and pass through constrictions much smaller than their nucleus, a mechanism that is not yet understood. Here, we propose a methodological framework to observe, quantify, and model this *nuclear translocation* phenomenon through a data-driven approach using microfluidic devices where cells migrate through controlled narrow spaces of sizes comparable to the ones encountered in physiological situations. Stochastic force inference is applied to experimental nuclear trajectories and nuclear shape descriptors, resulting in equations that effectively describe the kinematics of this *nuclear translocation* phenomenon. By employing a model where the channel geometry is an explicit parameter and by training it over experimental data with different sizes of constrictions, we ensure that the resulting equations are predictive. Altogether, the approach developed here paves the way for a mechanistic and quantitative description of dynamical cell complexity during its motility.

DOI: [10.1103/PhysRevResearch.6.043030](https://doi.org/10.1103/PhysRevResearch.6.043030)

### I. INTRODUCTION

The cell nucleus, three-to-four times stiffer than the cytoskeleton and twice as viscous, has traditionally been regarded as a mechanically passive compartment housing genetic information [1]. It is now established that in physiological conditions, the nucleus can experience large mechanical stresses that impact its shape and internal organization, affecting, for example, gene transcription [2]. In particular, when cells migrate through complex environments, the nucleus happens to experience large deformations, for instance, when passing through tight constrictions [3–6]. How these large deformations affect nucleus functioning and feed back into the behavior of the cell remain open questions. In fact, the overwhelming majority of cell migration studies focus on experiments on flat surfaces that were crucial to decipher the detailed mechanisms of cell motility [7–9]. However, the nucleus is only weakly altered in such experiments, which thus cannot be informative on the role of nuclear mechanics on cell motility, its passive mechanical resistance to deformation, and the mechanosensory pathways through which these deformations feed back into and actuate the cell behavior [2,10–13].

Addressing this problem through *in vivo* experimental observations of cell migration in tightly constraining environments such as the extracellular matrix and epithelial tissues represents a tremendous challenge. Indeed, one would have to disentangle the complexity of the environment from that of the migrating cell. For this reason, here we study an *in vitro* system of cells migrating in a microfabricated device that imposes three-dimensional mechanical constraints on spontaneously migrating eukaryotic cells [14,15]. We therefore focus on the influence of the geometry on squeezed cell migration. Specifically, cells migrate in an array of pillars designed to impose constrictions of controlled size, which incur large deformations of the nucleus. Remarkably, we find that cells with a nucleus of diameter of  $\sim 12\ \mu\text{m}$  in their rest state are able to spontaneously migrate through constrictions as tight as  $2\ \mu\text{m}$ . We refer to this process as *nuclear translocation*, in analogy with polymer translocation where a large macromolecule can pass through tight pores. Using bright-field and multichannel fluorescent imaging, we are able to track the trajectories of individual nuclei going through these constrictions. However, the analysis of the resulting trajectories poses multiple challenges due to their complexity, inherent stochasticity, and the limited amount of data: how does one extract quantitative models and mechanistic insights from such trajectories?

To tackle this challenge, here we develop and apply a data-driven approach to learn dynamical models directly from experimental nucleus trajectories. This contrasts with more traditional model-based approaches that postulate a model form and fit its parameters through the use of aggregate observables such as correlation functions: here we let the model emerge from the data, and the parameters are

\*These authors equally contributed to this work.

†Contact author: [cecile.sykes@phys.ens.fr](mailto:cecile.sykes@phys.ens.fr)

‡Contact author: [pierre.ronceray@univ-amu.fr](mailto:pierre.ronceray@univ-amu.fr)

optimized directly on the entire data set. Such approaches have recently received much attention, in particular due to the development of methods adapted to data-driven inference of deterministic dynamical models such as ordinary and partial differential equations [16,17]. These methods are well adapted for large-scale datasets such as tissue dynamics [18,19]. Importantly, however, single-cell dynamics are not deterministic: the inner complexity of these objects, coupled to the reliance to feedback pathways involving small numbers of signaling molecules, results in apparently erratic dynamics which is better captured by stochastic differential equations (SDEs) [20]. Data-driven approaches have been used to capture the dynamics of freely migrating cells [21–23], revealing a persistent random walk behavior. They have been used to quantify the dynamics of nonconstraining confined cell migration [24,25] and, recently, for constraining cell migration in an elastic environment [26]. Newly introduced inference methods for SDEs [27,28] have made it possible to efficiently learn such dynamics and have resulted in insights in cell-cell interactions during confined migration that would not have been possible with preexisting methods [24]. However, to our knowledge, such methods have not been applied to cell migration with mechanical constraints that lead to large deformations of the nucleus. To this aim, we define and measure quantitative descriptors of the cell shape and state, then use *stochastic force inference* (SFI) [27] to construct a model that captures the dynamics of these shape descriptors. By including the constriction shape as an explicit input of the model, we are able to extrapolate the model to other constriction sizes. Our inference analysis explicitly takes into account the spatial constraints of nucleus translocation and is applicable to other experimental designs.

## II. RESULTS

### A. Confined cell migration experiments

We use a CRISPR-modified mouse embryonic fibroblasts (MEFs) cell line that expresses nesprin-2 giant with a green fluorescent protein (GFP) sequence and lamin A/C with a red fluorescent protein (mCherry) sequence [29]. The lamin biopolymer shell that lies right underneath the nuclear envelope is linked to the cytoskeleton through the LINC complex, which includes nesprins [30–34]. Cells migrate through microfluidic devices that consist of a series of 5  $\mu\text{m}$  high pillar structures providing three sizes of constrictions (5, 3, and 2  $\mu\text{m}$ ) and larger channels (15  $\mu\text{m}$ ) [Figs. 1(a) and 1(b)]. Such migration devices are obtained by covalent assembly of a three-dimensional (3D)-imprinted block of polydimethylsiloxane (PDMS) with a glass coverslip [15]. Cells are placed on one side of the device with culture medium, before the pillars (see Appendices A 1–A 4). They exhibit global motion (on the  $x$  axis) towards the other side, empty of cells but filled with culture medium [Fig. 1(a)], a phenomenon we interpret as chemotaxis along a self-generated nutrient gradient [35]. The apparent width of MEF cell nuclei (on the  $y$  axis) is  $12 \pm 2 \mu\text{m}$  outside of constrictions (see Appendix B 1). It is, therefore, larger than constriction sizes and smaller than the large channel of 15  $\mu\text{m}$ . Note that in all conditions, nuclei shapes are mostly cylindrical (on the  $z$  axis), touching the

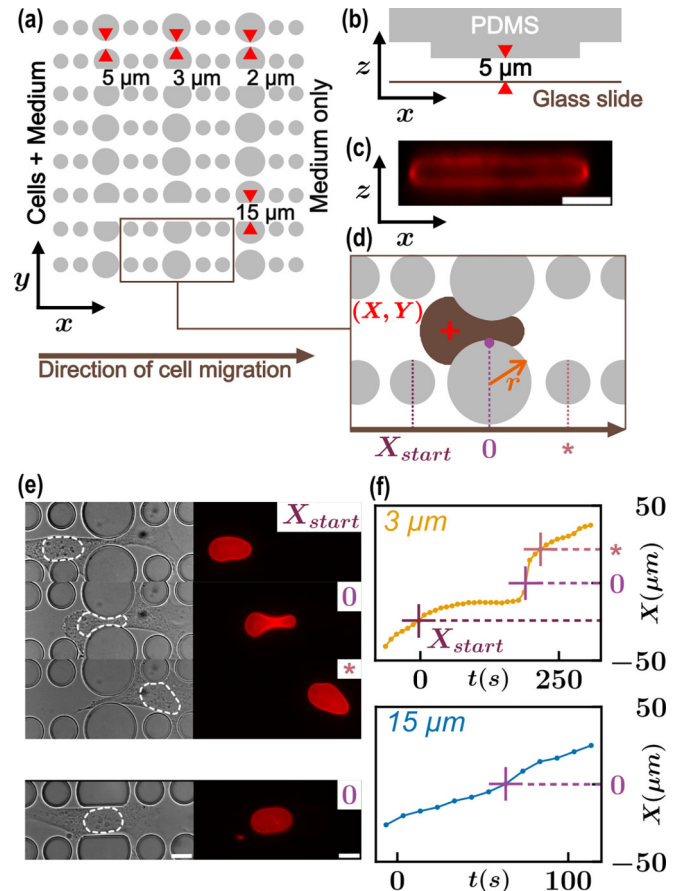


FIG. 1. CRISPR engineered MEFs are migrating in a microfluidic device made of constrictions. (a) Top view of the pattern used in the microfluidic device. It is composed of PDMS pillars of several widths in order to make three types of constrictions: 5, 3, and 2  $\mu\text{m}$  wide and a control channel of 15  $\mu\text{m}$ . (b) Side view of the microfluidic device. Height of the pillars is 5  $\mu\text{m}$ . (c) Side view ( $x, z$ ) of lamin A/C signal in mCherry of an engineered MEF in the middle of a constriction. (d) Representation of the origin points used for a nucleus (in brown) trajectory.  $(X, Y)$  is position of the nucleus (red star). (e) Epifluorescence images of an engineered MEF crossing a 3  $\mu\text{m}$  constriction (top) and a 15  $\mu\text{m}$  large channel (bottom). Left is the transmission signal and right is the mCherry signal (for Lamin A/C). (f) Examples of trajectories of one cell through a 3  $\mu\text{m}$  constriction (top) and one cell through a 15  $\mu\text{m}$  large channel (bottom). Annotated points correspond to specific positions illustrated in (d). Scale bars: 10  $\mu\text{m}$ .

bottom and the ceiling of the migration device [Fig. 1(c)]. We confirm previous observations [29] that during nucleus translocation through a constriction, the nesprin signal intensity increases at the front of the nucleus, while the lamin signal does not (see Appendix B 2).

### B. Extracting cell nucleus trajectories

We observe the movement of cell nuclei in the horizontal ( $x, y$ ) plane when cells migrate between vertical  $z$ -oriented pillars [Figs. 1(a) and 1(b)]. Nuclei are deformed when they translocate through narrow constrictions [14]. A constriction is defined by two facing big pillars of radius  $r$  [Fig. 1(d)]. For

each constriction, we define the spatial origin ( $x = 0, y = 0$ ) at the center of the constriction. The position of a nucleus is defined by its surface barycenter ( $X, Y$ ) detected through the lamin signal [Fig. 1(d), right, and Appendices A 5 and A 6]. An image is taken every 10 min. This time interval was optimized to limit fluorescence bleaching and cell phototoxicity.

For each nucleus trajectory, we define  $X_{\text{start}}$  as the middle of the small pillar that precedes the specific constriction [Fig. 1(d)]. The same definition is adapted to the 15  $\mu\text{m}$  channels with a truncated-disk pillar defining the ‘‘constriction.’’ The start of a trajectory (at time  $t = 0$  min) is defined either by  $X = X_{\text{start}}$  or by its interpolated value using a constant nuclear speed between the two available positions closest to  $X_{\text{start}}$ . The end of a trajectory is determined by the earliest of (i) the end of the overall acquisition, (ii) the start of a new trajectory in a new constriction, and (iii) half an hour before the cell starts to divide or die. Here, we exclude any trajectory corresponding to cells undergoing adherent cell-cell contact for more than an hour to exclusively address the migration of individual cells. Examples of recorded images of a nucleus translocating through a 3  $\mu\text{m}$  constriction and a nucleus migrating through a large 15  $\mu\text{m}$  channel are given in Fig. 1(e). The corresponding trajectories and origin points are displayed in Fig. 1(f). We do not observe nuclear rupture during this deformation, contrary to other mechanical studies of cell nuclei [13,36,37].

A typical nucleus trajectory  $X(t)$  through a 3  $\mu\text{m}$  constriction has a sigmoidlike shape, with a plateau soon after  $X_{\text{start}}$  when the nucleus reaches the entrance of the constriction, and a sharp speedup when it manages to pass through the center of the constriction at  $X = 0$ , followed by an unconstrained motion [Fig. 1(f), top]. A nucleus trajectory in a large 15  $\mu\text{m}$  channel displays a smooth movement [Fig. 1(f), bottom] at almost constant velocity. However, a fraction of cells does not translocate before the end of the trajectory recording. They are, nevertheless, included in our data set to avoid any statistical bias in the analysis. Overall, nucleus trajectories show some variability, both in the duration of the plateau and in the velocity of free migration, as can be seen in Fig. 2(a).

### C. Modeling the stochastic nature of nucleus translocation

The observed variability between different trajectories in the same constriction condition (either the 3  $\mu\text{m}$  or the 15  $\mu\text{m}$  wide constriction) and the fluctuations of velocity during a single trajectory reflect the internal complexity of the cells. This effective stochasticity is often modeled through a noise term by using stochastic differential equations (SDEs). A SDE describing the cell nucleus motility is typically of the form [38]

$$\dot{X}(t) = \underbrace{\Pi + f_{\text{ext}}(X)}_{\text{deterministic}} + \underbrace{\sqrt{2D_X} \cdot \eta_X(t)}_{\text{stochastic}}, \quad (1)$$

which consists of two deterministic terms ( $\Pi$  and  $f_{\text{ext}}$ ) reflecting the slow, predictable aspects of the dynamics and a stochastic noise that models the coupling of the observed position with fast, unobserved degrees of freedom. More specifically, in the deterministic contribution of nucleus dynamics,  $\Pi$  is a driving term, also called *polarity* of the cell, and captures the asymmetry in the internal organization of the cell that generates the motility [39]. The other deterministic

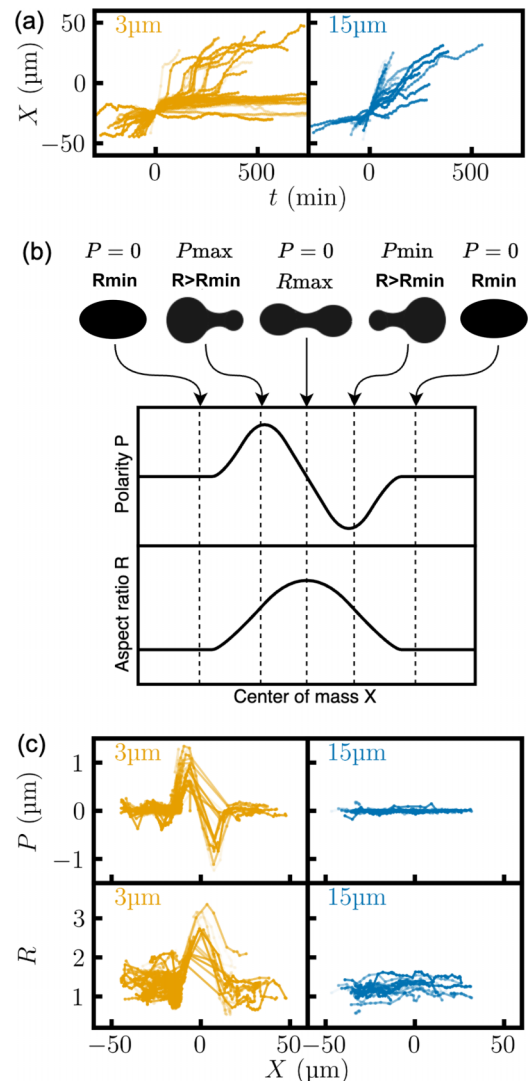


FIG. 2. Experimental trajectories of cell nucleus translocation. (a) Time series of nucleus position  $X$  during translocation. (b) Schematic of characteristic nuclei shapes and their corresponding position during the translocation. (c) Protusion vector  $P$  and aspect ratio  $R$  change along the translocation process. In (a) and (c), results from 3  $\mu\text{m}$  and 15  $\mu\text{m}$  constraints are shown for comparison.

term,  $f_{\text{ext}}$ , represents the direct effect of the environment on the cell nucleus dynamics. The noise term  $\sqrt{2D_X} \cdot \eta_X$  has an amplitude characterized by its diffusion coefficient  $D_X$ , which we assume here to be state independent, and  $\eta_X$  the noise, which, for simplicity, we assume to be white and Gaussian, and therefore  $\langle \eta_X(t) \rangle = 0$  and  $\langle \eta_X(t) \eta_X(t') \rangle = \delta(t - t')$ .

The polarity  $\Pi$  itself is dynamical and its dynamics describes the way cells sense their environment and actuate their self-propulsion accordingly. The dynamics of  $\Pi$  follows an SDE of the form

$$\dot{\Pi}(X, t) = f_{\Pi}(X, \Pi) + \sqrt{2D_{\Pi}} \cdot \eta_{\Pi}(t). \quad (2)$$

The drift term  $f_{\Pi}(X, \Pi)$  encodes the internal dynamics of  $\Pi$  as well as the feedback of the nucleus polarity to the external environment. Note that there are thus two ways in which the environment affects the dynamics: through direct



forces on  $X$  [term  $f_{\text{ext}}(X)$  in Eq. (1)] and through indirect feedback [ $f_{\Pi}(X, \Pi)$  in Eq. (2)], i.e., mechanosensing. Here again, fast internal processes of the cell are modeled as a Gaussian white noise  $\sqrt{2D_{\Pi}} \cdot \eta_{\Pi}(t)$  with diffusion coefficient  $D_{\Pi}$ , which determines, for instance, the persistence length of the free motion of the cell [23].

The class of cell motility models described by Eqs. (1) and (2) is very general and widely used. However, a key challenge to its applicability to experimental data is that the polarity  $\Pi$  is not directly measurable, as its molecular definition remains unknown. To bypass this difficulty, previous studies have relied on the use of underdamped dynamics: briefly, such approaches consist of differentiating Eq. (1) with respect to time, and plugging into Eq. (2) to eliminate  $\Pi$ , thus resulting in an effectively second-order dynamics for  $X$  [38]. This type of *embedding* approach exploits Taken's theorem and is popularly used for deterministic dynamical systems [40]. While this approach has been successful in quantifying, for instance, cell-cell interactions from data [24], it has several drawbacks. First, second-order inference is considerably more difficult and demanding in terms of data quality and precision than first-order inference [28]. Second, one has to neglect the noise on nucleus position in order for this approach to work, which is not always possible. Third, information about the nature of the polarity and its feedback mechanisms is lost in the process. An alternative approach was proposed recently, consisting of a model-driven treatment of data where the polarity is explicitly included as a hidden variable [25], but this requires strong assumptions on the motility mechanisms.

#### D. Data-driven modeling from geometric quantities

Here we propose an alternative approach to maintain the overdamped dynamics, which is more physical, and approximate  $\Pi$  with available information. Indeed, we have access to more than just the nuclear center  $X$ : using the lamin signal, we can track the precise contour of the nucleus and extract a richer set of geometrical quantities. In particular, when the cell engages into the constriction, the nucleus starts elongating and protruding toward the narrow part of the constriction, as schematized in Fig. 2(b). When exiting the constriction, the protrusion points backwards and the nucleus progressively recovers its oval shape.

From these observations, we define two variables to characterize nucleus deformations. First, to account for the geometrical shape change of the nucleus, we define its *protrusion vector*  $P = X_c - X$ , with  $X_c$  the barycenter of the contour of the nucleus (see the detailed expression in Appendix A 7). We choose to employ this variable  $P$  as it is directly measurable from the nuclear contour and gives a measure of how much and in which direction the nucleus boundary protrudes relative to the center of mass. A positive (negative) value of  $P$  corresponds to a forward (backward) extension of the nucleus relative to the center of mass [see Fig. 2(b)]. Second, we characterize the relative  $(x, y)$  stretch by defining the aspect ratio of the cell nucleus  $R$  (see the detailed expression in Appendix A 7). A perfectly circular disk would correspond to  $R = 1$ , whereas an ellipsoid oriented towards the  $x$  axis ( $y$  axis) would correspond to  $R > 1$  ( $R < 1$ ). Note that the microfluidic device slightly compresses the nucleus even in

the wider parts of the channel, as evident in Fig. 1(e): as a consequence, the nucleus elongates in the  $x$  direction and  $R$  is larger than 1 even out of the constricting region. When the nucleus starts to squeeze into the constriction,  $R$  increases to a maximum value reached right in the middle of the constriction [see Fig. 2(b)].

The quantities  $P$  and  $R$  describe two different and complementary aspects of nucleus deformation. As illustrated in Fig. 2(b),  $P$  does not distinguish a dumbbell from an oval shape, whereas  $R$  does;  $R$  cannot distinguish a front protrusion from a back protrusion, whereas  $P$  does. Note that our model, being overdamped, does not explicitly include the direction of movement. The breakdown of symmetry along the  $x$  axis could have been included in two ways: by including a shape descriptor internal to the cell that directly captures this orientation (e.g., by tracking the lamellipodium too) or by including terms in the dynamical model that break this symmetry externally (e.g., modeling a nutrient gradient in a chemotaxis model). We choose the latter here, motivated both by our aim to present a self-consistent dynamical model of the nucleus—the shape of which does not reflect its direction of motion—without referring to the extended cytoplasmic structures, and by the fact that chemotaxis is indeed suggested by cells consistently directing themselves from the cell-rich to the medium-rich side of the microfluidic device. Examples of the time series  $P$  and  $R$  against  $X$  when cells go through a 3  $\mu\text{m}$  constriction (left) or a 15  $\mu\text{m}$  channel (right) are displayed Fig. 2(c). Whereas  $P$  and  $R$  are constant in a 15  $\mu\text{m}$  large channel, they are significantly affected by the 3  $\mu\text{m}$  constriction. These complementary geometrical data will allow us to infer a kinematic model for nuclear translocation dynamics.

#### E. Inferring coupled dynamics of position and geometry

Our aim is to obtain a data-driven, quantitative, autonomous description of nuclear translocation using the position  $X$  and geometric descriptors  $P$  and  $R$ . More precisely, for each recorded nucleus trajectory in constraints 2, 3, 5, and 15  $\mu\text{m}$ , we extract three time series  $\{X_t, P_t, R_t\}$  at acquisition times  $t = 0, \Delta t, 2\Delta t, \dots$ . These data serve as the input in our inference analysis, from which we aim to extract coupled SDEs capturing the continuous-time dynamics of  $(X_t, P_t, R_t)$ . We postulate that including the geometric quantities  $P_t$  and  $R_t$ , on top of the nucleus position  $X_t$ , makes it possible to identify a set of such equations that is both autonomous (i.e., that does not couple to the dynamics of other, unobserved quantities, in contrast to the approach of Ref. [25]) and physically first order (i.e., that does not introduce emergent inertia as a polarity model, in contrast to most of the preexisting literature [23,24,38,41]).

To achieve this, we analyze the time series using a recently introduced framework, i.e., stochastic force inference (SFI) [27]. SFI allows us to reconstruct first-order SDEs from such time series by employing a data-efficient quasi-maximum-likelihood linear regression algorithm. In practice, it consists of approximating the drift term with estimators formed by a linear combination of basis functions. Here, we start from a relatively large basis that we construct based on symmetries and our physical understanding of the quantities we model, and that includes a systematic expansion of the geometrical

features of the system, i.e., the  $x$ -dependent constriction width. We then iteratively reduce this basis to select an appropriately minimal model for the dynamical equations that we aim to learn. More specifically, our starting model is

$$\dot{X}_t = \overbrace{C_X + \alpha_X P_t + \beta_X (R_t - R_t^{-1})}^{\text{internal driving}} + f_X(X_t, r) + \sqrt{2D_X} \cdot \eta_X(t), \quad (3)$$

$$\dot{P}_t = \overbrace{C_P + \alpha_P P_t + \beta_P (R_t - R_t^{-1})}^{\text{internal dynamics}} + f_P(X_t, r) + \sqrt{2D_P} \cdot \eta_P(t), \quad (4)$$

$$\dot{R}_t = \overbrace{C_R + \alpha_R P_t + \beta_R (R_t - R_t^{-1})}^{\text{internal dynamics}} + f_R(X_t, r) + \sqrt{2D_R} \cdot \eta_R(t). \quad (5)$$

Equation (3) connects to the general form presented in Eq. (1) by approximating the polarity  $\Pi$  with a linear combination of three terms:  $C_X$ , a constant drift representing the propensity of cells to migrate in the  $x$  direction, physically motivated by the imbalance in cell populations between the two sides of the device;  $\alpha_X P$ , which is a vectorlike term coupling the direction of motion and protrusion direction; and  $\beta_X (R - R^{-1})$  by which the nucleus shape modulates the self-propulsion velocity around its rest shape  $R = 1$ . The remainder,  $f_X$ , captures the effect of the environment and thus depends on the position  $X_t$ —we omit, for simplicity, the possibility that it depends on the geometry. Similarly, the dynamics of  $P$  is described by Eq. (4) [R by Eq. (5)] with the same decomposition into internal dynamics and external influence, and we use the same basis functions. Note that we use the combination  $(R - R^{-1})$  to reflect the fact that the aspect ratio  $R$  is a ratio of lengths which should remain positive at all times, and has average value of 1 in the absence of external constraints.

In a complex or unknown environment, the drifts  $f_X$ ,  $f_P$ , and  $f_R$  representing the influence of the environment on  $X$ ,  $P$ , and  $R$  would have to be expanded on a generic basis. Here, however, we take advantage of the fact that the geometry of the channel is known to simplify inference and allow for extrapolation of the model to other constriction sizes. Specifically, we symbolically include the shape of the constriction in the inference: our basis functions use the channel width  $w(X)$  as well as the vector coordinates  $\hat{n}(X) = [n_x(X), n_y(X)]$  of the normal to the constriction wall that contain information about its texture (see Appendix A 8 for a schematic of these quantities and their expressions). As we focus here on cylindrical pillars, we parametrize these functions by the radius  $r$  of the pillars that form the constriction [see Fig. 1(d)]; however, this approach could be directly generalized to other shapes described by the function  $w(X)$ . Using  $w(X, r)$ ,  $n_x(X, r)$ , and  $n_y(X, r)$  as ingredients, we approximate the environmental drift the nucleus experiences and reacts to, i.e.,  $f_X(X, r)$ ,  $f_P(X, r)$ , and  $f_R(X, r)$  in the constraint formed by pillars of radius  $r$ . Integrating the pillar radius as a control parameter into these functions allows us to infer a single model for the whole experimental data set of different constriction sizes, including the reference case where the channel does not have a constriction. It makes the model more

straightforward and easier to interpret and allows us to use the data more efficiently. As the influence of the pillars on the cell nucleus is expected to increase with decreasing channel width, we expand this geometrical influence in an inverse power series of the channel width in the basis, up to third order, i.e.,  $1/w$ ,  $1/w^2$ , and  $1/w^3$ , which we multiply by the geometrical quantities 1,  $n_x$ , and  $n_y$  that capture distinct features of the constriction. The full expression of our initial model is summarized in Appendix A 9. To summarize, the rationale for our proposed model consists in combining simple, first-order coupling between our geometric descriptors  $X$ ,  $P$ , and  $R$  with a systematic symbolic expansion of the pillar geometry.

The SFI algorithm provides estimators for the coefficients of the drift field as a linear combination of these basis functions. The initial model, which consists of the complete set of basis functions (in total, 36), is shown in Appendix A 9. A challenge to the use of stochastic inference techniques on cell migration data is that the time interval between frames  $\Delta t$  is typically of the same order as the typical translocation time and cannot be easily decreased as more frequent imaging would incur phototoxicity. To overcome this problem, we introduce an improvement on the SFI algorithm to accommodate large time steps, which uses a trapezoidal integration scheme that results in lower discretization biases than previous methods (see Appendix A 10 for details).

## F. Model selection algorithm

The learned model consisting of the full set of basis functions is constructed through physically motivated systematic expansion, and as such it is not minimal, which potentially leads to overfitting the data and precludes physical interpretation. To overcome this difficulty and obtain a more interpretable model, we improve this model through a sparsity-enforcing algorithm that consists in iteratively deleting the least statistically significant terms until a threshold significance is reached (throughout, “statistically significant” is taken to mean “with coefficient that is statistically distinguishable from zero”: our null hypothesis is that the term can be omitted from the model). This inference workflow, as schematized in Fig. 3, differs from popular sparse learning algorithms that include a penalization based on the values of the coefficients [16,42,43], which would not be appropriate here due to the fact that coefficients have distinct physical dimensions.

More specifically, this workflow consists of three iterative steps: *infer*, *bootstrap*, and *update*. The first step, *infer*, uses the SFI algorithm to learn coefficients using the current set of basis functions. In the second step, *bootstrap*, we assess the statistical significance of each inferred coefficient using the bootstrap method, running the inference again on sets of trajectories sampled with replacement and using the standard deviation of the coefficients as a confidence interval (see Appendix A 11 for details [44]). The significance of each basis function for our model is quantified by their signal-to-noise ratio, i.e., the ratio between the absolute value of the mean of the coefficient and its standard deviation. If one of these ratios is below a chosen significance threshold of 3 (corresponding to a  $3\sigma$  rule), we move to the third step, *update*, where we simplify the model by removing the least significant function

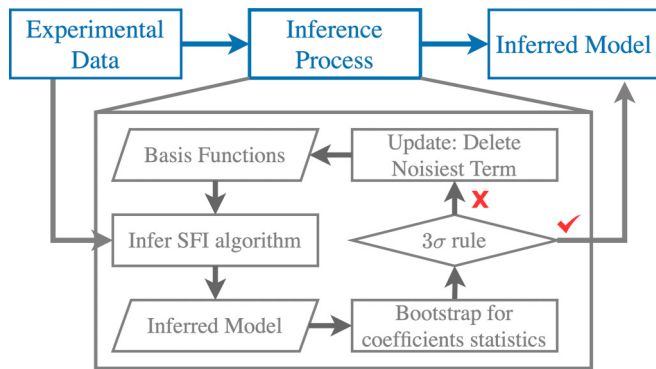


FIG. 3. Schematic of the inference workflow. To start, write an initial model—an overdamped Langevin equation of the problem at hand. Propose the basis functions that form the drift part of the equation and fix any known coefficients. Step 1. Infer: Input the experimental data and the candidate model into the SFI algorithm to obtain the most probable value of the unknown coefficients. Step 2. Evaluate: Use bootstrap to obtain the mean and the standard deviation for each coefficient. We evaluate the significance of each coefficient against the  $3\sigma$  rule. If one or more coefficients fails this test, then go to Step 3. Update: Update the model by removing the noisiest term. Repeat this process until a final model is reached, where any further elimination would deteriorate the model.

from the basis, and iterate the process. The outcome of this process is a final, minimal model where all terms are statistically significant.

### G. Resulting model for nuclear translocation dynamics

Applying this inference workflow to the whole nuclear translocation data set, we obtain the following model:

$$\dot{X}_t = C_X + \alpha_X P_t + \beta_X (R_t - R_t^{-1}) + \frac{a_X}{w^2(X_t, r)} + b_X \frac{n_y(X_t, r)}{w^2(X_t, r)} + \sqrt{2D_X} \cdot \eta_X(t), \quad (6)$$

$$\dot{P}_t = \alpha_P P_t + \beta_P (R_t - R_t^{-1}) + a_P \frac{1}{w(X_t, r)} + b_P \frac{1}{w^2(X_t, r)} + c_P \frac{n_x(X_t, r)}{w^2(X_t, r)} + \sqrt{2D_P} \cdot \eta_P(t), \quad (7)$$

$$\dot{R}_t = C_R + \alpha_R P_t + \beta_R (R_t - R_t^{-1}) + a_R \frac{1}{w(X_t, r)} + \sqrt{2D_R} \cdot \eta_R(t), \quad (8)$$

with a total of 14 drift terms. The values and standard deviations of the corresponding coefficients, as well as the inferred diffusion constants, are shown in Table I.

A representative selection of trajectories  $X(t)$  from the experiment and simulation is given in Fig. 4(a). Averaged trajectories  $(P, X)$  and  $(R, X)$  are given for experiments and simulations in Fig. 4(b) (average and standard deviation over  $N = 1000$  trajectories simulated with the learned coefficients from Table I). Position-binned curves  $P(X)$  and  $R(X)$  at different constraints can be differentiated in both simulated and experimental data: the deformation and protrusion increase significantly as the constraint becomes smaller. The starting and ending points of the experimental and simulated curves

TABLE I. Inferred coefficients for the minimal model, with corresponding terms in Eqs. (6)–(8). The confidence intervals correspond to the standard deviation obtained through bootstrapping. The last column indicates whether or not (Yes/No) this term breaks the left-right symmetry of the system.

Corr. term	Coeffs.	Value	Unit	Sym.
$\cdot$	$C_X$	$(2.6 \pm 0.5) \times 10^{-1}$	$\mu\text{m min}^{-1}$	Y
$P$	$\alpha_X$	$(-1.7 \pm 0.2) \times 10^{-1}$	$\text{min}^{-1}$	N
$R - 1/R$	$\beta_X$	$(6.7 \pm 0.7) \times 10^{-2}$	$\mu\text{m min}^{-1}$	Y
$1/w^2$	$a_X$	$(-4.7 \pm 1.1) \times 10^1$	$\mu\text{m}^3 \text{min}^{-1}$	Y
$n_y/w^2$	$b_X$	$(4.8 \pm 1.1) \times 10^1$	$\mu\text{m}^3 \text{min}^{-1}$	Y
$\eta_X$	$D_X$	$(5.8 \pm 0.8) \times 10^{-2}$	$\mu\text{m}^2 \text{min}^{-1}$	N
$P$	$\alpha_P$	$(-3.2 \pm 0.6) \times 10^{-2}$	$\text{min}^{-1}$	N
$R - 1/R$	$\beta_P$	$(3.0 \pm 0.7) \times 10^{-3}$	$\mu\text{m min}^{-1}$	Y
$1/w$	$a_P$	$(7.8 \pm 1.3) \times 10^{-2}$	$\mu\text{m}^2 \text{min}^{-1}$	Y
$1/w^2$	$b_P$	$-1.0 \pm 0.2$	$\mu\text{m}^3 \text{min}^{-1}$	Y
$n_x/w^2$	$c_P$	$-3.1 \pm 0.5$	$\mu\text{m}^3 \text{min}^{-1}$	N
$\eta_P$	$D_P$	$(8.0 \pm 1.4) \times 10^{-4}$	$\mu\text{m}^2 \text{min}^{-1}$	N
$\cdot$	$C_R$	$(-5.7 \pm 1.2) \times 10^{-3}$	$\text{min}^{-1}$	N
$P$	$\alpha_R$	$(2.2 \pm 0.3) \times 10^{-2}$	$\mu\text{m}^{-1} \text{min}^{-1}$	Y
$R - 1/R$	$\beta_R$	$(-8.7 \pm 1.4) \times 10^{-3}$	$\text{min}^{-1}$	N
$1/w$	$a_R$	$(1.1 \pm 0.2) \times 10^{-1}$	$\mu\text{m min}^{-1}$	N
$\eta_R$	$D_R$	$(8.6 \pm 1.8) \times 10^{-4}$	$\text{min}^{-1}$	N

$P(X)$  and  $R(X)$  coincide. Additionally, a qualitative agreement can be seen in their dynamics: for instance, the simulated  $P(X)$  curves recapitulate the asymmetry of the response before and after the constriction.

We now physically interpret the nature of the terms and value of the coefficients in the final model. The fact that  $C_X > 0$  indicates an average propensity of cells to migrate towards the nutrient-rich region. Interestingly, we find that  $\alpha_X < 0$  and  $\beta_X > 0$ : when entering the constricted region, the cell slows down as the nucleus first protrudes, then moves faster as it elongates. The fact that  $\alpha_P < 0$  confirms that in the absence of external forces, these quantities relax back to the equilibrium shape  $P = 0$ . Similarly,  $\beta_R < 0$ , inducing the relaxation of the aspect ratio outside the constriction; however, since the model also contains a constant term  $C_R < 0$ , the equilibrium value is  $R_{\min} > 1$ : the nucleus remains slightly elongated outside the constriction [see Figs. 1(e) and 2(b)]. Regarding the  $x$ -dependent external forces,  $f_X$  exhibits a repulsive term  $a_X/w^2$  that slows the cell near the entrance of the constriction, and an attractive term  $b_X n_y/w^2$  that speeds it up near  $x = 0$ , i.e., once it is engaged in the constriction. The protrusion force  $f_P$  exhibits a term  $c_P n_x/w^2$  that is odd under reflection symmetry and drives the rapid change of sign of the protrusion  $P$  as the nucleus crosses the tightest point of the constriction. Finally, the dynamics of  $R$  is captured by a single, elongation-driving term  $a_R/w$  with  $a_R > 0$ , consistent with the intuition that the nucleus elongates in tight constrictions. Combined with the fact that  $\beta_X > 0$ , this implies that cells tend to migrate faster when they are squeezed, reminiscent of previous works showing this effect for eukaryotic cells in constrictions, either mechanically constraining [13] or not [41]. We note that the breakdown of the left-right symmetry, self-generated by cells that migrate from the cell-rich to medium-rich side, results in

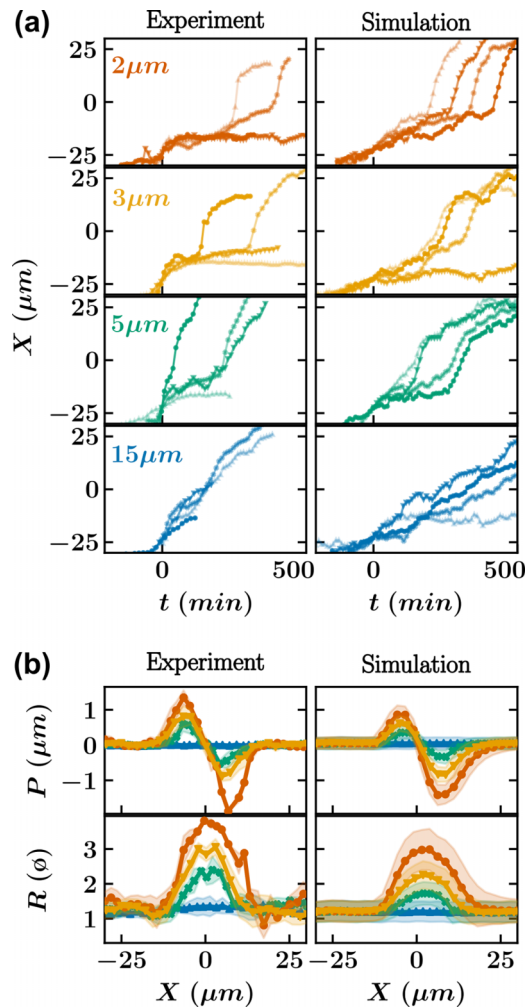


FIG. 4. Simulation results of the reduced model. (a) Comparison of four representative time series of cell nuclei position  $X(t)$  from the experimental data and simulation. (b) Experiment and simulation comparison of the averaged trajectories of nuclear protrusion against nucleus position  $P(X)$  and nucleus aspect ratio against position  $R(X)$ .

nonzero values for several coefficients whose corresponding terms explicitly break this symmetry, as indicated in Table I. Finally, the relaxation back to the equilibrium value at the constriction exit is sped up by the coupling  $\alpha_R P$  with negative  $P$  values. All in all, this model thus recapitulates, with a few terms, the directed migration of the cells through the channel and the way the nuclei stall when reaching the constriction entrance, then protrude, elongate, and, finally, pop through rapidly. In the final stage, the protrusion reverts and points backward, leading to a rapid relaxation of the aspect ratio and the exit from the constriction.

Our inference method also provides us with a physically interpretable estimate of the diffusion coefficients of the nucleus position  $D_X \sim 5.8 \times 10^{-2} \mu\text{m}^2 \text{min}^{-1}$ . This value is several orders of magnitude above the equilibrium expectations from the Stokes-Einstein equation for a purely passive particle in the highly viscous cellular environment,  $D_{\text{Stokes-Einstein}} \sim 8 \times 10^{-7} \mu\text{m}^2 \text{min}^{-1}$ , reflecting the fact that cellular motion is activity driven. Note that for simplicity of the analysis, we have assumed constant diffusion coefficients and

Gaussian white noise. To further investigate these assumptions, larger amounts of data with a higher time resolution would be needed.

Finally, while the learned model provides good agreement in terms of capturing the dynamical geometric change of the nucleus during translocation, with a single parametric model encompassing the available multiple constriction widths, we note that it also presents some limitations. Indeed, this model is trained on a population of cells and neglects any cell-to-cell variability due, e.g., to different sizes, genetic expression levels, and age of the cells. This inherent variability manifests itself in a different way from the dynamical stochasticity captured here by the diffusion terms. Taking into account such cell-to-cell variability is a major challenge, as the amount of data available for each cell is small: data-efficient methods such as SFI [27] or underdamped Langevin inference [28] provide a promising avenue towards this, but single-event processes such as nuclear translocation, studied here, remain intractable with these approaches. A further difficulty comes from the limited frame rate, which leads to trajectories that appear to “tunnel through” right at the end of the passage through the constriction [as evidenced by long straight lines connecting data points in Fig. 2(c)] and lower the resolution of the translocation event. We speculate that this effect leads to an underestimate of  $R$  and  $P$  near their peaks in the simulated model, compared to the experiment. Indeed, as these correspond to times when the cell goes fast through the constriction, the peak data points are sometimes missed, which tends to smooth out the inferred model. These challenges preclude the quantitative prediction of, e.g., mean translocation times, using the learned model. Despite these limitations, our model captures well the qualitative trends of the experimental data and is in semiquantitative agreement with each data set.

## H. Predictivity of the model

As our learned model takes the constriction geometry as an explicit parameter, we can extrapolate it to predict nuclear translocation dynamics in other constriction sizes defined by the radius of the pillars ( $r$ ). To assess the validity of this approach, we first test it on geometries for which experimental results are available: to this aim, we again perform the inference while masking one of our four constriction sizes (2, 3, 5, and 15 μm). We then use the model inferred from the other three constriction widths to make predictions on the fourth geometry, which includes interpolations (when masking the 3 and 5 μm sets) and extrapolations (when masking the 2 and 15 μm data). We then compare, in Fig. 5(a), the prediction on the fourth, masked geometry with the actual experimental results. The good general agreement confirms the validity of this approach and shows the usefulness of the learned model to predict behavior in geometries not used to train it. Note, however, that the simulation results exhibit smoother geometrical deformation than the experiments, in particular past the constriction. This discrepancy could be due to the small number of data points during the short time interval when the nuclei pass through the constraint.

Next, we extrapolate the model to other constriction sizes that were not studied experimentally: using the inferred model trained on the full data set, we simulate trajectories and



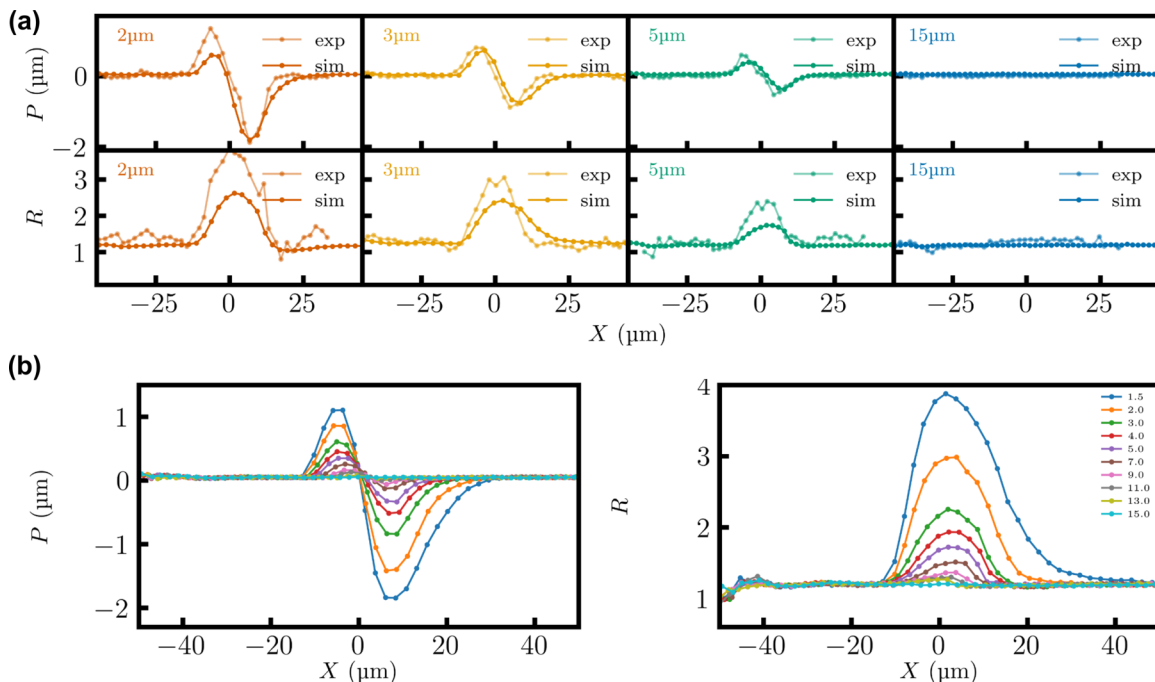


FIG. 5. Predictivity of the inferred model to other values of pillar radius  $r$ . (a) Extrapolation from partial data compared with the experimental data. (b) Comparison of the nuclear protrusion  $P$  against nuclei position  $X$  (first row) and aspect ratio against  $X$  (second row). (c) Extrapolation over a range of constraints with constriction size listed in the legend in  $\mu\text{m}$ .

compute the average geometric quantities  $P(X)$  and  $R(X)$  in constraints ranging from  $1.0\mu\text{m}$  to  $15\mu\text{m}$  [Fig. 5(b)]. Each curve is obtained by averaging over 1000 simulated trajectories. We observe a continuous increase in the maximum aspect ratio and protrusion as the constraint gets smaller. These predictions could be used for future experiment design, as a way to explore parameter space and focus experiments on the regions of interest. In principle, we could also extrapolate the model to other geometries, for instance, noncylindrical constrictions or channels of different widths, without further training, by directly modifying the position-dependent width function  $w(x)$  as well as the normal vector coordinates  $n_x(x) = \partial_x w / \sqrt{1 + \partial_x w^2}$  and  $n_y(x) = 1 / \sqrt{1 + \partial_x w^2}$  in the learned model. Furthermore, if data were available in such noncylindrical geometries, they could be included in the data set to learn this same model with better accuracy. In practice, however, here we focus on cylindrical constrictions of variable pillar radius.

### III. DISCUSSION

In this article, we have studied the spontaneous migration of individual cells in a microfluidic device that exerts tight three-dimensional constraints mimicking physiological scenarios where cells are able to migrate in strongly confined environments. Strikingly, cells can pass through constrictions much smaller than the rest diameter of their nucleus, leading to large deformations of the nucleus during translocation [3,14]. This controlled experimental setup differs from previous studies of 2D confined cell migration without three-dimensional constraints [41], in which the nucleus is not significantly deformed. We segment and track cell nuclei to obtain trajectories that we use to quantify the dynamics of

this nuclear translocation process. To this aim, we employ a data-driven approach that captures the stochastic nature of the motion and shape changes of the nucleus during cell motility in strongly constraining environments. In contrast with previous works where only the nucleus position was used [24], leading to effectively inertial dynamics, we include shape descriptors in our model that provide a proxy for the unobserved polarity of the cell. The outcome is an optimized set of overdamped equations that qualitatively and semiquantitatively captures the joint dynamics of nuclear position, protrusion, and elongation as coupled stochastic differential equations.

Our data-driven pipeline to infer these SDEs includes three main methodological developments compared to preexisting methods. First and most importantly, we symbolically include the geometry of the constriction in the inferred model, which allows us to train a single model on data for several constriction sizes. The resulting model can be successfully extrapolated to other constriction sizes and, potentially, other geometries. This contrasts with previous methods where the geometry is hard-coded into the inferred model, which both prevents extrapolation and precludes interpretability [24,41]. Second, we have introduced a sparsity-enforcing algorithm which simply consists in removing statistically insignificant terms from the set of basis functions, leveraging bootstrap estimates of significance to simplify the learned model. The desired level of significance (chosen here to be a standard  $3\sigma$ ) is the only parameter of this technique, contrary to popularly used sparse inference methods which include penalization terms with hyperparameters that require fine tuning [16,42,43]. Third, we propose a modification of the stochastic force inference algorithm that consists in using trapezoidal integration for normalization matrices. While minor, this modification significantly improves the performance of the

method when the time step  $\Delta t$  is large. The combination of these improvements allows us to robustly infer a minimal model for the geometrical kinematics of nuclear translocation.

Nuclear translocation involves a complex set of molecular mechanisms that enables cells to sense their mechanical environment and adapt their internal forces. Our study paves the way towards a data-driven understanding of this process, where the nucleus is considered as an actor of the dynamical process, rather than a passive tracer lagging behind. In the future, this description could be enriched with other cell state descriptors, in particular with the spatial distribution of cytoskeletal and nuclear components, such as protein complexes involved in the mechanotransduction process. A challenge towards this, however, consists in selecting appropriate quantitative descriptors to include in the dynamical model: for instance, while nesprin—the mechanical linking protein between cytoskeleton and nucleus—is observed to accumulate at the front of the nucleus, this is not recapitulated by a polarity defined in terms of the first moment of the protein distribution (see Appendix B 3).

### ACKNOWLEDGMENTS

We thank Thierry Dubois and his team (Breast Cancer Biology Group, Institut Curie—PSL Research University, Paris, France), as well as Julie Plastino and Inge Bos (Active Cell Matter team, LPENS, Paris, France), Chase Broedersz (Vrije Universiteit, Amsterdam, The Netherlands), David Brückner (Institute of Science and Technology Austria), and Jan Lammerding (Cornell University, USA) for their help, support, and fruitful discussions. This work was carried out in part at the Nikon Imaging Centre, Institut Curie, member of the French National Research Infrastructure France-BioImaging (Grant No. ANR10-INBS-04). The project leading to this publication was supported by an institutional Institut Curie grant and received funding from France 2030, the French Government program managed by the French National Research Agency (Grant No. ANR-16-CONV-0001) and from the Excellence Initiative of Aix-Marseille University—A\*MIDEX.

Co-funded by the European Union (ERC-SuperStoc-101117322). Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

## APPENDIX A: MATERIAL AND METHODS

### 1. Cell culture

Mouse embryonic fibroblasts (MEFs) were CRISPR-modified to create a new cell line: MEFs SYNE2-GFP LMNA-mCh, as described and validated in [29]. Cells are cultured at 37 °C in a humidified incubator with 5% CO<sub>2</sub>, in DMEM (Dulbecco's modified eagle medium: Gibco) supplemented with 10% (v/v) fetal bovine serum (FBS: Gibco).

### 2. Migration devices

The epoxy mold (R123/R614: Soloplast) we used was replicated from a polydimethylsiloxane (PDMS) imprinted

piece coming from the laboratory of Jan Lammerding (Cornell University, USA). A mix of PDMS (using a 10:1 ratio polymer:crosslinker) is vacuumed for 20 minutes to avoid bubbles, then poured into the epoxy mold and let to cure for 4 hours in a 60 °C oven. Imprinted PDMS pieces are cut using a scalpel and biopsy punches (2 and 5 mm in diameter). Glass coverslips are soaked overnight in a 0.2 M solution of HCl and rinsed with H<sub>2</sub>O and ethanol, then dried with Kim wipes. To form a migration device, an imprinted PDMS piece and a treated glass coverslip are placed in a plasma cleaner for 1 minute and gently stuck together. This process creates covalent bonds between the PDMS and the glass [45]. Devices are then directly put on a 100 °C hot plate for 5 minutes to help the sticking process.

### 3. Cell migration experiment

Microfluidic devices are sterilized and rinsed under a microbiological safety post: first once with ethanol (~250  $\mu$ L) then twice with phosphate-buffered saline (PBS: Gibco) and twice with DMEM (Gibco) supplemented with 10% (v/v) FBS. Cells are suspended at a concentration of 10 millions per mL in DMEM (Gibco) supplemented with 10% (v/v) FBS. They are seeded in the device by adding 5  $\mu$ L of the suspended solution in one of the two small ports of the device. After 6 hours, enough cells are in the constricted region of the device and acquisition can start. For that, the cell medium is changed to DMEM without phenol red and with HEPES (15 mM) (Gibco), supplemented with 10% FBS (Gibco), 100 units/mL penicillin, and 100  $\mu$ g/mL streptomycin (Life Technologies).

### 4. Image acquisition

Time-lapse acquisitions are performed on an epifluorescence microscope (Nikon Ti-E) equipped with a sCMOS camera (2048 ORCA Flash 4.0 V2, Hamamatsu or Prime BSI, Teledyne), a perfect focus system, a 60x oil objective (Nikon), and a temperature and gas control chamber (set on 37 °C, air at 5% CO<sub>2</sub>). Images are taken every 10 minutes.

### 5. Image analysis

Movies are analyzed using IMAGE J/FIJI and PYTHON. The projected nucleus surface is detected by using the “analyze particles” function on a threshold (median filter to 5.0 radius, normalized by 0.4% and autolocal threshold “Bernsen” 5) applied on the mCherry image (corresponding to a lamin A/C signal). The nucleus contour is defined by a band of 1  $\mu$ m width created from the detected nucleus projected surface (“reduce” and “make band” functions).

### 6. Definition of the spatial origin

The origin of the  $x$  axis is set at the center of the constriction pillar (2, 3, 5  $\mu$ m) or half pillar (15  $\mu$ m). The origin of the  $y$  axis is set at the top center of the bottom constriction pillar (2, 3, 5  $\mu$ m) or the top center of the fitted circle to the bottom half pillar (15  $\mu$ m).

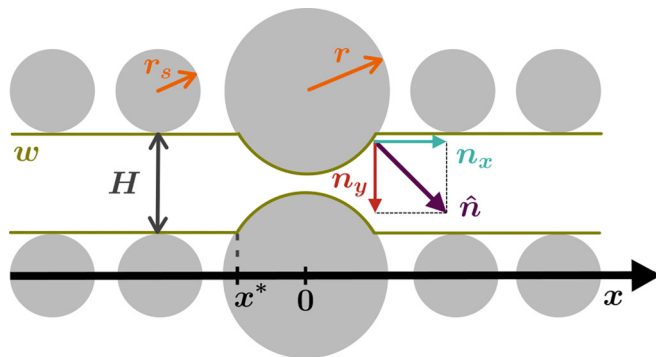


FIG. 6. Schematics of the geometric quantities describing the pillar shape used to construct force estimators. Note that for one series, the centers of the pillars are aligned on  $x$ .

### 7. Definition of $X$ , $X_c$ , $P$ , $R$

The position of a nucleus is defined by its surface barycenter  $(X, Y)$ , specifically,  $X = \sum_{i=0}^n x_i/n$  and  $Y = \sum_{i=0}^n y_i/n$ , with  $(x_i, y_i)$  the coordinates of each pixel  $i$  of the nucleus surface and  $n$  the number of pixels in the nucleus surface. The  $x$  coordinate of the center of the nucleus contour  $X_c$  is defined as  $X_c = \sum_{i=0}^{n_c} x_{c,i}/n_c$ , with  $x_{c,i}$  the  $x$  coordinates of each pixel of the nucleus contour and  $n_c$  the number of pixels in the nucleus contour. The nucleus protrusion vector is defined as  $P = X_c - X$ . The aspect ratio of the nucleus is defined by  $R = R_x/R_y$ , where  $R_x = \sqrt{\sum_{i=0}^n (x_i - X)^2/n}$  and  $R_y = \sqrt{\sum_{i=0}^n (y_i - Y)^2/n}$ .

### 8. Complementary expressions for basis functions

The direct and indirect effects of the environment,  $f_{\text{ext}}(X, r)$ ,  $f_P(X, r)$ , and  $f_R(X, r)$ , are approximated by combining the width function of the channel,  $w(X, r)$ , with the normal vector calculated from the shape of the pillar, which is a circle of radius  $r$ . The normal vector  $\hat{n}(X, r) = [n_x(X, r), n_y(X, r)]$  of a circle of radius  $r$  is given by

$$n_x(x, r) = \frac{x}{r}, \quad -x^* < x < x^*, \quad (\text{A1})$$

$$n_x(x, r) = 0 \quad \text{otherwise}, \quad (\text{A2})$$

$$n_y(x, r) = \frac{\sqrt{r^2 - x^2}}{r}, \quad -x^* < x < x^*, \quad (\text{A3})$$

$$n_y(x, r) = \frac{\sqrt{r^2 - x^{*2}}}{r} \quad \text{otherwise}, \quad (\text{A4})$$

where  $x^* = \sqrt{r^2 - r_s^2}$ , with  $r_s$  the small pillar radius. These quantities are schematized in Fig. 6. The channel width  $w(x)$  is given by

$$w(x, r) = H + 2r_s - 2\sqrt{r^2 - x^2}, \quad -x^* < x < x^*, \quad (\text{A5})$$

$$w(x, r) = H \quad \text{otherwise}, \quad (\text{A6})$$

where  $H = 15$  is the channel height (note that we neglect the texture of the small pillars here, as they do not constrict the nucleus).

### 9. Full expression of the initial model

The full model consisting of all basis functions, constructed by systematic expansion of the model over physically relevant variables, consists in linear combinations of the following basis functions:

$$\begin{aligned} & \{1, P, (R - R^{-1}), \\ & 1/w(X, r), n_x(X, r)/w(X, r), n_y(X, r)/w(X, r), \\ & 1/w^2(X, r), n_x(X, r)/w^2(X, r), n_y(X, r)/w^2(X, r), \\ & 1/w^3(X, r), n_x(X, r)/w^3(X, r), n_y(X, r)/w^3(X, r)\}, \quad (\text{A7}) \end{aligned}$$

and has 39 parameters ( $3 \times 12$  for the drift, 3 for the diffusion). The model inference and reduction framework described in Fig. 3 retains only 14 significant terms.

### 10. Improvement of the SFI algorithm for large time intervals $\Delta t$

To learn the dynamics of our system, characterized in this paragraph by the vector  $\mathbf{x}_t \equiv (X_t, P_t, R_t)$ , it is essential to estimate its discrete time derivative using  $\Delta \mathbf{x}_t = \mathbf{x}_{t+\Delta t} - \mathbf{x}_t$ . A challenge on applying SFI on cellular dynamics data is that the time interval  $\Delta t$  between frames is large (10 minutes) and of the same order of magnitude as the typical translocation time. It cannot be easily reduced due to phototoxicity: with the previously introduced algorithm [27], this incurs  $O(\Delta t)$  biases on the estimators. To adapt the method to this challenge, we propose a modification, which results in much smaller  $O(\Delta t^2)$  biases.

Specifically, we focus in the derivation on approximate time differences  $\Delta \mathbf{x}_t$  rather than the infinitesimal time difference  $d\mathbf{x}_t$ . Writing the dynamics in a generic form  $\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}_t) + \sqrt{2\mathbf{D}}\eta_t$ , we have, for discrete time increments,

$$\frac{\Delta \mathbf{x}_t}{\Delta t} = \frac{1}{\Delta t} \int_t^{t+\Delta t} \frac{d\mathbf{x}_{t'}}{dt'} dt' \quad (\text{A8})$$

$$= \frac{1}{\Delta t} \int_t^{t+\Delta t} [\mathbf{f}(\mathbf{x}_{t'}) + \sqrt{2\mathbf{D}}\eta_{t'}] dt'. \quad (\text{A9})$$

SFI consists in approximating the unknown deterministic drift field  $\mathbf{f}(\mathbf{x})$  by a linear combination of basis functions  $\mathbf{f}(\mathbf{x}) = \sum_{\alpha} F_{\alpha} \mathbf{b}_{\alpha}(\mathbf{x})$ , where  $F_{\alpha}$  are the coefficients to infer and  $\mathbf{b}_{\alpha}(\mathbf{x})$  are the basis functions. Thus, we can project the above equation on one of the basis functions  $\mathbf{b}_{\gamma}(\mathbf{x}_t)$  and derive its average in the Itô convention,

$$\left\langle \frac{\Delta \mathbf{x}_t}{\Delta t} \mathbf{b}_{\gamma}(\mathbf{x}_t) \right\rangle = \sum_{\alpha} F_{\alpha} \left\langle \mathbf{b}_{\gamma}(\mathbf{x}_t) \frac{1}{\Delta t} \int_t^{t+\Delta t} \mathbf{b}_{\alpha}(\mathbf{x}_{t'}) dt' \right\rangle, \quad (\text{A10})$$

where  $\langle \cdot \rangle$  represents the expectation over many realizations of the noise  $\eta_t$ , conditioned on the initial value  $\mathbf{x}_t$ . Since we only measure  $\mathbf{x}$  at discrete times  $t, t + \Delta t, \dots$ , we need to approximate  $\int_t^{t+\Delta t} \mathbf{b}_{\alpha}(\mathbf{x}_{t'}) dt' \approx \frac{\Delta t}{2} [\mathbf{b}_{\alpha}(\mathbf{x}_t) + \mathbf{b}_{\alpha}(\mathbf{x}_{t+\Delta t})]$ . Importantly, this *trapezoidal integration rule* is a more accurate approximation than the Riemann sum approximation  $\mathbf{b}_{\alpha}(\mathbf{x}_t)\Delta t$  used in Ref. [27]. We note that the trapezoidal approximation concerns only the right-hand side of Eq. (A10), which is a regular integral, and not the left-hand side, which is a stochastic integral and remains in the Itô convention.

By now averaging Eq. (A10) over all data points  $\{\mathbf{x}_t\}_{i=1,N}$ , represented by  $\langle \cdot \rangle_t$ , we obtain the *trapezoidal* approximation for the normalization matrix  $\langle \frac{1}{\Delta t} \int_t^{t+\Delta t} \mathbf{b}_\alpha(\mathbf{x}_{t'}) dt' \rangle \mathbf{b}_\gamma(\mathbf{x}_t)$ :

$$\mathbf{B}_{\alpha\gamma} = \left\langle \frac{1}{2} [\mathbf{b}_\alpha(\mathbf{x}_t) + \mathbf{b}_\alpha(\mathbf{x}_{t+\Delta t})] \mathbf{b}_\gamma(\mathbf{x}_t) \right\rangle_t, \quad (\text{A11})$$

which we use in the corrected Itô estimator of  $F_\alpha$  for large  $\Delta t$ ,

$$\hat{F}_\alpha = \sum_\gamma \mathbf{B}^{-1}_{\alpha\gamma} \left\langle \frac{\Delta \mathbf{x}_t}{\Delta t} \mathbf{b}_\gamma(\mathbf{x}_t) \right\rangle_t. \quad (\text{A12})$$

Finally, similarly to Ref. [27], we modify the integration convention of the stochastic integral towards the Stratonovich convention, in order to remove biases due to measurement noise, yielding the estimator used throughout this article,

$$\hat{F}_\alpha = \sum_\gamma \mathbf{B}^{-1}_{\alpha\gamma} \left\langle \frac{\Delta \mathbf{x}_t}{\Delta t} \left[ \frac{\mathbf{b}_\gamma(\mathbf{x}_t) + \mathbf{b}_\gamma(\mathbf{x}_{t+\Delta t})}{2} \right] - \mathbf{d}_t \nabla \mathbf{b}_\gamma(\mathbf{x}_t) \right\rangle_t, \quad (\text{A13})$$

where we use the instantaneous noise-corrected diffusion estimator [46],

$$\mathbf{d}_t = [(\Delta \mathbf{x}_t + \Delta \mathbf{x}_{t-\Delta t})^2 + 2\Delta \mathbf{x}_t \Delta \mathbf{x}_{t-\Delta t}] / 4\Delta t. \quad (\text{A14})$$

The use of the trapezoidal method for discrete differences thus results in a lower-order discretization bias compared to the original SFI method, and enables accurate inference with the data set considered in this article.

### 11. Bootstrap methods for coefficient error estimation

We estimate the mean and standard deviation of each coefficient and the diffusion constant using the bootstrap method. More specifically, we sample with replacements the set of small five-consecutive-point trajectories to generate an ensemble of trajectories [44]. For each sample, the drift coefficients are estimated with the above-mentioned modified SFI algorithm, while the diffusion coefficients are estimated using the method introduced by Vestergaard *et al.* [27,46]. We compute the average and standard deviation of coefficients over 20 bootstrapped data sets obtained from the initial set of trajectories, and use the resulting standard deviations as indicators of the confidence interval for our assessment of the statistical significance of these coefficients.

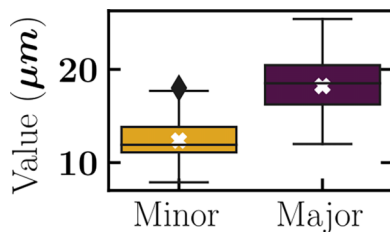


FIG. 7. MEFs nuclei size is measured through the minor and major axes of a fitted ellipse. Width (*minor axis of the fitted ellipse*) and length (*major axis*) of MEFs nuclei measured inside the microfluidic device when  $X = -20 \pm 5 \mu\text{m}$ , regardless of the constriction size. White cross represents the average value of the distribution.  $N = 75$ .

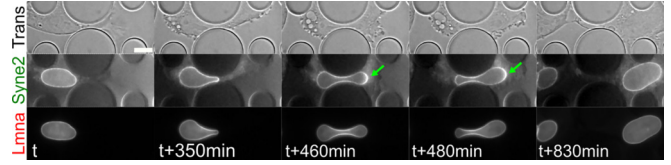


FIG. 8. MEF cell going through a  $2 \mu\text{m}$  wide constriction. Top: Transmission signal. Middle: GFP signal corresponding to the expression of Syne2-GFP gene for nesprin 2G proteins. Arrows point at hyperfluorescence of nesprin at the front of the nucleus during translocation. Bottom: mCherry signal corresponding to the expression Lmna-mCh gene for lamin A/C proteins. Scale bar:  $10 \mu\text{m}$ .

## APPENDIX B: COMPLEMENTARY RESULTS

### 1. Size of MEFs nuclei

The apparent size of the MEFs nuclei in the microfluidic device is calculated at  $X = -20 \mu\text{m}$ , before they are deformed by the constrictions. In fact, we allow a confidence interval of  $5 \mu\text{m}$ . We fit an ellipse to the detected nucleus and measure its major and minor axes (Fig. 7).

### 2. Time lapse of MEFs going through a $2 \mu\text{m}$ constriction

In agreement with previous results [29], we observe that when cells migrate through  $2 \mu\text{m}$  constrictions (Fig. 8, *Trans*), nesprin accumulates at the front of the nucleus (Fig. 8, *Syne 2*, green arrows), whereas lamin do not (Fig. 8, *Lmna*).

### 3. Nesprin distribution polarity vector

For each constriction size ( $2, 3,$  and  $5 \mu\text{m}$ ) and for the  $15 \mu\text{m}$  large channel, when the nucleus is in the middle of the constriction, we measure the intensity along a curvilinear abscissa ( $s$ ) around the nucleus (of perimeter  $L$ ) at  $X = 0 \mu\text{m}$

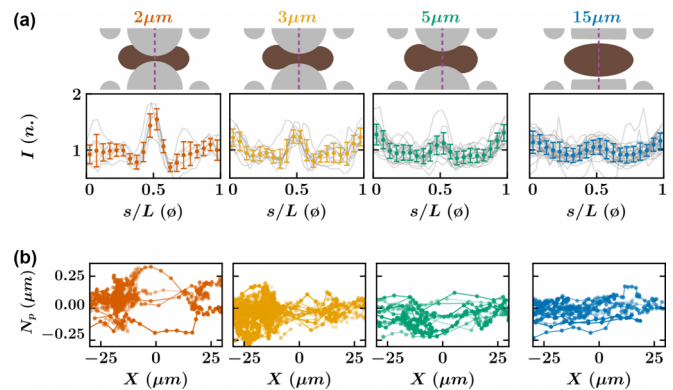


FIG. 9. Nesprins accumulate at the front of their nuclei, but the defined polarity vector does not show this tendency. (a) Nesprin 2G intensity signals measured along the curvilinear abscissa  $s$  when the nucleus (of perimeter  $L$ ) is in the middle of the constriction (large channel):  $X = 0 \pm 5 \mu\text{m}$ . Intensities are normalized by the contrast and divided by the average on the contour. From left to right, the number of cells shown is 7, 11, 23, and 32. Error bars correspond to the standard deviation. (b) Calculated nesprin polarity vector ( $N_p$ ) according to nucleus position ( $X$ ) for different constrictions sizes ( $2, 3,$  and  $5 \mu\text{m}$ ), as well as large channels ( $15 \mu\text{m}$ ). Each line represents a single cell. Dots are experimental data points. From left to right, the number of cells shown is 14, 36, 15, and 20.



[Fig. 9(a)]. At  $s/L = 0.5$ , which corresponds to the front of the nucleus, we confirm the increase of the nesprin signal intensity for the three constriction sizes (2, 3, and 5  $\mu\text{m}$ ) that does not happen in the 15  $\mu\text{m}$  large channel. This accumulation of nesprin at the front of the nucleus appears to be more dramatic as the constriction gets smaller.

In order to account for the nesprin distribution in a quantitative unidimensional way, we construct a *nesprin polarity vector*. First, we define  $X_n$  as the barycenter of the nucleus contour weighted by nesprin intensity. Explicitly,

$$X_n = \frac{\sum_{i=0}^{n_c} x_{c,i} I_{c,i}}{\sum_{i=0}^{n_c} I_{c,i}}, \quad (\text{B1})$$

with  $(I_{c,i})$  the corresponding intensity values of each of the pixels  $(x_{c,i})$  of the nucleus contour. Second, we define a

nesprin polarity vector  $N_p$  as  $N_p = X_n - X_c$ . It is calculated for cells migrating through small constrictions (2, 3, and 5  $\mu\text{m}$ ) and large channels (15  $\mu\text{m}$ ) along nucleus position  $X$  [Fig. 9(b)]. In fact,  $N_p$  is noisy, almost constant over nucleus position, and close to zero. In addition, there is no significant difference between the nesprin polarity of cells migrating through small constrictions and of cells migrating through large channels, when experimentally nesprins do not accumulate at the front of the nucleus in the 15  $\mu\text{m}$  case. This means that the barycenter of the contour weighted by the nesprin intensity  $X_n$  and the barycenter of the contour  $X_c$  are close in value. Therefore, the application of our method to dynamical changes in protein distribution will need further investigation, in particular to account for nesprin accumulation at the nucleus front.

- 
- [1] F. Guilak, J. R. Tedrow, and R. Burgkart, Viscoelastic properties of the cell nucleus, *Biochem. Biophys. Res. Commun.* **269**, 781 (2000).
- [2] C. Uhler and G. V. Shivashankar, Regulation of genome organization and gene expression by nuclear mechanotransduction, *Nat. Rev. Mol. Cell Biol.* **18**, 717 (2017).
- [3] P. M. Davidson, C. Denais, M. C. Bakshi, and J. Lammerding, Nuclear deformability constitutes a rate-limiting step during cell migration in 3-D environments, *Cell. Mol. Bioeng.* **7**, 293 (2014).
- [4] K. Wolf, M. Te Lindert, M. Krause, S. Alexander, J. Te Riet, A. L. Willis, R. M. Hoffman, C. G. Figdor, S. J. Weiss, and P. Friedl, Physical limits of cell migration: Control by ECM space and nuclear deformation and tuning by proteolysis and traction force, *J. Cell Biol.* **201**, 1069 (2013).
- [5] A. E. Patteson, A. Vahabikashi, K. Pogoda, S. A. Adam, K. Mandal, M. Kittisopikul, S. Sivagurunathan, A. Goldman, R. D. Goldman, and P. A. Janmey, Vimentin protects cells against nuclear rupture and DNA damage during migration, *J. Cell Biol.* **218**, 4079 (2019).
- [6] I. D. Estabrook, H. R. Thiam, M. Piel, and R. J. Hawkins, Calculation of the force field required for nucleus deformation during cell migration through constrictions, *PLoS Comput. Biol.* **17**, e1008592 (2021).
- [7] A. B. Verkhovsky, T. M. Svitkina, and G. G. Borisy, Self-polarization and directional motility of cytoplasm, *Curr. Biol.* **9**, 11 (1999).
- [8] P. Maiuri, J.-F. Rupprecht, S. Wieser, V. Rupprecht, O. Bénichou, N. Carpi, M. Coppey, S. De Beco, N. Gov, C.-P. Heisenberg, C. L. Crespo, F. Lautenschlaeger, M. Le Berre, A.-M. Lennon-Dumenil, M. Raab, H.-R. Thiam, M. Piel, M. Sixt, and R. Voituriez, Actin flows mediate a universal coupling between cell speed and cell persistence, *Cell* **161**, 374 (2015).
- [9] K. Keren, Z. Pincus, G. M. Allen, E. L. Barnhart, G. Marriott, A. Mogilner, and J. A. Theriot, Mechanism of shape determination in motile cells, *Nature (London)* **453**, 475 (2008).
- [10] D. E. Jaalouk and J. Lammerding, Mechanotransduction gone awry, *Nat. Rev. Mol. Cell Biol.* **10**, 63 (2009).
- [11] M. Maurer and J. Lammerding, The driving force: Nuclear mechanotransduction in cellular function, fate, and disease, *Annu. Rev. Biomed. Eng.* **21**, 443 (2019).
- [12] V. Venturini, F. Pezzano, F. Català Castro, H.-M. Häkkinen, S. Jiménez-Delgado, M. Colomer-Rosell, M. Marro, Q. Tolosa-Ramon, S. Paz-López, M. A. Valverde, J. Weghuber, P. Loza-Alvarez, M. Krieg, S. Wieser, and V. Ruprecht, The nucleus measures shape changes for cellular proprioception to control dynamic cell behavior, *Science* **370**, eaba2644 (2020).
- [13] A. J. Lomakin, C. J. Cattin, D. Cuvelier, Z. Alraies, M. Molina, G. P. F. Nader, N. Srivastava, P. J. Sáez, J. M. Garcia-Arcos, I. Y. Zhitnyak, A. Bhargava, M. K. Driscoll, E. S. Welf, R. Fiolka, R. J. Petrie, N. S. De Silva, J. M. González-Granado, N. Manel, A. M. Lennon-Dumenil, D. J. Müller *et al.*, The nucleus acts as a ruler tailoring cell responses to spatial constraints, *Science* **370**, eaba2894 (2020).
- [14] P. M. Davidson, J. Sliz, P. Isermann, C. Denais, and J. Lammerding, Design of a microfluidic device to quantify dynamic intra-nuclear deformation during cell migration through confining environments, *Integr. Biol.* **7**, 1534 (2015).
- [15] J. Keys, A. Windsor, and J. Lammerding, Assembly and use of a microfluidic device to study cell migration in confined environments, in *The LINC Complex*, edited by G. G. Gundersen and H. J. Worman (Springer, New York, 2018), Vol. 1840, pp. 101–118.
- [16] S. L. Brunton, J. L. Proctor, and J. N. Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems, *Proc. Natl. Acad. Sci. USA* **113**, 3932 (2016).
- [17] K. Champion, B. Lusch, J. N. Kutz, and S. L. Brunton, Data-driven discovery of coordinates and governing equations, *Proc. Natl. Acad. Sci. USA* **116**, 22445 (2019).
- [18] N. Romeo, A. Hastewell, A. Mietke, and J. Dunkel, Learning developmental mode dynamics from single-cell trajectories, *eLife* **10**, e68679 (2021).
- [19] M. S. Schmitt, J. Colen, S. Sala, J. Devany, S. Seetharaman, M. L. Gardel, P. W. Oakes, and V. Vitelli, Zyxin is all you need: Machine learning adherent cell mechanics, *arXiv:2303.00176*.
- [20] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, in Applications of Mathematics (Springer, Berlin, 2010), Vol. 23.
- [21] D. Selmeczi, S. Mosler, P. H. Hagedorn, N. B. Larsen, and H. Flyvbjerg, Cell motility as persistent random motion: Theories from experiments, *Biophys. J.* **89**, 912 (2005).

- [22] D. Selmeczi, L. Li, L. I. I. Pedersen, S. F. Nrrrelykke, P. H. Hagedorn, S. Mosler, N. B. Larsen, E. C. Cox, and H. Flyvbjerg, Cell motility as random motion: A review, *Eur. Phys. J.: Spec. Top.* **157**, 1 (2008).
- [23] L. Li, E. C. Cox, and H. Flyvbjerg, “Dicty dynamics”: Dictyostelium motility as persistent random motion, *Phys. Biol.* **8**, 046006 (2011).
- [24] D. B. Brückner, N. Arlt, A. Fink, P. Ronceray, J. O. Rädler, and C. P. Broedersz, Learning the dynamics of cell–cell interactions in confined cell migration, *Proc. Natl. Acad. Sci.* **118**, e2016602118 (2021).
- [25] D. B. Brückner, M. Schmitt, A. Fink, G. Ladurner, J. Flommersfeld, N. Arlt, E. Hannezo, J. O. Rädler, and C. P. Broedersz, Geometry adaptation of protrusion and polarity dynamics in confined cell migration, *Phys. Rev. X* **12**, 031041 (2022).
- [26] S. Stöberl, J. Flommersfeld, M. M. Kreft, M. Benoit, C. P. Broedersz, and J. O. Rädler, Nuclear deformation and dynamics of migrating cells in 3D confinement reveal adaptation of pulling and pushing forces, *bioRxiv* (2023), doi: 10.1101/2023.10.30.564765.
- [27] A. Frishman and P. Ronceray, Learning force fields from stochastic trajectories, *Phys. Rev. X* **10**, 021009 (2020).
- [28] D. B. Brückner, P. Ronceray, and C. P. Broedersz, Inferring the dynamics of underdamped stochastic systems, *Phys. Rev. Lett.* **125**, 058103 (2020).
- [29] P. M. Davidson, A. Battistella, T. Déjardin, T. Betz, J. Plastino, N. Borghi, B. Cadot, and C. Sykes, Nesprin-2 accumulates at the front of the nucleus during confined cell migration, *EMBO Rep.* **21**, e49910 (2020).
- [30] A. Vahabikashi, S. A. Adam, O. Medalia, and R. D. Goldman, Nuclear lamins: Structure and function in mechanobiology, *APL Bioengineer.* **6**, 011503 (2022).
- [31] Y. Gruenbaum and R. Foisner, Lamins: Nuclear intermediate filament proteins with fundamental functions in nuclear mechanics and genome regulation, *Annu. Rev. Biochem.* **84**, 131 (2015).
- [32] E. C. Tapley and D. A. Starr, Connecting the nucleus to the cytoskeleton by SUN–KASH bridges across the nuclear envelope, *Curr. Opin. Cell Biol.* **25**, 57 (2013).
- [33] M. Crisp, Q. Liu, K. Roux, J. B. Rattner, C. Shanahan, B. Burke, P. D. Stahl, and D. Hodzic, Coupling of the nucleus and cytoplasm: Role of the LINC complex, *J. Cell Biol.* **172**, 41 (2006).
- [34] J. A. Mellad, D. T. Warren, and C. M. Shanahan, Nesprins LINC the nucleus and cytoskeleton, *Curr. Opin. Cell Biol.* **23**, 47 (2011).
- [35] L. Tweedy, D. A. Knecht, G. M. Mackay, and R. H. Insall, Self-generated chemoattractant gradients: Attractant depletion extends the range and robustness of chemotaxis, *PLoS Biol.* **14**, e1002404 (2016).
- [36] C. M. Denais, R. M. Gilbert, P. Isermann, A. L. McGregor, M. Te Lindert, B. Weigelin, P. M. Davidson, P. Friedl, K. Wolf, and J. Lammerding, Nuclear envelope rupture and repair during cancer cell migration, *Science* **352**, 353 (2016).
- [37] C. R. Pfeifer, M. P. Tobin, S. Cho, M. Vashisth, L. J. Dooling, L. L. Vazquez, E. G. Ricci-De Lucca, K. T. Simon, and D. E. Discher, Gaussian curvature dilutes the nuclear lamina, favoring nuclear rupture, especially at high strain rate, *Nucleus* **13**, 130 (2022).
- [38] D. B. Brückner and C. P. Broedersz, Learning dynamical models of single and collective cell migration: A review, *arXiv:2309.00545*.
- [39] D. G. Drubin and W. J. Nelson, Origins of cell polarity, *Cell* **84**, 335 (1996).
- [40] J. P. Crutchfield and B. S. McNamara, Equations of motion from a data series, *Complex Syst.* **1**, 417 (1987).
- [41] D. B. Brückner, A. Fink, C. Schreiber, P. J. F. Röttgermann, J. O. Rädler, and C. P. Broedersz, Stochastic nonlinear dynamics of confined cell migration in two-state systems, *Nat. Phys.* **15**, 595 (2019).
- [42] L. Boninsegna, F. Nüske, and C. Clementi, Sparse learning of stochastic dynamical equations, *J. Chem. Phys.* **148**, 241723 (2018).
- [43] J. L. Callahan, J.-C. Loiseau, G. Rigas, and S. L. Brunton, Nonlinear stochastic modelling with Langevin regression, *Proc. R. Soc. A* **477**, 20210092 (2021).
- [44] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap* (Chapman and Hall/CRC, New York, 1994).
- [45] A. Borók, K. Laboda, and A. Bonyár, PDMS bonding technologies for microfluidic applications: A review, *Biosensors* **11**, 292 (2021).
- [46] C. L. Vestergaard, P. C. Blainey, and H. Flyvbjerg, Optimal estimation of diffusion coefficients from single-particle trajectories, *Phys. Rev. E* **89**, 022726 (2014).