Stochastic dynamics of migrating cells

A data-driven approach

David Benedikt Brückner

Munich 2021

Stochastic dynamics of migrating cells

A data-driven approach

David Benedikt Brückner

A dissertation

submitted to the Faculty of Physics

at the Ludwig–Maximilians–Universit"at München

for the degree of

Doctor rerum naturalium

Munich, 5th April 2021

First referee: Prof. Dr. Chase Broedersz
Second referee: Prof. Dr. Joachim Rädler
Day of the oral examination: 26th May 2021

Zusammenfassung **(Synopsis in German)**

Zellmigration ist ein Kernelement vieler physiologischer Phänomene wie der Embryogenese, dem Immunsystem und der Krebsmetastase. In all diesen Prozessen stehen Zellen vor einer physikalischen Herausforderung: Sie bewegen sich in beengten Umgebungen, in denen sie Engstellen passieren müssen. Die Zellbewegung wird von einer komplexen Maschinerie angetrieben, deren molekulare Komponenten immer besser verstanden werden. Demgegenüber fehlt ein quantitatives Verständnis des funktionalen Migrationsverhaltens der Zelle als Ganzes. Die verbindende Fragestellung der Projekte in dieser Arbeit lautet daher: gibt es emergente dynamische 'Gesetze', die die Verhaltensdynamik migrierender Zellen in beengten Umgebungen beschreiben?

Um dieser Frage nachzugehen, entwickeln wir datengetriebene Ansätze, die es uns erlauben, die Dynamik migrierender Zellen direkt aus experimentellen Daten zu inferieren. Wir untersuchen Zellmigration in künstlichen Systemen, in denen Zellen Engstellen wiederholt passieren müssen. Aus den experimentellen Zelltrajektorien inferieren wir eine Bewegungsgleichung, die die Dynamik in deterministische und stochastische Komponenten trennt. Diese Methode zeigt, dass sich Zellen deterministisch 'aktiv' in die Engstellen hineinbewegen, ganz entgegen der intuitiven Erwartung, dass Engstellen als Hindernis fungieren könnten. Dieser aktive Antrieb führt zu einer komplexen nichtlinearen Dynamik im Übergangsbereich zwischen einem bistabilen System und einem Grenzzyklus-Oszillator. Wir verallgemeinern diesen datenbasierten Ansatz, um die Varianz des Migrationsverhaltens innerhalb einer Zellpopulation zu quantifizieren, und analysieren, wie Zellen auf die Größe, Form und Orientierung ihrer Umgebung reagieren.

Darauf aufbauend untersuchen wir die zugrundeliegenden Mechanismen dieser Dynamik. Zellmigration basiert auf verschiedenen zellulären Komponenten, wie unter Anderem den Zellprotrusionen und der Adhäsion mit der Umgebung. Auf Basis der experimentellen Daten entwickeln wir ein mechanistisches Modell für Zellmigration in beengten Systemen, welches zeigt, dass der beobachtete aktive Antrieb eine Konsequenz zweier Effekte ist: Einer variierenden Adhäsion mit der Umgebung und einer Zellpolarität, die sich in Engstellen selbst verstärkt. Diese Ergebnisse deuten darauf hin, dass die Anpassung der Zellpolarität an die lokale Geometrie ein Schlüsselmechanismus in beengter Zellmigration ist.

Schließlich analysieren wir die Dynamik interagierender Zellen. Um Zell-Zell Interaktionen zu inferieren, entwickeln wir die *Underdamped Langevin Inference*, eine Inferenzmethode für stochastische hochdimensionale und interagierende Systeme. Wir wenden diese Methode auf Daten von eingeschlossenen Zellpaaren an, welche wiederholt miteinander kollidieren. Dies zeigt, dass gesunde (MCF10A) und krebsartige (MDA-MB-231) Zellen unterschiedliche Interaktionen aufweisen: Während gesunde Zellen mit Abstoßung und effektiver Reibung interagieren, zeigen Krebszellen Anziehung und eine überraschende 'Anti-Reibung'. Diese Interaktionen führen dazu, dass gesunde Zellen nach Kollisionen primär umkehren, während Krebszellen effizient aneinander vorbeigleiten. Darüberhinaus analysieren wir die Effekte von Cadherin-basierten Molekularkontakten auf Zell-Zell Interaktionen in kollektiver Migration.

Zusammenfassend könnten die in dieser Arbeit präsentierten datengetriebenen Ansätze dabei helfen, ein besseres Verständnis der emergenten stochastischen Dynamik migrierender Zellen zu erlangen. Wir zeigen, wie diese Methoden wichtige Erkenntnisse sowohl über die zugrundeliegenden Mechanismen als auch über das emergente Zellverhalten liefern können.

Synopsis

Cell migration is critical in many physiological phenomena, including embryogenesis, immune response, and cancer. In all these processes, cells face a common physical challenge: they navigate confining extra-cellular environments, in which they squeeze through thin constrictions. The motion of cells is powered by a complex machinery whose molecular basis is increasingly well understood. However, a quantitative understanding of the functional cell behaviours that emerge at the cellular scale remains elusive. This raises a central question, which acts as a common thread throughout the projects in this thesis: do migrating cells exhibit emergent dynamical 'laws' that describe their behavioural dynamics in confining environments?

To address this question, we develop data-driven approaches to infer the dynamics of migrating cells directly from experimental data. We study the migration of cells in artificial confinements featuring a thin constriction across which cells repeatedly squeeze. From the experimental cell trajectories, we infer an equation of cell motion, which decomposes the dynamics into deterministic and stochastic contributions. This approach reveals that cells deterministically drive themselves into the thin constriction, which is in contrast to the intuition that constrictions act as effective barriers. This active driving leads to intricate non-linear dynamics that are poised close to a bifurcation between a bistable system and a limit cycle oscillator. We further generalize this data-driven framework to detect and characterize the variance of migration behaviour within a cell population and to investigate how cells respond to varying confinement size, shape, and orientation.

We next investigate the mechanistic basis of these dynamics. Cell migration relies on the concerted dynamics of several cellular components, including cell protrusions and adhesive connections to the environment. Based on the experimental data, we systematically constrain a mechanistic model for confined cell migration. This model indicates that the observed deterministic driving is a consequence of the combined effects of the variable adhesiveness of the environment and a self-reinforcement of cell polarity in response to thin constrictions. These results suggest polarity feedback adaptation as a key mechanism in confined cell migration.

Finally, we investigate the dynamics of interacting cells. To enable inference of cell-cell interactions, we develop *Underdamped Langevin Inference*, an inference method for stochastic high-dimensional and interacting systems. We apply this method to experiments of confined pairs of cells, which repeatedly collide with one another. This reveals that non-cancerous (MCF10A) and cancerous (MDA-MB-231) cells exhibit distinct interactions: while the non-cancerous cells exhibit repulsion and effective friction, the cancerous cells exhibit attraction and a surprising 'anti-friction' interaction. These interactions lead to non-cancerous cells predominantly reversing upon collision, while the cancer cells are able to efficiently move past one another by relative sliding. Furthermore, we investigate the effects of cadherin-mediated molecular contacts on cell-cell interactions in collective migration.

Taken together, the data-driven approaches presented in this thesis may help to provide a new avenue to uncover the emergent laws governing the stochastic dynamics of migrating cells. We demonstrate how these approaches can provide key insights both into underlying mechanisms as well as emergent cell behaviours at larger scales.

Contents

List of publications

Publications directly relevant to this thesis (chronological order)

1. *Stochastic nonlinear dynamics of confined cell migration in two-state systems*
   D. B. Brückner*, A. Fink*, C. Schreiber, P. J. F. Röttgermann, J. O. Rädler and C. P. Broedersz
   Nature Physics 15, 595-601 (2019) [1]

2. *Area and geometry dependence of cell migration in asymmetric two-state micropatterns*
   A. Fink, D. B. Brückner, C. Schreiber, P. J. F. Röttgermann, C. P. Broedersz and J. O. Rädler
   Biophysical Journal 110, 1886-1895 (2020) [2]

3. *Disentangling the behavioral variability of confined cell migration*
   D. B. Brückner, A. Fink, J. O. Rädler and C. P. Broedersz
   Journal of the Royal Society Interface 17, 20190689 (2020) [3]

4. *Inferring the dynamics of underdamped stochastic systems*
   D. B. Brückner*, P. Ronceray* and C. P. Broedersz
   Physical Review Letters 125, 058103 (2020) [4]

5. *Learning the dynamics of cell-cell interactions in confined cell migration*
   D. B. Brückner, N. Arlt, A. Fink, P. Ronceray, J. O. Rädler and C. P. Broedersz
   Proceedings of the National Academy of Sciences 118, e2016602118 (2021) [5]

6. *Disentangling cadherin-mediated cell-cell interactions in collective cancer cell migration*
   T. Zisis*, D. B. Brückner*, T. Brandstätter, J. d'Alessandro, A. Vollmar, C. P. Broedersz and S. Zahler
   submitted for publication.

7. *Theory of protrusion and polarity dynamics in confined cell migration*
   D. B. Brückner, M. Schmitt, A. Fink, J. Flommersfeld, N. Arlt, E. Hannezo, J. O. Rädler and C. P. Broedersz
   manuscript in preparation for publication.

Other publications (chronological order)

8. *Guiding 3D cell migration in deformed synthetic hydrogel microstructures*
M. Dietrich, H. Le Roy, D. B. Brückner, H. Engelke, R. Zantl, J. O. Rädler and C. P. Broedersz
Soft Matter 14, 2816-2826 (2018) [6]

9. *Thermophoretic forces on a mesoscopic scale*
J. Burelbach, D. B. Brückner, D. Frenkel and E. Eiser
Soft Matter 14, 7446-7454 (2018) [7]

10. *Morphology and motility of cells on soft substrates*
A. Goychuk, D. B. Brückner, A. W. Holle, J. P. Spatz, C. P. Broedersz and E. Frey
arXiv: 1808.00314 (2018) [8]

11. *Autonomous microscale soft robotics based entirely on proteins*
H. Jia, J. Flommersfeld, M. Heymann, S. K. Vogel, H. G. Franquelim, D. B. Brückner, H. Eto, C. P. Broedersz and P. Schwille
submitted for publication.

12. *From collections of individual, independent, mindless robots to flexible, mobile and directional superstructures*
J. F. Boudet, J. Lintuvuori, C. Lacouture, T. Barois, A. Deblais, D. B. Brückner, J. C. Baret and H. Kellay
submitted for publication.

*equal contribution

Chapter 1

Introduction

Cell migration has rapidly become one of the most exciting fields in biophysics. The migration of cells is multi-faceted and deeply intertwined with diverse biological processes and signalling pathways. During development, migrating cells play an important role in ensuring that the correct tissue organization is achieved [9, 10]. In adult tissue, most cells adopt a stationary fate. Yet, some cells retain the ability to migrate, which allows immune cells to track down pathogens [11] and fibroblasts to close wounds [12–14]. Even seemingly non-migratory epithelial tissues have recently been shown to actively migrate, which drives epithelial turnover in the intestine [15]. On the flipside, the switch to a migratory phenotype is also exploited by cancer cells when they leave a tumor to invade surrounding tissue during metastasis [16–18]. All of these migratory processes involve a diverse variety of complex signalling pathways, allowing cells to respond to sensory inputs and adapt their migration behaviour [19–21]. These inputs can be of biochemical or mechanical nature, and provide cross-talk between cells and their extra-cellular environment, as well as from cell to cell. Yet, much of this machinery is also universal: the migration strategy of eukaryotic cells relies on actin flows, myosin contraction, and biochemical polarity cues [22, 23]. While the integration of these building blocks into the biological system at hand may be complex, they fulfil common functions and are highly conserved across organisms and tissues [24]. This apparent universality has caught the attention of physicists attempting to develop quantitative theoretical approaches to elucidate the principles governing cell migration.

Physics approaches to cell biology traditionally focus on those system properties that are clearly physical. Mechanics has been used to quantify and understand the material properties of the cellular cytoskeleton [25–27]; the theory of active hydrodynamics has been applied to the cytoskeletal dynamics of the actomyosin cortex [28–30]; reaction-diffusion frameworks have revealed the physical basis of protein pattern formation [31]; and statistical physics and stochastic thermodynamics have elucidated how molecular motors operate [32] and drive biological systems out of equilibrium at the cellular scale [33, 34]. In all of these examples, physics has had remarkable success in uncovering the rules governing biological machineries. However, at the cellular scale, the collective operation of these machineries gives rise to emergent, functional cell behaviours. An exciting perspective is therefore whether physics can go beyond explaining the physical components of cellular systems and provide frameworks to describe the emergent behaviour of cells as a whole. This raises a central question: are there simple, physical 'laws' that describe the behavioural dynamics of cells, and how do these behaviours adapt to external stimuli placed by their environments?

In the last two decades, tremendous progress has been made in this respect in the con-

text of collective cell migration comprising hundreds or thousands of cells. These systems operate on length scales that allow coarse-graining of the precise single-cell dynamics. This has allowed the application of hydrodynamic models [35, 36], active particle models with effective self-propulsion and interactions [37], and continuum mechanics models [38, 39] to collective cell migration. A key challenge in applying these approaches at the single-cell scale is that the length- and time-scales of interest are comparable to the length- and time-scales of the complex intra-cellular polarity-forming processes. This complicates the search for simple physical laws governing these behaviours due to two main reasons. Firstly, these intra-cellular processes are subject to intrinsic noise, meaning that behaviours at the single-cell level are inherently stochastic. This means that to identify simple laws governing cell behaviour, we must find ways to disentangle the deterministic (i.e. average or stereotypical) and stochastic (i.e. random) components of the dynamics. Secondly, these processes are subject to sensory inputs, causing cells to respond to their environment through complex regulatory networks [19–21]. As a consequence, cells may adapt their behaviours in different environments. Thus, to describe the behaviour of single cells, we need to go beyond soft matter, hydrodynamical, and mechanical models of cell migration. We need a 'Physics of Cell Behaviour', to provide principled system-level quantitative frameworks describing the stochastic dynamics of cellular behaviours.

The development of such quantitative frameworks for cell behaviour may help elucidate a number of critical open problems in the cell migration field: When cells navigate complex and confining environments, how does their dynamics adapt to the local surroundings? More specifically, how do cellular dynamics change when they are confined within a defined geometry? These emergent long time-scale behaviours of cells rely on underlying migratory mechanisms − so how can we connect these different levels of description quantitatively? How variable are migration behaviours within a population of cells, and what is the functional relevance of such variability? Going up from the single-cell level, how can we describe the stochastic interacting dynamics of colliding cell pairs? How do cell-cell adhesions and signalling molecules control the stereotypical collision behaviours exhibited by interacting cells? Can the collective dynamics of large groups of cells be described by the sum of all these interactions, or is there more to it?

In recent years, a number of studies have started applying ideas from the theory of stochastic processes, dynamical systems theory and statistical inference methods to answer some of these questions. A common theme across many of these works is a data-driven approach, where experimental data is used to rigorously constrain possible model candidates. However, a comprehensive overview over this emerging field is thus far lacking. In this introductory section, we will review recent developments of quantitative dynamical frameworks for the stochastic behaviour of single cell migration as well as interacting cells, and highlight how the work presented in this thesis connects to these approaches.

First, we will define what we mean by cell behaviour, and how we envision cell behaviour frameworks could be constructed in general (section 1.1). In section 1.2, we will review data-driven approaches to describe the simplest, archetypal migration problem: the stochastic dynamics of freely migrating cells on 2D substrates. Next, we will discuss how such approaches could be generalized to cell migration in structured extra-cellular environments (section 1.3), and systems exhibiting cell-to-cell variability (section 1.4). In section 1.5, we will provide a perspective on how data-driven approaches to emergent cell dynamics can be connected to underlying molecular mechanisms. Zooming out from the single cell level, we will then

review data-driven approaches to describe the interactions between cells. Finally, we provide an outlook on the open questions in the field and how data-driven methods could be used to address them.

1.1 Cell migration at the behavioural level

In multi-cellular organisms, individual cells migrate to execute functional tasks. Thus, cells are programmed to perform certain behaviours, which include net motion (migration), changes in cell shape (morphodynamics), exerting forces on the extra-cellular environment (traction forces), adaptation to external signals (stimulus response), or the degradation of surrounding matrix polymers (proteolysis). What all these examples have in common is that they are performed at the scale of the whole cell and often take place on long time-scales. Here, we refer to 'long time-scales' as those time intervals which are long compared to typical time-scales of molecular processes, for example the diffusion of signalling molecules across the cell. On these time- and length-scales, cellular behaviour emerges as a consequence of a large number of intra-cellular processes happening instantaneously. Therefore, a completely molecularly reductionistic approach may not be able to capture cell behaviours: precise knowledge of one particular signalling process and all its associated parameters may not be predictive for the emergent behaviour, as it integrates many processes, knowing all of which becomes unfeasible.

To circumvent this problem, minimal mechanistic models are often employed, which seek to identify the key mechanisms at play, and integrate them into a physical model. Such approaches typically postulate specific mechanisms and explore their implications, and we refer to them as *bottom-up* or mechanistic models in this review. Examples are computational models, including phase-field [40–42] and Cellular Potts models [8, 43, 44]. More minimal models include molecular clutch models [45, 46], which account for the coupling of adhesions to substrate mechanics, as well as active gel theories [47, 48] and models coupling actin flow, polarity cues, and focal adhesion dynamics [22, 23, 49–52]. However, these approaches suffer from two key limitations: computational models have many parameters that are difficult to constrain based on experimental data, and thus testing their predictive power for particular experiments remains challenging. The more minimal models are frequently tayloured to capture a particular aspect of the data, but it has often remained difficult to capture the full long time-scale dynamics of the cells, or how these dynamics adapt to external inputs.

An alternative to the mechanistic approach are data-driven *top-down* approaches, which systematically constrain model candidates using experimental data. These approaches provide a phenomenological description at the system-level, which allows coarse-graining over the molecular detail. In physics, phenomenological theories can often be constrained by known symmetries and conservation laws. These system-level descriptions of emergent phenomena can generate conceptual understanding that remains elusive in the reductionist approach, an idea that was famously articulated by Phil Anderson in his essay 'More is different' [53]. Phenomenological approaches are thus a promising avenue to develop quantitative frameworks for cell behaviour. We argue that these approaches should have the following properties:

- **Data-driven**: Fundamental principles that could be used to constrain theories of cell behaviour still remain elusive. Thus, for these frameworks to be viable, they need to directly connect to experimental data. This can be achieved by employing data-driven inference techniques, which rigorously constrain postulated models using input data.

- **Unbiased**: A central idea in data-driven approaches is that they should be agnostic with respect to the underlying molecular or mechanistic basis of the behaviour. Thus, a general model class is proposed, based e.g. on symmetry or causality arguments, but it should be constrained by experimental data in a principled manner, rather than using pre-conceived intuition. The hope is that an unbiased approach can yield a more general description of cell behaviours, which could then be used to systematically constrain mechanistic, bottom-up frameworks.
- **Predictive**: While a given model may be constrained using data, it should then also be able to predict new observations beyond the data that were used to constrain it. Tests of predictive power have two distinct roles: firstly, making predictions for the same experimental data set used to constrain the model, but for statistics that were not explicitly used in the inference, allows testing whether the model provides a meaningful representation of the cellular behaviour. Secondly, performing predictions for new experiments tests the usefulness of the model to provide a generalizable substrate for new systems.

Data-driven, unbiased approaches have a long history in biological systems at larger scales, where the separation between mechanism and phenomenology is much larger than in cell migration, for instance animal behaviour [54–57] or neural systems [58–61]. In recent years, due to advances in tracking and imaging technology, the diversity, accuracy, dimensionality, and size of these behavioural data-sets across the disciplines is rapidly increasing. This has led to the development of data-driven methods to rigorously infer the underlying dynamics of complex systems directly from experimental data, including stochastic trajectories [4, 57, 62–67], spatially extended fields [68, 69], and morphological dynamics [54–56]. We argue that applying principled inference approaches in the field of cell behaviour can make important contributions in answering key open questions in several ways:

- Owing to the intrinsic stochasticity and variability of cell behaviours, another key challenge is to identify what constitutes a 'typical' behaviour. Data-driven approaches could provide **analysis tools** for unbiased, quantitative characterization, classification, and observation of cellular behaviours. Examples for cellular readouts are cell persistence in 2D migration [22, 70], transition times and occupancy probabilities in confined migration [1, 2], movement biases in directional migration [71], and collision outcomes in cell-cell interactions [5, 72, 73].
- Due to the emergent nature and underlying complexity of cell behaviours, it is often unclear what the right quantitative concepts are to describe a particular observed behaviour. Data-driven approaches could yield **conceptual frameworks** to think about cell behaviours by identifying underlying quantitative concepts that can be used to describe cell dynamics. Examples for such concepts in the context of freely migrating cells on 2D substrates are the persistent random motion model [70, 74, 75], Lévy flights [76], and intermittent dynamics [22]. We will review these models, how to constrain them, and their biological implications, in section 1.2.
- Phenomenological models which are constrained in an unbiased and data-driven manner could furthermore yield strong **constraints for bottom-up models** for the underlying

mechanistic basis of the behaviour. These mechanistic models come in different flavours, from minimal mechanical models to active polar gel theories and complex computational implementations. A central difficulty in connecting these models to experiments is that they are frequently under-constrained and over-parameterized. Phenomenological descriptions could provide much more precise 'targets' for mechanistic approaches by introducing stronger constraints. Furthermore, they could be used to test conceptual modelling assumptions or approximations, and thus give insight into the key biological processes in a given system. We will discuss this connection in section 1.5.

- Finally, data-driven frameworks may provide a **generalizable substrate** to develop theories for new systems. For example, to describe the dynamics of interacting cells, it may be useful to have a theory for the dynamics of single migration cells. Can the interacting dynamics then be described by simply adding interaction terms to the single cell model? And can the dynamics of large groups of cells be captured by simply taking into account two-body interactions of pair-wise cell collisions? We will discuss these questions in section 1.6.

1.2 Learning equations of cell motion from data

The simplest possible experiment that could teach us something about cell migration behaviour is the motion of isolated single cells on a uniform two-dimensional (2D) substrate. This is of course not a common setting in physiological processes, in which cells encounter heterogeneous, confining three-dimensional (3D) environments − yet it is the archetypal cell migration experiment that has taught us much of what we know about migrating cells. We will turn our attention towards the description of systems that include spatial structures in the next section. Here, we will review what we have learnt from 2D cell migration, and how this may provide a generalizable basis to describe more complex systems.

Even in the a simple environment of a uniform 2D substrate, the migration of single cells is powered by a complex cytoskeletal machinery. A simple way to quantify the dynamics of this process is a reduction to a single degree of freedom: the position of the cell as a function of time, i.e. its trajectory $\mathbf{x}(t)$. The first cell tracking experiments were performed over a century ago [77, 78] (see ref. [75] for an excellent historical review). At this level, all other degrees of freedom, such as the cell shape, cytoskeletal organization, and traction forces, remain unobserved. The trajectory of the cell is thus a minimal representation of a behaviour: it is observed at the cellular scale, and over long times-periods compared to the time-scales of the internal dynamics. The underlying migratory processes give rise to a mix of deterministic trends, visible as persistent segments, and seemingly random, stochastic components. Accordingly, the mean-square-displacement (MSD) exhibits the signatures of ballistic motion at short time-scales and diffusive motion at long time-scales, and is well described by the formula [74, 78]

$$\langle [\mathbf{x}(t) - \mathbf{x}(0)]^2 \rangle = A(t/\tau_\mathrm{p} + e^{-t/\tau_\mathrm{p}} - 1) \tag{1.1}$$

By measuring the MSD, one can therefore recover two key parameters that characterize the behaviour: the persistence time τ_p, which quantifies the time over which correlations in the cell velocity decay, and the diffusion coefficient $D = A/2d\tau_\mathrm{p}$, where d is the dimensionality.

These parameters are frequently used to quantify cell migration, for example to determine the effect of pharmacological treatments of cells, or to contrast different cell types.

Yet, the trajectories $\mathbf{x}(t)$ give access to so much more information than just the MSD. Specifically, based on the cell trajectories, we can estimate the increments $\Delta\mathbf{x} = \mathbf{x}(t_1) - \mathbf{x}(t_2)$ at various time-scales, including the instantaneous velocities and accelerations of the cell. How can we think about the statistics provided by this additional short time-scale information?

A natural way to think about cell trajectories from a mathematical perspective is the framework of stochastic equations of motion, an idea that was pioneered in ref. [70]. A simple model that predicts an MSD of the form of Eq. (1.1) is an equation of motion for the cell velocity $\mathbf{v} = \mathrm{d}\mathbf{x}/\mathrm{d}t$:

$$\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = -\frac{1}{\tau_\mathrm{p}}\mathbf{v} + \sigma\boldsymbol{\eta}(t) \tag{1.2}$$

This equation of motion thus predicts the cell acceleration as a function of its velocity. It consists of two components: a deterministic contribution (first term on the right-hand side), which accounts for the cell persistestence, and a Gaussian white noise term (second term on the right-hand side), which accounts for the stochasticity of the motion. This equation predicts the MSD in Eq. (1.1) with $A = 2\sigma^2\tau_\mathrm{p}^2$. However, Eq. (1.2) also predicts many other features of the trajectory dynamics. Specifically, it predicts a Gaussian steady state probability distribution of velocities $p(\mathbf{v})$ with a variance $\tau_\mathrm{p}\sigma^2/2$, and a velocity auto-correlation function $\langle\mathbf{v}(t)\mathbf{v}(t')\rangle$ which decays as a single exponential with a time-scale τ_p. Furthermore, it makes a specific prediction about the *conditional average* of the observed cellular accelerations:

$$\left\langle \frac{\Delta v}{\Delta t}\middle| v \right\rangle \sim -\frac{1}{\tau_\mathrm{p}} v. \tag{1.3}$$

This conditional average corresponds to the average of the instantaneous acceleration for each observed instantaneous velocity.[1] These additional statistics beyond the MSD can thus be used to systematically constrain models for 2D cell migration in a data-driven manner. For example, calculating the conditional average on the left-hand side in Eq. (1.3) can constrain the deterministic term of the description: in principle, the dependence of acceleration on velocity could be non-linear, and this analysis would reveal this effect in a model-independent manner. Similarly, the magnitude of the stochastic noise term σ can be inferred from the fluctuations in the trajectories.

Such a data-driven approach was performed by Selmeczi et al. [70], and crucial differences to the standard persistent random motion framework (Eq. (1.2)) were identified: both keratinocytes and fibroblasts exhibit a double-exponential velocity auto-correlation function. Using the conditional average of the acceleration as a constraint, Selmeczi et al. determined the simplest model consistent with all the observed statistics, which exhibits an additional memory term in the velocities and a noise amplitude that grows with increasing speeds. Similar results were subsequently also found in 2D migration of the amoeba Dictyostelium [80–82] and breast cancer cells [1]. Alternatives to this persistent random motion framework also exist in the literature, including Lévy walk models [76], fractional diffusion equations [83], and switching between modes of movement [22, 84].

[1] Note that we do not write an exact equality as the conditional average gives rise to a bias term if discrete estimators are used for acceleration and velocities [79], as is the case when using discretely sampled experimental trajectories. This bias can be avoided by using the more rigorous inference techniques developed in refs. [4, 67] (see chapter 7 of this thesis).

The most remarkable feature of the persistent random motion model and other equations of cell motion is the drastic reduction in complexity achieved. Small and fast dynamics of the cell membrane appear as noise, and only a small number of parameters are necessary to accurately capture cell motion at the level of trajectories. Rigorously constraining these models based on experimental trajectories has also helped to turn persistent random cell motion from a concept into a theory. This provides a basis for connecting the terms of Eq. (1.2) to biological processes, which we will discuss in section 1.5. The data-driven inference of the simple 2D equation of cell motion (formulated Eq. (1.3)) may seem unimpressive. Yet, we argue that conditional averaging is a key technique whose full power is unleashed when used with rigorous inference methods [4, 64–67] and on complex data sets. We are by no means constrained to condition cell accelerations on velocities: what if the migration takes place in a complex structured environment? Suddenly, other degrees of freedom can be used as conditioning variables, such as the cell position. We can therefore infer how cellular responses (measured in accelerations) depend upon the local geometry or structure of the environment (measured by position). We will discuss such an approach in the next section. Furthermore, we can imagine tracking other degrees of freedom of the cell than just its position, for example protrusions and retractions, or even spatially extended variables such as shape or internal concentration fields. Rigorously deriving the equations of motion of these degrees of freedom, and their coupling to each other and to the environment could yield key insights into cell behaviour. This approach could provide more direct connections with mechanistic models, an idea which we discuss in section 1.5 of the introduction. Finally, new inference techniques allow inference also in high-dimensional and interacting systems [4, 66], which could be used to learn the dynamics of interacting cells in collective migration. We discuss this idea in section 1.6 of the introduction. The data-driven persistent random motion framework introduced in this section establishes a conceptual basis to understand these other approaches.

1.3 Cell migration in structured environments

Cell migration on unstructured 2D substrates provides an important benchmark for how to think about cell migration dynamics, and its simplicity has allowed significant theoretical progress. However, in physiological processes, cells do not encounter such artificial environments: the navigate complex, structured, and confining extra-cellular environments, including collagen matrices, bone marrow, or blood vessel linings [16]. Thus, if we want to understand cellular dynamics in physiological processes, we need to study confined cell migration. Restricting ourselves to free 2D migration risks studying cell behaviour in an artificial environment to which it has not adapted, and thus potentially missing key insights into how cells orchestrate their motion. To understand how cell behaviours respond to their external surroundings in a quantitative manner, it is therefore important to develop a data-driven quantitative framework for confined cell migration.

Cell migration in 3D extra-cellular matrices has been studied extensively (see reviews in refs. [85–87]). These matrices are spatially heterogeneous, and therefore single cells will only rarely encounter the same obstacle twice. Consequently, it is difficult to gather statistics on cellular interactions with their local microenvironment. Data-driven approaches however require large data sets to constrain model candidates, owing to the large variability in behaviours. To overcome the limitations of 3D matrices, cell migration is therefore frequently studied in standardized micro-fabricated environments. These include 2D

micropatterns [88, 89], microfluidic devices [90], 3D confinements [91–93], and suspended nanofibers [94–96]. These systems allow monitoring of large numbers of cells migrating in identical, standardized structured environments, yielding unprecedentedly large data sets on cell behaviour.

Micropatterns provide a simple way to confine cells, using differential surface coatings that define areas to which cells can adhere, surrounded by cell-repellent regions. With this technique, confinement regions with arbitrary geometrical shapes can be produced, giving access to a wide variety of systems. One of the simplest migration experiments using micropatterns is confinement to narrow lines [97]. In such effective one-dimensional confinements, cells typically perform persistent random motion in one dimension [98]. This one-dimensional mode of migration has been proposed as a good model for aspects of cell migration in 3D extracellular matrices: in 3D matrices, cells frequently encounter narrow channels through which they migrate, reminiscent of an effective one-dimensional (1D) confinement [95, 98, 99]. Indeed, the morphology of cells on narrow 1D lines is highly stretched, similar to morphologies observed in 3D, which do not feature the broad fan-like lamellipodia observed on 2D substrates [95, 99, 100]. However, physiological extra-cellular environments are also structured, meaning that studies using straight 1D lines may be missing important aspects of cell motility. An important feature of 3D matrices is the presence of thin constrictions through which cells need to squeeze during migration [18, 101–103]. Thus, studying migration in structured micro-confinements may give insights into how cells respond to such obstacles, which could relate to migration in 3D matrices, but also reveal important features of the intra-cellular processes driving migration.

A first step from straight 1D lines to structured environments is a line with periodic modulations, or gaps which cells need to overcome. In this spirit, ratchet-like confinement geometries were designed which were found to rectify the direction of motion of cells [71, 104–106], a process termed ratchetaxis (see ref. [107] for a review). Using a microfluidic confinement with walls featuring similar modulations, a novel mode of migration relying on friction with the local topography of the walls was revealed [93]. Another approach has been to study how cells make decisions at junctions featuring several thin constrictions of varying widths, which revealed the intra-cellular processes involved in cellular decision making in such systems [91, 92]. These experimental approaches using standardized confinements have given insight into intra-cellular processes [91, 92] and have yielded quantitative cellular readouts, for example the degree of directionality in ratchetaxis [71], or switching rates between run and rest states on 1D lines [108]. However, a dynamical theory for the stochastic dynamics of migrating cells in confining environments, akin to the persistent random motion model in 2D, has remained elusive. In chapter 3 of this thesis, we report on first steps we have made towards the development of such a framework, which we will put into context in the following subsection.

1.3.1 Dynamical systems frameworks for confined cell migration

As a minimal system to study how cells overcome thin constrictions in confining environments, we investigated cell migration in two-state micropatterns. These micropatterns consist of two square adhesive islands connected by a thin adhesive bridge. We find that migrating cells perform repeated stochastic transitions between these two islands, with large variability both over time but also between individual cells (Fig. 1.1a). By monitoring hundreds of cells

migrating in these patterns for up to 50 hours, we generated a large data set of confined cell migration trajectories.

Based on these trajectories, we then developed a generalization of the persistent random motion model (Eq. (1.2)) to the problem of confined migration. An important assumption in the persistent random motion model is the uniformity and isotropicity of space: the cellular dynamics are assumed to be independent of position, and the same in all directions. Clearly, these assumptions are no longer valid in structured systems. We therefore postulate a general formulation of an equation of cell motion for confined migration, in which the dynamics may also depend on the absolute position x of the cell:

$$\frac{\mathrm{d}v}{\mathrm{d}t} = F(x,v) + \sigma(x,v)\eta(t) \tag{1.4}$$

In the case of 2D migration, the form of the dynamical terms in Eq. (1.2) could be hypothesized based on the observed MSD, which coincides with the prediction of the canonical Ornstein-Uhlenbeck process.[2] In contrast, there is no *a priori* physical framework from which we could derive the structure of the dynamical terms in the equation for confined cell migration (Eq. (1.4)). We therefore employed a data-driven method to constrain these terms using the experimental trajectories. Specifically, to a first approximation, the deterministic term of this equation can be inferred using a conditional average of the observed cellular accelerations:

$$F(x,v) \sim \left\langle \frac{\Delta v}{\Delta t} \middle| x, v \right\rangle \tag{1.5}$$

which is the generalized formulation of Eq. (1.3) for an equation of motion with positional dependence. Thus, by inferring $F(x,v)$ from the data, we could determine the average acceleration of the cell as a function of its position x within the structured micropattern and its velocity v (Fig. 1.1b). Developing such an equation of motion model from experimental data in general involves four key steps, which we demonstrate using the example of the confined cell migration problem (Fig. 1.1):

1. **Observation**: in the first step, the important degrees of freedom (DOF) of the system have to be identified and observed. An equation of motion is a mathematical construct which predicts the future state of the system based on its current state, through a differential equation for the relevant DOF. To enable inference and interpretation of the model, this set of DOFs should ideally be low-dimensional and therefore provide a minimal representation of the system. For example, in the case of the confined migrating cells, we first restricted the analysis to the one-dimensional trajectories of the cell nucleus (Fig. 1.1a).

2. **Inference**: the second step in the development of data-driven cell migration models is the model inference. In this step, a general, unbiased formulation of a stochastic dynamical system for the tracked DOFs should be postulated, which can then be systematically constrained using the data. In the case of the confined migrating cell, a general, unbiased approach was to generalize the persistent random motion observed on unconfined substrates by postulating the general equation Eq. (1.4). We therefore make no *a priori* assumptions of what the dynamics of the system, given by $F(x,v)$ and $\sigma(x,v)$, should look like. Instead, we use the experimental trajectories to infer this term

[2] Note, however, that corrections to this simple form were derived in ref. [70], as described in section 1.2.

directly from the data. See Box 1 for a summary how stochastic inference approaches can be executed in practise.

3. **Prediction**: while we use the experimental data to constrain the shape and parameters of the deterministic dynamics $F(x, v)$, there is no guarantee that this approach yields an insightful representation of the dynamics of the system: the inference approach relies on the assumption that the dynamics of the system can in fact be described by the equation of motion Eq. (1.4). This assumption could fail in many ways: the dynamics could require additional memory terms [70], a time-dependent description [109], or an explicit description of the cell-to-cell variability [110]. Thus, to test the validity of this decription, we need to perform a test of predictive power. Specifically, to perform the inference, we constrained the equation of motion solely based on the short time-scale information provided by the experimental trajectories, including the velocities and accelerations of the cell. Thus, as an independent test of the model [1, 70], we predict statistics quantifying the cell behaviour on long time-scales, for example the distribution of transition times or the velocity auto-correlation function (Fig. 1.1c).

4. **Interpretation**: having determined a valid model for the observed dynamics, this model can be interpreted to gain insight into the system. This last step is of course very much system-dependent. An important aspect of the stochastic inference approach is the decomposition of the dynamics into deterministic and stochastic components, i.e. F and σ. Based on this decomposition, these components can be interpreted separately, and their respective contributions to the dynamics can be conceptualized. We will briefly summarize below how this approach provided a conceptual framework for the confined cell migration problem.

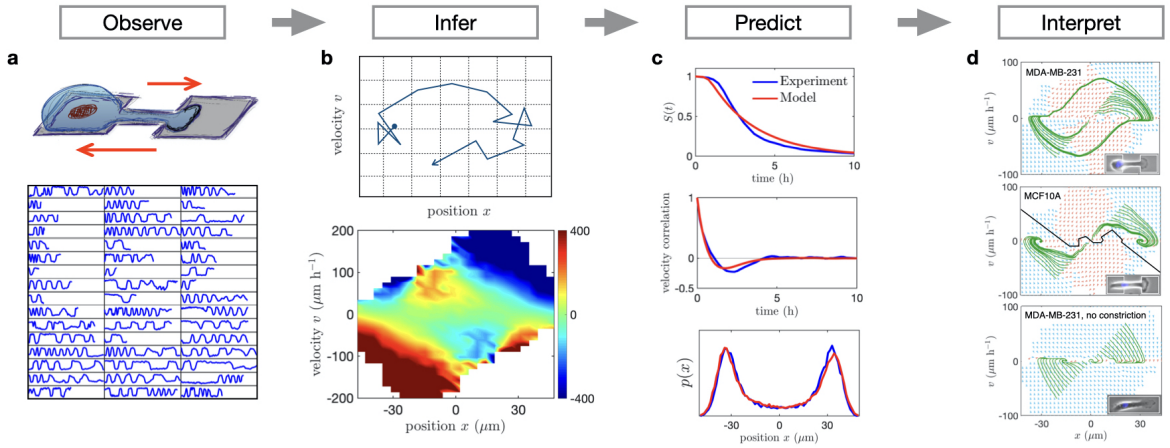


Figure 1.1: **Data-driven approach for confined migration. a.** Example data set of trajectories of the cell nucleus position of MDA-MB-231 cells migrating in a two-state micropattern (sketched). Trajectories are plotted as a function of time up to 50h. **b.** Top: sketch of a trajectory in xv-phase space with grid-based binning (dotted lines). Bottom: inferred deterministic term $F(x, v)$. **c.** Experimental (blue) and predicted (red) statistics, from top to bottom: survival probability distribution of states $S(t)$, defined as the probability that a cells has *not* made a transition after time t; velocity auto-correlation function; probability distribution of positions. **d.** Phase portraits obtained for various systems, from top to bottom: MDA-MB-231 cells migrating in a two-state pattern; MCF10A cells migrating in a two-state pattern; MDA-MB-231 cells migrating on confining strips with no constriction. Arrows show flow fields, including accelerating (orange) and decelerating (blue) components. Green lines are deterministic trajectories of the system.

In the example of the confined cell problem, we found that an insightful representation of the system can be achieved by examining the deterministic dynamics of the system in a phase-portrait of position and velocity. Intuitively, one might expect that the hopping behaviour across the thin constriction placed by the micropattern might be generated by a noisy cellular activity competing with an effective energy barrier placed by the constriction. Strikingly, however, the inferred map of the deterministic accelerations (Fig. 1.1b) reveals that cells have a tendency to deterministically accelerate into the constriction. In fact, the flow field of the deterministic dynamics exhibits an excitable flow, where a small noise-driven perturbation leads to a large excursion in the phase space due to a deterministic amplification of the cell speed (Fig. 1.1d). This amplification is observed in both cancerous (MDA-MB-231) and non-cancerous (MCF10A) cells, suggesting that it may be a generic cellular response to thin constrictions. Indeed, in systems in which the constriction is removed, the amplification vanishes (Fig. 1.1d). This approach also reveals that the non-linear dynamics are poised close to a bifurcation between a limit cycle and a bistable system. Interestingly, different cell lines exhibit behaviours on both sides of this transition: MDA-MB-231 cells exhibit a limit cycle, while MCF10A cells show excitable bistable dynamics. Thus, the deterministic phase-portrait implies that the cancerous cells have a stronger tendency to overcome the constriction, while the non-cancerous cells rely on stochastic fluctuations to perform transitions. In summary, this data-driven approach to confined cell migration provides a novel dynamical systems perspective on cell migration dynamics. In the next section, we will discuss how we can use these insights to quantify and characterize the striking variability in the observed cell behaviours, which are already apparent at the level of the cell trajectories (Fig. 1.1a). Moreover, this approach could help advance our understanding of locomotion at the molecular level by providing constraints for bottom-up models that connect microscopic rules to the system-level dynamics of cells. Finally, the insights gained based on this framework could provide a generalizable substrate to investigate the dynamics of assemblies of interacting cells. We will discuss both of these aspects in the following sections.

Box 1: Inference from stochastic trajectories

Here, we follow the general formalism proposed by Frishman and Ronceray [66]. Suppose we have tracked a set of N cellular DOFs, such as the position of the nucleus, which we denote by $\mathbf{q}(t) = \{q_\mu(t)\}$ where $\mu = 1...N$. Then a general dynamical system that could govern the dynamics of these DOFs is given by the stochastic equations of motion

$$\dot{q}_\mu = F_\mu(\mathbf{q}) + \sigma_{\mu\nu}(\mathbf{q})\eta_\nu(t) \tag{1.6}$$

where $\eta_\nu(t)$ is Gaussian white noise with $\langle \eta_\mu(t) \rangle = 0$, $\langle \eta_\mu(t)\eta_\nu(t') \rangle = \delta_{\mu\nu}\delta(t-t')$. Our aim is to infer the terms of this equation, F and σ, based on a set of observed experimental trajectories $\mathbf{q}(t)$. To break down this problem, one can express the dynamical terms as a truncated basis expansion

$$F_\mu(\mathbf{q}) \approx \sum_{\alpha=0}^{N_b} F_{\mu\alpha} c_\alpha(\mathbf{q}) \tag{1.7}$$

where $c_\alpha(\mathbf{q})$ are the basis functions used to represent the dynamics. A similar expression can be written for the stochastic term σ. Thus, the problem of inferring the equation of motion has reduced to estimating the parameters $F_{\mu\alpha}$. This can be done by projecting the dynamics onto the basis functions:

$$F_{\mu\alpha} = \langle \dot{q}_\mu c_\alpha(\mathbf{q}) \rangle \tag{1.8}$$

The simplest approach to such an inference problem is to perform a grid-based binning approach where the phase-space is divided into a regular grid of tiles (Fig. 1.1b). In this case, the basis functions $c_\alpha(\mathbf{q})$ become top-hat functions at regularly spaced locations in the phase-space. The parameters $F_{\mu\alpha}$ then correspond to the average velocities of the DOFs, $\dot{q}_\mu$, at that location in phase-space. However, this approach requires a large number of fitting parameters N_b and therefore suffers limitations in high-dimensional systems or experimental data sets with low statistics. Specifically, the inference error grows linearly with the number of parameters N_b [66]:

$$\text{inference error} \propto N_b/\tau \tag{1.9}$$

where τ is the total length of the trajectory. In the development of the equation of confined cell motion, we had access to a large data set of 1D nucleus trajectories (Fig. 1.1a), and thus the binning approach was feasible in this case. A data-efficient alternative to binning is projection onto a set of *smooth* basis functions, such as polynomials. In this case, fewer parameters are required and additional constraints such as symmetries of the system can be taken into account to further restrict the choice of basis functions. This also allows inference from interacting systems (section 1.6), which are inaccessible to binning approaches.

In addition to finite data, inference from experimental trajectories suffers from two key problems:

1. **Measurement errors**: Experimental trajectories are inevitably subject to random localization errors. Since stochastic inference relies on estimating the derivatives $\{\dot{q}_\mu\}$, this can lead to significant inference errors. To overcome this problem, error-corrected estimators can be derived [4, 66]. We develop such estimators in chapter 7 of this thesis.
2. **Discrete observations**: Tracking methods operate at a finite frequency, meaning that observed trajectories are sampled at a finite time interval Δt. This leads to inference errors when the dynamical system at hand is *underdamped* [4, 67, 79], meaning that the relevant DOFs to describe the system include time-derivatives, such as the velocity of the cell nucleus. This is the case in the typical cell migration models, including the persistent random motion model (Eq. (1.2)), and in confined cell motion (Eq. (1.4)). In chapter 7, we will introduce a way to circumvent this problem.

In summary, the presented stochastic inference approach allows us to infer the governing dynamical systems from observed experimental trajectories in an unbiased manner, and in a way that is robust to finite data limitations, measurement noise, and discretization errors.

1.4 Variability in cell behaviour

A key feature of migrating cells is the large variability of the observed behaviours within a cell population. A typical set of cell migration trajectories exhibits large variations between different individual trajectories, but also over time within a single trajectory. Tracing the origins of such variability is an open challenge that can be ideally addressed by data-driven approaches, since it naturally relies on the analysis of large ensembles of observations [111]. In this review, we propose to distinguish four distinct contributions to the behavioural variability of migrating cells:

- **Intrinsic stochasticity**: the intra-cellular machineries driving cell behaviours operate at the molecular level, and are thus subject to intrinsic noise. At larger scales, where these molecular degrees of freedom remain unobserved, this intrinsic noise leads to seemingly random patterns in behaviour, which we refer to as the intrinsic variability of cell behaviour.

- **Cell-to-cell variability**: even in populations of cells with identical genomes, the stochasticity of intra-cellular processes such as gene expression, cytoskeletal rearrangement and protein localization can lead to large differences in the proteomes of individual cells [112–116]. At the cellular scale, this diversity can lead to variations in cell behaviour, which is also referred to as phenotypic or population heterogeneity. The connection from molecular to behavioural heterogeneity has been demonstrated in cellular processes ranging from growth rate and drug response to morphology [117–120], and has been suggested to play an important role in collective cell migration [121, 122].

- **Temporal variability**: the behaviour of cells may also exhibit variations over time: as cells undergo the cell cycle, they grow, which may also affect other behaviours, including cell migration [123].

- **Extrinsic variability**: potentially unobserved changes in the extra-cellular environment may cause changes in behaviour, which could be mistaken for other types of variability. This can occur for example in cell migration in environments with unobserved structures, including porous 3D matrices [110].

Gaining insights into how these distinct contributions determine the overall variability of cell migration could be important for understanding physiological migration processes, as well as the mechanistic basis of the behaviour. However, disentangling these different contributions to the behavioural variability based on an observed data set can pose a formidable challenge. To this end, several data-driven approaches have been developed in the context of cell migration, which we will discuss here.

1.4.1 Detecting cell-to-cell variability in behaviour

Models for cell migration are typically formulated as stochastic processes, which is a natural way to capture processes exhibiting fluctuations. In the equations of cell motion introduced in previous sections (Eqs. (1.2) and (1.4)), the stochastic white noise term ensures that no two trajectories look alike. This is a model for the *intrinsic stochasticity* of the migration process. To determine the structure and parameters of cell migration models, the dynamics are typically averaged across different cells and over time, yielding ensemble- and time-averaged

stochastic models that describe the average member of a cell population. Therefore, these approaches fail to capture cell-to-cell and temporal variability. Similarly, bottom-up models for cell motility typically assume that all cells in a population can be described by a common set of parameters [39].

To develop a framework which can account for temporal and cell-to-cell variability in cell migration, Metzner et al. [109] developed a generalization of the persistent random motion framework (Eq. (1.2)) which allows for variations in the migration parameters. Indeed, previous work has suggested that to fully account for the statistics of cells migrating on 2D substrates, cell-to-cell variability has to be taken into account [110]. In the 'super-statistical' framework introduced by Metzner et al., both the persistence τ_p and the noise amplitude σ become functions of time which differ for each individual cell. The values of the parameters are inferred from experimental trajectories using a Bayesian maximum likelihood approach. With this method, 'run' and 'rest' states of migrating cells, which have been suggested to be due to distinct transient intra-cellular organizations [22], would be rigorously identified.

A second quantitative framework for migration variability was developed by Jordan et al. [124]. This approach is based on determining a characterization of the instantaneous motion, using the joint probability distribution of speed and turning moment, and then quantifying the change in this distribution over time and between individuals. Using clustering and dimensional reduction, this method leads to a low-dimensional behaviour space revealing which components of the cell motion contribute most to the variability. While this framework was thus far only applied to swimming protozoa, it has potential also for cell migration, provided that a sufficient time-resolution can be achieved experimentally. In summary, these works provide computational tools to rigorously characterize cell-to-cell and temporal variability in migration behaviours from trajectories alone.

While these frameworks provide a way to quantify and characterize cell migration variability over time and between individuals, they do not provide a method to determine whether such variability exists in the first place. Indeed, if the observed trajectories are short, as is often the case in cell migration experiments, they may appear variable simply due to the randomness introduced by intrinsic stochasticity. How then can real variability be distinguished from apparent variability due to the intrinsic stochasticity? In other words, how can we test if the data is consistent with ergodicity? This question has previously been raised in the context of collective cell migration [125], where it was suggested to compare the observed variability to an appropriate 'null-model'. Specifically, this means performing a direct comparison of variability-sensitive experimental observables, such as population variances, to the predictions by a parameter-optimized model without variability. Deviations from the variability-free model can then provide an indicator for real variability. A difficulty in applying this approach is that it requires both a large ensemble of migration trajectories in a standardized setting, and an appropriate theoretical framework to provide a null-model.

We have addressed this open problem for the case of confined migrating cells in micropatterns, which is presented in chapter 4. Specifically, we studied the trajectories of cells migrating in two-state micropatterns (Fig. 1.1a). For this system, we established that the ensemble- and time-averaged (ETA) statistics are well captured by an ETA equation of motion model (Eq. (1.4)). Thus, this equation of motion provides a null-model to which we can compare the experiment. We found that the variance in behaviour between individual cells was larger in the experiment than that measured in an ensemble of trajectories of similar length

predicted by the null-model. This indicated that there is real cell-to-cell variability in the system, beyond the random variations expected from a single, ergodic process. Interestingly, our analysis further revealed that within the cell population, there are qualitative differences in the class of dynamical systems describing the migration of individual cells. Specifically, larger cells exhibited limit-cycle dynamics, while slower cells exhibited bistability, with two stable fixed points. The emergence of distinct dynamical systems describing the migration within a population of cells has been rationalized to originate from a heterogeneity in microscopic migration parameters [49]. Specifically, it was suggested that tuning the elasticity and adhesiveness of cells could lead to distinct dynamical behaviours, including smooth migration, stick-slip migration, and bistability between these two modes.

Taken together, these results demonstrate that combining systematic inference tools that account for cell-to-cell variability with mechanistic models could in the future lead to novel insights into the behavioural variability of cell populations. An exciting approach in this respect would be to correlate variability at the molecular scale with variability at the behavioural scale, which could give insight into how molecular organization correlates with behaviour without relying on artificial perturbations of the system [126–129].

1.4.2 Extrinsic heterogeneity

In addition to cell-to-cell variability, migrating cells also encounter variability in their environment including heterogeneous extra-cellular matrices or contact with other cells. Indeed, apparent cell-to-cell variability in collective systems has in many cases been shown to be caused by environmental factors, including local cell density, cell-cell contacts and relative location in a cell cluster [117, 130–132]. An interesting special case of such extrinsic variability is *self-induced extrinsic variability*, where the cell itself causes changes to its environment which in turn affect its behaviour. Here, we discuss two examples of this case in which quantitative frameworks for the migration were developed.

First, in 3D migration through a matrix, some cells perform proteolysis, which is a mechanism that allows cells to locally digest the surrounding matrix to create a migration path. This behaviour was shown to lead to asymmetries in the preferred direction of motion of cells: cells were more likely to turn around by 180° than expected based on persistent random motion, thus backtracking on their previous path [110]. As a model for this process, the *anisotropic persistent random walk* model was proposed, which includes parameters that are non-isotropic in space and thereby account for this effect.

In the second example, it was found that migrating cells deposit material on the surface on which they migrate, causing them to behave differently when they return to a location that they previously visited [133]. Specifically, cells were observed to preferentially occupy previously visited areas. In this work, data-driven inference was used to generalize the phase-space analysis introduced in section 1.3.1 to the problem of self-attracting migration on a 1D line. This approach revealed that cells deterministically accelerate away from the boundaries of previously explored space. This observation motivated a quantitative description using a *persistent self-attracting walk* model, which quantifies the relative probabilities of turning back vs. exploring new areas. This effect leads to long-lived spatial memory in the migration, which can have dramatic consequences for the ways in which cells search and explore space.

In summary, these approaches identified important cell migration mechanisms using data-driven analysis of the migration trajectories which exhibited striking variability. The analysis

revealed that the observed variability is in fact due to extrinsic effects, albeit regulated by the cell itself. These findings are particularly interesting in the broader picture of regulated cell-to-cell variability proposed in [132], where it was suggested that deterministic, regulated variability could have functional importance in cell population, which is in contrast to cell-to-cell variability caused by random fluctuations of intra-cellular processes.

1.5 Connecting dynamics to mechanisms

In the previous sections, we have demonstrated how quantitative frameworks for cell migration can provide data analysis tools and yield conceptual frameworks to think about cell behaviour. A third important contribution such frameworks could make to the field is by providing constraints for mechanistic cell migration models. We refer to models as 'mechanistic' if they are based on a bottom-up approach in which model ingredients are postulated motivated by known cellular processes or physical intuition. This is in contrast to the data-driven, top-down approaches that we have focussed on so far in this review.

There is a long history of mechanistic biophysical modelling of cell migration (see e.g. [134] for a review). Here, we do not aim to provide a complete review of this broad field, but rather aim to highlight how mechanistic approaches could in the future be combined with top-down frameworks. First, we will discuss in more detail how the structure of the equation of cell motion models introduced in the previous sections can be understood in terms of the underlying cellular mechanisms. Secondly, we will summarize specific examples of bottom-up models which bridge the gap between underlying mechanisms and whole-cell behaviours.

1.5.1 Accelerations without force and mass

The equation for 2D persistent random motion (Eq. (1.2)) and the equation of motion for confined cell migration (Eq. (1.4)) share a key feature: both are stochastic differential equations that are second-order in time, and therefore a manifestation of the *underdamped* Langevin equation. Specifically, these equations predict the acceleration as a function of position and velocity. This is in contrast to first-order stochastic equations of motion which are frequently used to describe the motion of Brownian systems subject to thermal noise [135]. For Brownian systems, the effects of inertia can be neglected at time-scales larger than the relaxation time $\sqrt{\zeta/m}$, where ζ is the friction coefficient and m is the mass of the particle. Therefore, friction is directly equated with the sum of thermal and external forces, yielding a first-order, *overdamped* Langevin equation. A similar argument applies to cells: their mass is negligible to the friction − but still, we describe their dynamics with an underdamped equation of motion. Why? In other words: should we think of $F(x, v)$ in Eq. (1.4) as a 'force field'?

The answer is of course no: it is an acceleration field, if it is to match the left-hand side of the equation in dimension. Yet, Eq. (1.4) still has the same overall structure as Newton's second law, where accelerations are equated to forces. And indeed the answer goes deeper than this, as Eq. (1.4) is different to Newton's second law at a more fundamental level. Specifically, Newton's law posits $m\dot{v} = \sum_i F_i$ where $\sum_i F_i$ is the sum of all forces acting on the cell, and m is the inertial mass of the cell. However, in the case of a cell, this equation essentially reduces to $0 = 0$: the inertial mass of the cell is negligible compared to all other factors in the dynamics. Conversely, there is also no total force acting on the cell, $\sum_i F_i = 0$, as the traction forces cells exert on their substrate all balance each other. Therefore, its motion is quasi-stationary when compared to the scale of the traction forces.

Why then do cells accelerate if they have negligible mass and there are no total forces acting on them? These accelerations can be understood as a consequence of the joint action of the intra-cellular migration mechanisms that orchestrate correlations in the cell movement that extend to far longer time-scales than expected for a Brownian particle with comparable mass and friction. The speed of a migrating cell is to a good approximation proportional to the internal flow of actin [22], which is being polymerized at the leading edge, and depolymerized at the trailing edge [136]. The directionality of this actin flow is in turn determined by the concentration profiles of internal signalling cues within the cell, which reorient on long time-scales [23]. Reorientations of these polarity fields lead to changes of the cell velocity, and thus to accelerations. Therefore, cellular accelerations are changes of velocity that are determined by intra-cellular dynamics, and not by a net force acting on the cell. Consequently, the term $F(x, v)$ should not be interpreted as a force term, but as the deterministic contribution to the effective underdamped dynamics of the cell. Therefore, to understand the origin of the emergent cell migration dynamics, quantified by $F(x, v)$, we should consider how internal degrees of freedom of the cell, including the cell shape, protrusion formation and polarity determine the net movement of the cell, and how these degrees of freedom couple to the external environment.

1.5.2 Combining bottom-up and top-down approaches

Having understood the conceptual mechanistic basis of the emergent migratory dynamics of cells, we will discuss in more detail how these dynamics can be connected to mechanisms. Bottom-up biophysical models for cell migration traditionally focussed on particular aspects of the motility machinery, such the ratchet model for force generation by actin polymerization [137, 138], actin branching [139], and the molecular clutch model for adhesion dynamics [45]. However, to describe cell motility at the behavioural level, integrated models which connect the underlying mechanisms of actin flow and polarity dynamics are required. This is the philosophy behind complex computational models, such as Cellular Potts Model implementations that account for cell polarity [44, 140], as well as phase-field models [141, 142]. However, these computational models have many parameters that are difficult to constrain based on experimental data, and thus testing their predictive power and leveraging them for conceptual insight for particular experiments remains challenging. Indeed, ref. [143] showed that several conceptually different computational models with a distinct mechanistic basis are able to faithfully capture the migration of keratocytes. This apparent degeneracy of possible mechanisms may be a consequence of real degeneracy in biological mechanisms responsible for a given behaviour. However, this observation also points to a bigger problem with using complex bottom-up approaches whose parameters remain under-constrained for conceptual insight. In contrast, minimal models have frequently remained simplistic, and unable to capture the emergent stochastic behaviour of cells on long time-scales.

A breakthrough in this regard was made in a landmark study by Maiuri et al. [22], who proposed a mechanism for how the instantaneous actin flow speed determines the speed and persistence of migrating cells. Specifically, for a large number of cell types, it was shown empirically that their persistence is exponentially correlated with their speed, $\tau_{\mathrm{p}} \propto e^{\lambda \langle |v| \rangle}$, termed the universal coupling of cell speed and persistence (UCSP) law. Using a clever experimental approach, the authors then de-coupled the speeds of cell migration and actin flow speed, which are typically proportional to one another, and demonstrated that the persis-

tence is in fact proportional to the actin flow speed itself. This suggests that higher actin flow speed has a stabilizing effect on the motion, thus leading to more persistent trajectories. Using a minimal model coupling actin flow and polarity cue advection-diffusion, this relationship could be theoretically predicted. Importantly, this theoretical approach furthermore made predictions for the complete stochastic trajectory dynamics of freely migrating cells in terms of the microscopic actin-polarity parameters. This approach predicted that depending on cell parameters, migrating cells could exhibit three distinct behavioural regimes: random motion, highly persistent motion, and an intermittent mode exhibiting bistability between the two other regimes. This framework was later extended in refs. [49, 50] to account for the mechano-sensitive binding dynamics of focal adhesions at the trailing edge of cells. This extension predicts additional types of motility featuring periodic extension-retraction cycles, that have been observed as so-called 'stick-slip processes' in the biological literature [51, 144]. In summary, these studies have developed quantitative connection of underlying cellular mechanisms to the emergent cellular behaviours, and given key insights by revealing the mechanistic basis of important migration behaviours.

While these approaches have provided a quantitative connection from the mechanistic to the behavioural level, they have focussed on the dynamics of freely migrating cells on 2D substrates and unstructured 1D lines. Thus, it remained unclear how to describe the coupling of the intra-cellular mechanisms to external environments. In the context of immune cell migration in confinements featuring structured walls, an active gel theory approach has been developed, which relates the friction with the wall to the cell migration speed [93, 145]. However, this mechanism is specific to the amoeboid mode of migration displayed by immune cells.

In chapter 6 of this thesis, we develop an approach to connect a general mechanistic model to the long time-scale stochastic dynamics of confined migrating cells. To elucidate the dynamic interplay of the key cellular components that determine confined migration, including the formation of cell protrusions, adhesive connections to the environment and the positioning of the cell nucleus, we develop a hybrid mechanistic and data-driven theoretical approach. Specifically, we use a large data set of coupled protrusion and nucleus trajectories to systematically constrain a mechanistic model for confined cell migration. Interestingly, we find that cells confined to two-state micropatterns exhibit a stereotypical migration pattern, with protrusions growing to precede the transmigration of the cell nucleus across the constriction, which we term 'protrusion-nucleus cycling'. Interestinlgy, we find that the average dynamics of the cell nucleus in this system are determined by the locally available adhesive area. Furthermore, our model indicates that the protrusion dynamics are driven by a stochastic cell polarity that is sensitive to the local geometry. Specifically, under strong confinement, the polarity dynamics switches from a negative to a positive, self-reinforcing feedback loop. Strikingly, this model predicts, in agreement with experiments, that the protrusion-nucleus cycling disappears when the constriction is removed. This suggests that the positive polarity feedback loop emerges as a consequence of an adaptation of the cellular dynamics to the presence of the thin constriction. Our theoretical approach therefore suggests polarity feedback adaptation as a key mechanism in confined cell migration. Furthermore, we find that this mechanistic model fully captures the emergent dynamics of the nucleus trajectories alone, described by Eq. (1.4). This model further reveals that the observed excitable flow (Fig. 1.1) is a consequence of two combined effects: lower adhesiveness and enhanced polarity persistence in the constriction.

Taken together, these approaches demonstrate how combining bottom-up mechanistic modelling and top-down inference can advance our understanding of cell migration processes, and allow bridging the gap from mechanism to behaviour.

1.6 From the two-body to the many-body problem

In the previous sections, we focused on cell migration experiments where single cells are isolated and only interact with their environment but not with one another. While such simplified experimental systems can teach us much about the mechanistic basis of single cell motility, studying how cells interact with one another is crucial to understand migration in physiological settings. Indeed, in many of the physiological processes in which cell migration is important, cells constantly interact with one another to organize their behaviour [16, 146–148]. Cellular interactions depend on complex molecular mechanisms, including cadherin-dependent pathways and receptor-mediated cell-cell recognition [149–155]. These mechanisms lead to well-defined, stereotypical cell behaviours upon collision. The most prominent type of collision behaviour was discovered in the 1950s by Abercrombie and coworkers [156], and was termed *Contact Inhibition of Locomotion* (CIL). CIL refers to the tendency of cells to retract their lamellipodia, repolarize, and migrate apart upon contact. While these observations were made in a simple cell culture on 2D substrates, the relevance of CIL for physiological processes was later demonstrated in the development of the neural crest [149] (see [150, 157] for reviews).

At larger scales, these two-body interactions of cells lead to coordinated collective migration. At the collective level, the dynamics of groups of cells comprising tens, hundreds or thousands of individuals has been described with a variety of physical modelling approaches. These include active hydrodynamic theories [158], vertex [159, 160], mechanical [38], and mechano-chemical [161] models, cellular automata [44, 162], phase-field models [141, 142], as well as active particle models [163–168] (see [39] for an excellent review contrasting these different types of models). While all these modelling approaches are mechanistically distinct, they share two important features:

- With the exception of hydrodynamic approaches, which constrain models based on generic symmetry arguments, the modelling avenues highlighted above typically make *a priori* assumptions on the types of interactions between individual cells, and therefore classify as **bottom-up** approaches. Cell-cell interactions are frequently modelled using repulsive potentials as an implementation of excluded volume interactions, alignment terms [164–167], or explicit implementations of CIL-like reorientation events upon collision [163, 168]. However, the structure and parameters of these interactions is usually not derived directly from experimental data.

- These approaches have advanced our understanding at the collective level comprising many cells, which allows **coarse-graining over the microscopic features** of the system. Thus, an approximate implementation of the cell-cell interactions is sufficient to capture the key features of the system. For example, alignment interactions in active particle models subsume a number of underlying cellular processes which yield alignment between cells as an effective outcome. Thus, it is a challenge to connect these

approaches all the way from Abercrombie's archetypal two-body problem, with only two cells colliding, to the collective behaviour.

There are two key hurdles that have made the development of data-driven approaches for cell-cell interactions difficult until recently:

- The complexity of the biological settings in which cell-cell interactions take place make it difficult to **disentangle the distinct contributions of single-cell behaviour, interaction with the local micro-environment, cell proliferation, and cell-cell interactions**. To overcome this problem, simplified assays have been developed where small groups of cells are confined to micro-patterned patches [44, 169] or tracks [72, 73, 170–172], microfluidics [90], and suspended fibers [173].
- **The difficulty of the inference problem itself**: treating interacting systems using classical approaches based on grid-based binning of the phase space is unfeasible due to the prohibitively large data sets that would be required to constrain the high-dimensional phase space of the system. In the last few years, interaction inference methods which overcome these problems have been developed for deterministic systems [174, 175] and in the context of collective animal behaviour [176–178]. For stochastic systems, the general approach presented in Box 1 (developed in ref. [66] and chapter 7 of this thesis) can be extended to treat interacting systems by employing basis functions which exploit the (approximate) symmetries of the system, such as particle exchange symmetry or radial symmetry of the interactions. The recent development of these inference methods opens up new avenues to learn the dynamics of cell-cell interactions directly from observed data.

In chapter 8 of this thesis, we develop an approach for interacting cell dynamics which is fully data-driven and describes a system at the level to pair-wise cell interactions. To standardize the micro-environment of the cells, we developed a minimal 'cell collider', which ensures that pairs of cells repeatedly collide with one another. Based on the joint trajectories of the two cells, we sought to develop an interacting equation of motion for the system. Specifically, we postulate that the dynamics of the system can be described by a generalization of the equation of confined cell motion (Eq. 1.4):

$$\frac{\mathrm{d}v}{\mathrm{d}t} = F(x,v) + f(|\Delta x|)\Delta x + \gamma(|\Delta x|)\Delta v + \sigma\eta(t) \tag{1.10}$$

where $\eta(t)$ is Gaussian white noise as before. Thus, we propose that the deterministic dynamics of the system can be decomposed into two separate components: a single-cell term $F(x,v)$, similar to that inferred from single-cell experiments, and interactive components, which depend on the relative position Δx and the relative velocity Δv of the cells. The term $f(|\Delta x|)$ thus represents positional cell-cell interactions such as repulsion and attraction. In contrast, $\gamma(|\Delta x|)\Delta v$ depends on the relative motion of the cells, and has the mathematical form of an effective frictional interaction. For $\gamma < 0$, this interaction accounts for alignment between cells, as it seeks to minimize differences in relative velocity [164, 165].

An interesting problem to consider with this data-driven approach is to ask which types of interacting dynamics are required to describe the distinct types of collision behaviours of various cell types. Specifically, normal tissue cells typically undergo CIL upon collision, meaning that they typically reverse upon collision in a 1D confinement [72, 171]. CIL is

known to be suppressed in cancer cells [151, 179, 180], and therefore, these cells can migrate past each other more easily [73]. Furthermore, some cell types have also been observed to predominantly attach and follow each other upon contact, which was termed *Contact Following Locomotion* [170, 181–183].

In our experiments, we contrasted a non-cancerous (MCF10A) and a cancerous (MDA-MB-231) breast tissue cell line, and found that these cell lines indeed exhibit distinct collision phenotypes: while the non-cancerous cells predominantly reverse upon collision, the cancerous cells tend to slide past one another. Based on the recorded trajectories, we inferred the terms of the interacting equation of motion (Eq. (1.10)). We demonstrated that this inferred model is able to capture the experimentally observed statistics, suggesting that this framework provides an accurate representation of the system's behavioural dynamics. Strikingly, we found that the two cell types are described by distinct types of interactions: while the MCF10A cells exhibit repulsive and regular frictional interactions, the MDA-MB-231 attract at short distances and exhibit a positive friction term ($\gamma > 0$). This 'anti-friction' interaction ensures that rather than slowing down upon collision, cells deterministically accelerate, leading to the characteristic sliding events observed for this cell line. We generalized this inferred theoretical framework by predicting an 'interaction behaviour space', which relates the physical interaction terms to the emergent long-time behaviour of the system. This approach shows that the interactions we have inferred from our experiments can describe various cell-cell interaction modes known in the biological literature [73, 150, 170, 179, 182], including reversing, sliding and following interactions.

Applying a data-driven inference framework to interacting cells has therefore revealed novel, unexpected interactive dynamics and provided a unifying, quantitative framework to describe the various interaction phenotypes of cell pairs of different type. This work was made possible by previous progress on single confined cell migration, which provided a generalizable substrate for the interacting case. In the future, cell-cell interaction inference could provide a tool to assess the effect of pharmacological perturbations of the interacting cells. In this case, the inference methods could reveal how the interactions derived from unperturbed cells rely on particular molecular components of the cell-cell contacts.

The interaction behaviour space could furthermore provide a conceptual way to relate dynamics and behaviour also in more complex, collective settings, such as small clusters of cells [44, 166, 168], epithelial sheets [167, 184], or 3D organoids [185, 186]. Physical approaches to such multi-cellular systems thus far rely on modelling approaches where plausible types of interactions are assumed *a priori*. This is also the approach we follow in chapter 9, where we study the effect of cadherin-mediated cell-cell junctions on the collective spreading dynamics of cancer cell colonies. In this case, we use an indirect route to constrain the change in interactions upon disruption of the junctions by comparing experimental and theoretically predicted statistics. Being able to infer these interactions directly from the observed data could provide a much more direct route in the future, and might reveal unexpected types of interactions that would otherwise have gone unnoticed.

1.7 Outlook

In this introductory review, our aim was to introduce the relevant literature for the data-driven approach to cell migration pursued in this thesis, and to place the work presented in this thesis into context. In this last section, we aim to provide a brief outlook on how the data-driven approaches developed in this thesis could help answer open questions in the cell migration field in the future.

The first part of this review on single-cell dynamics contained two main themes: how to quantify and conceptualize the adaptation of cell behaviour to external confinements; and how to relate these behaviours to underlying molecular mechanisms. While we make first steps in this thesis to address these problems, much remains unknown. The basic molecular mechanisms that drive migration are increasingly well understood, but an integrated understanding of how these mechanisms interplay to set the emergent stochastic behaviours of cells at long time-scales still remains elusive. Furthermore, it remains unclear how cells respond to external confinements at the molecular level, and how these responses determine the emergent behaviour.

So far, data-driven cell migration models, from persistent random motion, confined cell migration, and interacting cells, have been limited to the treatment of low-dimensional sets of cellular degrees of freedom (DOFs), for example the trajectories of the cell nucleus. Clearly, the search for the principles governing the coupling from the molecular to the behavioural level calls for a description at a more microscopic level, which necessarily requires the tracking and analysis of further DOFs. A first step in this direction is made in chapter 6, where the effective protrusion position and the cell polarity are treated as additional DOFs. This minimal approach has yielded considerable additional insight compared to the treatment based on just the nucleus trajectories, highlighting how much more could be achieved by tracking further sub-cellular DOFs.

First steps towards descriptions of more complex cellular DOFs could be made based on analysis of the complete cell shape, rather than just the protrusions: shapes are easy to observe experimentally, and contain information about the overall state of the cell. For example, extending the approach of defining an effective protrusion position to the treatment of the complete protrusion-retraction shape components of the cell could give access to much more high-resolution information about how the cell reacts to its environment. These components corresponds to a 'ribbon' of alternating protrusion and retraction areas around the cell, quantifying the shape velocity at each moment in time. First approaches in this direction have been made in refs. [187, 188] where the temporal spectra of protrusions of freely migrating cells were quantified. In this case, it will likely be important to monitor the cellular dynamics at much higher time-resolution than in this thesis. Large time-intervals average out much of the short time-scale protrusive dynamics which could be required to understand this process. Applying a higher resolution approach to protrusion dynamics in confined geometries could provide additional insights into how these dynamics adapt to their environment.

A second approach to reveal the principles governing the interplay of cell shape and migration is to not only focus on the shape velocities, but to account for the entire shape as a function of time. The step from one-dimensional cell trajectories to the complete shape is a large jump in complexity. Thus, to determine the dominant contributions to the cell morphology, dimension reduction techniques may be required. By identifying the principle components of the cell shape, the morphodynamics could be studied in a low dimensional

space, similar to previous approaches in quantitative animal ethology [55, 56, 189]. A first step in this direction was made in ref. [190], where the application of different dimension reduction techniques to cell shapes were contrasted. This analysis revealed that cell shape alone is insufficient to predict the instantaneous velocity of the cell, suggesting that additional information beyond the shape is required to capture cell polarization. This observation furthermore suggests that cell shape may be thought of as a nematic DOF, which contains orientational information rather than polar information. Thus, to fully resolve cellular dynamics at the level of cell shapes will require the addition of further DOFs containing polar information, for instance shape velocities or information about the intra-cellular organization, such as the relative position of the cell nucleus or the traction forces. Determining a minimal set of DOFs required to predict cell polarization is an interesting problem on its own. Solving this puzzle may furthermore allow us to construct a morphodynamic space in which migration can be predicted based on a low-dimensional 'morphodynamic feature-vector'. Having a way to predict migration based on morphological features could then allow a more mechanistic understanding of cellular adaptation to external cues.

Changes in cell shape are driven by the polymerization and depolymerization of the actin network at the leading and trailing edges, respectively. The polymerization machinery in turn is controlled by a complex network of polarity cues, including Rac and Rho GTPases [191]. The local concentration of these cues determines the recruitment of actin regulators. Thus, the dynamics of the polarity cue concentration fields is a critical cellular DOF to understand migration and adaptation to geometries. While the general structure of the interaction network of polarity cues and actin regulators is increasingly well studied, the stochastic spatiotemporal dynamics of the polarity fields is not well understood. Furthermore, it remains unclear how these fields determine the overall migration behaviour on longer time-scales.

We argue that to address these open problems, data-driven inference tools can play a key role. How can we think about this problem mathematically? So far, we have used this type of approach to develop theories of the dynamics of the trajectory of net cell motion $\mathbf{x}_c(t)$. This trajectory in turn is determined by the dynamics of the cell shape, which can be represented as a discrete two-dimensional field $\mathcal{S}(\mathbf{x}, t)$, which takes values 1 inside the cell, and 0 outside. The underlying intra-cellular concentration fields in turn are continuous fields $\{\Phi_i(\mathbf{x}, t)\}$, where the index $i = 1...N$ refers to the different intra-cellular components such as the intensities of actin, myosin, Rac, Rho, CDC42, etc. These intensity fields could be obtained experimentally by staining particular components and tracking their spatiotemporal dynamics. Describing how these concentration fields couple to one another and to the cell shape would then required a formulation in terms of a stochastic field theory. Inferring field theories from observed data is currently an open theoretical problem, which would have to be addressed first. Methods to infer coupled partial differential equations have been developed for deterministic systems [68, 69], but no method for stochastic field inference exists thus far. Even if this method is developed, the problem has daunting complexity, and thus dimensional reduction of the relevant fields may be required to make it tractable. An approach in this direction has been developed in pioneering work by Danuser and coworkers [192–194], who measured the temporal correlations of signalling molecule recruitment and actin polymerization within cell protrusions. Such an approach has the potential to reveal causal relationships in the regulatory network in a perturbation-free manner. Extending these ideas to confined systems where cells are monitored on long time-scales, and combining it with stochastic inference methods, could yield key insights into the mechanistic basis of stochastic cell behaviours

and their adaptation to the environment.

Beyond actin polymerization and polarity cues, key players in confined migration are focal adhesions [71] and the cell nucleus [195–198]. These components are particularly important in 3D confined migration, where adhesions can be made at all sides of the cell, and the nucleus has to physically deform to overcome constrictions. Regarding cell adhesions, it would be particularly interesting to investigate how the adhesion-limited nucleus dynamics discovered in our mechanistic approach for 2D confined migration might change in 3D confinements. The switch from 2D to 3D could also have important consequences for the dynamics of cell protrusions and retractions, and the cell shape as a whole. While important progress has been made on how the nucleus can act as a sensor for external confinement [91, 197], it still remains unclear how nucleus deformations couple to the stochastic long time-scale dynamics of cell protrusions and polarity in 3D confinements. Thus, tracking the joint dynamics of nucleus deformation, protrusion formation, and cell migration could provide the basis for a data-driven approach to 3D confined migration.

In a similar way, data-driven approaches for cell-cell interactions could provide an avenue to better understand how molecular processes control interacting behaviours. One possible approach is to pharmacologically target molecular components that are known to be important in controlling cell-cell interactions, and then inferring the resulting change in dynamics. This may provide a way to identify the link between individual components with their role in the emergent behaviour. For example, a key question raised by the inference on two-cell collisions is how the separate positional and effective frictional components are controlled by molecular components. In particular, it is unclear what underlying mechanism controls the switch from friction or anti-friction interactions observed in non-cancerous and cancerous cells, respectively. Candidates are E-cadherin mediated cell-cell junctions, which are downregulated in cancer cells [73], or ephrins, which play a key role in cell-cell recognition [157]. Furthermore, to understand the emergence of the repulsive interaction between cells, which is responsible for Contact Inhibition of Locomotion, polarity cues, such as Rho GTPases, could be perturbed. These components are likely important in how cells change their direction of motion, an important process in the reversal events associated with CIL [72]. Thus, combining interaction inference with molecular perturbation in cell pair collision experiments could provide an avenue to link mechanisms and behaviour in interacting cellular systems.

Another interesting puzzle is how to describe the joint action of direct, contact-mediated cell-cell interactions, and indirect, substrate-mediated interactions. When cells migrate on mechanically soft, deformable substrates, their traction forces generate strain fields that can be sensed by other cells [199, 200]. By combining traction force microscopy [201] with trajectory inference in cell-cell interaction experiments on soft substrates could yield insight into how these two modes of communication interplay in cell collision events.

An exciting perspective is to take interaction inference beyond the level of the two-body problem. A key question that could be addressed this way is whether two-body interactions are sufficient to describe the dynamics of multi-body systems. For instance, can the dynamics of three confined and colliding cells be predicted purely based on the two-body interactions inferred from pairwise collisions? Or is an additional three-body interaction term required? Such higher order interaction types can be addressed with existing inference approaches [4, 66], and have been inferred from data in studies of animal behaviour, including interacting fish [177]. If three-body cell-cell interactions are discovered, this would raise the question

how it is mediated at the molecular level, for which currently no clear candidate mechanism exists.

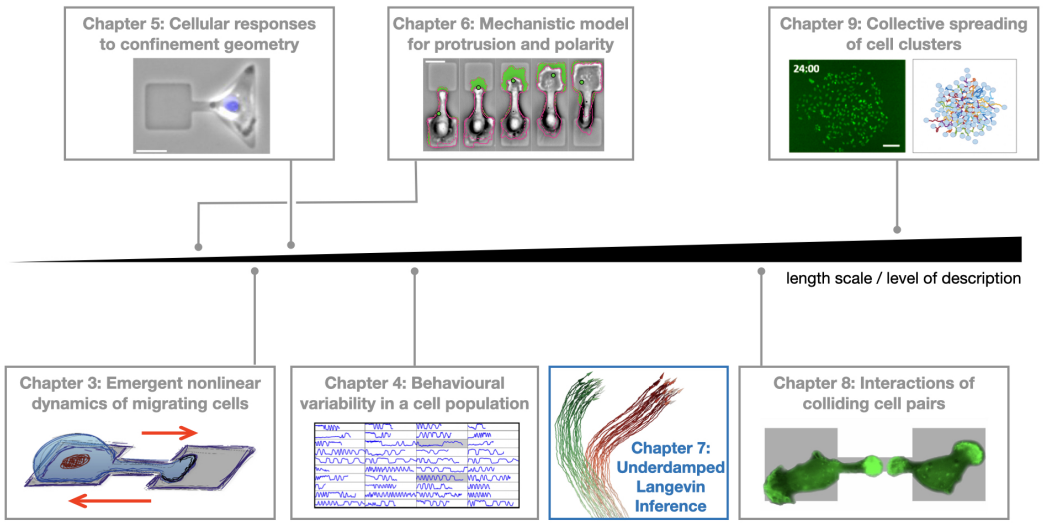
Interaction inference could also help in understanding the dynamics of larger cell assemblies. In many cases, active particle models which employ commonly observed types of interactions, such as repulsion and alignment, have been very successful at capturing the behaviour of collective cell migration [163–168]. However, there may be systems in which non-trivial types of interactions are at play. For example, cancerous tissue undergoes dramatic changes during metastasis, a process that can be recapitulated in *in vitro* cancer organoids [186, 202]. These changes happen simultaneously at the molecular, morphological, migratory, and mechanical level. A hallmark of this transition is the development of invasive branches along which cells migrate into the surrounding extra-cellular matrix [186, 202]. To understand this switch in migration behaviour, it would be interesting to disentangle the separate contributions of single-cell motility, cell-cell interactions, and the role of the overall morphology of the tumour, which acts as an effective confinement. To this end, an inference approach that can decompose the observed dynamics into these separate components could provide an important tool.

Taken together, these perspectives demonstrate how data-driven approaches have the potential to address key open questions in single and collective cell migration. A common thread in these ideas is that based on observed experimental trajectories, we no longer have to guess models, but we can infer them directly from data. Based on these inferred dynamics, we can then attempt to constrain underlying mechanisms, and predict emergent behaviours of the system. Our hope is that this philosophy could provide a basis for the development of a 'Physics of Cell Behaviour'.

Chapter 2

Outline of this thesis

The core of this thesis are self-contained manuscripts either prepared for a publication or already published in a peer-reviewed journal, which are presented in separate chapters (chapters 3-9). Here, I briefly summarize the open problem, the general approach, and key results of each manuscript.



Graphical Abstract. The projects in this thesis are presented roughly in order of increasing length scale and level of the description. In chapter 3, we develop a conceptual framework to analyze and interpret the dynamics of confined migrating cells. We then generalize this approach to account for the temporal and inter-individual variability in migratory behaviour (chapter 4) and to investigate how confined cells respond to adhesive geometries of different size, shape, and orientation (chapter 5). To gain insight into the underlying migratory mechanisms that control confined cell migration, we provide a theory for the protrusion and polarity dynamics of cell in chapter 6. In chapter 7, we develop a method to infer the dynamics of stochastic interacting systems from data, which we use in chapter 8 to make the first step from the single-cell level to interacting systems, by providing a direct inference of cell-cell interactions from trajectories of colliding cell pairs. Finally, in chapter 9, we investigate how the collective migration of cell clusters is controlled by cell-cell interactions.

Stochastic nonlinear dynamics of confined cell migration

In chapter 3, we introduce a combined experimental and theoretical approach to study the stochastic migration dynamics of cells in response to thin constrictions.

Reference

Stochastic nonlinear dynamics of confined cell migration in two-state systems
D. B. Brückner*, A. Fink*, C. Schreiber, P. J. F. Röttgermann, J. O. Rädler and C. P. Broedersz
Nature Physics 15, 595-601 (2019) [1]

Author contributions

A.F., C.S., P.J.F.R. and J.O.R designed experiments; A.F. and C.S. performed experiments; D.B.B., A.F. and C.S. analysed data. D.B.B. and C.P.B. developed the theoretical model. D.B.B., A.F., C.S., J.O.R. and C.P.B. wrote the manuscript.
This manuscript is also part of the PhD thesis of Alexandra Fink.

Features and Prizes

- Featured on the cover of Nature Physics in June 2019
- Featured in Nature Physics *News & Views* [203].
- Featured as a Nature Reviews Materials *Research Highlight* [204].
- Awarded with the LMU Center for NanoScience 'Best Interdisciplinary Publication Award' 2019.

Fun Fact

As of 12 February 2021, this article is the most tweeted article of all Nature Physics articles of a similar age (86 articles). It was tweeted 545 times. Among tracked articles of similar age in all scientific journals, it is in the 99^{th} percentile (ranked $1{,}246^{\text{th}}$ of 273,375 articles). Source: Altmetric via Nature Physics evaluated on 12 February 2021.

Abstract

Migrating cells in physiological processes, including development, homeostasis and cancer, encounter structured environments and are forced to overcome physical obstacles. Yet, the dynamics of confined cell migration remains poorly understood, and thus there is a need to study the complex motility of cells in controlled confining micro-environments. Here, we develop two-state micropatterns, consisting of two adhesive sites connected by a thin constriction, in which migrating cells perform repeated stochastic transitions. This minimal system enables us to obtain a large ensemble of single cell trajectories. From these trajectories, we infer an equation of cell motion, which decomposes the dynamics into deterministic and stochastic contributions in position-velocity phase space. Our results reveal that cells in two-state micropatterns exhibit intricate non-linear migratory dynamics, with qualitatively similar features for a cancerous (MDA-MB-231) and non-cancerous (MCF10A) cell line. In both cases, the cells drive themselves deterministically into the thin constriction; a process that is sped up by noise. Interestingly however, these two cell lines have distinct determin-

istic dynamics: MDA-MB-231 cells exhibit a limit cycle, while MCF10A cells show excitable bistable dynamics. Our approach yields a conceptual framework that may be extended to understand cell migration in more complex confining environments.

Key results

- In many cellular processes, the complex motility machinery of migrating cells reliably performs migration tasks, taking place in both 2D and 3D tissue [14, 16, 85, 101, 110]. In both cases, cells face a highly structured confining environment. To navigate such an environment, a migrating cell must overcome physical obstacles by squeezing through thin pores [18, 101–103]. While the dynamics of freely migrating cells is well understood in terms of persistent random walks [70, 74, 109, 110], a system-level understanding of the stochastic dynamics of cell migration in confined environments has remained elusive.
- To investigate the dynamics of confined migrating cells, we challenged cells with a minimal experimental migration problem: two adhesive islands on which the cell can rest, connected by a thin constriction. We observe that the cells repeatedly migrate between the islands, allowing us to characterize the statistics of this hopping process. Interestingly, we find that the migratory behaviour of the cells is highly variable, raising the question which aspects of the behaviour are deterministic, and which are stochastic.
- We disentangle the deterministic and stochastic components of the behaviour by inferring an equation of cell motion from the observed trajectories. While this approach is fully constrained by the short time-scale information of the trajectories, we find that the equation of motion accurately captures the key experimental long time-scale statistics, such as the steady-state probability distributions of position and velocity, the velocity correlation function, and the probability distribution of transition times.
- Our approach reveals that the cells have a deterministic tendency to overcome the thin constriction. In fact, the deterministic component of the migration is governed by intricate non-linear dynamics that are poised close to a bifurcation between a limit cycle and a bistable system: while MDA-MB-231 cells exhibit a limit cycle, MCF10A cells show excitable bistable dynamics.
- These results are in contrast with the intuitive expectation that the hopping behaviour might be generated by a noisy cellular activity competing with an effective energy barrier placed by the bridge. Instead, this framework suggests that the dynamics of confined cell migration are better understood as a non-linear dynamical system in position-velocity phase space.
- This approach could help advance our understanding of locomotion at the molecular level, as it provides strong constraints for bottom-up models that connect microscopic rules to the system-level dynamics of cells [22, 23, 134], and could provide a generalizable substrate to think about the dynamics of assemblies of interacting cells [39].

Quantifying the variability of cell migration behaviours

In chapter 4, we examine the variability in the behaviour of migrating cells both between individual cells and over time.

Reference
Disentangling the behavioral variability of confined cell migration
D. B. Brückner, A. Fink, J. O. Rädler and C. P. Broedersz
Journal of the Royal Society Interface 17, 20190689 (2020) [3]

Author contributions
D.B.B. and C.P.B. conceived the project; A.F. performed experiments; D.B.B. analysed data. D.B.B. and C.P.B. developed the theoretical model. D.B.B. and C.P.B. wrote the manuscript with input from all authors.

Abstract

Cell-to-cell variability is inherent to numerous biological processes, including cell migration. Quantifying and characterizing the variability of migrating cells is challenging, as it requires monitoring many cells for long time windows under identical conditions. Here, we observe the migration of single human breast cancer cells (MDA-MB-231) in confining two-state micropatterns. To describe the stochastic dynamics of this confined migration, we employ a dynamical systems approach. We identify statistics to measure the behavioural variance of the migration, which significantly exceed those predicted by a population-averaged stochastic model. This additional variance can be explained by the combination of an 'aging' process and population heterogeneity. To quantify population heterogeneity, we decompose the cells into subpopulations of slow and fast cells, revealing the presence of distinct classes of dynamical systems describing the migration, ranging from bistable to limit cycle behaviour. Our findings highlight the breadth of migration behaviours present in cell populations.

Key results

- Cell-to-cell variability is a key aspect of biological ensembles and has important implications for physiological processes such as collective migration of cell clusters, drug response, and cancer progression [117–121]. Numerous previous studies have provided quantitative methods for measuring variability of intra-cellular processes such as gene expression and protein translation [112–116]. Such cell-to-cell variability may in general propagate to the level of whole-cell behaviors − yet, it remains a key open question how to characterize and classify the diversity of phenotypes in complex migration dynamics from experimental data.
- A key challenge for such a characterization is that it requires both an appropriate theoretical framework as well as data sets where large numbers of migrating cells are monitored over long times under identical conditions. Here, we analyze the trajectories of hundreds of migrating cells confined in standardized micro-environments. Using these experimental cell trajectories, we develop a rigorous quantitative approach to show that the variance in behaviors between cells significantly exceeds that predicted by a single

stochastic process.

- We employ a dynamical systems approach, which we generalize for the case of heterogeneous ensembles and time-dependent 'aging' dynamics. While the ensemble-averaged statistics of confined cell migration are well captured by a stochastic equation of cell motion, described in chapter 3, this approach ignores cell-to-cell variability. We determine a simple way of capturing phenotypic diversity in this system, by optimizing the predictive power of the generalized model with respect to the observed variances.
- This procedure reveals the presence of distinct dynamical classes in the population, including stochastic bistability and limit cycle oscillations. We further account for a time-dependent effect in the form of a gradual deceleration of the cells over time. By combining these two effects, we obtain a quantitative model that can accurately predict the overall behavioral variability of migrating cells. This provides a simple and insightful way to quantitatively characterize heterogeneity in this system.
- These findings suggest that within a cell population, there are qualitative differences in the class of dynamical systems describing the migration of individual cells. These results provide a new way to characterize cell-to-cell variability in cell migration and give insights into the breadth of phenotypic diversity in how cells perform complex cell migration tasks.

Cellular responses to varying confinement size, shape and orientation

In chapter 5, we investigate how the migration dynamics of cells respond to variations in size, shape, and orientation of the local confining micro-environment.

Reference

Area and geometry dependence of cell migration in asymmetric two-state micropatterns
A. Fink, D. B. Brückner, C. Schreiber, P. J. F. Röttgermann, C. P. Broedersz and J. O. Rädler
Biophysical Journal 110, 1886-1895 (2020) [2]

Author contributions

A.F., C.S., P.J.F.R. and J.O.R designed experiments; A.F. performed experiments; A.F. and D.B.B. analyzed data. A.F., D.B.B., C.S., C.P.B. and J.O.R. interpreted the experiments and wrote the manuscript.
This manuscript is also part of the PhD thesis of Alexandra Fink.

Features

- Featured in Biophysical Journal *New and Notable* [205]

Abstract

Micro-structured surfaces provide a unique framework to probe cell migration and cytoskeletal dynamics in a standardized manner. Here, we report on the steady-state occupancy probability of cells in asymmetric two-state microstructures that consist of two fibronectin-coated adhesion sites connected by a thin guidance cue. In these dumbbell-like structures, cells transition between the two sites in a repeated and stochastic manner and average dwell times in the respective microenvironments are determined from the cell trajectories. We study the dynamics of human breast carcinoma cells (MDA-MB-231) in these microstructures as a function of area, shape and orientation of the adhesion sites. On square adhesive sites with different areas, we find that the occupancy probability ratio is directly proportional to the ratio of corresponding adhesion site areas. These asymmetries are well captured by a simple model for the stochastic nonlinear dynamics of the cells which reveals generic features of the motion. Sites of equal area but different shape lead to equal occupancy, if shapes are isotropic, e.g. squared or circular. In contrast, an asymmetry in the occupancy is induced by anisotropic shapes like rhombi, triangles or rectangles that enable motion in the direction perpendicular to the transition axis. Analysis of the 2D motion of cells between two rectangles with orthogonal orientation suggests that cellular transition rates depend on the cell polarisation induced by anisotropic micropatterns. Taken together, our results illustrate how two-state-micropatterns provide a dynamic migration assay with distinct dwell times and relative cell occupancy as readouts, which may be useful to probe cell-microenvironment interactions.

Key results

- Determining the factors and mechanisms that steer migrating cells in tissue and in the extracellular matrix is still an important open biophysical challenge. Cells can be guided by external cues, such as gradients in chemokine or ligand concentrations [90, 206, 207], stiffness [208–211], or through geometrical constraints [106, 107, 212]. However, it remains unclear how migrating cell respond to local geometrical cues through changes in cytoskeletal organization and migratory activity. Micropatterns offer highly reproducible microenvironments, allowing the extraction of extensive single cell statistics and are therefore well suited to reveal subtle changes in a quantifiable manner.
- Here, we extend the study of confined cell migration on micropatterned two-state systems presented in chapter 3 with new experiments assessing the relative dwell times in asymmetric patterns, in which cells probe adhesive patches with different geometrical properties, including size, shape and orientation. An advantage of this experimental approach is that by migrating repeatedly between the two sites, cells probe two microenvironments relative to each other, giving access to an occupation probability ratio, which could be used to quantify cellular preferences for local environments.
- In these asymmetric two-state systems, cells transition back- and forth between the two sites in a repeated and stochastic manner. We measured the mean dwell times, defined as the time between transitions, and the occupation probabilities, allowing us to quantify the cellular response to the distinct geometric aspects of the adhesion sites.
- We find that the dwell times are directly proportional to the adhesion site areas. Furthermore, in the case of isotropic confining shapes, with no difference in the x- and y-directions such as square or circular islands, sites of equal area but different shape lead to equal occupancy.
- Surprisingly, in anisotropic shapes such as rectangles arranged in orthogonal orientations, the occupancy is asymmetric. Using fluorescent staining of the cytoskeletal organization and focal adhesions, we find that this effect is likely related to cells polarizing in different directions on differently oriented sites.
- In the future, this assay could potentially be used for phenotyping behaviours of different cell types and to study cellular responses to other environmental cues such as cell-surface interactions.

Protrusion dynamics in confined cell migration

In chapter 6, we develop a theoretical framework for the joint dynamics of cell protrusions, polarity, and nucleus motion in confining micro-environments.

Reference

Theory of protrusion and polarity dynamics in confined cell migration
D. B. Brückner, M. Schmitt, A. Fink, J. Flommersfeld, N. Arlt, E. Hannezo, J. O. Rädler and C. P. Broedersz
manuscript in preparation for publication.

Author contributions

D.B.B., E.H. and C.P.B. conceived the project; A.F. performed experiments, with supervision by J.R.; M.S., J.F. and N.A. performed tracking and image segmentation; D.B.B. and M.S. analysed data. D.B.B., M.S. and C.P.B. developed the theory. D.B.B. and C.P.B. wrote the manuscript with input from all authors.
This project originated from preliminary work presented in the Master thesis by M. Schmitt, supervised by C. Broedersz and co-supervised by D. Brückner.

Abstract

Cell migration in many physiological processes relies on the concerted dynamics of several cellular components, including the formation of cell protrusions, adhesive connections to the environment, and the positioning of the cell nucleus. These components are coupled by the polarizable active cytoskeleton, and together play the dual role of driving net motion of the cell and sensing its local microenvironment. However, it remains poorly understood how the dynamic interplay of these components determines the emergent migration behavior at the cellular scale, and how these dynamics adapt to confining environments. Here, we develop a hybrid mechanistic and data-driven theoretical approach, where we use experimental data to systematically constrain a mechanistic model for confined cell migration. We measure a large data set of joint protrusion and nucleus trajectories of cells migrating in standardized micropatterned confinements featuring a thin constriction. Interestingly, we find that cells exhibit a stereotypical migration pattern, with protrusions growing to precede the transmigration of the cell nucleus across the constriction, which we term 'protrusion-nucleus cycling'. Based on a data-driven approach, we reveal that the average dynamics of the cell nucleus are determined by the locally available adhesive area. Furthermore, our model indicates that the protrusion dynamics are driven by a cell polarity that couples to the local geometry by switching from a negative to a positive, self-reinforcing feedback loop under strong confinement. Strikingly, this model predicts, in agreement with the experiment, that the protrusion-nucleus cycling disappears when the constriction is removed. This implies that the self-reinforcing polarity feedback loop emerges as a consequence of an adaptation of the cellular dynamics to the presence of the thin constriction. Our theoretical approach therefore suggests polarity feedback adaptation as a key mechanism in confined cell migration.

Key results

- The migration of cells in confining environments is intimately linked to changes of their shape: migrating cells generate protrusions which drive migration [136, 213], and help navigate complex, structured extracellular environments [213, 214]. However, it remains unclear how the underlying protrusion and polarity dynamics determine the emergent migration dynamics of cells in structured environments on long time-scales.
- Here, we develop a hybrid data-driven and mechanistic approach, where we use experimental data to rigorously constrain a mechanistic model for confined cell migration postulated on the basis of physical intuition and known cellular processes. Specifically, we measure a large data set of joint nucleus and protrusion trajectories of cells migrating in the two-state micropatterns introduced in chapter 3.
- The cell protrusions exhibit a stereotypical pattern of growing to precede the transmigration of the cell nucleus across the constriction, which we term 'protrusion-nucleus cycling'.
- To disentangle the distinct contributions to these dynamics, we follow an approach where we separately constrain the dynamics of the cell nucleus and the cell protrusion, which we postulate to be driven by a stochastic polarity force. This approach reveals two key contributions to the cellular dynamics:
  - We find that the dynamics of the cell nucleus are determined by the locally available adhesive area, which is modelled as a spatially variable friction coefficient.
  - Our model suggests that the protrusion dynamics are driven by a cell polarity which couples to the geometry of the local confinement by switching from a negative to a positive, self-reinforcing feedback loop, leading to strong polarities and persistent protrusion growth in the constriction.
- Together with a linear coupling between protrusion and nucleus, these two components of the system are key in generating the observed stereotypical protrusion-nucleus cycling.
- Strikingly, our model predicts, in agreement with experiments, that the protrusion-nucleus cycling disappears when the constriction is removed. This suggests that the self-reinforcing polarity feedback loop emerges as a consequence of an adaptation of the cellular dynamics to the presence of the thin constriction.
- A central challenge for our mechanistic theoretical approach is to capture the emergent stochastic nonlinear dynamics of the system that we discovered in chapter 3. Interestingly, we can directly map the equations of motion of the mechanistic model to the effective underdamped nucleus dynamics, revealing that the amplification behavior observed in chapter 3 is a consequence of two combined effects: lower adhesiveness and enhanced polarity persistence in the constriction.
- Taken together, these insights reveal the key mechanisms that give rise to the stochastic nonlinear dynamics introduce in chapter 3, and suggest polarity feedback adaptation as a key mechanism in confined cell migration.

Inferring the dynamics of underdamped stochastic systems

In chapter 7, we develop an inference method, Underdamped Langevin Inference, to infer the underlying dynamics of underdamped stochastic systems from observed experimental trajectories. We develop a rigorous way to deal with finite data, discrete observations, and experimental measurement noise. This method allows us to perform inference also for interacting and high-dimensional systems, which we will make use of in chapter 8.

Reference
Inferring the dynamics of underdamped stochastic systems
D. B. Brückner*, P. Ronceray* and C. P. Broedersz
Physical Review Letters 125, 058103 (2020) [4]

Author contributions
D.B.B., P.R. and C.P.B. conceived the project. D.B.B. and P.R. performed calculations, performed simulations and analysed data. D.B.B., P.R. and C.P.B. wrote the paper.

Features and Prizes
- Featured as PRL Editor's suggestion.
- Awarded with the LMU Center for NanoScience 'Scientific Breakthrough Award' 2020

Abstract

Many complex systems, ranging from migrating cells to animal groups, exhibit stochastic dynamics described by the underdamped Langevin equation. Inferring such an equation of motion from experimental data can provide profound insight into the physical laws governing the system. Here, we derive a principled framework to infer the dynamics of underdamped stochastic systems from realistic experimental trajectories, sampled at discrete times and subject to measurement errors. This framework yields an operational method, Underdamped Langevin Inference (ULI), which performs well on experimental trajectories of single migrating cells and in complex high-dimensional systems, including flocks with Viscek-like alignment interactions. Our method is robust to experimental measurement errors, and includes a self-consistent estimate of the inference error.

Key results

- Many complex systems, from migrating cells to interacting animal swarms, exhibit effective inertial dynamics described by second-order stochastic differential equations [1, 55, 56, 70, 82, 164, 168, 177, 215–220]. In recent years, experimental trajectory data on such systems have become readily available. By performing inference on such trajectories, one could recover the underlying equations of motion, yielding profound insight into the physical laws governing such systems. However, reliable inference from experimental data is hampered by two key problems: experimental data is inevitably discrete and subject to measurement noise. We show that a straightforward generalization of conventional inference methods for first-order systems [62–66] results in divergent systematic biases in second-order systems, which cannot be overcome by simply recording

more or better data.

- We provide a solution to this problem by deriving estimators for the deterministic and stochastic terms of the underlying dynamics, which are robust against discretization effects and measurement noise.
- Conceptually, we find that projecting the dynamics onto a smooth set of basis functions is key to construct unbiased estimators for second-order systems, as opposed to grid-based coarse-graining approaches typically employed in first-order inference [62–65].
- This method is valid for general non-linear inertial stochastic processes, including multiplicative stochastic terms, making it broadly applicable to complex experimental systems.
- We show that this approach is practical and data-efficient by applying it to short noisy experimental trajectories of single migrating cells.
- We test ULI on simulated trajectory data of flocks of active particles with Vicsek-like alignment interactions. We show that our approach reliably recovers the spatial structure of pairwise-interactions, and can disentangle alignment and central-force contributions to these interactions. Thus, ULI provides a tool to gain insight into the governing collective dynamics of many-body systems, ranging from clusters of migrating cells to flocks of birds.

Relation to other manuscripts

In chapters 3-5, we apply a simpler inference method, using a grid-based coarse-graining approach. In this case, we are performing inference on a very large set of trajectories in a low dimensional system, which makes the grid-based approach feasible. In contrast, as we show in chapter 7, to perform inference on a single-cell basis, the more data-efficient ULI method is required. To account for the bias due to discretization discussed in chapter 7, we perform an *a posteriori* empirical iterative scheme in chapter 3. Finally, ULI allows us to perform inference in interacting systems, which we make use of in chapter 8, where we infer the interactions of confined pairs of migrating cells.

Learning cell-cell interactions from pair-wise collisions

In chapter 8, we introduce a theoretical framework for the stochastic dynamics of interacting pairs of cells.

Reference

Learning the dynamics of cell-cell interactions in confined cell migration
D. B. Brückner, N. Arlt, A. Fink, P. Ronceray, J. O. Rädler and C. P. Broedersz
Proceedings of the National Academy of Sciences 118, e2016602118 (2021) [5]

Author contributions

D.B.B., J.R., and C.P.B. conceived the project. A.F. performed experiments and N.A. executed the tracking. D.B.B. and N.A. analyzed the experimental data. D.B.B., N.A. and C.P.B. developed the theoretical model. P.R. contributed code to perform Underdamped Langevin Inference. D.B.B. and C.P.B. wrote the paper with input from all authors.
This project originated from preliminary work presented in the Master thesis by N. Arlt, supervised by C. Broedersz and co-supervised by D. Brückner.

Abstract

The migratory dynamics of cells in physiological processes, ranging from wound healing to cancer metastasis, rely on contact-mediated cell-cell interactions. These interactions play a key role in shaping the stochastic trajectories of migrating cells. While data-driven physical formalisms for the stochastic migration dynamics of single cells have been developed, such a framework for the behavioral dynamics of interacting cells still remains elusive. Here, we monitor stochastic cell trajectories in a minimal experimental cell collider: a dumbbell-shaped micropattern on which pairs of cells perform repeated cellular collisions. We observe different characteristic behaviors, including cells reversing, following and sliding past each other upon collision. Capitalizing on this large experimental data set of coupled cell trajectories, we infer an interacting stochastic equation of motion that accurately predicts the observed interaction behaviors. Our approach reveals that interacting non-cancerous MCF10A cells can be described by repulsion and friction interactions. In contrast, cancerous MDA-MB-231 cells exhibit attraction and anti-friction interactions, promoting the predominant relative sliding behavior observed for these cells. Based on these experimentally inferred interactions, we show how this framework may generalize to provide a unifying theoretical description of the diverse cellular interaction behaviors of distinct cell types.

Key results

- In various physiological processes, cell-cell interactions manifest in the collision behavior of cells, taking distinct forms in various contexts: different types of cells are known to exhibit a variety of collision outcomes, from reversal and following to sliding behaviors [39, 73, 149–151, 156, 170, 179, 181, 182]. However, there is currently no theoretical framework that allows a direct inference of the underlying interactive dynamics of such cells from experimental data.

- To establish such a framework, we design a minimal 'cell collider': a controlled confining microenvironment in which two cells constantly collide into each other. Interestingly, we find that different cell types exhibit distinct collision behaviours: while non-cancerous (MCF10A) cells predominantly reverse upon collision, cancerous (MDA-MB-231) cells frequently slide past one another. This raises the question: can these distinct collision dynamics be described in a unified quantitative framework?
- To answer this question, we use the large trajectory data set provided by our cell collider experiments. Using Underdamped Langevin Inference, an inference method introduced in chapter 7, we infer the stochastic equation of motion that governs the two-body dynamics of interacting cells. This approach gives interesting insights into these dynamics:
  - First, we find that the dynamics of single-cell motility and cell-cell interactions can be decoupled in a simple way, by describing them as two separate terms: a one-body motility term and a two-body interaction term. Interestingly, the one-body term qualitatively matches that observed in single cell experiments [1].
  - Second, we find that this interaction term takes qualitatively different forms for non-cancerous (MCF10A) and cancerous (MDA-MB-231) cells: while non-cancerous cells exhibit simple repulsion and friction, the cancerous cells interact through an effective 'anti-friction'.
- We then generalize the inferred theoretical framework by predicting an 'interaction behavior space', which relates the physical interaction terms to the emergent long-time behavior of the system. This approach shows that the interactions we have inferred from our experiments can describe various cell-cell interaction modes known in the biological literature [73, 150, 170, 179, 182], including reversing, sliding and following interactions. This shows how this framework may generalize to describe the diverse cellular interaction behaviors of distinct cell types.

Cell-cell interactions in collective cell spreading

In chapter 9, we investigate the collective dynamics of spreading cell clusters, and how these dynamics are affected by the cadherin-mediated cell-cell adhesions.

Reference

Disentangling cadherin-mediated cell-cell interactions in collective cancer cell migration
T. Zisis*, D. B. Brückner*, T. Brandstätter, J. d'Alessandro, A. Vollmar, C. P. Broedersz, S. Zahler
submitted for publication.

Author contributions

T.Z., D.B.B., C.P.B. and S.Z. designed the study. T.Z. performed all experiments. J.A. contributed tracking software. D.B.B. and T.Z. analyzed data. D.B.B. and T.B. developed the theoretical model. T.Z. and D.B.B. wrote the paper with input from all authors.

Abstract

Cell dispersion from a confined area is fundamental in a number of biological processes, including cancer metastasis. To date, a quantitative understanding of the interplay of single cell motility, cell proliferation, and intercellular contacts remains elusive. In particular, the role of E- and N-Cadherin junctions, central components of intercellular contacts, is still controversial. Combining theoretical modeling with in vitro observations, we investigate the collective spreading behavior of colonies of human cancer cells (T24). Inhibition of E- and N-Cadherin junctions decreases colony spreading and average spreading velocities, without affecting the strength of correlations in spreading velocities of neighboring cells. Based on a biophysical simulation model for cell migration, we show that the behavioral changes upon disruption of these junctions can be explained by reduced repulsive excluded volume interactions between cells. This suggests that cadherin-based intercellular contacts sharpen cell boundaries leading to repulsive rather than cohesive interactions between cells, thereby promoting efficient cell spreading during collective migration.

Key results

- E- and N-Cadherins are central components of cell-cell adhesions, which play a critical role in cell migration during cancer metastasis [36, 221–223]. However, their distinct contribution to collective cell migration remains poorly understood, and quantitative frameworks to rigorously determine their impact on migration are currently lacking [39].
- To study the impact of cadherin junctions on collective migration, we develop an experimental system in which a cluster of cells is initially confined to a circular micropatterned region. By chemically activating the surrounding surface, the confinement is lifted, initiating collective spreading of the cluster in a standardized manner. This allows us to gather large data sets of many collective cell spreading events.
- By disrupting E- or N-Cadherin in collectively migrating T24 cancer cells through antibody blocking, we show that the absence of Cadherin-mediated contacts significantly reduces their spreading efficiency. Surprisingly however, the correlations in the velocity

fields of the cell sheet remain unaffected.

- To elucidate these findings, we develop a minimal active particle model for the collective migration dynamics. Our model shows that cell proliferation as well as repulsive excluded volume and Contact Inhibition of Locomotion interactions between cells drive tissue spreading.
- We systematically vary the types of cell-cell interactions included, and find that blocking either of the Cadherins has an effect akin to reducing repulsive excluded volume interactions in the model. In contrast, polarity interactions such as contact inhibition of locomotion, which control the velocity correlations, remain unaffected.
- Thus, our findings indicate that E- and N-Cadherins promote cell-cell repulsion by sharpening cellular boundaries rather than increasing cohesion, as one might intuitively expect. Our results thus suggest cell-cell repulsion as a decisive control parameter of collective cell migration.

Bibliography

[1] 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. *Nature Physics*, 15(6): 595–601, 2019.

[2] A. Fink, D. B. Brückner, C. Schreiber, P. J. Röttgermann, C. P. Broedersz, and J. O. Rädler. Area and Geometry Dependence of Cell Migration in Asymmetric Two-State Micropatterns. *Biophysical Journal*, 118(3):552–564, 2020.

[3] D. B. Brückner, A. Fink, J. O. Rädler, and C. P. Broedersz. Disentangling the Behavioural Variability of Confined Cell Migration. *J. R. Soc. Interface*, 17:20190689, 2020.

[4] D. B. Brückner, P. Ronceray, and C. P. Broedersz. Inferring the dynamics of underdamped stochastic systems. *Physical Review Letters*, 125(5):58103, 2020.

[5] 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. USA*, 118(7):e2016602118, 2020.

[6] M. Dietrich, H. Le Roy, D. B. Brückner, H. Engelke, R. Zantl, J. O. Rädler, and C. P. Broedersz. Guiding 3D cell migration in deformed synthetic hydrogel microstructures. *Soft Matter*, 14(15):2816–2826, 2018.

[7] J. Burelbach, D. B. Brückner, D. Frenkel, and E. Eiser. Thermophoretic forces on a mesoscopic scale. *Soft Matter*, 14(36):7446–7454, 2018.

[8] A. Goychuk, D. B. Brückner, A. W. Holle, J. P. Spatz, C. P. Broedersz, and E. Frey. Morphology and Motility of Cells on Soft Substrates. 2018.

[9] C. M. Franz, G. E. Jones, and A. J. Ridley. Cell Migration in Development and Disease. *Developmental Cell*, 2(2):153–158, 2002.

[10] E. Scarpa and R. Mayor. Collective cell migration in development. *The Journal of cell biology*, 212(2): 143–155, jan 2016.

[11] A. D. Luster, R. Alon, and U. H. von Andrian. Immune cell migration in inflammation: present and future therapeutic targets. *Nature Immunology*, 6(12):1182–1190, 2005.

[12] B. Li and J. H.-C. Wang. Fibroblasts and myofibroblasts in wound healing: force generation and measurement. *Journal of tissue viability*, 20(4):108–120, nov 2011.

[13] P. Bainbridge. Wound healing and the role of fibroblasts. *Journal of wound care*, 22(8):407–408,410–412, aug 2013.

[14] S. R. K. Vedula, H. Hirata, M. H. Nai, A. Brugués, Y. Toyama, X. Trepat, C. T. Lim, and B. Ladoux. Forces driving epithelial wound healing. *Nature Physics*, 10(1):683–690, 2014.

[15] D. Krndija, F. E. Marjou, B. Guirao, S. Richon, O. Leroy, Y. Bellaiche, E. Hannezo, and D. M. Vignjevic. Active cell migration is critical for steady-state epithelial turnover in the gut. *Science*, 710(August):705–710, 2019.

[16] P. Friedl and K. Wolf. Tumour-cell invasion and migration: diversity and escape mechanisms. *Nature reviews. Cancer*, 3(5):362–74, 2003.

[17] H. Yamaguchi, J. Wyckoff, and J. Condeelis. Cell migration in tumors. *Current opinion in cell biology*, 17(5):559–564, oct 2005.

[18] C. D. Paul, P. Mistriotis, and K. Konstantopoulos. Cancer cell motility: Lessons from migration in confined spaces. *Nature Reviews Cancer*, 17(2):131–140, 2017.

[19] C. C. DuFort, M. J. Paszek, and V. M. Weaver. Balancing forces: architectural control of mechanotransduction. *Nature reviews. Molecular cell biology*, 12(5):308–319, may 2011.

[20] J. D. Humphrey, E. R. Dufresne, and M. A. Schwartz. Mechanotransduction and extracellular matrix homeostasis. *Nature reviews. Molecular cell biology*, 15(12):802–812, dec 2014.

[21] P. Devreotes and A. R. Horwitz. Signaling Networks that Regulate Cell Migration. *Cold Spring Harb Perspect Biol*, 7:a005959, 2015.

[22] P. Maiuri, J. F. Rupprecht, S. Wieser, V. Ruprecht, O. Bénichou, N. Carpi, M. Coppey, S. De Beco, N. Gov, C. P. Heisenberg, C. Lage 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(2):374–386, 2015.

[23] A. C. Callan-Jones and R. Voituriez. Actin flows in cell migration: From locomotion and polarity to trajectories. *Current Opinion in Cell Biology*, 38:12–17, 2016.

[24] T. D. Pollard. The cytoskeleton, cellular motility and the reductionist agenda. *Nature*, 422(6933):741–5, 2003.

[25] A. R. Bausch and K. Kroy. A bottom-up approach to cell mechanics. *Nature Physics*, 2(4):231–238, 2006.

[26] K. E. Kasza, A. C. Rowat, J. Liu, T. E. Angelini, C. P. Brangwynne, G. H. Koenderink, and D. A. Weitz. The cell as a material. *Current opinion in cell biology*, 19(1):101–107, feb 2007.

[27] D. A. Fletcher and R. D. Mullins. Cell mechanics and the cytoskeleton. *Nature*, 463(7280):485–492, 2010.

[28] K. Kruse, J. F. Joanny, F. Jülicher, J. Prost, and K. Sekimoto. Generic theory of active polar gels: a paradigm for cytoskeletal dynamics. *Eur. Phys. J. E*, 16(16):5–16, 2005.

[29] J. Prost, F. Jülicher, and J. F. Joanny. Active gel physics. *Nature Physics*, 11(2):111–117, 2015.

[30] F. Jülicher, S. W. Grill, and G. Salbreux. Hydrodynamic theory of active matter. *Reports on Progress in Physics*, 81(7), 2018.

[31] J. Halatek, F. Brauns, and E. Frey. Self-organization principles of intracellular pattern formation. *arXiv*, 2018.

[32] A. B. Kolomeisky and M. E. Fisher. Molecular Motors: A Theorist's Perspective. *Annual Review of Physical Chemistry*, 58(1):675–695, apr 2007.

[33] C. Battle, C. P. Broedersz, N. Fakhri, V. F. Geyer, J. Howard, C. F. Schmidt, and F. C. MacKintosh. Broken detailed balance at mesoscopic scales in active biological systems. *Science*, 352(6285):604–607, apr 2016.

[34] F. S. Gnesotto, F. Mura, J. Gladrow, and C. P. Broedersz. Broken detailed balance and non-equilibrium dynamics in living systems: A review. *Reports on Progress in Physics*, 81(6), 2018.

[35] G. Duclos, C. Blanch-Mercader, V. Yashunsky, G. Salbreux, J. F. Joanny, J. Prost, and P. Silberzan. Spontaneous shear flow in confined cellular nematics. *Nature Physics*, 14(7):728–732, 2018.

[36] C. Pérez-González, R. Alert, C. Blanch-Mercader, M. Gómez-González, T. Kolodziej, E. Bazellieres, J. Casademunt, and X. Trepat. Active wetting of epithelial tissues. *Nature Physics*, 15(1):79–88, 2019.

[37] P. Romanczuk, M. Bär, W. Ebeling, B. Lindner, and L. Schimansky-Geier. Active Brownian particles: From individual to collective stochastic dynamics: From individual to collective stochastic dynamics. *European Physical Journal: Special Topics*, 202(1):1–162, 2012.

[38] X. Serra-Picamal, V. Conte, R. Vincent, E. Anon, D. T. Tambe, E. Bazellieres, J. P. Butler, J. J. Fredberg, and X. Trepat. Mechanical waves during tissue expansion. *Nature Physics*, 8(8):628–634, 2012.

[39] R. Alert and X. Trepat. Physical Models of Collective Cell Migration. *Annual Review of Condensed Matter Physics*, 11(1):77–101, 2020.

[40] D. Shao, W. J. Rappel, and H. Levine. Computational model for cell morphodynamics. *Physical Review Letters*, 105(10):2–5, 2010.

[41] D. Shao, H. Levine, and W. J. Rappel. Coupling actin flow, adhesion, and morphology in a computational cell motility model. *Proceedings of the National Academy of Sciences of the United States of America*, 109(18):6851–6856, 2012.

[42] F. Ziebert, S. Swaminathan, and I. S. Aranson. Model for self-polarization and motility of keratocyte fragments. *Journal of the Royal Society Interface*, 9(70):1084–1092, 2012.

[43] F. Graner and J. A. Glazier. Simulation of biological cell sorting using a two-dimensional extended Potts model. *Physical Review Letters*, 69(13):2013–2016, 1992.

[44] F. J. Segerer, F. Thüroff, A. Piera Alberola, E. Frey, and J. O. Rädler. Emergence and persistence of collective cell migration on small circular micropatterns. *Physical Review Letters*, 114(22):228102, 2015.

[45] C. Chan and D. Odde. Traction Dynamics of Filopodia on Compliant Substrates. *Science*, 322 (December):1687–1691, 2008.

[46] A. Elosegui-artola, X. Trepat, and P. Roca-Cusachs. Control of Mechanotransduction by Molecular Clutch Dynamics. *Trends in Cell Biology*, 28(5):356–367, 2018.

[47] K. Kruse, J. F. Joanny, F. Jülicher, and J. Prost. Contractility and retrograde flow in lamellipodium motion. *Physical Biology*, 3(2):130–137, 2006.

[48] P. Recho, T. Putelat, and L. Truskinovsky. Active gel segment behaving as an active particle. *Physical Review E*, 100(6):1–15, 2019.

[49] J. E. Ron, P. Monzo, N. C. Gauthier, R. Voituriez, and N. S. Gov. One-dimensional cell motility patterns. *Physical Review Research*, 2(3):1–27, 2020.

[50] P. Sens. Stick-Slip model for actin-driven cell protrusions, cell polarisation and crawling. *Proc. Natl. Acad. Sci. USA*, 117(40):24670–24678, 2020.

[51] K. Hennig, I. Wang, P. Moreau, L. Valon, S. DeBeco, M. Coppey, Y. A. Miroshnikova, C. Albiges-Rizo, C. Favard, R. Voituriez, and M. Balland. Stick-slip dynamics of cell adhesion triggers spontaneous symmetry breaking and directional migration of mesenchymal cells on one-dimensional lines. *Science Advances*, 6(1):1–13, 2020.

[52] C. Schreiber, B. Amiri, J. C. Heyn, J. O. Rädler, and M. Falcke. On the adhesionvelocity relation and length adaptation of motile cells on stepped fibronectin lanes. *Proceedings of the National Academy of Sciences of the United States of America*, 118(4), 2021.

[53] P. W. Anderson. More Is Different. *Science*, 177(4047):393 LP – 396, aug 1972.

[54] A. E. Brown and B. de Bivort. Ethology as a physical science. *Nature Physics*, 14:653–657, 2018.

[55] G. J. Stephens, B. Johnson-Kerner, W. Bialek, and W. S. Ryu. Dimensionality and Dynamics in the Behavior of C. elegans. *PLoS Comput Biol*, 4(4):e1000028, 2008.

[56] G. J. Stephens, M. Bueno, D. Mesquita, W. S. Ryu, and W. Bialek. Emergence of long timescales and stereotyped behaviors in Caenorhabditis elegans. *Proc. Natl. Acad. Sci. USA*, 108:7286–7289, 2011.

[57] A. C. Costa, T. Ahamed, and G. J. Stephens. Adaptive , locally linear models of complex dynamics. *Proc Natl Acad Sci U S A*, 116(5):1501–1510, 2019.

[58] E. Schneidman, M. J. Berry, R. Segev, and W. Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440(7087):1007–1012, 2006.

[59] G. Tkačik, O. Marre, T. Mora, D. Amodei, M. J. Berry, and W. Bialek. The simplest maximum entropy model for collective behavior in a neural network. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(3), 2013.

[60] M. Genkin and T. A. Engel. Moving beyond generalization to accurate interpretation of flexible models. *Nature Machine Intelligence*, 2(11):674–683, 2020.

[61] W. Młynarski, M. Hledík, T. R. Sokolowski, and G. Tkačik. Statistical analysis and optimality of neural systems. *Neuron*, mar 2021.

[62] S. Siegert, R. Friedrich, and J. Peinke. Analysis of data sets of stochastic systems. *Phys. Lett. A*, 243 (5-6):275–280, mar 1998.

[63] M. Ragwitz and H. Kantz. Indispensable Finite Time Corrections for Fokker-Planck Equations from Time Series Data. *Physical Review Letters*, 87(25):254501, 2001.

[64] M. E. Beheiry, M. Dahan, and J. B. Masson. InferenceMAP: Mapping of single-molecule dynamics with Bayesian inference. *Nature Methods*, 12(7):594–595, 2015.

[65] L. Pérez García, J. Donlucas Pérez, G. Volpe, A. V. Arzola, and G. Volpe. High-performance reconstruction of microscopic force fields from Brownian trajectories. *Nature Communications*, 9:5166, 2018.

[66] A. Frishman and P. Ronceray. Learning force fields from stochastic trajectories. *Physical Review X*, 10 (2):21009, 2020.

[67] F. Ferretti, V. Chardès, T. Mora, A. M. Walczak, and I. Giardina. Building General Langevin Models from Discrete Datasets. *Physical Review X*, 10(3):031018, 2020.

[68] S. H. Rudy, S. L. Brunton, J. L. Proctor, and J. N. Kutz. Data-driven discovery of partial differential equations. *Science Advances*, 3:e1602614, 2016.

[69] A. Borzou, A. E. Patteson, and J. M. Schwarz. A Data-Driven Statistical Field Theory for Active Matter. *Arxiv*, 2103.03461, 2021.

[70] D. Selmeczi, S. Mosler, P. H. Hagedorn, N. B. Larsen, and H. Flyvbjerg. Cell motility as persistent random motion: theories from experiments. *Biophysical journal*, 89(2):912–931, 2005.

[71] S. Lo Vecchio, R. Thiagarajan, D. Caballero, V. Vigon, L. Navoret, R. Voituriez, and D. Riveline. Collective Dynamics of Focal Adhesions Regulate Direction of Cell Motion. *Cell Systems*, 10:1–8, 2020.

[72] E. Scarpa, A. Roycroft, E. Theveneau, E. Terriac, M. Piel, R. Mayor, E. Scarpa, A. Roycroft, E. Theveneau, E. Terriac, M. Piel, and R. Mayor. A novel method to study contact inhibition of locomotion using micropatterned substrates. *Biology Open*, 2:901–906, 2013.

[73] D. F. Milano, N. A. Ngai, S. K. Muthuswamy, and A. R. Asthagiri. Regulators of Metastasis Modulate the Migratory Response to Cell Contact under Spatial Confinement. *Biophysical Journal*, 110(8):1886–1895, 2016.

[74] M. H. Gail and C. W. Boone. The Locomotion of Mouse Fibroblasts in Tissue Culture. *Biophysical Journal*, 10(10):980–993, 1970.

[75] D. Selmeczi, L. Li, L. I. I. Pedersen, S. F. Nrrelykke, P. H. Hagedorn, S. Mosler, N. B. Larsen, E. C. Cox, and H. Flyvbjerg. Cell motility as random motion: A review. *European Physical Journal: Special Topics*, 157(1):1–15, 2008.

[76] T. H. Harris, E. J. Banigan, D. A. Christian, C. Konradt, E. D. Wojno, K. Norose, E. H. Wilson, B. John, W. Weninger, A. D. Luster, A. J. Liu, and C. A. Hunter. Generalized Lévy walks and the role of chemokines in migration of effector CD8 + T cells. *Nature*, 486(7404):545–548, 2012.

[77] K. Przibram. Über die ungeordnete Bewegung niederer Tiere. *Pflügers Arch. Physiol.*, 153(8-10):401–405, 1917.

[78] R. Fürth. Die Brownsche Bewegung bei Berücksichtigung einer Persistenz der Bewegungsrichtung. Mit Anwendungen auf die Bewegung lebender Infusorien. *Z. Physik*, 2(3):244–256, 1920.

[79] J. N. Pedersen, L. Li, C. Gradinaru, R. H. Austin, E. C. Cox, and H. Flyvbjerg. How to connect time-lapse recorded trajectories of motile microorganisms with dynamical models in continuous time. *Physical Review E*, 94(6):062401, 2016.

[80] H. Takagi, M. J. Sato, T. Yanagida, and M. Ueda. Functional analysis of spontaneous cell movement under different physiological conditions. *PLoS ONE*, 3(7):e2648, 2008.

[81] H. U. Bödeker, C. Beta, T. D. Frank, and E. Bodenschatz. Quantitative analysis of random ameboid motion. *EPL (Europhysics Letters)*, 90(2):28005, 2010.

[82] L. Li, E. C. Cox, and H. Flyvbjerg. 'Dicty dynamics': Dictyostelium motility as persistent random motion. *Physical Biology*, 8(4):046006, 2011.

[83] P. Dieterich, R. Klages, R. Preuss, and A. Schwab. Anomalous dynamics of cell migration. *Proc. Natl. Acad. Sci. USA*, 105(2):459–63, 2008.

[84] A. A. Potdar, J. Jeon, A. M. Weaver, V. Quaranta, and P. T. Cummings. Human mammary epithelial cells exhibit a bimodal correlated random walk pattern. *PLoS ONE*, 5(3):e9636, 2010.

[85] S. Even-Ram and K. M. Yamada. Cell migration in 3D matrix. *Current Opinion in Cell Biology*, 17(5 SPEC. ISS.):524–532, 2005.

[86] M. K. Driscoll and G. Danuser. Quantifying Modes of 3D Cell Migration. *Trends in Cell Biology*, 25 (12):749–759, 2015.

[87] P. Vargas, L. Barbier, P. J. Sáez, and M. Piel. Mechanisms for fast cell migration in complex environments. *Current Opinion in Cell Biology*, 48:72–78, 2017.

[88] R. Singhvi, A. Kumar, G. P. Lopez, G. N. Stephanopoulos, D. C. Wang, G. M. Whitesides, and D. E. Ingber. Engineering Cell Shape and Function. *Science*, 264(5159):696–698, 1994.

[89] C. S. Chen, M. Mrksich, S. Huang, G. M. Whitesides, and D. E. Ingber. Micropatterned surfaces for control of cell shape, position, and function. *Biotechnology Progress*, 14(3):356–363, 1998.

[90] B. Lin, T. Yin, Y. I. Wu, T. Inoue, and A. Levchenko. Interplay between chemotaxis and contact inhibition of locomotion determines exploratory cell migration. *Nature Communications*, 6, 2015.

[91] J. Renkawitz, A. Kopf, J. Stopp, I. de Vries, M. K. Driscoll, J. Merrin, R. Hauschild, E. S. Welf, G. Danuser, R. Fiolka, and M. Sixt. Nuclear positioning facilitates amoeboid migration along the path of least resistance. *Nature*, 568(7753):546–550, 2019.

[92] A. Kopf, J. Renkawitz, R. Hauschild, I. Girkontaite, K. Tedford, J. Merrin, O. Thorn-Seshold, D. Trauner, H. Häcker, K.-D. Fischer, E. Kiermaier, and M. Sixt. Microtubules control cellular shape and coherence in amoeboid migrating cells. *The Journal of Cell Biology*, 219(6):e201907154, 2020.

[93] A. Reversat, F. Gaertner, J. Merrin, J. Stopp, S. Tasciyan, J. Aguilera, I. D. Vries, R. Hauschild, M. Hons, M. Piel, A. Callan-jones, R. Voituriez, and M. Sixt. Cellular locomotion using environmental topography. *Nature*, 582:582–585, 2020.

[94] Y. Z. Zhang, B. Su, J. Venugopal, S. Ramakrishna, and C. T. Lim. Biomimetic and bioactive nanofibrous scaffolds from electrospun composite nanofibers. *International journal of nanomedicine*, 2(4):623–638, 2007.

[95] C. Guetta-Terrier, P. Monzo, J. Zhu, H. Long, L. Venkatraman, Y. Zhou, P. P. Wang, S. Y. Chew, A. Mogilner, B. Ladoux, and N. C. Gauthier. Protrusive waves guide 3D cell migration along nanofibers. *Journal of Cell Biology*, 211(3):683–701, 2015.

[96] J. Singh, P. Sharma, and A. Nain. Rules of Contact Inhibition of Locomotion in Cells Migrating on ECM Mimicking Fibers. *PNAS*, 2021.

[97] P. Maiuri, E. Terriac, P. Paul-Gilloteaux, T. Vignaud, K. McNally, J. Onuffer, K. Thorn, P. A. Nguyen, N. Georgoulia, D. Soong, A. Jayo, N. Beil, J. Beneke, J. C. Hong Lim, C. Pei-Ying Sim, Y. S. Chu, A. Jiménez-Dalmaroni, J. F. Joanny, J. P. Thiery, H. Erfle, M. Parsons, T. J. Mitchison, W. A. Lim, A. M. Lennon-Duménil, M. Piel, and M. Théry. The first World Cell Race. *Current Biology*, 22(17): R673–R675, 2012.

[98] S. I. Fraley, Y. Feng, A. Giri, G. D. Longmore, and D. Wirtz. Dimensional and temporal controls of three-dimensional cell migration by zyxin and binding partners. *Nature Communications*, 3:713–719, 2012.

[99] A. D. Doyle, F. W. Wang, K. Matsumoto, and K. M. Yamada. One-dimensional topography underlies three-dimensional fibrillar cell migration. *Journal of Cell Biology*, 184(4):481–490, 2009.

[100] S. S. Chang, W. H. Guo, Y. Kim, and Y. L. Wang. Guidance of cell migration by substrate dimension. *Biophysical Journal*, 104(2):313–321, 2013.

[101] J. A. Park, J. H. Kim, D. Bi, J. A. Mitchel, N. T. Qazvini, K. Tantisira, C. Y. Park, M. McGill, S. H. Kim, B. Gweon, J. Notbohm, R. Steward, S. Burger, S. H. Randell, A. T. Kho, D. T. Tambe, C. Hardin, S. A. Shore, E. Israel, D. A. Weitz, D. J. Tschumperlin, E. P. Henske, S. T. Weiss, M. L. Manning, J. P. Butler, J. M. Drazen, and J. J. Fredberg. Unjamming and cell shape in the asthmatic airway epithelium. *Nature Materials*, 14(10):1040–1048, 2015.

[102] B. J. Green, M. Panagiotakopoulou, F. M. Pramotton, G. Stefopoulos, S. O. Kelley, D. Poulikakos, and A. Ferrari. Pore shape defines paths of metastatic cell migration. *Nano Letters*, 18:21402147, 2018.

[103] 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. *Journal of Cell Biology*, 201(7):1069–1084, 2013.

[104] G. Mahmud, C. J. Campbell, K. J. Bishop, Y. A. Komarova, O. Chaga, S. Soh, S. Huda, K. Kandere-Grzybowska, and B. A. Grzybowski. Directing cell motions on micropatterned ratchets. *Nature Physics*, 5(8):606–612, 2009.

[105] D. Caballero, R. Voituriez, and D. Riveline. Protrusion fluctuations direct cell motion. *Biophysical Journal*, 107(1):34–42, 2014.

[106] J. Comelles, D. Caballero, R. Voituriez, V. Hortigüela, V. Wollrab, A. L. Godeau, J. Samitier, E. Martínez, and D. Riveline. Cells as active particles in asymmetric potentials: Motility under external gradients. *Biophysical Journal*, 107(7):1513–1522, 2014.

[107] D. Caballero, J. Comelles, M. Piel, R. Voituriez, and D. Riveline. Ratchetaxis: Long-Range Directed Cell Migration by Local Cues. *Trends in Cell Biology*, 25(12):815–827, 2015.

[108] C. Schreiber, F. J. Segerer, E. Wagner, A. Roidl, and J. O. Rädler. Ring-Shaped Microlanes and Chemical Barriers as a Platform for Probing Single-Cell Migration. *Scientific Reports*, 6(February):26858, 2016.

[109] C. Metzner, C. Mark, J. Steinwachs, L. Lautscham, F. Stadler, and B. Fabry. Superstatistical analysis and modelling of heterogeneous random walks. *Nature communications*, 6(May):7516, 2015.

[110] P.-H. Wu, A. Giri, S. X. Sun, and D. Wirtz. Three-dimensional cell migration does not follow a random walk. *Proceedings of the National Academy of Sciences of the United States of America*, 111(11):3949–54, 2014.

[111] C. L. Vestergaard, J. N. Pedersen, K. I. Mortensen, and H. Flyvbjerg. Estimation of motility parameters from trajectory data: A condensate of our recent results. *European Physical Journal: Special Topics*, 224(7):1151–1168, 2015.

[112] M. Niepel, S. L. Spencer, and P. K. Sorger. Non-genetic cell-to-cell variability and the consequences for pharmacology. *Current Opinion in Chemical Biology*, 13:556–561, 2009.

[113] A. Sigal, R. Milo, A. Cohen, N. Geva-zatorsky, Y. Klein, Y. Liron, N. Rosenfeld, T. Danon, N. Perzov, and U. Alon. Variability and memory of protein levels in human cells. *Nature*, 444:28–31, 2006.

[114] A. Raj and A. van Oudenaarden. Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell*, 135(2):216–226, 2008.

[115] A. Wagner, A. Regev, and N. Yosef. Revealing the vectors of cellular identity with single-cell genomics. *Nature Biotechnology*, 34(11):1145–1160, 2016.

[116] S. J. Altschuler and L. F. Wu. Cellular Heterogeneity: Do Differences Make a Difference? *Cell*, 141(4): 559–563, 2010.

[117] A. A. Cohen, N. Geva-Zatorsky, E. Eden, M. Frenkel-Morgenstern, I. Issaeva, A. Sigal, R. Milo, C. Cohen-Saidon, Y. Liron, Z. Kam, L. Cohen, T. Danon, N. Perzov, and U. Alon. Dynamic Proteomics of Individual Cancer Cells in Response to a Drug. *Science*, 322(5907):1511–1516, dec 2008.

[118] O. Feinerman, J. Veiga, J. R. Dorfman, R. N. Germain, and G. Altan-Bonnet. Variability and Robustness in T Cell Activation from Regulated Heterogeneity in Protein Levels. *Science*, 321(5892):1081–1084, 2008.

[119] K. E. Gascoigne and S. S. Taylor. Cancer Cells Display Profound Intra- and Interline Variation following Prolonged Exposure to Antimitotic Drugs. *Cancer Cell*, 14(2):111–122, 2008.

[120] S. Wieser, J. Weghuber, M. Sams, H. Stockinger, and G. J. Schütz. Cell-to-cell variability in the diffusion constants of the plasma membrane proteins CD59 and CD147. *Soft Matter*, 5(17):3287–3294, 2009.

[121] B. A. Camley and W.-J. Rappel. Cell-to-cell variation sets a tissue-rheologydependent bound on collective gradient sensing. *Proceedings of the National Academy of Sciences*, 114(47):E10074 LP – E10082, nov 2017.

[122] X. Li, A. Das, and D. Bi. Mechanical Heterogeneity in Tissues Promotes Rigidity and Controls Cellular Invasion. *Physical Review Letters*, 123(5):58101, 2019.

[123] M. Boehm and N. E. G. Cell Cycle and Cell Migration. *Circulation*, 103(24):2879–2881, jun 2001.

[124] D. Jordan, S. Kuehn, E. Katifori, and S. Leibler. Behavioral diversity in microbes and low-dimensional phenotypic spaces. *Proceedings of the National Academy of Sciences*, 110(34):14018–14023, 2013.

[125] L. J. Schumacher, P. K. Maini, and R. E. Baker. Semblance of Heterogeneity in Collective Cell Migration. *Cell Systems*, 5(2):119–127, 2017.

[126] T. E. Chan, M. P. Stumpf, and A. C. Babtie. Gene Regulatory Network Inference from Single-Cell Data Using Multivariate Information Measures. *Cell Systems*, 5(3):251–267, 2017.

[127] H. Todorov, R. Cannoodt, W. Saelens, and Y. Saeys. Network Inference from Single-Cell Transcriptomic Data BT - Gene Regulatory Networks: Methods and Protocols. pages 235–249. Springer New York, New York, NY, 2019.

[128] B. Li and L. You. Predictive power of cell-to-cell variability. *Quantitative Biology*, 1(2):131–139, 2013.

[129] K. Sachs, O. Perez, and D. Pe. <Science 2005 Sachs K.pdf>. (August):523–530, 2005.

[130] A. Colman-Lerner, A. Gordon, E. Serra, T. Chin, O. Resnekov, D. Endy, C. Gustavo Pesce, and R. Brent. Regulated cell-to-cell variation in a cell-fate decision system. *Nature*, 437(7059):699–706, 2005.

[131] B. Snijder, R. Sacher, P. Rämö, E.-M. Damm, P. Liberali, and L. Pelkmans. Population context determines cell-to-cell variability in endocytosis and virus infection. *Nature*, 461:520, aug 2009.

[132] B. Snijder and L. Pelkmans. Origins of regulated cell-to-cell variability. *Nature Reviews Molecular Cell Biology*, 12(2):119–125, 2011.

[133] J. D'Alessandro, A. Barbier-Chebbah, V. Cellerin, O. Bénichou, R.-M. Mege, R. Voituriez, and B. Ladoux. Cell migration driven by long-lived spatial memory. *bioRxiv*, page 2021.01.05.425035, 2021.

[134] G. Danuser, J. Allard, and A. Mogilner. Mathematical Modeling of Eukaryotic Cell Migration: Insights Beyond Experiments. *Annual review of cell and developmental biology*, 4(164):501–528, 2013.

[135] E. Frey and K. Kroy. Brownian motion: a paradigm of soft matter and biological physics. *Annalen der Physik*, 14(13):20–50, feb 2005.

[136] T. D. Pollard and G. Borisy. Cellular Motility Driven by Assembly and Disassembly of Actin Filaments. *Cell*, 112:453–465, 2003.

[137] C. S. Peskin, G. M. Odell, and G. F. Oster. Cellular motions and thermal fluctuations: the Brownian ratchet. *Biophysical journal*, 65(1):316–324, jul 1993.

[138] A. Mogilner and G. Oster. Cell motility driven by actin polymerization. *Biophysical journal*, 71(6): 3030–3045, dec 1996.

[139] A. Mogilner and L. Edelstein-Keshet. Regulation of actin dynamics in rapidly moving cells: a quantitative analysis. *Biophysical journal*, 83(3):1237–1258, sep 2002.

[140] F. Thüroff, A. Goychuk, M. Reiter, and E. Frey. Bridging the gap between single-cell migration and collective dynamics. *eLife*, 8:e46842, 2019.

[141] B. A. Camley, Y. Zhang, Y. Zhao, B. Li, E. Ben-Jacob, H. Levine, and W.-J. Rappel. Polarity mechanisms such as contact inhibition of locomotion regulate persistent rotational motion of mammalian cells on micropatterns. *Proceedings of the National Academy of Sciences*, 111(41):14770–14775, 2014.

[142] T. Bertrand, J. D'Alessandro, A. Maitra, S. Jain, B. Mercier, R.-M. Mège, B. Ladoux, and R. Voituriez. Clustering and ordering in cell assemblies with generic asymmetric aligning interactions. 2020.

[143] C. W. Wolgemuth, J. Stajic, and A. Mogilner. Redundant mechanisms for stable cell locomotion revealed by minimal models. *Biophysical Journal*, 101(3):545–553, 2011.

[144] P. Monzo, Y. K. Chong, C. Guetta-Terrier, A. Krishnasamy, S. R. Sathe, E. K. F. Yim, W. H. Ng, B. T. Ang, C. Tang, B. Ladoux, N. C. Gauthier, and M. P. Sheetz. Mechanical confinement triggers glioma linear migration dependent on formin FHOD3. *Molecular Biology of the Cell*, 27(8):1246–1261, feb 2016.

[145] R. J. Hawkins, M. Piel, G. Faure-Andre, A. M. Lennon-Dumenil, J. F. Joanny, J. Prost, and R. Voituriez. Pushing off the walls: A mechanism of cell motility in confinement. *Physical Review Letters*, 102(5):1–4, 2009.

[146] M. Poujade, E. Grasland-Mongrain, A. Hertzog, J. Jouanneau, P. Chavrier, B. Ladoux, A. Buguin, and P. Silberzan. Collective migration of an epithelial monolayer in response to a model wound. *Proceedings of the National Academy of Sciences of the United States of America*, 104(41):15988–15993, 2007.

[147] B. Stramer, W. Wood, M. J. Galko, M. J. Redd, A. Jacinto, S. M. Parkhurst, and P. Martin. Live imaging of wound inflammation in Drosophila embryos reveals key roles for small GTPases during in vivo cell migration. *Journal of Cell Biology*, 168(4):567–573, 2005.

[148] H. Weavers, J. Liepe, A. Sim, W. Wood, P. Martin, and M. P. Stumpf. Systems Analysis of the Dynamic Inflammatory Response to Tissue Damage Reveals Spatiotemporal Properties of the Wound Attractant Gradient. *Current Biology*, 26(15):1975–1989, 2016.

[149] C. Carmona-Fontaine, H. K. Matthews, S. Kuriyama, M. Moreno, G. A. Dunn, M. Parsons, C. D. Stern, and R. Mayor. Contact inhibition of locomotion in vivo controls neural crest directional migration. *Nature*, 456(7224):957–961, 2008.

[150] B. Stramer and R. Mayor. Mechanisms and in vivo functions of contact inhibition of locomotion. *Nature reviews. Molecular cell biology*, 18(1):43–55, 2017.

[151] J. W. Astin, J. Batson, S. Kadir, J. Charlet, R. A. Persad, D. Gillatt, J. D. Oxley, and C. D. Nobes. Competition amongst Eph receptors regulates contact inhibition of locomotion and invasiveness in prostate cancer cells. *Nature Cell Biology*, 12(12):1194–1204, 2010.

[152] J. R. Davis, A. Luchici, F. Mosis, J. Thackery, J. A. Salazar, Y. Mao, G. A. Dunn, T. Betz, M. Miodownik, and B. M. Stramer. Inter-cellular forces orchestrate contact inhibition of locomotion. *Cell*, 161(2):361–373, 2015.

[153] R. Moore, E. Theveneau, S. Pozzi, P. Alexandre, J. Richardson, A. Merks, M. Parsons, J. Kashef, C. Linker, and R. Mayor. Par3 controls neural crest migration by promoting microtubule catastrophe during contact inhibition of locomotion. *Development*, 140(23):4763–4775, 2013.

[154] H. K. Matthews, L. Marchant, C. Carmona-Fontaine, S. Kuriyama, J. Larraín, M. R. Holt, M. Parsons, and R. Mayor. Directional migration of neural crest cells in vivo is regulated by Syndecan-4/Rac1 and non-canonical Wnt signaling/RhoA. *Development*, 135(10):1771–1780, 2008.

[155] S. Kadir, J. W. Astin, L. Tahtamouni, P. Martin, and C. D. Nobes. Microtubule remodelling is required for the front-rear polarity switch during contact inhibition of locomotion. *Journal of Cell Science*, 124 (15):2642–2653, 2011.

[156] M. Abercrombie and J. E. Heaysman. Observations on the social behaviour of cells in tissue culture. II. "Monolayering" of fibroblasts. *Experimental Cell Research*, 6(2):293–306, 1954.

[157] R. Mayor and C. Carmona-Fontaine. Keeping in touch with contact inhibition of locomotion. *Trends in Cell Biology*, 20(6):319–328, 2010.

[158] M. C. Marchetti, J. F. Joanny, S. Ramaswamy, T. B. Liverpool, J. Prost, M. Rao, and R. A. Simha. Hydrodynamics of soft active matter. *Reviews of Modern Physics*, 85(3), 2013.

[159] A. G. Fletcher, M. Osterfield, R. E. Baker, and S. Y. Shvartsman. Vertex models of epithelial morphogenesis. *Biophysical Journal*, 106(11):2291–2304, 2014.

[160] S. Alt, P. Ganguly, and G. Salbreux. Vertex models: From cell mechanics to tissue morphogenesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1720), 2017.

[161] D. Boocock, N. Hino, N. Ruzickova, T. Hirashima, and E. Hannezo. Theory of mechanochemical patterning and optimal migration in cell monolayers. *Nature Physics*, 2020.

[162] O. Ilina, P. G. Gritsenko, S. Syga, J. Lippoldt, C. A. La Porta, O. Chepizhko, S. Grosser, M. Vullings, G. J. Bakker, J. Starruß, P. Bult, S. Zapperi, J. A. Käs, A. Deutsch, and P. Friedl. Cellcell adhesion and 3D matrix confinement determine jamming transitions in breast cancer invasion. *Nature Cell Biology*, 22(9):1103–1115, 2020.

[163] B. Smeets, R. Alert, J. Pešek, I. Pagonabarraga, H. Ramon, and R. Vincent. Emergent structures and dynamics of cell colonies by contact inhibition of locomotion. *Proceedings of the National Academy of Sciences of the United States of America*, 113(51):14621–14626, 2016.

[164] N. Sepúlveda, L. Petitjean, O. Cochet, E. Grasland-Mongrain, P. Silberzan, and V. Hakim. Collective Cell Motion in an Epithelial Sheet Can Be Quantitatively Described by a Stochastic Interacting Particle Model. *PLoS Computational Biology*, 9(3), 2013.

[165] M. Basan, J. Elgeti, E. Hannezo, W. J. Rappel, and H. Levine. Alignment of cellular motility forces with tissue flow as a mechanism for efficient wound healing. *Proceedings of the National Academy of Sciences of the United States of America*, 110(7):2452–2459, 2013.

[166] K. Copenhagen, G. Malet-engra, W. Yu, G. Scita, N. Gov, and A. Gopinathan. Frustration-induced phases in migrating cell clusters. *Science Advances*, 4:eaar8483, 2018.

[167] S. Garcia, E. Hannezo, J. Elgeti, J.-F. Joanny, P. Silberzan, and N. S. Gov. Physics of active jamming during collective cellular motion in a monolayer. *Proceedings of the National Academy of Sciences*, 112 (50):15314–15319, 2015.

[168] J. D'alessandro, A. P. Solon, Y. Hayakawa, C. Anjard, F. Detcheverry, J. P. Rieu, and C. Rivière. Contact enhancement of locomotion in spreading cell colonies. *Nature Physics*, 13(10):999–1005, 2017.

[169] S. Huang, C. P. Brangwynne, K. K. Parker, and D. E. Ingber. Symmetry-breaking in mammalian cell cohort migration during tissue pattern formation: Role of random-walk persistence. *Cell Motility and the Cytoskeleton*, 61(4):201–213, 2005.

[170] D. Li and Y. L. Wang. Coordination of cell migration mediated by sitedependent cell-cell contact. *Proceedings of the National Academy of Sciences of the United States of America*, 115(42):10678–10683, 2018.

[171] R. A. Desai, S. B. Gopal, S. Chen, and C. S. Chen. Contact inhibition of locomotion probabilities drive solitary versus collective cell migration. *Journal of the Royal Society Interface*, 10(88), 2013.

[172] D. Mohammed, G. Charras, E. Vercruysse, M. Versaevel, J. Lantoine, L. Alaimo, C. Bruyère, M. Luciano, K. Glinel, G. Delhaye, O. Théodoly, and S. Gabriele. Substrate area confinement is a key determinant of cell velocity in collective migration. *Nature Physics*, 15(8):858–866, 2019.

[173] J. Singh, B. A. Camley, and A. S. Nain. Rules of Contact Inhibition of Locomotion for Cells on Suspended Nanofibers. *bioRxiv*, 2020.

[174] F. Lu, M. Zhong, S. Tang, and M. Maggioni. Nonparametric inference of interaction laws in systems of agents from trajectory data. *Proceedings of the National Academy of Sciences of the United States of America*, 116(29):14424–14433, 2019.

[175] J. Miller, S. Tang, M. Zhong, and M. Maggioni. Learning Theory for Inferring Interaction Kernels in Second-Order Interacting Agent Systems. *Arxiv*, page 2010.03729, 2020.

[176] R. Lukeman, Y. X. Li, and L. Edelstein-Keshet. Inferring individual rules from collective behavior. *Proceedings of the National Academy of Sciences of the United States of America*, 107(28):12576–12580, 2010.

[177] Y. Katz, K. Tunstrøm, C. C. Ioannou, C. Huepe, and I. D. Couzin. Inferring the structure and dynamics of interactions in schooling fish. *Proceedings of the National Academy of Sciences of the United States of America*, 108(46):18720–18725, 2011.

[178] R. Escobedo, V. Lecheval, V. Papaspyros, F. Bonnet, F. Mondada, C. Sire, and G. Theraulaz. A data-driven method for reconstructing and modelling social interactions in moving animal groups. *Phil. Trans. R. Soc. B*, 375:20190380, 2020.

[179] M. Abercrombie. Contact inhibition and malignancy. *Nature*, 281(5729):259–262, 1979.

[180] M. Abercrombie and J. E. Heaysman. Invasiveness of Sarcoma Cells. *Nature*, 174(4432):697–698, 1954.

[181] J. M. Teddy and P. M. Kulesa. In vivo evidence for short- and long-range cell communication in cranial neural crest cells. *Development*, 131(24):6141–6151, 2004.

[182] M. Hayakawa, T. Hiraiwa, Y. Wada, H. Kuwayama, and T. Shibata. Polar pattern formation induced by contact following locomotion in a multicellular system. *eLife*, 9:e53609, 2020.

[183] T. Fujimori, A. Nakajima, N. Shimada, and S. Sawai. Tissue self-organization based on collective cell migration by contact activation of locomotion and chemotaxis. *Proceedings of the National Academy of Sciences of the United States of America*, 116(10):4291–4296, 2019.

[184] T. E. Angelini, E. Hannezo, X. Trepat, J. J. Fredberg, and D. A. Weitz. Cell migration driven by cooperative substrate deformation patterns. *Physical Review Letters*, 104(16):1–4, 2010.

[185] A. Palamidessi, C. Malinverno, E. Frittoli, S. Corallino, E. Barbieri, S. Sigismund, G. V. Beznoussenko, E. Martini, M. Garre, I. Ferrara, C. Tripodo, F. Ascione, E. A. Cavalcanti-Adam, Q. Li, P. P. Di Fiore, D. Parazzoli, F. Giavazzi, R. Cerbino, and G. Scita. Unjamming overcomes kinetic and proliferation arrest in terminally differentiated cells and promotes collective motility of carcinoma. *Nature Materials*, 18:1252–1263, 2019.

[186] Y. L. Han, A. F. Pegoraro, H. Li, K. Li, Y. Yuan, G. Xu, Z. Gu, J. Sun, Y. Hao, S. K. Gupta, Y. Li, W. Tang, H. Kang, L. Teng, J. J. Fredberg, and M. Guo. Cell swelling, softening and invasion in a three-dimensional breast cancer model. *Nature Physics*, 16(1):101–108, 2020.

[187] M. Machacek and G. Danuser. Morphodynamic profiling of protrusion phenotypes. *Biophysical Journal*, 90(4):1439–1452, 2006.

[188] X. Ma, O. Dagliyan, K. M. Hahn, and G. Danuser. Profiling cellular morphodynamics by spatiotemporal spectrum decomposition. *PLoS Computational Biology*, 14(8):1–29, 2018.

[189] G. J. Berman, D. M. Choi, W. Bialek, and J. W. Shaevitz. Mapping the stereotyped behaviour of freely moving fruit flies. *Journal of The Royal Society Interface*, 11(99):20140672–20140672, 2014.

[190] C. K. Chan, A. Hadjitheodorou, T. Y. Tsai, and J. A. Theriot. Quantitative comparison of principal component analysis and unsupervised deep learning using variational autoencoders for shape analysis of motile cells. *bioRxiv*, page 2020.06.26.174474, 2020.

[191] N. Tapon and A. Hall. Rho, Rac and Cdc42 GTPases regulate the organization of the actin cytoskeleton. *Current Opinion in Cell Biology*, 9(1):86–92, 1997.

[192] E. S. Welf and G. Danuser. Using fluctuation analysis to establish causal relations between cellular events without experimental perturbation. *Biophysical Journal*, 107(11):2492–2498, 2014.

[193] K. Lee, H. L. Elliott, Y. Oak, C. T. Zee, A. Groisman, J. D. Tytell, and G. Danuser. Functional Hierarchy of Redundant Actin Assembly Factors Revealed by Fine-Grained Registration of Intrinsic Image Fluctuations. *Cell Systems*, 1(1):37–50, 2015.

[194] T. Isogai and G. Danuser. Discovery of functional interactions among actin regulators by analysis of image fluctuations in an unperturbed motile cell system. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1747), 2018.

[195] 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. *Cellular and Molecular Bioengineering*, 7 (3):293–306, 2014.

[196] 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(6283):353–358, 2016.

[197] A. Lomakin, C. Cattin, J. Garcia-Arcos, I. Zhitnyak, M. Driscoll, E. Welf, R. Petrie, A. Lennon-Duménil, and D. Müller. The nucleus acts as a ruler tailoring cell responses to spatial constraints. 2894(October), 2019.

[198] P. M. Davidson, A. Battistella, T. Dejardin, T. Betz, J. Plastino, B. Cadot, N. Borghi, and C. Sykes. Actin Accumulates Nesprin-2 at the Front of the Nucleus During Confined Cell Migration. *EMBO reports*, page e49910, 2020.

[199] Y. L. Han, P. Ronceray, G. Xu, A. Malandrino, R. Kamm, M. Lenz, C. P. Broedersz, and M. Guo. Cell contraction induces long-ranged stress stiffening in the extracellular matrix. *Proc Natl Acad Sci U S A*, 115(16):4075–4080, 2018.

[200] D. Rüdiger, K. Kick, A. Goychuk, A. M. Vollmar, E. Frey, and S. Zahler. Cell-Based Strain Remodeling of a Nonfibrous Matrix as an Organizing Principle for Vasculogenesis. *Cell Reports*, 32(6), 2020.

[201] M. Gómez-González, E. Latorre, M. Arroyo, and X. Trepat. Measuring mechanical stress in living tissues. *Nature Reviews Physics*, 2(6):300–317, 2020.

[202] W. Kang, J. Ferruzzi, C.-P. Spatarelu, Y. L. Han, Y. Sharma, S. A. Koehler, J. P. Butler, D. Roblyer, M. H. Zaman, M. Guo, Z. Chen, A. F. Pegoraro, and J. J. Fredberg. Tumor invasion as non-equilibrium phase separation. *bioRxiv*, page 2020.04.28.066845, jan 2020.

[203] U. S. Schwarz. Island hopping for cells. *Nature Physics*, 15(6):524–525, 2019.

[204] C.-M. Horejs. Determined to move. *Nature Reviews Materials*, 4(4):225, 2019.

[205] D. Riveline and S. Lo Vecchio. Stochastic Resonance for Individual Cells. *Biophysical Journal*, 118(3): 533–534, feb 2020.

[206] R. Snyderman and E. J. Goetzl. Molecular and cellular mechanisms of leukocyte chemotaxis. *Science*, 213(4510):830–837, aug 1981.

[207] E. Theveneau, L. Marchant, S. Kuriyama, M. Gull, B. Moepps, M. Parsons, and R. Mayor. Collective Chemotaxis Requires Contact-Dependent Cell Polarity. *Developmental Cell*, 19(1):39–53, 2010.

[208] C.-M. Lo, H.-B. Wang, M. Dembo, and Y.-L. Wang. Cell Movement Is Guided by the Rigidity of the Substrate. 79(October 1999):144–152, 2000.

[209] D. E. Discher, P. Janmey, and Y.-L. Wang. Tissue Cells Feel and Respond to the Stiffness of Their Substrate. *Science*, 310(5751):1139–1143, 2005.

[210] R. Sunyer, V. Conte, J. Escribano, A. Elosegui-Artola, A. Labernadie, L. Valon, D. Navajas, J. M. García-Aznar, J. J. Muñoz, P. Roca-Cusachs, and X. Trepat. Collective cell durotaxis emerges from long-range intercellular force transmission. *Science*, 5290(1988):5287–5290, 2016.

[211] M. Dietrich, H. Le Roy, D. B. Brückner, H. Engelke, R. Zantl, J. O. Rädler, and C. P. Broedersz. Guiding 3D cell migration in deformed synthetic hydrogel microstructures. *Soft Matter*, 2018.

[212] D. Caballero, R. Voituriez, and D. Riveline. Protrusion fluctuations direct cell motion. *Biophysical Journal*, 107(1):34–42, 2014.

[213] P. T. Caswell and T. Zech. Actin-Based Cell Protrusion in a 3D Matrix. *Trends in Cell Biology*, 28(10): 823–834, 2018.

[214] R. J. Petrie and K. M. Yamada. At the leading edge of three-dimensional cell migration. *Journal of Cell Science*, 125(24):5917–5926, 2012.

[215] J. Gautrais, F. Ginelli, R. Fournier, S. Blanco, M. Soria, H. Chaté, and G. Theraulaz. Deciphering Interactions in Moving Animal Groups. *PLoS Computational Biology*, 8(9), 2012.

[216] J. Jhawar, R. G. Morris, U. R. Amith-Kumar, M. Danny Raj, T. Rogers, H. Rajendran, and V. Guttal. Noise-induced schooling of fish. *Nature Physics*, 16:488–493, 2020.

[217] A. Cavagna, A. Cimarelli, I. Giardina, G. Parisi, R. Santagati, F. Stefanini, and M. Viale. Scale-free correlations in starling flocks. *Proceedings of the National Academy of Sciences*, 107(26):11865–11870, 2010.

[218] A. Attanasi, A. Cavagna, L. Del Castello, I. Giardina, T. S. Grigera, A. Jelic, S. Melillo, L. Parisi, O. Pohl, E. Shen, and M. Viale. Information transfer and behavioural inertia in starling flocks. *Nature Physics*, 10(9):691–696, 2014.

[219] J. Buhl, D. J. T. Sumpter, I. D. Couzin, J. J. Hale, E. Despland, E. R. Miller, and S. J. Simpson. From Disorder to Order in Marching Locusts. *Science*, 312(5778):1402–1406, 2006.

[220] A. Attanasi, A. Cavagna, L. Del Castello, I. Giardina, S. Melillo, L. Parisi, O. Pohl, B. Rossaro, E. Shen, E. Silvestri, and M. Viale. Collective Behaviour without Collective Order in Wild Swarms of Midges. *PLoS Computational Biology*, 10(7):e1003697, 2014.

[221] Y. Su, J. Li, C. Shi, R. H. Hruban, and G. L. Radice. N-cadherin functions as a growth suppressor in a model of K-ras-induced PanIN. *Oncogene*, 35(25):3335–3341, jun 2016.

[222] V. Padmanaban, I. Krol, Y. Suhail, B. M. Szczerba, N. Aceto, J. S. Bader, and A. J. Ewald. E-cadherin is required for metastasis in multiple models of breast cancer. *Nature*, 573(7774):439–444, 2019.

[223] R. M. Lee, M. I. Vitolo, W. Losert, and S. S. Martin. Distinct Roles of Tumor-Associated Mutations in Collective Cell Migration. *bioRxiv*, 2020.

Chapter 3

Stochastic nonlinear dynamics of confined cell migration

This chapter is based on the following publication:

Stochastic Nonlinear Dynamics of Confined Cell Migration in Two-State Systems

David B. Brückner*, Alexandra Fink*, Christoph Schreiber, Peter J. F. Röttgermann, Joachim O. Rädler†, Chase P. Broedersz†

* equal contribution
† corresponding author

Nature Physics 15, 595-601 (2019)

Stochastic Nonlinear Dynamics of Confined Cell Migration in Two-State Systems

David B. Brückner[1]*, Alexandra Fink[2]*, Christoph Schreiber[2], Peter J. F. Röttgermann[2], Joachim O. Rädler[2+], Chase P. Broedersz[1+]

[1] Arnold Sommerfeld Center for Theoretical Physics and Center for NanoScience, Ludwig-Maximilians-Universität, München
[2] Faculty of Physics and Center for NanoScience, Ludwig-Maximilians-Universität, München

* authors contributed equally
[+] corresponding authors

Migrating cells in physiological processes, including development, homeostasis and cancer, encounter structured environments and are forced to overcome physical obstacles. Yet, the dynamics of confined cell migration remains poorly understood, and thus there is a need to study the complex motility of cells in controlled confining micro-environments. Here, we develop two-state micropatterns, consisting of two adhesive sites connected by a thin constriction, in which migrating cells perform repeated stochastic transitions. This minimal system enables us to obtain a large ensemble of single cell trajectories. From these trajectories, we infer an equation of cell motion, which decomposes the dynamics into deterministic and stochastic contributions in position-velocity phase space. Our results reveal that cells in two-state micropatterns exhibit intricate non-linear migratory dynamics, with qualitatively similar features for a cancerous (MDA-MB-231) and non-cancerous (MCF10A) cell line. In both cases, the cells drive themselves deterministically into the thin constriction; a process that is sped up by noise. Interestingly however, these two cell lines have distinct deterministic dynamics: MDA-MB-231 cells exhibit a limit cycle, while MCF10A cells show excitable bistable dynamics. Our approach yields a conceptual framework that may be extended to understand cell migration in more complex confining environments.

In all stages of life and death, from embryogenesis and immune response to cancer, migrating cells are key players[1–3]. Single cells in these systems face a common physical challenge: Both in two- and three-dimensional systems, such as epithelial sheets[2,4] or extra-cellular matrices[5,6], cells migrate through complex confining environments, shaped by the surrounding tissue. Navigating such environments requires the cell to overcome physical

obstacles such as squeezing through thin pores[4,7–9]. Cell motility is powered by a complex cytoskeletal machinery containing a vast number of interacting molecular constituents that are subject to intrinsic noise. Nonetheless, at a larger scale, cells reliably perform vital motility functions. However, a system-level understanding of the emerging migratory behaviour in response to defined spatially structured environments has not yet been achieved. Thus, a fundamental question arises: Does cell migration, when confined to a well-defined geometry, exhibit emergent dynamical 'laws', and what are the roles of deterministic and stochastic contributions to these dynamics?

For over a century, the trajectories of migrating cells have been analysed quantitatively[10,11]. Such studies have led to the conceptual picture that on uniform two-dimensional surfaces, cells perform persistent random motion[5,12–14]. In recent years, micropatterning techniques have been developed[15–17], which confine cells to areas of a well-defined shape. By monitoring cells migrating on one-dimensional tracks[18–20], microratchets[21,22], and other geometries[23,24], useful measures such as cell speed and persistence have been established and used to quantify cell behaviour. In this context, confined cell migration has been considered a biophysical analogue of a particle bound to a confining potential[25–27]. However, it is unclear whether a potential landscape picture is adequate to describe a confined migrating cell. The search for simple laws that underlie cell migration is exacerbated by the intrinsically variable nature of living cells. Thus, to achieve a system-level understanding of confined cell migration, we need an approach that can disentangle the deterministic and stochastic contributions to the dynamics.

Here, we develop a theoretical framework that describes the stochastic migration of cells in structured environments. To this end, we designed a micropattern with two square adhesive islands connected by a thin bridge, yielding a minimal two-state system to investigate confined cell migration. We find that for all cell lines studied, in particular cancerous (MDA-MB-231) and non-cancerous (MCF10A) cells, this microenvironment leads to the emergence of a distinct migratory behaviour in the form of frequent stochastic transitions between the two islands. By inferring a stochastic equation of motion from the recorded short time-scale dynamics, we decompose the motion into deterministic and stochastic contributions. The resulting equation quantitatively captures various statistics of the cellular dynamics. Our analysis reveals that the cells generate a deterministic driving into the narrow connecting bridge. This driving emerges from intricate non-linear deterministic dynamics poised near the cross-over between a limit cycle and a bistable system. The intrinsic stochasticity of the system accelerates the transitions. Our approach yields an insightful representation of the system's behavioural dynamics, which may provide a basis for the understanding of the microscopic processes driving cell migration as well as the dynamics of cells in more complex environments.

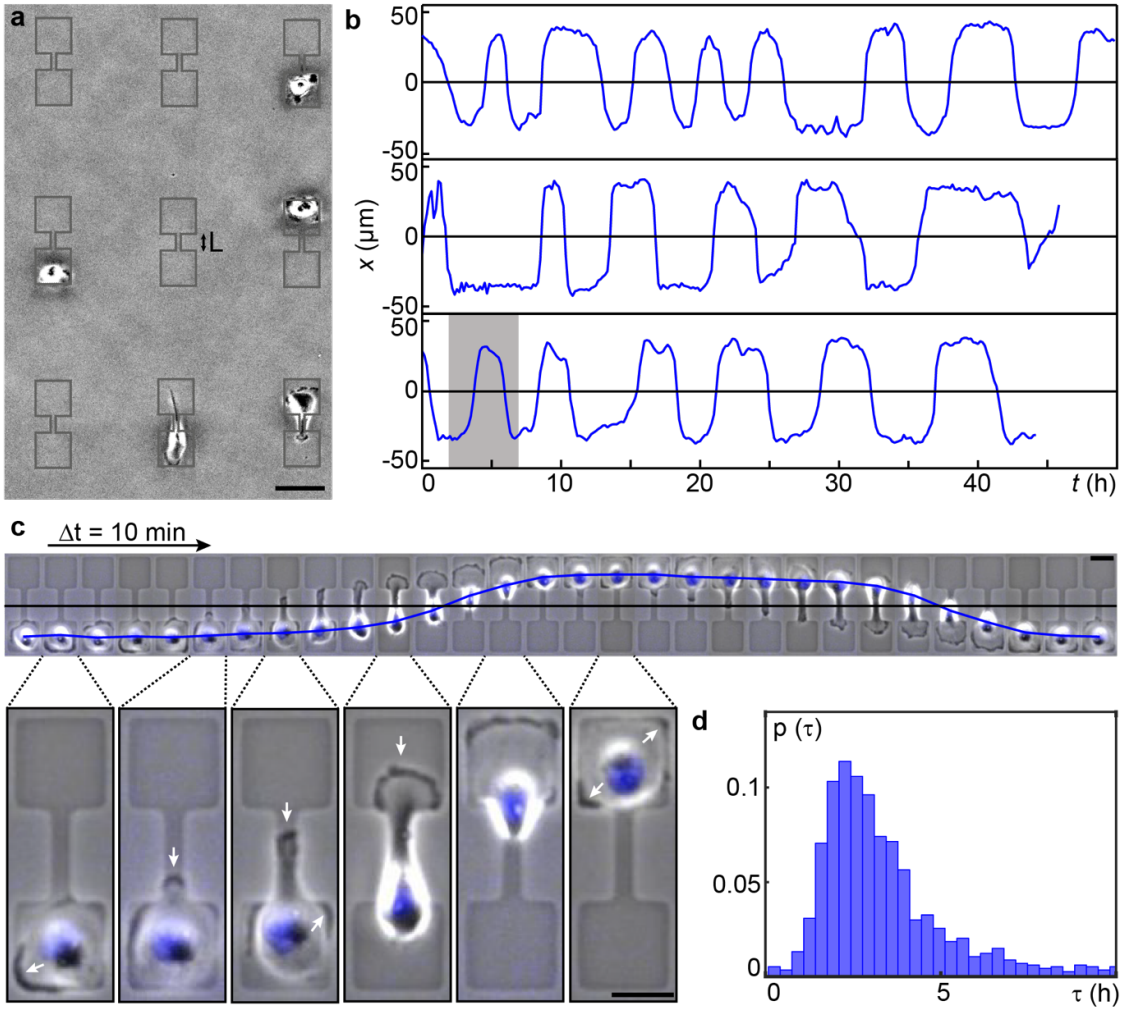


Figure 1 | **Experimental setup of two-state micropatterns with MDA-MB-231 cells. a**, Overview of single cells on a microstructured surface. The image is de-noised and the micropattern outline is shown in grey. The fibronectin-coated micropatterns consist of two square islands (37 µm x 37 µm), which are connected by bridges of varying lengths (L = 6-56 µm) (scale bar: 50 µm). **b**, Exemplary trajectories (along the long axis of the micropattern) of the nucleus' position plotted against time. **c**, Exemplary time series of a single cell track on a micropattern with bridge length L = 35 µm, corresponding to the section highlighted in grey in **b** (scale bar: 25 µm; time between images: 10 min). Key stages of the transition process are shown in the zoom in, with white arrows indicating regions of pronounced actin activity. **d**, Probability distribution of the dwell times τ (defined as the time between subsequent transitions of the cell nucleus across the bridge centre) of 149 cells (bin size 0.34 h).

Cells perform stochastic transitions on two-state micropatterns

We use time-lapse phase contrast microscopy to study confined migration of human metastatic breast cancer cells (MDA-MB-231) on two-state micropatterns (Fig. 1c, Supplementary Movies S1-S4). These micropatterns consist of two square adhesive islands connected by a thin bridge that are fibronectin-coated to promote cell adhesion, while the

surrounding area is passivated with cell-repellent PEG-PLL (Fig. 1a). Initially, single cells adhere to the micropattern and spread on one of the adhesive islands. However, cells remain highly motile within this confinement: Exploratory lamellipodia-like protrusions repeatedly form along the cell periphery, appearing as actin-driven activity in Lifeact-GFP transfected cells (see Supplementary Movies S5, 6) and dark regions in phase contrast imaging (Figure 1c). These short-lived protrusions form most frequently at the island corners. In contrast, lamellipodia appearing at the bridge entrance can grow into sustained protrusions: while some quickly retract, others reach the opposite, unoccupied island. Here, the protrusion broadens into a fan-like shape, followed by the transitioning of the cell body. After fully transitioning, the actin cytoskeleton appears to reorganise and the migration pattern repeats in the reverse direction, but with a large variability in time (Fig. 1c). The dwell time – the time between subsequent transitions of the cell nucleus across the bridge centre – displays a broad distribution (Fig. 1d). While the width of this distribution is partially due to the intrinsic cell-to-cell heterogeneity, we find that even for a single cell, the dwell times exhibit a large variability over time (Fig. 1b, Supplementary Section S2.5). The cellular 'hopping' behaviour therefore constitutes a stochastic process, which appears to be a generic migratory pattern also observed for various other cell lines (Supplementary Section 2.2).

An intuitive measure of the stochasticity is given by the survival probability: the probability that the cell has *not* transited after a given time. The survival probability decreases monotonically with time, as it becomes increasingly unlikely that a cell remains on the same island (Fig. 2a). For a Poisson process, the survival probability decays exponentially. By contrast, here we observe a prominent plateau at small times indicating that cells are less likely to immediately reverse after a transition, followed by a tail in the distribution that also exhibits non-exponential features (see Supplementary Fig. S7).

To explore the universal properties of the hopping process, we vary the length of the connecting bridge. We find that the survival probability function broadens with increasing length, indicating longer dwell times on the islands (Fig. 2a). Remarkably, when we rescale time by the average dwell time $\langle\tau\rangle$, all survival distributions approximately collapse onto a master curve (Inset Fig. 2a). This collapse suggests that, as we vary the micropattern dimensions, the cell migration remains to be governed by similar stochastic dynamics, with a characteristic timescale $\langle\tau\rangle$ set by the bridge length.

This timescale $\langle\tau\rangle$ varies linearly with the length of the connecting bridge (Fig. 2b). One might intuitively expect such a linear dependence, since the cells tend to migrate across the thin bridge nearly ballistically, so that the traversal time increases linearly with distance. However, the nucleus traversing the bridge is not the main time-limiting process and can only account for approximately 30% of the increase in dwell time (Supplementary Fig. S4). Thus, cells on

micropatterns with longer bridges also spend more time on the adhesive islands before performing transitions. These results indicate that there is an intricate stochastic process governing the hopping behaviour.

Having established the key observables of the hopping process, we next consider the properties of the cell trajectories themselves. The joint probability distribution $p(x, v)$ of position x and velocity v exhibits a double peaked structure in phase space, with highest occupation and lower speeds on the adhesive islands (Fig. 2c, i, ii). The marginal velocity probability distribution $p(v)$ further has a markedly non-Gaussian shape (Fig. 2c,iii), as observed in prior studies on various cell types[5,28–30]. These qualitative features emerge robustly on all bridge lengths that we have investigated (Supplementary Section S6).

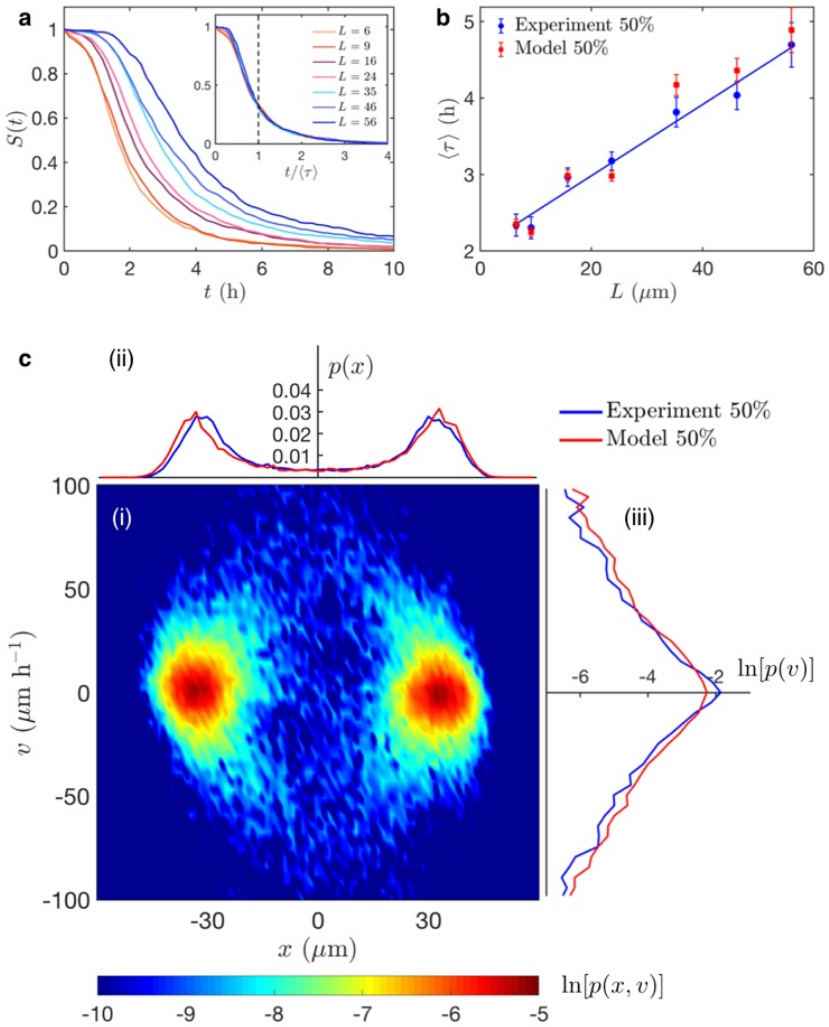


Figure 2 | Statistics of the hopping process (MDA-MB-231). a, Survival probability as a function of time for each bridge length (listed in units of µm). Inset: Distributions collapse on a master curve when

time is rescaled by the average dwell time $\langle\tau\rangle$. **b**, Average dwell time as a function of bridge length. Blue dots are averages calculated from 50% of the experimental data, while red squares correspond to predictions of the model trained on the excluded 50% of data. Error bars denote errors of the mean, obtained by bootstrapping. **c**,i, Joint probability density of position and velocity of cells on patterns with bridge length *L* = 35 µm, plotted logarithmically. Experimental (blue) and predicted (red) probability distributions of position (*ii*) and velocity (*iii*, plotted semi-logarithmically). As in **b**, experimental and predicted data are based on either half of the split data.

Confined cell migration is governed by a nonlinear stochastic equation of motion

The experimentally recorded statistics enable us to develop a quantitative framework for stochastic cell migration in structured environments. To this end, we generalize the equation of persistent random motion often used to describe 2D cell migration[5,12,13]. In particular, we postulate that the dynamics of the cell position x and velocity $v = \mathrm{d}x/\mathrm{d}t$ can be described by the stochastic equation of motion

$$\frac{dv}{dt} = F(x,v) + \sigma(x,v)\eta(t) \qquad (1)$$

where $\eta(t)$ represents Gaussian white noise with $\langle\eta(t)\rangle = 0$ and $\langle\eta(t)\eta(t')\rangle = \delta(t-t')$. Note that, unlike in persistent random motion, we also allow the noise strength $\sigma(x,v)$ to depend on the state of the system. A priori, there is no fundamental reason why this approach should succeed, since the dynamics of a cell in a structured environment might require a dynamical description more complex than that given by equation (1), including a memory kernel formulation as for free 2D migration[13,29,30] (Supplementary Section S4).

To investigate if the cell migration on our micropatterns can be captured by the basic stochastic dynamics in equation (1), we infer the terms in this equation for each bridge length. Thus, we adopt a data-driven approach, where we reconstruct the deterministic driving $F(x,v) = \langle\dot{v}|x,v\rangle$ (Fig. 3a) and the noise strength $\sigma(x,v) = \sqrt{\Delta t\langle[\dot{v} - F(x,v)]^2|x,v\rangle}$ (Fig. 3d) from measured trajectories (see Methods). Importantly, in this approach, we do not directly use any information about processes at long time-scales such as the transitions between sites; the deterministic and stochastic contributions to the equation of motion are completely constrained by the short-time scale behaviour of the measured trajectories. As a self-consistency test, we calculate the correlation function $\langle\Delta W(t+\theta)\Delta W(t)\rangle$ of the inferred noise $\Delta W(t) = \int_t^{t+\Delta t}\eta(t)\,\mathrm{d}t$. The correlation indeed decays within the experimental temporal resolution $\Delta t = 10\,\mathrm{min}$, confirming our initial white noise assumption on this timescale (Fig. 3g). Note, F and σ are derived from discrete-time data and should thus be interpreted within a finite difference form of equation (1), although the estimated continuous-time deterministic term[31] is very similar (Supplementary Section S3.6).

Next, we test the predictive capability of our model by generating trajectories from the inferred equation of motion (Fig. 3h). Remarkably, we find very good quantitative agreement of the dwell times that characterize cell hopping (Fig. 2b). Note that the model is trained on a random sample of 50% of the data, which is excluded from the experimental data shown in the graph. Similarly, we find good agreement in the steady-state probability distributions of position and velocity (Fig. 2c, ii, iii), the velocity correlation function (Fig. 3g), and the survival probability distributions (Supplementary Fig. S32). These observations hold similarly for all other bridge lengths as well as for asymmetric two-state systems (Supplementary Section S6), confirming that our model (equation (1)) reproduces the key statistical observables of confined cell migration in this setup.

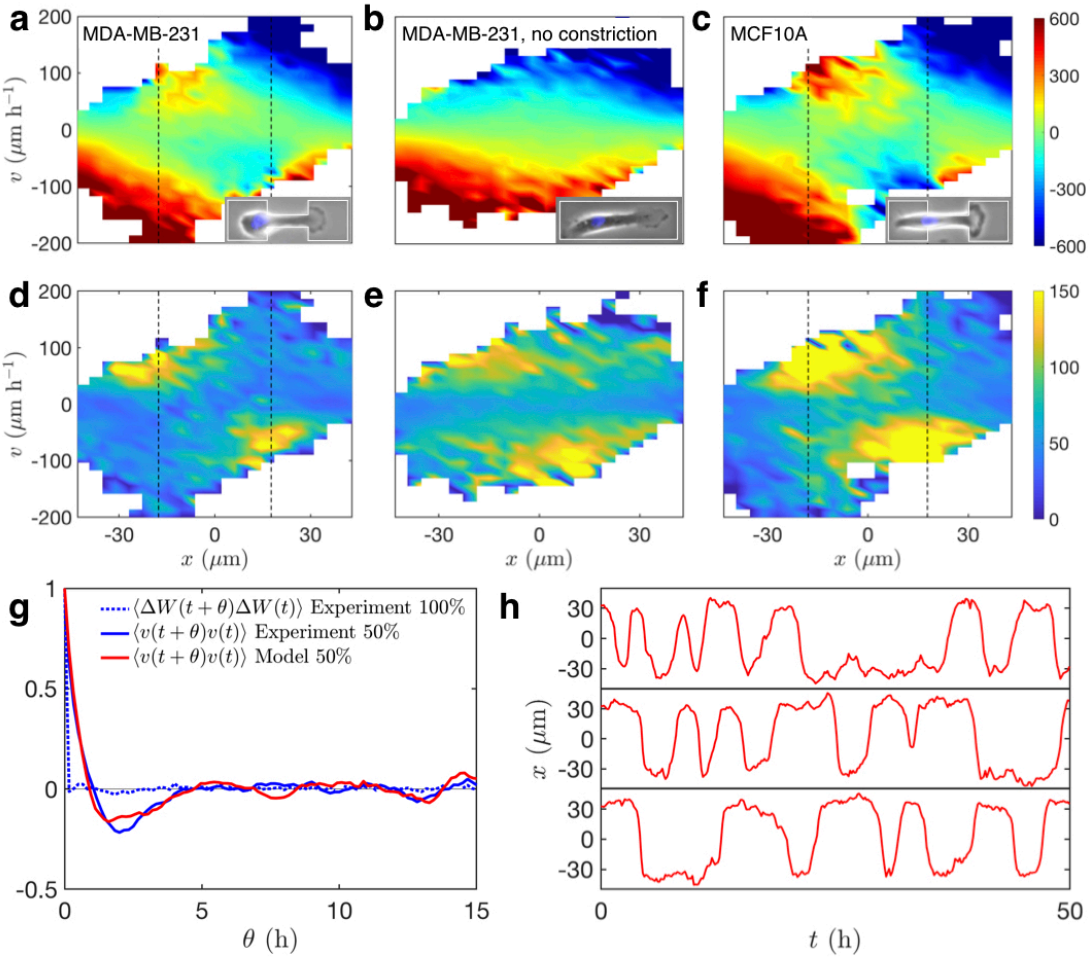


Figure 3 | Deterministic and stochastic contributions to the equation of motion for three systems: MDA-MB-231 cells with and without constriction, and MCF10A cells with constriction (L = 35 µm). **a**, **b**, **c**, The deterministic component $F(x, v)$ in units of µm/h^2. Insets show a phase contrast image of a cell in the micropattern geometry, with the pattern outlined in white. **d**, **e**, **f**, The noise strength

$\sigma(x, v)$ in units of μm/h$^{3/2}$. Dashed black lines correspond to the positions $x = \pm L/2$. Both F and σ are shown here as a linear interpolation. **g**, The correlation function of the inferred noise $\Delta W(t) = \int_t^{t+\Delta t} \eta(t)\,\mathrm{d}t$ (dotted blue line) together with the experimental velocity correlation function (solid blue line) and the associated model prediction (red line) for MDA-MB-231 cells. The experimental sampling interval is $\Delta t = 10$ min. **h**, Sample trajectories generated by the MDA-MB-231 model.

Cell dynamics in a two-state system resembles a non-linear oscillator

Our stochastic model gives unique insight into the system-level dynamics that govern the migration of MDA-MB-231 cells. The inferred deterministic term $F(x, v)$ has an intricate dependence that is additively non-separable in position and velocity (Fig. 3a). This observation excludes a dynamical picture in which the cells move in a potential landscape, which also provides predictions inconsistent with experiments (Supplementary Section S5). As expected, the symmetry of the confining microenvironment is reflected by the symmetry of the dynamical terms, $F(x, v) = -F(-x, -v)$ and $\sigma(x, v) = \sigma(-x, -v)$. The noise strength $\sigma(x, v)$ appears approximately constant except for two regions of elevated noise amplitude in the upper left and lower right quadrants of phase space (Fig. 3d). These regions correspond to states where the cell appears to initiate a transition across the bridge by developing a lamellipodium (Fig. 1c, Supplementary Movies S5, 6), suggesting increased fluctuations associated with these cellular dynamics.

The deterministic contribution of the equation of motion, defined by $\dot{v} = F(x, v)$, provides a conceptual understanding of the dynamics of confined cell migration. Remarkably, we find that even without noise, the inferred dynamical system exhibits regular transitions (Fig. 4a), implying that the transition behaviour is driven deterministically. Interestingly, if the deterministic system is initialised at different points in phase space, the trajectories always approach the same stable limit cycle; the deterministic dynamics encodes a self-sustained oscillation with a characteristic amplitude, period and waveform, independent of initial conditions. To contrast these findings, we perform experiments with patterns without the thin constriction (Supplementary Movie S7, Fig. 3b, e). In this case, the deterministic dynamics always relaxes to $v = 0$, constituting a line of stable fixed points (Fig. 4b).

To further understand these dynamics, we consider a cut through the deterministic contribution at $x = 0$. As a function of velocity, the deterministic term represents an effective frictional contribution (Fig. 4d). For cells migrating in confinement without constriction, this dependence is non-linear, and coincides with that of free 2D migration. This is in contrast to the case of simple persistent random motion (Ornstein-Uhlenbeck process), where this dependence is linear, $F \propto -v$. Strikingly, the presence of the constriction leads to a sign switch of F at low speeds $|v| < 50\ \mu\mathrm{m/h}$, giving rise to 'negative friction'. This phenomenon

is similar to the classic example of the van der Pol oscillator[32], but with important differences: In the van der Pol system, the sign and magnitude of the friction are controlled by position (Supplementary Section S5.4). By contrast, in our system, the onset of negative friction is dictated by velocity.

The qualitative structure of the cell's deterministic dynamics is characterised by the nullclines, corresponding to purely vertical ($\dot{x} = 0$) or horizontal flow ($\dot{v} = 0$) in phase space (Fig. 4e). The v-nullcline exhibits an intricate, non-linear shape: It intersects the x-nullcline at the origin, giving rise to an unstable fixed point, and partitions the upper half of phase space ($v > 0$) into distinct regions with different flow directions (arrows in Fig. 4e). Interestingly, the v-nullcline almost intersects the x-axis at the positions corresponding to the adhesive islands, suggesting that the system is poised close to a transition towards bistability. These results emerge robustly on all bridge lengths and are independent of the details of the inference procedure (Supplementary Sections S3, S6). Taken together, our results reveal that MDA-MB-231 cells exhibit dynamics analogous to a non-linear oscillator that generates deterministic forces to migrate across the "hurdle" imposed by the bridge.

Non-cancerous (MCF10A) cells exhibit excitable bistable dynamics

To explore the generality of these results, we also analysed the behaviour of non-cancerous breast cells (MCF10A). The deterministic and stochastic terms of the equation of motion exhibit similar qualitative features as for MDA-MB-231 cells (Fig. 3c, f, 4d). Interestingly however, the inferred MCF10A model does not exhibit a limit cycle, but two stable fixed points on either side of the bridge (Fig. 4c). The basins of attraction of these fixed points extend all the way to the other side of the micropattern, indicating that the bistable dynamics of MCF10A cells is excitable: a small noise-induced perturbation can result in a rapid deterministic excursion to the other side of the system.

Noise-induced excitation and deterministic amplification drive cellular transitions

The common features of the cellular response to thin constrictions are revealed by the deterministic flow in phase space: Both MDA-MB-231 and MCF10A cells exhibit a deterministic driving that controls the traversal of the thin constriction (orange arrows in Fig. 4a, c). In contrast, in the region of phase space where the cell reverses, the noise is much larger than the deterministic acceleration (marked in green in Fig. 4e, Supplementary Fig. S18). Beyond this noise-dominated region, the deterministic driving leads to an amplification of velocity, ensuring a rapid transition to the other side of the micropattern. For MDA-MB-231 cells migrating in confinement without constriction, there is no amplification of velocity, indicating that this may be a cellular response to the presence of the constriction (Fig. 4b). Beyond the centre of the constriction, the cell deterministically decelerates (relaxation) and reaches the other adhesive island.

The noise in the model plays a crucial role in speeding up the slow dynamics of reversals on the islands in the limit cycle dynamics of MDA-MB-231 cells and enables MCF10A cells to escape the stable fixed points. Thus, beyond abrogating long-time correlations (Supplementary Section 3.9), the noise reduces the key time-scale of the system: the average dwell time (Fig. 2b).

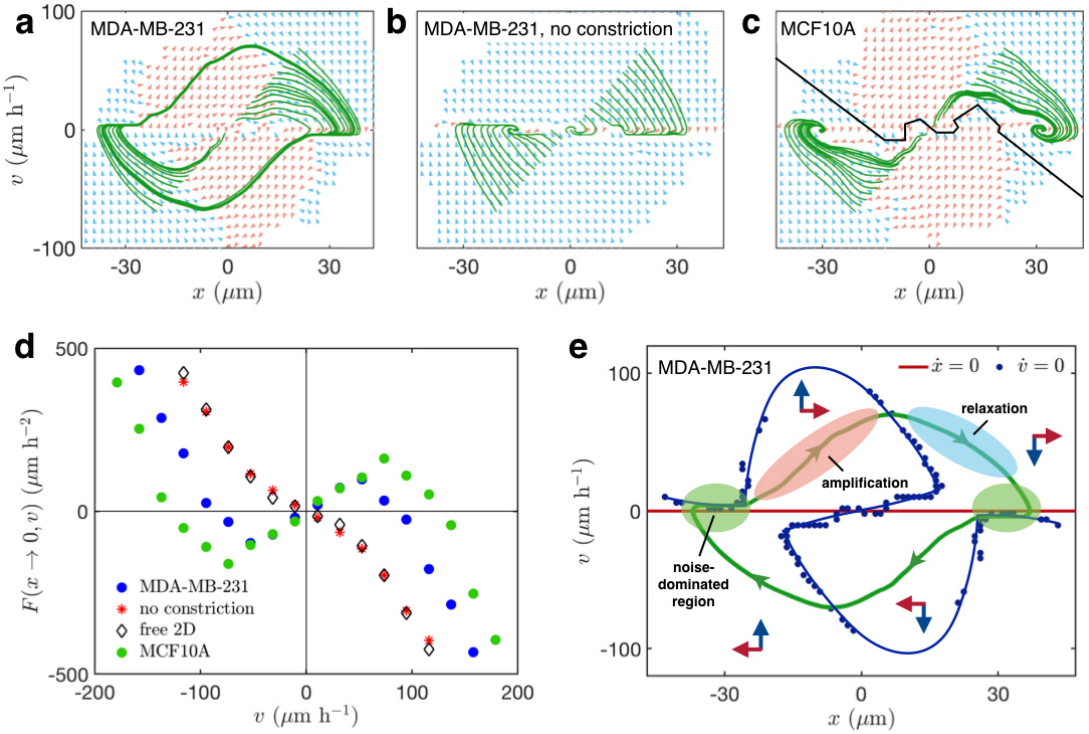


Figure 4 | **Nonlinear deterministic dynamics of the cell migration (*L* = 35 µm). a**, Trajectories (green) of the deterministic dynamics for a number of different initial conditions (MDA-MB-231). The flow field is shown by arrows, where accelation is orange and deceleration is blue. **b**, The same plot for MDA-MB-231 cells migrating in a system without constriction (See Fig. 3b). **c**, Deterministic trajectories and flow field of MCF10A cells. The black line indicates the divide between the basins of attraction of the two fixed points. **d**, The frictional component $F(x \to 0, v)$ is a non-linear function of velocity, here determined at the center of the system in a small interval around $x = 0$, for the three cases in **a**-**c**. Black diamonds show the velocity-dependent acceleration $F(v) = \langle \dot{v} | v \rangle$ of MDA-MB-231 cells migrating freely in 2D. **e**, The nullclines $\dot{x} = 0$ (red line) and $\dot{v} = 0$ (blue dots) for MDA-MB-231. The blue line is a guide to the eye for the v-nullcline. Red and blue arrows indicate the local direction of the flow, corresponding to the signs of $\dot{x}$ and $\dot{v}$, respectively. The flow gives rise to three distinct dynamical regimes: First, cells exit the slow, noise-dominated region on the adhesive island (blue area). This process is accelerated by noise-induced excitations. Subsequently, the velocity is deterministically amplified (green area) and finally the system deterministically relaxes (red area) onto the slow region on the other adhesive island. The results in **d**, **e** are obtained from a symmetrised version of the deterministic term (Supplementary sections S3.4 and S6).

A data-driven dynamical systems analysis of confined cell migration

We employed a high-throughput approach using arrays of two-state micropatterns to measure a large set of cell trajectories in a standardised confining environment. This approach enabled us to infer an equation of motion with deterministic and stochastic contributions that reproduces the observed migration statistics. Previous studies have argued that the accurate description of cell migration on two-dimensional surfaces requires complex equations of motion, including a memory kernel[13,29,30], time-dependent terms[14], fractional diffusion equations[33], or switching between modes of movement[34]. Here, we find that the migration of a confined cell is captured by a relatively simple second order Langevin equation with a two-dimensional non-linear deterministic term $F(x, v)$. We specifically ruled out simpler models, such as a first order equation of motion with white noise (Supplementary Section S5.1). Furthermore, we introduced a state-dependent noise $\sigma(x, v)$, because a constant noise amplitude does not quantitatively capture the statistics (Supplementary Section S5.3).

The deterministic and stochastic components of cell motion we extract may serve as a phenotypic characterisation of cell dynamics in confinement. The non-linear deterministic dynamics constitute novel cell behaviour, which emerges in response to the thin constriction in the two-state micropattern: we find limit cycle oscillations for MDA-MB-231 cells and excitable bistable dynamics for MCF10A cells, which are known to be less invasive[35,36]. In both cases, there is a deterministic tendency to invade the thin bridge of the micropattern, suggesting that such constrictions can provide guidance cues to cells[37,38]. For MDA-MB-231 cells, we find that the non-linear deterministic dynamics leads to self-sustained oscillations across the narrow constriction of the micropattern. This is in contrast with the intuitive expectation that the hopping behaviour might be generated by a noisy cellular activity competing with an effective energy barrier placed by the bridge. The transition dynamics we observe on the two-state pattern may relate to cellular transitions in more complex confining environments, such as cells squeezing through pores of a biopolymer meshwork[3] or the characteristic T1 transitions in jammed cell sheets[27,39,40]. Models of such processes frequently employ the language of energy potentials[26,27,40,41], which we quantitatively ruled out for models of the form of equation (1) in our system (Supplementary Section S5).

Our theoretical framework quantitatively captures distinct dynamical features of cell locomotion, and reveals how cellular dynamics adapts in response to physical obstacles. The characteristic dynamics of cells on two-state patterns may also provide an alternative approach to quantifying the migration of different cancer cell lines[42]. Finally, our top-down approach could help advance our understanding of locomotion at the molecular level, as it provides strong constraints for bottom-up models[25,43,44] that connect microscopic rules to the system-level dynamics of cells.

Methods

Micropatterning and sample preparation

For micropatterning, we employ microscale plasma-induced protein patterning[45]. Using photolithography, silicon masters bearing the desired shapes are prepared. Polydimethylsiloxane (Sylgard 184 1:10, Dow Corning) stamps are cast from these masters, placed in an ibidi µ-dish (ibidi GmbH) and exposed to oxygen plasma. For subsequent background passivation, a drop of 2 mg ml^{-1} PLL(20)-g[3.5]-PEG(2) (SuSoS) solution is added for 25 min. Afterwards, the sample is rinsed, the stamps removed, and the sample incubated with a 50 µg ml^{-1} human fibronectin (YO Proteins) solution for 50 min. Samples are stored in PBS at 4°C until cell seeding.

Cell culture

MDA-MB-231 human breast carcinoma epithelial cells (DSMZ) are cultured in Minimum Essential Medium (MEM, c.c. pro), supplemented with 10% FBS (Gibco) and 2mM L-Glutamine (c.c. pro). Cells are grown at 37°C up to 70-90% confluence, in an atmosphere with 5% CO_2, before being washed and trypsinised for 3 min. For experiments, the cell solution is centrifuged at 1000 rcf for 3 min, and then cells are re-suspended in MEM. Approximately 10,000 cells are added per µ-dish and left to adhere in the incubator. After 4h, the medium is exchanged to L-15 medium with L-glutamine (Gibco, supplemented with 10% FCS) which contains 25 nM Hoechst 33342 (Invitrogen) for staining of cell nuclei.

MCF10A cells (ATCC) are cultured at 5% CO_2 and at 37°C in DMEM/F-12 medium including Glutamax (Gibco) supplemented with 5% Horse Serum (Thermo Fisher), 20 ng/ml hEGF (Sigma), 500ng/ml Hydrocortisone (Sigma), 100ng/ml Cholera Toxin (Sigma) and 10 µg/ml Insulin (Sigma). For passaging, the supernatant is aspirated and centrifuged at 300rcf for 8 minutes. Cells are washed with PBS and detached by incubation with Accutase for 12 minutes at 37°C. After re-suspension with medium, the cell solution is centrifuged at 500rcf for 6 minutes. Cell pellets are resuspended in medium. For experiments, 10000 cells are seeded per µ-dish and left to adhere for at least 4h. Then, the medium is exchanged to culture medium with all supplements without phenol red. For nuclear staining, 15 nM Hoechst 33342 are added. During experiments, cells are kept in a 5% CO2-atmosphere and at 37°C.

Microscopy and Cell Tracking

Measurements are performed in time-lapse mode for up to 50 h on an IMIC digital microscope (TILL Photonics) or on a Nikon Eclipse Ti microscope. The samples are placed in a heated chamber (ibidi GmbH or Okolab) and kept at 37°C throughout the measurements. Brightfield and fluorescence images of the stained nuclei are acquired every 10 min. A band pass filter is applied to the images of the nuclei, then images are binarised and centre-of-mass positions are determined with *ImageJ*'s Analyze Particles plugin[46]. Due to the plasma treatment the two-state patterns are visible in the brightfield images and are thus used to manually determine the reference boundary of a pattern. For further details, please refer to Supplementary Section S1.

Theoretical Analysis

The position of the cell nucleus $x(t)$ is recorded at a time interval Δt = 10 min in the experiment. Velocity and acceleration are directly calculated as numerical derivatives. The terms $F(x,v)$ and $\sigma(x,v)$ of equation (1) are inferred by conditional averaging of the experimental trajectories[47–49]. By construction, the noise term averages to zero, $\langle\sigma(x,v)\eta(t)\rangle = 0$, and therefore the deterministic term is given by $F(x,v) = \langle\dot{v}|x,v\rangle$. The stochastic term is then estimated using $\sigma^2(x,v) = \Delta t\langle[\dot{v} - F(x,v)]^2|x,v\rangle$. The inferred model is subsequently integrated numerically with a time step Δt equal to the experimental sampling interval. To compare data from model and experiment, we use a hold-out method where we train the model on 50% of the recorded cell trajectories, which is then compared to the other 50% of the data. Both the training and validation set are randomly sampled. For more details, see Supplementary Section S3.

Acknowledgements

We thank E. Frey, F. Brauns, G. Gradziuk, D. Lubensky, P. Ronceray, K. Bassler and N. Wingreen for useful comments, C. Leu for the preparation of wafers, and A. Reiser for providing the transfection protocol. This work was supported by grants from the German Science Foundation (DFG) through the Collaborative Research Center (SFB) 1032 (Projects B01 and B12). D.B.B. is supported by a DFG fellowship within the Graduate School of Quantitative Biosciences Munich (QBM) and by the Joachim Herz Stiftung.

Author Contributions

A.F., C.S., P.J.F.R. and J.O.R designed experiments; A.F. and C.S. performed experiments; D.B.B., A.F. and C.S. analysed data. D.B.B. and C.P.B. developed the theoretical model. D.B.B., A.F., C.S., J.O.R. and C.P.B. wrote the manuscript.

References

1. Koser, D. E. *et al.* Mechanosensing is critical for axon growth in the developing brain. *Nat. Neurosci.* **19,** 1592–1598 (2016).
2. Vedula, S. R. K. *et al.* Forces driving epithelial wound healing. *Nat. Phys.* **10,** 683–690 (2014).
3. Friedl, P. & Wolf, K. Tumour-cell invasion and migration: diversity and escape mechanisms. *Nat. Rev. Cancer* **3,** 362–74 (2003).
4. Park, J. A. *et al.* Unjamming and cell shape in the asthmatic airway epithelium. *Nat. Mater.* **14,** 1040–1048 (2015).
5. Wu, P.-H., Giri, A., Sun, S. X. & Wirtz, D. Three-dimensional cell migration does not follow a random walk. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 3949–54 (2014).
6. Even-Ram, S. & Yamada, K. M. Cell migration in 3D matrix. *Curr. Opin. Cell Biol.* **17,** 524–532 (2005).
7. Green, B. J. *et al.* Pore shape defines paths of metastatic cell migration. *Nano Lett.* **18,** 2140−2147 (2018).
8. Paul, C. D., Mistriotis, P. & Konstantopoulos, K. Cancer cell motility: Lessons from migration in confined spaces. *Nat. Rev. Cancer* **17,** 131–140 (2017).
9. Wolf, K. *et al.* Physical limits of cell migration: Control by ECM space and nuclear deformation and tuning by proteolysis and traction force. *J. Cell Biol.* **201,** 1069–1084 (2013).
10. Przibram, K. Über die ungeordnete Bewegung niederer Tiere. *Pflügers Arch. Physiol.* **153,**

401–405 (1917).
11. Fürth, R. Die Brownsche Bewegung bei Berücksichtigung einer Persistenz der Bewegungsrichtung. Mit Anwendungen auf die Bewegung lebender Infusorien. *Z. Phys.* **2,** 244–256 (1920).
12. Gail, M. H. & Boone, C. W. The Locomotion of Mouse Fibroblasts in Tissue Culture. *Biophys. J.* **10,** 980–993 (1970).
13. Selmeczi, D., Mosler, S., Hagedorn, P. H., Larsen, N. B. & Flyvbjerg, H. Cell motility as persistent random motion: theories from experiments. *Biophys. J.* **89,** 912–931 (2005).
14. Metzner, C. *et al.* Superstatistical analysis and modelling of heterogeneous random walks. *Nat. Commun.* **6,** 7516 (2015).
15. Singhvi, R. *et al.* Engineering Cell Shape and Function. *Science* **264,** 696–698 (1994).
16. Chen, C., Mrksich, M., Huang, S., Whitesides, G. M. & Ingber, D. E. Geometric control of cell life and death. *Science* **276,** 1425–1428 (1997).
17. Thery, M. *et al.* Anisotropy of cell adhesive microenvironment governs cell internal organization and orientation of polarity. *PNAS* **103,** 19771–19776 (2006).
18. Maiuri, P. *et al.* Actin flows mediate a universal coupling between cell speed and cell persistence. *Cell* **161,** 374–386 (2015).
19. Schreiber, C., Segerer, F. J., Wagner, E., Roidl, A. & Rädler, J. O. Ring-Shaped Microlanes and Chemical Barriers as a Platform for Probing Single-Cell Migration. *Sci. Rep.* **6,** 26858 (2016).
20. Prentice-Mott, H. V. *et al.* Directional memory arises from long-lived cytoskeletal asymmetries in polarized chemotactic cells. *Proc. Natl. Acad. Sci.* **113,** 1267–1272 (2016).
21. Caballero, D., Voituriez, R. & Riveline, D. Protrusion fluctuations direct cell motion. *Biophys. J.* **107,** 34–42 (2014).
22. Mahmud, G. *et al.* Directing cell motions on micropatterned ratchets. *Nat. Phys.* **5,** 606–612 (2009).
23. Chang, S. S., Guo, W. H., Kim, Y. & Wang, Y. L. Guidance of cell migration by substrate dimension. *Biophys. J.* **104,** 313–321 (2013).
24. Kushiro, K. & Asthagiri, A. R. Modular design of micropattern geometry achieves combinatorial enhancements in cell motility. *Langmuir* **28,** 4357–4362 (2012).
25. Albert, P. J. & Schwarz, U. S. Dynamics of cell shape and forces on micropatterned substrates predicted by a cellular Potts model. *Biophys. J.* **106,** 2340–2352 (2014).
26. Camley, B. A. & Rappel, W. J. Velocity alignment leads to high persistence in confined cells. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **89,** 062705 (2014).
27. Bi, D., Lopez, J. H., Schwarz, J. M. & Manning, M. L. Energy barriers and cell migration in densely packed tissues. *Soft Matter* **10,** 1885–1890 (2014).
28. Czirók, A., Schlett, K., Madarász, E. & Vicsek, T. Exponential distribution of locomotion activity in cell cultures. *Phys. Rev. Lett.* **81,** 3038–3041 (1998).
29. Takagi, H., Sato, M. J., Yanagida, T. & Ueda, M. Functional analysis of spontaneous cell movement under different physiological conditions. *PLoS One* **3,** e2648 (2008).
30. Li, L., Cox, E. C. & Flyvbjerg, H. 'Dicty dynamics': Dictyostelium motility as persistent random motion. *Phys. Biol.* **8,** 046006 (2011).
31. Pedersen, J. N. *et al.* How to connect time-lapse recorded trajectories of motile microorganisms with dynamical models in continuous time. *Phys. Rev. E* **94,** 062401 (2016).
32. van der Pol, B. On relaxation-oscillations. *Philos. Mag.* **2,** 978–992 (1926).
33. Dieterich, P., Klages, R., Preuss, R. & Schwab, A. Anomalous dynamics of cell migration. *Proc. Natl. Acad. Sci. USA* **105,** 459–63 (2008).
34. Potdar, A. A., Jeon, J., Weaver, A. M., Quaranta, V. & Cummings, P. T. Human mammary epithelial cells exhibit a bimodal correlated random walk pattern. *PLoS One* **5,** e9636 (2010).
35. Mak, M., Reinhart-King, C. A. & Erickson, D. Microfabricated physical spatial gradients for

investigating cell migration and invasion dynamics. *PLoS One* **6,** e20825 (2011).
36. Kraning-Rush, C. M., Carey, S. P., Lampi, M. C. & Reinhart-King, C. A. Microfabricated collagen tracks facilitate single cell metastatic invasion in 3D. *Integr. Biol.* **5,** 606–616 (2013).
37. Berzat, A. & Hall, A. Cellular responses to extracellular guidance cues. *EMBO J.* **29,** 2734–2745 (2010).
38. Nakanishi, J. *et al.* Spatiotemporal control of migration of single cells on a photoactivatable cell microarray. *J. Am. Chem. Soc.* **129,** 6694–6695 (2007).
39. Oswald, L., Grosser, S., Smith, D. M. & Käs, J. A. Jamming transitions in cancer. *J. Phys. D. Appl. Phys.* **50,** 483001 (2017).
40. Bi, D., Lopez, J. H., Schwarz, J. M. & Manning, M. L. A density-independent rigidity transition in biological tissues. *Nat. Phys.* **11,** 1074–1079 (2015).
41. Le Berre, M. *et al.* Geometric friction directs cell migration. *Phys. Rev. Lett.* **111,** 198101 (2013).
42. Neve, R. M. *et al.* A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* **10,** 515–527 (2006).
43. Segerer, F. J., Thüroff, F., Piera Alberola, A., Frey, E. & Rädler, J. O. Emergence and persistence of collective cell migration on small circular micropatterns. *Phys. Rev. Lett.* **114,** 228102 (2015).
44. Camley, B. A., Zhao, Y., Li, B., Levine, H. & Rappel, W. J. Periodic migration in a physical model of cells on micropatterns. *Phys. Rev. Lett.* **111,** 158102 (2013).
45. Segerer, F. J. *et al.* Versatile method to generate multiple types of micropatterns. *Biointerphases* **11,** 011005 (2016).
46. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9,** 671–675 (2012).
47. Siegert, S., Friedrich, R. & Peinke, J. Analysis of data sets of stochastic systems. *Phys. Lett. A* **243,** 275–280 (1998).
48. Ragwitz, M. & Kantz, H. Indispensable Finite Time Corrections for Fokker-Planck Equations from Time Series Data. *Phys. Rev. Lett.* **87,** 254501 (2001).
49. Stephens, G. J., Johnson-Kerner, B., Bialek, W. & Ryu, W. S. Dimensionality and Dynamics in the Behavior of C. elegans. *PLoS Comput Biol* **4,** (2008).

Supplementary Information for

Stochastic Nonlinear Dynamics of Confined Cell Migration in Two-State Systems

David B. Brückner[1]*, Alexandra Fink[2]*, Christoph Schreiber[2], Peter J. F. Röttgermann[2], Joachim O. Rädler[2#], Chase P. Broedersz[1#]

[1] *Arnold Sommerfeld Center for Theoretical Physics and Center for NanoScience, Ludwig- Maximilians-Universität, München*

[2] *Faculty of Physics and Center for NanoScience, Ludwig-Maximilians-Universität, München*

* authors contributed equally

[#] correspondence to: C. P. Broedersz (c.broedersz@lmu.de) or J. O. Rädler (raedler@lmu.de)

Contents

0. Movie descriptions

Supplementary Movies 1-3
Single MDA-MB-231 cells transitioning repeatedly between the square adhesion sites of the two-state micropattern. Transitions are usually preceded by the formation of a protrusion along the bridge. The cell nucleus is fluorescently labelled to allow automated tracking of cell positions. The bridge lengths shown in movies 1-3 are L = 16, 35, 56 µm, respectively. Scale bar: 25 µm.

Supplementary Movie 4
Exemplary field of view of MDA-MB-231 cells migrating on two-state micropatterns of the same bridge length (L = 35 µm). All cells perform transitions between the square adhesion sites. Not all micropatterns are occupied, which is due to the low cell seeding density used to ensure single-cell occupancy. Cell nuclei are labelled for semi-automated detection of cell positions. Scale bar: 25 µm.

Supplementary Movies 5&6
Single MDA-MB-231 cells transfected with LifeAct-GFP to visualize actin on two-state micropatterns of bridge length L = 35 µm. The outline of the underlying micropattern is drawn as a reference up to scale. Actin hotspots are visible at the tip of the transition-mediating lamellipodium, as well as during the dynamic exploration of the square adhesion sites. Actin fibres reorganise dynamically.

Supplementary Movie 7
Single MDA-MB-231 cell on a stripe micropattern without constriction of total length 103 µm. The cell moves back and forth, repolarising upon contact with the pattern's borders. When the cell is positioned in the middle of the pattern, quick changes in the direction of lamellipodia formation can be seen. The cell nucleus is fluorescently labelled to allow automated tracking of cell positions.

Supplementary Movie 8
Sparsely seeded MDA-MB-231 cells freely migrating on a homogeneous fibronectin-coated 2D surface. Cells move randomly on the surface. Cell nuclei are fluorescently labelled for automated cell tracking. Scale bar: 100 µm.

Supplementary Movie 9
Single MCF10A cell transitioning repeatedly between the square adhesion sites of the two-state micropattern. Transitions are usually preceded by the formation of a protrusion along

the bridge. Several times, protrusions along the bridge are formed which do not lead to a transition. The cell nucleus is fluorescently labelled to allow automated tracking of cell positions. Bridge length L = 35 µm.

Supplementary Movies 10-13

Single cells of various cell lines (MDA-MB-436, MDCK, HuH7, A549) transitioning between the square adhesion sites of the two-state micropattern. The cell nucleus is fluorescently labelled to allow automated tracking of cell positions. Bridge length L = 35 µm.

1. Further experimental details

1.1. Photolithography

Silicon wafers are pre-treated with a solution of hydrofluoric acid (TECNIC). Then the wafer is spin-coated with photoresist AZ40XT (AZ Electronic Materials) and subsequently soft-baked.

The desired geometries are patterned into the photoresist by laser direct imaging (Protolaser, LPKF). After a post-exposure bake, the wafer is developed with AZ 726 MIF (AZ Electronic Materials). In the last step, the wafer is silanized with (Trichloro(1H,1H,2H,2H-perfluoro-octyl)silane (Sigma-Aldrich).

1.2. Stamp preparation

Stamps are created by mixing polydimethylsiloxane (PDMS) monomer and cross-linker (DC 184 elastomer kit, Dow Corning) in a 10:1 ratio. The polymer is then poured onto the silicon wafer bearing the desired geometries. After degassing the PDMS, it is cured overnight at 50 °C.

1.3. Micropattern design

All two-state micropatterns are designed to have equal square dimensions ((36.7 ± 0.6)^2 µm^2) and for the bridge to have the same width ((6.9 ± 0.6) µm). The measured bridge lengths L as well as the number of cell trajectories for each L are shown in table S1. The micropattern without constriction is designed to have a similar width and total length as the two-state pattern with L = 35 µm, and thus has dimensions ((103.4 ± 0.3) µm) x ((34.8 ± 0.2) µm).

Cell line	*L* / µm	Cells	Transitions	Time-points
MDA-MB-231	6.4 ± 0.3	98	1422	20955
	9.2 ± 0.3	101	1399	21044
	15.7 ± 0.3	169	1917	38538
	23.7 ± 0.4	216	2060	45639
	35.3 ± 0.5	149	1293	35103
	46.2 ± 0.4	127	947	26805
	56.0 ± 0.3	74	508	18190
	no constriction	212	-	42723
	2D	728	-	71341
MCF10A	35.3 ± 0.5	219	1125	46744

Table S1 | Bridge lengths *L* of the micropatterns and amount of statistics collected: number of cell trajectories, number of transitions for two-state systems, and total number of time-points recorded.

The quoted errors in *L* correspond to deviations in the dimensions of final protein patterns that are due to the intrinsic variance of the manual stamping process and the measurement uncertainty associated with the limited resolution of the brightfield images. Throughout the manuscript, we refer to the rounded values of the bridge lengths for simplicity.

1.4. Cell exclusion criteria

We track the positions of a large number of cells to determine the transition statistics of cells migrating on micropatterns. To limit the effects of ambiguous or abnormal migration behaviour we apply the following inclusion criteria in our analysis of migration in two-state micropatterns:

1. Only a single cell occupies the micropattern. Trajectories are cut when the cell rounds up for division.
2. The cell and its protrusions are entirely confined within the borders of the micropattern.
3. The cell shows no abnormalities such as multiple nuclei or the occurrence of cell death or detachment from the substrate at any time during the whole experiment.
4. Transition statistics are only included after the first and until the last observed bridge transition.
5. In the vast majority of cases the cell performs complete transitions. A complete transition requires that no parts of the cell adhere to the previous adhesion site once the nucleus has entered the new adhesion site.

Criteria 1-3 are basic conditions for single cell experiments on micropatterns, and only these three criteria are applied to cells migrating on the patterns without constriction. Criterion 4

ensures that start- and end-of-measurement artefacts in determining dwell times are avoided. Criterion 5 is specific to our two-state system and its importance may depend on the cell type. We therefore investigate the impact of condition 5 on a sample data set. We find that the experimental trend in dwell times is unaffected by condition 5 to within error (Fig. S1). Furthermore, the experiment-to-model agreement appears unaffected by this condition. We therefore conclude that applying condition 5 does not change our conclusions.

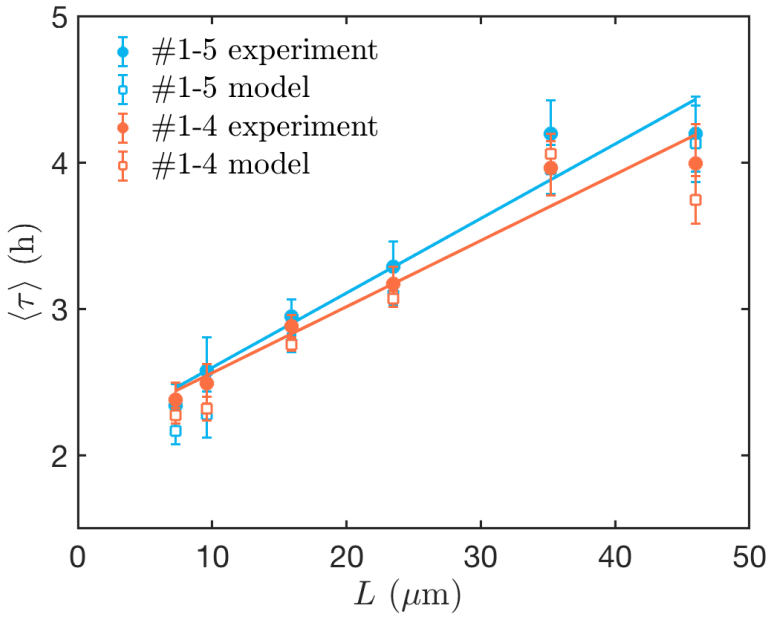


Figure S1 | Average dwell time as a function of bridge length (MDA-MB-231). Solid blue points correspond to trajectories that fulfil criteria #1-5 while solid orange points correspond to exclusion based on criteria #1-4. The corresponding model predictions are shown by empty square symbols in the respective colours. Error bars correspond to bootstrap errors.

1.5. Cell transfections

For life cell imaging of actin, approximately 10,000 MDA-MB-231 cells are seeded in patterned µ-dishes and left to adhere overnight. As a cell culturing medium, we use MEM including Glutamax (Gibco) supplemented with 10% FCS. 500ng LiveAct-GFP mRNA (in-house prepared) is resuspended in OptiMEM (Gibco) to a final volume of 150µl. This solution is then added to a mix of 1.25 µl Lipofectamine 2000 (Invitrogen) and 123.75 µl OptiMEM, and left to incubate for 20 minutes at room temperature. Subsequently, cells are rinsed once with PBS and the transfection mix is added and left on the cells for at least 5h, before being replaced by L-15 medium. Cells are imaged every 10 minutes on the Nikon Ti Eclipse microscope using a 60x oil-immersion objective.

1.6. Cell culture of additional cell lines

MDA-MB-436 human breast carcinoma cells (ATCC), MDCK-II dog kidney cells (ATTC), HuH7 human liver carcinoma cells (I.A.Z. Munich) and A549 human lung carcinoma cells (ATCC) are cultured at 37°C in an atmosphere with 5% CO_2. For MDA-MB-436 cells, DMEM-F12 medium including Glutamax (Gibco) supplemented with 10% FBS (Gibco) is used. MDCK cells, as well as A549 cells, are cultured in MEM medium including Glutamax (Gibco) and supplemented with 10% FBS. HuH7 cells are cultured in RPMI 1640 medium containing Glutamax (Gibco) supplemented with 10% FBS, 5mM HEPES (Gibco) and 1mM Sodiumpyrovate (Gibco).

1.7. Cell migration on homogeneous 2D substrates

For a characterisation of the free migration behaviour of MDA-MB-231 cells (Section S4), we performed experiments on uniform 2D surfaces that are prepared similarly to the micropatterned surfaces for comparability. First, a µ-dish (ibidi GmbH) is incubated with a 50 µg ml^{-1} human fibronectin (YO Proteins) solution for 50 min and subsequently rinsed with PBS. Approximately 1000-5000 MDA-MB-231 cells are seeded and left to adhere for 4h. For time-lapse measurements, the medium is exchanged to L-15 medium with L-glutamine (Gibco, supplemented with 10% FCS) containing 25 nM Hoechst 33342 (Invitrogen). Phase contrast and fluorescence images are acquired every 10 min for up to 48 h. To maximise the field of view and thereby allow cell tracking over larger distances, 4 images are acquired with 5% overlap are subsequently combined into a single image. All further analysis is performed in MATLAB (Mathworks). First, a blurred image is subtracted from the fluorescent images of the nuclei to correct for uneven illumination. Next, in analogy to data analysis on micropatterns, a band pass filter is applied and the resulting images are thresholded to achieve an adequate binary version of the fluorescence images. This enables an automated detection of cell nuclei and the localisation of the centres of the cell nuclei. Cells are tracked using an adaption of the IDL particle tracking software[1].

To minimise the influence of cell-cell interactions, only time points of each track with a minimum distance of 132 µm between neighbouring nuclei are considered. In accordance with previous work[2], we exclude obviously non-moving cells, defined as cells whose root mean square displacement never exceeds 25 µm. Furthermore, cells are checked manually to comply with cell exclusion criterion 3 (Section S1.4). To ensure comparability between experiments of freely migrating cells and confined cells, we only analyse trajectories with a minimum duration of 500 min. We analysed a total of 728 cell trajectories.

2. Analysis of the hopping process

2.1. Actin staining

To gain additional insight into the dynamics of the cytoskeletal activity and the morphological changes during the cellular 'hopping' process, we performed experiments with Lifeact-GFP transfected cells (see Section S1.5), which are included in Supplementary Movies S5&6 and Fig. S2. In addition to the actin hotspots at the leading cell edges, we sometimes observe the extension of filopodia at the cell periphery. Furthermore, retraction fibres appear to tail the cell body during transitions.

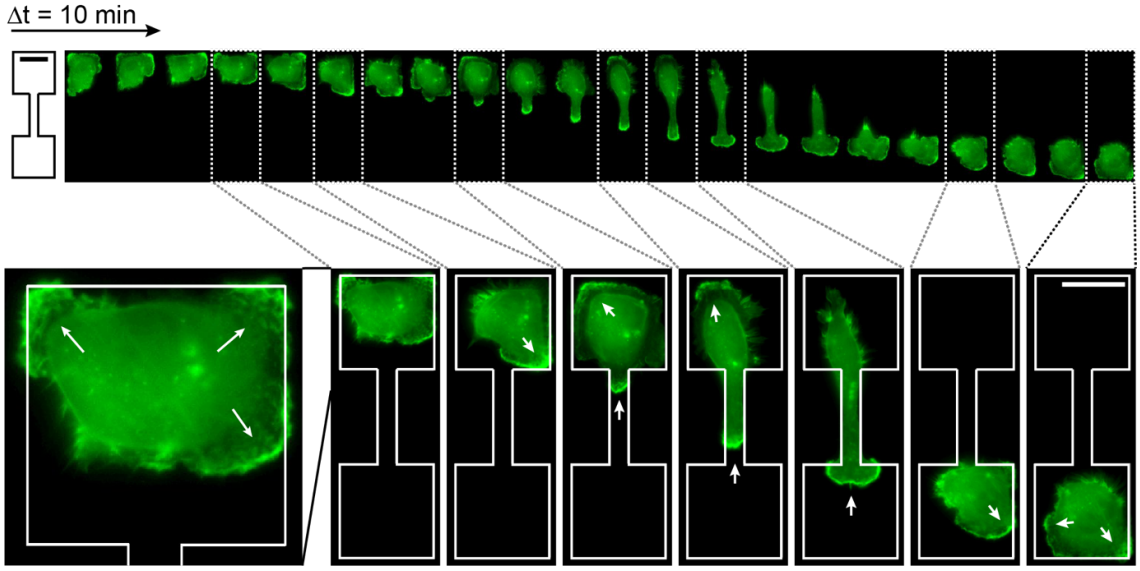


Figure S2 | Time-Series of an MDA-MB-231 cell on a two-state micropattern with LifeAct stain (*L* = 35 µm). Actin hotspots are visible at the lamellipodial tip, as well as at the leading edge of the cell during dynamic exploration of the square adhesion site, marked with white arrows. Corresponds to Supplementary Movie S5. Scale bars: 25 µm.

2.2. Generality of the hopping process for different cell lines

To investigate the generality of the hopping process that we observe for cancerous (MDA-MB-231) and non-cancerous (MCF10A) human breast cells, we also tested several other cell lines. Specifically, we tested MDA-MB-436 human breast cancer cells, Madin-Derby Canine Kidney cells (MDCK), human liver carcinoma cells (HuH7), and human lung carcinoma cells (A549) on two-state micropatterns with a bridge length L = 35 µm. In all cases, we found that the cells perform stochastic transitions between the two sites (Fig. S3), indicating that the hopping behaviour is a general migration pattern in motile cells. However, different cell

lines do not exhibit the same morphology during the transition phase (Fig. S3). Some cell types are less motile and thus limit the statistics generated per experiment. Not all cell lines have the same size, indicating that pattern size may need to be adjusted for optimal confinement. However, while some cell line-specific optimisation needs to be performed, the assay has a broad general applicability. The experimental details for this part are summarised in section 1.7.

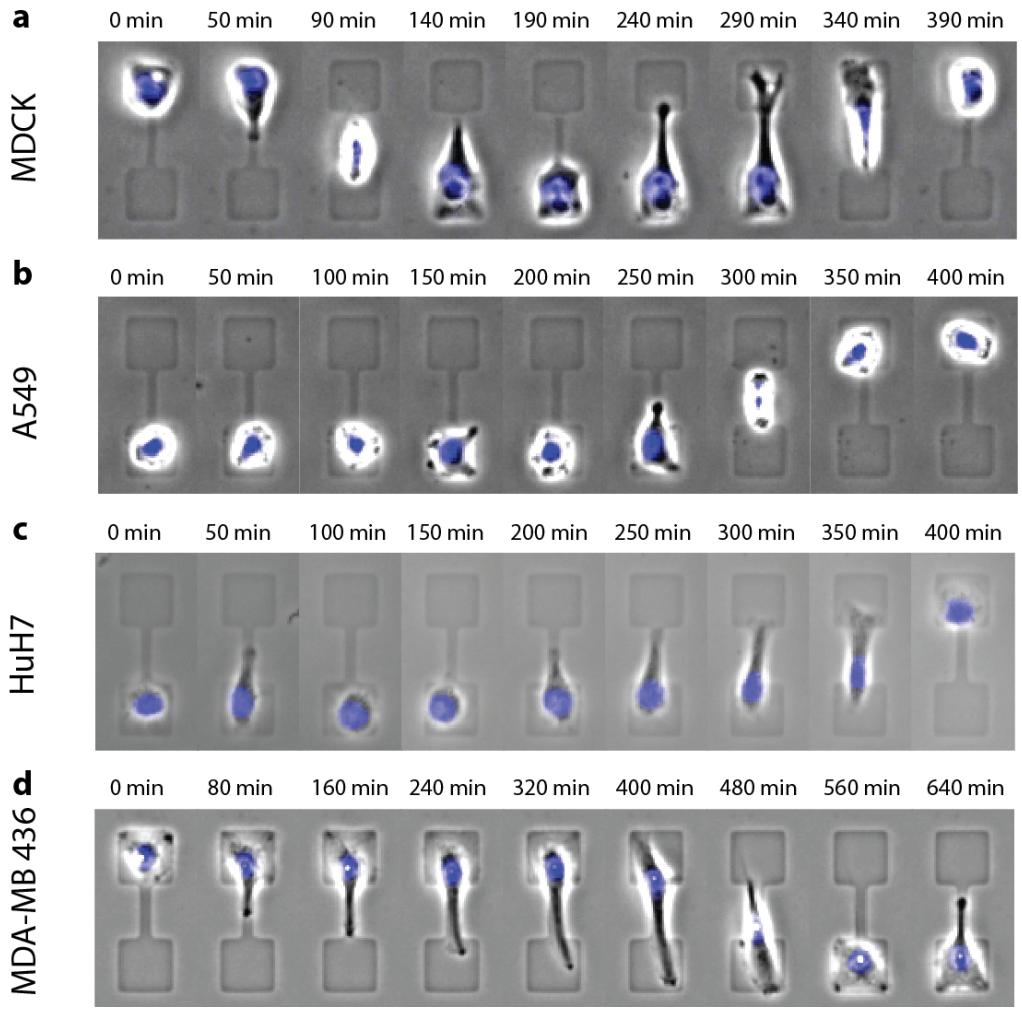


Figure S3 | Time-Series of different cell lines migrating on two-state micropatterns (*L* = 35 µm). Time series of phase contrast images and labelled nuclei (blue). **a**, Madin-Derby Canine Cells (MDCK). **b**, Human lung carcinoma cells (A549). **c**, Human liver carcinoma cells (HuH7). **d**, Human breast cancer cells (MDA-MB-436).

2.3. Calculation of the dwell times

To calculate the dwell times, the trajectories are binarised into two states, left and right of the centre of the connecting bridge. The centre of the bridge is determined by manually identifying the left border of each individual pattern, and then adding the sum of the mean left adhesion site edge length and half of the mean bridge length for each experiment. We subsequently calculate the time spent in either state between the transitions, yielding the dwell time τ.

Given the particular geometry of our system, there are two obvious choices of how to define the states 'left' and 'right': (i) treating the centre of the connecting bridge as the boundary or (ii) excluding any time spent on the bridge and simply measuring the time spent on the square adhesion sites. We find that both definitions yield a linear trend in average dwell time (Fig. S4). The slope of this linear dependence is smaller when using island boundaries, since this choice excludes the time it takes for the cells to traverse a longer bridge. We adopt the convention of the centre boundaries throughout, noting that this choice does not qualitatively affect any of our conclusions.

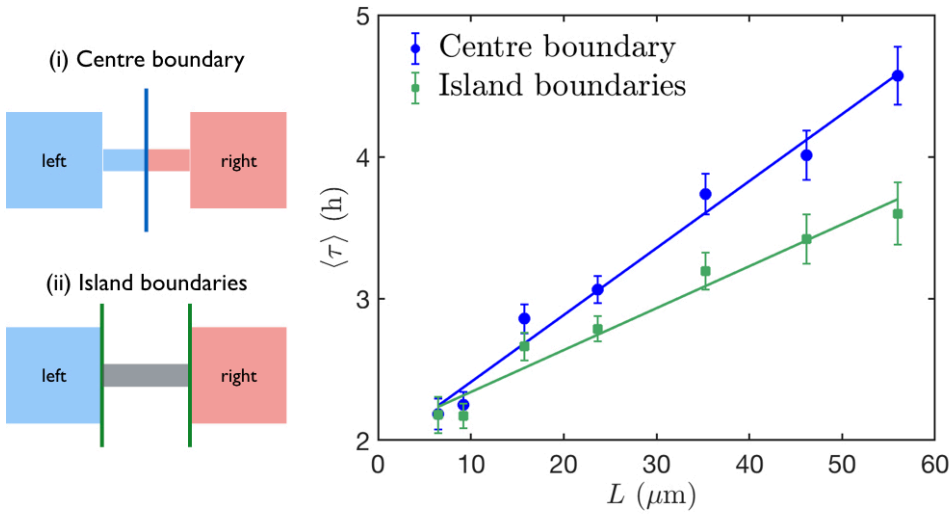


Figure S4 | Alternative definitions of the dwell time. Average dwell time of MDA-MB-231 cells as a function of bridge length, for two choices of defining the dwell time, sketched on the left. Error bars correspond to bootstrap errors.

2.4. Error analysis

To calculate the errors in the dwell times, we employ a bootstrapping procedure. The recorded cell trajectories are stochastic and contain long-time correlations. Subsequent dwell times may therefore also be correlated. A simple error calculation by determining the standard error of the dwell times will thus underestimate the error, since this procedure assumes that the measurements are independent. Therefore, we use a bootstrapping

procedure which estimates the errors including the correlations. An alternative related method for estimating averages of correlated data is the "blocking" method[3].

In general, the bootstrapping procedure to measure the error in the mean $\langle X \rangle$ of a dataset works as follows[4]: For a dataset $D = \{X_1, \ldots, X_N\}$ with N entries, a large number of realisations of the dataset is generated by randomly sampling its entries, *with replacement*. A given realisation is thus produced by picking N randomly selected entries from the data set, and a given entry may be picked multiple times. This procedure is repeated many times, producing many realisations. For each realisation, the mean $\langle X \rangle_{\text{realisation}}$ is calculated. The estimated error in the mean is then given by the standard deviation of the means of the realisations.

To take correlations into account, the same procedure is carried out for small groups of subsequent entries (windows) from the dataset. The window size w, i.e. the number of subsequent entries per group, must be much smaller than the total number of entries, $N \gg w$. We can then plot the standard deviation as a function of window size. The standard deviation will saturate as a function of w, and approximately reach a constant value $\sigma_{\max}$ when w exceeds the scale over which subsequent entries are correlated. This value $\sigma_{\max}$ is our final estimate of the error in the mean.

This procedure is illustrated in Fig. S5, using the example of our array of dwell times. For a particular bridge length, we generate the list $D = \{\tau_1, \ldots, \tau_N\}$ containing N dwell times. Given a window size w, we then pick groups of w subsequent dwell times until we have picked N dwell times, giving one realisation. For each window size, we generate 10^5 realisations. For each realisation i, we thus obtain an average dwell time $\langle \tau \rangle_i$.

As expected, the distribution of realisation averages $\langle \tau \rangle_i$ exhibits a Gaussian shape (Fig. S5a). As we change the window sizes, the mean of each distribution $\langle \langle \tau \rangle_i \rangle$ should remain constant at the experimentally observed average dwell time $\langle \tau \rangle_{\text{expt}}$, while the variance increases. The double bracket notation corresponds to taking the average of averages, as the distribution mean is given by the average of the realisation averages. Plotting the distribution mean and variance against window size, we indeed find very good agreement of the distribution averages $\langle \langle \tau \rangle_i \rangle$ with $\langle \tau \rangle_{\text{expt}}$ (Fig. S5b), as well as the expected saturating curve for the variance (Fig. S5c). As we monitor on the order of 10^3 transitions for each bridge length, we limit the window size to a maximum value of 60 in order to fulfil the criterion $N \gg w$. We then take the maximum value of the distribution variance as our final bootstrap error.

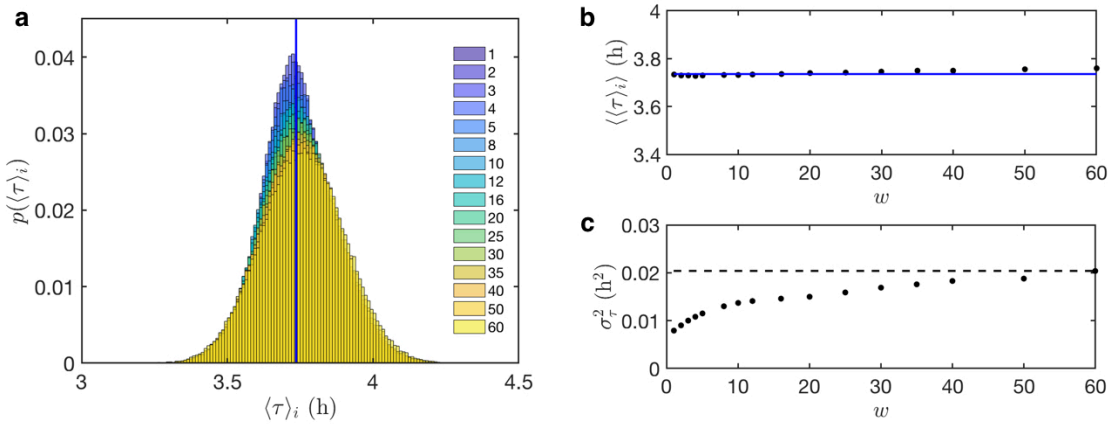


Figure S5 | Distributions of realisation averages $\langle\tau\rangle_i$ as a function of window size (*L* = 35 µm, MDA-MB-231). **a**, Probability distributions of the realisation averages $\langle\tau\rangle_i$ for different window sizes (indicated by colours as shown in the figure legend). **b**, Distribution averages $\langle\langle\tau\rangle_i\rangle$. In **a** and **b**, the average dwell time found experimentally, $\langle\tau\rangle_{\mathrm{expt}}$, is indicated by a blue line. **c**, Distribution variances σ_τ^2 as a function of window size. The dashed line indicates the maximum variance $\sigma_{\max}^2$that gives the estimate taken for the final error.

2.5. Stochasticity analysis

To investigate whether the width of the dwell time distribution $p(\tau)$ (Fig. 1d) is set by the cell-to-cell variability or the single cell stochasticity, we quantify the single-cell and the population variability.

For each bridge length, we obtain a set of dwell times $\{\tau\}$ which consists of smaller sets of dwell times $\{\tau\}_j$ generated by each cell j:

$$\{\tau\} = \left\{\{\tau\}_1, \ldots, \{\tau\}_j, \ldots, \{\tau\}_N\right\} \tag{S1}$$

We can therefore quantify the average single cell variance

$$\langle\sigma_{sc}^2\rangle = \langle Var\left(\{\tau\}_j\right)\rangle \tag{S2}$$

where the average is taken over different cells. This is a measure of the average variation of the hopping behaviour within the lifetime of single cells. To quantify the variation of dwell times across different cells, we can measure the cell-to-cell variance

$$\sigma_{cc}^2 = Var\left(\langle\{\tau\}_j\rangle\right) \tag{S3}$$

where the average is taken over all transitions of each cell j. As a reference, we also calculate the overall 'pooled' variance of all dwell times $\{\tau\}$ of all transitions of all cells:

$$\sigma_{pool}^2 = Var(\{\tau\}) \tag{S4}$$

If there was no stochasticity at the single-cell level, but each cell had a different typical dwell time, we would expect that

$$\sigma_{sc}^2 = 0, \qquad \sigma_{cc}^2 \approx \sigma_{pool}^2 \tag{S5}$$

Note that if all trajectories had the same length, then σ_{cc}^2 would exactly equal σ_{pool}^2. As different cells have different lifetimes, this is not the case here and we therefore only expect the two values to have a similar order of magnitude.

However, if each single cell performs stochastic transitions that give rise to the overall spread in dwell times, then we expect

$$\sigma_{sc}^2 \approx \sigma_{cc}^2 \approx \sigma_{pool}^2 \tag{S6}$$

Indeed, here we observe that all three variances have similar order of magnitudes, with a single-cell variance that exceeds both cell-to-cell and pooled variances on most of the tested bridge lengths (Fig. S6). We therefore conclude that there is an underlying stochasticity that is also relevant at the single-cell level, and not just at the population level.

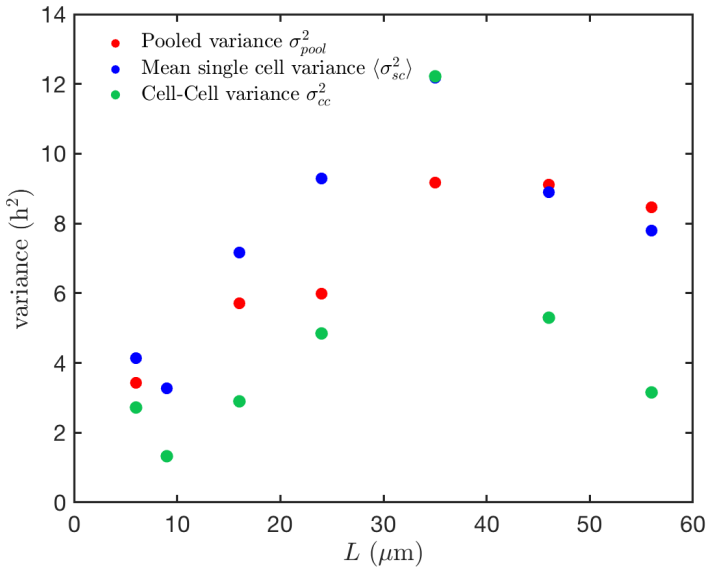


Figure S6 | Single-cell and cell-to-cell variability. Comparison of the dwell time variances at the single-cell level (blue), the cell-to-cell level (green) and the population level (red) for the different bridge lengths (MDA-MB-231).

2.6. Survival probability functions

Combining the dwell times from many cells, we obtain the probability distribution $p(\tau)$. The survival probability distribution is then calculated using

$$S(t) = 1 - \int_0^t p(\tau)\mathrm{d}\tau \qquad (S7)$$

Plotting the survival probability function in log-linear and log-log axes, we find that it follows neither a simple exponential decay nor a power law (Fig. S7).

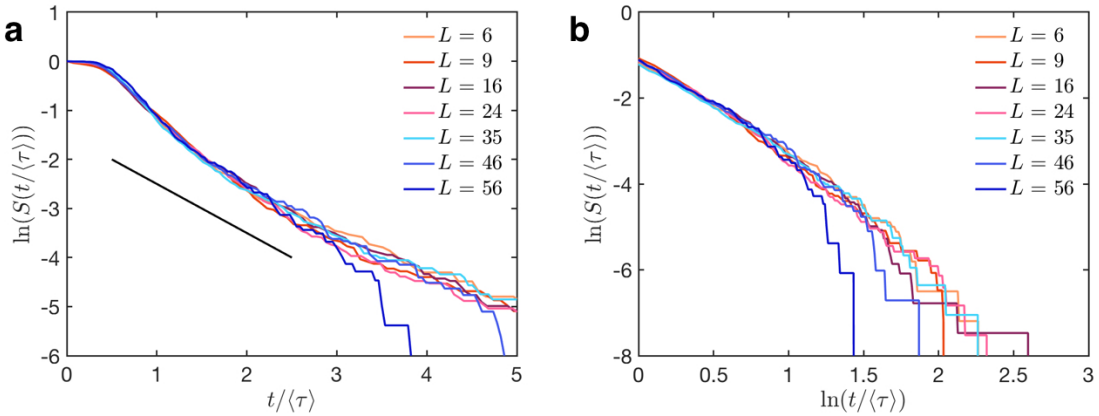


Figure S7 | Survival probability distributions normalised by average dwell time. a, Plotted in log-linear axes. As a guide to the eye, a black line with slope -1, corresponding to a single exponential, $S(t) = e^{-t/\langle\tau\rangle}$, is shown. **b**, Plotted in log-log axes.

The survival probability function $S(t)$ allows an insightful approximate interpretation of the cellular transition behaviour. For example, the initial plateau corresponds to a time-scale on which there are no transitions. An alternative way of analysing the hopping process is a two-step process of cell spreading and protrusion formation. After crossing the bridge, the cell spreads on the adhesion site and subsequently repolarises. During this time, the cell cannot perform another transition. The second decay time-scale is then likely related to the dynamics of the attempt and transition process. We therefore compare the experimental survival probability function to two different plausible types of processes governed by two timescales:

i. Plateau-and-decay Model: In this scenario, the spreading process occurs right after a transition, when the cell conforms to the geometry of the adhesive island and repolarises. During this process, which has a typical time-scale t_0, the cell cannot perform another transition, implying that the survival probability function should exhibit a flat plateau for

$t < t_0$. If the subsequent invasion is also governed by a single timescale, this implies a time-invariant probability of transitioning, resulting in an exponential decay of $S(t)$ for $t \geq t_0$:

$$S(t) = \begin{cases} 1 & t < t_0 \\ \exp(-k[t - t_0]) & t \geq t_0 \end{cases} \quad (S8)$$

By fitting this function to our experimental data, with k and t_0 as free parameters, we find that this functional form provides a reasonable approximation to the data (Fig. S8a). However, this approximation fails in the tail of the distribution.

ii. Multiple-timescale Poisson process: Cell spreading might also give rise to a probabilistic process. This would imply that the transition behaviour is governed by a two-step Poisson process, with the following survival probability function:

$$S(t) = \frac{k_2}{k_1 - k_2} e^{-k_1 t} + \frac{k_1}{k_2 - k_1} e^{-k_2 t} \quad (S9)$$

where $k_1, k_2 > 0$. However, both single- and two-step Poisson processes do not provide good fits (Fig. S8b). Certainly, including more time-scales will eventually result in a better fit, but may not allow a simple interpretation of these timescales.

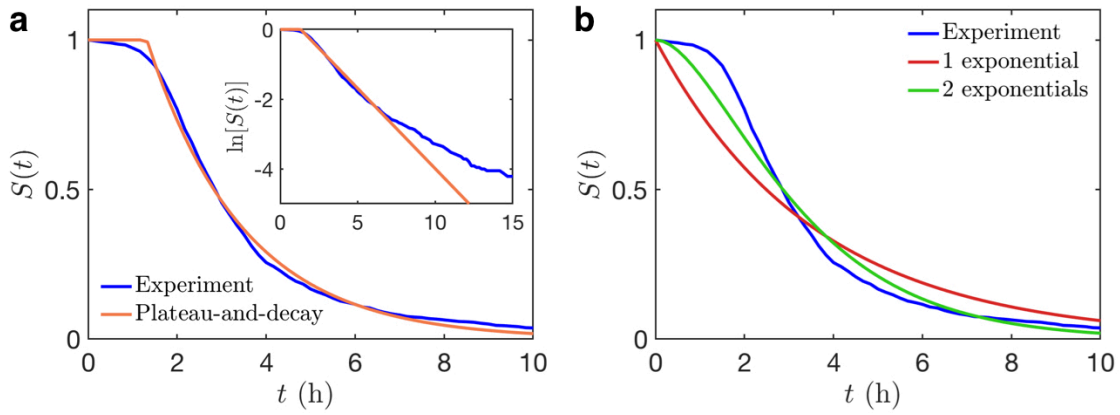


Figure S8 | Fits to the survival probability function (*L* = 35 µm, MDA-MB-231). **a**, Experimental survival probability function with the fitted plateau-and-decay model expression given by equation (S8). Inset: Corresponding log-linear plot. **b**, Single and double exponential decay functions fitted to the experimental survival probability function (*L* = 35 µm).

2.7. Importance of the bridge length

For all bridge lengths that we tested (6 - 56 µm), we found similar qualitative deterministic and stochastic contributions to the dynamics (Supplementary Figs. S28 and 29). Here, the

lamellipodium of the cell is able to reach the adhesive island the cell is migrating towards, while the nucleus still resides on the other island. For larger *L*, this may no longer be possible, implying that if the cell is still able to migrate across the thin constriction, it will do so by employing a different migration mechanism. Indeed, previous studies have shown that even on micropatterned stripes as thin as 1 µm cells are still able to migrate[5], implying that even on very long bridges, cells can in principle perform transitions. However, we expect these transitions to follow different dynamics on these length-scales.

3. Inference and analysis of the stochastic model

3.1. Inference procedure

The position of the cell nucleus is recorded with a time interval Δt = 10 min in the experiment, and velocity and acceleration are directly calculated as numerical derivatives:

$$v(t) = \frac{x(t) - x(t-\Delta t)}{\Delta t} \tag{S10}$$

$$a(t) = \frac{v(t+\Delta t) - v(t)}{\Delta t} = \frac{x(t+\Delta t) - 2x(t) + x(t-\Delta t)}{(\Delta t)^2} \tag{S11}$$

Note that we employ the symmetric numerical derivative, where $a(t)$ is a linear combination of $x(t+\Delta t)$, $x(t)$, and $x(t-\Delta t)$. Our proposed model has the form

$$\frac{dv}{dt} = F(x,v) + \sigma(x,v)\eta(t) \tag{S12}$$

where $\eta(t)$ represents a Gaussian white noise with zero mean and correlation $\langle \eta(t)\eta(t')\rangle = \delta(t-t')$. Interpreting the stochastic differential equation in the Itô-sense, a discrete Euler integration at time step δt yields

$$\begin{aligned} v(t+\delta t) - v(t) &= F\big(x(t),v(t)\big)\delta t + \sigma\big(x(t),v(t)\big)\int_t^{t+\delta t} \eta(s)\mathrm{d}s \\ &= F\big(x(t),v(t)\big)\delta t + \sigma\big(x(t),v(t)\big)\delta W(t) \end{aligned} \tag{S13}$$

where

$$\delta W(t) = \int_t^{t+\delta t} \eta(s)\mathrm{d}s \tag{S14}$$

The separation of the noise magnitude from the integral in the first equality of equation (S13) is valid in the Itô-picture. Conditional averaging then gives the deterministic and stochastic terms in the Itô-picture[6–8], which coincide with the drift and diffusion terms of the corresponding Fokker-Planck equation for the system:

$$F(x,v) = \left\langle \frac{v(t+\delta t) - v(t)}{\delta t} \,\middle|\, x = x(t), v = v(t) \right\rangle \tag{S15}$$

and

$$\sigma^2(x,v) = \delta t \left\langle \left[\frac{v(t+\delta t) - v(t)}{\delta t} - F(x,v) \right]^2 \,\middle|\, x = x(t), v = v(t) \right\rangle \tag{S16}$$

and

$$\delta W(t) = \frac{\delta t}{\sigma(x,v)} \left\langle \left[\frac{v(t+\delta t) - v(t)}{\delta t} - F(x,v) \right] \middle| x = x(t), v = v(t) \right\rangle \qquad (S17)$$

The conditional averaging requires a discrete binning of the phase space. Here, we use a uniform binning of 30x30 bins throughout. The binning structure is shown explicitly in Supplementary Figs. S28 and S29. None of our conclusions depend sensitively on the number of bins used. Furthermore, we set $\delta t = \Delta t$ throughout. From the trajectories of the inferred noise $\Delta W(t) = \delta W(t)$ with $\delta t = \Delta t$, we can directly calculate the noise correlation function. As a self-consistency check, we verify that this correlation function decays within a time scale less than Δt. This is indeed what we find for this system (See Fig. S9). The small negative correlation that we detect for time-separations equal to the sampling interval ($t = \theta$) can be accounted for by localisation errors in the positions (see Section 3.7 and ref. 9).

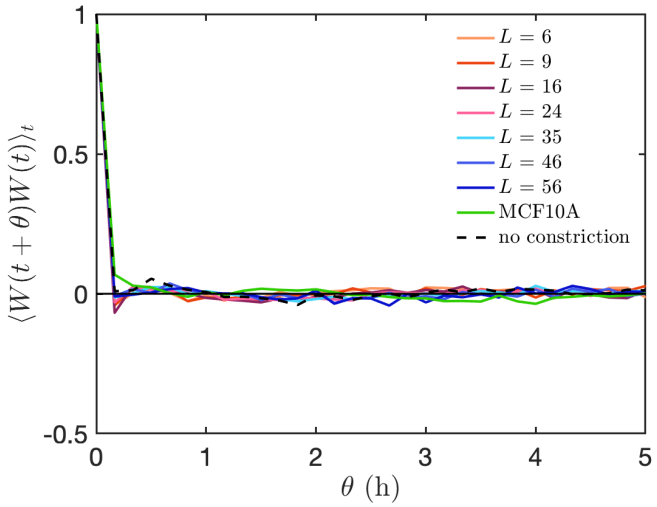


Figure S9 | Correlation functions of the inferred noise. See also Fig. S31.

Throughout this work, we interpret the stochastic differential equation (1) in the Itô-picture. This has several advantages. Firstly, to integrate the dynamics in a step-wise manner (equation S19), it is convenient to employ the Itô formulation, as the noise in the step-wise simulation can only depend on the current, known state of the system.

Secondly, in the Itô-picture, we can recover the deterministic dynamics of the system by simply setting the noise to zero:

$$\frac{d\langle v \rangle}{dt} = \langle F(x,v) \rangle + \langle \sigma(x,v)\eta(t) \rangle = F(x,v) \qquad (S18)$$

where $\langle \dots \rangle$ corresponds to an ensemble average. If we used the alternative Stratonovich interpretation, then $\langle \sigma(x,v)\eta(t) \rangle \neq 0$ and we would have to deduct an additional noise-induced drift[10]. The consistent use of the Itô-interpretation of equation (1) avoids these issues.

3.2. Integration of the stochastic model

The inferred model is subsequently integrated numerically with a time step Δt equal to the experimental sampling interval, using a standard Euler scheme:

$$\begin{aligned} x_i &= x_{i-1} + v_i \Delta t \\ v_{i+1} &= v_i + F(x_i, v_i)\Delta t + \sigma(x_i, v_i) r_i \sqrt{\Delta t} \end{aligned} \qquad (S19)$$

where r_i is a random number drawn from a Gaussian distribution with mean 0 and variance 1. To determine the values $F(x_i, v_i)$ and $\sigma(x_i, v_i)$, we bin the simulated phase space using the same grid as in the inference scheme and take the inferred raw values for F and σ in each bin. If a particle leaves the phase space sampled in the experiment, the simulation terminates.

For the comparison between experiment and model, we employ the hold-out method. The terms $F(x,v)$ and $\sigma(x,v)$ are inferred by conditional averaging of 50% of the cell trajectories recorded in experiment, forming the training data set. The cells included in this set are randomly sampled from the entire data set. The model prediction is then compared to the remaining 50% of the experimental results that were excluded from the model inference (validation data set). This comparison gives similar results if the training and validation sets are interchanged.

3.3. Integration of the deterministic dynamics

The deterministic dynamics is more sensitive to the binning of phase space than the stochastic model. Since the binned function $F(x,v)$ is not differentiable, trajectories can intersect in phase space. To avoid these artefacts, we use a cubic interpolation of $F(x,v)$ for the integration of the purely deterministic dynamics. However, the same deterministic dynamics still emerges robustly if a non-interpolated deterministic term is used (Fig. S10). For the integration of the deterministic dynamics, a simulation time interval of 0.01 h is used.

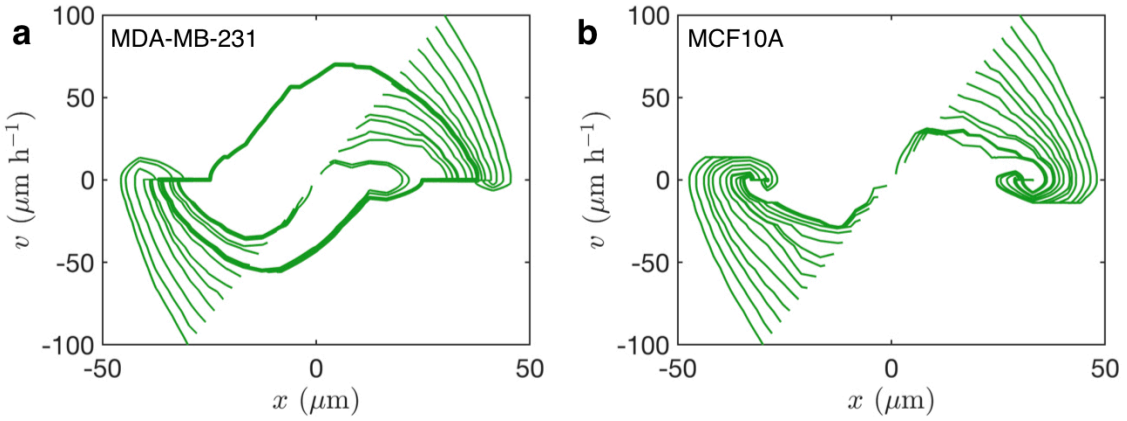


Figure S10 | Deterministic dynamics in coarse-grained phase space (*L* = 35 µm). Trajectories in xv-space for a number of different initial conditions, integrated using the raw binned values of $F(x, v)$. **a**, MDA-MB-231. **b**, MCF10A.

3.4. Symmetrisation procedure for the detection of the nullclines

The nullcline $\dot{v} = 0$ is given by the set of points where $F(x, v) = 0$. Since only a discrete sampling of F is accessible, we find the intersections of $F(x, v)$ with the zero-plane by finding all points of $F(x, v)$ that satisfy the condition $-b < F(x, v) < b$. Here, we set $b = 4$ µm/h^2.

This procedure is rather sensitive to the measurement noise in $F(x, v)$. To improve the statistics, we therefore exploit the inversion symmetry of the system. By construction, our micropatterns are symmetric under the operation $x \to -x$. This means that our model must have the following properties:

$$\begin{aligned} F(x, v) &= -F(-x, -v) \\ \sigma(x, v) &= \sigma(-x, -v) \end{aligned} \qquad (S20)$$

To test whether this symmetry is indeed present in our data, we first verify that both the dynamical properties, i.e. the dwell times on the right and left sides, and the static probability distribution are symmetric (Fig. S11a, b). Next, to verify the inversion symmetry of the deterministic and stochastic terms (equation (S20)), we define the relative symmetry deviations in each case:

$$\delta_F(x, v) = \frac{F(-x, -v) + F(x, v)}{F(x, v)} \qquad (S21)$$

$$\delta_\sigma(x, v) = \frac{\sigma(-x, -v) - \sigma(x, v)}{\sigma(x, v)} \qquad (S22)$$

These fractional deviations are shown as a function of x and v in Fig. S11c, d. We find that there are no systematic deviations from the postulated symmetries. The only significant deviations are in the deterministic term close to the nullclines, which may be due to random measurement noise which is larger in this region. We thus use this symmetry by inferring only the top half of phase space ($v > 0$) using both the left-right ($v > 0$) and the right-left ($v < 0$) transitions. The bottom half is then given by a point reflection of the top half around the origin. We found similar results for the other bridge lengths as well as with MCF10A cells.

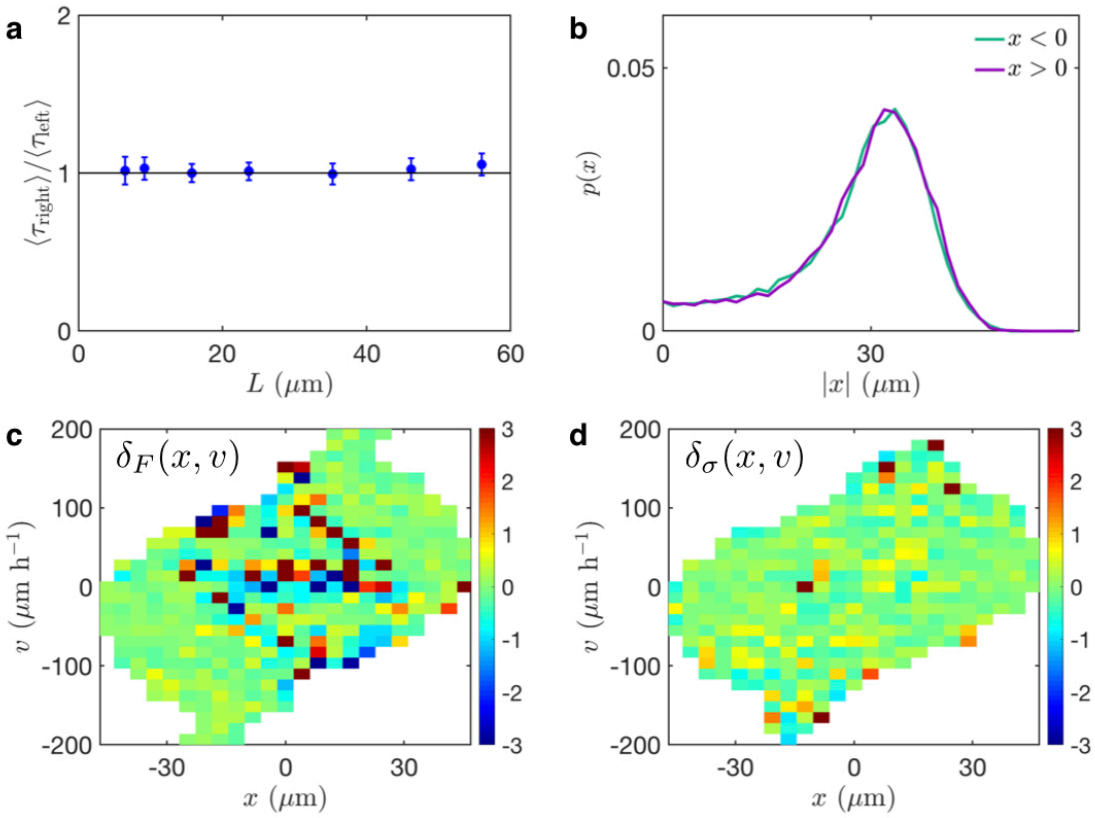


Figure S11 | Test of the inversion symmetry (MDA-MB-231). a, The ratio of average dwell times on the right and left side of the micropattern as a function of bridge length. Error bars denote bootstrap errors. **b**, The position probability distribution $p(x)$ as a function of $|x|$, shown for both $x > 0$ and $x < 0$. **c**, The relative symmetry deviation $\delta_F(x, v)$ of the deterministic term. **d**, The relative symmetry deviation $\delta_\sigma(x, v)$ of the noise strength. **b**, **c**, **d** correspond to data from *L* = 35 µm.

3.5. Comparison to a global inference of the deterministic term

To test the robustness of our results, we also employ an alternative method to estimate the deterministic term from the trajectories[11,12], which does not require a discrete binning of the phase space. Specifically, we define the quantity

$$\epsilon^2 = \langle\langle \left[\frac{v(t+\delta t) - v(t)}{\delta t} - G(x, v; \{p\})\right]^2 \rangle_t \rangle_{\text{cells}} \qquad (S23)$$

As the averages run over both time and cells, ϵ^2 is a sum of the square deviation of the fitting function from all individual observations. By minimising ϵ^2 with respect to the parameters $\{p\}$ of the fitting function $G(x, v; \{p\})$, we perform a global fit across the whole sampled phase space. This different from the binning procedure outlined previously, which gives local best estimates at each phase space position (x, v).

Next, we show explicitly that $G(x, v; \{p\})$ converges to $F(x, v)$ by substituting the discrete Itô-differential equation (S11):

$$\epsilon^2 = \langle\langle \left[F\big(x(t), v(t)\big) + \frac{\sigma\big(x(t), v(t)\big)\Delta W(t)}{\Delta t} - G(x, v; \{p\})\right]^2 \rangle_t \rangle_{\text{cells}} \qquad (S24)$$

The minimisation procedure then implies that for each parameter p,

$$\frac{\partial \epsilon^2}{\partial p} = \langle\langle\, 2\big[F\big(x(t), v(t)\big) - G(x, v; \{p\})\big]\, \rangle_t \rangle_{\text{cells}} = 0 \qquad (S25)$$

since $\langle\langle\, \sigma(x(t), v(t))\Delta W(t)/\Delta t\, \rangle_t \rangle_{\text{cells}} = 0$. The stochastic term therefore does not affect the minimisation procedure. Since our system is bounded, we use a Fourier expansion of the form

$$G(x, v; \{p\}) = \sum_{n=0}^{N} \sum_{m=0}^{M} \big[p_{1,n}\sin(nk_x x) + p_{2,n}\cos(nk_x x)\big]\big[p_{3,m}\sin(mk_v v) + p_{4,m}\cos(mk_v v)\big] \quad (S26)$$

where $k_{x,v} = \pi/\Omega_{x,v}$. $\Omega_{x,v}$ are typical system sizes in x and v that define the resolution of the expansion. Here, we use Ω_x = 100 µm and Ω_v = 250 µm/h as typical system sizes. To limit the number of terms in the expansion, we again enforce the symmetry constraint of the system (equation (S20)).

This leads to the symmetrised fit function

$$G_{\mathrm{sym}}(x,v;\{\alpha,\beta\}) = \sum_{n=0}^{N}\sum_{m=0}^{M}[\alpha_{nm}\sin(nk_x x)\cos(mk_v v) + \beta_{nm}\cos(nk_x x)\sin(mk_v v)] \quad (S27)$$

Importantly, we find that this alternative inference procedure reproduces all of the key qualitative features of the deterministic dynamics that we determined using the binning procedure. Indeed, the qualitative features of $F(x,v)$ are well captured by the expansion, for all tested hyperparameters *N* and *M* (Fig. S12). Similarly, the nullclines of the function show similar features to the binned results within the sampled phase space, including the characteristic proximity of the $\dot{x}$ and $\dot{v}$ nullcline at the positions corresponding to the adhesive islands, indicating that the inferred dynamics is close to a transition between limit cycle and bistable dynamics (Fig. S13). In agreement with the binned dynamics, the deterministic trajectories also perform limit cycle oscillations in the case of MDA-MB-231 cells, while we find two stable fixed points in the dynamics of MCF10A.

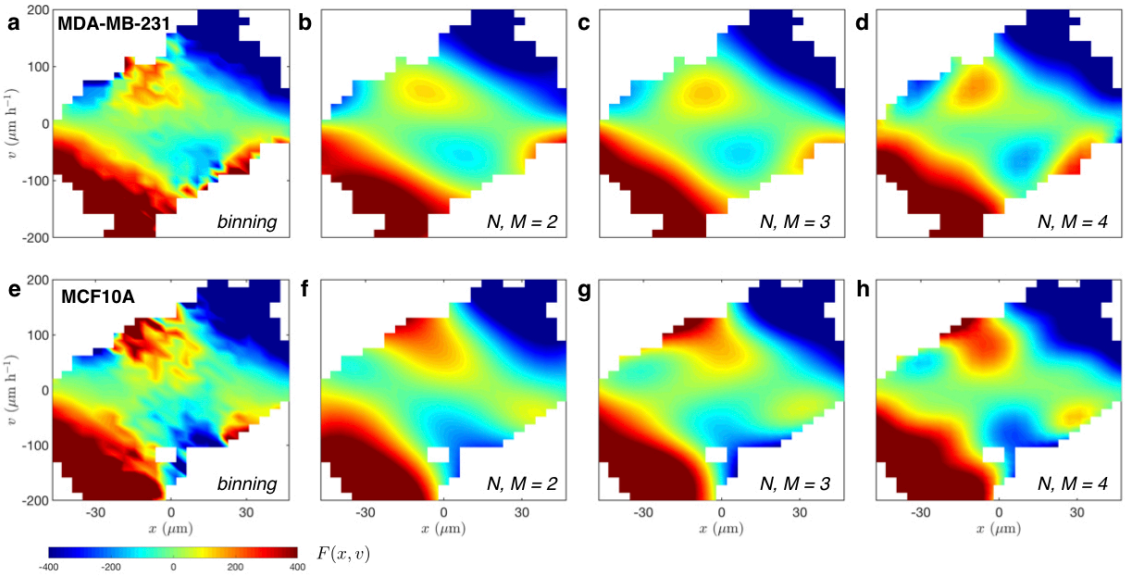


Figure S12 | Inferred deterministic term using different procedures (*L* = 35 µm). a, **e**, The result the binning procedure, shown as a linear interpolation. **b-d**, **f-h**, Results from the Fourier expansion of the deterministic dynamics up to orders *N, M* = 2, 3, 4 respectively. All panels are in units of µm/h^2. Panels **a-d** correspond to MDA-MB-231, and panels **e-h** to MCF10A data. For the Fourier expansion, we show only those regions of the phase space that are sampled in experiment, to avoid extrapolation and to allow a direct comparison to the results obtained by binning.

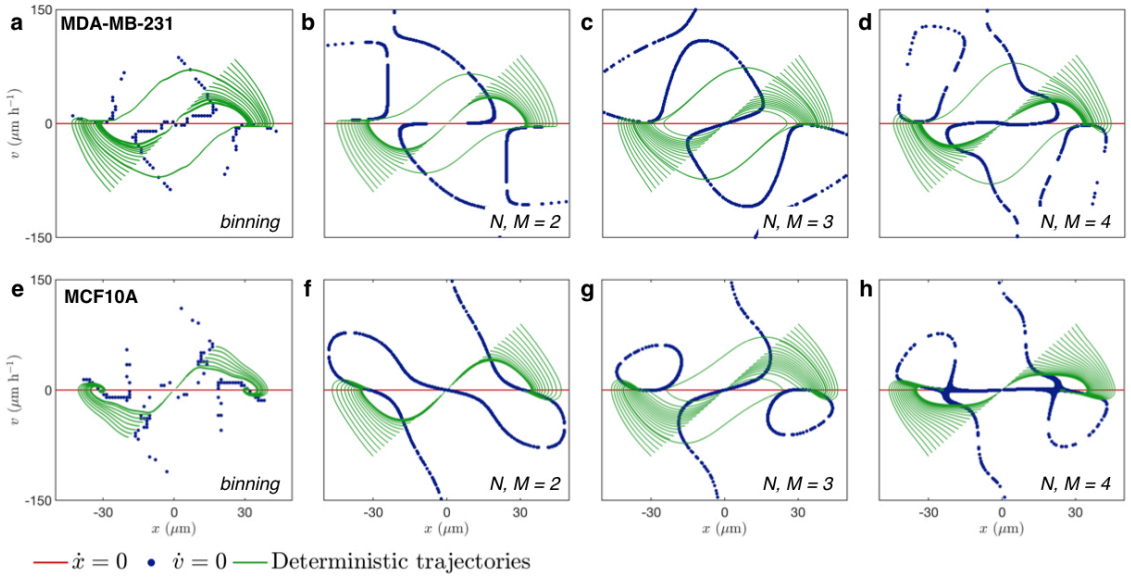


Figure S13 | **Nullclines and sample trajectories of the deterministic term inferred using different procedures (*L* = 35 µm). a**, **e**, The nullclines (blue, red) of the deterministic term obtained using the binning approach which has been interpolated as outlined in section S3.3. **b-d**, **f-h**, Nullclines of the Fourier expansion of the deterministic dynamics up to orders *N, M* = 2, 3, 4 respectively. The deterministic trajectories (green) have been calculated as specified in section S3.3. Panels **a-d** correspond to MDA-MB-231, and panels **e-h** to MCF10A data.

3.6. Consequences of the finite sampling rate

Throughout this work, cell trajectories are sampled at a time interval Δt = 10 min. In general, the parameters of inferred stochastic equations of motion may depend on the value of Δt, which was shown previously for the Ornstein-Uhlenbeck process[9]. Here, we derive numerical finite-time corrections for the deterministic term, to show that our qualitative conclusions are not affected by the discretisation effects.

Suppose the deterministic term of the "true" stochastic process, that occurs in continuous time, is given by $F\left(\left\{p_i^{(\mathrm{true})}\right\}\right)$, with N parameters $\left\{p_i^{(\mathrm{true})}\right\}$, $i = 1 \ldots N$. Then we assume that the model that is inferred from experimental data taken at a finite time interval Δt takes the form

$$F_{\mathrm{expt}} = F\left(\left\{p_i^{(\mathrm{expt})}\right\}\right) = F\left(\left\{\alpha_i^{(\mathrm{true})}(\Delta t) p_i^{(\mathrm{true})}\right\}\right) \qquad (S28)$$

where the parameter-specific rescaling factor $\alpha_i^{(\mathrm{true})}$ in general depends on both the time interval Δt and the value of the true parameter $p_i^{(\mathrm{true})}$. For the Ornstein-Uhlenbeck process, an analytical treatment is possible, and the rescaling factor $\alpha_\gamma^{(\mathrm{true})}$ for the inverse

persistence time γ can be found[9]. For more complex, numerical models this exact treatment is not feasible. To estimate the finite-time corrections for our model, we therefore develop a numerical iterative three-step scheme to estimate $\left\{\alpha_i^{(\mathrm{true})}\right\}$.

Specifically, in step 1, we simulate $F\left(\left\{p_i^{(\mathrm{expt})}\right\}\right)$ with a small time interval $dt = 0.001\mathrm{h}$, then sample the simulated trajectories at a finite time interval $\Delta t = 0.16\mathrm{h} \approx 10\mathrm{min}$, and then infer a new model $F\left(\left\{\tilde{p}_i^{(1)}(\Delta t)\right\}\right)$. We denote this sequence of steps as SSI, for Simulate-Sample-Infer. The resulting relation between $\tilde{p}_i^{(1)}(\Delta t)$ and $p_i^{(\mathrm{expt})}$ then gives us a first estimate of the finite-time correction factor for each parameter $\alpha_i^{(0)}(\Delta t)$ (step 2). We then obtain a first estimate for the corrected model parameters $\left\{p_i^{(1)}\right\}$ in step 3, which we then use in the next iteration as input parameters in step 1

$$F\left(\left\{p_i^{(n-1)}\right\}\right) \xrightarrow{\mathrm{SSI}} F\left(\left\{\tilde{p}_i^{(n)}(\Delta t)\right\}\right) = F\left(\left\{\alpha_i^{(n-1)}(\Delta t)p_i^{(n-1)}\right\}\right) \qquad \text{(Step 1)}$$

Then, for each i,

$$\alpha_i^{(n-1)}(\Delta t) = \frac{\tilde{p}_i^{(n)}(\Delta t)}{p_i^{(n-1)}} \qquad \text{(Step 2)}$$

$$p_i^{(n)} = \frac{p_i^{(n-1)}}{\alpha_i^{(n-1)}(\Delta t)} \qquad \text{(Step 3)}$$

where $p_i^{(0)} = p_i^{(\mathrm{expt})}$. If enough iterations are performed, we should find that $\alpha_i^{(\infty)} = \alpha_i^{(\mathrm{true})}$ and $p_i^{(\infty)} = p_i^{(\mathrm{true})}$.

To test our iterative scheme, we first verify that it matches the analytical result for the Ornstein-Uhlenbeck process[9] (Fig. S14). If this procedure is performed using conditional averaging for processes with two-dimensional phase space, like our model (equation 1), the results are highly sensitive to the measurement noise present in the initial experimental parameters. We therefore employ the functional expansion introduced in section S3.5. We verified numerically that the discretisation effects in the functional expansion and the conditional averaging are identical for the Ornstein-Uhlenbeck process (Fig. S14).

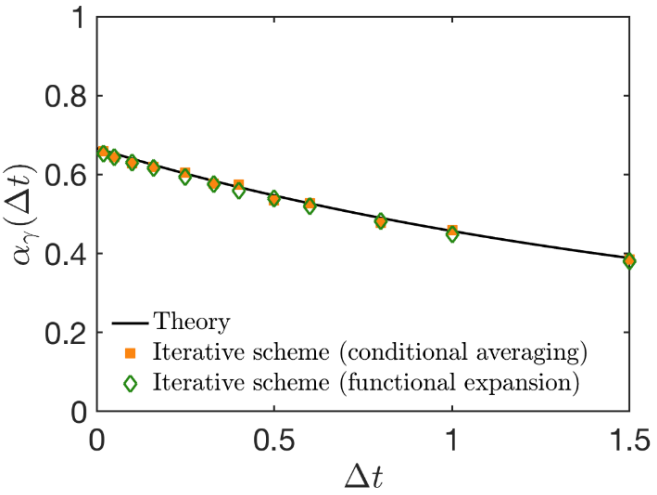


Figure S14 | Finite-time correction factor for the Ornstein-Uhlenbeck (OU) process. We infer the correction factor $\alpha_\gamma(\Delta t)$ for the frictional component γ of the OU process $\dot{v} = -\gamma v + \sigma\eta(t)$ (here $\gamma = \sigma = 1$) using the iterative scheme with two inference methods: conditional averaging (equation S15) and a functional expansion (equation S23) with a polynomial basis. Both results coincide with the theoretical prediction[9].

Next, we simulated a process given by the functional expansion with the same terms as in our experimental system, but with known parameters and linear noise. To keep the number of fitting parameters at a minimum, we use the Fourier expansion (equation S27) with hyperparameters $M, N = 1$:

$$F(x,v) = p_1 \sin(k_x x) + p_2 \sin(k_v v) + p_3 \sin(k_x x)\cos(k_v v) + p_4 \cos(k_x x)\sin(k_v v) \quad (S29)$$

To test our iterative scheme for this model, we simulate this model with known parameters $p_i^{(\mathrm{true})}$, sample the trajectories and infer "experimental" parameters $p_i^{(\mathrm{expt})}$. By applying our iterative scheme, we verify that the corrected parameters converge to the known true parameters after just a few iterations (Fig. S15a).

Finally, we derive the finite-time correction for the experimentally inferred model, which converges similarly to the previous example, and thus yields the correction of the deterministic term (Fig. S15d). Note, the corrected dynamics is qualitatively very similar to the experimentally inferred dynamics, with only a small shift in parameters. Applying the same iterative scheme to the multiplicative noise appears to be technically more involved, and we therefore restrict our analysis here to the deterministic term. However, we note that sampling tends to lead to an underestimation of the noise strength[9] and we thus expect the magnitude of $\sigma(x,v)$ to be increased by a similar finite-time correction. Throughout the manuscript, we show the non-corrected terms obtained directly from conditional averaging,

which should therefore be interpreted as the terms of the finite-difference version of equation (1). The resulting finite-time effects in the integrated dynamics predicted by this equation may also explain the small mismatch seen in the velocity correlation function (Fig. 3g, S31).

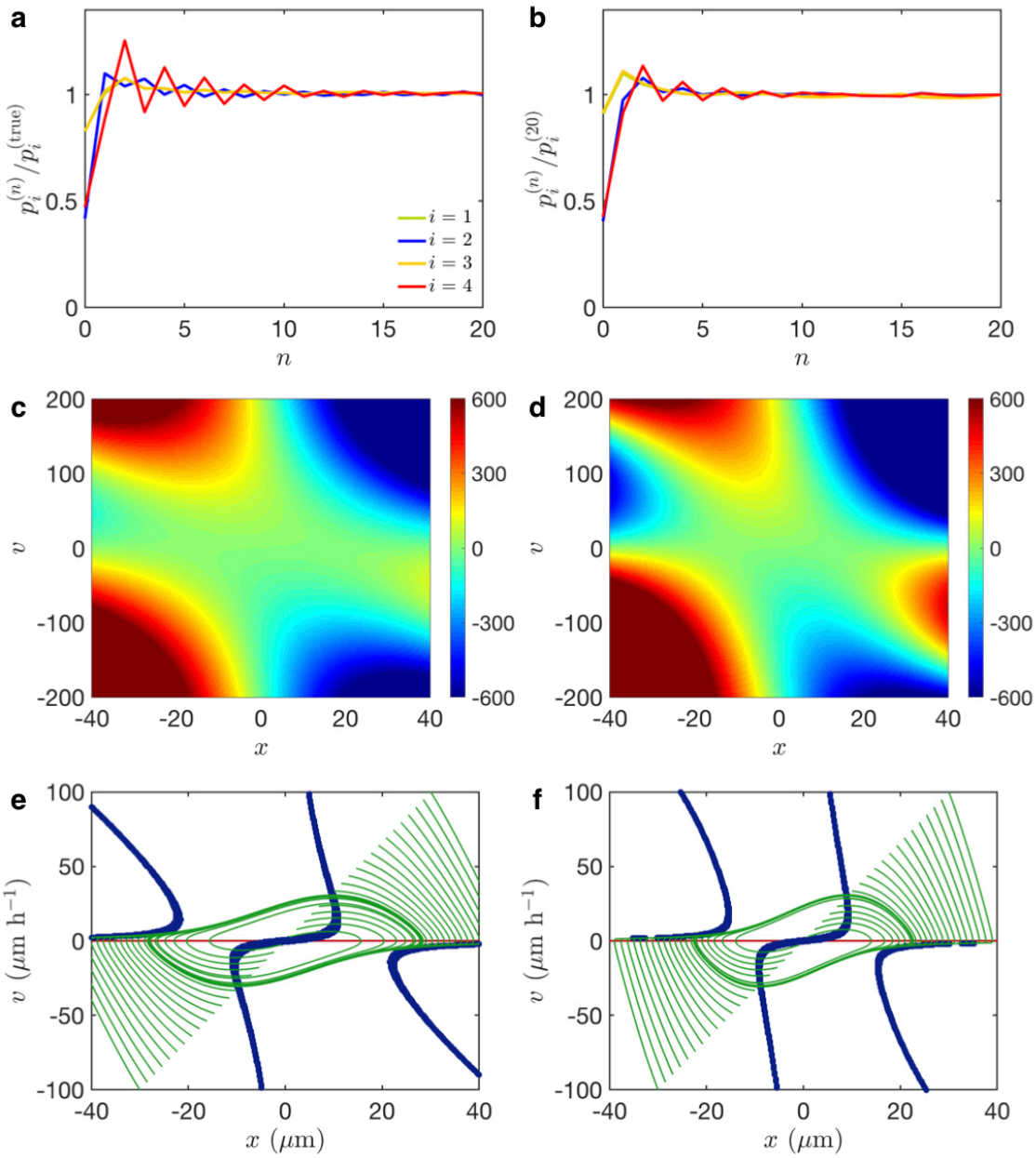


Figure S15 | Finite-time corrections for the deterministic term. a, First, we verify that for known model parameters in a Fourier expansion (equation S29), we can recover the correct parameters after iterating. **b**, For the experimental data (MDA-MB-231, L = 35 µm), the parameters similarly approach a steady value after within a few iterations, indicating that the iterative scheme has

converged. **c**, Experimentally measured deterministic term $F\left(\left\{p_i^{(\mathrm{expt})}\right\}\right)$ in units of μm/h^2. **d**, Similar plot for the corrected term $F\left(\left\{p_i^{(20)}\right\}\right)$. **e**, **f**, Trajectories (green) and nullclines (red, blue) of the deterministic dynamics shown in **c** and **d** respectively.

To show explicitly by how much the corrected term varies from the finite-time version, we calculate the relative difference

$$\delta(x,v) = \frac{F\left(x,v;\left\{p_i^{(\mathrm{expt})}\right\}\right) - F\left(x,v;\left\{p_i^{(20)}\right\}\right)}{F\left(x,v;\left\{p_i^{(20)}\right\}\right)} \qquad (S30)$$

This measure shows in large regions of the sampled phase space, the relative deviation is small, while there is a large fractional variation around the nullclines where $F\left(\left\{p_i^{(20)}\right\}\right) = 0$ and $\delta(x,v)$ thus diverges (Fig. S16). To quantify the average deviation, we calculate

$$\langle\delta(x,v)\rangle_{x,v} = \frac{\left\langle\left|F\left(x,v;\left\{p_i^{(\mathrm{expt})}\right\}\right) - F\left(x,v;\left\{p_i^{(20)}\right\}\right)\right|\right\rangle_{x.v}}{\left\langle\left|F\left(x,v;\left\{p_i^{(20)}\right\}\right)\right|\right\rangle_{x,v}} \approx 0.39$$

where we only included the region of phase space that are sampled in the experiment (Fig. S16). Thus, we conclude that average deviation in magnitude of the deterministic term due to finite time effects is of the order of 40%. However, we also note that the finite time effects appear to mainly constitute a small shift in the positions of the nullclines (Fig. S15e, f), and the average fractional change thus arguably gives an overestimate of "how different" the finite time and continuous time terms are in most of phase space.

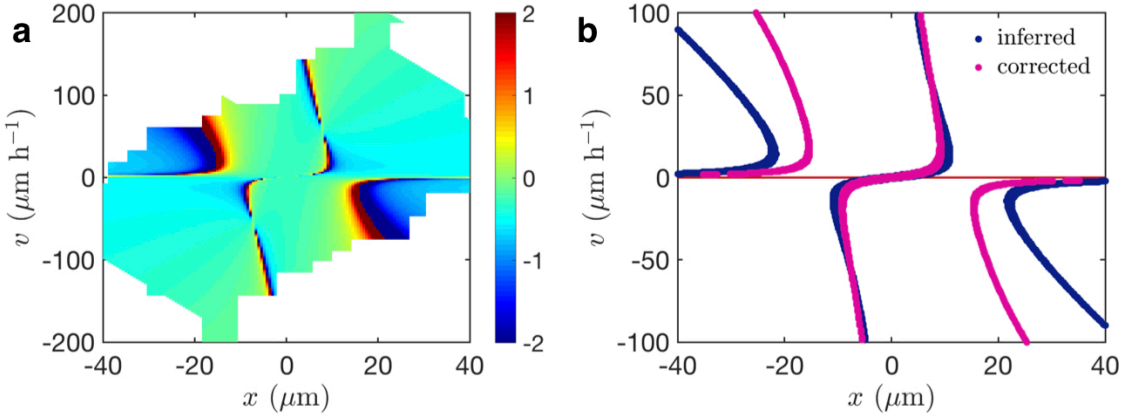


Figure S16 | Relative finite time effects in the deterministic term a, The relative deviation $\delta(x, v)$ calculated using equation (S30), and shown for the region of phase space that is sampled in experiment. **b**, The nullclines of the inferred (blue) and corrected (purple) deterministic term.

3.7. Consequences of localisation errors

We define the cell position as the position of the centre of the nucleus, which is recorded using automated tracking of the fluorescence signal. The detection of the position of the nucleus centre is subject to detection errors. We estimate an upper bound of $\sigma_{\mathrm{pos}} = 2\mu\mathrm{m}$ for this error. To investigate whether this localisation error affects the inference of the stochastic model, we perform a simulation of the inferred model with an added stochastic localisation error:

$$\begin{aligned} x_{i+1} &= x_i + v_i\Delta t + \sigma_{\mathrm{pos}} w_i \\ v_{i+1} &= v_i + F(x_i, v_i)\Delta t + \sigma(x_i, v_i) r_i \sqrt{\Delta t} \end{aligned} \qquad (S31)$$

where we model the localisation error as a Gaussian random variable w_i with mean 0 and variance 1 that is added to the positions. None of the central migration statistics which depend directly on the recorded positions are affected by localisation errors of this magnitude (Fig. S17 a, b, d). The correlation function of the inferred noise exhibits a small negative correlation for time-separations equal to the sampling interval ($t = \theta$), which is not present in the simulation without localisation errors. For a thorough discussion of this effect, we refer the reader to ref. 9.

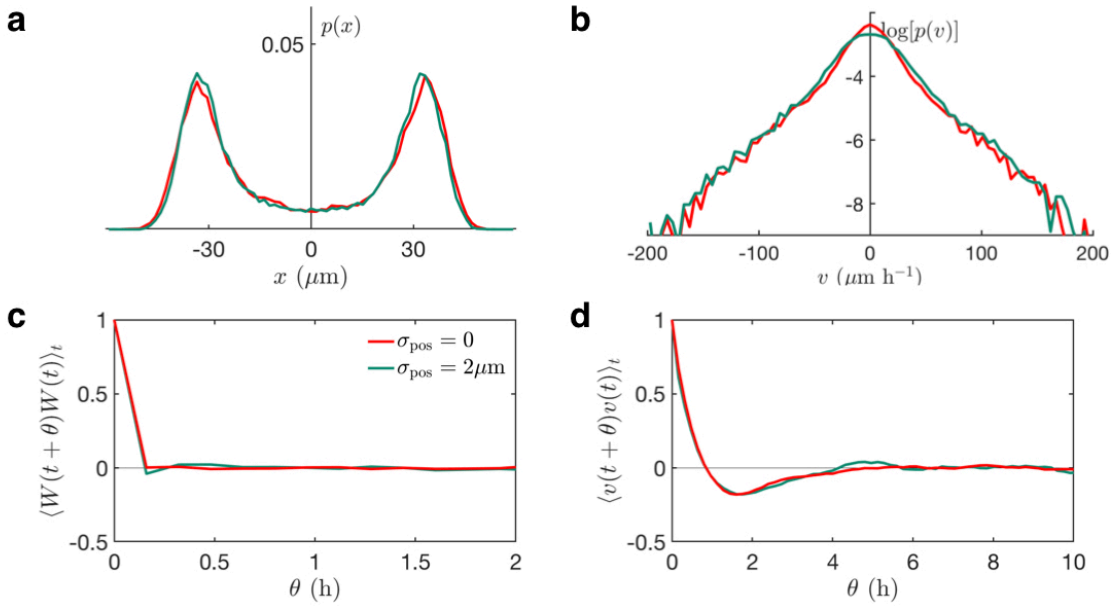


Figure S17 | Comparison of simulated statistics with and without localisation errors (*L* = 35 µm, MDA-MB-231). a, The position probability distribution $p(x)$. The red curves in all panels correspond

to the standard model simulation (equation (S19)), while the green curves correspond to a simulation including localisation errors (equation (S31)). **b**, The velocity probability distribution $p(v)$, plotted semi-logarithmically. **c**, The correlation function of the inferred noise from each model simulation. **d**, The velocity correlation function.

3.8. Relative magnitude of deterministic and stochastic terms

To quantify the relative magnitudes of deterministic and stochastic contributions, we define the root-mean-square ratio $r(x,v)$ as the ratio of the deterministic and stochastic terms of the discrete equation of motion (equation S13):

$$r(x,v) = \frac{\sqrt{\langle F\big(x(t),v(t)\big)\Delta t\rangle^2}}{\sqrt{\langle \sigma^2\big(x(t),v(t)\big)\Delta W^2(t)\rangle}} = \frac{|F(x,v)|\sqrt{\Delta t}}{\sigma(x,v)} \qquad (S32)$$

where we have used Ito's lemma $\langle \Delta W(t)^2\rangle = \Delta t$. We find that the noise dominates in the regions corresponding to the adhesive islands (Fig. S18), as indicated schematically in Fig. 4e.

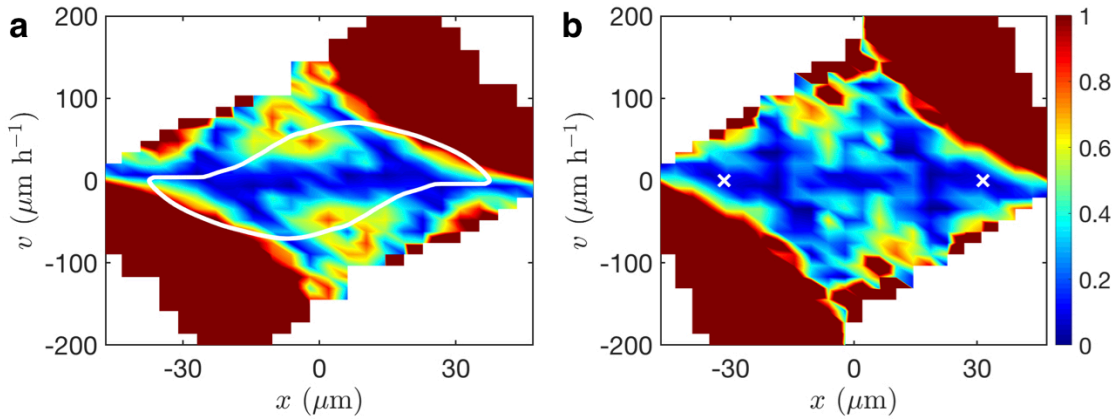


Figure S18 | Ratio of deterministic and stochastic contributions (*L* = 35 µm). Derived from the symmetrised functions F and σ. To clearly show the regions where the noise dominates, the colourscheme has a cutoff at $r = 1$. **a**, MDA-MB-231. The limit cycle is indicated by the white line. **b**, MCF10A. The two stable fixed points are marked by white crosses. The colour scale applies to both panels.

3.9. Role of the noise

For MCF10A cells, the role of the stochastic term is self-evident: without noise, the system relaxes to the stable fixed points, and there are no transitions.

In contrast, for MDA-MB-231, the deterministic dynamics still leads to oscillations even in the absence of noise. To investigate the role of stochasticity in this system, we therefore scale the stochastic term $\sigma(x, v)$ by a pre-factor λ in the range from $\lambda = 0$ (deterministic motion) to $\lambda = 1$ (fully stochastic system) (Fig. S19). In the purely deterministic system, the limit cycle dynamics leads to repeated oscillations of the velocity correlation. As we increase λ, the complex oscillatory velocity correlation of the deterministic dynamics is smoothly deformed into the highly damped correlation function observed experimentally. At intermediate λ, we find that the correlation function performs decaying oscillations. While there likely is no experimental procedure to tune λ, this theoretical analysis allows us to compare the role of the stochastic contributions in the oscillatory MDA-MB-231 dynamics. Specifically, these findings reveal that the noise abrogates the long-time correlations encoded in the deterministic contribution beyond a single oscillation. In addition, the noise plays a crucial role by speeding up the slow dynamics of reversals on the islands in MDA-MB-231, and by providing excitations out of the basins of attraction of the fixed points in MCF10A. In both cases, the noise thus contributes to setting the key time-scale of the system – the average dwell time (Fig. 2b).

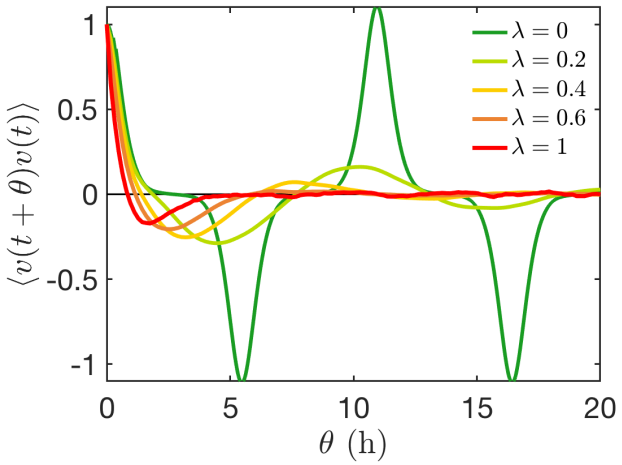


Figure S19 | Velocity correlation functions of the dynamics with different noise strengths (MDA-MB-231). The magnitude of the stochastic term $\sigma(x, v)$ is scaled by a prefactor λ in the range from $\lambda = 0$, corresponding to purely deterministic dynamics (green), to $\lambda = 1$, corresponding to the full stochastic system (equation (1)) (red). The deterministic correlation function is obtained from the symmetrised results (Section 3.4)

4. Characterisation of free 2D cell migration

Cell migration is often studied on unconfined 2D surfaces. The migration behaviour in such an unstructured environment can therefore serve as a motility benchmark for a given cell type. To characterise the free migration behaviour of our standard cell type (MDA-MB-231) on a 2D surface, we analyse the trajectories of cells migrating on isotropic 2D surfaces with the same surface properties as in the case of the two-state confinement (Supplementary Movie S8; for experimental details see Section 1.7).

Here, we again adopt a data-driven approach to find a stochastic equation of motion that describes the migration, and we follow the method first proposed by *Selmeczi et al.*[2] First, we verify that the experimental data fulfils rotational as well as space- and time-translational symmetry (Fig. S20a-c). Therefore, we can average any observations over orientation, space, and time, and any model describing the data should also possess these symmetries. Next, we find that the velocity correlation of the migration follows a bi-exponential decay, which was previously found in both keratinocytes (HaCaT) and fibroblasts (NHDF)[2] (Fig. S20d). In contrast to HaCaT and NHDF cells, however, velocities are much more broadly distributed here, up to speeds of 150 µm h^{-1} (Fig. S20e). Measuring the average acceleration as a function of speed $\langle \dot{v}|v\rangle$ in the reference frame of the cell, we find that the acceleration along the direction perpendicular to the direction of motion vanishes, while the acceleration along the direction of motion takes a non-linear form that is well fitted by a third-order polynomial (Fig. S20f). We also measure the noise strength of fluctuations around the deterministic trajectories, defined as $\sigma(v) = \sqrt{\Delta t \langle [\dot{v} - \langle \dot{v}|v\rangle]^2 | v\rangle}$. We find that the parallel and perpendicular components of the noise collapse onto a single curve, which is well described by a fit function of the form

$$\sigma(v) = \begin{cases} \sigma_0 + \sigma_1 v & v < v_{\text{plateau}} \\ \sigma_{\text{plateau}} & v \geq v_{\text{plateau}} \end{cases} \qquad (S34)$$

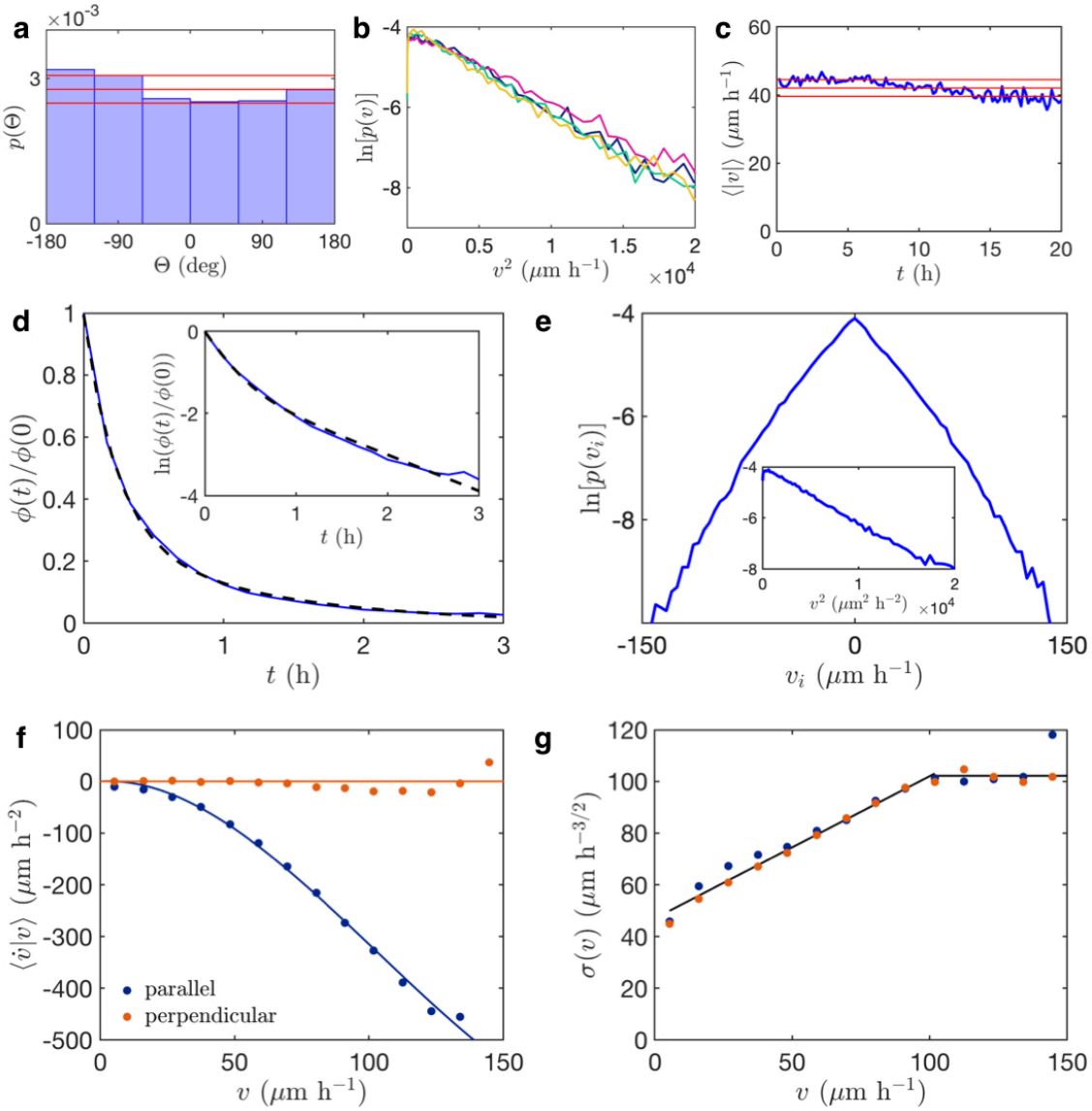


Figure S20 | Experimental characterisation of free cell migration. a, Probability distribution of angles in which the cells move in the laboratory frame of reference (lab frame). Here, the angles in which cells have moved (angle of the vector connecting initial and final position) for *N* = 728 cells are shown together with the average and standard deviations from the average of the numbers shown by the columns (red lines). **b**, Probability distributions of speeds in the lab frame measured in four different regions of the substrate (corresponding to the four colours). Plotted semi-logarithmically against the square of the speed. **c**, Average migration speed as a function of time. Red lines indicate the average and standard deviation of the speed. Together, **a**-**c** indicate that the migration statistics are isotropic and uniform in space and time. **d**, Velocity correlation function $\phi(t) = \langle v(s+t)v(s)\rangle_s$. The black dashed line indicates a bi-exponential fit, $\phi_{\mathrm{fit}}(t) = \phi_1 e^{-t/P_1} + \phi_2 e^{-t/P_2}$. Inset: semi-logarithmic plot of the same quantities. **e**, Marginal probability distribution $p(v_i) = \int p(v_i, v_j)\mathrm{d}v_j$ of velocities v_i in both the x- and y-directions ($i = x, y$). Inset: Probability distribution of speeds, plotted semi-logarithmically against the square of the speed. **f**, Average

acceleration as a function of speed in the frame of reference of the moving cell. The parallel and perpendicular directions are defined by the velocity of the cell. The solid lines are fits of the form $F_{\parallel} = p_0$ and $F_{\perp} = p_1 v + p_2 v^2 + p_3 v^3$. **g**, Noise component $\sigma(v) = \sqrt{\Delta t \langle [\dot{v} - \langle \dot{v}|v \rangle]^2 | v \rangle}$ as a function of speed in the two directions. The black line is a fit of equation (S34) to the average of the parallel and perpendicular components.

The experimental data in Fig. S20 allows us to constrain the equation of motion that describes the migration. To do so, we systematically increase the complexity of models fitted to the dynamics. We start with a Langevin equation with velocity dependent noise,

$$\frac{\mathrm{d}\vec{v}}{\mathrm{d}t} = \vec{F}(\vec{v}) + \sigma(\vec{v})\vec{\eta}(t) \qquad \text{(Model A)}$$

Using the empirical fits of $\vec{F}(\vec{v})$ and $\sigma(\vec{v})$, a simulation of this model shows that it is not able to capture the bi-exponential correlation function, as it only captures the first time-scale of the correlation (Fig. S21a). We therefore increase the model complexity to a non-Markovian model that includes a memory kernel. A relatively simple memory kernel with a bi-exponential velocity correlation function that was found to describe the dynamics of HaCaT cells[2] is:

$$\frac{\mathrm{d}\vec{v}}{\mathrm{d}t} = -\beta\vec{v} + \alpha^2 \int_{-\infty}^{t} e^{-\gamma(t-s)}\vec{v}(s)\mathrm{d}s + \sigma(\vec{v})\vec{\eta}(t) \qquad \text{(Model B)}$$

However, this model has a strictly linear dependence of the average acceleration on speed[2]. In our case, this dependence is non-linear, as was previously found for NHDF[2] and *Dictyostelium discoideum*[13] cells. To accommodate this non-linearity, we follow previous work[2,13] by introducing a speed-dependent $\beta(v) = \beta_0 + \beta_1 v$, which we call Model C. This model reasonably fits all the observables (Fig. S21).

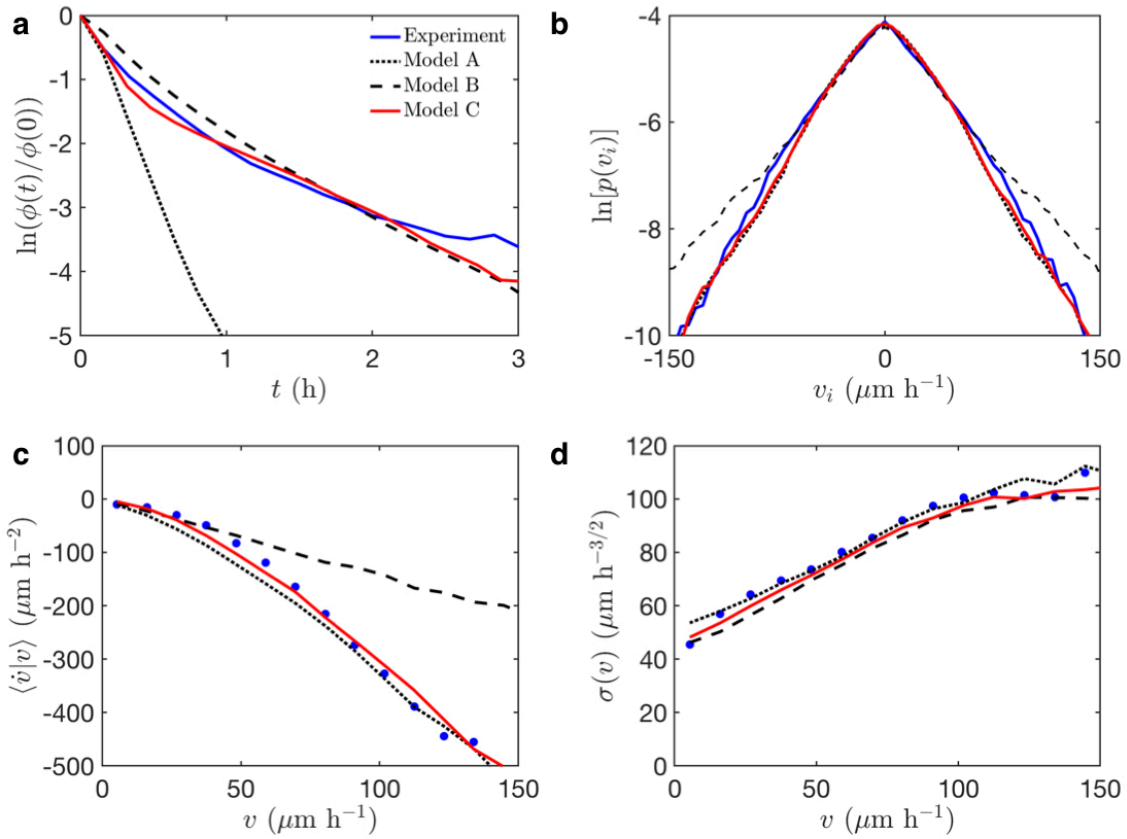


Figure S21 | Experimental-to-model comparison for various model candidates for a freely migrating cell in 2D. a, Velocity correlation function on a semi-logarithmic plot. **b**, Probability distribution of velocities. **c**, Average acceleration as a function of speed. **d**, Noise strength as a function of speed.

In table S2, we summarise the parameters of Model C for MDA-MB-231 cells and also compare them to the parameters found previously for HaCaT and NHDF cells that are described by a similar memory kernel model[2].

Cell line	**MDA-MB-231**	**HaCaT**[2]	**NHDF**[2]
α (h^{-1})	2.2	1.5	2.9
$\beta(v)$ (h^{-1})	$0.8 + 0.07v$	4.1	$2.1 + 0.06v$
γ (h^{-1})	2.2	1.0	2.1
$\sigma(v)$ (μm h$^{-3/2}$)	$\begin{cases} 42 + 1.8v & v < 100\ \mu\text{m h}^{-1} \\ 350 & v \geq 100\ \mu\text{m h}^{-1} \end{cases}$	$11 + 1.3v$	$\sigma_{\parallel}(v) = 7.3 + 2.8v$ $\sigma_{\perp}(v) = 7.3 + 0.9v$

Table S2 | Model parameters for Model C of our cancer cell line compared to those of keratinocytes (HaCaT) and fibroblasts (NHDF) from the literature[2]. These model parameters are corrected for finite-time effects, since they are obtained by parameter sweeps of the model equation, integrated with a small time step, and subsequently sampled at $\Delta t = 10$min[2,9].

5. Comparison to other models

To arrive at the general form of our model (equation (1)), we aim to systematically increase the model complexity until we reach a model that describes the experimental statistics well. In particular, we test the applicability of stochastic equations of motion in the following order:

i. First order equations of motion
ii. Second order model with a deterministic term separable in position and velocity
iii. Second order model with constant noise strength

In the next three subsections, we show that none of these models are able to describe the data well. Within this scheme, we view equation (1) as the next-simplest stochastic equation of motion. However, alternative descriptions of the data may be possible, such as a generalisation of the description for 2D motility in the form of a memory kernel with position-dependent parameters. In the last subsection, we also discuss a comparison of the deterministic term of equation (1) to the van der Pol oscillator, which is the classic textbook example of a limit cycle oscillator.

5.1. First order equations of motion

An alternative theoretical ansatz to equation (1) would be to consider a first order equation of motion, in analogy to the description of overdamped systems such as Brownian particles. As a simple, intuitive model, one might think that the cell behaves like a Brownian particle diffusing in an external potential set by the geometry of the micropattern, with an equation of motion for each bridge length L

$$\dot{x} = -\partial_x V_L(x) + \sigma\eta(t) \qquad (S35)$$

where $\langle\eta(t)\rangle = 0$. The hopping process would then correspond to stochastic transitions in this energy landscape (Fig. S22a). A physical interpretation of this effective potential could be elastic energy constraints placed by the membrane deformations that occur during the transition, while the noise would be generated by the motility machinery of the cell[14,15].

To make progress analytically, we assume that the noise is uncorrelated: $\langle\eta(t)\eta(t')\rangle = \delta(t-t')$. The steady state probability distribution $p_L(x)$ is then given by the Boltzmann distribution. Since we have measured the probability distribution, we can invert the usual Boltzmann expression to yield an inferred potential

$$V_L(x)/\sigma = -\ln p_L(x) + C \qquad (S36)$$

The resulting potentials (Fig. S22b) do indeed have the qualitative features one would expect in a typical escape problem in a double-well potential. However, we observe that the effective activation energies of the escape do not satisfy the condition $Q_L/\sigma \gg 1$. The standard Kramers expression for the transition rates in terms of potential curvatures is therefore invalid in this case[16]. We can, however, make progress by integrating the corresponding Fokker-Planck equation, which yields the mean-first passage time over the barrier[10,17]

$$\langle \tau_L \rangle = \frac{1}{\sigma} \int_{x_0}^{x_r} \mathrm{d}y \exp\left(\frac{V_L(y)}{\sigma}\right) \int_{x_l}^{y} \mathrm{d}z \exp\left(-\frac{V_L(z)}{\sigma}\right) \qquad (S37)$$

where x_0 is the starting position of the cell, and x_l and x_r are reflecting and absorbing boundaries respectively, with $x_l < x_0 < x_r$. We take x_l as the boundary of the left adhesive island, and $x_r = 0$. To estimate the dwell times, we use quartic fits to the inferred potentials and subsequently use numerical integration to evaluate equation (S37).

Since we are directly comparing dynamical with steady-state quantities, this yields an independent consistency check of the potential landscape model (equation (S35)). However, the model fails dramatically (Fig. S23), both quantitatively in terms of the relative change of dwell times with changing bridge length, as well as qualitatively. The theoretical model predicts an exponential increase of dwell times while the experimental data follows a linear trend.

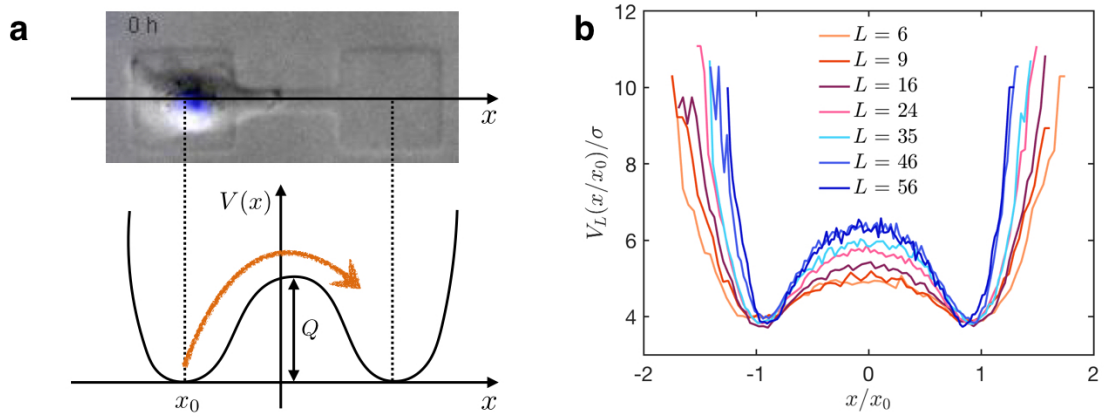


Figure S22 | Effective potential model (MDA-MB-231). a, Scheme of the effective potential model. The process is interpreted as an escape problem over a barrier of height Q. **b**, Effective potentials calculated from the experimental steady state probability distribution. For presentation purposes, the x-axis is normalised by $x_0 = (L + a)/2$, where $a = (36.7 \pm 0.6)\ \mu$m is the side length of the adhesive square of the micropattern.

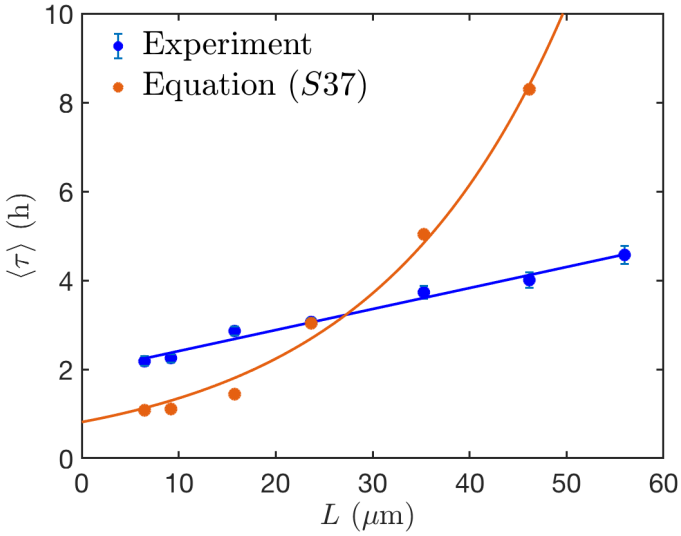


Figure S23 | Experimental dwell times compared to the prediction from the effective potential model (MDA-MB-231). Note that the overall magnitude of the predicted rates is not determined, as the noise strength σ is unknown within this model. Here, we arbitrarily rescale the dwell times using $\sigma = 450\ \mu\mathrm{m\ h}^{-1/2}$ to match the order of magnitude of the observed rates.

Next, we relax some of the assumptions of the model and write down a more general first order equation of motion:

$$\dot{x} = F_L(x) + \sigma_L(x)\eta(t) \qquad (S38)$$

where again $\langle \eta(t) \rangle = 0$ and $\langle \eta(t)\eta(t') \rangle = \delta(t - t')$. To infer the terms of this equation, we take conditional averages: $F_L(x) = \langle \dot{x} \mid x \rangle$ and $\sigma_L^2(x) = \Delta t \langle [\dot{x} - F(x)]^2 \mid x \rangle$. We are indeed able to find these terms (Fig. S24a). However, we find that the inferred noise is correlated over timescales comparable to the correlation timescale of the system (Fig. S24b), given by the correlation time of the velocity correlation function (Fig. 3g). This implies that the white noise assumption of the inference scheme is violated and it is therefore not valid. We therefore conclude that a first order model with white noise is inconsistent with the experimentally observed dynamics.

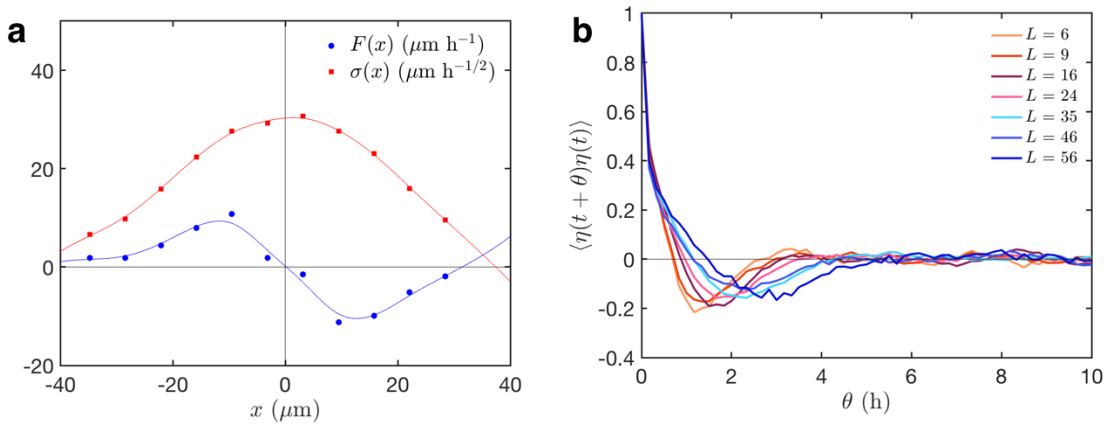


Figure S24 | Inferred first order model (MDA-MB-231). a, Deterministic term (blue dots) and noise strength (red squares) as a function of position for a bridge length of L = 35 µm. **b**, Correlation functions of the inferred noise.

5.2. Second order model with an additively separable deterministic term

In the main text, we state that the deterministic term $F(x,v)$ has a non-separable dependence upon position and velocity. Here, we demonstrate this by showing that an additively separable deterministic term deviates from our inferred function, and that it provides predictions inconsistent with experiment.

An additively separable deterministic term can be expressed as

$$F_{\text{sep}}(x,v) = F_x(x) + F_v(v) \tag{S39}$$

From the symmetry of the system, $F_{\text{sep}}(x,v) = -F_{\text{sep}}(-x,-v)$, and thus $F_x(x) = -F_x(-x)$ and similarly $F_v(v) = -F_v(-v)$. This implies that $F_x(0) = F_v(0) = 0$ and therefore $F_x(x) = F(x, v=0)$ and $F_v(v) = F(x=0, v)$. We can therefore directly reconstruct from $F(x,v)$ what F_{sep} should be, if the deterministic term has the separable form of equation (S30):

$$F_{\text{sep}}(x,v) = F(x, v=0) + F(x=0, v) \tag{S40}$$

This separable model exhibits deterministic dynamics that is qualitatively different to the dynamics of $F(x,v)$ (Fig. S25), and predicts stochastic dynamics inconsistent with experiment, as the simulated trajectories rapidly exit the sampled phase space.

We thus conclude that the inferred deterministic dynamics cannot be captured by a model of this form with a deterministic term that is additively separable in position and velocity. In Supplementary Section S5.1, we ruled out an energy potential formulation based on a first

order equation of motion. Here, we also show that the confined cell migration we observe is incompatible with a second order model which includes an energy potential, as such a model would have to be separable in position and velocity. Our theory therefore rather belongs to the class of non-linear limit cycle oscillators, such as the classic van der Pol oscillator (See section 5.4), which is also not additively separable.

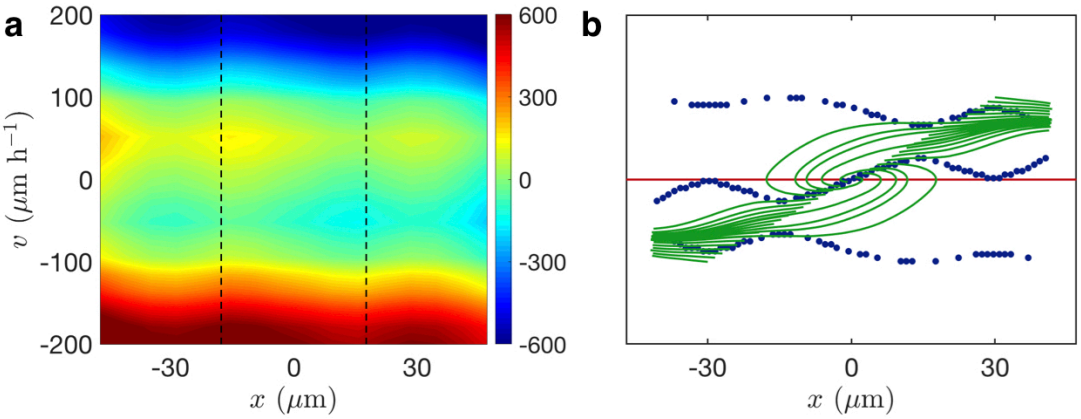


Figure S25 | Inferred dynamics of a model with an additively separable deterministic term. a, The reconstructed separable deterministic term $F_{\mathrm{sep}}(x,v) = F_x(x) + F_v(v)$ in units of µm/h². **b**, The corresponding x- (red line) and v-nullclines (blue dots) together with the deterministic trajectories (green lines).

5.3. Second order model with constant noise strength

Within the framework of the second order model (equation (1)), we consider an approximation of the noise, where we replace the function $\sigma(x,v)$ by simple additive noise with strength σ_0:

$$\frac{dv}{dt} = F(x,v) + \sigma_0\eta(t) \qquad (S41)$$

We therefore infer the constant noise amplitude σ_0 from the experimental data and subsequently compare the model predictions with the experimental observations. We find that $p(x)$ is qualitatively reproduced, but that the time spent on the bridge is overestimated, as we overestimate the probability in the centre of the system but underestimate the probabilities on the adhesive islands (Fig. S26a). The velocity distribution also exhibits slight discrepancies (Fig. S26b).

The qualitative functional form of $S(t)$ is also captured (Fig. S26c), but the tail of the distributions is underestimated, leading to an underestimation of the average dwell time (Fig. S26d). These trends are reproduced on all bridge lengths as well as for the case of

MCF10A cells. Taken together, these results show that a model with linear noise captures the qualitative but not the quantitative features of the dynamics. To fully capture the cell migration on the two-state micropatterns, we thus need a multiplicative noise term.

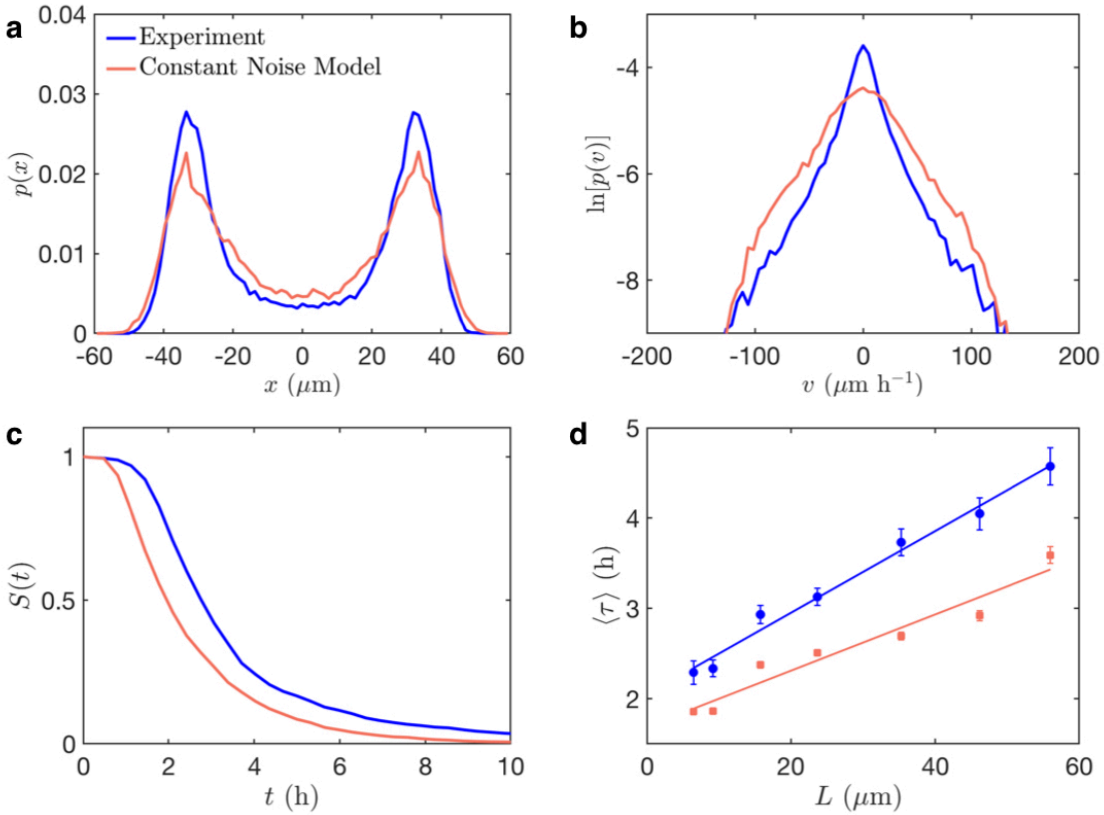


Figure S26 | Statistics of the constant noise strength model (*L* = 35 µm, MDA-MB-231). **a**, Position probability distribution. **b**, Velocity probability distribution, plotted semi-logarithmically. **c**, Survival probability distribution. **d**, Average dwell times as a function of bridge length. Error bars denote bootstrap errors.

5.4. The van der Pol oscillator

The textbook example for a non-linear oscillator that exhibits a limit cycle is the van der Pol oscillator[18]. In the deterministic case, $\dot{v} = F_{\mathrm{vdP}}(x, v)$, where

$$F_{\mathrm{vdP}}(x, v) = \mu(1 - a^2x^2)v \; - \; x \qquad (S42)$$

To compare our inferred deterministic term to this well-known example, we show the deterministic map as well as the nullclines and deterministic trajectories of equation (S30) with parameters μ and a chosen such that the oscillator has a similar period and amplitude as our inferred dynamics for *L* = 35 µm (Fig. S27).

Similar to the deterministic term of the hopping cells, the v-nullcline of the van der Pol oscillator,

$$v = \frac{x}{\mu(1 - a^2x^2)} \tag{S43}$$

gives rise to an unstable fixed point at $x = 0$, and closely approaches the x-nullcline at the maximum of the oscillation, giving rise to the slow dynamics that is characteristic in relaxation oscillators.

As we discuss in the main text, the way the effective friction is controlled in our system is qualitatively different to the van der Pol oscillator. Specifically, $F_{\mathrm{vdP}}(x, v)$ is always a linear function of v, and the constant of proportionality and is determined by the position x. Thus, $F_{\mathrm{vdP}}(x \to 0, v) = \mu v$, corresponding to 'negative' friction for all speeds. In contrast, the cellular oscillator exhibits a non-linear dependence upon v when $x \to 0$, with 'negative' friction only occurring at small speeds.

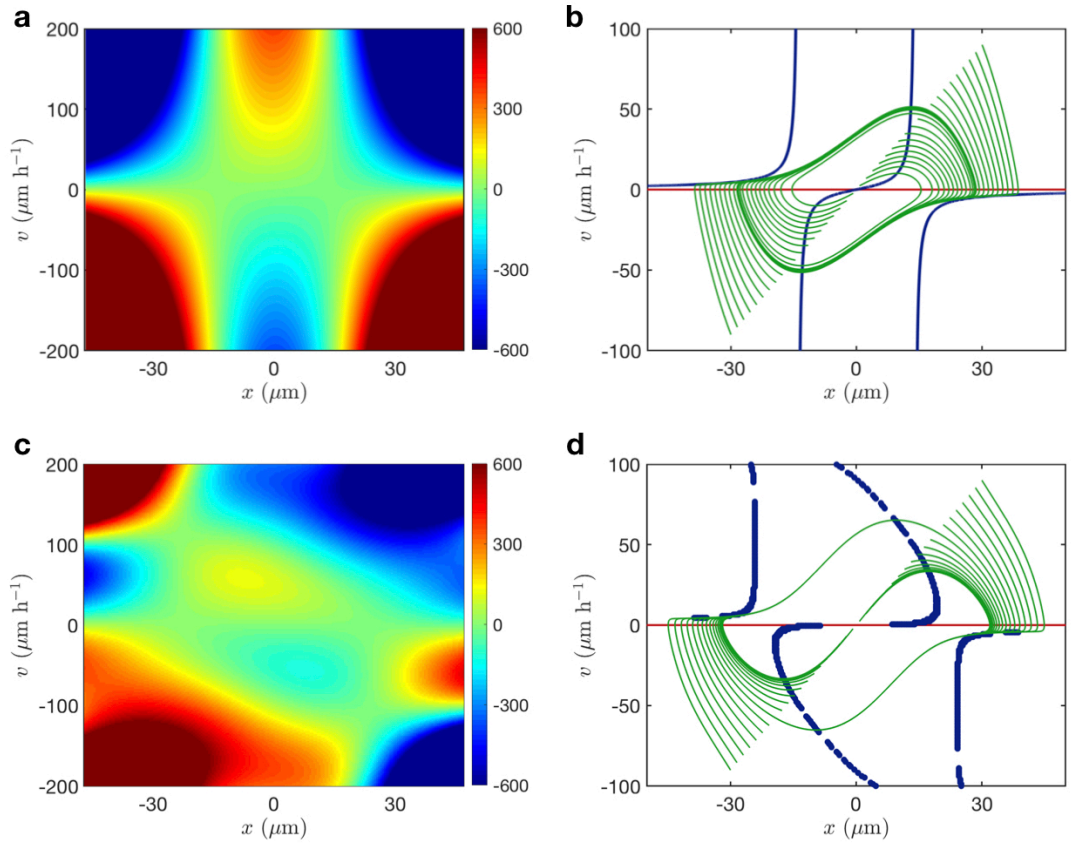


Figure S27 | Comparison of $F(x, v)$ to the van der Pol osillator equation of motion. a, **b**, the van der Pol dynamics $F_{\mathrm{vdP}}(x, v)$ given by equation (S42). **c**, **d**, The inferred deterministic dynamics

$F(x, v)$ of the confined cells (MDA-MB-231, *L* = 35 µm), calculated by fitting to a Fourier expansion as described in section S3.5 (*M, N* = 2). In **a**, **c**, we show the deterministic acceleration as a function of x and v in units of µm/h^2. **b**, **d** show the nullclines $\dot{x} = 0$ (red) and $\dot{v} = 0$ (blue) together with the deterministic trajectories starting from a range of initial conditions. In $F_{\mathrm{vdP}}(x, v)$, the parameters are $\mu = 1.86$ and $a = 0.07$.

6. Supplementary Results

To evaluate the agreement between experiment and model (equation (1)), we compare the survival probability distributions (Fig. S15). In the main text, a number of plots (Figures 2c; 3a, c, e; 4a, d, e) are shown for a single bridge length. Here, we show the same plots for all bridge lengths (Figs. S16-S22).

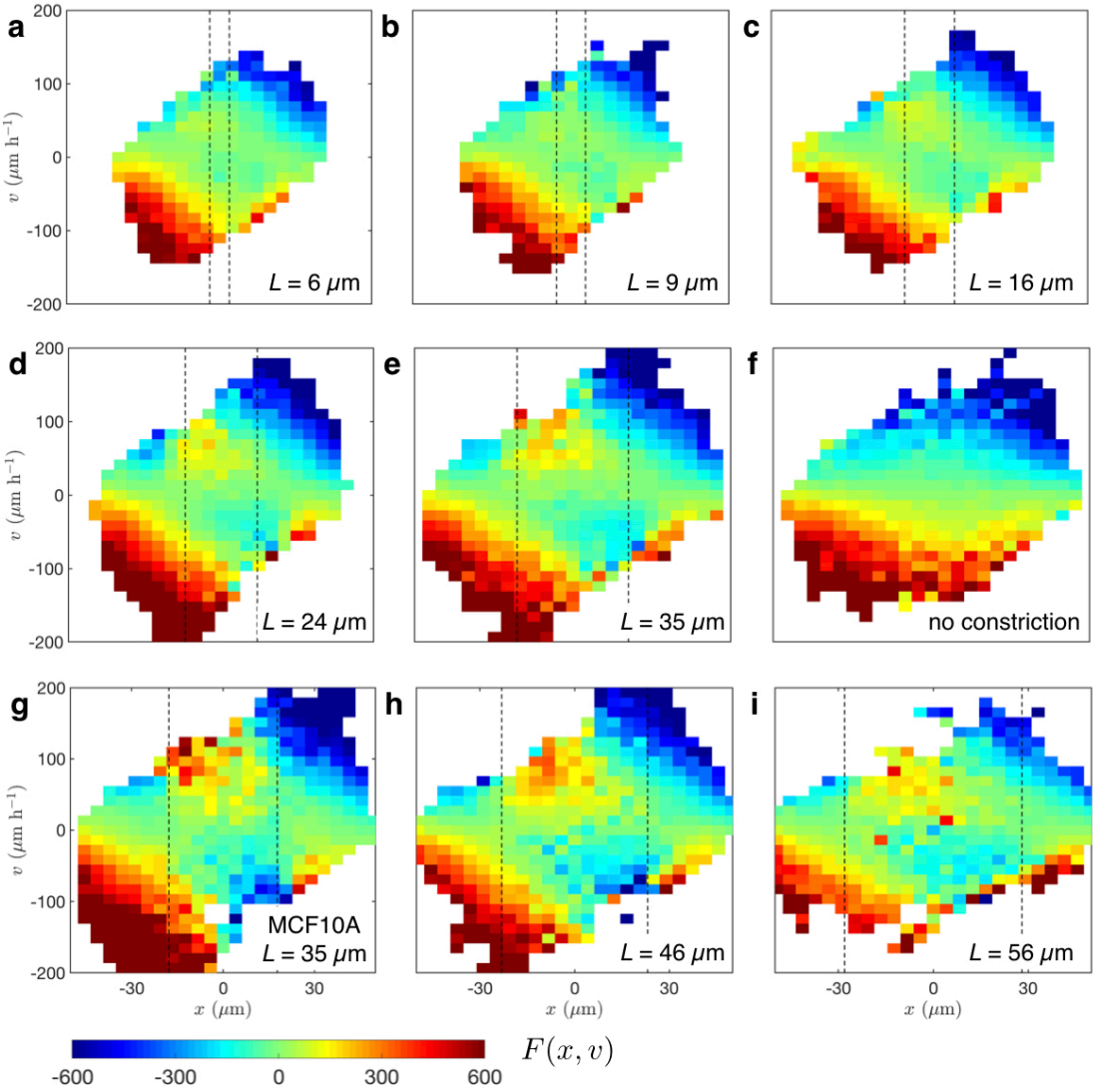


Figure S28 | Deterministic contributions $F(x, v)$ in units of μm/h². In the main text, a linearly interpolated version is shown for presentation purposes. Here we show the raw data as obtained from the inference procedure. Dashed black lines correspond to the positions $x = \pm L/2$.

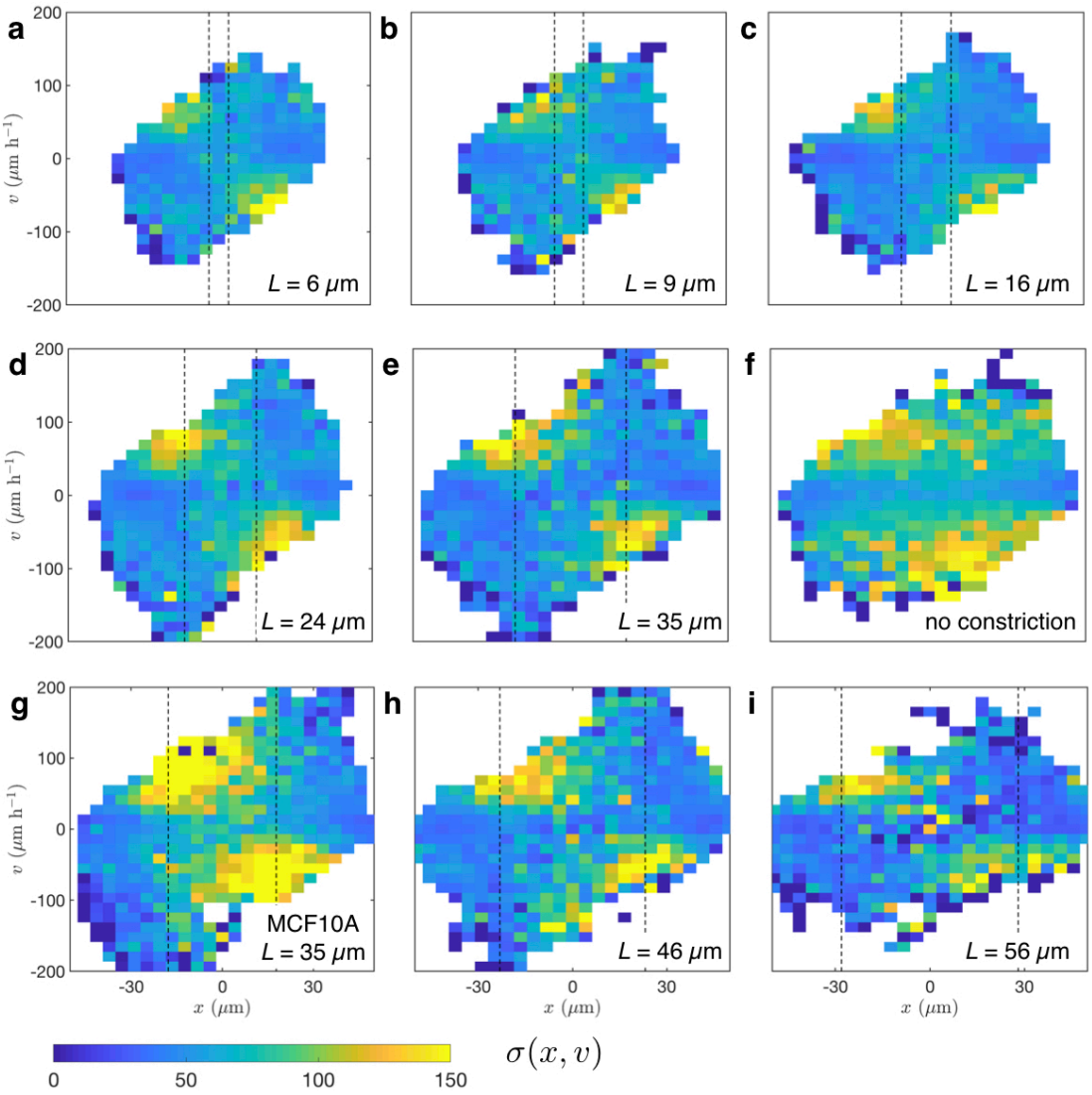


Figure S29 | The noise strength $\sigma(x, v)$ in units of μm/h$^{3/2}$. In the main text, a linearly interpolated version is shown for presentation purposes. Here we show the raw data as obtained from the inference procedure. Dashed black lines correspond to the positions $x = \pm L/2$.

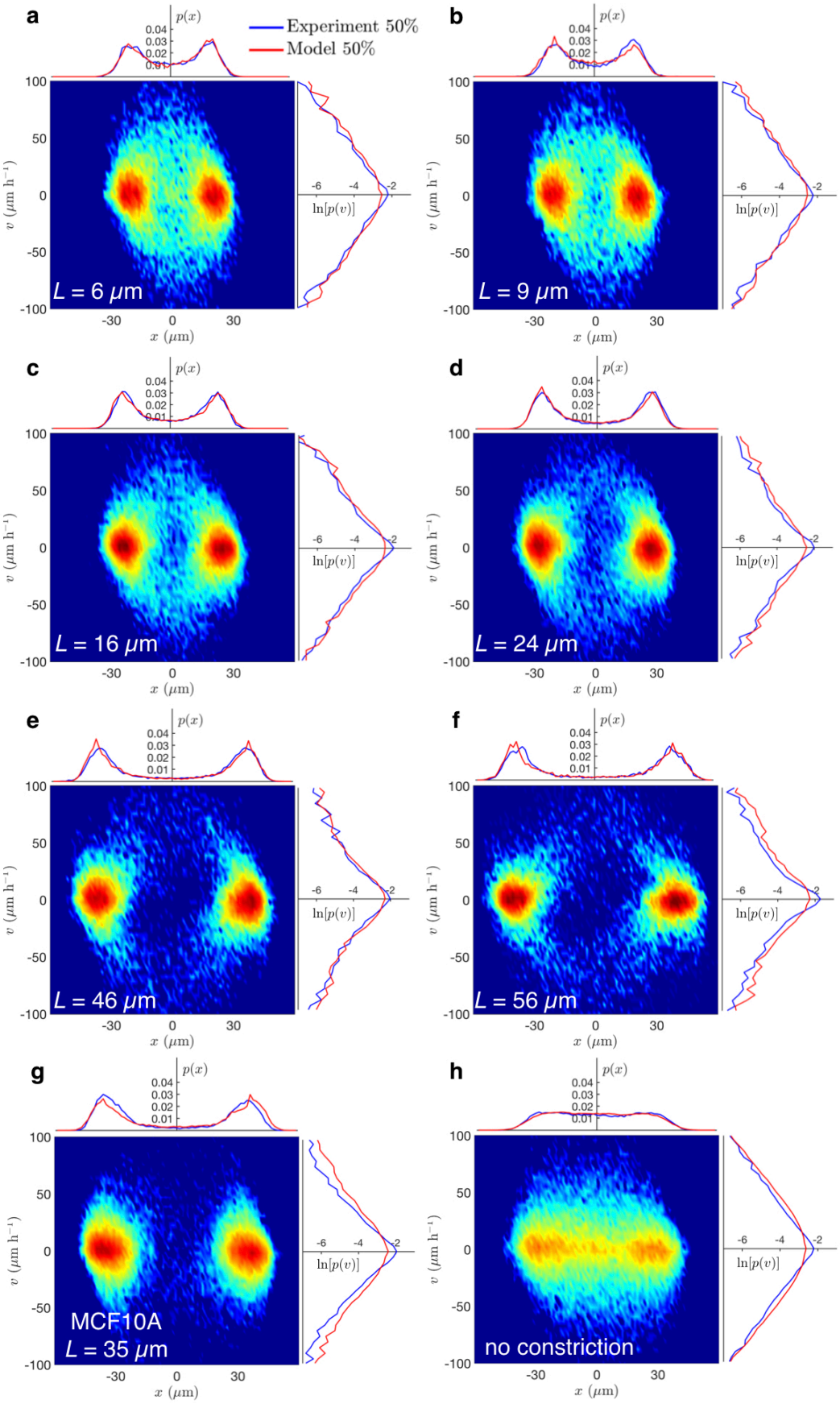

a
b
c
d
e
f
g
h
Experiment 50%
Model 50%
p(x)
v (μm h⁻¹)
x (μm)
ln[p(v)]
L = 6 μm
L = 9 μm
L = 16 μm
L = 24 μm
L = 46 μm
L = 56 μm
MCF10A
L = 35 μm
no constriction

Figure S30 | Probability distributions of position and velocity. Shown here for each of the six bridge lengths not shown in the main text, the MCF10A data, and the system without constriction. Each subfigure shows the the joint probability distribution of position and velocity $p(x, v)$ (colorplot, plotted logarithmically). The line plots display the marginal probability distributions of position $p(x)$ and velocity $p(v)$ (semi-logarithmically) given by the experiment (blue) and the model (red).

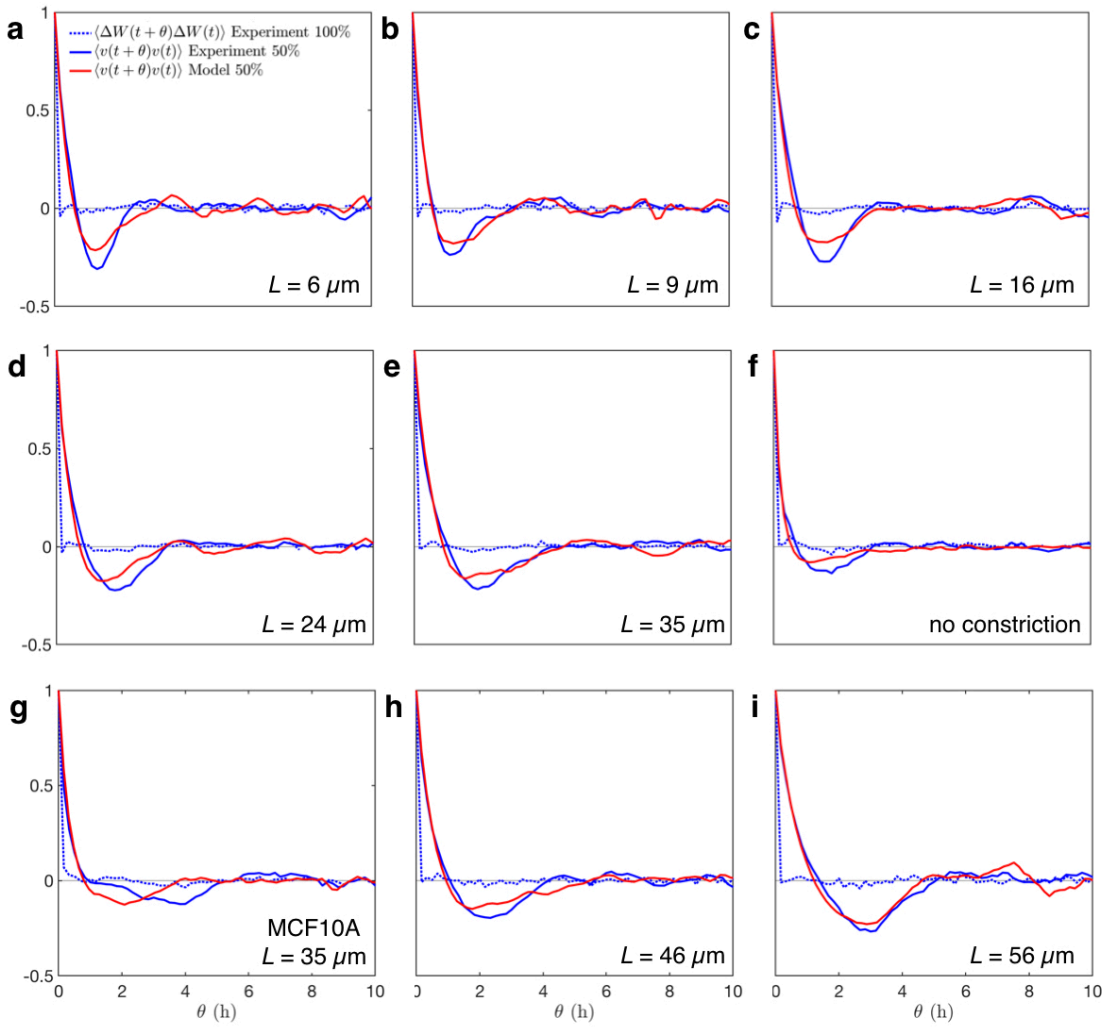


Figure S31 | The correlation functions of velocity and inferred noise. The correlation function of the inferred noise $\Delta W(t) = \int_t^{t+\Delta t} \eta(t)\,\mathrm{d}t$ is shown by the dotted blue line. The velocity correlation functions from experiment (solid blue line), and the associated model prediction (red line) are each based on one of two halves of the entire data set. $\Delta t = 10$ min is the experimental sampling interval.

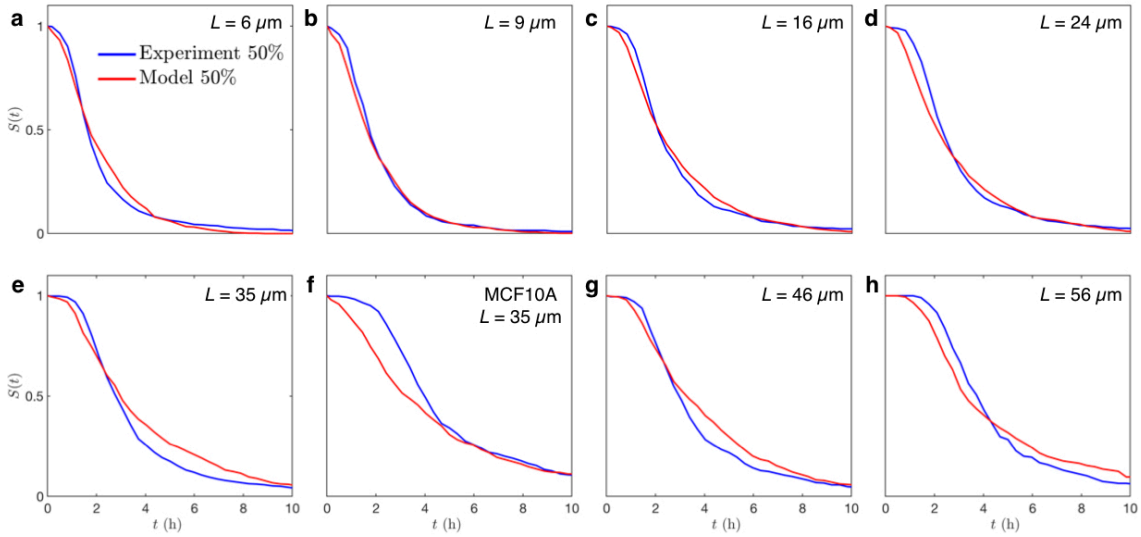


Figure S32 | **Survival probability distributions.** Comparison of the experimental observation (blue) and the model prediction (red) for all two-state systems. Experimental data and model prediction are based on mutually exclusive halves of the recorded data, as explained in Supplementary Section S3.

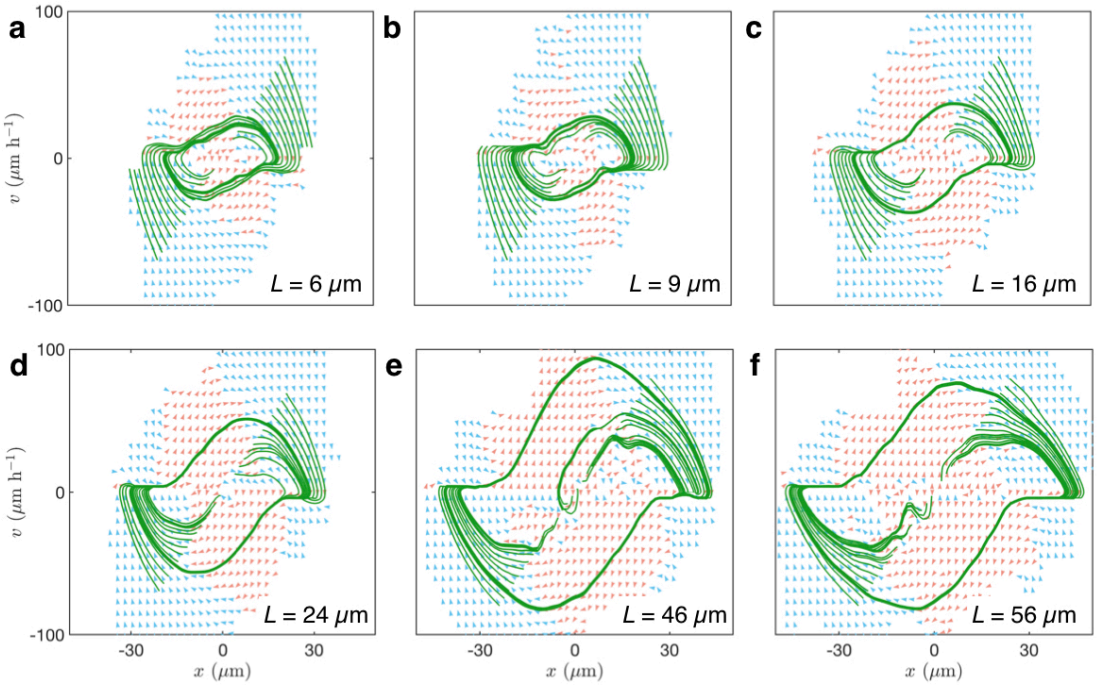


Figure S33 | **Trajectories and flow fields of the deterministic term.** Shown for the bridge lengths not shown in the main text. Arrows indicate the local direction of flow, where accelation is orange and deceleration is blue, and green lines correspond to trajectories of the deterministic dynamics.

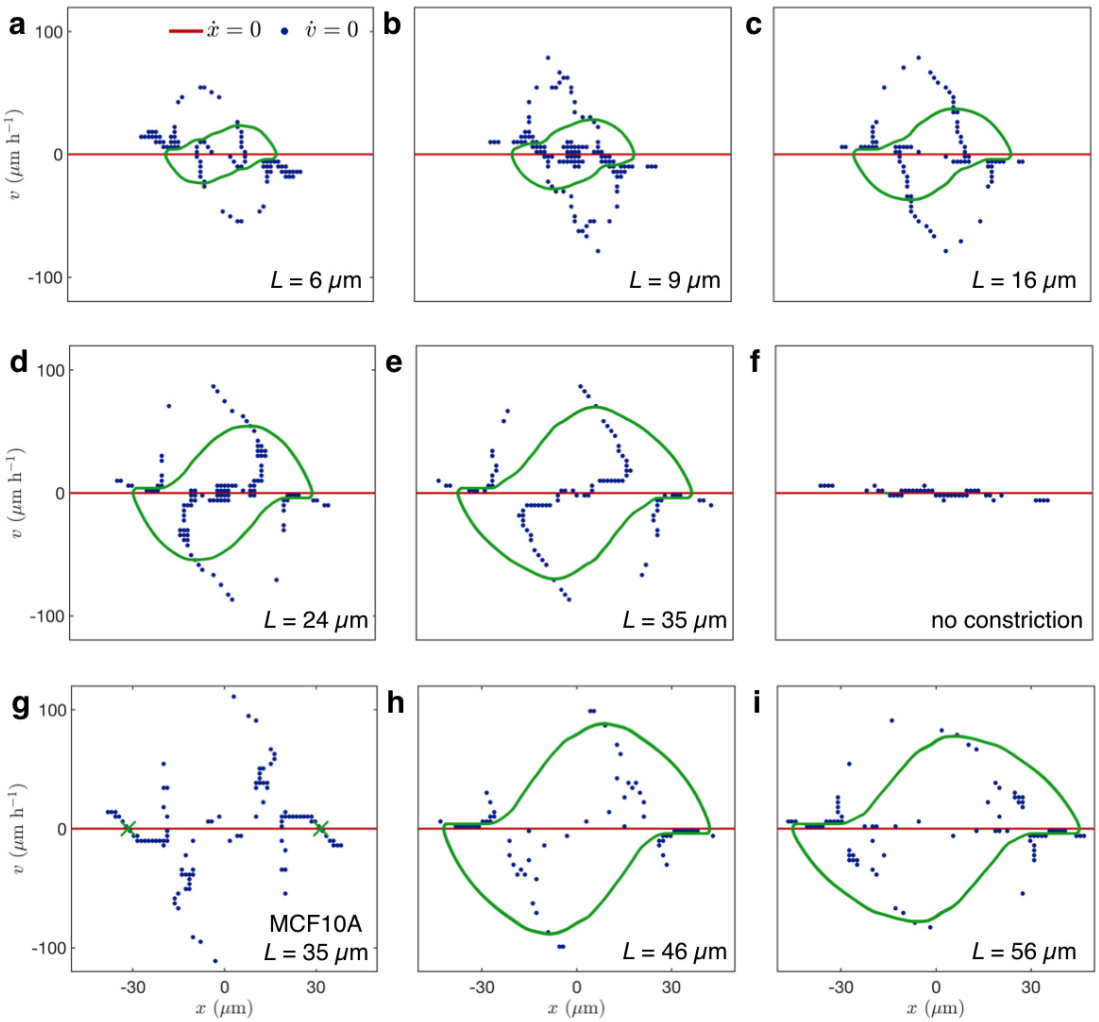


Figure S34 | Nullclines of the deterministic dynamics. The nullclines $\dot{x} = 0$ and $\dot{v} = 0$ are shown by the red line and blue dots respectively. They are obtained from the symmetrised version of the deterministic term (see Section S3.4). For MDA-MB-231 (**a**-**e**, **h**, **i**), the green lines indicate the stable limit cycle attractor, while green crosses show the locations of the stable fixed points for MCF10A (**g**). The result for L = 35 µm is shown in Fig. 4e in the main text.

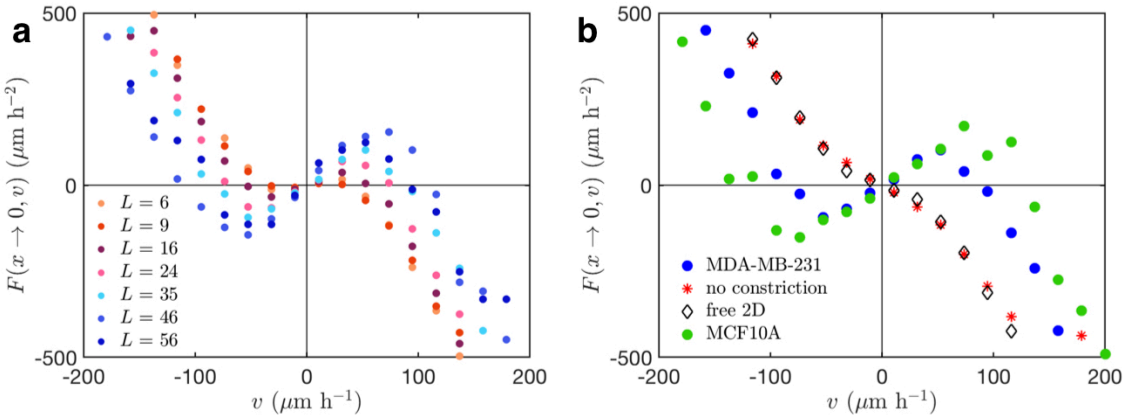


Figure S35 | The frictional component $F(x \to 0, v)$. In the main text, the frictional component of the symmetrised dynamics is shown. Here we shown the non-symmetrised versions, which show the same qualitative features. **a**, Here shown for all seven bridge lengths used for MDA-MB-231 cells. **b**, Equivalent to Fig. 4d, but with non-symmetrised data.

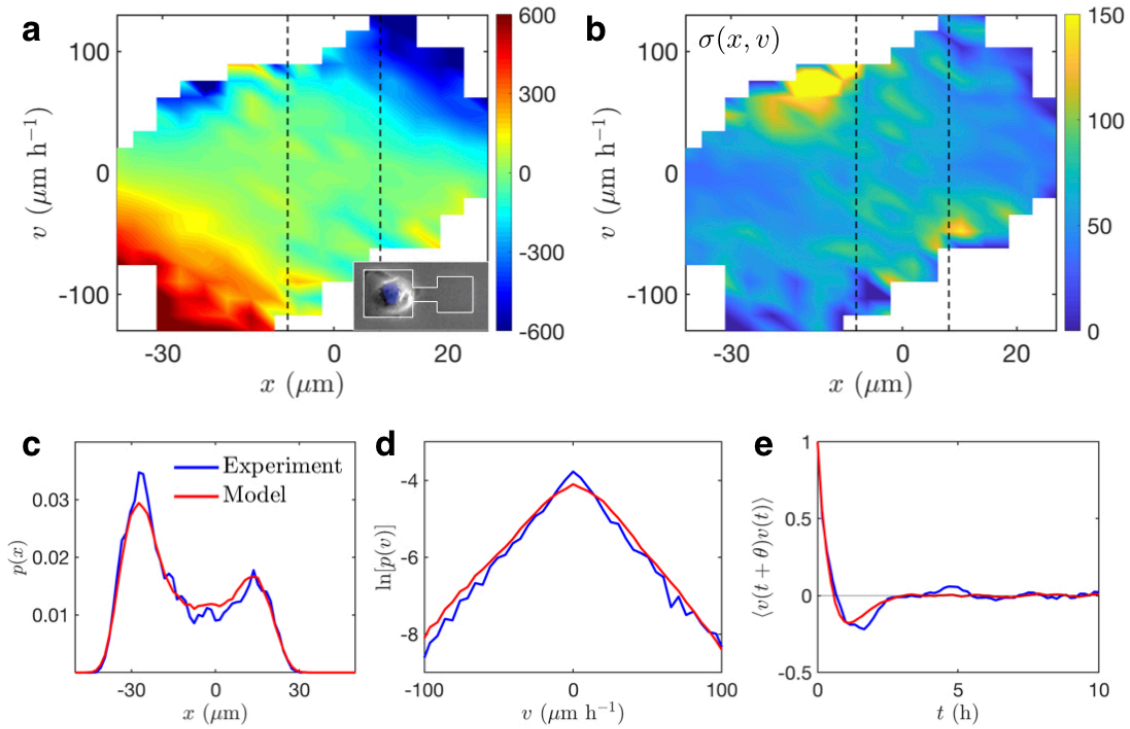


Figure S36 | Cell dynamics in an asymmetric two-state system with square edge lengths 42 µm and 27 µm (MDA-MB-231). **a**, The deterministic component $F(x, v)$ in units of µm/h^2. **b**, The noise strength $\sigma(x, v)$ in units of µm/h$^{3/2}$. Dashed black lines correspond to the positions $x = \pm L/2$. Both F and σ are shown here as a linear interpolation. **c**, **d**, Experimental (blue) and predicted (red) probability distributions of position (**c**) and velocity (**d**, plotted semi-logarithmically). **e**, Velocity correlation function.

References

1. http://site.physics.georgetown.edu/matlab/index.html.
2. Selmeczi, D., Mosler, S., Hagedorn, P. H., Larsen, N. B. & Flyvbjerg, H. Cell motility as persistent random motion: theories from experiments. *Biophys. J.* **89,** 912–931 (2005).
3. Flyvbjerg, H. & Petersen, H. G. Error estimates on averages of correlated data. *J. Chem. Phys.* **91,** 461–466 (1989).
4. Efron, B. & Tibshirani, R. J. *An Introduction to the Bootstrap*. (CRC Press, 1994).
5. Doyle, A. D., Wang, F. W., Matsumoto, K. & Yamada, K. M. One-dimensional topography underlies three-dimensional fi brillar cell migration. *J. Cell Biol.* **184,** 481–490 (2009).
6. Siegert, S., Friedrich, R. & Peinke, J. Analysis of data sets of stochastic systems. *Phys. Lett. A* **243,** 275–280 (1998).
7. Ragwitz, M. & Kantz, H. Indispensable Finite Time Corrections for Fokker-Planck Equations from Time Series Data. *Phys. Rev. Lett.* **87,** 254501 (2001).
8. Friedrich, R., Peinke, J., Sahimi, M. & Reza Rahimi Tabar, M. Approaching complexity by stochastic methods: From biological systems to turbulence. *Phys. Rep.* **506,** 87–162 (2011).
9. Pedersen, J. N. *et al.* How to connect time-lapse recorded trajectories of motile microorganisms with dynamical models in continuous time. *Phys. Rev. E* **94,** 062401 (2016).
10. Gardiner, C. *Handbook of Stochastic Methods*. (Springer, 1985).
11. Stephens, G. J., Johnson-Kerner, B., Bialek, W. & Ryu, W. S. Dimensionality and Dynamics in the Behavior of C. elegans. *PLoS Comput Biol* **4,** e1000028 (2008).
12. Stephens, G. J., Bueno, M., Mesquita, D., Ryu, W. S. & Bialek, W. Emergence of long timescales and stereotyped behaviors in Caenorhabditis elegans. *Proc. Natl. Acad. Sci. USA* **108,** 7286–7289 (2011).
13. Takagi, H., Sato, M. J., Yanagida, T. & Ueda, M. Functional analysis of spontaneous cell movement under different physiological conditions. *PLoS One* **3,** e2648 (2008).
14. Albert, P. J. & Schwarz, U. S. Dynamics of cell shape and forces on micropatterned substrates predicted by a cellular Potts model. *Biophys. J.* **106,** 2340–2352 (2014).
15. Albert, P. J. & Schwarz, U. S. Modeling cell shape and dynamics on micropatterns. *Cell Adhes. Migr.* **10,** 516–528 (2016).
16. Kramers, H. A. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **7,** 284–304 (1940).
17. Abkenar, M., Gray, T. H. & Zaccone, A. Dissociation rates from single-molecule pulling experiments under large thermal fluctuations or large applied force. *Phys. Rev. E* **95,** 042413 (2017).
18. van der Pol, B. On relaxation-oscillations. *Philos. Mag.* **2,** 978–992 (1926).

Chapter 4

Quantifying the variability of cell migration behaviours

This chapter is based on the following publication:

Disentangling the Behavioural Variability of Confined Cell Migration

David B. Brückner, Alexandra Fink, Joachim O. Rädler, Chase P. Broedersz†

† corresponding author

Journal of the Royal Society Interface 17, 20190689 (2020)

Disentangling the Behavioural Variability of Confined Cell Migration

David B. Brückner[1], Alexandra Fink[2], Joachim O. Rädler[2], Chase P. Broedersz[1+]

[1] Arnold-Sommerfeld-Center for Theoretical Physics and Center for NanoScience, Ludwig-Maximilians-Universität, München
[2] Faculty of Physics and Center for NanoScience, Ludwig-Maximilians-Universität, München

[+] corresponding author (c.broedersz@lmu.de)

Cell-to-cell variability is inherent to numerous biological processes, including cell migration. Quantifying and characterizing the variability of migrating cells is challenging, as it requires monitoring many cells for long time windows under identical conditions. Here, we observe the migration of single human breast cancer cells (MDA-MB-231) in confining two-state micropatterns. To describe the stochastic dynamics of this confined migration, we employ a dynamical systems approach. We identify statistics to measure the behavioural variance of the migration, which significantly exceed those predicted by a population-averaged stochastic model. This additional variance can be explained by the combination of an 'aging' process and population heterogeneity. To quantify population heterogeneity, we decompose the cells into subpopulations of slow and fast cells, revealing the presence of distinct classes of dynamical systems describing the migration, ranging from bistable to limit cycle behaviour. Our findings highlight the breadth of migration behaviours present in cell populations.

On all scales of life, the behaviour of single organisms is intrinsically variable, both as a function of time and between individuals. A number of studies have quantified the inter-individual variations in behaviour in a broad range of systems, from bacteria [1–4] to protozoans [5,6], fruit flies [7], mice [8], and humans [9,10]. However, even in populations of single mammalian cells with identical genomes, the intrinsic stochasticity of intra-cellular processes such as gene expression, cytoskeletal rearrangement and protein localization can lead to large differences in the proteomes of individual cells [11–15]. Further downstream, this diversity leads to cell-to-cell variability (CCV), i.e. phenotypic or population heterogeneity, at the level of whole-cell behaviours, such as growth rate, drug response, morphology, and migration [16–20]. In the case of migrating cells, CCV has profound implications at larger scales for the behaviour of cell clusters, such as their ability to chemotax [21], invade surrounding tissue [22] and perform collective motion [23,24]. Thus, CCV is increasingly well understood at the molecular level, and its implications for collective

behaviour are becoming clearer. However, the detection and quantitative characterization of this variability in ensembles of individual migrating cells remains challenging; such a characterization requires both an appropriate theoretical framework and data sets where migrating cells are monitored over a sufficiently long time under identical conditions.

Cell migration is a key feature of many cell types, including immune, epithelial and cancer cells. The movement of such cells is powered by an intricate machinery that relies on the coordination of vast numbers of molecular constituents, whose collective dynamics are key in generating the persistent motion of cells [25,26]. Despite this underlying complexity, a number of studies have shown that the emergent migratory dynamics at the cellular scale can be quantitatively captured by relatively simple stochastic equations of motion [6,27–34]. This approach has been successfully applied to migration on uniform, two-dimensional surfaces [6,27–30,35], and recently to cells migrating in confinement [32,33]. However, to determine the structure and parameters of such stochastic models, the dynamics are typically averaged across different cells and over time, yielding ensemble- and time-averaged (ETA) models that describe the average member of a cell population, and thus fail to capture phenotypic heterogeneity [15,36]. Similarly, bottom-up models for cell motility typically assume that all cells in a population can be described by a common set of parameters [37–45].

To characterize heterogeneity in cell migration, a super-statistical framework was previously applied to cell migration [46]. By fitting a persistent random motion model with time-dependent parameters to migration in a variety of environments, large variations in cell behaviour were revealed both in time and between individuals. Another approach has been to construct a phenotypic space through dimensional reduction, which identifies the aspects of behaviour that vary most between individuals [5]. However, in both approaches, the presence of time- and population-heterogeneity is assumed from the start. It thus remains unclear how much of the observed variability is due to real heterogeneity, and how much is due to the intrinsic stochasticity of the migration process. A central difficulty in detecting phenotypic heterogeneity in cell migration is the short time window of a trajectory of an individual cell, which is limited by cell division. Thus, even for a hypothetical population of identical cells whose migration can be described by a single equation of motion, one would expect to observe statistical differences between individual trajectories due to the limited observation time of the experiment. A key challenge is therefore to distinguish such apparent variability in short trajectories from real inter-individual variability. To detect and characterize CCV in behaviour, it would therefore be helpful to have a model system for which a quantitative theory for the ensemble-averaged behaviour of the cells has been developed. Such an ETA model can then provide a reference to identify observables that are sensitive to potential phenotypic heterogeneity in the population.

In previous work [32,33], we analyzed the time-trajectories of single breast cancer cells (MDA-MB-231) confined to two-state micropatterns, in which these cells repeatedly transit across a thin constriction. This system mimics the extra-cellular environment cells face in physiological settings, in which they navigate confining structures and squeeze through thin channels and constrictions [47–59]. An important advantage of this setup is that it allows us to quantitatively monitor the behaviour of a large population of cells in identical, standardized surroundings, suppressing extrinsic sources of heterogeneity in the trajectories. The overall dynamics of this two-state migration are well described by a stochastic equation of cell motion, with a deterministic and stochastic contribution that can be inferred from the short time-scale data observed experimentally. This model accurately reproduces the ensemble- and time-averaged statistics including the dwell time distribution, the distributions of position and velocity, and the velocity autocorrelation function.

Here, we develop a theoretical approach to address the behavioural variability of confined cell migration arising through phenotypic heterogeneity and time-dependent dynamics. To this end, we identify a statistic that reveals the intrinsic variability of the migration in two-state micropatterns: the variance in the number of transitions a single cell performs in a given time. The sensitivity of this statistic to variability arises because it measures single-cell behaviour accumulated over long periods of time, beyond the time-scale of a single transition, and compares this to the population average. Indeed, we find that this variance significantly supersedes the prediction of the ensemble-averaged model. We account for this observation by two effects: a time-dependent 'aging' effect and phenotypic heterogeneity. From the trajectory data, we show that the average acceleration of the cells gradually decreases over time, leading to a decrease in hopping activity. Furthermore, we propose to capture phenotypic heterogeneity by inferring migration behaviour for subpopulations of cells. Specifically, we find that by splitting the cell population by transition time into only two subpopulations, we can capture most of the inter-cell variance. These subpopulations are described by distinct classes of dynamical systems, with fast cells performing limit cycle oscillations while slow cells exhibit excitable bistability. Thus, our approach reveals a striking breadth of migratory behaviours within a cell population, paving the way to investigate the origins and functional implications of such variability.

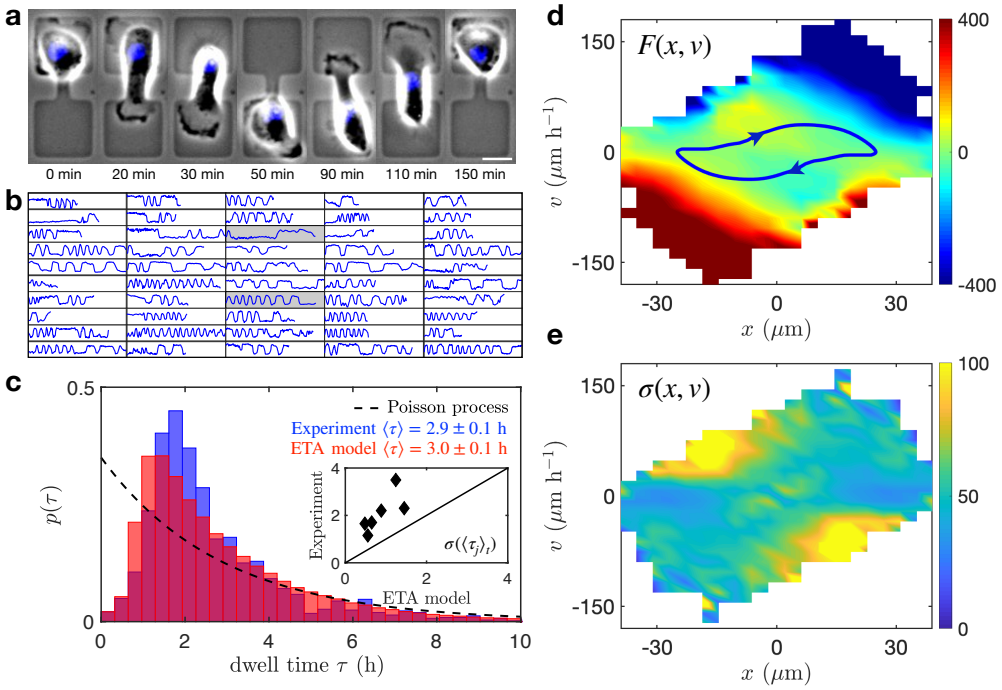


Fig 1. Equation of motion for an ensemble of cells migrating on two-state micropatterns (L = 16 µm). (a) Time series of a single cell migrating on a two-state micropattern with bridge length L = 16 µm (scale bar: 20 µm). (b) Trajectories of the cell nucleus as a function of time, plotted from t = 0 to t = 50 h, for a collection of cells. The two trajectories shaded in grey correspond to supplementary videos S1 and S2, providing examples of a 'slow' and 'fast' cell, respectively. (c) Distribution of dwell times observed experimentally (blue) and predicted by an ensemble-averaged equation of motion (red). The dashed line indicates the probability distribution of a Poisson process with the same mean dwell time, defined by $p(\tau) = ve^{-v\tau}$, with $v = 1/\langle\tau\rangle_{t,e}$. Inset: inter-cellular variance $\sigma(\langle\tau(j)\rangle_t)$ observed in experiment, plotted against the prediction by the ETA model. (d) The deterministic term $F(x,v)$ of the ETA model in units of µm/h^2. (e) The ETA model noise strength $\sigma(x,v)$ in units of µm/h$^{3/2}$.

Cells perform non-Poissonian transitions on two-state micropatterns

To develop a quantitative framework for the variability of migrating cells, we employ two-state micropatterns, in which cells are confined to a standardized geometry consisting of two square adhesive sites connected by a bridge of length L (Fig. 1a). In this environment, we previously found that MDA-MB-231 breast cancer cells perform repeated transitions between the adhesive sites [32]. This setup provides a minimal system to study how cells respond to the presence of thin constrictions and ensures that all cells in the population interact with an identical micro-environment.

Interestingly, the trajectories of the cells, which we define by the position of the nucleus center, exhibit large variations both in time and across individuals in the population (Fig. 1b; supplementary movies S1 and S2). Accordingly, the probability distribution of dwell times

exhibits a large spread (Fig. 1c). This raises a key question: Can the experimentally observed variation be explained by a single stationary stochastic process with time-independent parameters identical for all individuals?

The two-state migration assay provides an ideal platform to address this question. Previously, we have identified a stochastic equation of cell motion that captures the ensemble- and time-averaged behaviour of cells on this micropattern [32]. This equation takes the form

$$\frac{dv}{dt} = F(x,v) + \sigma(x,v)\eta(t) \tag{1}$$

where $\eta(t)$ represents Gaussian white noise with $\langle\eta(t)\rangle = 0$ and $\langle\eta(t)\eta(t')\rangle = \delta(t-t')$. Here, the deterministic term $F(x,v) = \langle\dot{v}|x,v\rangle_{t,\mathrm{e}}$ (Fig. 1d) and the noise strength $\sigma^2(x,v) = \Delta t\langle[\dot{v} - F(x,v)]^2|x,v\rangle_{t,\mathrm{e}}$ (Fig. 1e) are inferred from the short time-scale cell dynamics measured experimentally [60–62], where $\langle\Theta\rangle_{t,\mathrm{e}}$ indicates time- and ensemble-averaging of the observable Θ (see Methods). The terms $F(x,v)$ and $\sigma(x,v)$ thus show the average acceleration of the cell as a function of its position in xv-phase space, and the magnitude of stochastic fluctuations, respectively. In the case of the MDA-MB-231 cell line considered here, the deterministic term exhibits non-linear dynamics in the form of a limit cycle, indicating that the cellular transitions can be interpreted as noisy non-linear oscillations (Fig. 1d). We have shown that this model accurately captures the key statistics of the cell migration, including the dwell time distribution (Fig. 1c), as well as the distributions of position and velocity and the velocity correlation function [32].

To probe the limits of the ETA approach, we first test whether it also quantitatively predicts the dwell time distributions at the single-cell level. To do so, we simulate the model with the same observation time per cell as recorded experimentally. To measure cell-to-cell variability, we obtain the average dwell time $\langle\tau(j)\rangle_t$ for each single cell j and then determine the standard deviation $\sigma(\langle\tau(j)\rangle_t)$ of these single-cell averages, giving a measure of the inter-cell variance. Clearly, in the limit of infinitely long trajectories, the inter-cell variance in the ensemble-averaged model should vanish; therefore, the finite variance we find here for simulated trajectories is simply due to finite trajectory length. Interestingly, however, we find that the experimental inter-cell variance significantly exceeds that predicted by the ETA model for this trajectory length (inset Fig. 1c). This suggests that there are significant variations from cell to cell, which cannot be accounted for by the ETA model. We therefore hypothesize that the cell population exhibits phenotypic heterogeneity detectable at the level of cell trajectories.

The presence of phenotypic heterogeneity could lead to apparent temporal correlations in behaviour that persist for the whole life of a cell, up to division. To further quantify the overall variability in the transition behaviour, we thus need to consider statistics that could be sensitive to correlations over a range of time-scales, spanning from a single transition time up to the division time. A common way to quantify the overall variability of transition processes is to determine the ratio between mean and variance in the number of transitions within a given time interval. Thus, we measure the distribution function of the hopping statistics, which is given by the probability distribution $p(N,T)$ to observe N transitions of a cell in a given time interval T (Fig. 2a). Importantly, to determine the statistics of the observable N, we monitor the behaviour of a single cell over a wide range of time-scales, extending well beyond the average transition time. We therefore expect this statistic to be particularly sensitive to behavioural heterogeneity in a population, where some cells might have the tendency to consistently perform slower- or faster-than-average transitions.

As the observation time interval is increased, both the mean $\mu(T)$ and variance $\Sigma^2(T)$ of these distributions increase (inset Fig. 2a). The simplest 'hopping' process with such a distribution is the Poisson process, for which $\mu = \Sigma^2$. Interestingly, this is not the case here (Fig. 2b). Instead, at short time-scales, the transitions are sub-Poissonian with a mean that exceeds the variance ($\mu > \Sigma^2$), while beyond a cross-over time-scale T^*, the variance exceeds the mean, $\mu < \Sigma^2$, indicating a super-Poissonian regime. In fact, as we vary the length L of the connecting bridge, this behaviour is qualitatively reproduced in all cases (Fig. 2b). The cross-over time-scale T^* scales linearly with the average dwell time $\langle\tau\rangle_{t,\mathrm{e}}$, and typically exceeds it by a factor of approximately 4 (inset Fig. 2b).

These results show that the ratio between mean and variance in this transition process exhibits a subtle dependence on the time-scale on which the system is observed. At short time-scales, up to a few times the average dwell time, we observe a smaller variance than in a Poisson process, which is defined by a transition probability that is constant in time. Similar sub-Poissonian behaviour is found, for instance, in the stepping behaviour of molecular motors, where it was rationalized as a consequence of an underlying multi-step scheme that has to be executed before a large step can be made [63]. Analogously, here we might argue that during the cellular transitions in the two-state confinement, a number of cytoskeletal and shape rearrangements have to occur between transitions, which leads to fewer short dwell times than would be expected from a Poisson process (Fig. 1c). On long time-scales, however, the transition variance starts to significantly exceed the mean. Such super-Poissonian statistics have previously been observed, e.g. in RNA-transcription [64], and were attributed to cell-to-cell variations in the transcription rates [65].

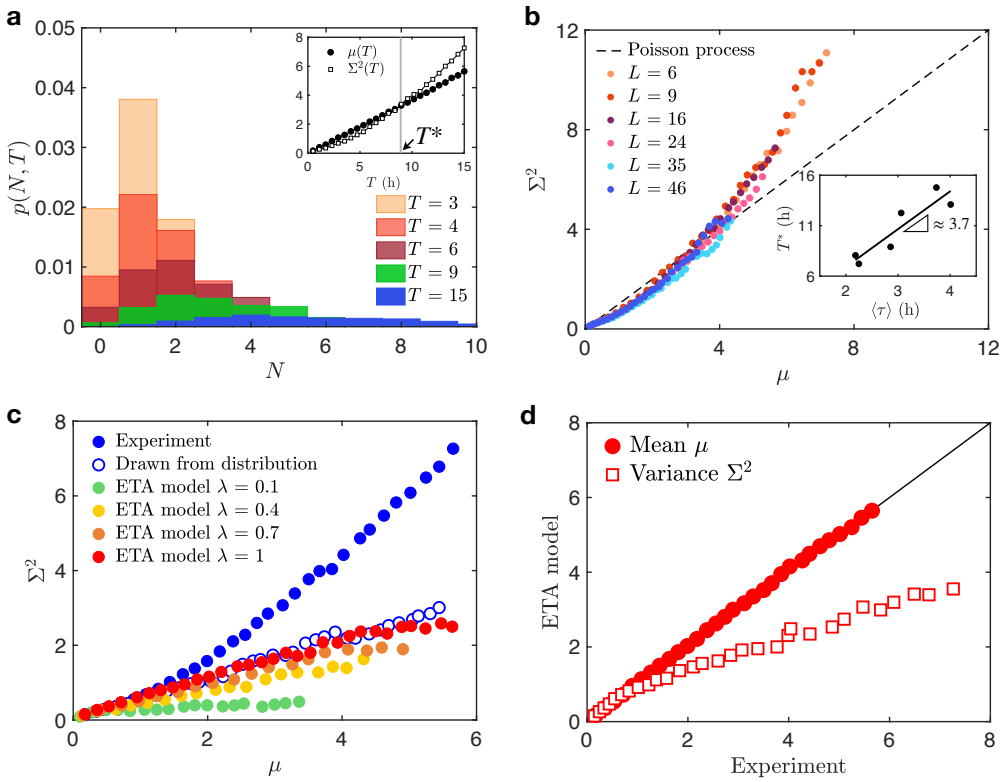


Fig 2. Transition number statistics of hopping cells. (a) Probability distributions of the number of transitions N of a single cell in a given time interval T, indicated by the colours. Inset: Variance Σ^2 and mean μ of the number of transitions as a function of the time interval T. The grey vertical bar indicates the cross-over time-scale T^*. (b) Variance plotted against the mean for all bridge lengths. The dashed line corresponds to the Poisson behaviour $\mu = \Sigma^2$. Inset: Cross-over time-scale T^* plotted as a function of $\langle\tau\rangle_{t,e}$. Black line corresponds to a linear fit. (c) Mean-variance curve observed experimentally (blue) compared to the prediction by the stochastic model (red) and the expectation from random sampling of the experimental dwell time distribution (open symbols). In green, orange and yellow, we show predictions of the model with a noise amplitude scaled by factors $\lambda = 0.1, 0.4, 0.7$. (d) Mean and variance predicted by the model plotted against experimental values. Panels (a), (c), (d) correspond to bridge length L = 16 µm.

Observed variability exceeds that of a single stochastic process

The observation of non-Poissonian statistics does not necessarily indicate cell-to-cell variability. To investigate the origin of the non-Poissonian statistics in the hopping behaviour of cells, we test our inferred ETA model (Eq. (1)) by comparing the predicted mean-variance curve to that measured experimentally. Strikingly, the ETA model predicts a qualitatively different, purely sub-Poissonian trend (Fig. 2c). While this model accurately predicts the average dwell time $\langle\tau\rangle_{t,e}$, and thus also the mean of the hopping distribution μ, it underestimates the variance (Fig. 2d).

The behaviour of the statistics predicted by the ETA model can be understood intuitively: in a system with no noise ($\sigma(x,v) = 0$), we observe regular limit cycle oscillations, with zero variance. To investigate how different noise levels theoretically affect the hopping variance, we introduce an artificial prefactor λ to scale the noise: $\dot{v} = F(x,v) + \lambda\sigma(x,v)\eta(t)$. As λ is increased from 0 to 1, we find that the variance gradually increases (Fig. 2c). At the experimentally inferred noise level $\lambda = 1$, the system still appears to be dominated by the underlying deterministic oscillation period, giving rise to a sub-Poissonian process at all time-scales. Thus, while the ETA model predicts many statistics of the motion correctly, including the short time-scale variability, it falls short of capturing variability in the hopping behaviour on longer time-scales.

Next, we investigate the origin of the super-Poissonian regime directly from experimental data. If we assume that all cells follow the same equation of motion with identical parameters, and that these parameters do not vary in time, then each cell can be described by the same time-independent dwell time distribution $p(\tau)$. In other words, the hopping distribution $p(N,T)$ should be fully constrained by $p(\tau)$ and the stationary correlations between hops [63]. Interestingly, if we generate a trajectory of 'hops' by random independent sampling from the population dwell time distribution, we recover the same sub-Poissonian mean-variance relation as predicted by the inferred ETA model (Fig. 2c). This indicates that the transition time of subsequent hops in the ETA model have negligible correlations, while the experimental data may exhibit correlations extending over multiple hops.

By time- and ensemble-averaging, we miss this aspect that influences the long time-scale variability of the motion. In the next sections, we therefore carefully assess the validity of our assumptions of time-invariance and population homogeneity.

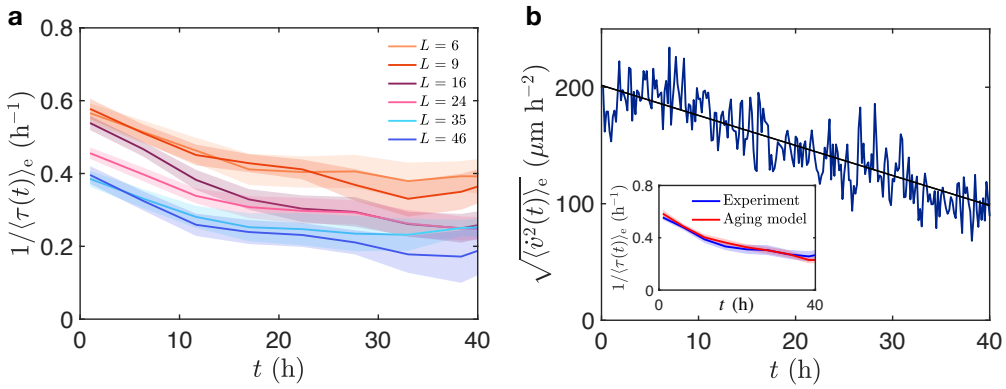


Fig 3. Testing time-invariance. (a) Hopping rates $1/\langle\tau(t)\rangle_e$ as a function of time for all bridge lengths. Here, for a stay from t_1 to t_2, with $\tau = t_2 - t_1$, we define the time at which the stay occurred as $t_{stay} = (t_1 + t_2)/2$. The quantity $\langle\tau(t)\rangle_e$ is then averaged over the cell population and in time windows of width 5 h (Supplementary Section S5). Shaded intervals correspond to standard error in the mean. Note that we cannot extrapolate this trend beyond $t \approx 40$h, as this is the time-scale of cell division. (b) Root-mean-square acceleration of the cells as a function of time, averaged across the ensemble in each time frame (L = 16 µm). Black line corresponds to a linear fit. Inset: Hopping rates as a function of time predicted by the aging model compared to the experimental observation (L = 16 µm).

Attenuation of cell dynamics enhances transition variability

To investigate if the cellular hopping process indeed violates time-invariance, we measure the hopping rates $1/\langle\tau(t)\rangle_e$ as a function of time. The rate slowly decreases with time, indicating that the cell migration is slowing down over time (Fig. 3a). Similarly, the root-mean-square (r.m.s.) acceleration decreases roughly linearly as a function of time, by almost a factor of two (Fig. 3b). Note, however, that the cell population in this experiment is not cell-cycle-synchronized, such that t = 0 does not necessarily correspond to the same moment in the cell cycle for all cells. The decrease in acceleration we identify thus simply indicates the approximate trend observed in this experiment, giving the average behaviour of an aging population.

Clearly, the average acceleration is changing significantly over time. To incorporate this into the model, we postulate that most of the time-dependence can be accounted for by introducing an overall time-dependent pre-factor into the equation of motion, which we term the aging model:

$$\frac{dv}{dt} = \alpha(t)[F(x,v) + \sigma(x,v)\eta(t)] \qquad (2)$$

where $\alpha(t) = \sqrt{\langle\dot{v}^2(t)\rangle_e}/\sqrt{\langle\dot{v}^2\rangle_{t,e}}$. Here, $\sqrt{\langle\dot{v}^2(t)\rangle_e}$ is the ensemble averaged r.m.s. acceleration as a function of time which is determined in every time frame, and is normalized

by $\sqrt{\langle \dot{v}^2 \rangle_{t,e}}$, the r.m.s. acceleration averaged over all time and cells. This implementation accurately reproduces the time-dependence of the hopping rates (inset Fig. 3b). Remarkably, this simple approach also significantly improves our estimate of the hopping variance Σ^2 (Fig. 4e). An alternative approach to incorporate time-dependence into our model would be to infer, for example, separate early- and late-stage models, corresponding to the first and second half of the experiment. We tested such an approach and found that the resulting predictions did not improve the estimate of the transition variability (Supplementary Section S1). However, even the more successful pre-factor aging approach underestimates the experimentally observed variance, suggesting that another effect needs to be accounted for.

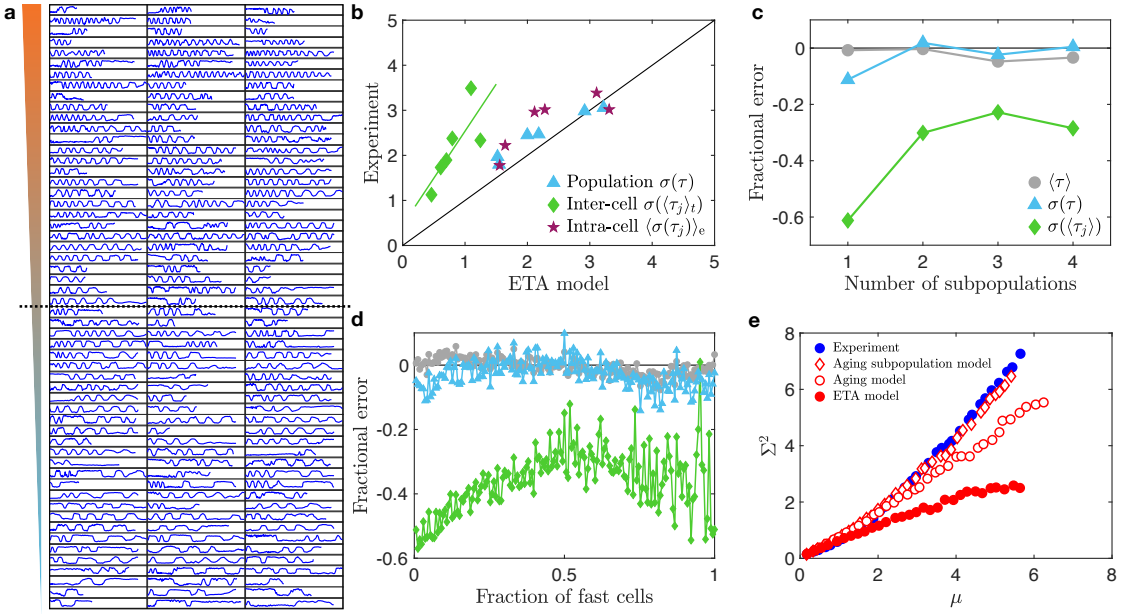


Fig 4. Detection and characterization of cell-to-cell variability. (a) All cell trajectories $x(t)$ plotted as a function of time, from t = 0 to t = 50 h. The cells are ranked by their average single-cell dwell time $\langle\tau(j)\rangle_t$, with the fastest cells at the top and the slowest at the bottom. The blacked dashed line indicates the 50:50 subpopulation split. Note, we exclude the first and last stay of each trajectory from the analysis throughout, as they are defined by the start and end of measurement, respectively, or by cell division, and are therefore not part of the hopping process. (b) Experimental observations plotted against ETA model predictions for various variance measures for all six bridge lengths: the population variance of all dwell times $\sigma(\tau)$, the inter-cell variance $\sigma(\langle\tau(j)\rangle_t)$, and the intra-cell variance $\langle\sigma(\tau(j))\rangle_e$. All variances are plotted as standard deviations. The green line indicates a linear fit to the inter-cell variance points, which has a slope of 2. (c) Fractional error in the prediction of various quantities as a function of the number of subpopulation models. In each case, the population is divided evenly into subpopulations of equal size. (d) Fractional error in various quantities as a function of the location of the subpopulation split, defined as the number of fast cells in the two-type subpopulation model. Symbols are defined in the legend of panel (c). (e) Mean-variance curves of the ETA model (Eq. (1); red dots), the aging model (Eq. (2); open dots) and the aging subpopulation model (open diamonds) in which two subpopulation models are combined with the time-dependent

prefactor. Experimental data is shown by blue dots. Panels (a), (c)-(e) correspond to bridge length L = 16 µm.

Population heterogeneity is captured by a decomposition into subpopulations

In the previous section, we showed that by time averaging, we miss an important contribution to the hopping variance caused by the gradual decrease in the acceleration of the cells. Similarly, in the inference of the ETA model, we have employed ensemble averaging, and have thus assumed that all cells follow the same equation of motion with identical parameters. However, individual cells of the same cell line may have variations in their genome [66], but even in an isogenic population of cells, we expect there to be phenotypic differences between cells [11,12,17,67]. Thus, we next consider the impact of phenotypic heterogeneity on our model.

The ETA model performs well in predicting the variance of the population distribution $\sigma(\tau)$ (Fig. 4b). Moreover, the ETA model gives a good estimate of the intra-cellular variance, defined as the average variance of dwell times within the life time of a single cell $\langle\sigma(\tau(j))\rangle_{\mathrm{e}}$. This suggests that the ETA model accurately captures the variance of behaviour within the life time of a single cell. We therefore do not consider switching between different 'modes' of movement [35,46,68]. However, we found that the inter-cellular variance $\sigma(\langle\tau(j)\rangle_t)$ is larger in experiment than predicted by the ETA model (inset Fig. 1c), indicating that the cell population may exhibit phenotypic heterogeneity.

In principle, a natural approach to tackle inter-cellular variability within the framework of our stochastic nonlinear model would be to infer single-cells models. For a population of n cells, we would then infer n models, and we could investigate how these would distribute over model space. However, due to the finite division time of the cells, reliably inferring the deterministic and stochastic components for a single cell is challenging. Instead, we therefore propose to rank the population of cells by their average dwell time $\langle\tau(j)\rangle_t$, and then split the population into subpopulations of cells with similar average hopping rates (Fig. 4a). Importantly, the ranking of the cells is uncorrelated with the trajectory length, suggesting that it is not determined by the cell cycle stage of the cell. We then infer separate stochastic nonlinear models for each of the subpopulations, generate trajectories for each model, and finally analyze a population of trajectories made up of these subpopulation-trajectories.

Interestingly, performing this procedure for only two subpopulations already significantly improves our estimate of the inter-cellular variance (reducing the relative error by more than a factor of 2), while not significantly affecting the prediction of other quantities such as the population mean and variance (Fig. 4c). However, further increasing the number of subpopulations does not appear to significantly improve predictions, suggesting that the

dominant contribution to this heterogeneity can be captured by two phenotypes. These results indicate that CCV in this system is well approximated by a decomposition into two subpopulations.

As an alternative approach to subpopulation models, we tested a model in which we measured an overall rescaling prefactor for each single cell, in a similar spirit to our approach to the 'aging' effect. However, this approach does not predict the correct population averaged dwell time $\langle\tau\rangle_{t,\mathrm{e}}$ and furthermore does not lead to an improvement in the estimate of the transition variance (Supplementary Section S3).

Given that a splitting into two subpopulations appears to be the simplest effective approach, we next asked whether there is an optimal way of doing so. To answer this question, we vary the fraction of cells included in the fast model. Interestingly, we find that the model performs optimally for a 50:50 split of the cell population (Fig. 4d). Thus, through the quantitative nature of our approach, we are able to identify the most natural way of decomposing the cell population into subpopulations, by optimizing the predictive power of the model. In general, one could imagine other scenarios than a predictive optimum at 50:50, for example, if there are two clearly distinct phenotypes in the population with an uneven distribution. In contrast, the actual variation of cellular identities in the experimental data appears to be continuous from very slow to very fast cells (Fig. 4a). This is further supported by the observation that the distributions of single cell observables, such as dwell times and speeds, exhibit unimodal distributions, with no clear peak structure that would indicate subpopulations (Supplementary Section S2). However, while the ETA model cannot capture the width of these unimodal distributions, the subpopulation model captures it well (Supplementary Section S2). The 50:50 decomposition into subpopulations therefore represents a first approximation to characterize the continuous variability in the ensemble, which we will thus use in the remainder of the paper.

To put the two-type decomposition to a test, we construct a model that accounts for both CCV and aging by using the slow and fast models, both including the same time-dependent prefactor $\alpha(t)$. Strikingly, this aging subpopulation model can accurately predict the mean-variance curve observed experimentally, indicating that both effects need to be accounted for to capture the behavioural variability of confined cell migration (Fig. 4e). As expected, the mean-variance curves of the subpopulations are thus also accurately predicted by the subpopulation models when aging is included (Supplementary Section S4). Furthermore, the aging subpopulation model accurately captures the other statistics that were already well-predicted by the ETA model, such as the dwell time, the position and velocity distribution and the velocity auto-correlation function (Supplementary Section S4). Taken together, these results indicate that the cell migration in this system exhibits phenotypic heterogeneity

detectable at the level of cell trajectories that can be captured by an equation of motion with two sets of parameters, for a 'fast' and a 'slow' subpopulation of approximately equal sizes; on top of this, the subpopulations homogenously age over time, in the form of a gradual slowing down which is similar for all cells.

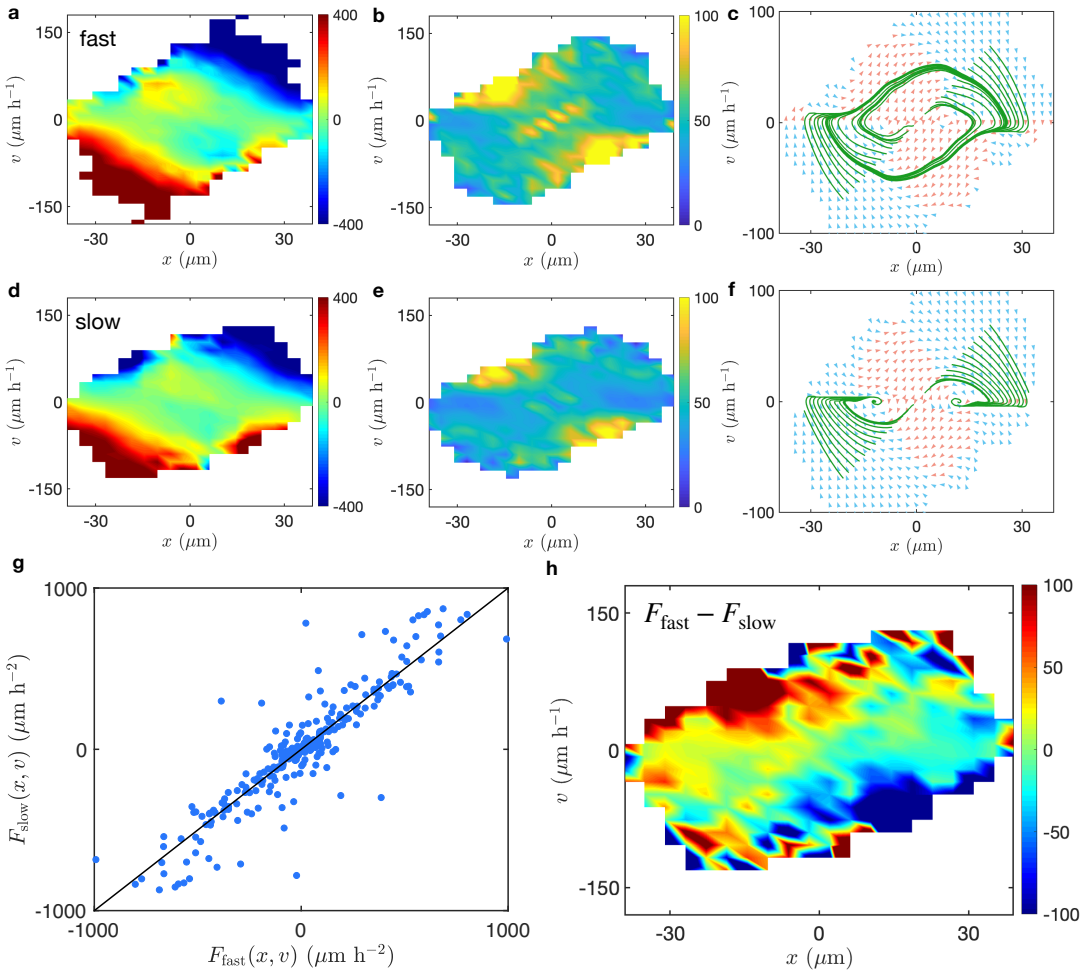


Fig 5. Subpopulation models (*L* = 16 µm). (a), (d) Deterministic terms $F(x, v)$ in units of µm/h^2. (b), (e) The noise strengths $\sigma(x, v)$ in units of µm/h$^{3/2}$. (c), (f) Deterministic flow fields and trajectories. (g) Local deterministic accelerations in the slow model plotted against those in the fast model. (h) The difference between the slow and fast deterministic terms as a function of position in phase space. Panels (a)-(c) correspond to the fast subpopulation, while panels (d)-(f) correspond to the slow subpopulation.

Subpopulations exhibit distinct non-linear dynamics

Our theoretical framework allows us to pinpoint the key aspects of the cellular dynamics that are sensitive to population heterogeneity. To conceptualize the deterministic dynamics, we plot the flow fields in xv-phase space that describe how the purely deterministic system evolves in time (Fig. 5c, f). Here, the region of phase space in which the cell is deterministically accelerating is coloured in orange. This shows that the cells exhibit a deterministic amplification that drives the transitions across the constriction, as we observe deterministic acceleration just before the cell enters the thin constriction. Interestingly, we find that both subpopulation models exhibit similar qualitative features as the ETA model, including this characteristic amplification stage (Fig. 5a, c, d, f). Moreover, the dynamics exhibit similar noise hotspots as in the ETA model (Fig. 5b, e) in states where the cells initiate transition attempts. However, we find that the deterministic acceleration in the amplification phase is larger in the fast model, and similarly the noise hotspot is more pronounced in this case. Interestingly, the differences between the fast and slow models cannot be explained by an overall rescaling factor (Fig. 5g), but rather manifest as local deviations in the phase space, which are largest in the regions of phase space corresponding to states where the cells initiate the transition (Fig. 5h).

The two subpopulation models furthermore exhibit distinct non-linear dynamics: while the fast model exhibits a limit cycle, the slow model is bistable (Fig. 5c, f). These bistable dynamics are however highly excitable, as the amplification region extends nearly all the way to the stable fixed points, meaning that a small perturbation suffices for the system to perform a large deterministic excursion to the other fixed point. The transition from limit cycle to bistable dynamics is therefore the consequence of a relatively small shift in parameters, where only a subtle change in flow is required to modify the system from bistable to limit cycle dynamics. Interestingly, while the subpopulations exhibit distinct deterministic dynamics, the ETA model is dominated by the limit cycle behaviour of the fastest cells (Fig. 1d). This appears to be a robust result, as both the early- and late-stage models inferred separately for the first and second half of the experiment exhibit limit cycle dynamics, despite the slowing down of the average dynamics (Supplementary Section 1).

Discussion

In this work, we have developed a quantitative approach for detecting and characterizing temporal and inter-individual variability in the behaviour of migrating cells. To detect cell-to-cell variability (CCV), we used an ensemble-averaged stochastic equation of cell motion as a reference to determine the inherent variability of single cells migrating within a confining two-state micropattern. This ensemble-based approach successfully captures the key statistics of the cell trajectories, such as the dwell time distribution, the distributions of position and velocity and the velocity correlations. However, here we identify statistics aimed at quantifying the variability of behaviour between cells in the population, and show that they are underestimated by the ensemble-averaged model. We extend our framework to capture this variability, by performing a decomposition of the cell population into subpopulations of similarly behaving cells. This approach then captures the overall variability of the motion and reveals that the subpopulations exhibit distinct deterministic nonlinear dynamics. Taken together, our results show which statistics of the migration are sensitive to time- and population-heterogeneity, and which are well explained by an ensemble-averaged approach.

Using the variance of transition times as a marker for CCV, we found that the experimentally observed variability significantly exceeds that predicted by a single stochastic process. This indicates that migrating cells exhibit CCV that affects the behaviour at the level of cell trajectories. Previous studies have shown that the proteome of single human cells in a population exhibits a broad distribution [12], and this variability is frequently interpreted as a consequence of the inherent stochasticity of various intra-cellular processes [11]. However, CCV at the level of cell behaviour in a population context have in many cases been shown to be determined by environmental factors [17,67,69,70], such as external cues like local cell density, cell-cell contacts and relative location in a cell cluster. Moreover, previous measurements of the phenotypic variability of single cells on micropatterns have shown that the distributions of subcellular organelles are remarkably constant [71], a finding that has been interpreted to indicate that there is little intrinsic variability in isolated cells [67]. In contrast, here we show that isolated migrating cells in identical confining micro-environments do exhibit significant variability that is detectable at the behavioural level.

The CCV-sensitive predictions of our model are improved significantly by decomposing the cell population into subpopulations. Interestingly, evenly dividing cells into two subpopulations of slow and fast cells leads to a model that performs well in predicting inter-cellular variance. This decomposition reveals subtle differences in the system-level dynamics of the subpopulations: the deterministic dynamics of the fast cells exhibit a limit cycle, while those of the slow cells are an excitable bistable system. However, an inspection of our data does not indicate that there are two discrete types of cell behaviour, but rather suggest a

continuum of variation to which a two-type decomposition is a reasonable coarse-grained approximation (Supplementary Section S2). At the single-cell level, we would thus expect a continuous spread of the population over the parameter space. However, this spread spans across the transition line from limit cycle to bistable behaviour, and thus separates the population into distinct subpopulations with different non-linear dynamics. This distinction is similar to our previous finding that two different cell lines exhibit these two classes of non-linear dynamics [32]. Specifically, we found that the dynamics of the non-cancerous and less invasive [72,73] MCF10A cell line can be described as a bistable system, which is in contrast to the limit cycle behaviour of the cancerous MDA-MB-231 cell line that we consider here (Fig. 1d). The presence of these two distinct behaviours thus appears to be a more common feature of confined two-state cell migration. Our observation here that both behaviours are represented among cells of a single cell line highlights the breadth of phenotypic diversity present in this population.

Furthermore, we showed that cells gradually slow down over their life-time, an effect that can be captured by an overall rescaling factor of the equation of motion. The observed time-dependence could be related to the fact that cells grow in size over time [74,75] or could reflect a mechano-sensitive adaption of cells to their confining environment [76]. Similar 'aging' effects have been observed in cell populations [77], due to the gradual 'jamming' of the growing cell sheet. Our results show that there may be an additional contribution at the single-cell level to such observations. Together, the 'aging' effect and the two-type CCV implementation lead to a model that quantitatively captures the long time-scale transition distribution which characterizes the migratory behaviour of cells in the two-state micropattern.

In summary, our work provides a quantitative assessment of the diversity of behavioural phenotypes in a cell population. The unique geometry of the two-state micropattern allows a natural definition of variability in terms of the variances of the transition behaviour, which can be compared to a reference theory for this system. Our quantitative framework may help our understanding of cell-to-cell variability in collective systems [21–24,67,77] and provide useful tools for inverse approaches that attempt to correlate heterogeneity in behaviour to differences in the proteomes of single cells [67,78–81].

Methods

Experiments

We use microscale plasma-induced protein patterning [82] to create standardized cell micro-environments for MDA-MB-231 human breast carcinoma epithelial cells (DSMZ). The cells are cultured in Minimum Essential Medium (MEM, c.c. pro), which is supplemented with 10% FBS (Gibco) and 2mM L-Glutamine (c.c. pro). Cells are grown at 37°C in an atmosphere with 5% CO_2. For passaging, cells are washed and trypsinized for 3 min. This cell solution is centrifuged for 3 min at 1000 rcf, and then cells are re-suspended in MEM. Approximately 10,000 cells are added per µ-dish (ibidi GmbH) and left to adhere for 4h in the incubator. Subsequently, the medium is exchanged to L-15 medium containing L-glutamine (Gibco, supplemented with 10% FCS) and 25 nM Hoechst 33342 (Invitrogen) for fluorescent staining of cell nuclei. Time-lapse measurements are performed in on an IMIC digital microscope (TILL Photonics) or on a Nikon Eclipse Ti microscope. Images in brightfield and fluorescent mode are acquired every 10 min for up to 50 h. The samples kept at 37°C in a heated chamber (ibidi GmbH or Okolab) throughout the measurements. To extract cell trajectories, first a band pass filter is applied to the images of the nuclei. Then images are binarized and *ImageJ*'s Analyze Particles plugin [83] is used to determine the center of the nuclei. A reference boundary of the micropattern is extracted from the brightfield images to yield absolute cell positions. Further details can be found in ref. [32].

Inference of the equation of motion

The position of the cell nucleus $x(t)$ is recorded at a time interval Δt = 10 min in the experiment. Velocity and acceleration are directly calculated as numerical derivatives, given by $v(t) = (x(t) - x(t - \Delta t))/\Delta t$ and $a(t) = (v(t + \Delta t) - v(t))/\Delta t$, respectively. We interpret equation (1) in the Itô-sense throughout, and infer $F(x, v) = \langle \dot{v} | x, v \rangle$ and $\sigma^2(x, v) = \Delta t \langle [\dot{v} - F(x, v)]^2 | x, v \rangle$ by conditional averaging [61,62,84] in a coarse-grained grid of 30x30 bins. To make optimal use of the available data, we exploit the inversion symmetry of the micropattern, which renders the dynamical terms (anti-) symmetric: $F(x, v) = -F(-x, -v)$ and $\sigma(x, v) = \sigma(-x, -v)$. More details can be found in [32].

Acknowledgements

We thank C. Schreiber, D. S. Fischer, E. Frey, J. M. Lober, J. Messelink, G. Gradziuk, M. Schmitt, N. Arlt, P. Ronceray, and G. Stephens and his group for inspiring discussions. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 201269156 – SFB 1032 (Projects B01 and B12). D.B.B. is supported by a DFG fellowship within the Graduate School of Quantitative Biosciences Munich (QBM) and by the Joachim Herz Stiftung.

References

1. Spudich JL, Koshland DE. 1976 Non-genetic individuality: chance in the single cell. *Nature* **262**, 467–471.
2. Aldridge BB, Fernandez-suarez M, Heller D, Ambravaneswaran V, Irimia D, Toner M, Fortune SM. 2012 Growth and Antibiotic Susceptibility. *Science.* **335**, 100–104.
3. Umehara S, Inoue I, Wakamoto Y, Yasuda K. 2007 Origin of individuality of two daughter cells during the division process examined by the simultaneous measurement of growth and swimming property using an on-chip single-cell cultivation system. *Biophys. J.* **93**, 1061–1067. (doi:10.1529/biophysj.106.098061)
4. Min TL, Mears PJ, Chubiz LM, Rao C V., Golding I, Chemla YR. 2009 High-resolution, long-term characterization of bacterial motility using optical tweezers. *Nat. Methods* **6**, 831–835. (doi:10.1038/nmeth.1380)
5. Jordan D, Kuehn S, Katifori E, Leibler S. 2013 Behavioral diversity in microbes and low-dimensional phenotypic spaces. *Proc. Natl. Acad. Sci.* **110**, 14018–14023. (doi:10.1073/pnas.1308282110)
6. Li L, Cox EC, Flyvbjerg H. 2011 'Dicty dynamics': Dictyostelium motility as persistent random motion. *Phys. Biol.* **8**, 046006. (doi:10.1088/1478-3975/8/4/046006)
7. Berman GJ, Choi DM, Bialek W, Shaevitz JW. 2014 Mapping the stereotyped behaviour of freely moving fruit flies. *J. R. Soc. Interface* **11**, 20140672–20140672. (doi:10.1098/rsif.2014.0672)
8. Forkosh O, Karamihalev S, Roeh S, Engel M, Alon U, Nussbaumer M, Flachskamm C, Kaplick P, Shemesh Y. 2018 Identity domains in complex behavior: Toward a biology of personality. (doi:http://dx.doi.org/10.1101/395111)
9. Eysenck HJ. 1953 *The structure of human personality.* New York, NY, US: Methuen.
10. McCrae RR, Costa Jr. PT. 2003 Personality in adulthood: A five-factor theory perspective, 2nd ed. *Personal. adulthood A five-factor theory Perspect. 2nd ed.* (doi:10.4324/9780203428412)
11. Niepel M, Spencer SL, Sorger PK. 2009 Non-genetic cell-to-cell variability and the consequences for pharmacology. *Curr. Opin. Chem. Biol.* **13**, 556–561. (doi:10.1016/j.cbpa.2009.09.015)
12. Sigal A *et al.* 2006 Variability and memory of protein levels in human cells. *Nature* **444**, 28–31. (doi:10.1038/nature05316)
13. Raj A, van Oudenaarden A. 2008 Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell* **135**, 216–226. (doi:10.1016/j.cell.2008.09.050)
14. Wagner A, Regev A, Yosef N. 2016 Revealing the vectors of cellular identity with single-cell genomics. *Nat. Biotechnol.* **34**, 1145–1160. (doi:10.1038/nbt.3711)
15. Altschuler SJ, Wu LF. 2010 Cellular Heterogeneity: Do Differences Make a Difference? *Cell* **141**, 559–563. (doi:10.1016/j.cell.2010.04.033)
16. Holland AJ, Cleveland DW. 2008 Beyond Genetics: Surprising Determinants of Cell Fate in Antitumor Drugs. *Cancer Cell*. (doi:10.1016/j.ccr.2008.07.010)
17. Cohen AA *et al.* 2008 Dynamic Proteomics of Individual Cancer Cells in Response to a Drug. *Science.* **322**, 1511–1516. (doi:10.1126/science.1160165)
18. Feinerman O, Veiga J, Dorfman JR, Germain RN, Altan-Bonnet G. 2008 Variability and Robustness in T Cell Activation from Regulated Heterogeneity in Protein Levels. *Science.* **321**, 1081–1084. (doi:10.1126/science.1158013)
19. Gascoigne KE, Taylor SS. 2008 Cancer Cells Display Profound Intra- and Interline Variation following Prolonged Exposure to Antimitotic Drugs. *Cancer Cell* **14**, 111–122. (doi:10.1016/j.ccr.2008.07.002)
20. Wieser S, Weghuber J, Sams M, Stockinger H, Schütz GJ. 2009 Cell-to-cell variability in the diffusion constants of the plasma membrane proteins CD59 and CD147. *Soft Matter* **5**, 3287–

3294. (doi:10.1039/B902266J)
21. Camley BA, Rappel W. 2017 Cell-to-cell variation sets a tissue-rheology – dependent bound on collective gradient sensing. *Proc. Natl. Acad. Sci. USA* **114**, E10074–E10082. (doi:10.1073/pnas.1712309114)
22. Li X, Das A, Bi D. 2019 Mechanical heterogeneity in tissues promotes rigidity and controls cellular invasion.
23. Copenhagen K, Malet-engra G, Yu W, Scita G, Gov N, Gopinathan A. 2018 Frustration-induced phases in migrating cell clusters. , 1–10.
24. Lee RM, Yue H, Rappel WJ, Losert W. 2017 Inferring single-cell behaviour from largescale epithelial sheet migration patterns. *J. R. Soc. Interface* **14**. (doi:10.1098/rsif.2017.0147)
25. Petrie RJ, Doyle AD, Yamada KM. 2009 Random versus directionally persistent cell migration. *Nat. Rev. Mol. Cell Biol.* **10**, 538–49. (doi:10.1038/nrm2729)
26. Danuser G, Allard J, Mogilner A. 2013 Mathematical Modeling of Eukaryotic Cell Migration: Insights Beyond Experiments. *Annu Rev Cell Dev Biol.* **29**, 501–528. (doi:10.1146/annurev-cellbio-101512-122308.Mathematical)
27. Gail MH, Boone CW. 1970 The Locomotion of Mouse Fibroblasts in Tissue Culture. *Biophys. J.* **10**, 980–993. (doi:10.1016/S0006-3495(70)86347-0)
28. Wu P-H, Giri A, Sun SX, Wirtz D. 2014 Three-dimensional cell migration does not follow a random walk. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 3949–54. (doi:10.1073/pnas.1318967111)
29. Selmeczi D, Mosler S, Hagedorn PH, Larsen NB, Flyvbjerg H. 2005 Cell motility as persistent random motion: theories from experiments. *Biophys. J.* **89**, 912–931. (doi:10.1529/biophysj.105.061150)
30. Selmeczi D, Li L, Pedersen LII, Nrrelykke SF, Hagedorn PH, Mosler S, Larsen NB, Cox EC, Flyvbjerg H. 2008 Cell motility as random motion: A review. *Eur. Phys. J. Spec. Top.* **157**, 1–15. (doi:10.1140/epjst/e2008-00626-x)
31. Takagi H, Sato MJ, Yanagida T, Ueda M. 2008 Functional analysis of spontaneous cell movement under different physiological conditions. *PLoS One* **3**, e2648. (doi:10.1371/journal.pone.0002648)
32. Brückner DB, Fink A, Schreiber C, Röttgermann PJF, Rädler JO, Broedersz CP. 2019 Stochastic nonlinear dynamics of confined cell migration in two-state systems. *Nat. Phys.* **15**, 595–601. (doi:10.1038/s41567-019-0445-4)
33. Fink A, Brückner DB, Schreiber C, Röttgermann PJF, Broedersz CP, Rädler JO. 2019 Area and Geometry Dependence of Cell Migration in Asymmetric Two-State Micropatterns Condensed running title : Geometry Dependence of Cell Migration. *Biophys. J.*
34. Lavi I, Piel M, Lennon-Duménil A-M, Voituriez R, Gov NS. 2016 Deterministic patterns in cell motility. *Nat. Phys.* **12**, 1146–1152. (doi:10.1038/nphys3836)
35. Potdar AA, Jeon J, Weaver AM, Quaranta V, Cummings PT. 2010 Human mammary epithelial cells exhibit a bimodal correlated random walk pattern. *PLoS One* **5**, e9636. (doi:10.1371/journal.pone.0009636)
36. Vestergaard CL, Pedersen JN, Mortensen KI, Flyvbjerg H. 2015 Estimation of motility parameters from trajectory data: A condensate of our recent results. *Eur. Phys. J. Spec. Top.* **224**, 1151–1168. (doi:10.1140/epjst/e2015-02452-5)
37. Thüroff F, Goychuk A, Reiter M, Frey E. 2019 Bridging the gap between single cell migration and collective dynamics. (doi:http://dx.doi.org/10.1101/548677)
38. Dietrich M, Le Roy H, Brückner DB, Engelke H, Zantl R, Rädler JO, Broedersz CP. 2018 Guiding 3D cell migration in deformed synthetic hydrogel microstructures. *Soft Matter* (doi:10.1039/C8SM00018B)
39. Albert PJ, Schwarz US. 2016 Dynamics of Cell Ensembles on Adhesive Micropatterns: Bridging the Gap between Single Cell Spreading and Collective Cell Migration. *PLoS Comput. Biol.* **12**,

1–34. (doi:10.1371/journal.pcbi.1004863)
40. Segerer FJ, Thüroff F, Piera Alberola A, Frey E, Rädler JO. 2015 Emergence and persistence of collective cell migration on small circular micropatterns. *Phys. Rev. Lett.* **114**, 228102. (doi:10.1103/PhysRevLett.114.228102)
41. Ziebert F, Swaminathan S, Aranson IS. 2012 Model for self-polarization and motility of keratocyte fragments. *J. R. Soc. Interface* **9**, 1084–1092. (doi:10.1098/rsif.2011.0433)
42. Camley BA, Zhang Y, Zhao Y, Li B, Ben-Jacob E, Levine H, Rappel W-J. 2014 Polarity mechanisms such as contact inhibition of locomotion regulate persistent rotational motion of mammalian cells on micropatterns. *Proc. Natl. Acad. Sci.* **111**, 14770–14775. (doi:10.1073/pnas.1414498111)
43. Goychuk A, Brückner DB, Holle AW, Spatz JP, Broedersz CP, Frey E. 2018 Morphology and Motility of Cells on Soft Substrates.
44. Albert PJ, Schwarz US. 2016 Modeling cell shape and dynamics on micropatterns. *Cell Adhes. Migr.* **10**, 516–528. (doi:10.1080/19336918.2016.1148864)
45. Maiuri P *et al.* 2015 Actin flows mediate a universal coupling between cell speed and cell persistence. *Cell* **161**, 374–386. (doi:10.1016/j.cell.2015.01.056)
46. Metzner C, Mark C, Steinwachs J, Lautscham L, Stadler F, Fabry B. 2015 Superstatistical analysis and modelling of heterogeneous random walks. *Nat. Commun.* **6**, 7516. (doi:10.1038/ncomms8516)
47. Green BJ, Panagiotakopoulou M, Pramotton FM, Stefopoulos G, Kelley SO, Poulikakos D, Ferrari A. 2018 Pore shape defines paths of metastatic cell migration. *Nano Lett.* **18**, 2140−2147.
48. Paul CD, Mistriotis P, Konstantopoulos K. 2017 Cancer cell motility: Lessons from migration in confined spaces. *Nat. Rev. Cancer* **17**, 131–140. (doi:10.1038/nrc.2016.123)
49. Davidson PM, Battistella A, Déjardin T, Betz T, Plastino J, Cadot B, Borghi N, Sykes C. 2019 Actin accumulates nesprin-2 at the front of the nucleus during confined cell migration. (doi:https://doi.org/10.1101/713982)
50. Renkawitz J *et al.* 2019 Nuclear positioning facilitates amoeboid migration along the path of least resistance. *Nature* **568**, 546–550. (doi:10.1038/s41586-019-1087-5)
51. Patteson AE, Pogoda K, Byfield FJ, Charrier EE, Peter A, Deptu P, Bucki R, Janmey PA. 2019 Loss of vimentin intermediate filaments decreases peri-nuclear stiffness and enhances cell motility through confined spaces.
52. Wolf K *et al.* 2013 Physical limits of cell migration: Control by ECM space and nuclear deformation and tuning by proteolysis and traction force. *J. Cell Biol.* **201**, 1069–1084. (doi:10.1083/jcb.201210152)
53. Paul CD, Shea DJ, Mahoney MR, Chai A, Laney V, Hung WC, Konstantopoulos K. 2016 Interplay of the physical microenvironment, contact guidance, and intracellular signaling in cell decision making. *FASEB J.* **30**, 2161–2170. (doi:10.1096/fj.201500199R)
54. Wilson K, Lewalle A, Fritzsche M, Thorogate R, Duke T, Charras G. 2013 Mechanisms of leading edge protrusion in interstitial migration. *Nat. Commun.* **4**, 1–12. (doi:10.1038/ncomms3896)
55. Liu YJ, Le Berre M, Lautenschlaeger F, Maiuri P, Callan-Jones A, Heuzé M, Takaki T, Voituriez R, Piel M. 2015 Confinement and low adhesion induce fast amoeboid migration of slow mesenchymal cells. *Cell* **160**, 659–672. (doi:10.1016/j.cell.2015.01.007)
56. Charras G, Sahai E. 2014 Physical influences of the extracellular environment on cell migration. *Nat. Rev. Mol. Cell Biol.* **15**, 813–824. (doi:10.1038/nrm3897)
57. Wolf K, Alexander S, Schacht V, Coussens LM, von Andrian UH, van Rheenen J, Deryugina E, Friedl P. 2009 Collagen-based cell migration models in vitro and in vivo. *Semin. Cell Dev. Biol.* **20**, 931–941. (doi:https://doi.org/10.1016/j.semcdb.2009.08.005)

58. Friedl P, Alexander S. 2011 Cancer invasion and the microenvironment: Plasticity and reciprocity. *Cell* **147**, 992–1009. (doi:10.1016/j.cell.2011.11.016)
59. Gritsenko PG, Ilina O, Friedl P. 2012 Interstitial guidance of cancer invasion. *J. Pathol.* **226**, 185–199. (doi:10.1002/path.3031)
60. Siegert S, Friedrich R, Peinke J. 1998 Analysis of data sets of stochastic systems. *Phys. Lett. Sect. A Gen. At. Solid State Phys.* **243**, 275–280. (doi:10.1016/S0375-9601(98)00283-7)
61. Ragwitz M, Kantz H. 2001 Indispensable Finite Time Corrections for Fokker-Planck Equations from Time Series Data. *Phys. Rev. Lett.* **87**, 254501. (doi:10.1103/PhysRevLett.87.254501)
62. Stephens GJ, Johnson-Kerner B, Bialek W, Ryu WS. 2008 Dimensionality and Dynamics in the Behavior of C. elegans. *PLoS Comput Biol* **4**, e1000028. (doi:10.1371/journal.pcbi.1000028)
63. Svoboda K, Mitra PP, Block SM. 1994 Fluctuation analysis of motor protein movement and single enzyme kinetics. *Proc. Natl. Acad. Sci. USA* **91**, 11782–11786. (doi:10.1073/pnas.91.25.11782)
64. Taniguchi Y, Choi PJ, Li G-W, Chen H, Babu M, Hearn J, Emili A, Xie XS. 2010 Quantifying E. coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. *Science (80-.).* **329**, 533–538. (doi:10.1126/science.1188308)
65. Sherman MS, Lorenz K, Lanier MH, Cohen BA. 2015 Cell-to-Cell Variability in the Propensity to Transcribe Explains Correlated Fluctuations in Gene Expression. *Cell Syst.* **1**, 315–325. (doi:10.1016/j.cels.2015.10.011)
66. Ben-David U *et al.* 2018 Genetic and transcriptional evolution alters cancer cell line drug response. *Nature* **560**, 325–330. (doi:10.1038/s41586-018-0409-3)
67. Snijder B, Pelkmans L. 2011 Origins of regulated cell-to-cell variability. *Nat. Rev. Mol. Cell Biol.* **12**, 119–125. (doi:10.1038/nrm3044)
68. Schreiber C, Segerer FJ, Wagner E, Roidl A, Rädler JO. 2016 Ring-Shaped Microlanes and Chemical Barriers as a Platform for Probing Single-Cell Migration. *Sci. Rep.* **6**, 26858. (doi:10.1038/srep26858)
69. Snijder B, Sacher R, Rämö P, Damm E-M, Liberali P, Pelkmans L. 2009 Population context determines cell-to-cell variability in endocytosis and virus infection. *Nature* **461**, 520.
70. Colman-Lerner A, Gordon A, Serra E, Chin T, Resnekov O, Endy D, Gustavo Pesce C, Brent R. 2005 Regulated cell-to-cell variation in a cell-fate decision system. *Nature* **437**, 699–706. (doi:10.1038/nature03998)
71. Schauer K, Duong T, Bleakley K, Bardin S, Bornens M, Goud B. 2010 Probabilistic density maps to study global endomembrane organization. *Nat. Methods* **7**, 560–566. (doi:10.1038/nmeth.1462)
72. Mak M, Reinhart-King CA, Erickson D. 2011 Microfabricated physical spatial gradients for investigating cell migration and invasion dynamics. *PLoS One* **6**, e20825. (doi:10.1371/journal.pone.0020825)
73. Kraning-Rush CM, Carey SP, Lampi MC, Reinhart-King CA. 2013 Microfabricated collagen tracks facilitate single cell metastatic invasion in 3D. *Integr. Biol.* **5**, 606–616. (doi:10.1039/c3ib20196a)
74. Théry M. 2010 Micropatterning as a tool to decipher cell morphogenesis and functions. *J. Cell Sci.* **123**, 4201–4213. (doi:10.1242/jcs.075150)
75. Suffoletto K, Ye N, Meng F, Verma D, Hua SZ. 2015 Intracellular forces during guided cell growth on micropatterns using FRET measurement. *J. Biomech.* **48**, 627–635. (doi:10.1016/j.jbiomech.2014.12.051)
76. Gegenfurtner FA, Jahn B, Wagner H, Ziegenhain C, Enard W, Geistlinger L, Rädler JO, Vollmar AM, Zahler S. 2018 Micropatterning as a tool to identify regulatory triggers and kinetics of actin-mediated endothelial mechanosensing. *J. Cell Sci.* **131**, jcs212886. (doi:10.1242/jcs.212886)

77. Garcia S, Hannezo E, Elgeti J, Joanny J-F, Silberzan P, Gov NS. 2015 Physics of active jamming during collective cellular motion in a monolayer. *Proc. Natl. Acad. Sci.* **112**, 15314–15319. (doi:10.1073/pnas.1510973112)
78. Chan TE, Stumpf MPH, Babtie AC. 2017 Gene Regulatory Network Inference from Single-Cell Data Using Multivariate Information Measures. *Cell Syst.* **5**, 251–267. (doi:10.1016/j.cels.2017.08.014)
79. Todorov H, Cannoodt R, Saelens W, Saeys Y. 2019 Network Inference from Single-Cell Transcriptomic Data BT - Gene Regulatory Networks: Methods and Protocols. In (eds G Sanguinetti, VA Huynh-Thu), pp. 235–249. New York, NY: Springer New York. (doi:10.1007/978-1-4939-8882-2_10)
80. Li B, You L. 2013 Predictive power of cell-to-cell variability. *Quant. Biol.* **1**, 131–139. (doi:10.1007/s40484-013-0013-3)
81. Sachs K, Perez O, Pe'er D, Lauffenburger DA, Nolan GP. 2005 Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data. *Science.* **308**, 523–529. (doi:10.1126/science.1105809)
82. Segerer FJ, Röttgermann PJF, Schuster S, Piera Alberola A, Zahler S, Rädler JO. 2016 Versatile method to generate multiple types of micropatterns. *Biointerphases* **11**, 011005. (doi:10.1116/1.4940703)
83. Schneider CA, Rasband WS, Eliceiri KW. 2012 NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675. (doi:10.1038/nmeth.2089)
84. Siegert S, Friedrich R, Peinke J. 1998 Analysis of data sets of stochastic systems. *Phys. Lett. A* **243**, 275–280. (doi:10.1016/S0375-9601(98)00283-7)

Supplementary Information for

Disentangling the Behavioural Variability of Confined Cell Migration

David B. Brückner[1], Alexandra Fink[2], Joachim O. Rädler[2], Chase P. Broedersz[1+]

[1] Arnold-Sommerfeld-Center for Theoretical Physics and Center for NanoScience, Ludwig- Maximilians-Universität, München
[2] Faculty of Physics and Center for NanoScience, Ludwig-Maximilians-Universität, München

[+] corresponding author (c.broedersz@lmu.de)

0. Movie and data descriptions

Supplementary Movies S1 and S2. Single MDA-MB-231 cells transitioning repeatedly between the square adhesion sites of the two-state micropattern. The cell nucleus is fluorescently labelled to allow automated tracking of cell positions. The corresponding trajectories are shown in Fig. 1b, highlighted by grey shading. Movie S1 corresponds to the trajectory in row 3, column 3, while movie S2 corresponds to row 7, column 3. The bridge length is L = 16 µm. Scale bar: 20 µm.

Supplementary Data Sets S1 to S6. The data sets S1 - 6 include the trajectories of nuclear positions $x(t)$ for all six bridge lengths L = 6 - 46 µm. In each file, different rows correspond to different cells. As trajectories have different lengths, they are padded with NaN entries at the end. The time interval is Δt = 10 min.

1. Description of time-variability with early and late stage models

In the main text, we incorporated the time variability of the cell migration as an overall time-dependent prefactor to the equation of motion, which we measured from the trajectories. An alternative approach to this procedure would be to decompose the temporal dynamics into separate stages. The simplest decomposition is a two-stage model, with separate parameter sets for the first and second half of the experiment. Inferring these two models, we find that they exhibit similar dynamics and that in both cases the deterministic dynamics exhibits a limit cycle. However, our simulation of this two-stage model did not improve the prediction of the transition variability beyond that of the time- and ensemble-averaged model. More fine-grained decompositions, starting with a three-stage decomposition, turned out to be problematic as there was then not enough data to accurately infer the latest stage, due to the variable length of cell trajectories. We thus do not rule out this approach, but find that its execution is hindered by the amount of data available.

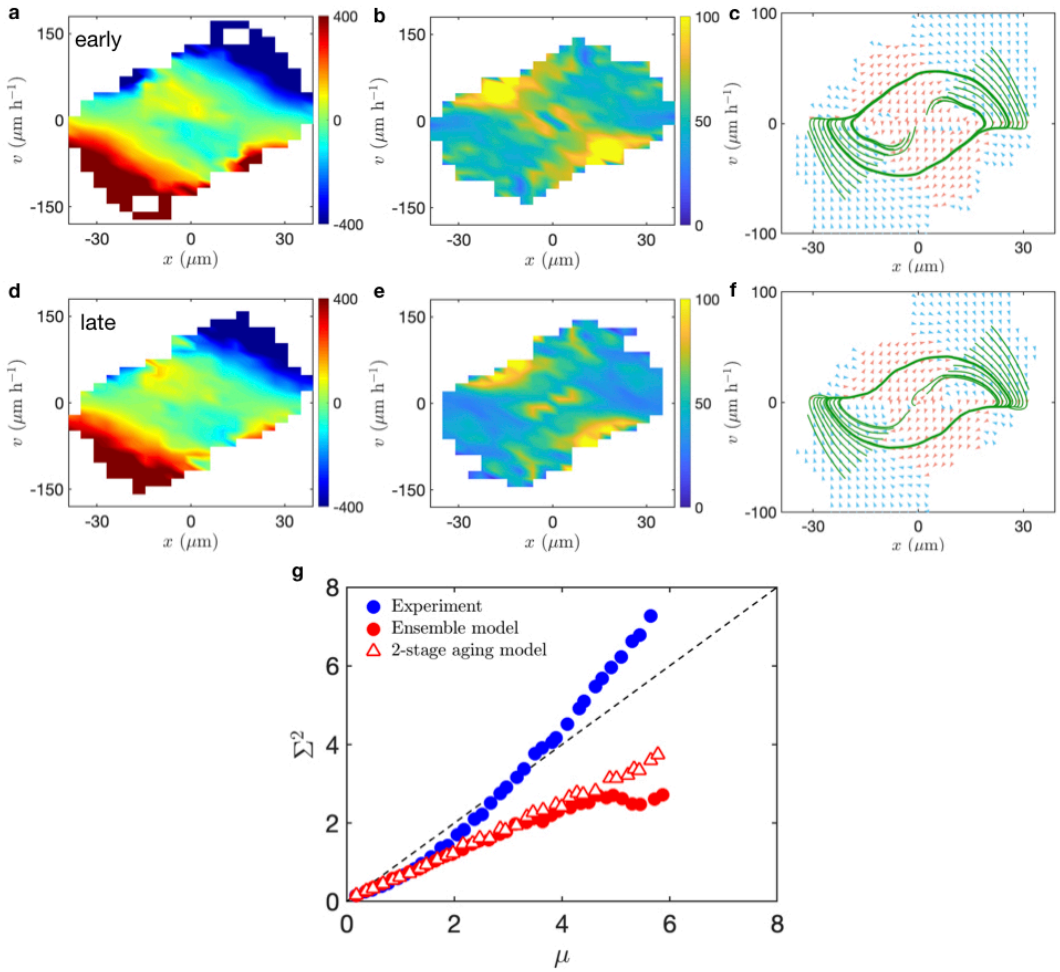


Fig S1. Early- and late-stage models (*L* = 16 µm). (a), (d) Deterministic terms $F(x, v)$ in units of µm/h^2. (b), (e) The noise strengths $\sigma(x, v)$ in units of µm/h$^{3/2}$. (c), (f) Deterministic flow fields and trajectories. (g) Mean-variance curves of the ensemble model (red dots) and the 2-stage aging model (open triangles) compared to experiment (blue dots).

2. Distributions of single cell observables

As we discuss in the main text, the distribution of single cell behaviours appears continuous across the population (Fig. 4a). To further quantify this finding, we plot the probability distributions of single cell parameters, namely the single cell average dwell times and the single cell average speeds (Fig. S2a, b). Indeed, both of these distributions exhibit a unimodal structure, with no clear peaks that would indicate subpopulations. However, the ETA model significantly underestimates the spread of this peak (Fig. S2c), while our approach of inferring two subpopulation models significantly improves the prediction of this distribution. Indeed, as shown in Fig. S2d, the decomposition of the dynamics into a slow and fast subpopulation gives rise to two significantly distinct subpopulation distributions with distinct peaks, which combined give the population distribution (Fig. S2a).

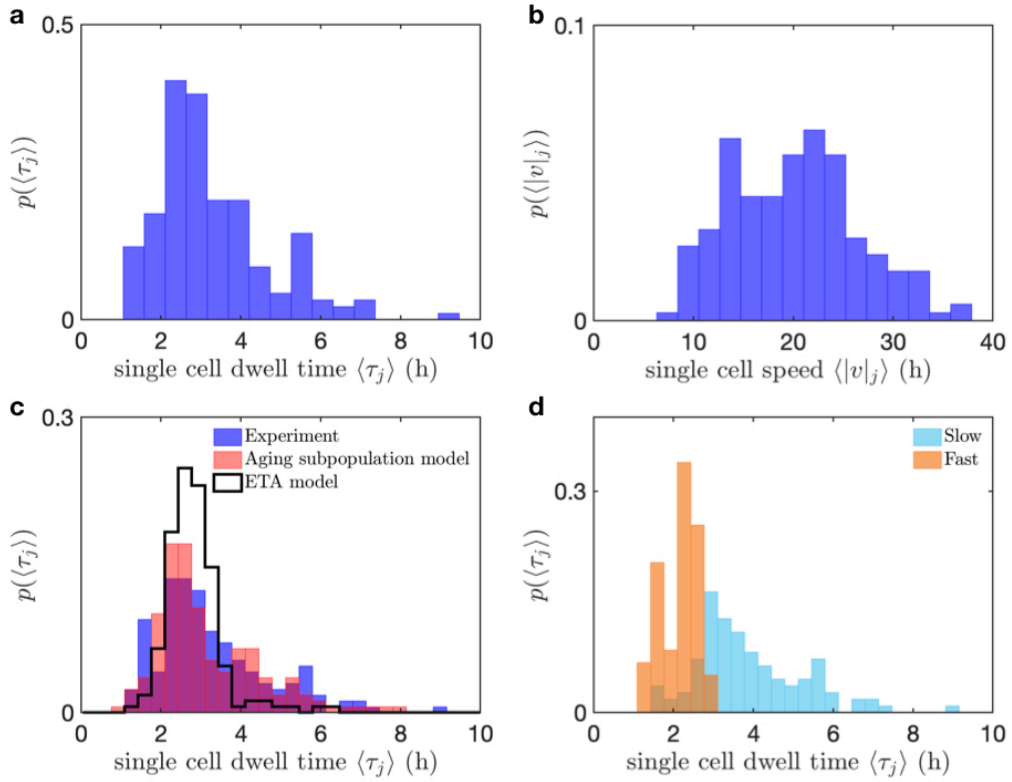


Fig S2. Distributions of single cell observables (*L* = 16 µm). (a) Probability distribution of single cell average dwell times $\langle\tau(j)\rangle_t$. (b) Probability distribution of single cell average speeds $\langle|v|_j\rangle_t$. (c) Probability distributions of single cell average dwell times $\langle\tau(j)\rangle_t$ compared between experiment (blue), ETA model (black outline), and the aging subpopulation model (red). (d) Probability distributions of the experimentally observed single cell average dwell times $\langle\tau(j)\rangle_t$ in the slow (light blue) and fast (orange) subpopulations.

3. Fitting single-cell prefactors

As an alternative approach to inferring subpopulation models to capture CCV with little data, we tested a model where we measure a single-cell prefactor, in a similar spirit to the prefactor that accounts for the 'aging' effect. Specifically, we postulate that the equation of motion for each cell *j* takes the form

$$\frac{dv}{dt} = \gamma(j)[F(x,v) + \sigma(x,v)\eta(t)] \qquad (S1)$$

where we estimate γ using $\gamma(j) = \sqrt{\langle \dot{v}^2(j) \rangle}/\sqrt{\langle \dot{v}^2 \rangle}$. The experimental data exhibits a broad distribution of γ's, significantly broader than that found for a simulation of the ensemble-averaged model (Fig. S3a). This suggests that the prefactor γ is able to detect CCV. However, it does not correlate strongly with the average dwell time of each cell (Fig. S3b), and a model with a combination of the aging prefactor and the single cell prefactor does not predict the transition variance well (Fig. S3c) and leads to significant systematic errors in ensemble averaged quantities such as the average dwell time (Fig. S3d). We therefore conclude that fitting such single-cell prefactors does not provide a quantitatively accurate description of CCV in this system. This also makes sense in light of the fact that the key differences in the fast and slow subpopulation models are local features in phase space rather than an overall rescaling of the whole dynamical landscape (Fig. 5g,h).

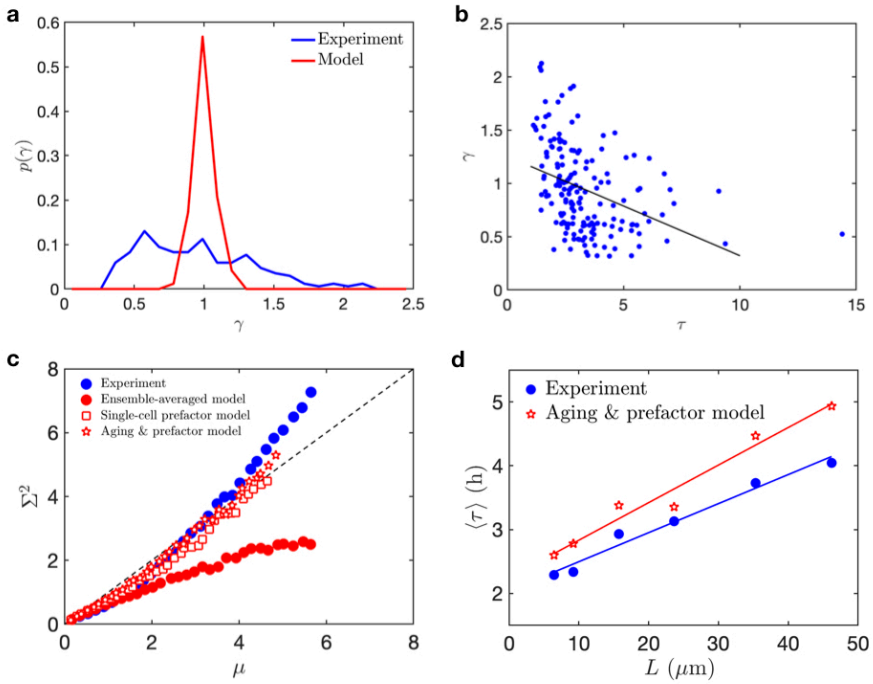


Fig S3. Single-cell prefactor model. (a) Probability distributions of single-cell prefactors in experiment (blue) and the ensemble-averaged model, simulated with the same amount of statistics as in the experiment (red). (b) Correlation plot of the single-cell prefactor with the single-cell average dwell time. Each dot represents a single cell. Black line is a linear fit to the data. (c) Mean-variance curves of

the ensemble model (red dots), the single-cell prefactor model (open squares) and the single-cell prefactor model combined with the prefactor aging model (open pentagons), compared to experiment (blue dots). (d) Ensemble-averaged dwell time as a function of the length of the connecting bridge, for experiment (blue dots) and the single-cell prefactor model combined with the prefactor aging model (open pentagons). Lines correspond to linear fits. Panels (a)-(c) correspond to bridge length L = 16 µm.

4. Testing the aging subpopulation model

By calculating the mean-variance curves for each subpopulation, we find that neither reaches the cross-over to super-Poissonian behaviour in the experimentally accessible time-scales. However, both exhibit a qualitatively similar non-linear trend as the population averaged result (Fig. 2c). The subpopulation curves are not captured by non-aging models for each subpopulation (Fig. S4a,b), but if aging is taken into account, the curves are well predicted (Fig. S4c,d).

As a consistency check, we further validate the aging subpopulation model by showing that it correctly predicts the time- and ensemble-averaged statistics that were already well predicted by the ETA model (Fig. S5).

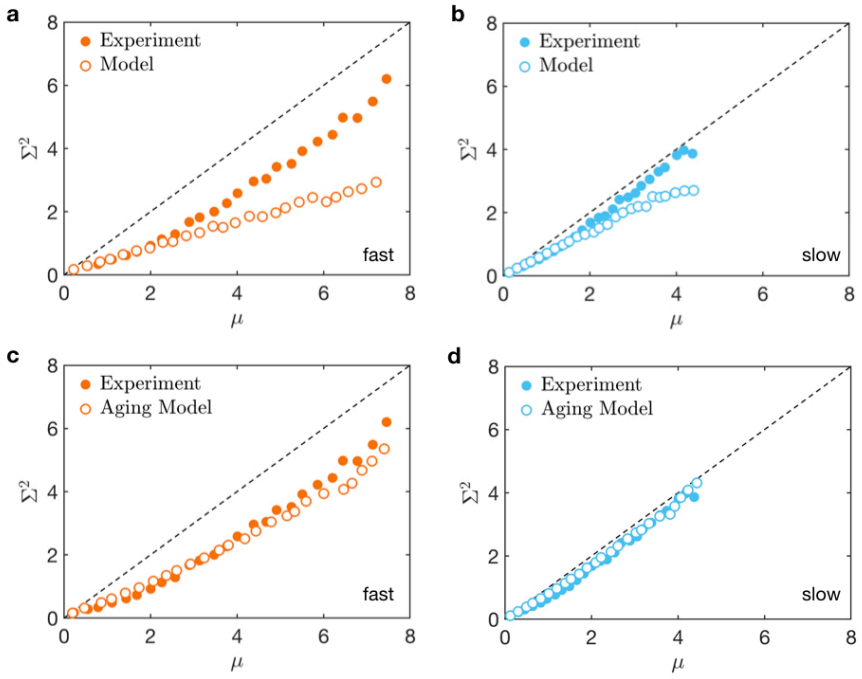


Fig S4. Mean-variance curves of the subpopulations (L = 16 µm). (a) Fast subpopulation compared to the non-aging fast model. (b) Slow subpopulation compared to the non-aging slow model. (c) Fast subpopulation compared to the aging fast model. (d) Slow subpopulation compared to the non-aging slow model.

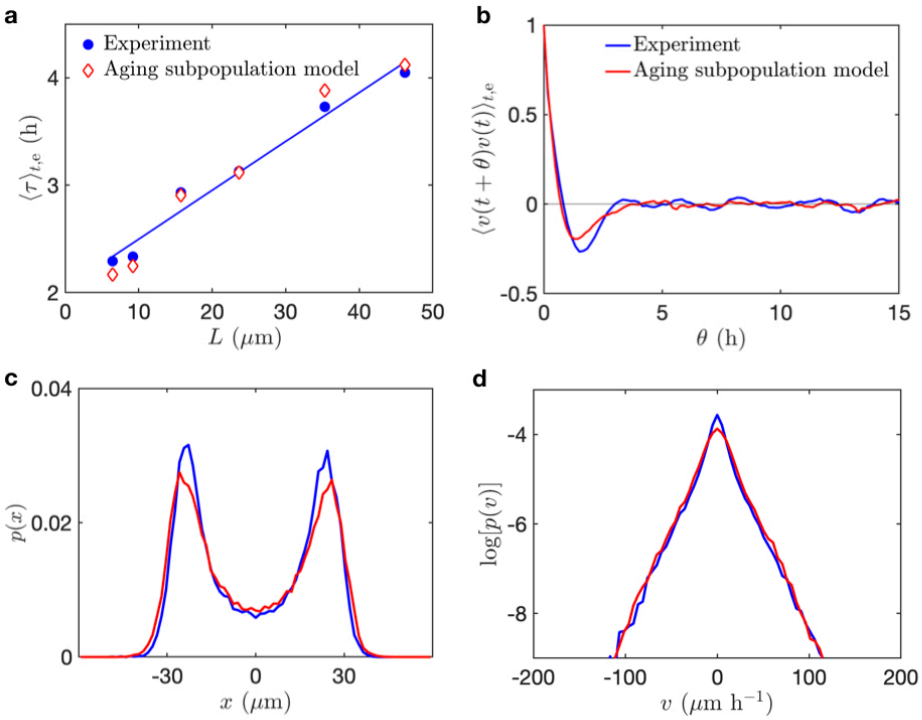


Fig S5. Prediction of time- and ensemble averaged statistics by the aging subpopulation model. (a) Average dwell time as a function of bridge length. (b) Velocity auto-correlation. (c) Probability distribution of cell positions. (d) Probability distribution of cell velocity, plotted on log-linear axes.

5. Choice of time window for time-dependence of the dwell times

In Fig. 3a, we show the average dwell time plotted as a function of time. While the root-mean-square acceleration (Fig. 3b) can be determined in every time frame, we employ averaging over a finite time window for the dwell times. This is because the average dwell time is of the order of hours, and there is therefore no well-defined measure of an instantaneous dwell time. In Fig. S1, we show that our results are not sensitive to the choice of the time window used for the averaging.

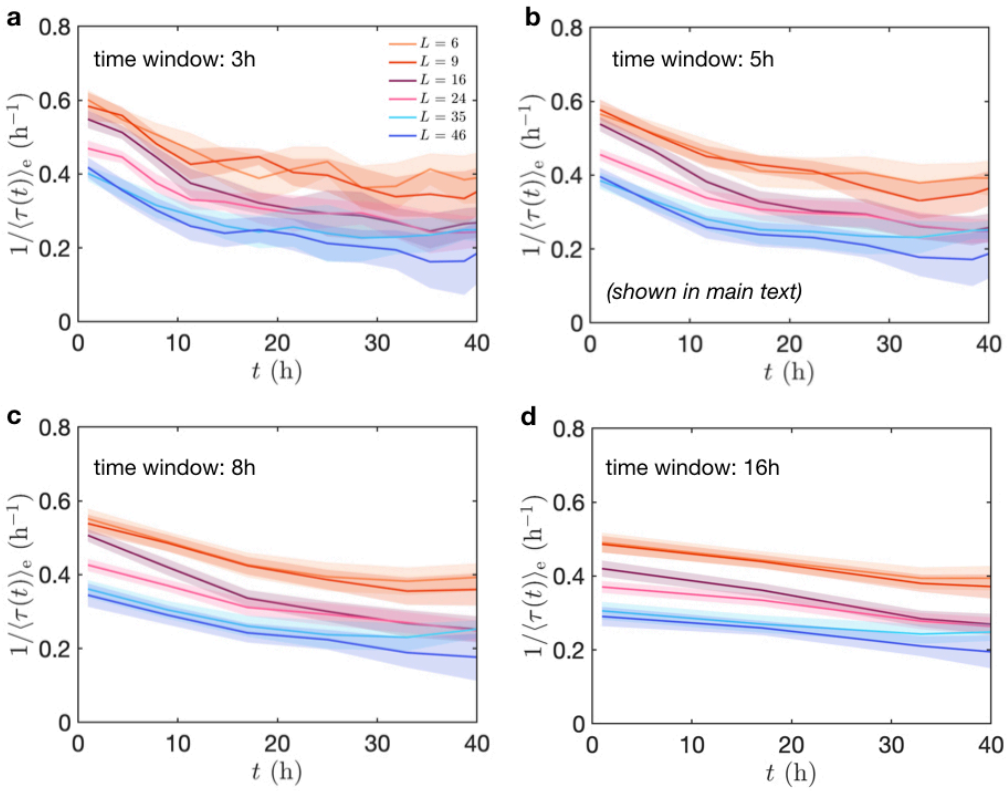


Fig S6. Inverse dwell times as a function of time for different time window averages. Panel (b) is shown in the main text.

Chapter 5

Cellular responses to varying confinement size, shape and orientation

This chapter is based on the following publication:

Area and geometry dependence of cell migration in asymmetric two-state micropatterns

Alexandra Fink, David B. Brückner, Christoph Schreiber, Peter J. F. Röttgermann, Chase P. Broedersz, Joachim O. Rädler†

† corresponding author

Biophysical Journal 110, 1886-1895 (2020)

Area and Geometry Dependence of Cell Migration in Asymmetric Two-State Micropatterns

Alexandra Fink[1], David B. Brückner[2], Christoph Schreiber[1], Peter J. F. Röttgermann[1], Chase P. Broedersz[2] and Joachim O. Rädler[1+]

[1] Faculty of Physics and Center for NanoScience, Ludwig-Maximilians-Universität, München
[2] Arnold Sommerfeld Center for Theoretical Physics and Center for NanoScience, Ludwig-Maximilians-Universität, München

[+] corresponding author

Micro-structured surfaces provide a unique framework to probe cell migration and cytoskeletal dynamics in a standardized manner. Here, we report on the steady-state occupancy probability of cells in asymmetric two-state microstructures that consist of two fibronectin-coated adhesion sites connected by a thin guidance cue. In these dumbbell-like structures, cells transition between the two sites in a repeated and stochastic manner and average dwell times in the respective microenvironments are determined from the cell trajectories. We study the dynamics of human breast carcinoma cells (MDA-MB-231) in these microstructures as a function of area, shape and orientation of the adhesion sites. On square adhesive sites with different areas, we find that the occupancy probability ratio is directly proportional to the ratio of corresponding adhesion site areas. These asymmetries are well captured by a simple model for the stochastic nonlinear dynamics of the cells which reveals generic features of the motion. Sites of equal area but different shape lead to equal occupancy, if shapes are isotropic, e.g. squared or circular. In contrast, an asymmetry in the occupancy is induced by anisotropic shapes like rhombi, triangles or rectangles that enable motion in the direction perpendicular to the transition axis. Analysis of the 2D motion of cells between two rectangles with orthogonal orientation suggests that cellular transition rates depend on the cell polarisation induced by anisotropic micropatterns. Taken together, our results illustrate how two-state-micropatterns provide a dynamic migration assay with distinct dwell times and relative cell occupancy as readouts, which may be useful to probe cell-microenvironment interactions.

Significance — Disentangling the determinants driving cell migration is ultimately important for understanding cell migration in cancer, development and inflammation. To investigate the role of cellular microenvironments, we use novel two-state micropatterns consisting of two adhesion sites connected by a thin stripe. These dumbbell-like micropatterns form a unique framework on which cells are found to repeatedly migrate between the two sites. The assay thus provides numerous time-resolved readouts from single-cell transitions and yields a population-averaged steady-state distribution cellular dwell times in each site. The dwell times depend on the area, shape and orientation of the adhesion sites. We quantify relative cellular preferences and rank absolute escape rates, paving the way for applications such as cell-based choice assays for biophysical studies.

In many fundamental biological processes like early development, cancer metastasis and inflammation, cells are guided by external cues. Understanding the factors and mechanisms that steer cells in tissue and in the extracellular matrix is still an important open biophysical challenge. Many cell-based assays provide platforms to quantify cell behaviour in purposefully designed settings. Well studied are the mechanisms of cell guidance by external cues, for example chemotaxis enabling cells to follow soluble chemical gradients, or durotaxis, in which a motile cell is guided by mechanical stiffness gradients in the surrounding matrix (1-5). Likewise, gradients of surface bound integrin ligands and differences in extracellular matrix protein density can guide cell migration in two dimensions (6-8). Cellular preferences or decision making is typically probed on the population level. In such experiments, cell-substrate interactions are tuned over a large surface area and the distribution of cells is measured as endpoint readout rather than in a time-resolved manner (9-15). A more direct way to probe cell decision making in the presence of guidance cues is to provide assays where single cells are given the choice to spread on one of two different kinds of surfaces. In neuronal research, biochemical assays comprising adjacent lines of different proteins are known to study the preferential alignment and adhesion of neurons (16, 17). Also, the response of sensory neurites and growth cones to different substrates and patterns of different proteins has been studied (18, 19). Thus, micropatterning has advanced as a tool for single cell adhesion and migration studies providing standardized microenvironments.

Various micro-structuring techniques enable the spatially controlled deposition of proteins in defined areas (20-23). Cells seeded on micropatterns spread and adapt to the shape and size of the adhesive area (24). Regular lattices of micropatterns have been used to capture and strain single cells for a multitude of studies, including spindle-orientation (25, 26), the influence of cell shape on cell polarization (27-31), and to monitor cell responses over time (32, 33). For the investigation of cell migration dynamics, micropatterned lanes have proven useful as they confine cell motion to one dimension (34-37). Variations of microlanes with symmetry-breaking triangular or teardrop-shaped patterns demonstrate control over the direction of cell migration by geometry (37-44). One key question in these studies of cell spreading and migration on micropatterns is how geometry affects the cytoskeletal organization, e.g. the distribution of stress fibres and focal adhesions, and consequently how

this relates to changes in the migration dynamics. Micropatterns offer highly reproducible microenvironments and allow the extraction of extensive single cell statistics and are therefore well suited to reveal subtle changes in a quantifiable manner.

Recently, we reported on a novel type of confined migration of cells in two-state micropatterns (45). The symmetric micropatterns consist of two approximately cell-sized adhesive sites, connected by a thin bridge. On these dumbbell-shaped structures, cells transition between the two sites in a repeated and stochastic manner. Using a theoretical approach, we decomposed cell motion in this confinement into distinct deterministic and stochastic components and demonstrated that different cell lines can exhibit distinct nonlinear deterministic dynamics. Furthermore, we showed that these dynamics scale with the length of the connecting bridge (45). As cells also adapt their shape to the geometry of the adhesion site, it is an open question, whether the geometry of the pattern affects the escape rates in a dynamic assay. Two-state micropatterns represent an ideal system to study escape dynamics from distinct adhesion sites. However, so far only the dynamics of cells transitioning between adhesion sites of the same shape has been studied.

In this study, we explore the dwell times and steady-state distribution of cells on asymmetric two-state micropatterns with adhesive sites that feature different geometric properties. This system represents a simple implementation of a confining microenvironment, which is capable of probing the preference of cells for defined surface geometries. The micropatterns consist of two fibronectin-coated adhesion sites that are connected by a thin fibronectin-coated stripe. In these structures, cells transition repeatedly between the two adhesion sites. We monitor the trajectories of the fluorescently labelled nuclei and quantify the occupation probabilities on each adhesion site. Instead of primarily focussing on the short time-scale dynamics of the migration, here we identify coarse-grained cellular read-outs such as dwell times and occupancy probabilities to characterise the migration. First, we investigate how adhesion area affects transit rates and show that the relative dwell times between subsequent transitions on differently sized square adhesion sites scale approximately linearly with the ratio of the adhesion areas. Next, we study pairs of sites with equal area but different isotropic geometries, such as squares and circles, result in equal distributions indicating a weak effect of site shapes on dwell times. However, patterns that include anisotropic sites, e.g. squares versus rhombi or triangles show skewed occupancies. We attribute this effect to induced cell polarization, which we quantify by analysing the directional motion and by imaging the actin stress fibres and focal adhesions. We also apply this analysis to equally sized rectangular adhesion sites arranged in different orientations with respect to the connecting bridge. In this case, we observe a preferred occupancy in the rectangle that is oriented perpendicular to the direction of transitions. Our observations are consistent with a simple kinetic rate model and we discuss possible applications of dynamic assays to measure relative affinities of cells in artificially designed asymmetric microenvironments.

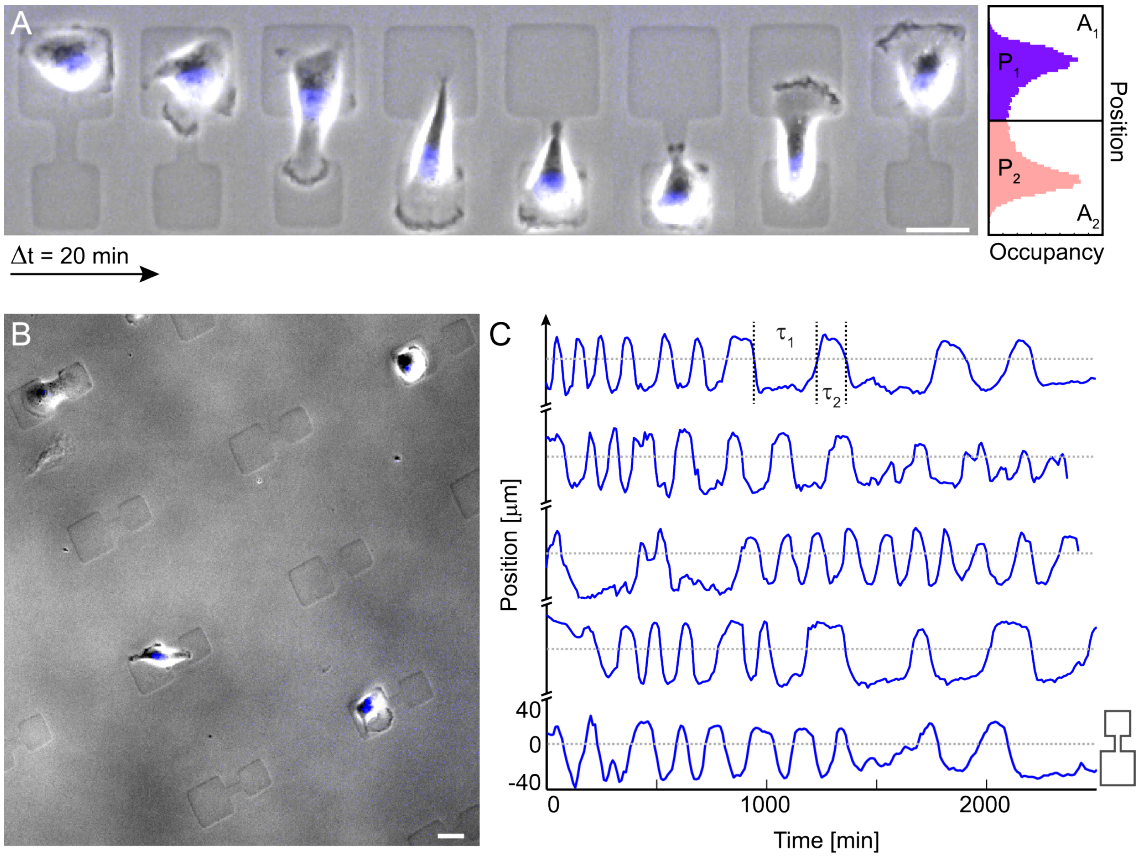


Figure 1. **A:** Time series of a single MDA-MB-231 cell transitioning between two differently sized square adhesion sites. Time intervals between the snapshots are Δt = 20 min. The dynamic exploration of the adhesion sites is visible by the extended lamellipodia seen as dark rims along the cell periphery. On the right: Schematic of the probability distribution of the cell nucleus positions along the long axis of the dumbbell. Scale bar: 25 µm. **B**: Overview over experimental setup. Single cells seeded at low density adhere to asymmetric dumbbell-shaped micropatterns. Cell nuclei are stained for semi-automated cell tracking. Scale bar: 25 µm. **C**: Trajectories, as described by the centre-of-mass of the stained cell nucleus, are plotted for several cells on dumbbell patterns of the shown geometry. The definition of dwell times (τ_1, τ_2) is shown. Dwell times allow for absolute quantification of cell response.

Cells Repeatedly Transition Between Adhesion Sites on Asymmetric Two-State Micropatterns

To study how the shape of microenvironments affects the migration of single cells, we designed artificial two-state micropatterns consisting of two adhesion sites as previously reported (45). In contrast to the previous study, the micropatterns used here consist of two sites with different geometries, which are connected by a thin bridge of a constant length. The surface of the micropatterns is coated with fibronectin, while the surrounding space is

passivated using PLL-PEG. As a result, cells only spread within the fibronectin-coated area and remain effectively confined in the micropatterns. We seed MDA-MB-231 human breast carcinoma cells at a concentration low enough to ensure sparse filling of the surface with an average occupancy of one cell per micropattern (Fig. 1B) (for further details, see Material and Methods). After seeding, the cells settle and spread on one of the two square adhesion sites in each pattern. Due to their intrinsic motility, cells then repeatedly transition between the two connected sites in a stochastic manner (Fig. 1A and 1C and Supplementary Movie M1).

Transitions between adhesion sites are characterised by a typical order of morphological sequences: Cells that are positioned on one of the two sites exhibit dynamic ruffling of the cell contour and form filopodia which primarily explore the cell-repellent regions surrounding the micropatterns (see LifeAct staining and high time-resolution phase-contrast images in Fig. 5, Fig. S1, Supplementary Movies M8, M9 and M10). Interestingly, short-lived unsuccessful attempts to push lamellipodia beyond the boundaries of the micropattern are clearly visible at the periphery of the cell (black regions in phase contrast images Fig. 1A), likely in conjunction with filopodia extension (Supplementary Section 2). On the fibronectin-coated micropatterns, however, cell motion seems to be dominated by lamellipodia. At the entrance to the bridge, stable lamellipodia form that can proceed to grow across the bridge, which acts as a guidance cue for cell transitions towards the other site. Once a lamellipodium reaches the opposite adhesion site, a fan-like broadening of the lamellipodial tip is observed, which is usually followed by the cell body transitioning across the bridge to the other side (Fig. 1A). Retraction fibres can be seen along the cellular track (movie S10). On the opposite site, the cell spreads into the available adhesion site area and the process repeats. All these behaviours have been previously observed on symmetric two-state micropatterns (45).

This cellular hopping process is well captured by the trajectories of the fluorescently-labelled cell nuclei (45). The trajectories span time intervals of up to 50h – their length is limited by cell division or the end of a measurement. Within this time interval, the transition process is stochastic and individual tracks vary from each other. While the cells spend a significant amount of time on the adhesive islands, the transitions themselves occur abruptly (Fig. 1C). These distinct transitions allow an effective reduction of cell positions to two states, separated by the centre of the connecting bridge. A 'stay' is defined as the total amount of time the nucleus spends on one side between two crossings of the middle of the bridge. Each of these stays yields a corresponding dwell time, τ.

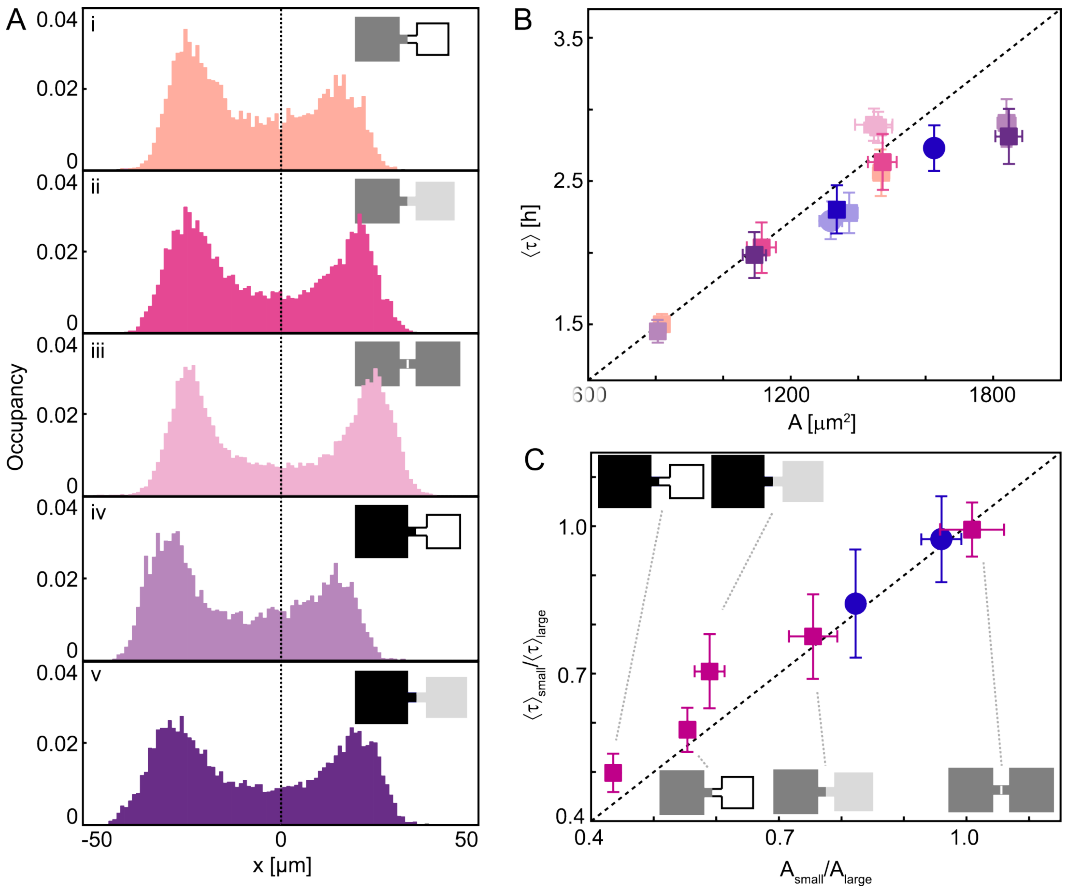


Figure 2. **A**: Occupancies of cells on square adhesion sites with different area. Adhesion sites are depicted in grey shades, same tones denoting the same area. The dotted line represents the middle of the bridge. The occupancies reflect the asymmetry of the underlying micropattern. Third in row (iii) is the symmetric case. (i) and (iii) are adapted from (45). **B**: Mean dwell times, $\langle\tau\rangle$, of cells as a function of square area, A. Dwell times increase with adhesion site area. The dotted line is a guide to the eye. Same colours denote the same two-state geometry, colours correspond to the colours of occupation probability distributions in **A**. Blue circles and squares correspond to dwell times on square-circle micropatterns. Y-Errors are bootstrapping errors, x-errors are standard deviations of weighted area means. **C**: The ratio of mean dwell times plotted against adhesion site area ratio. The dashed line indicates equality. Blue circles correspond to data from square-and circle micropatterns. Y-Errors are bootstrapping errors.

Micropatterns with Unequally Sized Adhesion Sites Exhibit Asymmetric Occupation Probabilities

First, we ask the question how adhesion site area influences the transition statistics. To this end, we perform experiments with square adhesion sites of different edge lengths in the range of 27.2 to 42.3 µm. The connecting bridge has a fixed length of about 16 µm. To characterise cellular response to the imposed geometries we use an easily accessible and intuitive read-out: We determine the occupancy probabilities as a function of position along

the main axis of the micropattern with the centre of the bridge defined as $x = 0$. Here, we include all time-scales of the migration into the data analysis rather than mainly utilising the short timescale dynamics (45). As expected from symmetry arguments, on equally sized adhesion sites, the occupation probabilities are symmetric (Fig. 2A (iii)). In contrast, the probability distributions corresponding to asymmetric two-state patterns are asymmetric with a clear shift towards higher occupation probabilities on larger adhesion sites (Fig. 2A).

To quantify the dependence of the transition dynamics on the adhesion site area, we plot the mean dwell times $\langle\tau\rangle$, averaged over all time traces of the entire ensemble of single cell trajectories, against the corresponding adhesion site areas. We find an almost linear increase of dwell times with indications for a saturation effect for areas larger than approximately 1500 µm^2. It appears that the dwell times are largely determined by the area of the confining adhesion site. Dwell times are also independent of the size of the opposite adhesion area (Fig. 2B). In Fig. 2C, the ratio of mean dwell times is plotted against the ratio of adhesion site areas and compared to the direct proportionality relationship between dwell times and adhesion site area (dotted line).

$$\langle\tau_s\rangle/\langle\tau_l\rangle = A_s/A_l \qquad \text{(Eq. 2)}$$

where $\langle\tau_s\rangle$, $\langle\tau_l\rangle$ denote the mean dwell times and A_s, A_l the area of the small and large adhesion sites, respectively. For area ratios smaller than 0.5, the data slightly deviate from the linear relation. In particular, the two data points with the largest tested adhesion site (A_l ≈ 1780 µm^2) deviate most from linearity (the corresponding dumbbells are shown as inserts with black areas in the left upper part of Fig. 2C). The mean cell area of unconfined cells at the start of a measurement is about 940 ± 37 µm^2 and hence slightly smaller than the majority of adhesion sites. Therefore, cells only cover the entire adhesion site on average, through dynamic exploration as well as after growth (Supplementary Section 3 and Fig. S2). Thus, the linear relation between dwell times and area appears to be a consequence of the confinement imposed by the adhesion sites on cell spreading and motion.

It is also noteworthy that, for the consideration described above, we can either use the mean values of dwell times or occupation probabilities. We determined the occupancy probabilities as a function of position along the long axis of the dumbbell (Fig. 2A). The integral of the occupancy probability densities on the left and right sides yield a left versus right occupation ratio, which is equivalent to the ratio of left/right mean dwell times in the limit of long time traces with an equal number of left and right stays. As further detailed in Material and Methods, in a steady-state situation and for sufficiently large statistics, we can equate without further assumptions:

$$P_{left}/\,P_{right} = \langle\tau_l\rangle/\langle\tau_r\rangle \qquad \text{(Eq. 3)}$$

This relation provides a link between dynamic and static readouts. Also, while occupation probabilities present a relative measure of preferential cell localisation, the mean dwell times are absolute quantities, which provide a time-scale of the transition dynamics. In the following analysis, we will mostly discuss mean dwell times, which may also be considered as reverse escape rates, $\langle\tau\rangle_{l,r} = k_{l,r}^{-1}$, yet without making a statement on the detailed dynamics and nature of the transition mechanism. Taken together, the results so far show that cell occupancies are related to the adhesive area through an approximately linear proportionality relation, and that the adhesion site area can act as a key control parameter for confined cell dynamics.

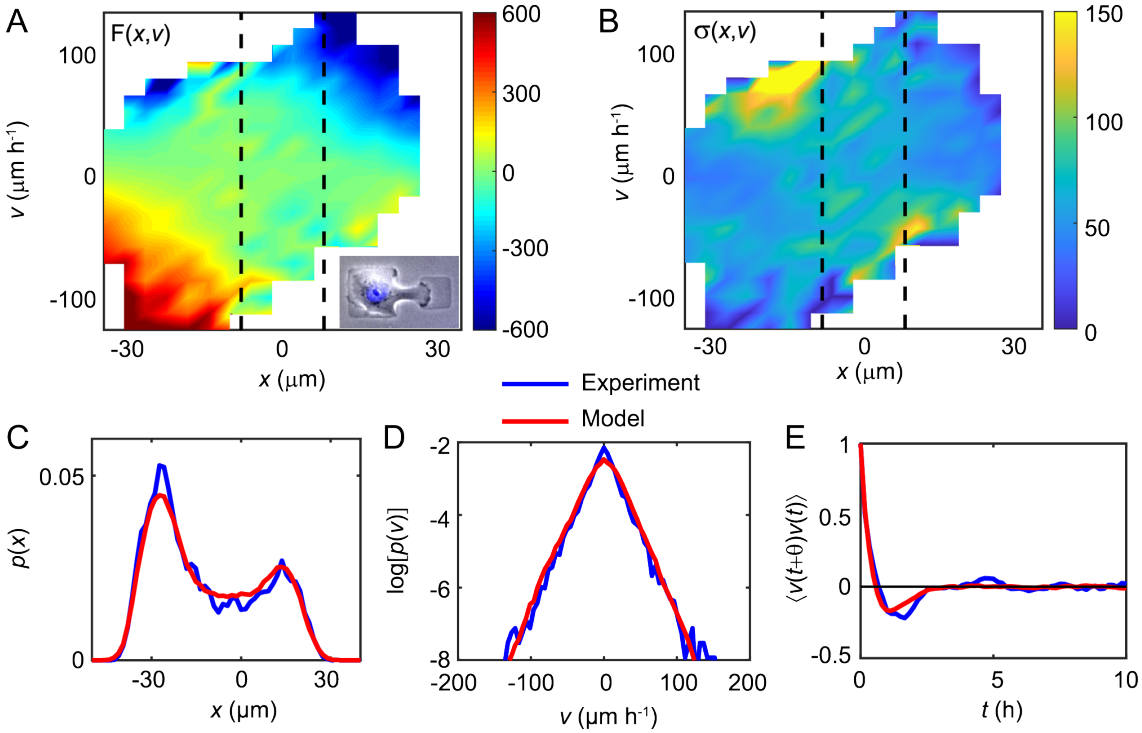


Figure 3. A: The deterministic component $F(x,v)$ in units of μm/h^2 of cell dynamics in a two-state system with square edge lengths 42 μm and 27 μm. **B:** The corresponding noise strength $\sigma(x,v)$ in units of μm/h$^{3/2}$. The dashed black mark both ends of the bridge. Both F and σ are shown here as a linear interpolation. **C, D, E:** Experimental (blue) and predicted (red) probability distributions of position (c), velocity (d, plotted semi-logarithmically) and e, velocity correlation function for all asymmetric two-state micropatterns presented in Fig. 2. For a comparison between model predictions and experimental data for the other two-state system geometries see Supplementary Section S4.

A Stochastic Nonlinear Equation of Motion Captures the Asymmetries

To investigate if we can find a simple quantitative description of the stochastic dynamics of the cell trajectories in this confinement, we infer an equation of motion from the

experimentally recorded dynamics. This provides a connection between the short time-scale dynamics of the cell trajectories and the long time-scale cellular readouts, such as dwell times and occupation asymmetries. In previous work (45), we showed that the dynamics of cells migrating on symmetric two-state micropatterns can be described by an equation of motion (Eq. 1) which disentangles the deterministic and stochastic contributions of the motion. Specifically, the deterministic term *F(x,v)* gives the average acceleration of the cell nucleus when at position *x*, and moving with a velocity *v* (Fig. 3A). Similarly, the noise strength σ(*x*,*v*) indicates the strength of the stochastic fluctuations of the motion (Fig. 3B). Interestingly, we find that in asymmetric patterns, both functions share many qualitative features with those observed for symmetric systems. Importantly, similar to symmetric systems, the deterministic function shows that the cell deterministically accelerates into the thin constriction before transitions (Fig. 3A). Furthermore, we find that the equation of motion approach yields accurate predictions for the experimentally observed statistics, such as occupation probabilities, velocities and velocity correlation functions (Fig. 3 C-E and Fig. S3). These results indicate that the stochastic equation of motion approach is generalizable to asymmetric patterns, highlighting the generic features of cell migration in two-state micropatterns.

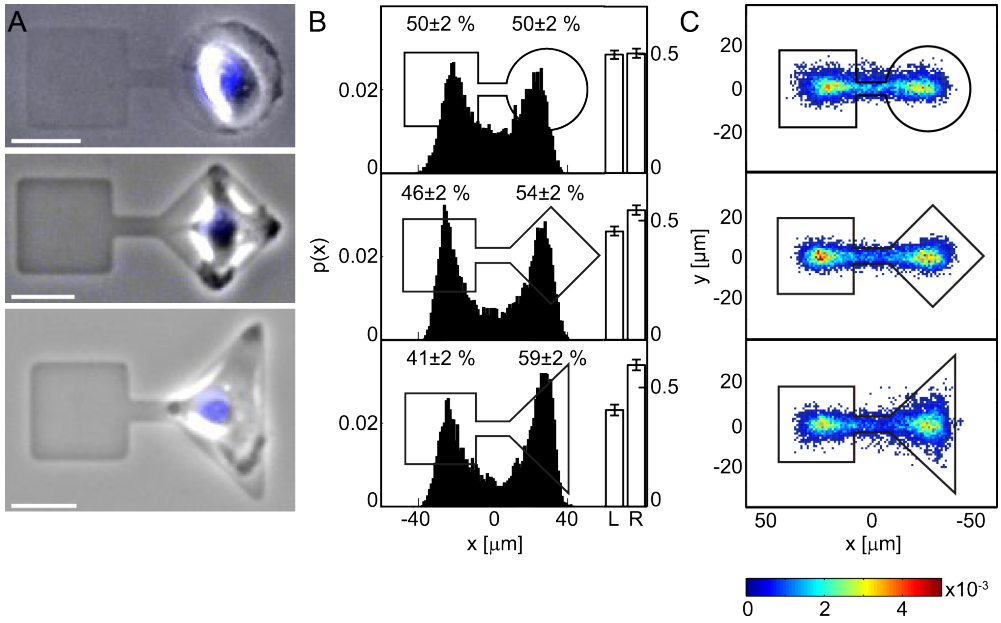


Figure 4. A: Snapshots of single MDA-MB-231 cells migrating on two-state patterns with adhesion sites of different geometries and approximately equal area. **B**: Probability distributions of cell positions along the x-axis of the two-state patterns and the corresponding left-right occupation probabilities normalised by the corresponding adhesion site area. An occupation asymmetry is detectable for the square versus rhombus pattern and clearly visible for square versus triangle. Errors are bootstrapping errors. **C**: Two-dimensional probability maps of cell positions with the corresponding micropatterns drawn to scale.

Anisotropic Adhesion Site Shapes Bias Occupancies

We next investigate how the geometrical shape of adhesion sites affects cell migration. To this end, we designed asymmetric two-state micropatterns that connect a square adhesion site to sites with the same area, but of different shape, including circular, rhombical, and triangular sites (Fig. 4). In all cases, the cells perform repeated stochastic transitions, adapt to the shape of the underlying adhesion site and dynamically explore the available area (Fig. 4A and Supplementary Movies M2-M4). To correct for differences in adhesion site areas introduced in the production process, occupation probabilities are normalised by adhesion site areas for square-circle, square-rhombus and square-triangle micropatterns.

Using circles, we can test the dependence of dwell times on area or perimeter while keeping the other respective parameter constant. On square-circle micropatterns with equal areas, we observe a symmetric occupation probability distribution (Fig. 4B), and both types of square-circle patterns follow the general trend (blue circles in Fig. 2B, 2C, Fig. S4 and Supplementary Section 5). To relate this finding to the underlying cytoskeletal dynamics driving the migration, we observe LifeAct-GFP transfected cells. Visible as actin hotspots along the cell periphery, a dynamic exploratory motion of the cell within the adhesion sites is apparent. While the distribution of actin appears to be homogenous along the periphery of circular adhesive sites, on square adhesion sites, protrusions form preferentially at the corners. However, overall, very few differences are visible in the actin dynamics, in the type and localisation of actin fibres and locations of protrusion activity between cells adhering to the square and circular adhesion sites (Fig. 5 (i) and (ii) and Supplementary Movie M8). Cells migrating onto the adhesion sites and exploring them adopt a variety of shapes. Where the adhesive shapes impose their shape onto the cells, differences in actin and paxillin localisation in fixed cells become apparent. Thus, on squares, paxillin is mainly accumulated in the square corners, whereas on circles, focal adhesions are more evenly distributed along the cell edge. Rhombical adhesion sites, specifically squares rotated by 45°, allow us to investigate the question whether the corner facing the bridge entrance is able to bias the cell occupancies towards the square site as suggested by previous research (38, 54, 55). Interestingly, the occupation probabilities on square-rhombus patterns show a small bias towards stays on the rhombus (Fig. 4B).

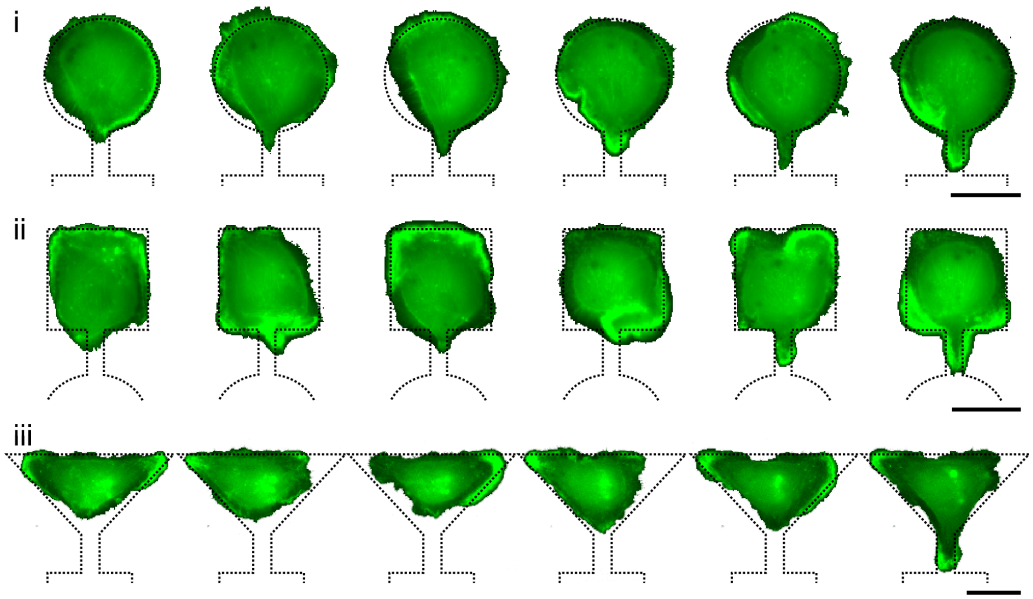


Figure 5. Typical morphological sequences of LifeAct-GFP stained single cells for the asymmetric dumbbells shown in Fig. 4. The distribution of protrusions is seen as bright spots of localized actin along the cell periphery. Actin activity appears more pronounced at adhesion sites' corners. For improved visibility, cell outlines were selected with the wand tool in Photoshop, and the background colours inverted. Pattern outlines are drawn up to scale. Scale bars: 25 µm.

To challenge this unexpected finding, we designed dumbbells consisting of a square paired with a right-angled triangle. Two-state systems of this kind show a bias towards increased dwell times on the triangular site (Fig. 4B). This finding coincides with the observation that the 2D occupancy distribution on the triangle exhibits a larger spread in the vertical direction, while the spatial distributions on the square, circular and rhombical shapes are very similar (Fig. 4C). In addition, cells on the triangular shape exhibit pronounced protrusion formation at the triangle's corners (Fig. 4A, Fig. 5 (iii), and Supplementary Movie M9).

Hence, we recapitulate that for sites that are isotropic in the x- and y-directions (square and circle), only adhesive area and not the geometrical shape determines the relative cell occupancy. In contrast, we found that cells take a longer time to escape from rhombical and triangular patterns. Clearly, the triangle has a lower symmetry than the square and circular structures and the protrusions appear to be biased in the vertical direction towards the pointed corners. Also, as this triangle's base extends considerably in the y-direction, cell motion is slightly more pronounced orthogonally to the direction of transitions, thus leading to longer dwell times (Fig. 4C (ii) and Fig. S6). The rhombus seems to represent an intermediate case, as only a small bias is observed (see also Fig. S5) and the angular distribution of cell velocities looks very similar to that on square adhesion sites (Fig. S6).

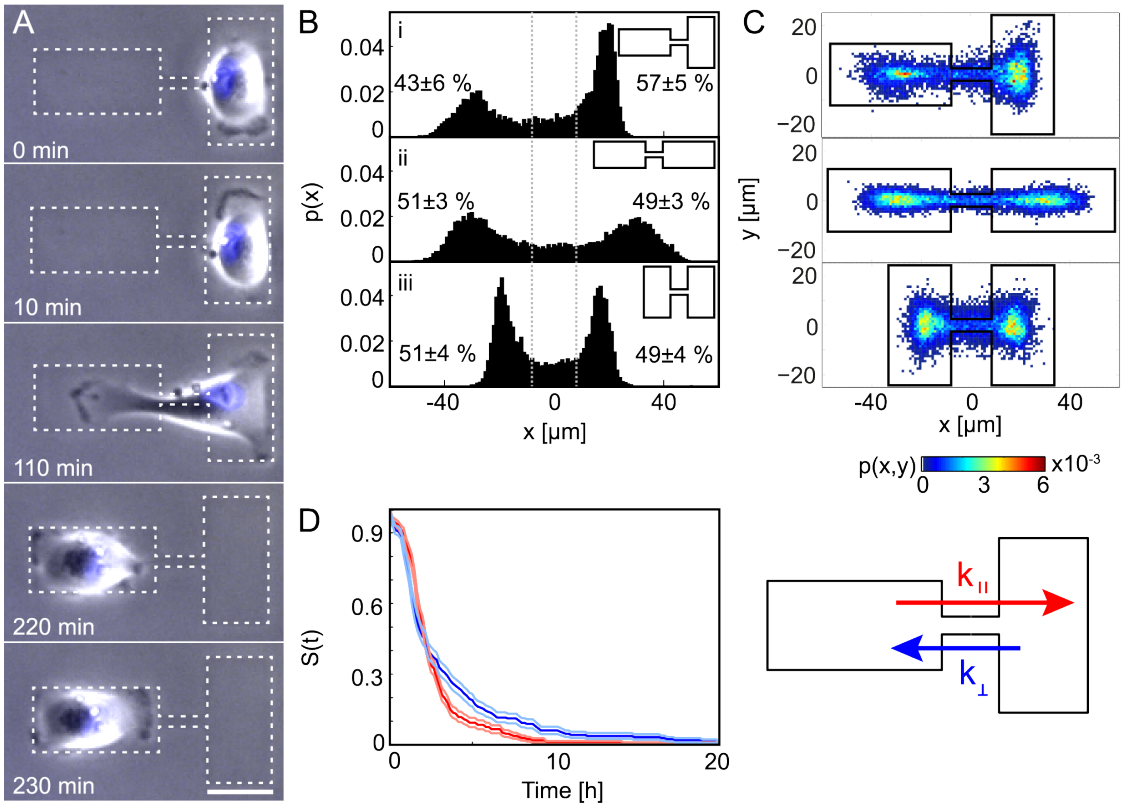


Figure 6. **A**: Time series of a single cell on a rectangular micropattern of orthogonal orientation. The cell preferentially polarizes along the long axes of the rectangles as seen by the formation of lamellipodia in these directions. Scale bar: 25 µm. **B**: The cell probability distribution and left-right occupancies on two-state patterns with equally sized but differently oriented rectangles. The distribution is biased in the case of orthogonal orientation (i). Errors given are bootstrapping errors. **C**: Two-dimensional probability maps of cell positions. A clear difference in the extent of localisation along the y-axis is visible between horizontal and vertical rectangles. **D**: Survival probability functions S(t) of stays in the horizontal and vertical site respectively as indicated in the schematic drawing. Errors are shown as lightly coloured lines. A significant difference in decay behaviour of S(t) is visible for times > 4h.

Different Orientations of Rectangular Adhesion Sites Result in Biased Occupancies

The observed asymmetry on rhombical and triangular adhesive sites leads to the question of whether adhesion site orientation can determine dwell times. To resolve this question, we designed patterns of equal shape and area but of different orientation. These systems consist of rectangular sites with an aspect ratio of 2:1 in three arrangements: two horizontal rectangles, i.e. with their long axes parallel to the direction of the connecting bridge; two vertical rectangles, i.e. with their long axes perpendicular to the connecting bridge; and a mixed, asymmetric configuration with a horizontal and a vertical rectangle. For the two

symmetric arrangements of the rectangular adhesion sites, the observed occupancies are symmetric as expected (Fig. 6B (ii) and (iii)). In contrast, in the case of the asymmetric combination, we find a significant bias towards stays on the upright site (Fig. 6B (i)), supporting our previous assumption that adhesion site anisotropy biases occupancies.

The observed occupancy asymmetries on the mixed-configuration micropattern can be explained in terms of the steady-state equation Eq. 3, which relates the relative occupancies to the dwell time ratio or, in other words, the ratio of the dynamical escape rates from a horizontal and a vertical state. We observe that on rectangular adhesion sites cells tend to polarize along the long axis of the rectangle, which is visible as pronounced lamellipodia activity in the direction of the rectangle's long axis indicated by the black regions in brightfield images (Fig. 6A, 0-10 minutes and 220-230 minutes, Fig. S8 in the Supporting Material and Supplementary Movies M5-M7), biases the direction of cell motion on the adhesion sites and is also confirmed by the orientation of stress fibres in fixed cells (Fig. 7 and Fig. S12). Thus, one may expect that the escape rate along the long axis is faster than along the short axis, resulting in a biased steady-state distribution.

The relation of cell polarization and transition rates is also demonstrated in a more detailed analysis of the two-dimensional cell trajectories. Specifically, we observe a more pronounced motion in the y-direction on vertical rectangles than on horizontal rectangles (Fig. 6C, Fig. S9, Fig. S11A) which occurs more frequently than on squares or circles (Fig. S6). The distribution of velocities in the y-direction is broader on vertical states, indicating that the cells polarize and accelerate along the long axis of the rectangle, orthogonal to the axis of the micropattern (see also Supplementary Section 8, Fig. S10). Moreover, we find that whenever there is pronounced displacement in the y-direction, the extent of motion in the x-direction decreases (Fig. S11B). This suggests that the cell polarization in one direction suppresses motion in the orthogonal direction. Thus, escape rates from vertical rectangles are reduced by the induced polarization orthogonal to the direction of the transitions. For a more detailed discussion of escape times on rectangular dumbbell micropatterns we refer to the Supplementary Section 8.

The full dynamical dependence of the escape process is best represented by the survival probability distribution S(t) which gives the probability that a transition has *not* occurred after time *t* (Fig. 6D). Initially, S(t) for upright rectangles (blue curve) decays faster than S(t) for horizontal rectangles (red curve). However, for stay times longer than 4h, the trend reverses and the blue curve decays significantly slower. This suggests that when transitioning onto the upright rectangle, cells either quickly repolarise or become trapped. The former is possible when cells are reflected by the edge of the micropattern and thus have to turn by 180°. Trapping happens when lamellipodia form along the long axis of the rectangle, making the

formation of a lamellipodium along the bridge, a pre-requisite for escape, less likely because the cell is polarized orthogonally to the bridge. In summary, we found that the orientation of anisotropic adhesion sites can bias cell occupancies since migrating cells tend to polarise along the long axis of these sites, leading to a reduced escape rate when this axis is orthogonal to the transition axis.

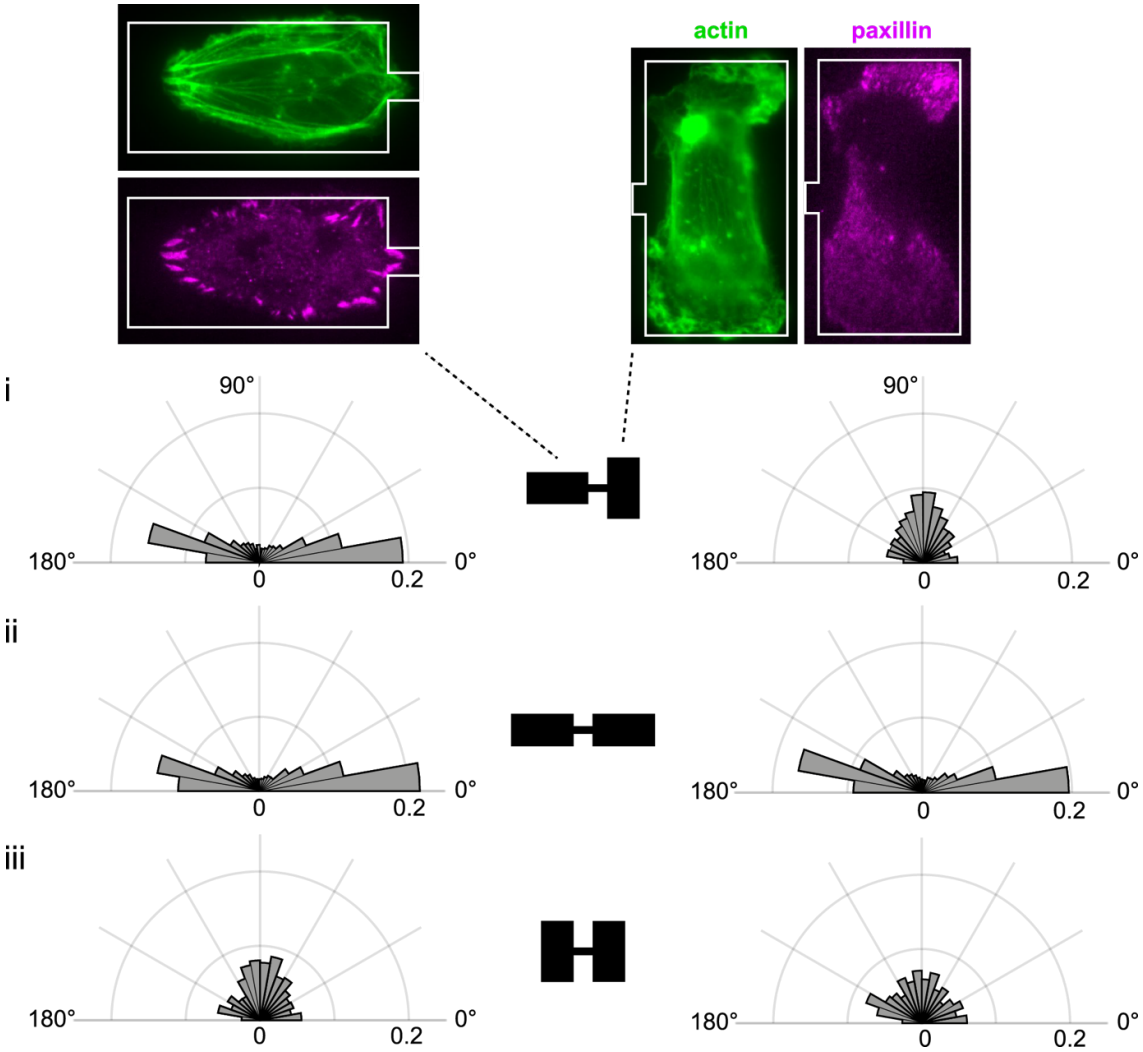


Figure 7. Images of fixed cells on mixed-orientation rectangle micropatterns stained for actin and paxillin (for (ii) and (iii) see Fig. S12). Actin fibres are mostly oriented along the long axis of the rectangular adhesion sites, indicating cell polarisation along that axis. Below, angular probability distributions of the directions of cell velocities on rectangular adhesion sites are plotted. The corresponding micropatterns are shown in the middle. The directions are stated with respect to the long axis of the micropattern (i.e. the axis of transitions). Each bin contains the counts for the shown angle and for motion in angle+180°. Only cell motion within adhesion sites is included in the analysis.

Discussion

In this work, we studied cell migration on artificial two-state micropatterns, which allow for a statistical analysis of the relative preference of cells for two opposing adhesion sites with different geometric properties. Occupancies are described in terms of relative occupation probabilities or absolute mean dwell times. We observed a preferred occupancy with increasing adhesion site area in the case of two square adhesion sites. The finding is in agreement with previously reported 'dimension sensing' of cells on thin 1D lines interspersed with wider rectangles (36). Larger adhesive areas induce longer dwell times, which is in qualitative agreement with the observation that freely moving cells exhibit reduced cell motility with larger spreading area (56, 57). Also, recently, Guo et al. reported that a large cell adhesion area is linked to smaller cell volume and an increase in cell stiffness due to water efflux in a spread out state (58). That cell stiffening corresponds to cell spreading has also been reported for endothelial cells (47). While in a confluent cell layer, MDCK II cells follow the same trend. Interestingly, no linear dependence of cell size and stiffness has been found for individual cells on micropatterns (59). A stiffer cell cortex is likely associated with slower migration (60). Thus, a potential explanation for the area dependence of escape dynamics (Fig. 2B and 2C) might be a transient stiffening of the cell cortex on sites with larger adhesion area. Also, larger adhesive areas provide more adhesion ligands, and it was observed that high levels of cell-substrate adhesiveness reduce cell migration speed (61). Furthermore, we found that a stochastic equation of motion framework, developed in previous work for symmetric two-state patterns (45), can be generalised to quantitatively capture the statistics of cell motion also in asymmetric systems. Interestingly, the qualitative features of the deterministic and stochastic contributions to the cell dynamics exhibit similar features to symmetric systems, indicating that these dynamics include generic features insensitive to the sizes of the adhesive islands.

We found only a weak dependence of occupation probabilities on shape in sites that are symmetric under 90 degree rotations. In particular, cells on circular, square and rhombical adhesion sites exhibit very similar transition behaviour (Fig. 2B, 2C, Supplementary Section 6, Fig. S5, Fig. S6). This is remarkable since some of these sites have different perimeters. The cell perimeter is a parameter that frequently enters elastic cell models which assume a global line tension (62-64). While previous studies suggest a difference in the localized distribution of cellular protrusions on square and circular micropatterns (28, 65), for the geometries we tested, we did not observe any influence of square or circular shape on occupancies, which should be sensitive to the likelihood of directional protrusions. This finding, however, is consistent with the reported lack of cell polarity on square and circle micropatterns (28). While cells may transiently polarise even on squares and circles, the absence of induced cell polarisation, defined as pre-requisite for cell displacement, is also demonstrated in experiments where cells showed no preferred direction of motion when released from

squares or circular micropatterns by removal of a micropatterned frame (29). Mechanical parameters like stiffness and contractility were previously found to be the same for endothelial cells patterned on squares and circles (47).

The role of acute corners and stress fibres on the transition dynamics needs further inspection. It was reported that stress fibres running along straight edges impede protrusion formation, while stress fibres meeting in corners are easier to penetrate (38, 54, 55). Intriguingly, in the presence of other guidance cues such as corners located orthogonally to the axis of transitions, we did not observe increased transition rates where corners were pointed towards the connecting bridge. However, this is in line with by Caballero et al. (42, 43) reporting more protrusions forming on the base of the triangle. Specifically, the presence of acute angles in triangles seems to stretch and orient cells in a direction perpendicular to the transition axis (Supplementary Section S7 and Fig. S6), a phenomenon which has been previously suggested to increase friction during cell migration (43). This could also be related to the increased likelihood of cells to extend lamellipodia from acute angles (30).

Dwell times on triangles and on vertical rectangles deviate strongest from the linear area dependence observed for more isotropic adhesion sites (Fig. S13 and Supplementary Section S9). Specifically, in the rectangular system, the importance of pattern orientation became clear where a directed formation of lamellipodia orthogonal to the direction of transitions between the adhesion sites is achieved, and therefore the motion becomes anisotropic (28, 29, 40). Cells appear to be polarised along the long axes of the rectangles as evidenced by the mean orientation of stress fibres in fixed cells (Fig. 7 and Fig. S12). The orientation of long stress fibres and cell elongation are considered an indicator for cell polarity (35, 66). Actual quantification of stress fibre orientation is possible in (semi-)automated ways (67-69). The rectangles act in the same way as one-dimensional stripe patterns: the persistence of cells in the direction of the stripes is increased (37) and polarity can be established (35, 70).

Interestingly, the escape rates in rectangular systems in particular cannot be explained by geometrical considerations with respect to the ratio of the bridge width and the adhesion site edge length (Fig. S14). Another intriguing feature of the rectangle system is that the escape rates from the vertical rectangle in the mixed system differ from those in the symmetric setup. While internal cell organisation in fixed cells (Fig. 7 and Fig. S12) and the distributions of cell velocities in equally oriented rectangles look similar, the difference in escape rates could indicate the presence of memory effects. Such memory effects have been reported previously for migration on topological ratchet patterns (71), cell spreading on fibronectin micropatterns (64) and cell migration in the presence of chemoattractants (72). For all other micropatterns, we do not observe similar effects as dwell times for adhesion sites of similar areas but from different two-state systems are equal within errors (Fig. 2B, Fig. S13).

Conclusion

In summary, we created asymmetric two-state micropatterns, which allow for a quantitative assessment of geometric determinants in confined cell migration. The comparison of mean dwell times and steady-state occupancy probabilities in different adhesive geometries identifies relevant factors, such as adhesion site area and geometry, and quantifies their effect on cell migration. The statistics of individual trajectories and transition rates also suggest that escape rates could be an indirect measure of protrusion dynamics and cell polarisation in confining microenvironments. In general, single-cell microarrays enable a highly parallel assessment of individual trajectories of an ensemble of cells in defined boundaries. This is all the more important as cellular heterogeneity is understood to be an intrinsic property of cell cultures (73, 74). Two-state microarrays could also be extended to probe cellular affinities for protein-coated surfaces by creating patterns with different chemical functionalities. In addition, as the two-state micropatterns allow for repeated observation of transition events of individual cells, the leading-edge dynamics could be studied under defined boundary conditions and with high statistical certitude. The forced transitions in artificial microenvironments could also, in future work, be related to relevant, disease-related states of cells like deformability (75) and, for example, serve to characterize migratory phenotypes such as invasiveness (76). Hence, two-state micropatterns represent a unique single-cell migration assay with the capacity to assess relative preferences of cells for microenvironments and to define novel metrics for cell-migration phenotypes.

Materials and Methods

Micropatterning and Sample Preparation

We use microscale plasma-induced protein patterning (46). Briefly, silicon masters with the desired shapes are prepared using photolithography. Polydimethylsiloxane (PDMS) monomer and crosslinker (DC 184 elastomer kit, Dow Corning, Midland, Michigan) are mixed in a 10:1 ratio and subsequently poured onto the silicon wafer. After degassing the PDMS, it is cured overnight at 50 °C.

PDMS stamps are placed, with the features facing down, in an ibidi µ-dish (ibidi GmbH, Martinsried, Germany): The dish with the stamps is then exposed to oxygen plasma. For background passivation, a drop of 2 mg ml^{-1} PLL(20)-g[3.5]-PEG(2) (SuSoS AG, Dübendorf, Switzerland) solution is added and left to incubate for 25 minutes. After rinsing the sample, the stamps are removed, and the sample is incubated with a 50 µg ml^{-1} human fibronectin (YO Proteins, Ronninge, Sweden) solution for 50 minutes. After final rinsing, samples are stored in PBS at 4°C.

Micropattern Design

The two-state micropatterns have square adhesion sites of edge lengths between 27.3 ± 0.4 µm to 42.2 ± 0.5 µm. Adhesive area dimensions were chosen such as to provide enough adhesive area for cells to be able to fully occupy them. A limit was found for adhesion sites smaller than approximately

25 µm, where cells would still transition between adhesion sites but partially remain on the bridge and the larger adhesion site. No adhesion sites with edge lengths larger than 42.2 µm were used.

Square-circle micropatterns were designed to have the circular sites either probe a same-area or a same-perimeter case with respect to the square adhesion sites. Micropatterns with triangles are designed to offer adhesion sites of equal areas and also to both have a right angle pointing towards the bridge to test the influence of shape while keeping the funnelling angle constant. Rectangular adhesion sites have equal areas and were designed to offer similar adhesion site areas to the symmetric square-square setup. The aspect ratio of rectangles was chosen to be 2:1, with the small side to be longer than the minimum edge length determined for squares.

For all geometries, a similar bridge length was used. The length of the bridge was chosen such that cells would transition frequently between adhesion sites to ensure sufficient statistics while being separated far enough to ensure cells mainly occupied a single adhesion site at any time (see exclusion criterion 5, Supplementary Section 1.2.). The influence of bridge length was also tested for rectangular patterns (see Supplementary Figure S7). Details of that and exact measures for all patterns, as well as cell statistics, can be found in the Supplementary Sections 1 and 8.

Errors given for the pattern dimensions are weighted standard deviations to account for the variation in statistics gained in different experiments. The dimensions of final protein patterns are subject to experiment-to-experiment and wafer-to-wafer variations. Experiment-to-experiment variability is mainly due to the intrinsic variance of the manual stamping process. Also, measurement uncertainty is added due to the limited resolution of images.

Cell Culture

MDA-MB-231 human breast carcinoma epithelial cells (DSMZ, Braunschweig, Germany) are cultured in Minimum Essential Medium (MEM, c.c. pro, Oberdorla, Germany) with 10% FBS (Gibco, Paisley, United Kingdom) and 2mM L-Glutamine (c.c. pro). Cells are grown in an atmosphere with 5% CO_2 and at 37°C up to 70-90% confluence. For splitting, cells are rinsed with PBS and subsequently trypsinised for 3 minutes. For experiments, the trypsinised cell solution is centrifuged at 1000 rcf for 3 minutes, the cell pellet is resuspended in MEM and approximately 10,000 cells are seeded per µ-dish. Cells are left to adhere for 4h in the incubator, before the medium is exchanged to L-15 medium (containing L-Glutamine (Gibco, Grand Island, New York), and supplemented with 10% FCS). The L-15 also contains 25 nM Hoechst 33342 (Invitrogen, Eugene, Oregon) for the staining of cell nuclei.

Microscopy and Cell Tracking

Measurements of up to 50h are performed in time-lapse mode, either on an IMIC digital microscope (TILL Photonics, Kaufbeuren, Germany) or on a Nikon Eclipse Ti microscope using a 10x objective. To keep samples at 37°C throughout the measurement, dishes are placed in a heated chamber (ibidi GmbH or Okolab, Pazzuoli, Italy). Brightfield and DAPI-images are acquired every 10 minutes. It is interesting to note, that under these experimental conditions we observe that the doubling time of cells is slightly larger than in cell culture. The influence of micropattern size and geometry on cell proliferation has been reported before (24, 47, 48).

For analysis, a band pass filter is applied to the fluorescence images and a threshold is used to binarise the images using ImageJ (49). This allows subsequent automated tracking of the centre-of-mass of the cell nuclei by ImageJ's Analyze Particles plugin. The two-state patterns are visible in the brightfield images and are used to manually determine the reference boundary of a pattern, defined as edge of the left adhesive island, for each single cell. Using this value it is possible to convert nuclear coordinates to absolute positions. All further analysis is performed in MATLAB (The MathWorks, Natick, Massachusetts).

For every geometry, at least 46 cells, chosen from a minimum of three experiments, were analysed. The criteria for the determination of suitable cells, as well as the number of analysed cells per geometry, are detailed in Supplementary Section S1. Specifically, to probe the response to each adhesion site separately, we only included cells which, in most cases, made complete transitions between adhesion sites. A complete transition is characterised by all parts of the cell being confined to one adhesion site before the next transition is initiated. Also, to limit abnormal migration behaviour, we excluded all cells which underwent cell death at any time during the experiment.

Cell Fixation and Immunostaining

Cells are cultured and seeded as described above. 16-24h after seeding, cells are rinsed once with warm PBS and fixed in warm 3.8% formaldehyde in PBS (Sigma, Saint Louis, Missouri) for 15 minutes at room temperature. Cells are permeabilised with 0.5% Triton X-100 (Roth, Karlsruhe, Germany) for 5 minutes and then incubated in a blocking solution consisting of 10% normal goat serum, 0.2% Triton X-100 and PBS. The primary antibody (5 µg/ml anti-paxillin, Invitrogen, Rockford, Illinois) is added in the same blocking solution for 1h at room temperature. After washing, 5 µg/ml of the secondary antibody (goat anti-mouse conjugated with Alexa Fluor 488, Abcam, Cambridge, United Kingdom) and 100 nM rhodamine phalloidin (Molecular Probes, Eugene, Oregon) are added and left to incubate for 30 minutes at room temperature. The cells are then washed, 0.5 µg/ml DAPI (Sigma) is added for 5 minutes and after a last wash, cells are stored in PBS for imaging on the same day. Imaging is performed on the Nikon Ti Eclipse microscope, using a 60x oil-immersion objective in both, epifluorescence and TIRF mode.

LifeAct-GFP Transfection

Approximately 12,500 MDA-MB-231 cells are seeded in patterned µ-dishes and left to adhere overnight. Cells are cultured in MEM including Glutamax (Gibco, Paisley, United Kingdom) supplemented with 10% FCS.

For life-cell actin imaging, 500ng LifeAct-GFP mRNA (in-house prepared) is resuspended in OptiMEM (Gibco, Grand Island, New York) to a final volume of 150 µl. In parallel, a mix of 1.25 µl Lipofectamine 2000 (Invitrogen, Carlsbad, California) and 123.75 µl OptiMEM is prepared. Then, the mRNA solution is added to the Lipofectamine-mix, and left to incubate for 20 minutes at room temperature. Before adding the transfection mix, cells are rinsed once with PBS. Cells are incubated with the transfection mix for at least 3h, before the mix is replaced by L-15 medium. Cells are imaged every 10 minutes on the Nikon Ti Eclipse microscope using a 60x oil-immersion objective.

Dwell Times and Occupation Probabilities

For dwell time calculation, the trajectories of the nuclei are binarised into two states, left and right of the middle of the connecting bridge. Furthermore, in order to avoid effects of a finite observation time window, all analysis is performed on cropped trajectories, starting from the time point of the first transition and ending after the last fully observed dwell time. The middle of the bridge is determined by adding the sum of the mean left adhesion site edge length and half of the mean bridge length for each experiment to the individually determined left border reference points. The time spent in either state in between transitions over the middle of the bridge, is the dwell time τ.

To assess the relative occupation probabilities, we calculate the sum of dwell times, τ_i, on adhesion site i divided by the total observation time, T_{tot} of all trajectories, which can also be expressed by the number of stays N_i and the mean stay time, $\langle\tau_i\rangle$, divided by total time:

$$p_i = \frac{\sum \tau_i}{T_{tot}} = \frac{N_i\langle\tau_i\rangle}{T_{tot}}$$

The ratio of the observed left/right mean dwell times $\langle\tau_{left}\rangle/\langle\tau_{right}\rangle$ thus converges to the ratio of probabilities p_{left}/p_{right}, if the number of total stays is large, and thus $N_{left} \approx N_{right}$.

To correct for unequal adhesion site areas in the cases where adhesion sites were designed to have the same area but were found to deviate from the expected areas, we normalise the occupation probabilities by adhesion site areas A:

$$p_1 = \frac{\frac{\sum \tau_1}{A_1}}{\frac{\sum \tau_1}{A_1} + \frac{\sum \tau_2}{A_2}}$$

Survival Probability Functions

The survival probability distribution S(t) is defined as

$$S(t) = 1 - \int_0^t p(\tau)\mathrm{d}\tau$$

where p(τ) is the probability distribution of dwell times for all cells evaluated on a certain adhesion site.

Theoretical Analysis

To obtain a simple equation of motion model for the system, the experimental trajectories of the nucleus position $x(t)$, recorded at a time interval $\Delta t = 10$ min, are analysed using a data-driven inference method. Velocity v and acceleration a are directly calculated as numerical derivatives. We postulate that the dynamics can be described by a stochastic differential equation of the form

$$\frac{dv}{dt} = F(x,v) + \sigma(x,v)\eta(t) \qquad \text{(Eq. 1)}$$

Here, $\eta(t)$ represents a Gaussian white noise process with zero mean and no correlation. The terms $F(x,v)$ and $\sigma(x,v)$ are inferred by conditional averaging of the experimental trajectories (50-52). Since the noise term averages to zero, $\langle\sigma(x,v)\eta(t)\rangle=0$, the deterministic term is given by $F(x,v)=\langle a|x,v\rangle$. The

strength of the noise is then estimated using $\sigma^2(x,v)=\Delta t\langle[a-F(x,v)]^2|x,v\rangle$. To test the model predictions, the equation is integrated numerically with a time step Δt equal to the experimental sampling interval. For more details, see (45).

Direction of Cell Motion

To analyse the direction of cell motion cellular trajectories are analysed during each stay: Once the centre of the cell nucleus has crossed the inner edge of an adhesion site, and until it does so again to leave the adhesion site, the difference in x and y-location between two subsequent time points (Δt = 10 min) is calculated. For each velocity vector, the angle with the x-axis is determined using the following relation:

$$\varphi = \cos^{-1}\left(\frac{\vec{v}\cdot\vec{x}}{|\vec{v}||\vec{x}|}\right)$$

with $\vec{v}$ the velocity vector, $\vec{x}$ the unit vector in x-direction.

Error Analysis

We employ a bootstrapping procedure to estimate errors (45, 53). We generate a large number of realisations of a given dataset $D = \{X_1, \ldots, X_N\}$ with N entries by randomly sampling the dataset's entries with replacement. The mean $\langle X\rangle_{\text{realisation}}$ is calculated for each realisation. The standard deviation of all $\langle X\rangle_{\text{realisation}}$ gives the estimated error in the mean.

The same procedure is carried out for small groups of subsequent entries of the dataset to account for correlations. For each window size, 50000 realisations are generated. We limit the window size to a maximum value of 60. The maximum value of the standard deviation calculated for each window size is our final bootstrap error.

Acknowledgments

Financial support of the German Science Foundation (DFG) for the collaborative research centre SFB 1032 project B01 and B12 is acknowledged. D.B.B. is supported by a DFG fellowship within the Graduate School of Quantitative Biosciences Munich and by the Joachim Herz Stiftung. We thank C. Leu for the preparation of wafers, A. Reiser for providing the transfection protocol, G. Schwake for preparation of the LifeAct-mRNA, S. Reinhardt for measuring cell areas and E. Petrov, A. Heuer-Jungemann and T. Neuß for helpful discussions.

Author contributions

A.F., C.S., P.J.F.R. and J.O.R designed experiments; A.F. performed experiments; A.F. and D.B.B. analyzed data. A.F., D.B.B., C.S., C.P.B. and J.O.R. interpreted the experiments and wrote the manuscript.

References

Reference (77) appears in the Supporting Material.

1. Snyderman, R., and E. Goetzl. 1981. Molecular and cellular mechanisms of leukocyte chemotaxis. Science 213:830-837.
2. Lo, C.-M., H.-B. Wang, M. Dembo, and Y.-l. Wang. 2000. Cell Movement Is Guided by the Rigidity of the Substrate. Biophysical Journal 79:144-152.
3. Friedl, P., and D. Gilmour. 2009. Collective cell migration in morphogenesis, regeneration and cancer. Nature Reviews Molecular Cell Biology 10:445.
4. Charras, G., and E. Sahai. 2014. Physical influences of the extracellular environment on cell migration. Nature Reviews Molecular Cell Biology 15:813.
5. Dietrich, M., H. Le Roy, D. B. Brückner, H. Engelke, R. Zantl, J. O. Rädler, and C. P. Broedersz. 2018. Guiding 3D cell migration in deformed synthetic hydrogel microstructures. Soft Matter 14:2816-2826.
6. Gunawan, R. C., J. Silvestre, H. R. Gaskins, P. J. A. Kenis, and D. E. Leckband. 2006. Cell Migration and Polarity on Microfabricated Gradients of Extracellular Matrix Proteins. Langmuir 22:4250-4258.
7. Gunawan, R. C., E. R. Choban, J. E. Conour, J. Silvestre, L. B. Schook, H. R. Gaskins, D. E. Leckband, and P. J. A. Kenis. 2005. Regiospecific Control of Protein Expression in Cells Cultured on Two-Component Counter Gradients of Extracellular Matrix Proteins. Langmuir 21:3061-3068.
8. Arnold, M., V. C. Hirschfeld-Warneken, T. Lohmüller, P. Heil, J. Blümmel, E. A. Cavalcanti-Adam, M. López-García, P. Walther, H. Kessler, B. Geiger, and J. P. Spatz. 2008. Induction of Cell Polarization and Migration by a Gradient of Nanoscale Variations in Adhesive Ligand Spacing. Nano Letters 8:2063-2069.
9. Arnold, M., E. A. Cavalcanti-Adam, R. Glass, J. Blümmel, W. Eck, M. Kantlehner, H. Kessler, and J. P. Spatz. 2004. Activation of Integrin Function by Nanopatterned Adhesive Interfaces. ChemPhysChem 5:383-388.
10. Cavalcanti-Adam, E. A., T. Volberg, A. Micoulet, H. Kessler, B. Geiger, and J. P. Spatz. 2007. Cell Spreading and Focal Adhesion Dynamics Are Regulated by Spacing of Integrin Ligands. Biophysical Journal 92:2964-2974.
11. Deeg, J. A., I. Louban, D. Aydin, C. Selhuber-Unkel, H. Kessler, and J. P. Spatz. 2011. Impact of Local versus Global Ligand Density on Cellular Adhesion. Nano Letters 11:1469-1476.
12. Cavalcanti-Adam, E. A., A. Micoulet, J. Blümmel, J. Auernheimer, H. Kessler, and J. P. Spatz. 2006. Lateral spacing of integrin ligands influences cell spreading and focal adhesion assembly. European Journal of Cell Biology 85:219-224.
13. Huang, J., S. V. Gräter, F. Corbellini, S. Rinck, E. Bock, R. Kemkemer, H. Kessler, J. Ding, and J. P. Spatz. 2009. Impact of Order and Disorder in RGD Nanopatterns on Cell Adhesion. Nano Letters 9:1111-1116.
14. Schvartzman, M., M. Palma, J. Sable, J. Abramson, X. Hu, M. P. Sheetz, and S. J. Wind. 2011. Nanolithographic Control of the Spatial Organization of Cellular Adhesion Receptors at the Single-Molecule Level. Nano Letters 11:1306-1312.

15. Arnold, M., M. Schwieder, J. Blümmel, E. A. Cavalcanti-Adam, M. López-Garcia, H. Kessler, B. Geiger, and J. P. Spatz. 2009. Cell interactions with hierarchically structured nano-patterned adhesive surfaces. Soft Matter 5:72-77.
16. Vielmetter, J., B. Stolze, F. Bonhoeffer, and C. A. O. Stuermer. 1990. In vitro assay to test differential substrate affinities of growing axons and migratory cells. Experimental Brain Research 81:283-287.
17. Evans, A. R., S. Euteneuer, E. Chavez, L. M. Mullen, E. E. Hui, S. N. Bhatia, and A. F. Ryan. 2007. Laminin and fibronectin modulate inner ear spiral ganglion neurite outgrowth in an in vitro alternate choice assay. Developmental Neurobiology 67:1721-1730.
18. Gundersen, R. W. 1987. Response of sensory neurites and growth cones to patterned substrata of laminin and fibronectin in vitro. Developmental Biology 121:423-431.
19. Yamada, A., M. Vignes, C. Bureau, A. Mamane, B. Venzac, S. Descroix, J.-L. Viovy, C. Villard, J.-M. Peyrin, and L. Malaquin. 2016. In-mold patterning and actionable axo-somatic compartmentalization for on-chip neuron culture. Lab on a Chip 16:2059-2068.
20. Picone, R., B. Baum, and R. McKendry. 2014. Chapter 5 - Plasma Microcontact Patterning (PμCP): A Technique for the Precise Control of Surface Patterning at Small-Scale. In Methods in Cell Biology. P. Matthieu, and T. Manuel, editors. Academic Press. 73-90.
21. Alom Ruiz, S., and C. S. Chen. 2007. Microcontact printing: A tool to pattern. Soft Matter 3:168-177.
22. Singhvi, R., A. Kumar, G. Lopez, G. Stephanopoulos, D. Wang, G. Whitesides, and D. Ingber. 1994. Engineering cell shape and function. Science 264:696-698.
23. Lautenschläger, F., and M. Piel. 2013. Microfabricated devices for cell biology: all for one and one for all. Current Opinion in Cell Biology 25:116-124.
24. Chen, C. S., M. Mrksich, S. Huang, G. M. Whitesides, and D. E. Ingber. 1997. Geometric Control of Cell Life and Death. Science 276:1425-1428.
25. Théry, M., V. Racine, A. Pépin, M. Piel, Y. Chen, J.-B. Sibarita, and M. Bornens. 2005. The extracellular matrix guides the orientation of the cell division axis. Nature Cell Biology 7:947-953.
26. Théry, M., A. Jimenez-Dalmaroni, V. Racine, M. Bornens, and F. Jülicher. 2007. Experimental and theoretical study of mitotic spindle orientation. Nature 447:493-496.
27. Théry, M., V. Racine, M. Piel, A. Pépin, A. Dimitrov, Y. Chen, J.-B. Sibarita, and M. Bornens. 2006. Anisotropy of cell adhesive microenvironment governs cell internal organization and orientation of polarity. Proceedings of the National Academy of Sciences 103:19771-19776.
28. James, J., E. D. Goluch, H. Hu, C. Liu, and M. Mrksich. 2008. Subcellular curvature at the perimeter of micropatterned cells influences lamellipodial distribution and cell polarity. Cell Motility and the Cytoskeleton 65:841-852.
29. Jiang, X., D. A. Bruzewicz, A. P. Wong, M. Piel, and G. M. Whitesides. 2005. Directing cell migration with asymmetric micropatterns. Proceedings of the National Academy of Sciences of the United States of America 102:975-978.
30. Brock, A., E. Chang, C.-C. Ho, P. LeDuc, X. Jiang, G. M. Whitesides, and D. E. Ingber. 2003. Geometric Determinants of Directional Cell Motility Revealed Using Microcontact Printing. Langmuir 19:1611-1617.

31. Mandal, K., I. Wang, E. Vitiello, L. A. C. Orellana, and M. Balland. 2014. Cell dipole behaviour revealed by ECM sub-cellular geometry. Nature Communications 5:5749.
32. Ferizi, M., C. Leonhardt, C. Meggle, M. K. Aneja, C. Rudolph, C. Plank, and J. O. Rädler. 2015. Stability analysis of chemically modified mRNA using micropattern-based single-cell arrays. Lab on a Chip 15:3561-3571.
33. Röttgermann, P. J. F., K. A. Dawson, and J. O. Rädler. 2016. Time-Resolved Study of Nanoparticle Induced Apoptosis Using Microfabricated Single Cell Arrays. Microarrays 5:8.
34. Irimia, D., and M. Toner. 2009. Spontaneous migration of cancer cells under conditions of mechanical confinement. Integrative Biology 1:506-512.
35. Pouthas, F., P. Girard, V. Lecaudey, T. B. N. Ly, D. Gilmour, C. Boulin, R. Pepperkok, and E. G. Reynaud. 2008. In migrating cells, the Golgi complex and the position of the centrosome depend on geometrical constraints of the substratum. Journal of Cell Science 121:2406-2414.
36. Chang, Stephanie S., W.-h. Guo, Y. Kim, and Y.-l. Wang. 2013. Guidance of Cell Migration by Substrate Dimension. Biophysical Journal 104:313-321.
37. Kushiro, K., and A. R. Asthagiri. 2012. Modular Design of Micropattern Geometry Achieves Combinatorial Enhancements in Cell Motility. Langmuir 28:4357-4362.
38. Mahmud, G., C. J. Campbell, K. J. M. Bishop, Y. A. Komarova, O. Chaga, S. Soh, S. Huda, K. Kandere-Grzybowska, and B. A. Grzybowski. 2009. Directing cell motions on micropatterned ratchets. Nature Physics 5:606-612.
39. Kumar, G., C. C. Co, and C.-C. Ho. 2011. Steering Cell Migration Using Microarray Amplification of Natural Directional Persistence. Langmuir 27:3803-3807.
40. Ko, Y.-G., C. C. Co, and C.-C. Ho. 2013. Directing cell migration in continuous microchannels by topographical amplification of natural directional persistence. Biomaterials 34:353-360.
41. Kushiro, K., S. Chang, and A. R. Asthagiri. 2010. Reprogramming Directional Cell Motility by Tuning Micropattern Features and Cellular Signals. Advanced Materials 22:4516-4519.
42. Caballero, D., R. Voituriez, and D. Riveline. 2014. Protrusion Fluctuations Direct Cell Motion. Biophysical Journal 107:34-42.
43. Caballero, D., R. Voituriez, and D. Riveline. 2015. The cell ratchet: Interplay between efficient protrusions and adhesion determines cell motion. Cell Adhesion & Migration 9:327-334.
44. Lo Vecchio, S., R. Thiagarajan, D. Caballero, V. Vigon, L. Navoret, R. Voituriez, and D. Riveline. 2019. Cell motion as a stochastic process controlled by focal contacts dynamics. bioRxiv:750331.
45. Brückner, D. B., A. Fink, C. Schreiber, P. J. F. Röttgermann, J. O. Rädler, and C. P. Broedersz. 2019. Stochastic nonlinear dynamics of confined cell migration in two-state systems. Nature Physics 15:595-601.
46. Segerer, F. J., P. J. F. Röttgermann, S. Schuster, A. P. Alberola, S. Zahler, and J. O. Rädler. 2016. Versatile method to generate multiple types of micropatterns. Biointerphases 11:011005.
47. Roca-Cusachs, P., J. Alcaraz, R. Sunyer, J. Samitier, R. Farré, and D. Navajas. 2008. Micropatterning of Single Endothelial Cell Shape Reveals a Tight Coupling between Nuclear Volume in G1 and Proliferation. Biophysical Journal 94:4984-4995.

48. Thakar, R. G., Q. Cheng, S. Patel, J. Chu, M. Nasir, D. Liepmann, K. Komvopoulos, and S. Li. 2009. Cell-Shape Regulation of Smooth Muscle Cell Proliferation. Biophysical Journal 96:3423-3432.
49. Schneider, C. A., W. S. Rasband, and K. W. Eliceiri. 2012. NIH Image to ImageJ: 25 years of image analysis. Nature Methods 9:671–675.
50. Siegert, S., R. Friedrich, and J. Peinke. 1998. Analysis of data sets of stochastic systems. Physics Letters A 243:275-280.
51. Ragwitz, M., and H. Kantz. 2001. Indispensable Finite Time Corrections for Fokker-Planck Equations from Time Series Data. Physical Review Letters 87:254501.
52. Stephens, G. J., B. Johnson-Kerner, W. Bialek, and W. S. Ryu. 2008. Dimensionality and Dynamics in the Behavior of C. elegans. PLOS Computational Biology 4:e1000028.
53. Efron, B., and R. J. Tibshirani. 1994. An Introduction to the Bootstrap. Chapman and Hall/CRC.
54. Théry, M. 2010. Micropatterning as a tool to decipher cell morphogenesis and functions. Journal of Cell Science 123:4201-4213.
55. Maiuri, P., E. Terriac, P. Paul-Gilloteaux, T. Vignaud, K. McNally, J. Onuffer, K. Thorn, P. A. Nguyen, N. Georgoulia, D. Soong, A. Jayo, N. Beil, J. Beneke, J. C. Hong Lim, C. Pei-Ying Sim, Y.-S. Chu, A. Jiménez-Dalmaroni, J.-F. Joanny, J.-P. Thiery, H. Erfle, M. Parsons, T. J. Mitchison, W. A. Lim, A.-M. Lennon-Duménil, M. Piel, and M. Théry. The first World Cell Race. Current Biology 22:R673-R675.
56. Chen, B., G. Kumar, C. C. Co, and C.-C. Ho. 2013. Geometric Control of Cell Migration. Scientific Reports 3:2827.
57. Webb, K., V. Hlady, and P. A. Tresco. 2000. Relationships among cell attachment, spreading, cytoskeletal organization, and migration rate for anchorage-dependent cells on model surfaces. Journal of Biomedical Materials Research 49:362-368.
58. Guo, M., A. F. Pegoraro, A. Mao, E. H. Zhou, P. R. Arany, Y. Han, D. T. Burnette, M. H. Jensen, K. E. Kasza, J. R. Moore, F. C. Mackintosh, J. J. Fredberg, D. J. Mooney, J. Lippincott-Schwartz, and D. A. Weitz. 2017. Cell volume change through water efflux impacts cell stiffness and stem cell fate. Proceedings of the National Academy of Sciences 114:E8618-E8627.
59. Nehls, S., H. Nöding, S. Karsch, F. Ries, and A. Janshoff. 2019. Stiffness of MDCK II Cells Depends on Confluency and Cell Size. Biophysical Journal 116:2204-2211.
60. Swaminathan, V., K. Mythreye, E. T. O'Brien, A. Berchuck, G. C. Blobe, and R. Superfine. 2011. Mechanical Stiffness Grades Metastatic Potential in Patient Tumor Cells and in Cancer Cell Lines. Cancer Research 71:5075-5080.
61. Palecek, S. P., J. C. Loftus, M. H. Ginsberg, D. A. Lauffenburger, and A. F. Horwitz. 1997. Integrin-ligand binding properties govern cell migration speed through cell-substratum adhesiveness. Nature 385:537-540.
62. Bi, D., X. Yang, M. C. Marchetti, and M. L. Manning. 2016. Motility-Driven Glass and Jamming Transitions in Biological Tissues. Physical Review X 6:021011.
63. Albert, Philipp J., and Ulrich S. Schwarz. 2014. Dynamics of Cell Shape and Forces on Micropatterned Substrates Predicted by a Cellular Potts Model. Biophysical Journal 106:2340-2352.

64. Kassianidou, E., D. Probst, J. Jäger, S. Lee, A.-L. Roguet, U. S. Schwarz, and S. Kumar. 2019. Extracellular Matrix Geometry and Initial Adhesive Position Determine Stress Fiber Network Organization during Cell Spreading. Cell Reports 27:1897-1909.
65. Parker, K. K., A. L. Brock, C. Brangwynne, R. J. Mannix, N. Wang, E. Ostuni, N. A. Geisse, J. C. Adams, G. M. Whitesides, and D. E. Ingber. 2002. Directional control of lamellipodia extension by constraining cell shape and orienting cell tractional forces. The FASEB Journal 16:1195-1204.
66. Prager-Khoutorsky, M., A. Lichtenstein, R. Krishnan, K. Rajendran, A. Mayo, Z. Kam, B. Geiger, and A. D. Bershadsky. 2011. Fibroblast polarization is a matrix-rigidity-dependent process controlled by focal adhesion mechanosensing. Nature Cell Biology 13:1457-1465.
67. Boudaoud, A., A. Burian, D. Borowska-Wykręt, M. Uyttewaal, R. Wrzalik, D. Kwiatkowska, and O. Hamant. 2014. FibrilTool, an ImageJ plug-in to quantify fibrillar structures in raw microscopy images. Nature Protocols 9:457–463.
68. Eltzner, B., C. Wollnik, C. Gottschlich, S. Huckemann, and F. Rehfeldt. 2015. The Filament Sensor for Near Real-Time Detection of Cytoskeletal Fiber Structures. PLOS ONE 10:e0126346.
69. Rogge, H., N. Artelt, N. Endlich, and K. Endlich. 2017. Automated segmentation and quantification of actin stress fibres undergoing experimentally induced changes. Journal of Microscopy 268:129-140.
70. Doyle, A. D., F. W. Wang, K. Matsumoto, and K. M. Yamada. 2009. One-dimensional topography underlies three-dimensional fibrillar cell migration. The Journal of Cell Biology 184:481-490.
71. Comelles, J., D. Caballero, R. Voituriez, V. Hortigüela, V. Wollrab, Amélie L. Godeau, J. Samitier, E. Martínez, and D. Riveline. 2014. Cells as Active Particles in Asymmetric Potentials: Motility under External Gradients. Biophysical Journal 107:1513-1522.
72. Prentice-Mott, H. V., Y. Meroz, A. Carlson, M. A. Levine, M. W. Davidson, D. Irimia, G. T. Charras, L. Mahadevan, and J. V. Shah. 2016. Directional memory arises from long-lived cytoskeletal asymmetries in polarized chemotactic cells. Proceedings of the National Academy of Sciences 113:1267-1272.
73. Ben-David, U., B. Siranosian, G. Ha, H. Tang, Y. Oren, K. Hinohara, C. A. Strathdee, J. Dempster, N. J. Lyons, R. Burns, A. Nag, G. Kugener, B. Cimini, P. Tsvetkov, Y. E. Maruvka, R. O'Rourke, A. Garrity, A. A. Tubelli, P. Bandopadhayay, A. Tsherniak, F. Vazquez, B. Wong, C. Birger, M. Ghandi, A. R. Thorner, J. A. Bittker, M. Meyerson, G. Getz, R. Beroukhim, and T. R. Golub. 2018. Genetic and transcriptional evolution alters cancer cell line drug response. Nature 560:325-330.
74. Brückner, D. B., A. Fink, J. O. Rädler, and C. P. Broedersz. 2019. Disentangling the Behavioural Variability of Confined Cell Migration. bioRxiv:797555.
75. Guck, J., S. Schinkinger, B. Lincoln, F. Wottawah, S. Ebert, M. Romeyke, D. Lenz, H. M. Erickson, R. Ananthakrishnan, D. Mitchell, J. Käs, S. Ulvick, and C. Bilby. 2005. Optical Deformability as an Inherent Cell Marker for Testing Malignant Transformation and Metastatic Competence. Biophysical Journal 88:3689-3698.
76. Schwarz, U. S. 2019. Island hopping for cells. Nature Physics.
77. LiveWire Plugin Download. 2009. 01.05.2018. https://sourceforge.net/projects/ivussnakes/.

Supplementary Information for

Area and Geometry Dependence of Cell Migration in Asymmetric Two-State Micropatterns

Alexandra Fink[1], David B. Brückner[2], Christoph Schreiber[1], Peter J. F. Röttgermann[1], Chase P. Broedersz[2] and Joachim O. Rädler[1+]

[1] Faculty of Physics and Center for NanoScience, Ludwig-Maximilians-Universität, München
[2] Arnold Sommerfeld Center for Theoretical Physics and Center for NanoScience, Ludwig-Maximilians-Universität, München
[+] corresponding author

0. Movie Description

Supplementary Movie M1:
Single MDA-MB-231 cell migrating repeatedly back and forth between two square adhesion sites of different areas (A_{left} = 1405 µm^2 and A_{right} = 755 µm^2). The nucleus is stained with Hoechst 33342 to allow semi-automated cell tracking.

Supplementary Movies M2-M4:
Single MDA-MB-231 cells migrating repeatedly back and forth between two adhesion sites of comparable areas but of different shapes: square and circle, square and rhombus and square and triangle.

Supplementary Movies M5-M7:
Single MDA-MB-231 cells migrating on rectangle micropatterns of different orientations. It is visible how the cells polarize along the long axes of the rectangles. Bridge length = 16.2 µm.

Supplementary Movies M8-M9
LifeAct-GFP stained single MDA-MB-231 cells migrating on square-circle (equal areas) or square-triangle micropatterns. Actin hotspots along the cell periphery are visible which are preferentially localized in the patterns' corners.

Supplementary Movie M10
Single MDA-MB-231 cell migrating on square-circle micropattern with adhesion sites of equal area. This movie was acquired at high magnification (60x) and with a time resolution of 30s to visualise filopodia dynamics.

1. Additional Materials and Methods

1.1. Detailed Micropattern Dimensions

x-dimensions left site [µm]*	x-dimensions right site [µm]*	Area left [µm^2]	Area right [µm^2]	Further information	Number of cells analysed	Bridge length x width [µm]
37.5 ± 0.3	27.5 ± 0.2	1405 ± 22.5	754.7 ± 11.0	Square-square	47	(16.0 ± 0.3) x (7.8 ± 0.3)
37.6 ± 0.6	32.4 ± 0.7	1410.6 ± 45.1	1052 ± 45.4	Square-square	68	
37.2 ± 0.8	37.4 ± 0.6	1385.1 ± 59.5	1397.3 ± 44.9	Square-square (symmetric)	169	
42. 1 ± 0.3	27.3 ± 0.4	1776.2 ± 25.3	743.1 ± 21.8	Square-square	58	
42.2 ± 0.5	32.1 ± 0.6	1782.1 ± 42.2	1028.8 ± 38.5	Square-square	74	
36.5 ± 0.4	40.3 ± 0.6	1329.7 ± 29.2	1275 ± 38.0	Square-circle (equal area)	67	(14.5 ± 0.2) x (6.1 ± 0.1)
35.9 ± 0.2	44.6 ± 0.3	1291.9 ± 14.4	1580.2 ± 21.0	Square-circle (equal perimeter)	98	
37.6 ± 0.2	46.7 ± 0.5 49.6 ± 0.6	1413.8 ± 14.3	1243.2 ± 24.7	Square-rhombus	63	(17.7 ± 0.1) x (7.7 ± 0.1)
37.8 ± 0.2	33.5 ± 0.3 67.7 ± 0.7	1428.8 ± 17.3	1263.8 ± 16.5	Square-triangle	82	
48.9 ± 0.7 24.9 ± 0.4	48.9 ± 0.8 24.9 ± 0.3	1217.6 ± 26.2	1217.6 ± 24.6	Lying rectangles	78	(16.2 ± 0.1) x (5.0 ± 0.2)
24.9 ± 0.4 47.9 ± 0.5	24.8 ± 0.3 47.8 ± 0.4	1192.7 ± 22.1	1185.4 ± 15.3	Upright rectangles	65	
48.2 ± 0.7 24.3 ± 0.4	25.2 ± 0.2 47.8 ± 0.5	1171.3 ± 24.0	1204.6 ± 16.9	Mixed rectangles	59	

Table S1. Measures of all used dumbbell geometries, calculated adhesion site areas and numbers of cells analysed for each geometry. Orange shading denotes adhesion sites designed to have comparable areas**.** Errors are weighted standard deviations as different statistics were gathered from each experiment and an experiment-to-experiment variation between pattern dimensions is unavoidable. * Y-dimensions of longest axis are given for anisotropic shapes (bottom numbers)**.**

The following channel lengths were probed with different rectangle orientations: 8.2 ± 0.3 µm, 16.2 ± 0.1 µm (as presented in the main text), 24.8 ± 0.2 µm, and 34.7 ± 0.1 µm.

1.2. Cell Exclusion Criteria

To limit the effects of abnormal migration behaviour, we deem cells suitable for further analysis if they comply with the following criteria (1):

1. Only a single cell occupies the micropattern. Observations are stopped when the cell rounds up for division.
2. We only include trajectories of cells which, including their protrusions, are entirely confined within the borders of the micropattern.
3. Transition statistics are only included after the first and last observed transition. This avoids start- and end-of-measurement artefacts in determining dwell times.
4. The cell shows no abnormalities such as multiple nuclei or the occurrence of cell death at any time during the whole experiment.
5. In the vast majority of cases the cell performs complete transitions. A complete transition is defined by the fact that no parts of the cell adhere to the previous adhesion site once the nucleus has entered the new adhesion site.

Criteria 1-4 are basic requirements for single cell studies on micropatterns. We try to limit the influence of abnormal migration behaviour by applying criterion 4, as it would be expected that dying cells, in contrast to healthy cells, respond differently to external cues. Criterion 5 is specific to our dumbbell setup and, depending on the cell type, may be a rather strong constraint. We previously found that criterion 5 can be relaxed without affecting general conclusions on cell dynamics (1). We have decided to apply criterion 5 to the data presented in this work as we intend to probe cellular response to either of two geometries *separately*.

1.3. Filopodia Movies

For filopodia movies, images in the brightfield channel were taken every 30s and, to keep experimental conditions as similar as possible to other experiments, every 10 minutes an image in the DAPI channel was acquired.

1.4. Cell Area Determination

Cell areas were determined manually with ImageJ's Ivussnakes Plugin (2) by tracing the cell outlines visible in the phase contrast images. Cell areas were determined for cells complying with the above exclusion criteria and only at times where the cell was fully confined (i.e. where there were no protrusions inside the channel). As this approach overestimates the influence of smaller cells, which are naturally better confined, a mean cell area was

calculated for every stay (in contrast to calculating a mean over all frames used) and an average of these areas was taken as final cell area.

2. Filopodia

To investigate filopodia dynamics in our setup, we acquired images of cells migrating on the two-state system every 30s with high magnification (60x). Filopodia are mainly visible exploring the passivated area surrounding the micropatterns. Sometimes, their exploratory motion is followed by a lamellipodium spreading in the same direction (cf. time series Fig. S1). Filopodia were only rarely observed on the fibronectin-coated areas of the micropatterns, indicating that cell motion in our system is dominated by lamellipodia.

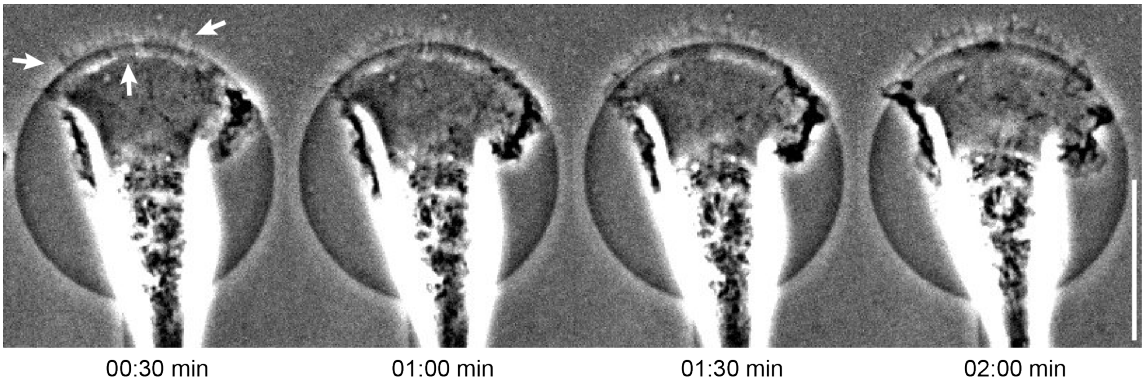


Figure S1. Time series of a single cell on a circular adhesion site highlighting filopodia activity. Typically, filopodia can be seen vividly exploring the passivated surroundings of micropatterns. The white arrows indicate some of the filopodia visible.

3. Cell Areas on Square Adhesion Sites

To test the influence of cell area on the mean dwell times, we measure the average cell areas that cells actually cover on the differently sized square adhesion sites. In Fig. S2, the dependence of cell area on available adhesion site area is shown. It is clearly visible that only for the two smallest adhesion sites cells fully occupy the adhesion site, and thereby reflect in their area the area of the adhesion sites. For larger adhesion sites, the cell area still increases but does so more slowly. This means that on large adhesion sites, cells frequently do not fully fill all available area of the confining site.

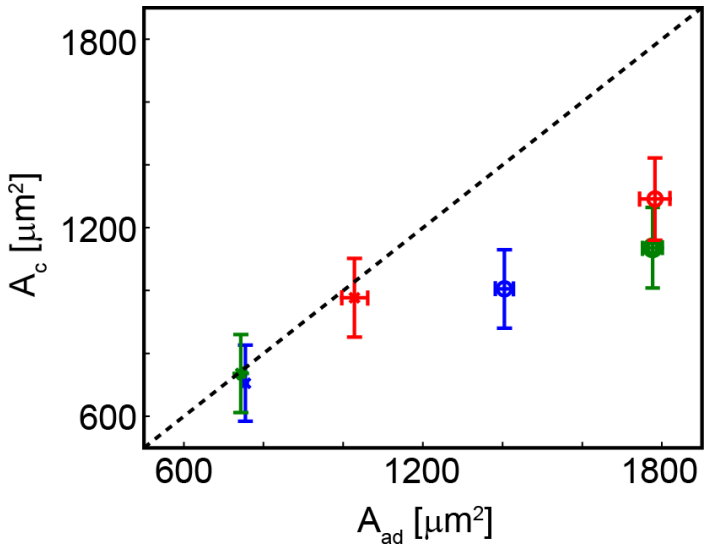


Figure S2. Cell areas A_c on adhesion sites of different areas A_{ad}. The cell area increases with provided adhesion area. The increase in cell area is not equal to the increase of provided adhesion areas. The dashed line is a guide to the eye corresponding to $A_c = A_{ad}$. Same colours denote cell areas taken from the same dumbbell geometry. Error bars of the cell area are obtained by bootstrapping.

4. Comparison between Model and Experiment for Square Adhesion Sites

The comparison between model predictions and experimental data using the predictions by our model based on the short timescale dynamics of the cells (1) is shown in Fig. S3.

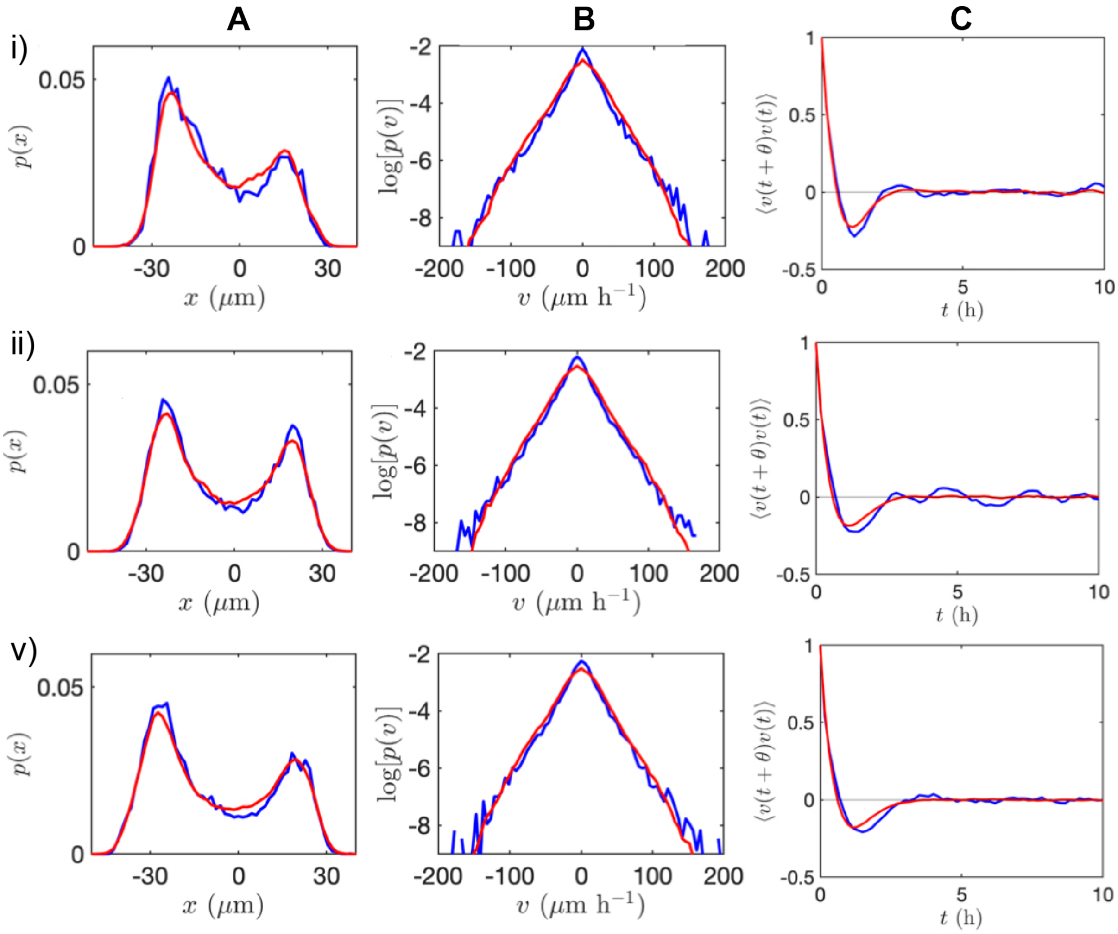


Figure S3. Experimental (blue) and predicted (red) probability distributions of position (A), velocity (B) plotted semi-logarithmically) and (C) velocity correlation function for all asymmetric two-state micropatterns presented in Fig. 2 except (iv).

5. Square-Circle Pattern of Equal Perimeter

To probe the influence of shape on cellular dwell times, we have created square-circle micropatterns with adhesion sites of either equal area (data shown in main text) or the same perimeter (Fig. S4). We find that occupation probabilities on square-circle patterns are biased towards the circle if the square and circular adhesion sites have the same perimeters. When the corresponding mean dwell times are added to the plot of mean dwell times against adhesion site areas, the data points follow the same linear dependence on area as seen for square-only micropatterns (Fig. 2B and 2C).

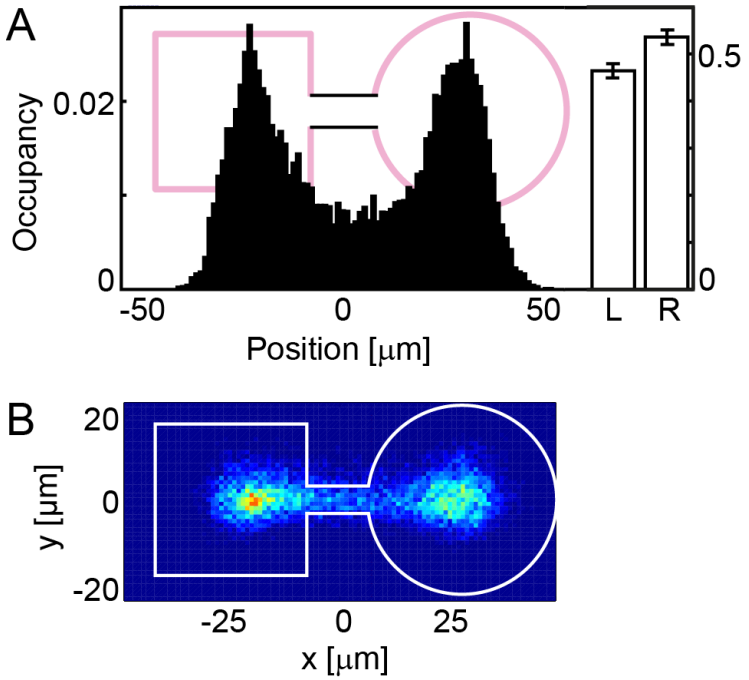


Figure S4. A: Probability distribution in x-direction of cell positions on square-circle micropattern with adhesion sites of equal perimeter. A significant asymmetry of stay probabilities is visible, with stays biased towards the circular adhesion site. Cells spend (46 ± 1) % of their stays on the left side, and (54 ± 1) % of their stays on the right side. **B**: 2D probability distribution of cell positions on the same geometry. A broader spread of positions is visible on the circular site.

6. Survival Functions on Adhesion Sites of Different Shapes

We measured the dwell times in repeated experiments for square, circular, rhombical and triangular adhesion sites of similar areas. While for the circle the survival function of stay times, S(t), does not differ from those measured on square adhesion sites, a gradual broadening of S(t) is visible for the rhombical and triangular geometry (Fig. S5).

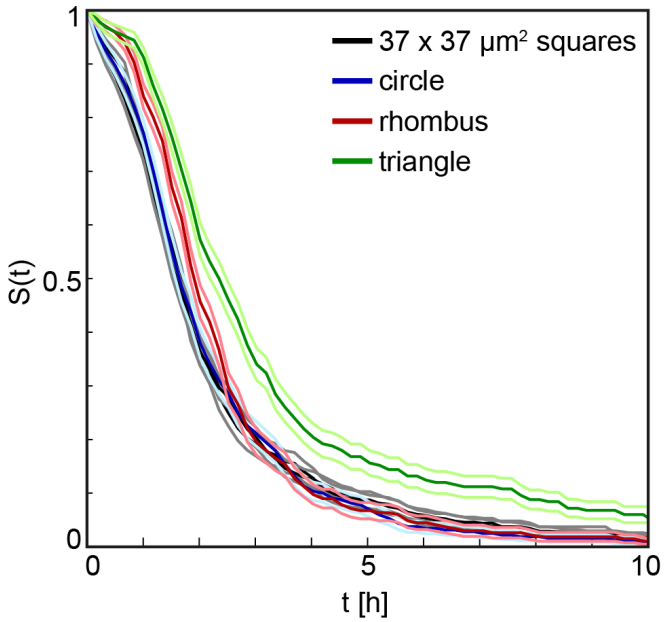


Figure S5. Variability of survival functions S(t) for adhesion sites of approximately the same area. S(t) for stays on square adhesion sites are plotted in black. In blue, the survival function for stays on circular adhesion sites with the same area is shown. The red curve shows S(t) for the rotated square (rhombus), the green curve shows S(t) for triangular adhesion sites. Bootstrapping errors are plotted in light colours.

7. Cell Organisation and Polarisation on Differently Shaped Adhesion Sites

To study cell organisation on a range of adhesion site shapes, we have fixed cells and stained their actin cytoskeleton and focal adhesions. Focal adhesions are mainly localised under lamellipodia or in cell corners. If cells are migrating within adhesion sites, actin stress fibres organise along the direction of motion and indicate cell polarisation.

To test whether cells polarise on the adhesion sites provided, we plot a histogram of the direction of the cell velocity between two subsequent time points of each stay. Between squares of different areas, a circle and a rhombus of similar areas very few differences in the preferential direction of nuclear displacement on the adhesion site are visible. Only for the triangle, which is anisotropic under rotations of 90°, is elongated in y-direction and sports acute angles located orthogonally to the axis of transitions, the angular distributions looks

different (lowest polar plot, Fig. S6). The extent of cell motion, and interrelatedly, cell polarisation in y-direction is increased compared to the other studied adhesion site geometries presented in Fig. S6.

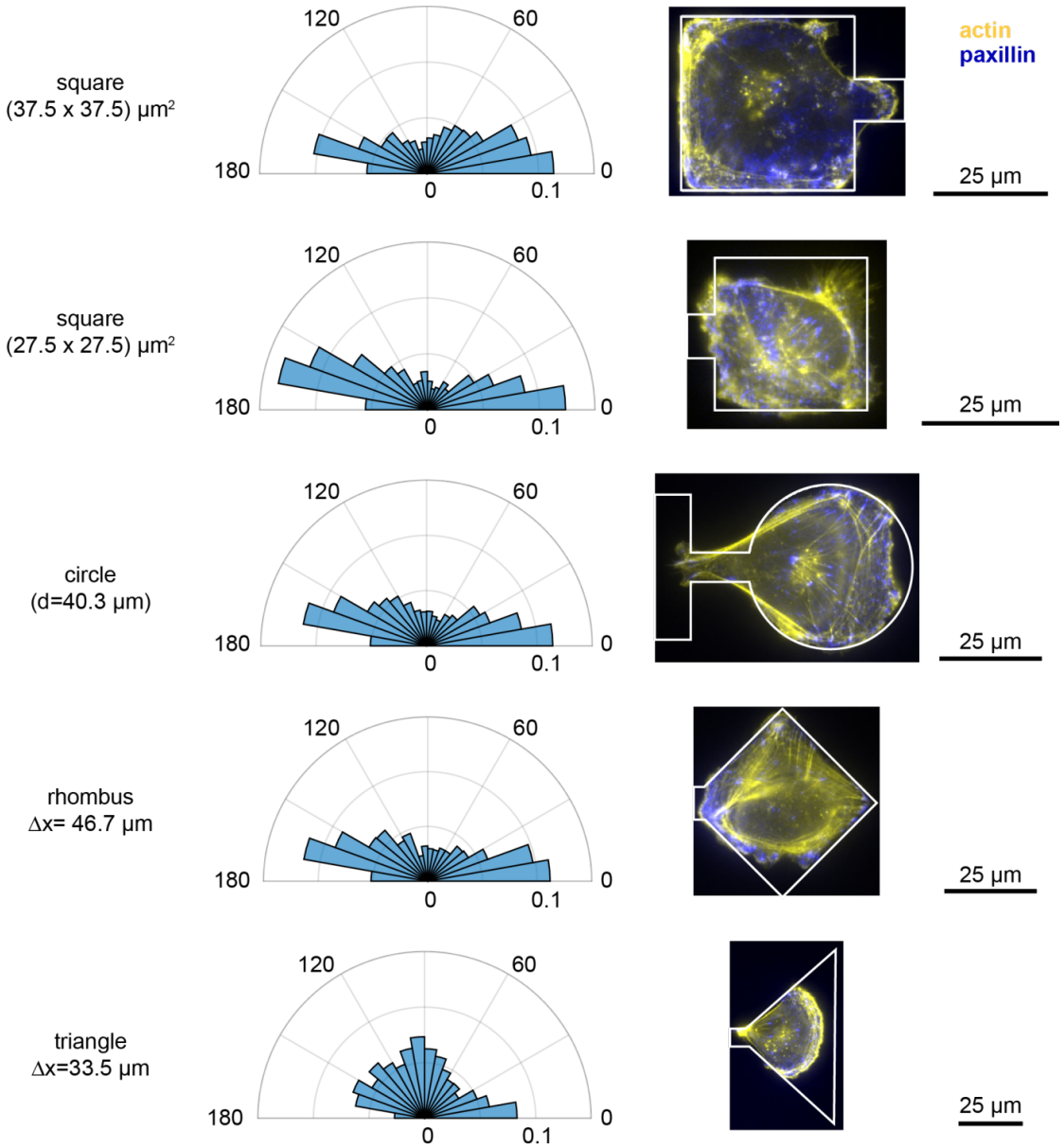


Figure S6. Angular probability distributions of the directions of cell velocities on adhesion sites only. The corresponding micropatterns are detailed on the left and shown as overlays on the right. The directions are stated with respect to the long axis of the micropattern (i.e. the axis of transitions). Each bin contains the counts for the shown angle and for motion in angle+180°, as the angles are calculated using the scalar product. Only cell motion within adhesion sites is included in the analysis. Also, images of fixed cells on the respective micropatterns stained for actin and paxillin (merged) are shown.

8. Rectangle Micropatterns

8.1. Survival Functions as a Function of Bridge Length

To further explore the migration of cells between anisotropic states, we also vary the length of the connecting bridge between the two rectangles. Due to their identical orientation, we expect the survival probability functions on the symmetric upright rectangle dumbbell and on the upright rectangle from the mixed orientation setup to be identical. The same should be true for horizontal rectangles. From Fig. S7A, where survival probability functions for corresponding rectangle orientations are shown in the same plot and for different channel lengths, it is evident that cells behave differently for the same rectangle orientations, depending on the overall composition of the two-state system. This could be related to memory effects such as have previously been suggested in ratchetaxis (3). Interestingly, the time-scales of dwell times in the rectangles are longer than the 50 minutes reported as stress fibre orientation memory (4), possibly indicating more determinants for cellular memory. However, when fixed cells are stained for their actin cytoskeleton and focal adhesions, no obvious differences are visible between cells in the mixed orientation and symmetric setups (Fig. 7 and Fig. S12).

Furthermore, we observe an increase of the mean dwell times with channel length for all rectangle orientations and setups. This behaviour is in qualitative agreement with our observations on symmetric square systems (1). Strikingly, also the dwell times on upright rectangles in the mixed setup are generally longer than the corresponding stay times on upright rectangles in the symmetric setup. Furthermore, there is no significant difference in stay times between the horizontal rectangles and the upright rectangles from the mixed orientation setup (Fig. S7B).

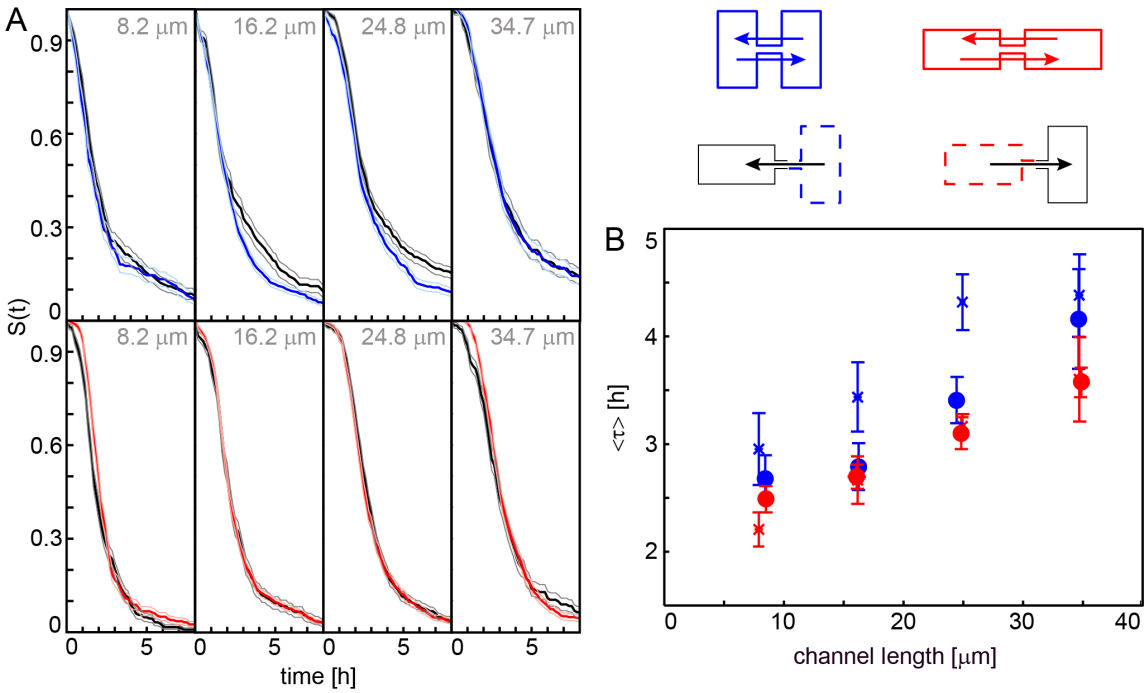


Figure S7. **A**: Survival functions S(t) for different channel lengths. In each plot, the survival function is shown for rectangles of the same orientation but from different setups, the black line always corresponding to mixed-orientation dumbbells. Despite the rectangles having the same orientation, deviations in the decay behaviour of S(t) are visible for longer times (t>3h) for upright rectangles connected by channels of medium lengths. In contrast, some differences in S(t) are visible for horizontally-oriented rectangles (lower plots) for short stay times. **B**: Mean stay time $\langle\tau\rangle$ plotted against channel length for all tested rectangle orientations. An approximately linear increase of mean dwell times with increasing channel length is visible for all orientations. The blue cross datapoints correspond to stay times measured on the upright rectangle, in the mixed orientation setup, the solid blue circle data correspond to the symmetric setup of upright rectangles, the red crosses corresponds to data from the lying rectangle of a mixed-orientation setup and the solid red circles correspond to stay times measured on dumbbells with equally oriented, horizontal rectangular adhesion sites.

8.2. Migration Behaviour on Rectangle Micropatterns

In all rectangle orientations, cells are observed to form protrusions along the long axis of the rectangle (Fig. S8). This results in a broadening of the probability distribution of y-positions on the vertical rectangles relative to the horizontal case (Fig. S9). This is the case in all vertical rectangles, both in the symmetric and the mixed setup. The anisotropies also significantly affect the probability distributions of velocities, which are always broader along the long axis of each rectangle (Fig. S10). Taken together, these results show that there is a significant difference in migratory activity along the long and short axes of the rectangles.

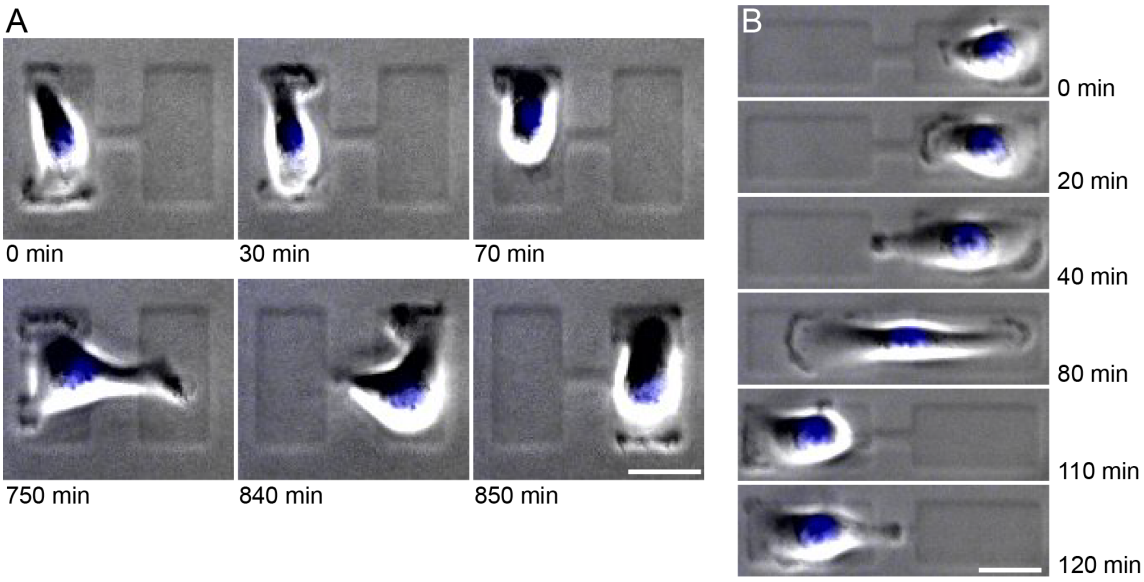


Figure S8. A: Time series of a single cell migrating on a dumbbell made of vertical rectangles (bridge length L = 16.2 µm). It can be seen how lamellipodia are formed along the long axis of the rectangle. Scale bar: 25 µm. **B**: Time series of a single cell transitioning on horizontal rectangles. The cell is seen to migrate within each rectangle until it meets the dumbbell border. Scale bar: 25 µm.

Furthermore, when inspecting cellular trajectories in x- and y-directions, it becomes apparent that x- and y-motion are often mutually exclusive (Fig. S11A-B). Thus, y-motion is pretty much suppressed on horizontal rectangles, and if the extent of motion in y-direction is large on vertical rectangles, transitions along the bridge seem to be less frequent (cf. Fig. S11A). Obviously, this is a consequence of the adhesion sites extending very little along the short axis of the rectangle but it also agrees well with the observation of stress fibre orientation and directionality of cell motion as indicators of polarisation (Fig. 7, Fig. S12).

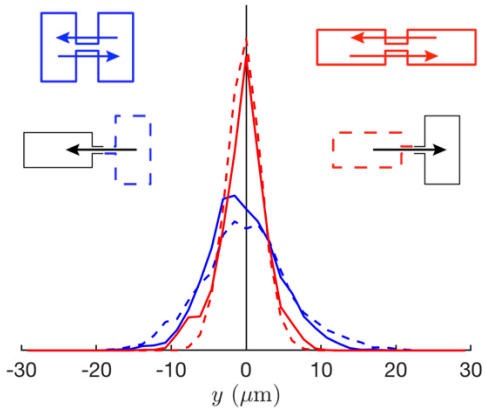


Figure S9. Probability distributions of positions in the vertical direction on rectangle micropatterns with bridge length L=16.2 µm. The cartoons serve as a legend to the four distributions.

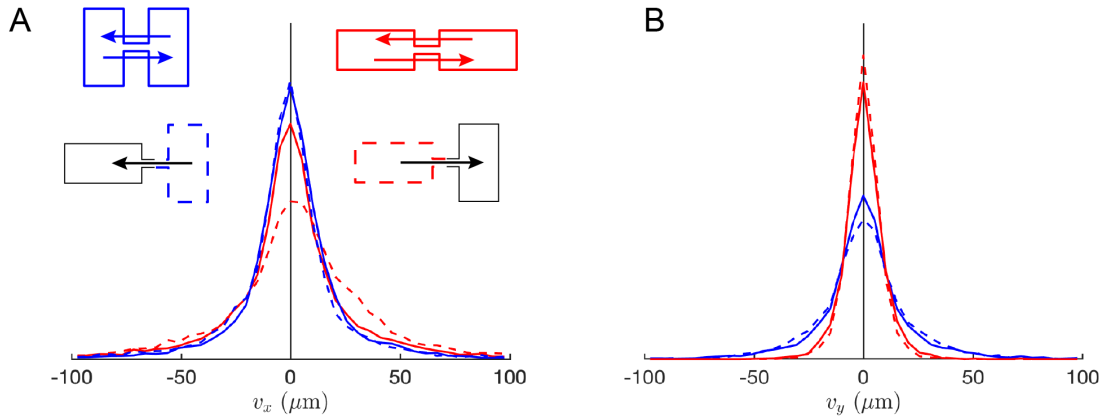


Figure S10. Probability distributions of velocities in horizontal and vertical directions (bridge length L = 16.2 µm). **A**: Distribution of velocities along the micropattern's main direction (x-direction). **B**: Distribution of velocities in the direction orthogonal to the main axis of the micropattern (y-direction). The colour code is indicated by the cartoons.

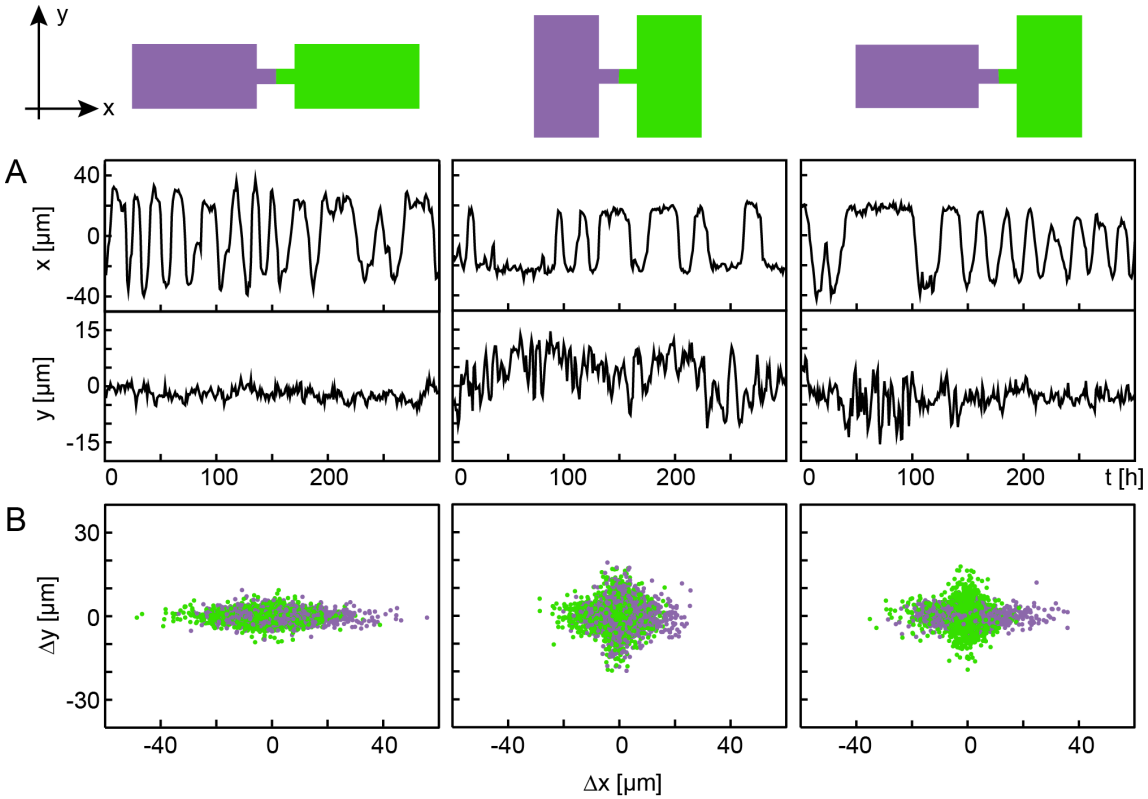


Figure S11. **A**: Comparison between x- and y-motion: Exemplary single-cell trajectories in x- and y-direction for each of the three rectangle setups. The displacement fluctuations along the y-axis are larger for vertical rectangles than for horizontal rectangles. **B**: Displacements in x- and y-direction with respect to the centres of the rectangles. The colours of the dots correspond to the left or right adhesion site, as indicated in the schematic drawings. It is clearly visible that where the extent of x-motion is large, motion in the y-direction is limited, and vice versa.

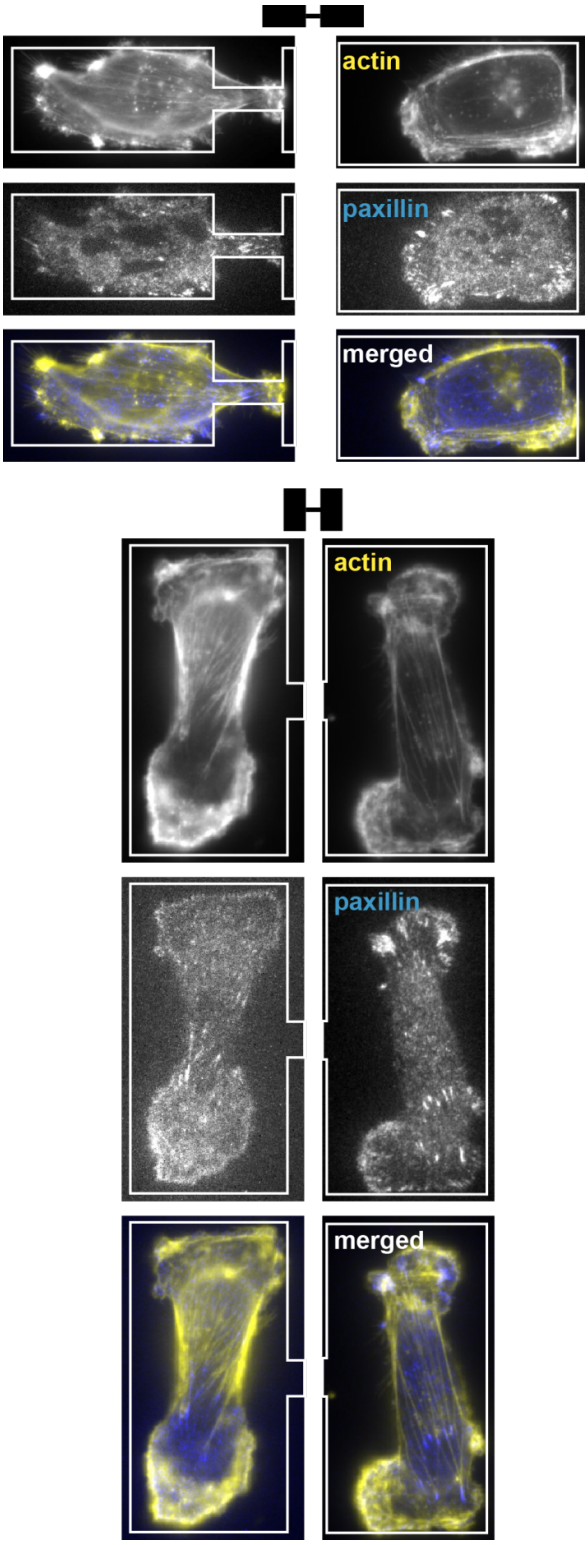


Figure S12. Images of fixed cells on rectangular adhesion sites stained for actin (yellow) and paxillin (blue). The underlying micropatterned adhesion sites are superimposed as a reference.

9. All Dwell Times

When all dwell times measured on the two-state systems are put into perspective, it is visible that the vertical rectangle and triangle data points deviate from the rest of the data, indicating that determinants other than area affect dwell times. For all other adhesion site geometries, the mean dwell times for differently shaped motifs and equally sized adhesion sites from different micropatterns agree within error, suggesting that there are no memory effects (Fig. S13). Memory effects in the rectangular system are discussed in section S8.

We can also test the dependence of escape rates, i.e. inverse mean dwell times, on the ratio of bridge width and adhesion site edge length. Intuitively, one would expect a correlation between the likelihood of escape and the relative size of an opening providing an exit from confinement. When plotting escape rates against the bridge width/edge length ratio, correlated and uncorrelated regions in the data are visible (Fig. S10)

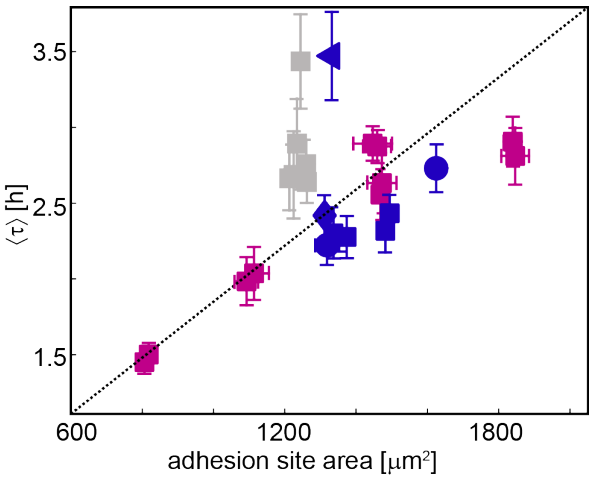


Figure S13: Mean dwell times plotted against adhesion site area for all tested adhesion sites. Both, the vertical rectangle from the mixed setup (topmost grey square) and the triangle (blue triangular data point) clearly deviate from the linear trend observed for square, circular and rhombical adhesion sites.

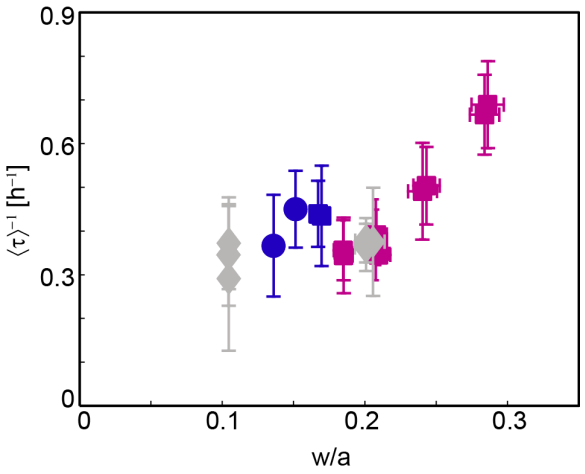


Figure S14: Plot of escape rates (inverse mean dwell times) against the ratio of bridge width and adhesion site edge length. Pink data points correspond to square-square micropatterns, blue data points correspond to square-circle micropatterns with the symbol shape depicting the particular adhesion site shape, and grey data points correspond to rectangular systems of all reported orientations.

Supporting References

1. Brückner, D. B., A. Fink, C. Schreiber, P. J. F. Röttgermann, J. O. Rädler, and C. P. Broedersz. 2019. Stochastic nonlinear dynamics of confined cell migration in two-state systems. Nature Physics 15:595-601.
2. LiveWire Plugin Download. 2009. 01.05.2018. https://sourceforge.net/projects/ivussnakes/.
3. Comelles, J., D. Caballero, R. Voituriez, V. Hortigüela, V. Wollrab, Amélie L. Godeau, J. Samitier, E. Martínez, and D. Riveline. 2014. Cells as Active Particles in Asymmetric Potentials: Motility under External Gradients. Biophysical Journal 107:1513-1522.
4. Kassianidou, E., D. Probst, J. Jäger, S. Lee, A.-L. Roguet, U. S. Schwarz, and S. Kumar. 2019. Extracellular Matrix Geometry and Initial Adhesive Position Determine Stress Fiber Network Organization during Cell Spreading. Cell Reports 27:1897-1909.

Chapter 6

Protrusion and polarity dynamics in confined cell migration

This chapter is based on the following manuscript:

Theory of protrusion and polarity dynamics in confined cell migration

David B. Brückner, Matthew Schmitt, Alexandra Fink, Johannes Flommersfeld, Nicolas Arlt, Edouard Hannezo, Joachim O. Rädler and Chase P. Broedersz[†]

[†] corresponding author

This manuscript is in preparation for publication.

Theory of protrusion and polarity dynamics in confined cell migration

David B. Brückner[1], Matthew Schmitt[1], Alexandra Fink[2], Johannes Flommersfeld[1], Nicolas Arlt[1], Edouard Hannezo[3], Joachim O. Rädler[2] and Chase P. Broedersz[1,4,†]

[1]Arnold Sommerfeld Center for Theoretical Physics and Center for NanoScience, Department of Physics, Ludwig-Maximilian-University Munich, Theresienstr. 37, D-80333 Munich, Germany, [2]Faculty of Physics and Center for NanoScience, Ludwig-Maximilian-University, Geschwister-Scholl-Platz 1, D-80539 Munich, Germany, [3]Institute of Science and Technology, Am Campus 1, 3400 Klosterneuburg, Austria, [4]Department of Physics and Astronomy, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands, [†]corresponding author

Cell migration in many physiological processes relies on the concerted dynamics of several cellular components, including the formation of cell protrusions, adhesive connections to the environment, and the positioning of the cell nucleus. These components are coupled by the polarizable active cytoskeleton, and together play the dual role of driving net motion of the cell and sensing its local microenvironment. However, it remains poorly understood how the dynamic interplay of these components determines the emergent migration behavior at the cellular scale, and how these dynamics adapt to confining environments. Here, we develop a hybrid mechanistic and data-driven theoretical approach, where we use experimental data to systematically constrain a mechanistic model for confined cell migration. We measure a large data set of joint protrusion and nucleus trajectories of cells migrating in standardized micropatterned confinements featuring a thin constriction. Interestingly, we find that cells exhibit a stereotypical migration pattern, with protrusions growing to precede the transmigration of the cell nucleus across the constriction, which we term 'protrusion-nucleus cycling'. Based on a data-driven approach, we reveal that the average dynamics of the cell nucleus are determined by the locally available adhesive area. Furthermore, our model indicates that the protrusion dynamics are driven by a cell polarity that couples to the local geometry by switching from a negative to a positive, self-reinforcing feedback loop under strong confinement. Strikingly, this model predicts, in agreement with the experiment, that the protrusion-nucleus cycling disappears when the constriction is removed. This implies that the self-reinforcing polarity feedback loop emerges as a consequence of an adaptation of the cellular dynamics to the presence of the thin constriction. Our theoretical approach therefore suggests polarity feedback adaptation as a key mechanism in confined cell migration.

The ability of cells to migrate is essential for many physiological processes, including embryogenesis, immune response, and cancer [1–4]. The migration of cells is intimately linked to changes of their shape: migrating cells generate protrusions that drive migration [5, 6]. Protrusion formation is the result of complex molecular processes, including the polymerization of actin which in turn is regulated by an interconnected network of diffusible polarity cues and actin regulators [5, 7–9]. At the cellular scale, this molecular machinery leads to coordinated, functional migration, which manifests as persistent random motion on uniform two-dimensional substrates [10, 11]. However, in physiologically relevant environments, cells must

navigate complex, structured extracellular environments [6,12], featuring obstacles such as thin constrictions [4,13]. Thus, migrating cells need to adapt their migration strategy by responding to the structure of their local micro-environment. However, it remains unclear how the underlying protrusion and polarity dynamics determine the emergent migration dynamics of cells in structured environments on long time-scales.

Achieving a quantitative understanding of how the dynamics of protrusions and polarity control migration in structured environments could yield key insights into both the underlying molecular mechanisms, and the adaptive system-level behaviors. Previous biophysical approaches to develop quantitative cell migration models can be broadly classified into two categories: top-down inference approaches, which systematically constrain model candidates using experimental data; and bottom-up mechanistic models, which postulate specific mechanisms and explore their implications. Top-down approaches have been used to describe the dynamics of migrating cells at the level of cell trajectories using equations of cell motion. This has been successfully applied to quantify the persistent random motion of freely migrating cells [11], migration in structured confinements [14–17] and 3D matrices [14,18], as well as pair-wise interactions of cells [19]. At the mechanistic level, complex computational models that include polarity processes and protrusion formation have been developed, including phase-field [20–22] and Cellular Potts models [23–25]. More minimal models include molecular clutch models [26,27], which account for the coupling of adhesions to substrate mechanics, as well as active gel theories [28,29] and models coupling actin flow, polarity cues, and focal adhesion dynamics [30–35]. However, these approaches suffer from two key limitations: computational models have many parameters that are difficult to constrain based on experimental data, and thus testing their predictive power for particular experiments remains challenging. The more minimal models are frequently taylored to capture a particular aspect of the data, but it has often remained difficult to capture the full long time-scale dynamics of the cells, or how these dynamics adapt to external inputs. A key difficulty in connecting such approaches to experimentally observed migration dynamics is the intrinsic stochasticity of migrating cells, which results in highly variable behaviors. Thus, there is currently a fundamental divide in the field that has yet to be bridged: how to connect mechanistic approaches to the underlying protrusion and polarity dynamics of confined migrating cells with their emergent, stochastic long time-scale motion.

Here, we develop a hybrid data-driven and mechanistic approach, where we use experimental data to rigorously constrain a mechanistic model for confined cell migration postulated on the basis of physical intuition and known cellular processes. Specifically, we measure a large data set of joint nucleus and protrusion trajectories of cells migrating in a standardized micropatterned confinement. We find that the stochastic dynamics of protrusions are a key driver of migration across thin constrictions imposed by the micropattern. Specifically, we find a stereotypical pattern of 'protrusion-nucleus cycling', where protrusions grow to precede the transmigration of the cell nucleus across the constriction. To account for these intricate coupled protrusion-nucleus dynamics, we use a systematic approach to determine a mechanistic model which connects all the way from the short time-scale protrusion dynamics to the long time-scale stochastic cell behavior. Remarkably, we find that this model is able to capture cellular dynamics in systems with varying constriction width and length, indicating that the model has predictive power beyond the specific confinement geometry used to constrain it. This approach reveals two key insights: first, we find that the dynamics of the cell nucleus in the confinement is determined by a spatially variable friction coefficient determined by the adhesive area locally available to the cell. Second, in response to thin constrictions, the polarity dynamics switch from a negative to a positive feedback loop, leading to strong polarities and persistent protrusion growth in the constriction. Our model therefore reveals that cellular protrusion dynamics adapt to the local confinement geometry, leading to the emergence of stereotypical protrusion-nucleus migration cycles in response to thin constrictions.

Protrusion dynamics drive confined cell migration

To investigate the dynamics of cell shapes and protrusions in confined migration, we study the migration dynamics of single MDA-MB-231 breast carcinoma cells confined to two-state micropatterns (Fig. 1a). These patterns consist of two adhesive islands connected by a thin adhesive bridge, which acts as a constriction, allowing us to study how migrating cells respond to constraints in the extra-cellular environment. A minimal description of these dynamics is provided by the trajectories of the cell nuclei. In previous work [15], we found that these nucleus trajectories exhibit intricate stochastic non-linear dynamics in position-velocity phase space. This is a generalization of the persistent random motion framework previously applied to cells freely migrating on 2D substrates [10, 11]. However, it remains unclear how the underlying shape changes control the overall migration dynamics of confined cells. Here, we use time-lapse phase-contrast microscopy and fluorescent staining of the cell nuclei to investigate the joint dynamics of cell shape and nucleus motion. Interestingly, we find that the motion of the cell nucleus is correlated with the growth of a cell protrusion across the constriction of the pattern (Fig. 1a), suggesting that the protrusion dynamics of these cells are key to understanding cell migration dynamics.

To quantify these protrusive dynamics, we first isolate cell shapes from bright-field microscopy image stacks using a convolutional neural network with a U-Net architecture [36] (Methods). This segmentation procedure allows us to accurately determine the 2D shape of the cells as a function of time (Fig. 1a). To identify protrusions, we classify those components of the cell shape added in each time step as protrusive areas (Fig. 1b,c, Methods, Supplementary Movie S3) [37]. As expected, we find that during the transmigration of cells across the constriction, protrusive areas are located at the growing tip of the cell. Importantly, due to the geometry of the micropattern, we find that most of the protrusive activity is captured by the x-direction along the long axis of the pattern (Supplementary Section S2). To provide a low-dimensional representation of the protrusion dynamics, we therefore define the effective protrusion position $x_\mathrm{p}(t)$ as the geometric center of protrusive area (Fig. 1c), which we refer to as the protrusion from here on. The trajectories of the protrusion serve as an indicator of the protrusive dynamics of the cells, as shown by an overlay with the kymograph of the microscopy images (Fig. 1d, Supplementary Movie S3). Thus, our analysis pipeline gives access to a large data set of low-dimensional trajectories of cell nucleus and protrusion dynamics, allowing in-depth statistical analysis of the cellular dynamics.

The joint nucleus and protrusion trajectories reveal that these cells migrate across the constriction in a stereotypical manner: first, the protrusion grows slowly across the constriction, after which the nucleus rapidly follows (Fig. 1e). Interestingly, the nucleus motion is less stochastic than the protrusions. The nucleus trajectory $x_\mathrm{n}(t)$ furthermore appears to be a time-lagged version of the protrusion trajectory $x_\mathrm{p}(t)$. This feature is also represented in the joint probability distribution of nucleus and protrusion positions, $p(x_\mathrm{n}, x_\mathrm{p})$, which exhibits a characteristic ring-like structure (Fig. 1f). While the most likely states are where both nucleus and protrusion occupy one adhesive island, there is significant probability along the path where the protrusion first crosses the constriction and reaches the other island, followed by the traversal of the nucleus. In contrast, there is very low probability of observing both protrusion and nucleus in the constriction. To further quantify these time-lag dynamics, we calculate the cross-correlation function $\langle x_\mathrm{n}(t) x_\mathrm{p}(t+T)$ which exhibits a peak at negative time shifts, indicating that the protrusion leads the nucleus by a typical time-shift $T_\mathrm{np} \approx 0.6$ h (Fig. 1g). While the cells also perform retractions, we find that these are strongly correlated with the motion of the nucleus with near-zero time-lag, and therefore do not contain significant additional information (Fig. 1g, Supplementary Section S2). Thus, we here restrict our analysis to the protrusion dynamics. To-

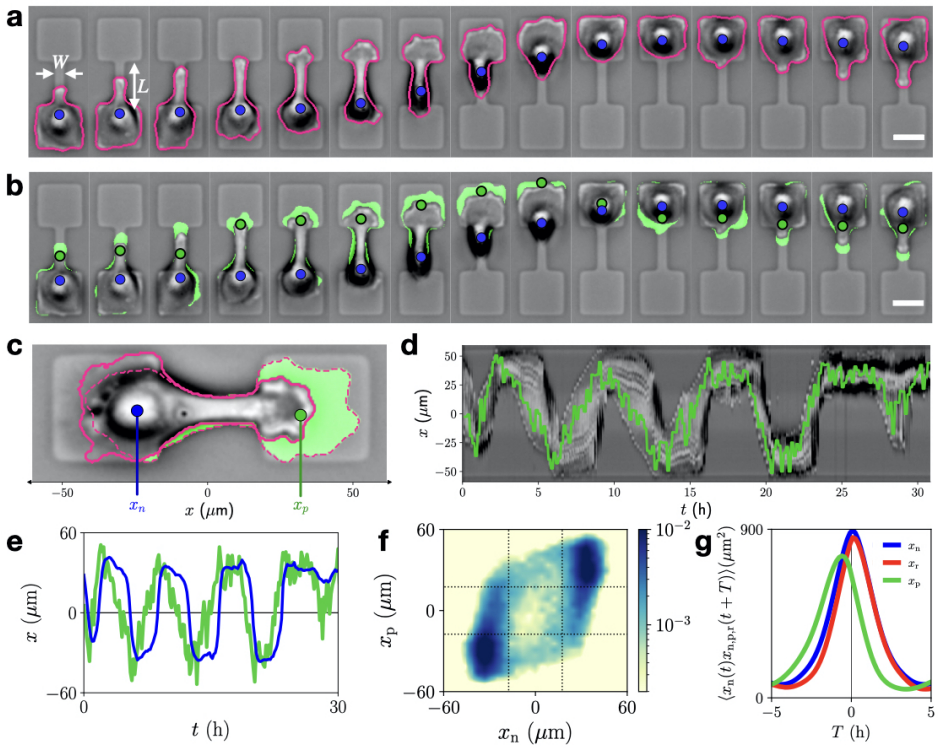


Figure 1: **Protrusion-nucleus cycling in confined migration. a.** Exemplary brightfield microscopy image series of an MDA-MB-231 breast cancer cell migrating in a two-state micropattern with constriction width $W = 7$ µm and length $L = 35$ µm, indicated by arrows. Images are inverted for better visibility of cell shapes. Tracked cells shapes are shown as pink outlines. Blue dots indicate tracked nucleus position x_n. **b.** Same time-series as in **a**, with protrusive areas marked in green, blue dot indicates the nucleus position x_n, green dot effective protrusion position x_p. **c.** Example image showing how protrusion areas are calculated and projection of nucleus and protrusion positions on the x-axis at time t. The solid pink line shows the current boundary of the cell area at time t, and the dashed line is the boundary at $t + \Delta t$. **d.** Kymograph of the brightfield microscopy images, with superimposed protrusion trajectory $x_\mathrm{p}(t)$ in green. **e.** Joint trajectory of nucleus $x_\mathrm{n}(t)$ (blue) and protrusion $x_\mathrm{p}(t)$ (green). **f.** Joint probability distribution $p(x_\mathrm{n}, x_\mathrm{p})$ of the x-positions of nucleus and protrusion, plotted logarithmically and interpolated. Dotted lines indicate the boundaries of the adhesive islands. **g.** Position cross-correlations between nucleus and protrusion $\langle x_\mathrm{n}(t)x_\mathrm{p}(t+T)\rangle$ (green), and nucleus and retraction $\langle x_\mathrm{n}(t)x_\mathrm{r}(t+T)\rangle$ (red). The retraction position x_r is determined in a similar way to the protrusion (Supplementary Section S2). Blue line shows the nucleus position auto-correlation, $\langle x_\mathrm{n}(t)x_\mathrm{n}(t+T)\rangle$. All scale bars: 25 µm.

gether, these results indicate that the cellular dynamics in this system exhibit a stereotypical 'protrusion-nucleus cycling' which can be represented as paths in $x_\mathrm{n}x_\mathrm{p}$-space. This data set could thus allow insight into the crosstalk between protrusion-nucleus coupling and the feedback between geometry and polarity dynamics. To elucidate these features, we aim to develop a mechanistic physical model for the stochastic dynamics of cell protrusions and polarity and their coupling to the nucleus dynamics.

Confined cells migrate in an adhesiveness landscape

We aim to develop a mechanistic theory to describe how the coupled stochastic dynamics of cell nucleus and protrusion determine the confined migration of cells. To this end, we start with a

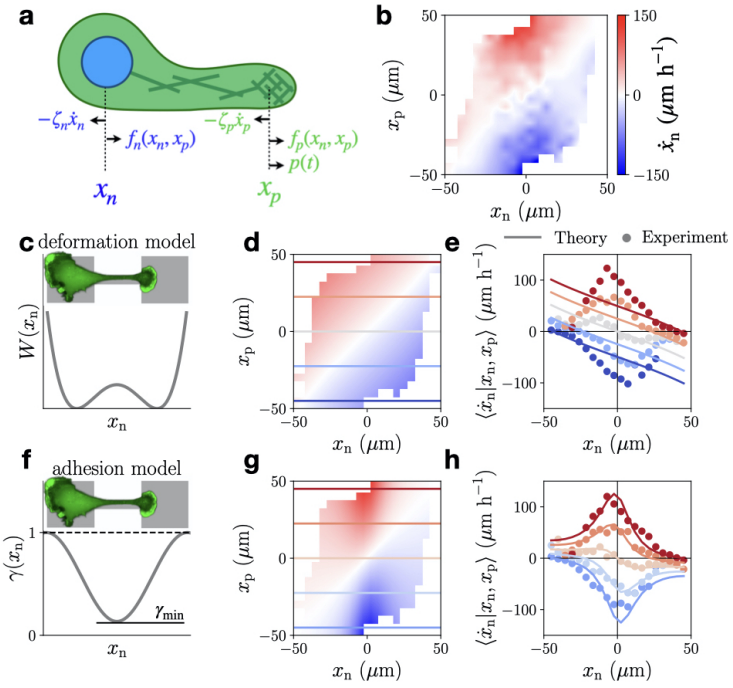


Figure 2: **Nuclear velocity maps constrain model candidates. a.** Schematic of the model. Arrows indicate the forces acting on the cell at positions x_n and $\dot{x}_p$. **b.** Experimental nuclear velocity map (NVM), calculated as the conditional average of the nucleus velocity as a function of nucleus and protrusion positions, $\langle \dot{x}_n | x_n, x_p \rangle$, shown with interpolation. **c.** Double-well potential $W(x_n) = Q(1-(x_n/x_0)^2)^2$ used in the deformation model, where Q determines the height of the energy barrier, and x_0 the positions of the minima. Image indicates the dimensions of the micropattern, and shows a fluorescence microscopy image of a LifeAct-stained cell in the pattern. **d.** NVM predicted by the deformation model. Parameters are determined by a best fit to the experimental NVM (Supplementary Section S4). **e.** Cuts of the NVM along the horizontal lines indicated in panel **d**, showing $\dot{x}_p$ as a function of x_n for different x_p. Dots: Experiment, solid lines: deformation model prediction. **f.** Spatially variable friction $\gamma(x_n) = \gamma_{min} + \frac{1}{2}(1-\gamma_{min})\left(1-\cos\left(x_n \pi / L_{system}\right)\right)$ used in the adhesion model, where γ_{min} is the friction at the center of the constriction, and L_{system} is the total length of the micropattern. **g.** NVM predicted by the adhesion model. Parameters are determined by a best fit to the experimental NVM (Supplementary Section S4). **h.** Same plot as in panel **d** for the adhesion model.

general set of equations of motion for the nucleus and protrusion dynamics, which we systematically constrain using the experimental data. Our strategy will be to postulate simple model candidates based on known cellular processes, physical intuition, and symmetry arguments, whose predictions we will test against data.

Forces on the nucleus can arise due to two main contributions: coupling to the cell protrusion [6, 38–40], and the effect of the confining micropattern. Similarly, protrusions couple to the cell nucleus [6], and may be sensitive to the external environment. Additionally, we assume the protrusion to be driven by a stochastic active force $P(t)$, which serves as a minimal implementation of the time-dependent forces driving protrusion formation, such as the active pushing force due to actin polymerization [5, 31]. This active force determines the instantaneous direction of polarization in which protrusions are generated, and we therefore refer to it as the cell polarity. Thus, considering force balance at x_n and x_p (Fig. 2a), we obtain the general equations

$$\zeta_n \dot{x}_n = f_n(x_n, x_p) \tag{1}$$

$$\zeta_p \dot{x}_p = f_p(x_n, x_p) + P(t) \tag{2}$$

where ζ_n, ζ_p are the friction coefficients of nucleus and protrusion, respectively. In this model, we assume the origin of the intrinsic stochasticity of the system to be the polarity dynamics driving the protrusion. In contrast, the nucleus dynamics are given as a deterministic function of x_n and x_p. These assumptions are supported by the experimental data: we find that the nucleus velocities $\dot{x}_n$ have a signal-to-noise ratio > 1, indicating that deterministic components dominate over stochastic components. In contrast, the stochastic contribution to the protrusion velocities $\dot{x}_p$ is much larger than that of the nucleus, and the protrusion velocities have a signal-to-noise ratio < 1 (Supplementary Section S3).

To constrain our model step-by-step, we start with the dynamics of the cell nucleus given by $f_n(x_n, x_p)$. In migrating cells, the motion of the nucleus is coupled to the dynamics of the leading edge, for example through elastic stresses in the cytoskeleton which connects protrusion and nucleus [6, 38, 39, 41], or through active processes coupling the leading and trailing edge of the cell [42]. As a minimal model for this coupling, we consider a linear elastic spring, similar to previous work [32, 33].

It is less clear however, how to incorporate the effect the confining micropattern might have on the dynamics. A key step in the migration across the constriction is the deformation of the cell body near the nucleus (Fig. 1a, Supplementary Movie S2). Such deformation dynamics of cells is frequently modelled using elastic Hamiltonians including the surface and line tension of the cell [24, 25, 43–45], which would suggest that the deformed state of the cell in the constriction is associated with an increased energy. A simple description for this confinement contribution is a double-well potential $W(x_n)$, where the cell has minimal energy when it is on the adhesive islands (Fig. 2 c). We therefore postulate the equation of motion for the cell nucleus

$$\zeta_n \dot{x}_n = k(x_p - x_n) - \partial_{x_n} W(x_n) \tag{3}$$

This equation makes a concrete prediction for how the nucleus velocity $\dot{x}_n$ varies with the positions of nucleus and protrusion. To test this prediction directly on the experimental data, we determine the average velocity of the cell nucleus as a function of x_n and x_p, $\langle \dot{x}_n | x_n, x_p \rangle$, which we term *nuclear velocity maps* (NVM) (Fig. 2 b). Strikingly however, we find that the NVM predicted by this model qualitatively fails to capture the experimental data (Fig. 2 d,e). Notably, the model (Eq. (3)) predicts that the speed of the nucleus should decrease as it crosses the constriction when the protrusion has reached the other side. In contrast, in the experiment we find that the nucleus accelerates as it crosses the constriction. The model similarly fails for more general non-linear elastic couplings between nucleus and protrusion (Supplementary Section S3). Therefore, we conclude that such potential energy models alone are not able to recover the cellular dynamics in this setup.

A second feature of the confinement that may play a role in the migration dynamics is the difference in adhesive area available to the cell on the island and in the constriction. Mesenchymal migration relies on the formation of mature focal adhesions at the cell rear, where the nucleus typically resides [46, 47]. These adhesions can only form within the micropatterned area, and we therefore expect the adhesiveness to be largest on the islands and smallest at the center of the constriction. A simple model for such differential adhesiveness is a spatially variable friction coefficient (Fig. 2f):

$$\zeta_n \gamma(x_n) \dot{x}_n = k(x_p - x_n) \tag{4}$$

Remarkably, this model provides an excellent fit to our data, and captures the characteristic increase in nucleus speeds during transmigration (Fig. 2 g,h). This observation suggests that the spatially variable adhesiveness plays a dominant role in the confinement dynamics. The acceleration of the nucleus could therefore be understood as a consequence of the reduced number

of adhesions formed by the cell body around the nucleus when it is in the constriction. Importantly, by constraining our model purely based on the nucleus velocities, we can constrain the nucleus dynamics (Eq. (1)) without making any assumptions on the protrusion and polarity dynamics (Eq. (2)). Taken together, these results indicate that a spatially variable adhesiveness is a key component of the effect of the thin constriction on migration dynamics, which dominates over possible contributions due to cellular deformations in our setup.

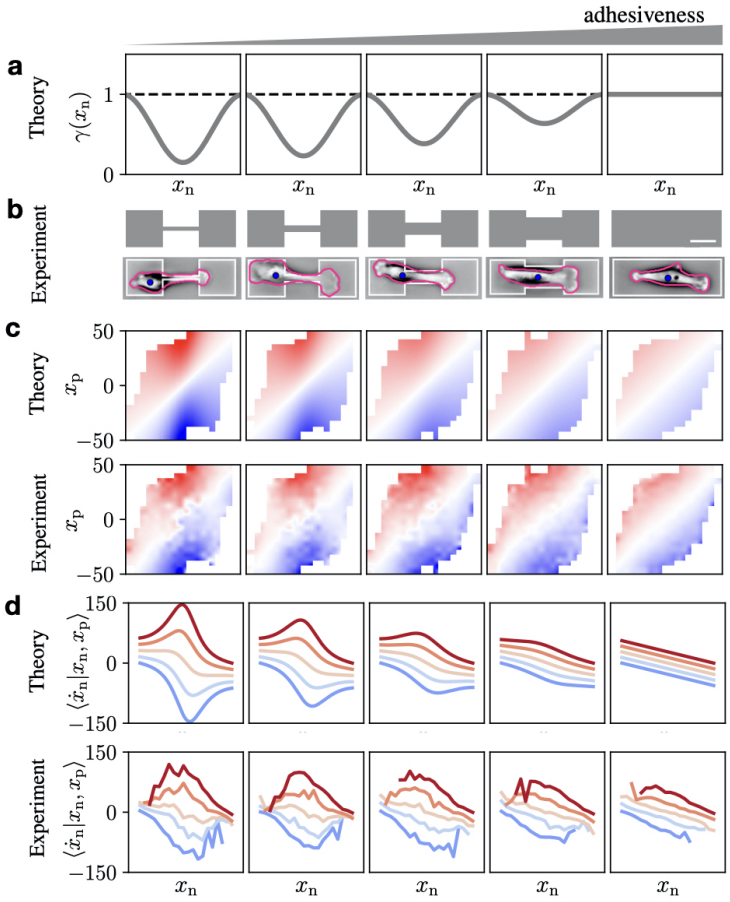


Figure 3: **Adhesion model predicts dynamics with varying constriction width. a.** Friction profiles $\gamma(x_n)$ with increasing γ_{min} as a model for increasing bridge width. The value of γ_{min} is fitted for the narrowest bridge ($W = 4$ µm. For the widest system without constriction, we take a flat profile and interpolate the profiles of intermediate widths (Supplementary Section S4). **b.** Sketch of confinement geometries with increasing bridge widths $W = 4, 7, 12, 22, 35$ µm (from left to right); brightfield microscopy images of MDA-MB-231 cells migrating in these geometries with cell outline in pink and nucleus position in blue, and geometry in white. Scale bar: 25 µm. **c.** Predicted and experimental nuclear velocity maps (NVM) $\langle \dot{x}_n | x_n, x_p \rangle$. Plotted with the same color axis as shown in Fig. 2. **d.** Predicted and experimental cuts of the NVM, showing $\dot{x}_p$ as a function of x_n for different x_p as described in Fig. 2.

Adhesion model captures dependence of nucleus dynamics on varying constriction width

Our adhesion model (Eq. (4)) makes a simple, intuitive prediction. As we widen the constricting bridge of the micropattern, more adhesive area becomes available, and thus the variations in

the friction profile should become smaller (Fig. 3a). In the extreme limit where the constriction has the same width as the islands, we expect a uniform adhesiveness profile. Accordingly, we predict the acceleration of the cell nucleus observed on thin bridges (Fig. 2) to decrease with the increasing adhesiveness of a wider bridge, and to completely disappear for constant adhesiveness (Fig. 3c,d). In this case, we expect dynamics that are completely determined by the linear elastic coupling between nucleus and protrusion (last panel Fig. 3d).

To challenge the predictive power of the adhesion model, we perform experiments with varying bridge width (Fig. 3b). Remarkably, we find that the nuclear velocity maps inferred from these experiments exhibit a similar trend as predicted by the adhesion model, with a decreasing maximum nucleus speed in the constriction (Fig. 3c,d). On the rectangular micropattern without any constriction, we find an almost linear profile of the nucleus speed with position, as predicted theoretically. This further supports our model of the nucleus-protrusion coupling as a linear elastic spring. In summary, the adhesion model has predictive power for confining geometries with varying constriction width.

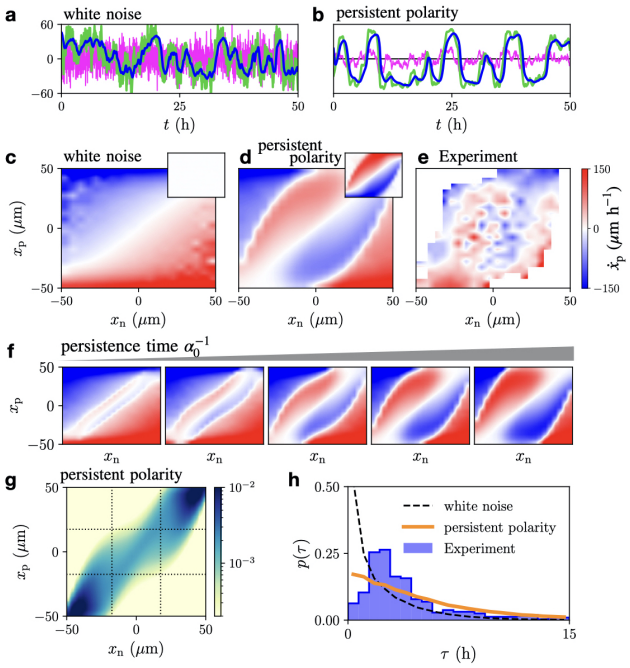


Figure 4: **Dynamics of time-correlated polarities. a.** Stochastic trajectory $x_n(t)$ (blue), $x_p(t)$ (green), and $P(t)$ (pink) of the white noise model. **b.** Trajectory of the persistent polarity model (Eq. (7)). **c,d.** Protrusion velocity maps (PVM) $\langle \dot{x}_p | x_n, x_p \rangle$ predicted by the white noise model, and the persistent polarity model. In both models, we use a potential to enforce the overall system boundaries, $V(x_p) = (x/x_{boundary})^8$ (see Supplementary Section S4). *Inset:* polarity contribution to the PVM, given by $\langle P | x_n, x_p \rangle$. **e.** PVM inferred from experiments with bridge width $W = 7$ µm. **f.** PVMs predicted by the persistent polarity model for varying polarity persistence time α_0^{-1}. From left to right, persistence times $\alpha_0^{-1} = 0.1, 0.2, 0.5, 1, 3$ h are shown. **g.** Joint probability distribution $p(x_n, x_p)$ predicted by the persistent polarity model, plotted logarithmically. Dotted lines indicate the boundaries of the adhesive islands. **h.** Probability distribution of the dwell times τ, defined as the time between subsequent transitions of the cell nucleus across the bridge centre; observed experimentally ($W = 7$ µm, blue), and predicted by the white noise (black dashed line) and the persistent polarity model (orange).

Observed protrusion dynamics suggest time-correlated cell polarity

Having determined how the dynamics of the nucleus couples to the confinement and protrusion, we next investigate the dynamics of the protrusion itself (Eq. (2)). As a minimal model for the protrusion dynamics, we postulate a coupling to the cell nucleus which is equal and opposite to the coupling introduced for the cell nucleus (Eq. (4)). In addition, we enforce a confinement to the overall system size, using soft-wall boundary conditions enforced by a potential $V(x_\mathrm{p})$:

$$\zeta_\mathrm{p}\dot{x}_\mathrm{p} = -k(x_\mathrm{p} - x_\mathrm{n}) - \partial_{x_\mathrm{p}} V(x_\mathrm{p}) + P(t) \tag{5}$$

Thus, we assume that the protrusion is insensitive to the presence of the constriction and is able to move through it unhindered without variable adhesiveness. This is also supported by the observation that the protrusion, unlike the nucleus, spends significant time in the constriction, as shown by the probability distribution $p(x_\mathrm{n}, x_\mathrm{p})$ (Fig. 1f). However, we anticipate that the polarity P may couple to the confinement, as it models the active driving of the protrusion by the molecular migration machinery, including actin polymerization and the diffusion of polarity cues [5, 31], which may be sensitive to the geometry of the confinement. Furthermore, we also expect that the polarity may exhibit time-correlations to account for the spatiotemporal dynamics of the migration machinery.

Similar to our approach to the nucleus dynamics, this model provides a prediction for the average protrusion velocity as a function of x_n and x_p, which we term *protrusion velocity map* (PVM) $\langle \dot{x}_\mathrm{p} | x_\mathrm{n}, x_\mathrm{p} \rangle$. According to our general model ansatz (Eq. (2)), unlike the NVM, the PVM consists of two components, one due to protrusion-nucleus coupling, and the other due to the polarity dynamics:

$$\langle \dot{x}_\mathrm{p} | x_\mathrm{n}, x_\mathrm{p} \rangle = f_\mathrm{p}(x_\mathrm{n}, x_\mathrm{p}) + \langle P | x_\mathrm{n}, x_\mathrm{p} \rangle \tag{6}$$

Since the polarity term $\langle P | x_\mathrm{n}, x_\mathrm{p} \rangle$ does not average to zero for time-correlated polarities, we cannot in general infer the function $f_\mathrm{p}(x_\mathrm{n}, x_\mathrm{p})$ directly from the PVM [48]. However, we can postulate simple models for the polarity dynamics and explore their predictions for the PVM and other statistics.

We first show that the data cannot be captured by the simplest possible stochastic polarity dynamics: a Gaussian white noise process $P = \sigma\xi(t)$ with $\langle \xi(t) \rangle = 0$ and $\langle \xi(t)\xi(t') \rangle = \delta(t - t')$ (Fig. 4a). In this case, the polarity has no time-correlations, and thus $\langle P | x_\mathrm{n}, x_\mathrm{p} \rangle = 0$ (Inset Fig. 4c), and we directly recover the expected contractile elastic coupling (Eq. (5)) in the PVM (Fig. 4c). In striking contrast, the PVM inferred from experiments shows a non-trivial dependence of the protrusion velocities as a function of x_n and x_p (Fig. 4d). This dependence cannot be accounted for by the white noise model. The overall structure of the experimental PVM is in line with a contractile coupling between nucleus and protrusion: it exhibits negative velocities in the upper diagonal and positive velocities in the lower diagonal. These features correspond to the protrusion being pulled back towards the nucleus. Interestingly, however, when the protrusion extends into the bridge, the protrusion velocity takes the opposite sign, corresponding to an unexpected extensile average driving force, pushing the protrusion away from the nucleus. This 'polarity driving' is a striking observation, since it cannot be accounted for even by a non-linear contractile coupling to the nucleus. These results indicate that the simple white noise model does not capture our experimental data. In fact, we can rule out a white noise polarity more generally, using the most general functions f_n and f_p inferred directly from experiment under the white noise assumption. This approach similarly fails to capture the experimental statistics, allowing us to rule out the possibility that the protrusion is driven by a white noise process (Supplementary Section S3). This suggests that to account for the experimentally observed dynamics, we need to account for time-correlations in the cell polarity.

To investigate how time-correlated polarity dynamics affect the behavior, we begin with the simplest choice of an exponentially correlated polarity governed by the equation

$$\dot{P} = -\alpha_0 P + \sigma\xi(t) \tag{7}$$

which we term the *persistent polarity model*; with $\alpha_0 > 0$. The polarity evolving according to Eq. (7) thus experiences negative feedback, $\dot{P} \propto -P$, and exhibits time-correlations decaying exponentially on a persistence time-scale α_0^{-1}. Interestingly, these polarity dynamics have significant correlations with the state of the system: the conditional average $\langle P|x_n, x_p\rangle$ no longer vanishes as in the white noise case (Inset Fig. 4d). Since the PVM is given by the sum of the polarity and the nucleus-protrusion coupling term (Eq. (6)), these correlations lead to a clear deviation from the linear elastic coupling in the PVM. Specifically, the predicted PVM exhibits a polarity driving similar to the experimental PVM (Fig. 4d,e). The correlation time-scale of the polarity α_0^{-1} determines the strength of the driving (Fig. 4f). Taken together, these results indicate that cell protrusions are driven by time-correlated polarity dynamics.

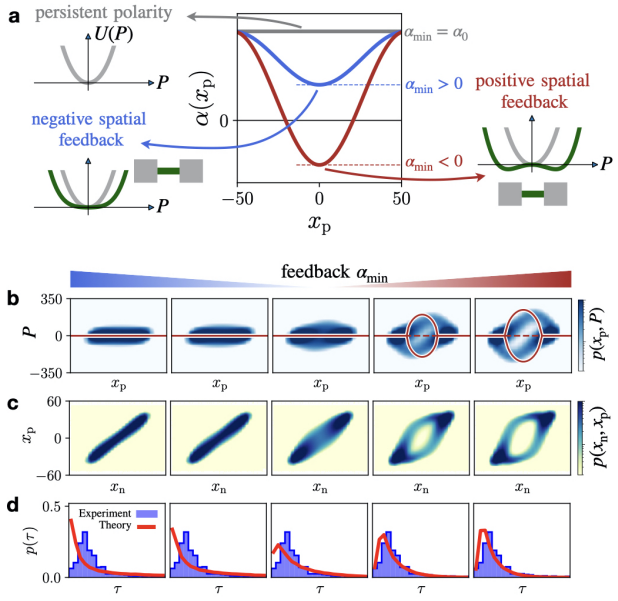


Figure 5: **Spatial feedback model. a.** Schematic illustration of the polarity models. The persistent polarity model corresponds to a flat feedback profile, $\alpha(x_p) = \alpha_0$, where the polarity is effectively confined to a constant harmonic potential $U(P)$ (left). In the spatial feedback model (Eq. (8)), α_{min} controls the sign of the feedback. For $\alpha_{min} > 0$, the feedback is negative and the protrusion confined to harmonic potentials with spatially varying stiffness (left). For $\alpha_{min} < 0$, the feedback locally becomes positive, leading to two stable fixed points (right). **b.** From left to right, we vary the value of $\alpha_{min} = \{6, 4, 0, -4, -6\}$ h^{-1}. The joint probability distribution of protrusion position and polarity $p(x_p, P)$ is shown. Solid red lines indicate the position of the stable fixed points, dashed red lines indicate unstable fixed points. **c.** Joint probability distributions $p(x_n, x_p)$ predicted by the spatial feedback model with varying α_{min}. **d.** Probability distribution of the dwell times τ predicted by the model (red) and observed experimentally ($W = 7$ µm, blue).

Constrictions trigger polarity self-reinforcement

While the persistent polarity model (Eq. (7)) captures the qualitative features of the protrusion velocities (Fig. 4b), it predicts stochastic dynamics that do not capture the key features of the experiment. Specifically, it fails to capture the stereotypical protrusion-nucleus cycling indicated by the ring structure in the experimental probability distribution $p(x_n, x_p)$(Figs. 1f, 4g).

Furthermore, the model does not capture the distribution of dwell times, defined as the time between subsequent transitions of the cell nucleus across the bridge centre. In the experiment, this distribution exhibits a marked peak at finite times, indicating that there is a typical timescale for re-orientation on the island and subsequent traversal between transitions. This feature is not captured by the persistent polarity model (Fig. 4e).

The persistent polarity model relies on several simplifying assumptions. Firstly, we assumed the polarity to be insensitive to the local confinement, as the polarity dynamics does not explicitly depend on the position of the protrusion. Secondly, we assumed negative feedback, $\dot{P} \propto -P$. This means that the polarity is effectively confined to a harmonic potential $U(P) = \alpha_0 P/2$ (Fig. 5a).

To relax these assumptions, we extend the persistent polarity model with a spatial feedback, where the strength and sign of the polarity feedback depend on the local geometry of the confinement. Thus, we propose that the feedback may vary with the position of the protrusion, $\alpha = \alpha(x_\mathrm{p})$. Physically, we expect that the polarity may become more persistent when the protrusion is in the constriction. Such an increase in persistence could be due to increased alignment of actin fibers [49–52], or more stable polarity cue gradients [31, 53, 54] when the protrusion is confined to a narrow constriction. To ensure that the polarity remains bounded, we include the next-order term allowed by symmetry $-\beta P^3$, with $\beta > 0$, and allowing $\alpha < 0$ locally:

$$\dot{P} = -\alpha(x_\mathrm{p})P - \beta P^3 + \sigma\xi(t) \tag{8}$$

To account for larger persistence in the constriction, we choose a feedback function $\alpha(x_\mathrm{p})$ which takes a minimal value α_min in the center of the bridge (Fig. 5a). If $\alpha_\mathrm{min} > 0$, the polarity dynamics exhibits a stable fixed point at $P = 0$ everywhere (Fig. 5a). In contrast, if $\alpha_\mathrm{min} < 0$, the polarity is still driven back to $P = 0$ on the islands, but in the constriction, a stable fixed point $P^* = \pm\sqrt{|\alpha|/\beta}$ emerges. Consequently, when the protrusion is in the constriction and the polarity is small ($P < P^*$), a positive feedback mechanism $\dot{P} \propto P$ is activated, which leads to self-reinforcement of the polarity into the current direction of polarization. Thus, the symmetry of the system is broken, which then features two preferred states of polarization in opposite directions.

We now explore the predictions of the spatial feedback model by varying α_min (Fig. 5b). As expected, for negative polarity feedback ($\alpha_\mathrm{min} > 0$), we find a polarity distribution $p(x_\mathrm{p}, P)$ with polarities centered around $P = 0$ at all positions x_p. Similar to the persistent polarity model, this model predicts a joint position distribution $p(x_\mathrm{n}, x_\mathrm{p})$ with no annular structure, and no peak in the dwell time distribution. Remarkably however, when the feedback becomes positive in the constriction, we observe the protrusion-nucleus cycling in the probability distribution $p(x_\mathrm{n}, x_\mathrm{p})$ and the model predicts a marked peak in the dwell time distribution, both in line with the experiment. These features emerge in conjunction with a ring-structure in the polarity distribution $p(x_\mathrm{p}, P)$, as polarities preferentially take finite values in the constriction. These findings suggest that to capture the stereotypical protrusion-nucleus cycling, we require a geometry-sensitive polarity feedback.

Predicting dynamics with varying constriction width

We challenge the predictive power of our model by testing its predictions for the stochastic trajectory dynamics in systems with increasing constriction width. Increasing constriction width has a clear implication for the spatial feedback model (Eq. (8)): in addition to the adhesiveness profile $\gamma(x_\mathrm{n})$ becoming flatter (Fig. 3a), we also expect the positive polarity feedback to

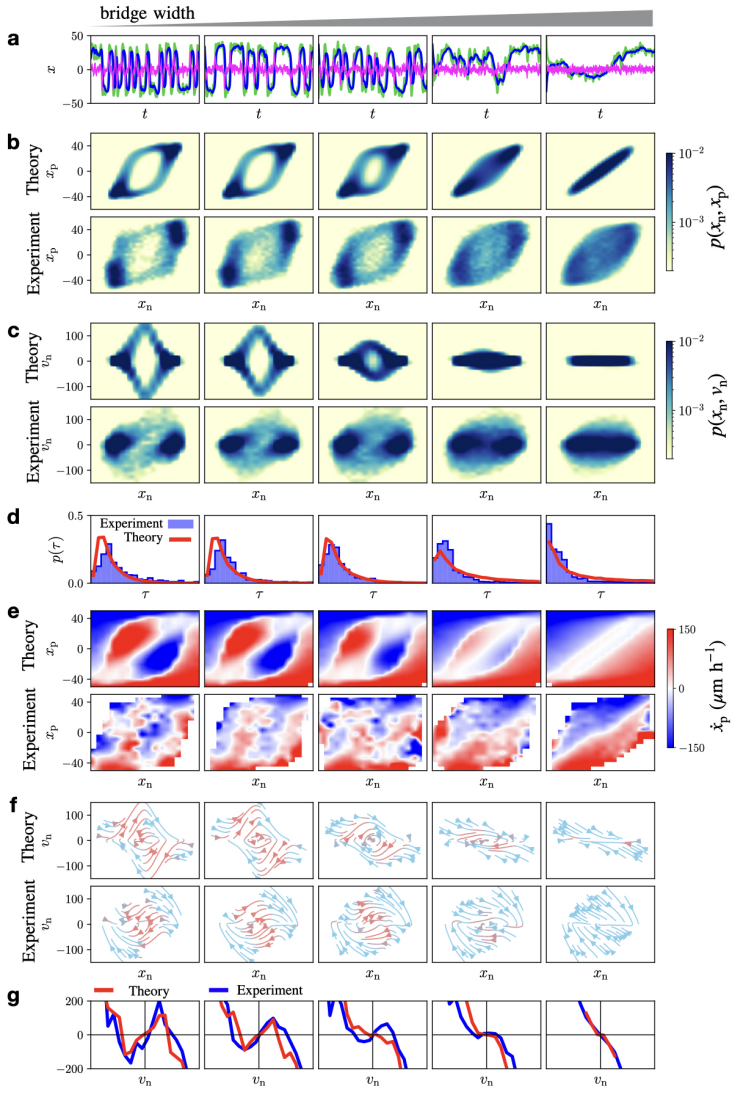


Figure 6: **Model predicts dynamics with varying constriction width. a.** Stochastic trajectory $x_n(t)$ (blue), $x_p(t)$ (green), and $P(t)$ (pink) predicted by the spatial feedback model (Eq. (8)) with increasing γ_{min} and α_{min} as a model for increasing bridge width W (Supplementary Section S4). **b.** Joint probability distributions $p(x_n, x_p)$. **c.** Joint probability distributions $p(x_n, v_n)$. **d.** Predicted (red) and experimental (blue) dwell time distributions $p(\tau)$. **e.** Protrusion velocity maps (PVM) $\langle \dot{x}_p | x_n, x_p \rangle$. **f.** Flow field $(\dot{x}_n, \dot{v}_n) = (v_n, F(x_n, v_n))$ indicated by arrows. Arrow color indicates the direction of the local flow: acceleration is orange and deceleration is blue. **g.** Predicted (red) and experimental (blue) effective friction at the bridge center $F(x_n \to 0, v_n)$. In panels **b, c, e, f**, the top row corresponds to the spatial feedback model prediction, the bottom row to experimental observations. Experimental observations correspond to $W = 4, 7, 12, 22, 35$ µm (from left to right).

diminish. Remarkably, based on these assumptions, this model predicts that the stereotypical protrusion-nucleus cycling observed in systems with thin constrictions gradually disappears with increasing width (Fig. 6b). Specifically, we observe that the ring-structure in the position probability distribution $p(x_n, x_p)$ gradually closes, with a uniform distribution in the limiting case of a flat profile. Simultaneously, the predicted acceleration of the nucleus in the constriction decreases: on narrow constrictions, an annular probability distribution of nucleus position

and speed $p(x_n, v_n)$ is predicted, which gradually closes for wider bridges (Fig. 6c). Interestingly, as the stereotypical migration pattern in these probability distributions disappears, we also predict that the typical transition time-scale, indicated by the peak in the dwell time distribution, disappears (Fig. 6g). These observations can be understood by examining the protrusion dynamics in the model: we find that the predicted PVM exhibits a polarity driving of decreasing magnitude with increasing bridge width (Fig. 6d). Notably, the driving disappears in the widest system with no constriction, where the protrusion velocities are thus determined by the elastic nucleus-protrusion coupling. This coupling pulls nucleus and protrusion together, inhibiting the stereotypical cycle. In summary, our model suggests that the weaker polarity self-reinforcement expected for wider constrictions leads to the disappearance of the protrusion-nucleus cycling. The model therefore suggests that these stereotypical cycles rely on the adaptation of the cell polarity dynamics to its local environment. Remarkably, we find that these predicted features of the dynamics are confirmed by the experimental data of cells migrating in pattern with increasing width (Fig. 6a-d).

We further test the validity of our model by exploring geometries with varying the constriction length L. We find that our model also captures the changes in dynamics observed in this case (Supplementary Section S3). Taken together, these results indicate that our model has striking predictive power beyond the specific confinement geometry which we used to constrain it.

Mechanistic model captures emergent nonlinear dynamics

A central challenge for our mechanistic theoretical approach is to capture the emergent long time-scale stochastic dynamics of the system. Based on the trajectories of the cell nucleus alone, we showed in previous work [15], that the dynamics of these trajectories can be described by an underdamped stochastic equation of motion: $\dot{v}_n = F(x_n, v_n) + \sigma(x_n, v_n)\eta(t)$ where $\eta(t)$ is Gaussian white noise. This model represents an effective description of the cellular dynamics, with no direct connection to interpretable cellular degrees of freedom such as the protrusion and polarity, which we consider here. The deterministic contribution to these dynamics $F(x_n, v_n)$ exhibits intricate non-linear dynamics, which can be represented in a phase-space portrait (Fig. 6e). This analysis revealed that the nucleus deterministically accelerates into the thin constriction, and therefore exhibits a deterministic amplification of velocity (orange arrows Fig. 6e). This effect is driven by the emergence of an effective 'negative friction' at the center of the thin constriction. Specifically, at low speeds, the nucleus accelerates with increasing velocity, such that $F(v_n) \propto v_n$ locally (Fig. 6h) [15]. These observations encapsulate the emergent long time-scale stochastic dynamics of the system, yet a mechanistic model that can capture these dynamics has thus far remained elusive.

Remarkably, we find that our model predicts a phase-space flow similar to that inferred from experiments (Fig. 6e). Our mechanistic model for the coupling of the cell nucleus motion to the cellular protrusion and polarity dynamics (Eqns. (4), (5), (8)) therefore recovers the effective dynamical description for the stochastic dynamics of the nucleus alone. Interestingly, we can directly map the equations of motion of the mechanistic model to the effective underdamped nucleus dynamics, revealing that the observed amplification behavior is a consequence of two combined effects: lower adhesiveness and enhanced polarity persistence in the constriction (Supplementary Section S3). Accordingly, our model predicts that when the constriction is widened, corresponding to higher adhesiveness and lower polarity persistence, the amplification of nucleus velocities gradually disappears (Fig. 6e). This is indeed what we observe experimentally. Interestingly, we find that within the experimentally relevant parameter regimes, the observed effective anti-friction in the constriction only emerges when positive feedback is acti-

vated, independent of the differential adhesiveness (Supplementary Section S3). Consequently, we predict that the friction relation changes from effective anti-friction to regular friction with widening constrictions, which is confirmed by the experiments (Fig. 6f). In summary, we find that the mechanistic model for the nucleus, protrusion, and polarity degrees of freedom of the cell gives insights into the origin of the nonlinear dynamics inferred directly from the nuclear dynamics.

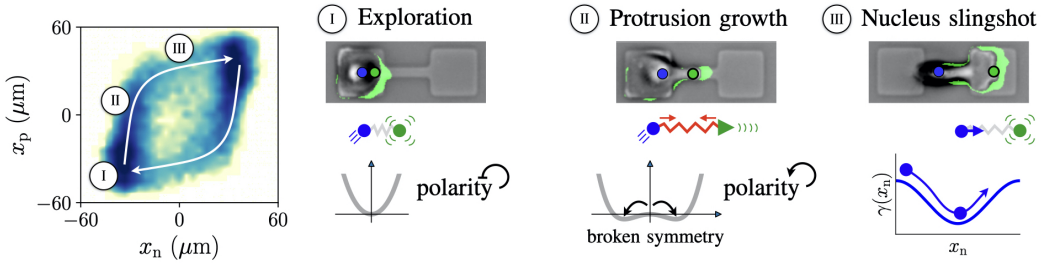


Figure 7: **Three stages of protrusion-nucleus cycling.** In the joint position probability distribution of nucleus x_n and protrusion x_p (left, same color code as in Fig. 1f), we indicate the typical evolution of the system with white arrows, and identify three distinct stages of the process. Typical brightfield microscopy images with overlayed protrusive areas, and the positions x_n and x_p indicated are shown for each of the three stages. Schematics indicate the physical mechanisms that dominate each phase according to our theoretical framework.

Discussion

In this work, we develop a mechanistic theoretical framework to describe the joint stochastic dynamics of cell nucleus, protrusion, and polarity, and their coupling to the extracellular microenvironment. Experimentally, we find that cells migrating in confinements with a thin constriction exhibit a stereotypical protrusion-nucleus cycling, where the protrusion grows to precede the transition of the nucleus across the bridge. Using a large data set of joint protrusion and nucleus trajectories, we systematically constrain a mechanistic model for confined cell migration. In our model, we identify three distinct stages of the protrusion-nucleus cycling (Fig. 7). First, we observe an initial exploration phase, where both nucleus and protrusion are located on the same island (Stage I). At this stage, the polarity is subject to negative feedback, causing the protrusion to frequently change direction and explore its surroundings. Stochastic polarity excitations can trigger the protrusion to penetrate the constriction. Within the constriction, the protrusion becomes highly confined, causing the polarity dynamics to switch from a negative to a positive feedback loop. This positive feedback reinforces the polarity, driving the protrusion growth into the constriction (Stage II). At the same time, tension builds up due to the elastic coupling to the nucleus, which is held back on the island due to the enhanced adhesion with the substrate. Once the protrusion reaches the other end of the system, the nucleus is pulled across the constriction, relaxing the tension in the elastic coupling, reminiscent of a slingshot (Stage III). The three stages of the transition process emerge as a consequence of the interplay of the three key physical mechanisms in the system: the adhesiveness landscape, the nucleus-protrusion coupling, and the polarity self-reinforcement.

To constrain our theoretical approach, we first studied the stochastic dynamics of the cell nucleus. We show that the dynamics of the nucleus are inconsistent with movement in a simple double-well potential, as might be expected from cellular deformation arguments. The deformation dynamics of cells is frequently modelled using elastic Hamiltonians [24,25,43–45] which would suggest a higher energy of the cell membrane associated with the stretched shape of the

cell when it is in the constriction compared to the compact shape on the islands. Indeed, active particles confined to double-well potentials can exhibit excitable dynamics similar to those observed in the experimental trajectories of the nucleus alone [55], making the double well a promising model candidate. However, based on the observed joint dynamics of nucleus and protrusion, we find that this energy barrier model is unable to capture the experiment, suggesting that it is not the dominating effect in the dynamics of the cell nucleus.

Instead, our model suggests that the movement of the nucleus is determined by the locally available adhesive area, which manifest as a spatially variable friction coefficient. We therefore interpret the observed speed increase of the nucleus in the constriction as a consequence of the differential adhesiveness of the pattern. This is in line with experimental observations showing that in mesenchymal migration, the movement of the cell rear, where the nucleus is typically located, is limited by the unbinding of mature adhesions [56, 57]. The dependence on local adhesive area is also consistent with our previous observation that cells spend less time on smaller adhesive islands [16]. The adhesion-limited nucleus motion is reminiscent of stick-slip processes that have been observed in cell migration on 1D lines [32, 34, 58], which are a consequence of the mechanosensitive binding and unbinding dynamics of adhesions [32, 33]. Our analysis suggests that the constriction enforces a 'spatially structured stick-slip process', that is not only driven by the internal dynamics of the cell as is the case in unstructured systems [32], but has time- and length-scales externally enforced by the confinement. Specifically, as the protrusion grows across the bridge, the back of the cell 'sticks' due to the high adhesiveness on the island. Once the protrusion has crossed the bridge, the elastic coupling to the cell nucleus causes the 'slip', or contraction stage, where the cell quickly contracts and the nucleus moves across the bridge with high speed. In the model, the tension in the elastic coupling rapidly relaxes during the slip phase, similar to a slingshot. Such 'slingshot' dynamics have also been observed in confined 3D migration in fibrous matrices [59].

In addition to the spatially variable adhesiveness, we find that a crucial component of the protrusion-nucleus dynamics is the geometry-dependent dynamics of the cell polarity. Remarkably, we find that we can capture all key features of the data by allowing the polarity dynamics to sense the local geometry by switching from a negative to a positive feedback loop when the protrusion enters the constriction. Thus, this position-dependent polarity feedback model posits that the cell polarity dynamics adapts to the presence of the constriction by activating a self-reinforcing feedback loop. This positive feedback leads to a broken-symmetry state, in which there is a non-zero preferred polarity. Previous work on unconfined 2D migration has suggested a similar symmetry breaking mechanism in the cell velocity dynamics [30]. However, in that case, this state emerged for particular cell parameters, leading to a long-lived state of spontaneous polarization. In contrast, our work suggests that such states may also emerge as a consequence of adaptation to an external confinement.

An alternative model to the position-dependent feedback are polarity dynamics which are insensitive to the external geometry, but depend on the extension of the cell. An interesting aspect of comparing these two models are their conceptually distinct implications. The position-dependent feedback implies a direct coupling to the external environment, where the cell may sense and adapt to the external geometry. In contrast, an extension-dependent feedback is translationally invariant, meaning that the polarity dynamics do not depend upon the absolute position of the cell within its environment, but only on its internal state. Such a model has previously been proposed for protrusion formation in keratocytes, where the distance from the cell center to the protrusion was suggested to be decisive for protrusion growth [60]. Interestingly however, we find that an extension-dependent polarity model robustly fails to account for the observed changes in the dynamics with widening constrictions (Supplementary Section S3). In contrast, the spatial feedback model captures the changes in dynamics with both changing

width and length of the constriction. These results suggest that the positive feedback loop in the polarity is a response to the geometry of the local microenvironment rather than to the overall extension of the cell.

There are a number of cellular processes that could contribute to the coupling of polarity to local geometry. First, based on the physics of active gels, which describe, for example, the actomyosin cortex in the protrusion, we expect a greater degree of alignment of actin fibers in a narrow constriction [49, 50]. Increased alignment of actin is associated with higher contractility [61, 62] and the emergence of spontaneous cell polarization [51, 52, 63–65]. A further key determinant of cell polarization are diffusable polarity cues [9, 31, 66], whose spatiotemporal organization may couple to external geometries, for example through focal adhesions [67], or the cell shape itself [53, 54]. Finally, changes in membrane tension may also provide feedback to the cell polarity [33, 68]. Our theoretical framework provides a new way to analyze and interpret the joint nucleus and protrusion dynamics of migrating cells, and to connect observed features in experimental trajectories to underlying mechanisms. Thus, our work could provide a tool to analyze the dynamics of cells in experiments where these distinct contributions to the polarity dynamics are targeted through molecular perturbations.

Based on experiments in which we varied the dimensions of the constriction, we found that our model has predictive power beyond the specific confinement geometry used to constrain it. Thus, this model could potentially be used to make predictions for new experiments, such as cell migration on patterned lines [35, 70–72]. Previous work has investigated the effect of differential protein patterning on 1D lines [35, 70], and the effect of asymmetric periodic patternings, which led to a rectification of cell migration in one direction [71–73]. This rectification has been interpreted to be a consequence of the asymmetry in locally available adhesive area [74], consistent with our findings. Our work suggests that the regulation of cell polarity may also play an important role in such processes. Furthermore, migrating cells face structured surroundings and thin constrictions in 3D extra-cellular matrices [4, 18, 75], in which potrusion and polarity dynamics are critical [6]. Finally, protrusion and polarity dynamics are key also to pair-wise interactions of cells [19, 76, 77], which in turn control the collective dynamics of cells [78]. Our framework could provide a way to analyze and interpret cellular protrusion and polarity dynamics in these more complex systems.

Methods

Sample preparation
Fibronectin micropatterns are made by microscale plasma-initiated protein patterning as described previously [15]. All two-state micropatterns are designed to have adhesive island with square dimensions ($(36.7 \pm 0.6)^2$ µm^2). For patterns with varying bridge width, we use a standard bridge length $L = 35.3 \pm 0.5$ µm and widths $W = 3.9 \pm 0.5, 6.9 \pm 0.6, 9.5 \pm 0.5, 12.4 \pm 0.5, 19.1 \pm 0.5, 21.7 \pm 0.5, 34.8 \pm 0.2$ µm. For patterns with varying bridge length, we use standard bridge width $W = 6.9 \pm 0.6$ and lengths $L = 6.4 \pm 0.3, 9.2 \pm 0.3, 15.7 \pm 0.3, 23.7 \pm 0.4, 35.3 \pm 0.5, 46.2 \pm 0.4, 56.0 \pm 0.3$ µm. We refer to the rounded values for W and L throughout the text.

Cell culture and microscopy
MDA-MB-231 cells (DSMZ) are cultured in Minimum Essential Medium (MEM, c.c. pro), containing 10% FBS (Gibco) and 2mM L-Glutamine (c.c. pro). Cells are grown in a 5% CO_2 atmosphere at 37°C. For passaging and experiments, cells are washed once with PBS and trypsinised for 3 min. This cell solution is centrifuged at 1000 rcf for 3 min. The cell pellet is re-suspended in MEM and 10,000 cells are added per µ-dish and left to adhere in the incubator for 4h. The medium is then exchanged to L-15 medium containing L-glutamine (Gibco, supplemented with 10% FCS) and 25 nM Hoechst 33342 (Invitrogen) for staining cell nuclei. Experiments are performed at 37°C without CO_2. All measurements are performed in time-lapse mode for up to 50 h on an IMIC digital microscope (TILL Photonics) or on a Nikon Eclipse Ti microscope using a 10x objective. The samples are kept in a heated chamber (ibidi GmbH or Okolab) at 37°C throughout the measurements. Images (brightfield and DAPI) are acquired every 10 mins.

Nucleus tracking and cell segmentation
The trajectories of the cell nuclei are obtained by applying a band pass filter to the images of the nuclei, binarising, and tracking the binarised images using a Matlab tracking algorithm [79]. For cell segmentation, bright-field images of cells were segmented using a convolutional neural network with a U-Net architecture [36]. We created a data set of 372 manually segmented images, and used 80% for training and 20% for validation; both data sets were augmented with random rotations, shifts, shears, zooms, and reflections. For training, we used the Adam optimization algorithm [80] with a learning rate of 10^{-4}. No learning rate scheduling was used. After training for 20 epochs, we attain a pixel classification accuracy of $> 96\%$ on the validation dataset. Segmented cell images, which take values between 0 and 1, are subsequently binarized via thresholding. For further details, see Supplementary Section S2.

Protrusion tracking
Our image segmentation pipeline provides the 2D shape of the cells $\mathcal{S}(t)$ as a function of time. To identify protrusions, we classify the positive contributions to the shape velocities $\mathcal{V}(t) = \mathcal{S}(t + \Delta t) - \mathcal{S}(t)$ as protrusive components $\mathcal{P}(t)$. To recover a low-dimensional representation of the protrusive dynamics from these areas, we define the x-position of the protrusion x_p as the geometric center of the protrusive shape $x_\mathrm{p}(t) = \int x\mathcal{P}(t)\mathrm{d}x$. For further details, see Supplementary Section S2.

Acknowledgements

We thank Gregorz Gradziuk, Steven Riedijk, and Janni Harju for inspiring discussions, and Andriy Goychuk for advice on the image segmentation. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project-ID 201269156 - SFB 1032 (Project B12). D.B.B. is supported in part by a DFG fellowship within the Graduate School of Quantita-

tive Biosciences Munich (QBM) and by the Joachim Herz Stiftung.

References

[1] C. M. Franz, G. E. Jones, and A. J. Ridley, "Cell Migration in Development and Disease," *Developmental Cell*, vol. 2, no. 2, pp. 153–158, 2002.

[2] E. Scarpa and R. Mayor, "Collective cell migration in development," *The Journal of cell biology*, vol. 212, pp. 143–155, jan 2016.

[3] A. D. Luster, R. Alon, and U. H. von Andrian, "Immune cell migration in inflammation: present and future therapeutic targets," *Nature Immunology*, vol. 6, no. 12, pp. 1182–1190, 2005.

[4] P. Friedl and K. Wolf, "Tumour-cell invasion and migration: diversity and escape mechanisms.," *Nature reviews. Cancer*, vol. 3, no. 5, pp. 362–74, 2003.

[5] T. D. Pollard and G. Borisy, "Cellular Motility Driven by Assembly and Disassembly of Actin Filaments," *Cell*, vol. 112, pp. 453–465, 2003.

[6] P. T. Caswell and T. Zech, "Actin-Based Cell Protrusion in a 3D Matrix," *Trends in Cell Biology*, vol. 28, no. 10, pp. 823–834, 2018.

[7] A. J. Ridley, M. A. Schwartz, K. Burridge, R. A. Firtel, M. H. Ginsberg, G. Borisy, J. T. Parsons, and A. R. Horwitz, "Cell Migration: Integrating Signals from Front to Back," *Science*, vol. 302, no. 5651, pp. 1704–1709, 2003.

[8] R. H. Insall and L. M. Machesky, "Actin Dynamics at the Leading Edge: From Simple Machinery to Complex Networks," *Developmental Cell*, vol. 17, no. 3, pp. 310–322, 2009.

[9] H. Warner, B. J. Wilson, and P. T. Caswell, "Control of adhesion and protrusion in cell migration by Rho GTPases," *Current Opinion in Cell Biology*, vol. 56, pp. 64–70, 2019.

[10] M. H. Gail and C. W. Boone, "The Locomotion of Mouse Fibroblasts in Tissue Culture," *Biophysical Journal*, vol. 10, no. 10, pp. 980–993, 1970.

[11] D. Selmeczi, S. Mosler, P. H. Hagedorn, N. B. Larsen, and H. Flyvbjerg, "Cell motility as persistent random motion: theories from experiments.," *Biophysical journal*, vol. 89, no. 2, pp. 912–931, 2005.

[12] R. J. Petrie and K. M. Yamada, "At the leading edge of three-dimensional cell migration," *Journal of Cell Science*, vol. 125, no. 24, pp. 5917–5926, 2012.

[13] C. D. Paul, P. Mistriotis, and K. Konstantopoulos, "Cancer cell motility: Lessons from migration in confined spaces," *Nature Reviews Cancer*, vol. 17, no. 2, pp. 131–140, 2017.

[14] C. Metzner, C. Mark, J. Steinwachs, L. Lautscham, F. Stadler, and B. Fabry, "Superstatistical analysis and modelling of heterogeneous random walks.," *Nature communications*, vol. 6, no. May, p. 7516, 2015.

[15] 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," *Nature Physics*, vol. 15, no. 6, pp. 595–601, 2019.

[16] A. Fink, D. B. Brückner, C. Schreiber, P. J. Röttgermann, C. P. Broedersz, and J. O. Rädler, "Area and Geometry Dependence of Cell Migration in Asymmetric Two-State Micropatterns," *Biophysical Journal*, vol. 118, no. 3, pp. 552–564, 2020.

[17] D. B. Brückner, A. Fink, J. O. Rädler, and C. P. Broedersz, "Disentangling the Behavioural Variability of Confined Cell Migration," *J. R. Soc. Interface*, vol. 17, p. 20190689, 2020.

[18] P.-H. Wu, A. Giri, S. X. Sun, and D. Wirtz, "Three-dimensional cell migration does not follow a random walk.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 11, pp. 3949–54, 2014.

[19] 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. USA*, vol. 118, no. 7, p. e2016602118, 2020.

[20] D. Shao, W. J. Rappel, and H. Levine, "Computational model for cell morphodynamics," *Physical Review Letters*, vol. 105, no. 10, pp. 2–5, 2010.

[21] D. Shao, H. Levine, and W. J. Rappel, "Coupling actin flow, adhesion, and morphology in a computational cell motility model," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 18, pp. 6851–6856, 2012.

[22] F. Ziebert, S. Swaminathan, and I. S. Aranson, "Model for self-polarization and motility of keratocyte fragments," *Journal of the Royal Society Interface*, vol. 9, no. 70, pp. 1084–1092, 2012.

[23] F. Graner and J. A. Glazier, "Simulation of biological cell sorting using a two-dimensional extended Potts model," *Physical Review Letters*, vol. 69, no. 13, pp. 2013–2016, 1992.

[24] F. J. Segerer, F. Thüroff, A. Piera Alberola, E. Frey, and J. O. Rädler, "Emergence and persistence of collective cell migration on small circular micropatterns," *Physical Review Letters*, vol. 114, no. 22, p. 228102, 2015.

[25] A. Goychuk, D. B. Brückner, A. W. Holle, J. P. Spatz, C. P. Broedersz, and E. Frey, "Morphology and Motility of Cells on Soft Substrates," 2018.

[26] C. Chan and D. Odde, "Traction Dynamics of Filopodia on Compliant Substrates," *Science*, vol. 322, no. December, pp. 1687–1691, 2008.

[27] A. Elosegui-artola, X. Trepat, and P. Roca-Cusachs, "Control of Mechanotransduction by Molecular Clutch Dynamics," *Trends in Cell Biology*, vol. 28, no. 5, pp. 356–367, 2018.

[28] K. Kruse, J. F. Joanny, F. Jülicher, and J. Prost, "Contractility and retrograde flow in lamellipodium motion," *Physical Biology*, vol. 3, no. 2, pp. 130–137, 2006.

[29] P. Recho, T. Putelat, and L. Truskinovsky, "Active gel segment behaving as an active particle," *Physical Review E*, vol. 100, no. 6, pp. 1–15, 2019.

[30] P. Maiuri, J. F. Rupprecht, S. Wieser, V. Ruprecht, O. Bénichou, N. Carpi, M. Coppey, S. De Beco, N. Gov, C. P. Heisenberg, C. Lage 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*, vol. 161, no. 2, pp. 374–386, 2015.

[31] A. C. Callan-Jones and R. Voituriez, "Actin flows in cell migration: From locomotion and polarity to trajectories," *Current Opinion in Cell Biology*, vol. 38, pp. 12–17, 2016.

[32] J. E. Ron, P. Monzo, N. C. Gauthier, R. Voituriez, and N. S. Gov, "One-dimensional cell motility patterns," *Physical Review Research*, vol. 2, no. 3, pp. 1–27, 2020.

[33] P. Sens, "Stick-Slip model for actin-driven cell protrusions, cell polarisation and crawling," *Proc. Natl. Acad. Sci. USA*, vol. 117, no. 40, pp. 24670–24678, 2020.

[34] K. Hennig, I. Wang, P. Moreau, L. Valon, S. DeBeco, M. Coppey, Y. A. Miroshnikova, C. Albiges-Rizo, C. Favard, R. Voituriez, and M. Balland, "Stick-slip dynamics of cell adhesion triggers spontaneous symmetry breaking and directional migration of mesenchymal cells on one-dimensional lines," *Science Advances*, vol. 6, no. 1, pp. 1–13, 2020.

[35] C. Schreiber, B. Amiri, J. C. Heyn, J. O. Rädler, and M. Falcke, "On the adhesionvelocity relation and length adaptation of motile cells on stepped fibronectin lanes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 118, no. 4, 2021.

[36] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation BT - Medical Image Computing and Computer-Assisted Intervention MICCAI 2015," (Cham), pp. 234–241, Springer International Publishing, 2015.

[37] M. Machacek and G. Danuser, "Morphodynamic profiling of protrusion phenotypes," *Biophysical Journal*, vol. 90, no. 4, pp. 1439–1452, 2006.

[38] 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," *Journal of Cell Biology*, vol. 172, no. 1, pp. 41–53, 2006.

[39] P. M. Davidson, A. Battistella, T. Dejardin, T. Betz, J. Plastino, B. Cadot, N. Borghi, and C. Sykes, "Actin Accumulates Nesprin-2 at the Front of the Nucleus During Confined Cell Migration," *EMBO reports*, p. e49910, 2020.

[40] P. M. Davidson and B. Cadot, "Actin on and around the Nucleus," *Trends in Cell Biology*, vol. 31, no. 3, pp. 211–223, 2021.

[41] P. M. Davidson and B. Cadot, "Actin on and around the Nucleus," *Trends in Cell Biology*, dec 2020.

[42] T. Y. Tsai, S. R. Collins, C. K. Chan, A. Hadjitheodorou, P. Y. Lam, S. S. Lou, H. W. Yang, J. Jorgensen, F. Ellett, D. Irimia, M. W. Davidson, R. S. Fischer, A. Huttenlocher, T. Meyer, J. E. Ferrell, and J. A. Theriot, "Efficient Front-Rear Coupling in Neutrophil Chemotaxis by Dynamic Myosin II Localization," *Developmental Cell*, vol. 49, no. 2, pp. 189–205.e6, 2019.

[43] P. J. Albert and U. S. Schwarz, "Dynamics of cell shape and forces on micropatterned substrates predicted by a cellular Potts model," *Biophysical Journal*, vol. 106, no. 11, pp. 2340–2352, 2014.

[44] D. Bi, J. H. Lopez, J. M. Schwarz, and M. L. Manning, "Energy barriers and cell migration in densely packed tissues," *Soft Matter*, vol. 10, no. 12, pp. 1885–1890, 2014.

[45] D. Bi, J. H. Lopez, J. M. Schwarz, and M. L. Manning, "A density-independent rigidity transition in biological tissues," *Nature Physics*, vol. 11, no. 12, pp. 1074–1079, 2015.

[46] D. Lehnert, "Cell behaviour on micropatterned substrata: limits of extracellular matrix geometry for spreading and adhesion," *Journal of Cell Science*, vol. 117, no. 1, pp. 41–52, 2004.

[47] N. Q. Balaban, U. S. Schwarz, D. Riveline, P. Goichberg, G. Tzur, I. Sabanay, D. Mahalu, S. Safran, A. Bershadsky, L. Addadi, and B. Geiger, "Force and focal adhesion assembly: a close relationship studied using elastic micropatterned substrates.," *Nature cell biology*, vol. 3, no. 5, pp. 466–72, 2001.

[48] B. Lehle and J. Peinke, "Analyzing a stochastic process driven by Ornstein-Uhlenbeck noise," *Arxiv*, pp. 1–14, 2016.

[49] M. Soares E Silva, J. Alvarado, J. Nguyen, N. Georgoulia, B. M. Mulder, and G. H. Koenderink, "Self-organized patterns of actin filaments in cell-sized confinement," *Soft Matter*, vol. 7, no. 22, pp. 10631–10641, 2011.

[50] F. Bonelli, G. Gonnella, A. Tiribocchi, and D. Marenduzzo, "Spontaneous flow in polar active fluids: the effect of a phenomenological self propulsion-like term," *European Physical Journal E*, vol. 39, no. 1, pp. 1–10, 2016.

[51] X. Jiang, D. A. Bruzewicz, A. P. Wong, M. Piel, and G. M. Whitesides, "Directing cell migration with asymmetric micropatterns," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 4, pp. 975–978, 2005.

[52] M. Prager-Khoutorsky, A. Lichtenstein, R. Krishnan, K. Rajendran, A. Mayo, Z. Kam, B. Geiger, and A. D. Bershadsky, "Fibroblast polarization is a matrix-rigidity-dependent process controlled by focal adhesion mechanosensing," *Nature Cell Biology*, vol. 13, no. 12, pp. 1457–1465, 2011.

[53] A. F. Marée, V. A. Grieneisen, and L. Edelstein-Keshet, "How cells integrate complex stimuli: The effect of feedback from phosphoinositides and cell shape on cell polarization and motility," *PLoS Computational Biology*, vol. 8, no. 3, 2012.

[54] A. Vasilevich, S. Vermeulen, J. D. Boer, A. Carlier, and K. Eroume, "On the influence of cell shape on dynamic reaction-diffusion polarization patterns," *PLoS ONE*, vol. 16, no. 3, p. e0248293., 2021.

[55] L. Caprini, U. Marini Bettolo Marconi, A. Puglisi, and A. Vulpiani, "Active escape dynamics: The effect of persistence on barrier crossing," *Journal of Chemical Physics*, vol. 150, no. 2, 2019.

[56] S. L. Gupton and C. M. Waterman-Storer, "Spatiotemporal Feedback between Actomyosin and Focal-Adhesion Systems Optimizes Rapid Cell Migration," *Cell*, vol. 125, no. 7, pp. 1361–1374, 2006.

[57] G. Giannone, R. M. Mège, and O. Thoumine, "Multi-level molecular clutches in motile cell processes," *Trends in Cell Biology*, vol. 19, no. 9, pp. 475–486, 2009.

[58] P. Monzo, Y. K. Chong, C. Guetta-Terrier, A. Krishnasamy, S. R. Sathe, E. K. F. Yim, W. H. Ng, B. T. Ang, C. Tang, B. Ladoux, N. C. Gauthier, and M. P. Sheetz, "Mechanical confinement triggers glioma linear migration dependent on formin FHOD3," *Molecular Biology of the Cell*, vol. 27, pp. 1246–1261, feb 2016.

[59] W. Y. Wang, C. D. Davidson, D. Lin, and B. M. Baker, "Actomyosin contractility-dependent matrix stretch and recoil induces rapid cell migration," *Nature Communications*, vol. 10, no. 1, pp. 1–12, 2019.

[60] F. Raynaud, M. E. Ambühl, C. Gabella, A. Bornert, I. F. Sbalzarini, J. J. Meister, and A. B. Verkhovsky, "Minimal model for spontaneous cell polarization and edge activity in oscillating, rotating and migrating cells," *Nature Physics*, vol. 12, no. 4, pp. 367–373, 2016.

[61] J. Prost, F. Jülicher, and J. F. Joanny, "Active gel physics," *Nature Physics*, vol. 11, no. 2, pp. 111–117, 2015.

[62] F. Jülicher, S. W. Grill, and G. Salbreux, "Hydrodynamic theory of active matter," *Reports on Progress in Physics*, vol. 81, no. 7, 2018.

[63] L. Trichet, J. Le Digabel, R. J. Hawkins, S. R. K. Vedula, M. Gupta, C. Ribrault, P. Hersen, R. Voituriez, and B. Ladoux, "Evidence of a large-scale mechanosensing mechanism for cellular adaptation to substrate stiffness.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 18, pp. 6933–8, 2012.

[64] M. Gupta, B. R. Sarangi, J. Deschamps, Y. Nematbakhsh, A. Callan-Jones, F. Margadant, R. M. Mège, C. T. Lim, R. Voituriez, and B. Ladoux, "Adaptive rheology and ordering of cell cytoskeleton govern matrix rigidity sensing," *Nature Communications*, vol. 6, no. May, 2015.

[65] B. Ladoux, R. M. Mège, and X. Trepat, "Front-Rear Polarization by Mechanical Cues: From Single Cells to Tissues," *Trends in Cell Biology*, vol. 26, no. 6, pp. 420–433, 2016.

[66] R. G. Hodge and A. J. Ridley, "Regulating Rho GTPases and their regulators," *Nature Reviews Molecular Cell Biology*, vol. 17, no. 8, pp. 496–510, 2016.

[67] K. A. Demali and K. Burridge, "Coupling membrane protrusion and cell adhesion," 2003.

[68] S. Saha, T. L. Nagy, and O. D. Weiner, "Joining forces: Crosstalk between biochemical signalling and physical forces orchestrates cellular polarity and dynamics," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 373, no. 1747, 2018.

[69] T. Heck, D. A. Vargas, B. Smeets, H. Ramon, P. van Liedekerke, and H. van Oosterwyck, *The role of actin protrusion dynamics in cell migration through a degradable viscoelastic extracellular matrix: Insights from a computational model*, vol. 16. 2020.

[70] C. Schreiber, F. J. Segerer, E. Wagner, A. Roidl, and J. O. Rädler, "Ring-Shaped Microlanes and Chemical Barriers as a Platform for Probing Single-Cell Migration," *Scientific Reports*, vol. 6, no. February, p. 26858, 2016.

[71] D. Caballero, R. Voituriez, and D. Riveline, "Protrusion fluctuations direct cell motion," *Biophysical Journal*, vol. 107, no. 1, pp. 34–42, 2014.

[72] D. Caballero, J. Comelles, M. Piel, R. Voituriez, and D. Riveline, "Ratchetaxis: Long-Range Directed Cell Migration by Local Cues," *Trends in Cell Biology*, vol. 25, no. 12, pp. 815–827, 2015.

[73] G. Mahmud, C. J. Campbell, K. J. Bishop, Y. A. Komarova, O. Chaga, S. Soh, S. Huda, K. Kandere-Grzybowska, and B. A. Grzybowski, "Directing cell motions on micropatterned ratchets," *Nature Physics*, vol. 5, no. 8, pp. 606–612, 2009.

[74] S. Lo Vecchio, R. Thiagarajan, D. Caballero, V. Vigon, L. Navoret, R. Voituriez, and D. Riveline, "Collective Dynamics of Focal Adhesions Regulate Direction of Cell Motion," *Cell Systems*, vol. 10, pp. 1–8, 2020.

[75] S. I. Fraley, Y. Feng, R. Krishnamurthy, D.-h. Kim, A. Celedon, G. D. Longmore, D. Wirtz, and S. Louis, "A distinctive role for focal adhesion proteins in three-dimensional cell motility," *Nat. Cell Biol.*, vol. 12, no. 6, pp. 598–604, 2010.

[76] M. Abercrombie and J. E. Heaysman, "Observations on the social behaviour of cells in tissue culture. I. Speed of movement of chick heart fibroblasts in relation to their mutual contacts," *Experimental Cell Research*, vol. 5, no. 1, pp. 111–131, 1953.

[77] C. Carmona-Fontaine, H. K. Matthews, S. Kuriyama, M. Moreno, G. A. Dunn, M. Parsons, C. D. Stern, and R. Mayor, "Contact inhibition of locomotion in vivo controls neural crest directional migration," *Nature*, vol. 456, no. 7224, pp. 957–961, 2008.

[78] R. Alert and X. Trepat, "Physical Models of Collective Cell Migration," *Annual Review of Condensed Matter Physics*, vol. 11, no. 1, pp. 77–101, 2020.

[79] D. Blair and E. Dufresne, "The Matlab Particle Tracking Code Repository," *http://site.physics.georgetown.edu/matlab/*, 2008.

[80] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, International Conference on Learning Representations, ICLR, 2015.

Supplementary Material:

Theory of protrusion and polarity dynamics in confined cell migration

David B. Brückner, Matthew Schmitt, Alexandra Fink, Johannes Flommersfeld, Nicolas Arlt, Edouard Hannezo, Joachim O. Rädler and Chase P. Broedersz

Contents

1 Movie descriptions

Supplementary Movie S1
Single MDA-MB-231 cell migrating in a two-state micropattern with constriction length $L = 35$ µm and width $W = 7$ µm. The cell nucleus is fluorescently labelled to allow automated tracking of nucleus positions. Scale bar: 25 µm.

Supplementary Movie S2
Cell shape segmentation of the brightfield microscopy images from Supplementary Movie S1, with the detected cell boundary marked in yellow. Scale bar: 25 µm.

Supplementary Movies S3-S10
Protrusion dynamics of MDA-MB-231 cells in two state micropatterns with widths $W = 4, 7, 10, 12, 19, 22, 35$ µm respectively. Protrusive areas are marked in green. The blue dot corresponds to the x-position of the nucleus, and the green dot to the x-position of the center of protrusive area. Scale bar: 25 µm.

2 Image analysis

2.1 Cell segmentation

In the bright-field images obtained from experiment, attempts to isolate the cell from its background are easily confounded by the nontrivial structure of the micropatterned substrate. Additionally, the cell boundary is often difficult to pinpoint by eye. For these reasons, we found that attempts to segment the cell images using traditional methods of image binarization failed and thus turned to more advanced machine learning techniques.

To segment our images, we utilized convolutional neural networks, which allow for high pixel classification accuracy by accounting for local properties of the image. In particular we use a U-Net architecture, which has been found to be very successful in image segmentation tasks in biology [1]. In principle, the network combines an encoder/decoder structure with skip connections across the latent layers. The encoder/decoder structure allows for efficient recognition of large-scale features in the image, while the skip connections effectively propagate local, low-level information forward in the network. The encoder and decoder branches of our network are three layers deep, with 64 channels in the first layer which are doubled after every max pool layer, similar to previous implementations [1].

For training, the network is fed augmented data which has undergone random rotations, shifts, shears, zooms, and reflections. We use 80% of the original labeled data set of ($N = 372$) images for training, and withhold 20% for validation. Each epoch then consists of 2000 steps of batch size 16, and training is stopped after 20 epochs to pre-

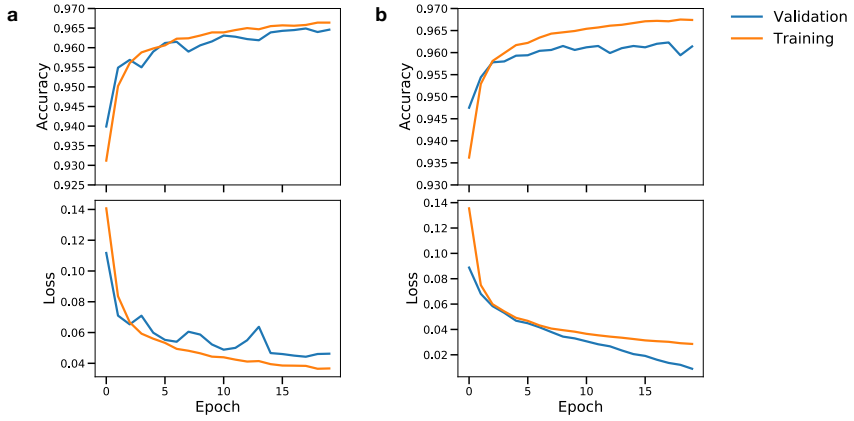


Figure S1: **Accuracy and loss curves of the training process for videos with (a) high and (b) low contrast.** The network for high contrast videos was trained with constant $\alpha = 1$, while the network for low contrast videos was trained with varying α.

vent overfitting. Gradient updates are performed using the Adam optimizer [2] with a constant learning rate of 10^{-4}. We used the binary cross-entropy as a loss function to optimize the pixel classification accuracy. For videos with low contrast between the cells and the background, resulting from the use of a different microscope, we adjusted the loss function throughout the training to increase the focus on the cell edges, which improved the segmentation quality, which has been found to have a similar effect in previous work [3]. Specifically, we use the total loss function

$$\mathcal{L} = \alpha \mathcal{L}_{\text{BCE}} + (1 - \alpha)\mathcal{L}_{\text{BCE, edge}}. \tag{S1}$$

Here, $\mathcal{L}_{\text{BCE}}$ is the binary cross entropy loss for the entire image, and $\mathcal{L}_{\text{BCE, edge}}$ is the binary cross entropy only applied to pixels near the edge of the cell. The factor α is deterministically reduced in each epoch to force the network to specialize and focus on the cell boundary in the later phase of the training, which makes up a comparatively small number of pixels compared to the cell as a whole. The parameter α is initialized to 1 and then gradually reduced by 0.05 with each epoch, which we found improved training compared to a fixed alpha.

Training according to the above protocol results in a pixel classification accuracy of 96.5% for videos with high contrast and 96.1% for videos with low contrast on the validation dataset. We note an apparent slight overfitting, with predictions on the training set achieving a slightly higher accuracy of 96.6% for both high and low contrast videos (Fig. S1).

Finally, the predicted segmentations are converted to binary images by applying a threshold. Consequently, pixels with predicted values above 0.12 are mapped to 1, else to 0. This pipeline yields an accurate segmentation of the cell shape for the vast majority of frames (Fig. S2).

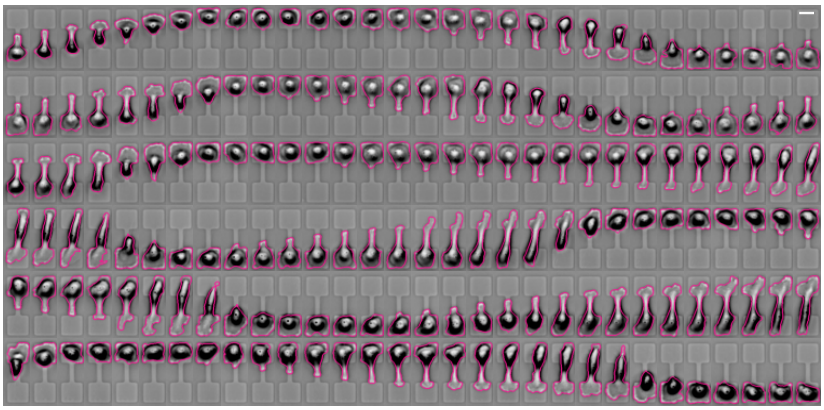


Figure S2: **Exemplary brightfield time-series, with segmented cell shapes shown in pink.** Each image is a frame from a video which is sampled every $\Delta t = 10$ min; time flows from the left to the right, and each row is the continuation of the row above it. Bright-field images are inverted for better visibility. Scale bar: 25 µm.

2.2 Protrusion tracking

To quantify the joint dynamics of nucleus and protrusion motion, we seek a minimal, low-dimensional representation of the cell protrusions. Our image segmentation pipeline gives access to the 2D shape of the cells $\mathcal{S}(t)$ as a function of time. To identify protrusions, we classify the positive contributions to the shape velocities $\mathcal{V}(t) = \mathcal{S}(t + \Delta t) - \mathcal{S}(t)$ as the shape of the protrusion $\mathcal{P}(t)$ (green areas in Fig. S3).

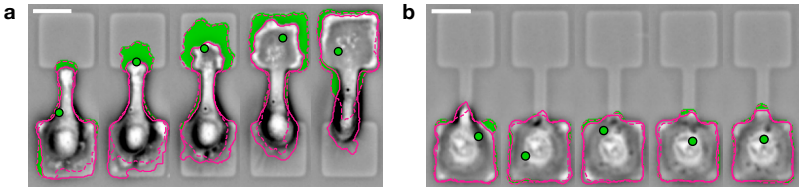


Figure S3: **Dynamics of protrusive areas in two frame sequences.** The solid pink line shows the current boundary of the cell area $\mathcal{S}(t)$, and the dashed line is the boundary of $\mathcal{S}(t + \Delta t)$. The protrusive shape (green) is the area which is added between these two frames, $\mathcal{P}(t)$. The geometric center of the protrusive area $\mathbf{x}_\mathrm{p}$ is shown as a green dot. Scale bars: 25 µm.

As a low-dimensional representation of the protrusive dynamics, we define an effective position of the protrusion $\mathbf{x}_\mathrm{p}$ as the geometric center of the protrusive shape $\mathbf{x}_\mathrm{p}(t) = \int \mathbf{x}\mathcal{P}(t)\mathrm{d}\mathbf{x}$ (green dot in Fig. S3). The two-state micropattern is designed in such a way that most of the behavior occurs in the x-direction along the long axis of the micropattern. Indeed, we find that, similar to the nucleus dynamics [4], most of the protrusive behaviour is captured by the x-component of $\mathbf{x}_\mathrm{p}$ (Fig. S4): the variance in y-motion is small (Fig. S4a), and the joint probability distribution $p(y_\mathrm{n}, y_\mathrm{p})$ is peaked around $(0, 0)$ and exhibits no special structure, unlike the probability distribution for x-components $p(x_\mathrm{n}, x_\mathrm{p})$ (Fig. S4b,c). In the following, we will therefore take the x-component x_p as a minimal representation of the protrusive dynamics in this system.

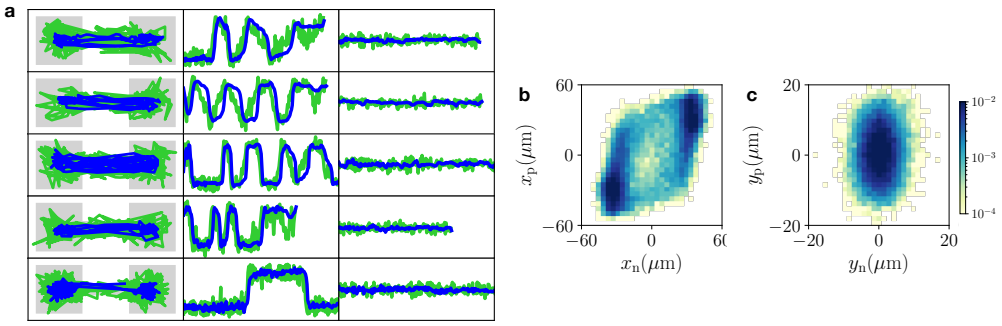


Figure S4: **2D motion of nucleus and protrusions. a.** Several examples of 2D trajectories. Left: xy-trajectories plotted on top of the micropattern dimension (shown in grey). Axis limits are -50 µm $< x <$ 50 µm and -20 µm $< y <$ 20 µm; $(x = 0, y = 0)$ corresponds to the center of the constriction. Middle: x-trajectories as a function of time t. Axis limits are -50 µm $< x <$ 50 µm and $0 < t <$ 30 h. Right: y-trajectories as a function of time t. Axis limits are -50 µm $< y <$ 50 µm and $0 < t <$ 30 h, to allow direct comparison with the x-trajectories. Blue: nucleus, green: protrusion. **b.** Joint probability distribution $p(x_n, x_p)$ of the x-positions, plotted logarithmically. Here shown without the Gaussian interpolation employed in Fig. 1 in the main text. **c.** Joint probability distribution $p(y_n, y_p)$ of the y-positions, plotted logarithmically. Note the smaller axis range compared to panel **b**.

We find that this definition captures the characteristic features of the protrusive dynamics during the cell-hopping process: as the protrusion grows into the constriction, the effective protrusion position also moves into the channel (Fig. S5 and Fig. 1d in the main text). Thus, $\mathbf{x}_p$ typically precedes $\mathbf{x}_n$ in the constriction, as expected from the experimental observations (Supplementary Movies S1-3, Fig. 1d in the main text). Furthermore, we find that when protrusions form randomly and uniformly around the cell boundary, x_p is located near the cell centroid (Fig. S3).

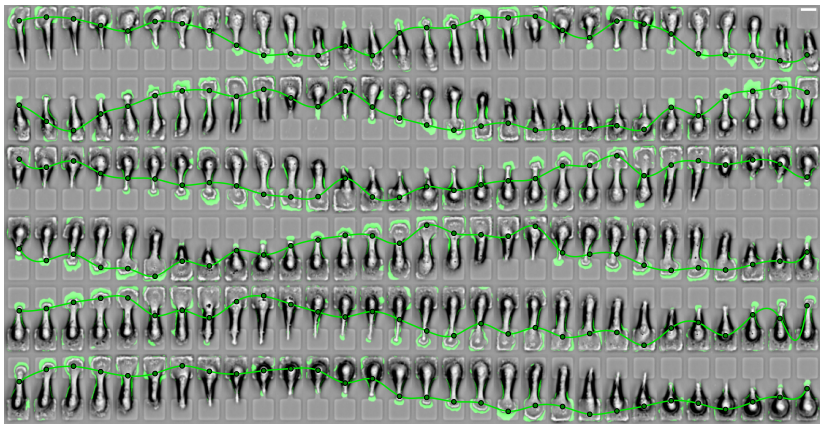


Figure S5: **Time series of x_p dynamics overlaid on images of cells with protrusions.** Each image is a frame from a video which is sampled every 10 minutes; time flows from the left to the right, and each row is the continuation of the row above it. Time series curve is an interpolation of the circular points to serve as a guide for the eye. Scale bar: 25 µm.

In addition to the protrusive dynamics, the cell also performs retractions, corresponding to the negative components of the shape velocities, $\mathcal{R}(t)$ (Fig. S6a). Using a similar analysis of the retractive dynamics by defining the effective position of the retractions,

$x_r(t) = \int x\mathcal{R}(t)dx$, we find however that the retractions are well correlated with the position of the nucleus, which typically resides at the rear end of the cell (Fig. S6b). Specifically, the cross-correlation of nucleus and retraction positions exhibits almost no time-lag, in contrast to the correlation between nucleus and protrusion (Fig. S6c). Furthermore, the cross-correlation between nucleus and retractions is very similar in magnitude and shape to the nucleus position auto-correlation, indicating that the retraction trajectories do not contain significant additional information to the nucleus trajectories. Furthermore, the joint probability distribution of nucleus and retraction positions has maximal probability around the diagonal, with little additional structure, in contrast to the distribution of nucleus and protrusion positions (Fig. S6d,e). Therefore, to achieve a minimal, low-dimensional description for the coupled dynamics of shape and nucleus motion, we restrict our analysis to the protrusions.

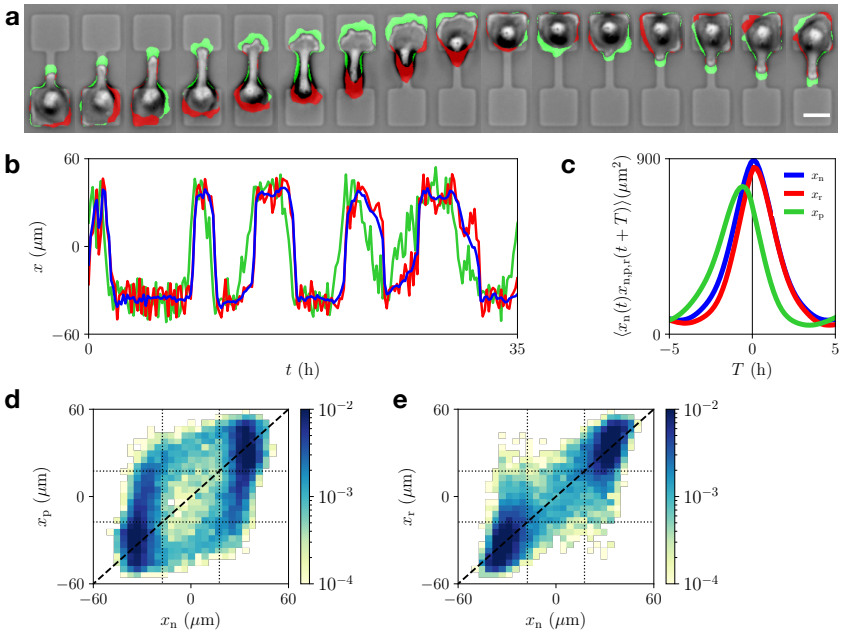


Figure S6: **Dynamics of cell retractions.** **a.** Exemplary brightfield microscopy image series with protrusive shape velocity components $\mathcal{P}(t)$ indicated in green, and retraction components $\mathcal{R}(t)$ in red. **b.** Trajectories of the protrusion $x_p(t) = \int x\mathcal{P}(t)dx$ (green), retraction $x_r(t) = \int x\mathcal{R}(t)dx$ (red), and the cell nucleus $x_n(t)$ (blue). **c.** Position cross-correlations between nucleus and protrusion $\langle x_n(t)x_p(t+T)\rangle$ (green), between nucleus and retraction $\langle x_n(t)x_r(t+T)\rangle$ (red), and nucleus position auto-correlation, $\langle x_n(t)x_n(t+T)\rangle$ (blue). **d.** Joint probability distribution $p(x_n, x_p)$ of the x-positions of nucleus and protrusion, plotted logarithmically. Here shown without the Gaussian interpolation employed in Fig. 1 in the main text. Dashed line indicates the diagonal; dotted lines indicate the boundaries of the adhesive islands. **e.** Joint probability distribution $p(x_n, x_r)$ of the x-positions of nucleus and retraction, plotted logarithmically.

3 Supplementary results

3.1 Inferred white noise model does not capture experimental dynamics

In this section, we show that a general model with white noise polarity dynamics is unable to capture the experimental dynamics. Specifically, we consider a model of the form

$$\dot{x}_\mathrm{n} = f_\mathrm{n}(x_\mathrm{n}, x_\mathrm{p}) + \sigma_\mathrm{n}(x_\mathrm{n}, x_\mathrm{p})\xi(t) \tag{S2}$$
$$\dot{x}_\mathrm{p} = f_\mathrm{p}(x_\mathrm{n}, x_\mathrm{p}) + \sigma_\mathrm{p}(x_\mathrm{n}, x_\mathrm{p})\xi(t) \tag{S3}$$

Here, we assume that $\xi(t)$ is a white noise with $\langle\xi(t)\rangle = 0$ and $\langle\xi(t)\xi(t')\rangle = \delta(t - t')$. Under this assumption, we can infer the terms $f_\mathrm{n,p}$ and $\sigma_\mathrm{n,p}$ directly from the observed data [5, 6]. Specifically, we use the estimators $f_\mathrm{n}(x_\mathrm{n}, x_\mathrm{p}) \approx \langle\dot{x}_\mathrm{n}|x_\mathrm{n}, x_\mathrm{p}\rangle$ and $\sigma_\mathrm{n}^2(x_\mathrm{n}, x_\mathrm{p}) \approx \Delta t\langle[\dot{x}_\mathrm{n} - f_\mathrm{n}(x_\mathrm{n}, x_\mathrm{p})]^2|x_\mathrm{n}, x_\mathrm{p}\rangle$ and similarly for the protrusion terms. These inferred terms provide the best fit estimates for a general model inferred under the white noise assumption. In this case, the inferred functions f_n and f_p are given by the NVM and the PVM by definition (Fig. S7a,b).

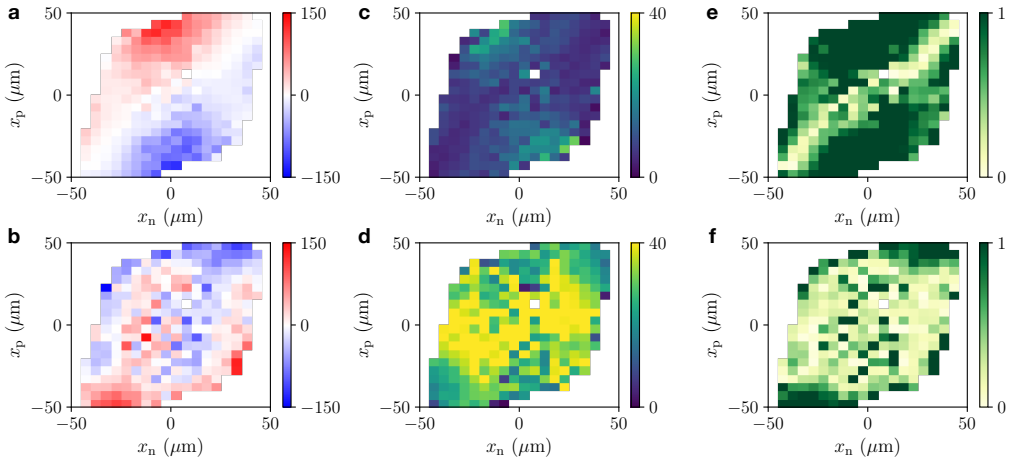


Figure S7: **Inferred model terms based on white noise assumption. a.** Inferred deterministic nucleus term $f_\mathrm{n}(x_\mathrm{n}, x_\mathrm{p}) \approx \langle\dot{x}_\mathrm{p}|x_\mathrm{n}, x_\mathrm{p}\rangle$ in units of µm h^{-1}. **b.** Inferred deterministic protrusion term. **c.** Inferred multiplicative noise term on the nucleus $\sigma_\mathrm{n}(x_\mathrm{n}, x_\mathrm{p}) \approx (\Delta t\langle[\dot{x}_\mathrm{n} - f_\mathrm{n}(x_\mathrm{n}, x_\mathrm{p})]^2|x_\mathrm{n}, x_\mathrm{p}\rangle)^{1/2}$ in units of µm h$^{-1/2}$. **d.** Inferred multiplicative noise term on the protrusion. **e.** Signal-to-noise ratio of the nucleus velocities, for an increment in a time-step Δt, given by $|f_\mathrm{n}(x_\mathrm{n}, x_\mathrm{p})|\sqrt{\Delta t}/\sigma_\mathrm{n}(x_\mathrm{n}, x_\mathrm{p})$. **f.** Signal-to-noise ratio of the protrusion velocities.

Interestingly, we find that the noise on the protrusion significantly exceeds that on the nucleus (Fig. S7c,d). Specifically, the average estimated noise magnitudes are $\hat{\sigma}_\mathrm{n} \approx 8.4$ µm h$^{-1/2}$ and $\hat{\sigma}_\mathrm{p} \approx 33$ µm h$^{-1/2}$. Accordingly, we find that the nucleus dynamics is dominated by its deterministic component, with a signal-to-noise ratio > 1 everywhere except where nucleus and protrusion are very close together (Fig. S7e). In contrast, the

protrusion dynamics is dominated by noise, with at signal-to-noise ratio < 1 in most regions of phase-space (Fig. S7f). In the mechanistic model introduced in the main text, we assume that the source of stochasticity in the system acts on the protrusion, which is further supported by these observations.

Using the inferred terms $f_{\mathrm{n,p}}$ and $\sigma_{\mathrm{n,p}}$, we perform simulations to assess the validity of the model postulated by Eqs. (S2), (S3). However, we find that this model yields predictions that are inconsistent with our experimental observations. The model does not recover the ring-structure in the probability distribution $p(x_{\mathrm{n}}, x_{\mathrm{p}})$ (Fig. S8a,b). Furthermore, it fails to predicts the peak in the dwell time distribution (Fig. S8e). Taken together, these results indicate that the model postulated in Eqs. (S2), (S3) does not provide a good representation of the experimental dynamics, ruling out the white noise protrusion model for all $f_{\mathrm{n,p}}$.

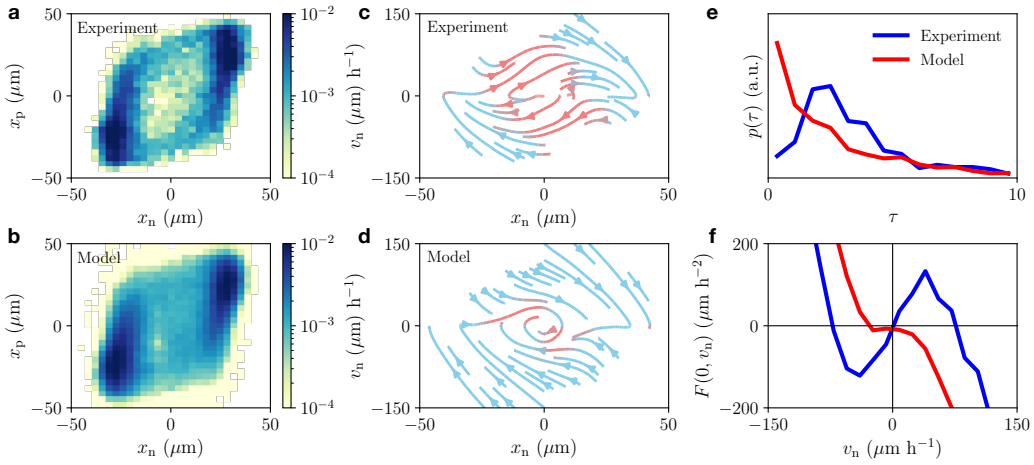


Figure S8: **Predictions of the inferred white noise model. a.** Experimental distribution $p(x_{\mathrm{n}}, x_{\mathrm{p}})$. **b.** Model prediction for $p(x_{\mathrm{n}}, x_{\mathrm{p}})$. **c.** $x_{\mathrm{n}}v_{\mathrm{n}}$ phase-space portrait inferred from experiments. **d.** Phase-space portrait predicted by model. **e.** Experimental (blue) and predicted (red) dwell time distribution. **f.** Experimental (blue) and predicted (red) effective friction relation $F(x_{\mathrm{n}} \to 0, v_{\mathrm{n}})$.

3.2 Non-linear nucleus-protrusion couplings

In the main text, we show that a double-well potential as a model for the effect of the confinement on the nucleus dynamics provides predictions that are inconsistent with the experimental data. To test if a possible nonlinearity in the nucleus-protrusion coupling could provide a better fit, we consider the next order coupling term allowed by symmetry in the deformation model:

$$\dot{x}_\mathrm{n} = k_\mathrm{n}^{(1)}(x_\mathrm{p} - x_\mathrm{n}) + k_\mathrm{n}^{(2)}(x_\mathrm{p} - x_\mathrm{n})^3 - \partial_{x_\mathrm{n}} W(x_\mathrm{n}) \tag{S4}$$

Fitting this model to the experimental NVM, we find that it does provide a slightly improved fit to the nucleus velocities as a function of x_n compared to the linear deformation presented in the main text (Fig. S9b). Specifically, unlike the linear coupling deformation model, the non-linear model (Eq. (S4)) exhibits an acceleration of the nucleus for large x_p. This is because the best fit parameters for the coupling yield a negative spring constant $k_\mathrm{n}^{(2)} < 0$ for the third order term, indicating that for large extensions, the spring is no longer contractile, but becomes extensile. However, for smaller x_p values, this model still performs a significantly worse fit than the first-order coupling adhesion model (Eqn. (S10)), which has fewer fitting parameters. Furthermore, this model is unable to capture the profiles of the nucleus velocities as a function of x_p (Fig. S9c), unlike the first-order coupling adhesion model (Fig. S17 c).

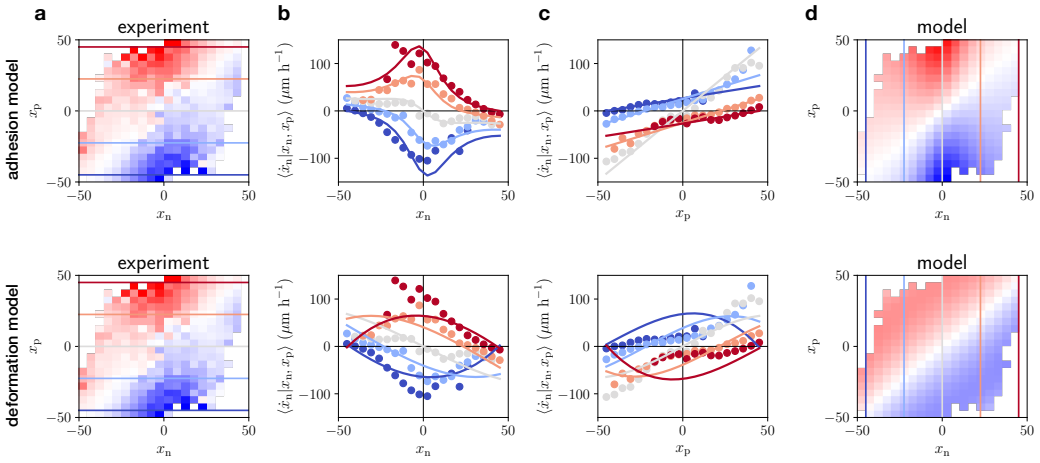


Figure S9: **Parameter estimation by fitting to the NVM for non-linear nucleus-protrusion coupling models. a.** Experimental NVM for bridge width $W = 3$ µm. **b.** Cuts of the NVM along the horizontal lines indicated in panel **a**, i.e. $\dot{x}_\mathrm{p}$ as a function of x_n for different x_p.. Dots: Experiment, Line: Fitted non-linear coupling model. **c.** Cuts of the NVM along the vertical lines indicated in panel **d**, i.e. $\dot{x}_\mathrm{p}$ as a function of x_p for different x_n. Dots: Experiment, Line: Fitted non-linear coupling model. **d.** Fitted NVM. *Top row:* adhesion model. *Bottom row:* deformation model.

Fig. S9 also demonstrates that adding the third-order coupling term to the adhesion model does not yield a significant improvement of the fit. This model is defined by the

equation

$$\dot{x}_\mathrm{n} = \left[k_\mathrm{n}^{(1)}(x_\mathrm{p} - x_\mathrm{n}) + k_\mathrm{n}^{(2)}(x_\mathrm{p} - x_\mathrm{n})^3\right] / \gamma(x_\mathrm{n}) \tag{S5}$$

Fitting this model, we infer coupling constants $|k_\mathrm{n}^{(2)}|/|k_\mathrm{n}^{(1)}| \approx 10^{-6}$, indicating that the third order term is negligible compared to the first order term. Thus, we conclude that the first-order coupling adhesion model (Eqn. (S10)) is the simplest mechanistic model that captures our data accurately.

3.3 Connecting the mechanistic model to emergent stochastic nonlinear dynamics

A central challenge for our mechanistic approach is to capture the emergent long timescale stochastic dynamics of the system. In previous work [4], we discovered that the stochastic dynamics of the nucleus trajectories $x_\mathrm{n}(t)$ of these cells can be described by an equation of motion for the velocity of the cell nucleus v_n of the form

$$\dot{v}_\mathrm{n} = F(x_\mathrm{n}, v_\mathrm{n}) + \sigma(x_\mathrm{n}, v_\mathrm{n})\eta(t) \tag{S6}$$

where $\eta(t)$ is Gaussian white noise, with $\langle\eta(t)\rangle = 0$ and $\langle\eta(t)\eta(t')\rangle = \delta(t - t')$. This is an effective description of the dynamics of the nucleus alone, with unobserved degrees of freedom, such as the protrusion and polarity, integrated out. Thus, in contrast to our mechanistic model (Eqns. 1,2 in the main text), the dynamics of the nucleus alone is described by an *underdamped* equation of motion with the velocity v_n as an additional degree of freedom.

Here, we provide a direct mapping between the mechanistic and the effective underdamped model for this system. Specifically, we recast Eqns. 1,2 in the main text into a single differential equation for v_n. Then, using the definition $F(x_\mathrm{n}, v_\mathrm{n}) = \langle\dot{v}_\mathrm{n}|x_\mathrm{n}, v_\mathrm{n}\rangle$, we find

$$F(x_\mathrm{n}, v_\mathrm{n}) = \underbrace{f_\mathrm{p}\frac{\partial f_\mathrm{n}}{\partial x_\mathrm{p}} + v_\mathrm{n}\frac{\partial f_\mathrm{n}}{\partial x_\mathrm{n}}}_{F_\mathrm{cc}(x_\mathrm{n}, v_\mathrm{n})} + \underbrace{\frac{\partial f_\mathrm{n}}{\partial x_\mathrm{p}}\langle P|x_\mathrm{n}, v_\mathrm{n}\rangle}_{F_\mathrm{pol}(x_\mathrm{n}, v_\mathrm{n})} \tag{S7}$$

Thus, we expect the deterministic dynamics of the nucleus to be determined by two components. A component $F_\mathrm{cc}(x_\mathrm{n}, v_\mathrm{n})$ determined by the confinement and coupling dynamics, and a component $F_\mathrm{pol}(x_\mathrm{n}, v_\mathrm{n})$ determined by the polarity dynamics.

For white noise polarities, the second term vanishes, as $\langle P|x_\mathrm{n}, v_\mathrm{n}\rangle = 0$ (Insets Fig. S10a), and thus the phase space flow is due to the combined effects of nucleus-protrusion coupling and the space-dependent adhesiveness acting on the nucleus. Interestingly, for the white noise model, we find a small region of deterministic amplification where the nucleus enters the constriction - however, the amplification only sets in at high speeds, while there is no amplification for low speeds (Fig. S10a). The amplification in the flow is due to the differential adhesiveness, as it vanishes for a flat adhesiveness profile. In contrast, in the experiments, we found that the excitable amplification regime

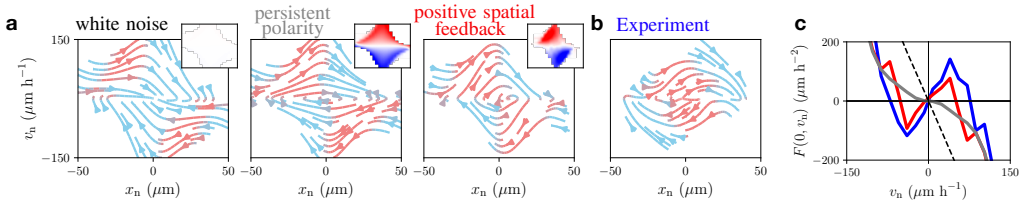


Figure S10: **Phase-space portraits of the non-linear dynamics of nucleus trajectories.** **a.** Flow field $(\dot{x}_n, \dot{v}_n) = (v_n, F(x_n, v_n))$ indicated by arrows. Arrow color indicates the direction of the local flow: acceleration is orange and deceleration is blue. From left to right, the predictions for the white noise model, the persistent polarity model, and the spatial feedback model (with $\alpha_{min} < 0$). *Insets:* conditional average of the polarity as a function of nucleus position and velocity, $\langle P|x_n, v_n\rangle$, which determines the polarity component of the phase space flow $F_{pol}(x_n, v_n)$ (see Eq. (S7)). **b.** Flow field inferred from the experiment for bridge width $W = 7$ µm (see ref. [4]). **b.** Effective friction at the bridge center $F(x_n \to 0, v_n)$ for all three models and the experiment: white noise model (dashed black line), persistent polarity model (grey), spatial feedback model (red), and experiment (blue).

sets in already at low speeds. Furthermore, the effective friction acting on the nucleus in the white noise model is a simple linear friction, $F(x_n \to 0, v_n) \propto -v_n$, in contrast to the non-linear anti-friction observed experimentally (Fig. S10b).

In contrast, persistent polarities lead to a significant contribution to the deterministic dynamics, with $\langle P|x_n, v_n\rangle$ exhibiting positive values for positive velocities. Consequently, the polarity component of the dynamics F_{pol} fundamentally changes the phase space flow, leading to amplification even at small velocities, similar to the experiment (Fig. S10a). However, while the model predicts a non-linear effective friction relation $F(x_n \to 0, v_n)$, it does not predict a sign-change, corresponding to anti-friction, in any parameter regime we investigated (Fig. S10b; see Supplementary Section 4.3.2).

Finally, for the spatial feedback model, we find that the polarity component similarly leads to an amplification of velocity, and yields excitable dynamics similar to those observed experimentally. However, in contrast to the persistent polarity model, we find that the spatial feedback model captures the effective anti-friction at the center of the constriction. In the parameter regime relevant to the experiments (see Supplementary Section 4), we find that the effective anti-friction emerges for $\alpha_{min} < 0$ and $\gamma_{min} \lesssim 0.3$ (Fig. S11). To model the effects of increasing constriction width, we simultaneously increase γ_{min} and α_{min} (red arrow Fig. S11). We observe that this leads to the disappearance of the effective anti-friction, first giving rise to a flat non-linear friction, and finally an almost linear regular friction (Insets Fig. S11). We observe very similar changes in the effective friction in the experiment (Supplementary Section 3.4).

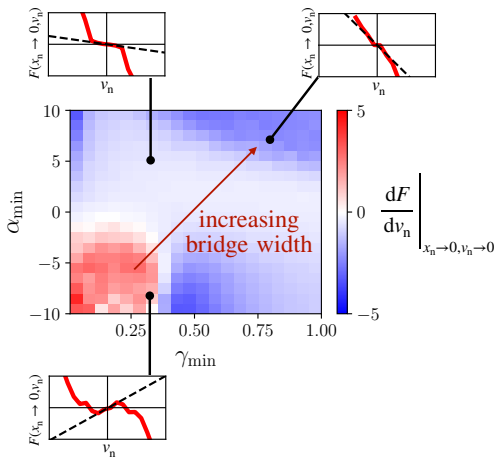


Figure S11: **Emergence of effective anti-friction for positive feedback polarities.** We vary the parameters $\gamma_{\min}$ and $\alpha_{\min}$ and determine the effective friction relationship at the center of the constriction $F(x_n \to 0, v_n)$. The gradient of the effective friction at $v_n \to 0$ is indicated by the color. Red arrow corresponds to simultaneously increasing $\gamma_{\min}$ and $\alpha_{\min}$ which we do as a model of increasing bridge width. Insets: effective friction relationships at the indicated locations.

In summary, we find that the effective non-linear dynamics of the nucleus trajectories put strong constraints on our mechanistic model, and in contrast to the white noise and persistent polarity models, the spatial feedback model is able to capture the experimentally observed dynamics. Our model furthermore gives insight into the origin of the non-linear dynamics: $F(x_n, v_n)$ is composed of a confinement-coupling and a polarity component. The effective anti-friction exhibited by the inferred dynamics is reproduced for parameters corresponding to positive polarity feedback, indicating that such a feedback mechanism may be required to explain the emergence of effective anti-friction in the underdamped nuclear dynamics.

3.4 Results for all constriction widths

In Fig. S12, the same data as in main text Fig. 6 is shown, but with additional model-experiment comparisons. Here, we show all seven bridge widths observed experimentally (with additional widths $W = 9, 18$ µm). Heatmaps are plotted without the interpolation implemented in the main text, and the nucleus velocity auto-correlation is shown as an additional comparison.

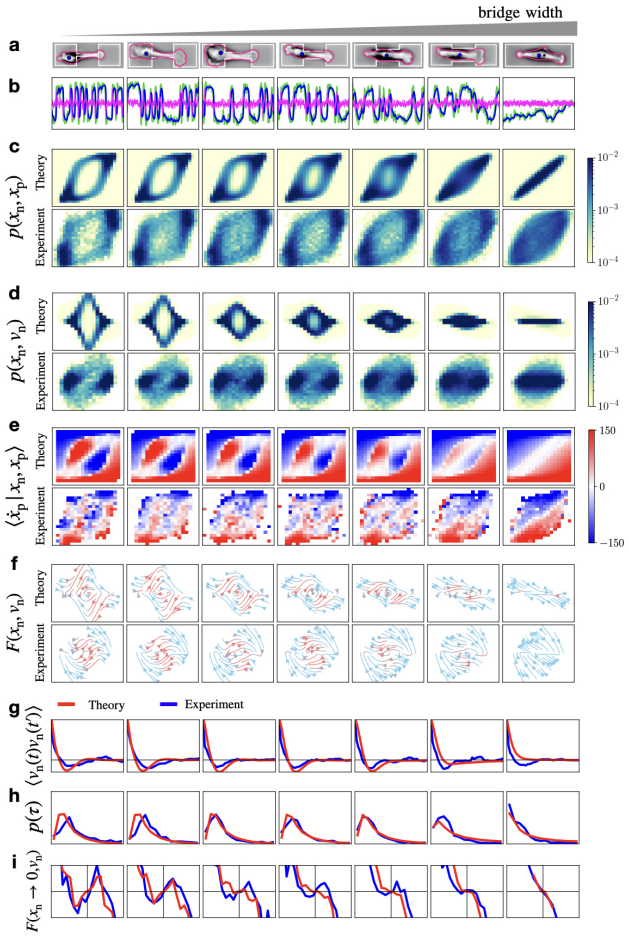


Figure S12: **Dynamics for all constriction widths. a.** Brightfield microscopy images of MDA-MB-231 cells migrating in two-state micropatterns with increasing bridge width W with cell outline in pink, nucleus position in blue, and geometry in white. **b.** Stochastic trajectory $x_n(t)$ (blue), $x_p(t)$ (green), and $P(t)$ (pink) predicted by the spatial feedback model (Eq. (S14)) with increasing γ_{min}, but constant polarity dynamics, as a model for increasing bridge width (Supplementary Section S4). **c.** Joint probability distributions $p(x_n, x_p)$. **d.** Joint probability distributions $p(x_n, v_n)$. **e.** Protrusion velocity maps (PVM) $\langle \dot{x}_p | x_n, x_p \rangle$. **f.** Flow field $(\dot{x}_n, \dot{v}_n) = (v_n, F(x_n, v_n))$ indicated by arrows. Arrow color indicates the direction of the local flow: acceleration is orange and deceleration is blue. **g.** Predicted (red) and experimental (blue) nucleus velocity auto-correlation function $\langle v_n(t) v_n(t') \rangle$, plotted as a function of $|t - t'|$. **h.** Predicted (red) and experimental (blue) dwell time distributions $p(\tau)$. **i.** Predicted (red) and experimental (blue) effective friction at the bridge center $F(x_n \to 0, v_n)$ In panels **c-f**, the top row corresponds to the spatial feedback model prediction, the bottom row to experimental observations.

3.5 Results for varying constriction length

To further test the spatial feedback model, we also change the length L of the constriction (Fig. S13a). To implement this change in the model, we change the total length of the system, $L_{\mathrm{sys}} = 2a + L$, where a is the side length of the square islands (see Supplementary Section S4 for exact implementation). We find that the model captures the main qualitative changes observed in the experiment. Interestingly, we find that the 'polarity driving' becomes more strongly pronounced in the longest constrictions, which provides additional evidence for this effect.

3.6 Extension-dependent polarity model

In this section, we show that an alternative model in which the polarity feedback is sensitive to $\Delta x = |x_{\mathrm{p}} - x_{\mathrm{n}}|$ instead of the absolute position of the protrusion x_{p} is unable to capture our experimental observations. Such a model can be formulated by using polarity dynamics of the form

$$\dot{P} = -\alpha(\Delta x)P - \beta P^3 + \sigma\xi(t) \tag{S8}$$

We expect the polarity to become more persistent for stretched states, with a possible switch to positive feedback at large extensions. As a simple implementation of this dependence, we take α to be a linear function of Δx:

$$\alpha(\Delta x) = \alpha_0 - \alpha_1 \Delta x \tag{S9}$$

A switch to positive feedback therefore occurs at a critical extension $\Delta x_{\mathrm{critical}} = \alpha_0/\alpha_1$. Note that since this model does not couple to geometry, α_0 and α_1 are assumed to be intrinsic cell parameters, which do not adapt to the environment. Therefore, in this model, the bridge width is implemented only through the adhesiveness profile. Of course, we could in principle also consider extension-dependent feedback which changes for changing bridge widths, e.g. using $\alpha_1 = \alpha_1(x_{\mathrm{p}})$. This could be implemented such that the critical extension at which positive feedback is activated increases with increasing bridge width - with similar arguments as those used to motivate the spatial feedback model. However, the spatial feedback model provides simpler implementation of such a geometry dependence, with fewer parameters. Thus, by contrasting these models, we do not seek to rule out a dependence of the polarity dynamics on the extension of the cell, but investigate whether the geometry-sensitive component to these dynamics dominates over the extension-dependent one.

Interestingly, we find that there are large regions in parameter space in which the extension feedback model (Eq. (S14)) captures many of the qualitative features of the experiments with thin constrictions ($w = 3,7\,\mu\mathrm{m}$) (Fig. S14). Specifically, the model exhibits a ring-structure in $p(x_{\mathrm{n}}, x_{\mathrm{p}})$ and $p(x_{\mathrm{n}}, v_{\mathrm{n}})$, amplification in the phase space flow with effective anti-friction, a peaked dwell time distribution, and a matching velocity velocity auto-correlation function. Based on the trajectories, we observe that spikes in the polarity precede transitions, indicating that the activation of positive feedback in this model

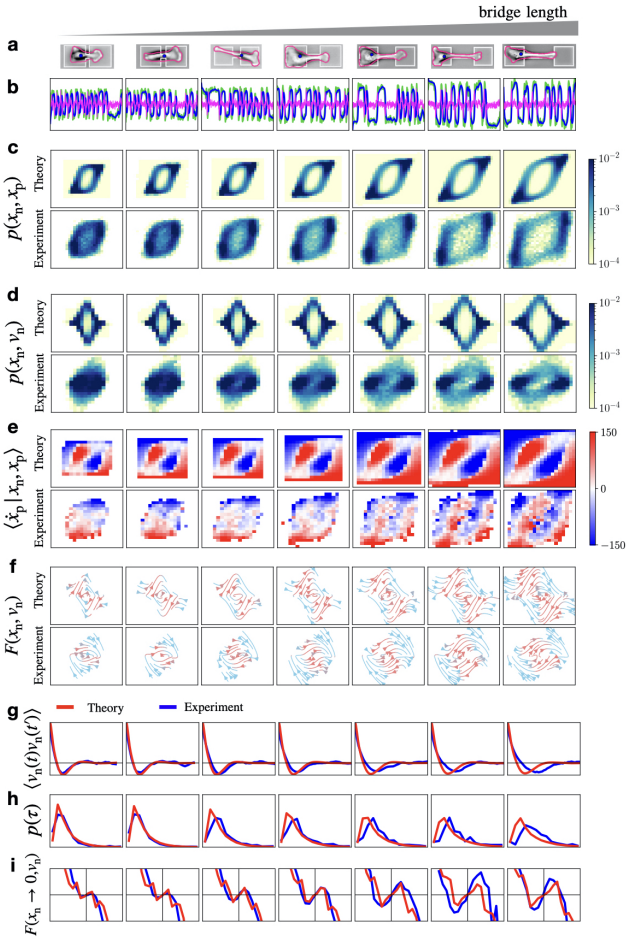


Figure S13: **Model predicts dynamics with varying constriction length. a.** Brightfield microscopy images of MDA-MB-231 cells migrating in two-state micropatterns with increasing bridge length L with cell outline in pink, nucleus position in blue, and geometry in white. **b.** Stochastic trajectory $x_n(t)$ (blue), $x_p(t)$ (green), and $P(t)$ (pink) predicted by the spatial feedback model (Eq. (S14)) with increasing L_{sys} as a model for increasing bridge length (Supplementary Section S4). **c.** Joint probability distributions $p(x_n, x_p)$. **d.** Joint probability distributions $p(x_n, v_n)$. **e.** Protrusion velocity maps (PVM) $\langle \dot{x}_p | x_n, x_p \rangle$. **f.** Flow field $(\dot{x}_n, \dot{v}_n) = (v_n, F(x_n, v_n))$ indicated by arrows. Arrow color indicates the direction of the local flow: acceleration is orange and deceleration is blue. **g.** Predicted (red) and experimental (blue) nucleus velocity auto-correlation function $\langle v_n(t) v_n(t') \rangle$, plotted as a function of $|t - t'|$. **h.** Predicted (red) and experimental (blue) dwell time distributions $p(\tau)$. **i.** Predicted (red) and experimental (blue) effective friction at the bridge center $F(x_n \to 0, v_n)$ In panels **c-f**, the top row corresponds to the spatial feedback model prediction, the bottom row to experimental observations.

is also correlated with the extension preceding bridge entry of the protrusion. Accordingly, the model predicts an extensile pushing effect in the PVM, although, unlike in the experimental PVM, it is not restricted to positions in the bridge.

Since α_0 and α_1 do not change with bridge width, even cells in systems without constrictions experience positive feedback when they are in a stretched state $\Delta x > \Delta x_{critical}$. Experimentally, we find that the distribution of cell extensions $p(\Delta x)$ does not change significantly with bridge width, suggesting that based on the extension feedback model,

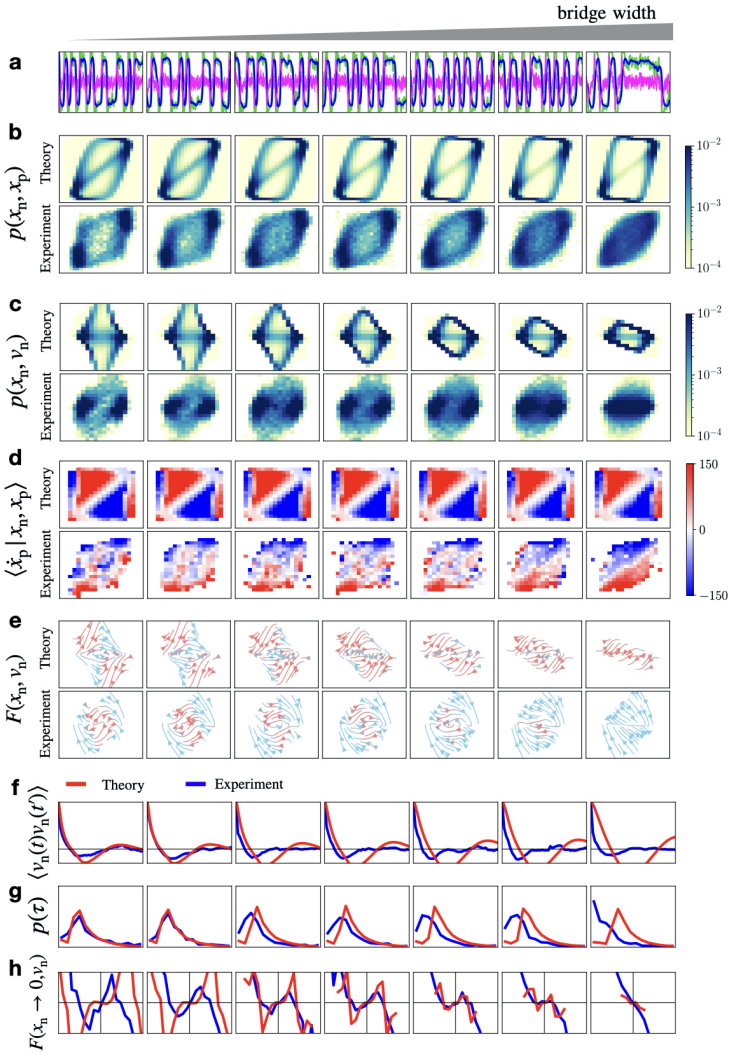


Figure S14: **Extension feedback fails to predict dynamics with varying constriction width. a.** Stochastic trajectory $x_n(t)$ (blue), $x_p(t)$ (green), and $P(t)$ (pink) predicted by the spatial feedback model (Eq. (S14)) with increasing γ_{min} and α_{min} as a model for increasing bridge width W (Supplementary Section S4). **b.** Joint probability distributions $p(x_n, x_p)$. **c.** Joint probability distributions $p(x_n, v_n)$. **d.** Protrusion velocity maps (PVM) $\langle \dot{x}_p | x_n, x_p \rangle$. **e.** Flow field $(\dot{x}_n, \dot{v}_n) = (v_n, F(x_n, v_n))$ indicated by arrows. Arrow color indicates the direction of the local flow: acceleration is orange and deceleration is blue. **f.** Predicted (red) and experimental (blue) nucleus velocity auto-correlation function $\langle v_n(t) v_n(t') \rangle$, plotted as a function of $|t - t'|$. **g.** Predicted (red) and experimental (blue) dwell time distributions $p(\tau)$. **h.** Predicted (red) and experimental (blue) effective friction at the bridge center $F(x_n \to 0, v_n)$. In panels **b-e**, the top row corresponds to the spatial feedback model prediction, the bottom row to experimental observations.

we expect similar polarity dynamics, including positive feedback states also on wide bridges (Fig. S15). This already suggests that this model is unlikely to capture the observed dynamics with increasing bridge width.

To explore the predictions of this model for increasing bridge width directly, we focus on those parameter regimes of $\{\alpha_0, \alpha_1, \sigma\}$ in which the dynamics on thin bridges

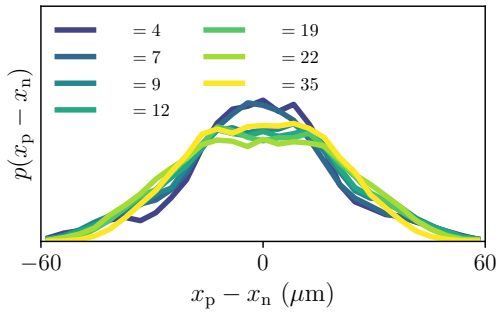


Figure S15: **Experimental probability distributions of protrusion-nucleus extension for all bridge widths**

is captured. We find that the predicted dynamics qualitatively fails to capture the experiments with wide bridges (Fig. S14). Specifically, the model predicts a pronounced ring-like probability distribution even on the rectangular patterns without a constriction, in contrast to the uniform experimental distribution. The model also predicts a peaked dwell time distribution on the rectangles, in contrast to the decaying distribution observed in the experiment. Furthermore, the presence of positive feedback makes a contribution to the polarity component of the phase space flow (see section 3.3), and thus leads to significant amplification in systems with wide bridges, in contradiction to the experiments.

To further pinpoint the shortcomings of the model, we investigate the dependence of the protrusion velocities on the instantaneous protrusion position and protrusion-nucleus extension, $\langle \dot{x}_p | x_p - x_n, x_p \rangle$ (Fig. S16). As expected, we observe that in the extension feedback model, the protrusion velocities are predominantly determined by the protrusion-nucleus extension $x_p - x_n$. This dependence does not change very much with changing bridge width. In contrast, the spatial feedback model predicts a more complex dependence of the protrusion velocity on $x_p - x_n, x_p$: for narrow bridges, for x_p outside of the constriction, the elastic coupling dominates. When x_p is inside the bridge, we observe the pushing effect, whose sign depends on $x_p - x_n$, as positive polarities, which lead to positive $\dot{x}_p$ drive the system towards $x_p - x_n > 0$. As the bridge widens, the pushing contribution disappears. For the rectangular systems without constriction, the elastic coupling dominates, and the protrusion velocities are completely determined by $x_p - x_n$. These features are qualitatively very similar to those seen in the experiments, further supporting the hypothesis that the spatial feedback model is a better description of the data than the extension feedback model.

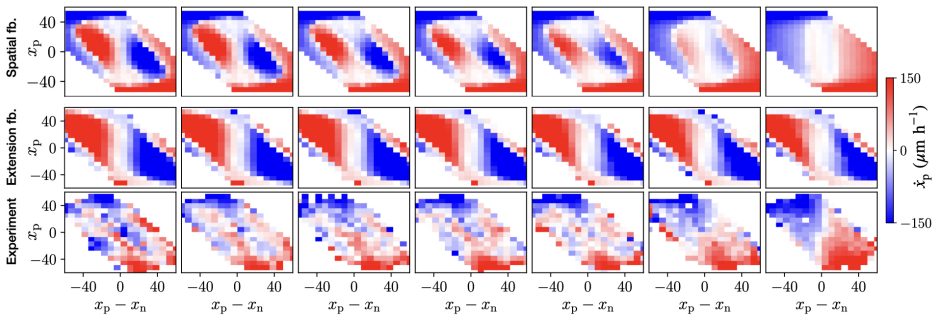


Figure S16: **Protrusion velocities as a function of x_p and $x_p - x_n$ for two model candidates**. The conditional average $\langle \dot{x}_p | x_p - x_n, x_p \rangle$ is shown for the spatial feedback model (top row), the extension feedback model (Eq. (S14)) (center row), and the experiment (bottom row), for all 7 bridge widths (in increasing order from left to right).

4 Model implementation and choice of parameters

4.1 Nucleus dynamics

For the nucleus dynamics, we postulate an equation of motion including a spatially variable adhesiveness:

$$\dot{x}_n = k_n(x_p - x_n)/\gamma(x_n) \tag{S10}$$

where $k_n = k/\zeta_n$ and the dimensionless adhesiveness profile

$$\gamma(x_n) = \frac{1-\gamma_{min}}{2}\left(1 - \cos\left(\frac{x_n \pi}{L_{system}}\right)\right) + \gamma_{min} \tag{S11}$$

Here, $\gamma(x_n)$ varies between γ_{min} at $x_n = 0$ and 1 on the islands. The magnitude of the adhesiveness is accounted for the by the parameter ζ_n. Throughout the paper, we use dimensionful parameters, such that the simulation results can be directly compared to the experimental statistics on the same axes. However, we reduce the number of parameters by constructing parameter combinations such as k_n. For all length-scale parameters, we directly use the known dimension of the experimental confinement, i.e. $L_{system} = a + L/2 = 52.5$ µm, where $a \approx 35$ µm is the side length of the adhesive islands, L the bridge length; for systems with confinement length $L \approx 35$ µm. Thus, Eqn. (S10) has only two free parameters: k_n and γ_{min}. We determine these parameters by fitting Eqn. (S10) to the experimentally observed NVM (Fig. S17a). In Fig. 2 of the main text, we show the result for our standard constriction width $W = 7$ µm. To constrain the parameters used for all constriction widths throughout, we first fit the thinnest constriction width $W = 3$ µm, and obtain $k_n \approx 0.6\ s^{-1}$, $\gamma_{min} \approx 0.2$ (Fig. S17). We find good agreement for all aspects of the NVM (see Fig. S17 for additional cuts through the map not shown in the main text). To make predictions for varying bridge width (Fig. 3 in the main text), we fix k_n, and linearly interpolate γ_{min} between 0.2 and 1 according to bridge width, to achieve a flat adhesiveness profile for systems without constrictions (see plots in Fig. 3a in the main text).

For the alternative deformation model, we take $W(x_n) = Q(1-(x_n/x_0)^2)^2$ with $x_0 = (L+a)/2 = 35$ µm, such that the potential minimum is located in the center of the adhesive island. Using best fit parameters for k_n and Q, we find that this model robustly fails to capture the NVM for various constriction widths (Fig. S17).

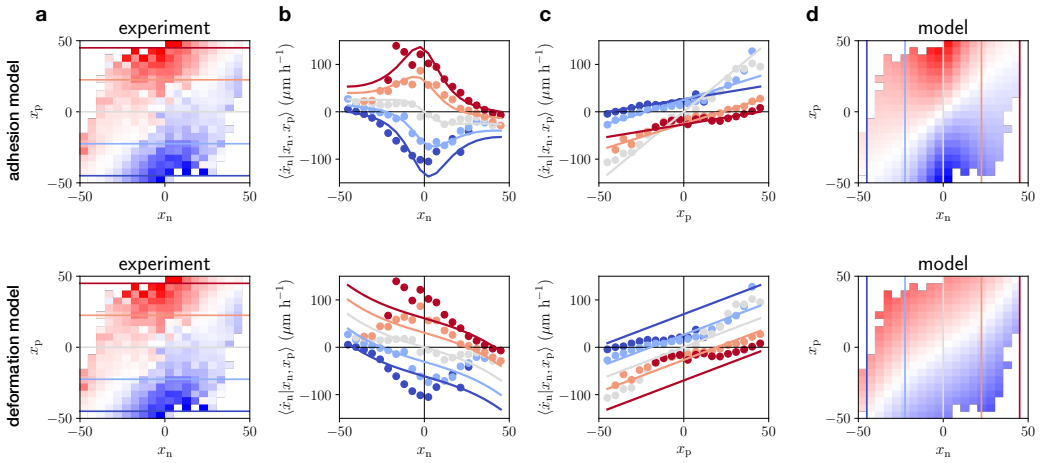


Figure S17: **Parameter estimation by fitting to the NVM. a.** Experimental NVM for bridge width $W = 3$ µm. **b.** Cuts of the NVM along the horizontal lines indicated in panel **a**, i.e. $\dot{x}_p$ as a function of x_n for different x_p.. Dots: Experiment, Line: Fitted model. **c.** Cuts of the NVM along the vertical lines indicated in panel **d**, i.e. $\dot{x}_p$ as a function of x_p for different x_n. Dots: Experiment, Line: Fitted model. **d.** Fitted NVM. *Top row:* adhesion model. *Bottom row:* deformation model.

4.2 Protrusion dynamics

For the protrusion dynamics, we use the equation of motion

$$\dot{x}_\mathrm{p} = -k_\mathrm{p}(x_\mathrm{p} - x_\mathrm{n}) - \partial_{x_\mathrm{p}} V(x_\mathrm{p}) + P(t) \tag{S12}$$

where $k_\mathrm{p} = k/\zeta_\mathrm{p}$ and ζ_p is absorbed into the boundary potential and our definition of the polarity P. We use soft-wall boundary conditions at the system boundaries, using the potential $V(x_\mathrm{p}) = (x_\mathrm{p}/x_\mathrm{boundary})^{2n}$. We find that within a reasonable range, the boundary potential parameters do not strongly affect the results, and take $n = 4$ and $x_\mathrm{boundary} = 0.4 * L_\mathrm{system}$ throughout. Similarly, we find that the choice of k_p does not strongly affect the results. Physically, we expect the friction on the nucleus to be larger than on the protrusion, i.e. $\zeta_\mathrm{p} < \zeta_\mathrm{n}$, and thus $k_\mathrm{p} > k_\mathrm{n}$. We therefore take $k_\mathrm{p} = 1.2\ \mathrm{s}^{-1}$, which we also find to accurately capture the PVM inferred from systems without constrictions, which is dominated by the elastic coupling.

4.3 Polarity dynamics

Unlike for the nucleus and protrusion dynamics, for the polarity, we do not have access to experimental trajectories to constrain our choice of parameters. We therefore perform parameter sweeps over the parameters of the polarity dynamics. Generally, we find that the model behaviour varies smoothly and as physically expected with the parameters, and none of our conclusions depend on fine tuning parameters.

4.3.1 White noise polarity

For the white noise polarity model, $P = \sigma\xi(t)$, there is only a single free parameter, the noise strength σ. As expected, the probability distributions broaden with increasing noise; however they robustly fail to capture the ring structure in $p(x_\mathrm{n}, x_\mathrm{p})$, the peak in the dwell time-distribution, and the flow field $F(x_\mathrm{n}, v_\mathrm{n})$.

4.3.2 Persistent polarity

The persistent polarity model (with $\alpha_0 > 0$)

$$\dot{P} = -\alpha_0 P + \sigma\xi(t) \tag{S13}$$

has two parameters: the noise strength σ and the persistence time α_0^{-1}. Since P is now an integrated noise, it scales differently with time than the white noise model (leading order term scales with $\Delta t^{3/2}$, as opposed to $\Delta t^{1/2}$ in the white noise model). Thus, we take larger noise amplitudes to achieve similar magnitudes of P. Still, we find that this model does not capture the experimentally observed features irrespective of our choice of σ and α_0^{-1} (Fig. S18, S19). In Fig. 4 of the main text, we use $\sigma = 100$ and $\alpha_0 = 1$.

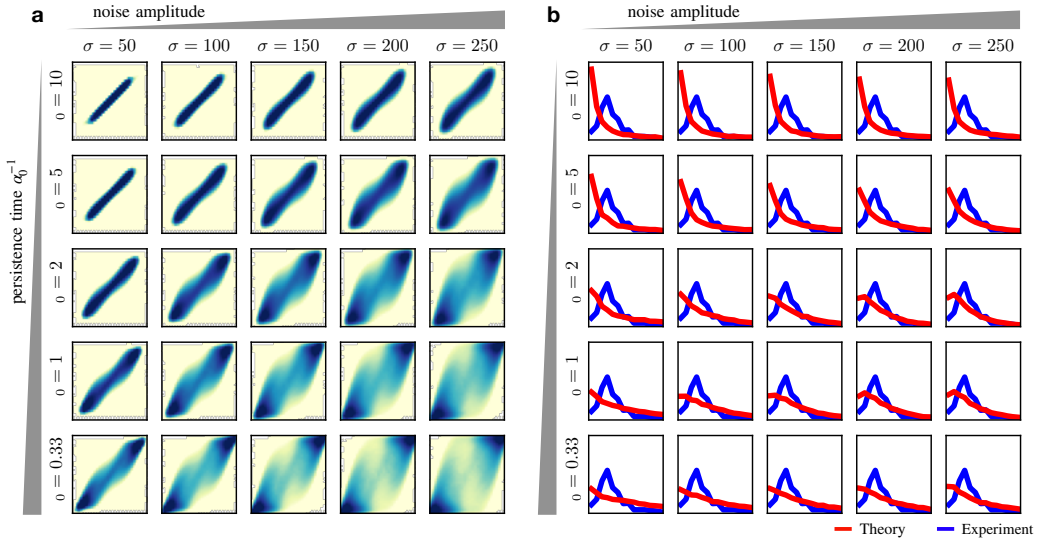


Figure S18: **Predictions of the persistent polarity model with varying noise amplitude and persistence times.** Here, we use $\gamma_{\min} = 0.23$ and compare to the experimental data for bridge width $W = 7$ µm. **a.** Predicted distributions $p(x_n, x_p)$, plotted logarithmically with the same colour axis as in Fig. 1 in the main text. **b.** Experimental (blue) and predicted (red) dwell time distribution. Axis limits are the same as shown in the main text for all panels.

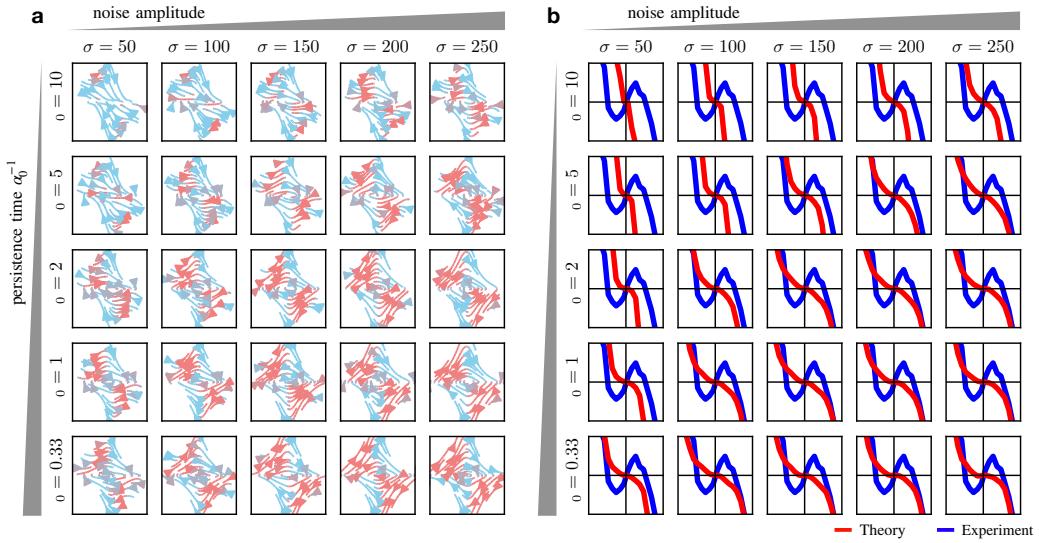


Figure S19: **Predictions of the persistent polarity model with varying noise amplitude and persistence times. a.** Phase-space portrait predicted by model. **b.** Experimental (blue) and predicted (red) effective friction relation $F(x_n \to 0, v_n)$.

4.3.3 Spatial feedback polarity

For the spatial feedback model, we vary the feedback on the polarity as a function of the protrusion position:

$$\dot{P} = -\alpha(x_\mathrm{p})P - \beta P^3 + \sigma\xi(t) \tag{S14}$$

Here, we use a similar spatial profile as for the adhesiveness of the nuclear dynamics, with a minimal value α_min at $x_\mathrm{p} = 0$ and a maximal value α_0 on the adhesive islands:

$$\alpha(x_\mathrm{p}) = \frac{\alpha_0 - \alpha_\mathrm{min}}{2} - \frac{\alpha_0 - \alpha_\mathrm{min}}{2}\cos\left(\frac{x_\mathrm{p}\pi}{L_\mathrm{system}}\right) \tag{S15}$$

Thus, this model has 4 parameters: $\{\alpha_0, \alpha_\mathrm{min}, \beta, \sigma\}$. Importantly, for positions where $\alpha(x_\mathrm{p}) < 0$ locally, the preferred polarity is $P_0 = \pm\sqrt{|\alpha|/\beta}$. Thus, we take values for β which give a reasonable order of magnitude of the preferred polarity compared to the typical order of magnitude of the protrusion velocities. Specifically, we take $\beta = 0.0001$ throughout. On the islands, we postulate that protrusions move with low persistence, and therefore take $\alpha_0 = 10$. This choice yields accurate results for the PVM in systems with no constriction, where we take $\alpha_\mathrm{min} = \alpha_0$, corresponding to a flat profile. Taking much smaller values of α_0 leads to a pushing effect in the PVM for the system with no constriction, which is not observed experimentally. For α_min, we broadly find the behaviour shown in Fig. 5 of the main text: ring structures for $\alpha_\mathrm{min} < 0$, but not for $\alpha_\mathrm{min} > 0$. Similarly, we find that we require a negative α_min to capture the effective anti-friction in the nucleus dynamics (see Supplementary section 3.3 for a parameter sweep). We find that $\alpha_\mathrm{min} = -6.5$ yields accurate predictions for the thinnest constriction, and systematically increase α_min for wider bridges, up to $\alpha_\mathrm{min} = \alpha_0$ for the system without constriction. Finally, the model predictions do not sensitively depend on the choice of the noise amplitude; we take $\sigma = 100$ throughout.

4.3.4 Extension feedback polarity

For the extension feedback model, we vary the feedback on the polarity as a function of the extension of the protrusion away from the nucleus, $\Delta x = |x_\mathrm{p} - x_\mathrm{n}|$:

$$\dot{P} = -\alpha(\Delta x)P - \beta P^3 + \sigma\xi(t) \tag{S16}$$

Physically, we expect the polarity to become more persistent for stretched states, with a possible switch to positive feedback at large extensions. As a simple implementation of this dependence, we take α to be a linear function of Δx:

$$\alpha(\Delta x) = \alpha_0 - \alpha_1\Delta x \tag{S17}$$

A switch to positive feedback therefore occurs at a critical extension $\Delta x_\mathrm{critical} = \alpha_0/\alpha_1$. Thus, this model has 4 parameters: $\{\alpha_0, \alpha_1, \beta, \sigma\}$. In section 3.6, we take $\beta = 0.0005$, $\sigma = 200$, and $\alpha_0 = 10$, to be consistent with the spatial feedback model at small extensions, and $\alpha_1 = 1$, such that the critical extension is $\Delta x_\mathrm{critical} = 10$ µm.

References

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation BT - Medical Image Computing and Computer-Assisted Intervention MICCAI 2015," (Cham), pp. 234–241, Springer International Publishing, 2015.

[2] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, International Conference on Learning Representations, ICLR, 2015.

[3] H. Kervadec, J. Bouchtiba, C. Desrosiers, E. Granger, J. Dolz, and I. B. Ayed, "Boundary loss for highly unbalanced segmentation," in *International conference on medical imaging with deep learning*, pp. 285–296, PMLR, 2019.

[4] 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," *Nature Physics*, vol. 15, no. 6, pp. 595–601, 2019.

[5] S. Siegert, R. Friedrich, and J. Peinke, "Analysis of data sets of stochastic systems," *Phys. Lett. A*, vol. 243, pp. 275–280, mar 1998.

[6] A. Frishman and P. Ronceray, "Learning force fields from stochastic trajectories," *Physical Review X*, vol. 10, no. 2, p. 21009, 2020.

Chapter 7

Inferring the dynamics of underdamped stochastic systems

This chapter is based on the following publication:

Inferring the dynamics of underdamped stochastic systems

David B. Brückner*, Pierre Ronceray*, Chase P. Broedersz†

* equal contribution
† corresponding author

Physical Review Letters 125, 058103 (2020)

Inferring the dynamics of underdamped stochastic systems

David B. Brückner[1⋆], Pierre Ronceray[2⋆] and Chase P. Broedersz[1,3,†]

[1]Arnold Sommerfeld Center for Theoretical Physics and Center for NanoScience, Department of Physics, Ludwig-Maximilian-University Munich, Theresienstr. 37, D-80333 Munich, Germany, [2]Center for the Physics of Biological Function, Princeton University, Princeton, NJ 08544, USA, [3]Department of Physics and Astronomy, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands, ⋆equal contribution †corresponding author

Many complex systems, ranging from migrating cells to animal groups, exhibit stochastic dynamics described by the underdamped Langevin equation. Inferring such an equation of motion from experimental data can provide profound insight into the physical laws governing the system. Here, we derive a principled framework to infer the dynamics of underdamped stochastic systems from realistic experimental trajectories, sampled at discrete times and subject to measurement errors. This framework yields an operational method, Underdamped Langevin Inference (ULI), which performs well on experimental trajectories of single migrating cells and in complex high-dimensional systems, including flocks with Viscek-like alignment interactions. Our method is robust to experimental measurement errors, and includes a self-consistent estimate of the inference error.

Across the scientific disciplines, data-driven methods are used to unravel the dynamics of complex systems. These approaches often take the form of inverse problems, aiming to infer the underlying governing equation of motion from observed trajectories. This problem is well understood for deterministic systems [1–3]. For a broad variety of physical systems, however, a deterministic description is insufficient: fast, unobserved degrees of freedom act as an effective dynamical noise on the observable quantities. Such systems are described by Langevin dynamics, and inferring their equation of motion is notoriously harder: one must then disentangle the stochastic from the deterministic contributions, both of which contribute to shape the trajectory. In molecular-scale systems described by the overdamped Langevin equation, a first-order stochastic differential equation, recently developed techniques make it possible to efficiently reconstruct the dynamics from observed trajectories [4–8]. Many complex systems at larger scales, however, exhibit stochastic dynamics governed by the *underdamped* Langevin equation, a second-order stochastic differential equation. Examples include cell motility [9–13], postural dynamics in animals [14, 15], movement in interacting swarms of fish [16–18], birds [19, 20], and insects [21, 22], as well as dust particles in a plasma [23]. Due to recent advances in tracking technology, the diversity, accuracy, dimensionality, and size of these behavioral data-sets is rapidly increasing [24], resulting in a growing need for accurate inference approaches for high-dimensional underdamped stochastic systems. However, there is currently no rigorous method to infer the dynamics of such underdamped stochastic systems.

Inference from underdamped stochastic systems suffers from a major challenge absent in the overdamped case. In any realistic application, the accelerations of the degrees of freedom must be obtained as discrete second derivatives from the observed position trajectories, which are sampled at discrete intervals Δt. Consequently, a straightforward generalization of the estimators for the force and noise fields of overdamped systems fails: these estimators do not converge to the correct values, even in the limit $\Delta t \to 0$ [25, 26]. To make matters worse, real data is always subject to measurement errors, leading to divergent biases in the discrete estimators [27]. These problems have so far precluded reliable inference in underdamped stochastic systems.

Here, we introduce a general framework, Underdamped Langevin Inference (ULI), that conceptually explains the origin of these biases, and provides an operational scheme to reliably infer the equation of motion of underdamped stochastic systems governed by non-linear force fields and multiplicative noise amplitudes. To provide a method that can be robustly applied to realistic experimental data, we rigorously derive estimators that converge to the correct values for discrete data subject to measurement errors. We demonstrate the power of our method by applying it to experimental trajectories of single migrating cells, as well as simulated complex high-dimensional data sets, including flocks of active particles with Viscek-style alignment interactions.

We consider a general d-dimensional stationary stochastic process $\mathbf{x}(t)$ with components $\{x_\mu(t)\}_{1\leqslant\mu\leqslant d}$ governed by the underdamped Langevin equation

$$\begin{aligned}\dot{x}_\mu &= v_\mu \\ \dot{v}_\mu &= F_\mu(\mathbf{x},\mathbf{v}) + \sigma_{\mu\nu}(\mathbf{x},\mathbf{v})\xi_\nu(t)\end{aligned} \tag{1}$$

which we interpret in the Itô-sense. Throughout, we employ the Einstein summation convention, and $\xi_\mu(t)$ represents a Gaussian white noise with the properties $\langle\xi_\mu(t)\xi_\nu(t')\rangle = \delta_{\mu\nu}\delta(t-t')$ and $\langle\xi_\mu(t)\rangle = 0$. Our aim is to infer the force field $F_\mu(\mathbf{x},\mathbf{v})$ and the noise amplitude $\sigma_{\mu\nu}(\mathbf{x},\mathbf{v})$ from an observed finite trajectory of the process [1].

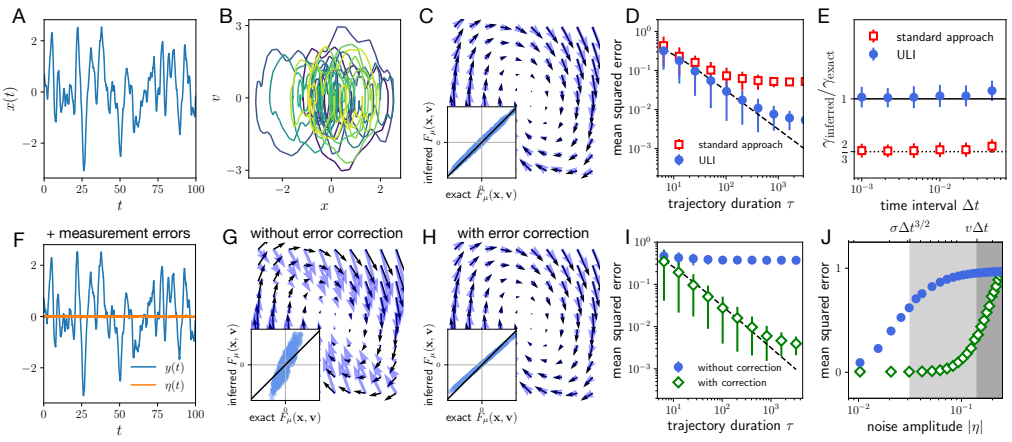


Figure 1: **Inference from discrete time series subject to measurement error. A.** Trajectory $x(t)$ of a stochastic damped harmonic oscillator, $F(x,v) = -\gamma v - kx$. **B.** The same trajectory represented in xv-phase space. Color coding indicates time. **C.** Force field in xv-space inferred from the trajectory in A using ULI with basis functions $b = \{1, x, v\}$ (blue arrows), compared to the exact force field (black arrows). *Inset:* inferred components of the force along the trajectory *versus* the exact values. **D.** Convergence of the mean squared error of the inferred force field, obtained using ULI (circles) and with the previous standard approach [13, 14, 25, 27] (squares). Dashed lines indicate the predicted error $\delta\hat{F}^2/\hat{F}^2 \sim N_b/2\hat{I}_b$. **E.** Inferred friction coefficient γ divided by the exact one as a function of the sampling time interval Δt, comparing the previous standard approach to ULI. **F.** Trajectory $y(t) = x(t) + \eta(t)$ (blue) corresponding to the same realization $x(t)$ in A, with additional time-uncorrelated measurement error $\eta(t)$ (orange) with small amplitude $|\eta| = 0.02$. **G,H.** Force field inferred from $y(t)$ using estimators without and with measurement error corrections, respectively. **I.** Inference convergence for data subject to measurement error using estimators without (circles) and with (diamonds) measurement error corrections. **J.** Dependence of the inference error on the noise amplitude $|\eta|$ (same symbols as in I).

[1] Since we interpret eqn. (1) in the Itô-sense, the inferred force field $F_\mu(\mathbf{x},\mathbf{v})$ corresponds to this convention.

We start by approximating the force field as a linear combination of n_b basis functions $b = \{b_\alpha(\mathbf{x},\mathbf{v})\}_{1\leqslant\alpha\leqslant n_b}$, such as polynomials, Fourier modes, wavelet functions, or Gaussian kernels [14]. From these basis functions, we construct an empirical orthonormal basis $\hat{c}_\alpha(\mathbf{x},\mathbf{v}) = \hat{B}^{-1/2}_{\alpha\beta} b_\beta(\mathbf{x},\mathbf{v})$ such that $\langle \hat{c}_\alpha \hat{c}_\beta\rangle = \delta_{\alpha\beta}$, an approach that was recently proposed for overdamped systems [8]. Here and throughout, averages correspond to time-averages along the trajectory. We can then approximate the force field as $F_\mu(\mathbf{x},\mathbf{v}) \approx F_{\mu\alpha}\hat{c}_\alpha(\mathbf{x},\mathbf{v})$. Similarly, we perform a basis expansion of the noise amplitude $\sigma^2_{\mu\nu}(\mathbf{x},\mathbf{v})$. Thus, the inference problem reduces to estimating the projection coefficients $F_{\mu\alpha}$ and $\sigma^2_{\mu\nu\alpha}$.

Dealing with discreteness

In practice, only the configurational coordinate $\mathbf{x}(t)$ is accessible in experimental data, sampled at a discrete time-interval Δt. We therefore only have access to the discrete estimators of the velocity $\hat{\mathbf{v}}(t) = [\mathbf{x}(t) - \mathbf{x}(t-\Delta t)]/\Delta t$ and acceleration $\hat{\mathbf{a}}(t) = [\mathbf{x}(t+\Delta t) - 2\mathbf{x}(t) + \mathbf{x}(t-\Delta t)]/\Delta t^2$. Our goal is to derive an estimator $\hat{F}_{\mu\alpha}$, constructed from the discrete velocities and accelerations, which converges to the exact projections $F_{\mu\alpha}$ in the limit $\Delta t \to 0$.

An intuitive approach would be to simply generalize the estimators for overdamped systems [8] and calculate the projections of the accelerations $\langle \hat{a}_\mu \hat{c}_\alpha(\mathbf{x},\hat{\mathbf{v}})\rangle$. This expression has indeed previously been used for underdamped systems [13,14,25,27]. We derive the correction term to this estimator by expanding the basis functions $\hat{c}_\alpha(\mathbf{x},\hat{\mathbf{v}}) = \hat{c}_\alpha(\mathbf{x},\mathbf{v}) + (\partial_{v_\mu}\hat{c}_\alpha)(\hat{v}_\mu - v_\mu) + ...$, where the leading order contribution to the second term is a fluctuating (zero average) term of order $\Delta t^{1/2}$. Similarly, we perform a stochastic Itô-Taylor expansion of the discrete acceleration $\hat{\mathbf{a}}(t)$, which has a leading order fluctuating term of order $\Delta t^{-1/2}$. Thus, while each of these terms individually averages to zero, their product results in a bias term with non-zero average of order Δt^0: $\langle \hat{a}_\mu \hat{c}_\alpha(\mathbf{x},\hat{\mathbf{v}})\rangle = F_{\mu\alpha} + \frac{1}{6}\left\langle \sigma^2_{\mu\nu}\partial_{v_\nu} c_\alpha(\mathbf{x},\mathbf{v})\right\rangle + \mathcal{O}(\Delta t)$ (Supplementary Information). As expected, this bias vanishes in the limit $\sigma \to 0$, and therefore does not appear in deterministic systems. However, it poses a problem wherever a second derivative of a stochastic signal is averaged conditioned on its first derivative. The occurrence of such a bias was observed in linear systems [25,26]. Specifically, for a linear viscous force $F(v) = -\gamma v$, it was found that $\langle \hat{a}c(\hat{v})\rangle = -\frac{2}{3}\gamma + \mathcal{O}(\Delta t)$, which is recovered by our general expression for the systematic bias (Supplementary Information).

Previous approaches to correct for this bias rely on *a priori* knowledge of the observed stochastic process [25], are limited to simple parametric forms [26], or perform an *a posteriori* empirical iterative scheme [13]. In contrast, by simply deducting the general form of the bias, we obtain our Underdamped Langevin Inference (ULI) estimator (Supplementary Information):

$$\hat{F}_{\mu\alpha} = \langle \hat{a}_\mu \hat{c}_\alpha(\mathbf{x},\hat{\mathbf{v}})\rangle - \frac{1}{6}\left\langle \widehat{\sigma^2}_{\mu\nu}(\mathbf{x},\hat{\mathbf{v}})\partial_{v_\nu}\hat{c}_\alpha(\mathbf{x},\hat{\mathbf{v}})\right\rangle \tag{2}$$

The presence of the derivative of a basis function in the estimator highlights the importance of projecting the dynamics of underdamped systems onto a set of *smooth* basis functions, in contrast to the traditional approach of taking conditional averages in a discrete set of bins [4,5], equivalent to a basis of non-differentiable top-hat functions.

Similarly to the force field, we expand the noise amplitude as a sum of basis functions, and derive an unbiased estimator for the projection coefficients (Supplementary Information)

$$\widehat{\sigma^2}_{\mu\nu\alpha} = \frac{3\Delta t}{2}\langle \hat{a}_\mu \hat{a}_\nu \hat{c}_\alpha(\mathbf{x},\hat{\mathbf{v}})\rangle \tag{3}$$

To test our method, we start with a simulated minimal example, the stochastic damped harmonic oscillator $\dot{v} = -\gamma v - kx + \sigma\xi$ (Fig. 1A-E). Indeed, we find that even for such a simple system, the intuitive acceleration projections $\langle \hat{a}_\mu \hat{c}_\alpha(\mathbf{x}, \hat{\mathbf{v}}) \rangle$ yield a biased result (Fig. 1E). In contrast, ULI, defined by Eqs. (3) and (2), provides an accurate reconstruction of the force field (Fig. 1C,E). To test the convergence of these estimators in a quantitative way, we calculate the expected random error due to the finite length τ of the input trajectory, $\delta\hat{F}^2/\hat{F}^2 \sim N_b/2\hat{I}_b$, where we define $\hat{I}_b = \frac{\tau}{2}\hat{\sigma}^{-2}_{\mu\nu}\hat{F}_{\mu\alpha}\hat{F}_{\nu\alpha}$ as the empirical estimate of the information contained in the trajectory, and $N_b = dn_b$ is the number of degrees of freedom in the force field [8]. We confirm that the convergence of our estimators follows this expected trend, in contrast to the biased acceleration projections (Fig. 1D). Therefore, ULI provides an operational method to accurately infer the dynamical terms of underdamped stochastic trajectories.

Treatment of measurement errors

A key challenge in stochastic inference from real data is the unavoidable presence of time-uncorrelated random measurement errors $\boldsymbol{\eta}(t)$, which can be non-Gaussian: the observed signal in this case is $\mathbf{y}(t) = \mathbf{x}(t) + \boldsymbol{\eta}(t)$. This problem is particularly dominant in underdamped inference, where the signal is differentiated twice, leading to a divergent bias of order Δt^{-3} (Supplementary Information). Thus, for small Δt, even small measurement errors can lead to prohibitively large systematic inference errors, which cannot be rectified by simply recording more data.

To overcome this challenge, we derive estimators which are robust against measurement error. These estimators are constructed such that the leading-order bias terms cancel. For the force estimator, we find that this is achieved by using the local average position $\overline{\mathbf{x}}(t) = \frac{1}{3}(\mathbf{x}(t-\Delta t) + \mathbf{x}(t) + \mathbf{x}(t+\Delta t))$ and the symmetric velocity $\hat{\mathbf{v}}(t) = [\mathbf{x}(t+\Delta t) - \mathbf{x}(t-\Delta t)]/(2\Delta t)$ in Eq. (2) [2]. Similarly, we derive an unbiased estimator for the noise term, which is constructed using a linear combination of four-point increments (Supplementary Information).

Remarkably, these modifications result in a vastly improved inference performance in the presence of measurement error (Fig. 1F-J). Specifically, while the bias becomes dominant at an error magnitude $|\eta| \sim \sigma\Delta t^{3/2}$ in the standard estimators, the bias-corrected estimators only fail when the measurement error becomes comparable to the displacement in a single time-step, $|\eta| \sim v\Delta t$ (Fig. 1J) (Supplementary Information). Thus, our method has a significantly larger range of validity extending up to the typical displacement in a single time-frame.

Non-linear dynamics

Since our method does not assume linearity, we can expand the projection basis to include higher order functions to capture the behavior of systems with non-linear dynamics. As a canonical example, we study the stochastic Van der Pol oscillator $\dot{v} = \kappa(1 - x^2)v - x + \sigma\xi$, a common model for a broad range of biological dynamical systems [28]. We simulate a short trajectory of this process, with added artificial measurement error (Fig. 2A). Indeed, we find that ULI reliably infers the underlying phase-space flow (Fig. 2B). This is not limited to one-dimensional systems, as shown by studying convergence of higher-dimensional oscillators (Fig. 2C).

[2] Note that due to the change of definition of $\hat{\mathbf{v}}$, the prefactor of the correction term in Eq. (2) changes from $1/6$ to $1/2$.

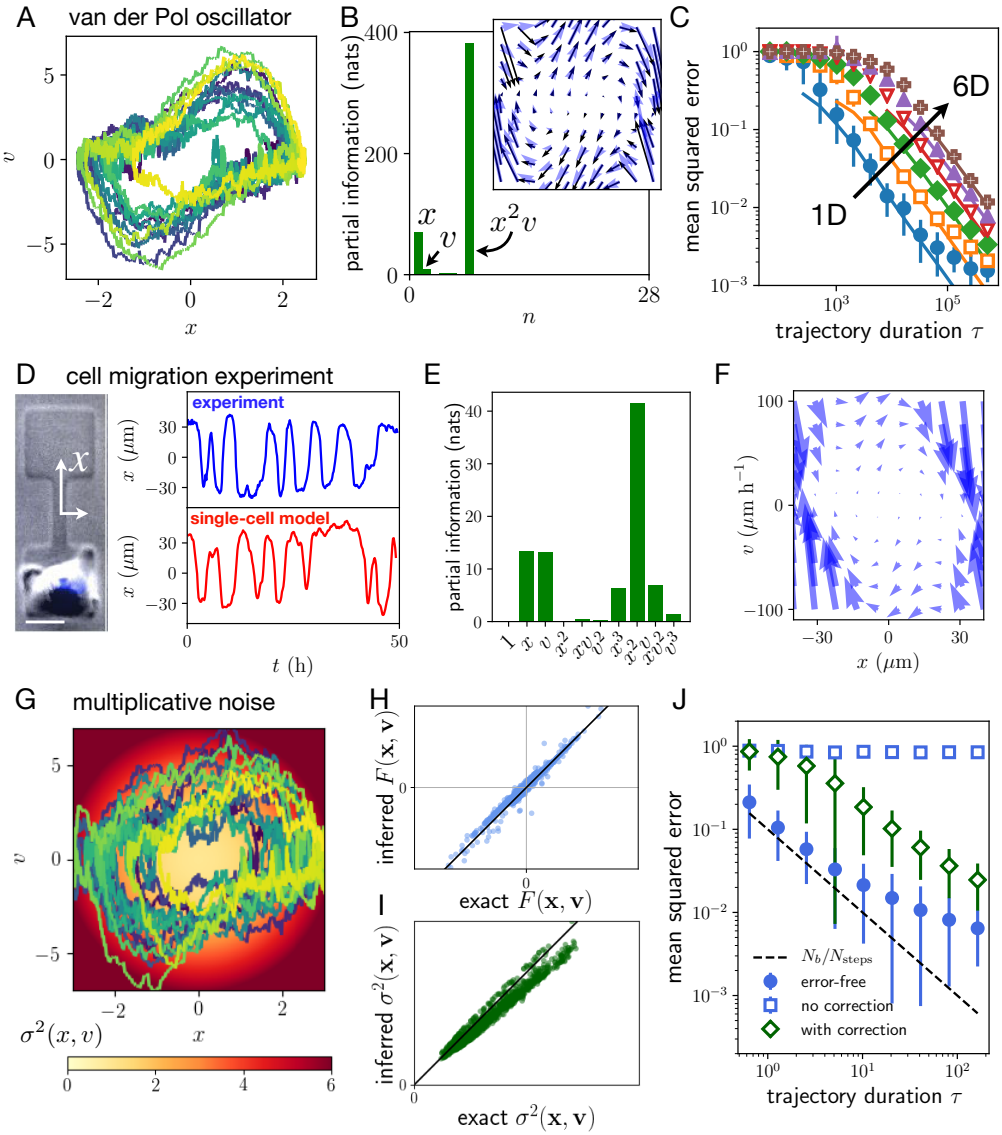


Figure 2: **Inferring non-linear dynamics and multiplicative noise. A.** xv-trajectory of the stochastic Van der Pol oscillator, $F(x,v) = \kappa(1-x^2)v - x$ with measurement error. **B.** Partial information of the 28 basis functions of a 6th order polynomial basis in natural information units (1 nat $= 1/\log 2$ bits), inferred from the trajectory in A. *Inset:* Corresponding force field reconstruction. **C.** Convergence of the inference error for the d-dimensional Van der Pol oscillator $F_\mu(\mathbf{x},\mathbf{v}) = \kappa_\mu(1-x_\mu^2)v_\mu - x_\mu$ (no summation, $1 \leqslant \mu \leqslant d$) with $d = 1...6$, using a third-order polynomial basis. **D.** Microscopy image of a migrating human breast cancer cell (MDA-MB-231) confined in a two-state micropattern (scale bar: 20μm). Experimental trajectory of the cell nucleus position, recorded at a time-interval $\Delta t = 10$ min (blue), and simulated trajectory using the inferred model (red). **E.** Partial information for the experimental trajectory in D, projected onto a third-order polynomial basis. **F.** Deterministic flow field inferred from the experimental trajectory in D. **G.** Trajectory of a Van der Pol oscillator with multiplicative noise $\sigma^2(x,v) = \sigma_0 + \sigma_x x^2 + \sigma_v v^2$ (colormap). **H,I.** Inferred *versus* exact components of the force and noise term, respectively, for the trajectory in G. **J.** Inference convergence of the multiplicative noise amplitude, using Eq. (3) without measurement error (circles), with measurement error (squares), and using the error-corrected estimator (diamonds). The error saturation at large τ is due to the finite time-step. Dashed line: predicted error $\delta\widehat{\sigma^2}/\widehat{\sigma^2} \sim \sqrt{N_b \Delta t/\tau}$ [8].

Importantly, this good performance does not rely on using a polynomial basis to fit a polynomial field: employing a non-adapted basis, such as Fourier components, yields similarly good results (Supplementary Information).

To capture the Van der Pol dynamics, only the three basis functions $\{x, v, x^2v\}$ are required. But can these functions be identified directly from the data without prior knowledge of the underlying force field? To address this question, we introduce the concept of partial information. We can estimate the information contained in a finite trajectory as $\hat{I}_b(n_b) = \frac{\tau}{2}\hat{\sigma}^{-2}_{\mu\nu}\hat{F}_{\mu\alpha}\hat{F}_{\nu\alpha}$, where $\hat{F}_{\nu\alpha}$ are the projection coefficients onto the basis b with n_b basis functions [8]. To assess the importance of the n^{th} basis function in the expansion, we calculate the amount of information it contributes:

$$\hat{I}_b^{(\text{partial})}(n) = \hat{I}_b(n) - \hat{I}_b(n-1) \tag{4}$$

which we term the partial information contributed by the basis function b_n. This approach successfully recovers the relevant terms in large basis sets (Inset Fig. 2B). Thus, the partial information provides a useful heuristic for detecting the relevant terms of the force field.

To illustrate that ULI is practical and data-efficient, we apply it to experimental trajectories of cells migrating in two-state confinements (Fig. 2D). Within their lifetime, these cells perform several transitions between the two states, resulting in relatively short trajectories. Previously, we inferred dynamical properties by averaging over a large ensemble of trajectories [13,29,30]. In contrast, with ULI, we can reliably infer the governing equation of motion from single cell trajectories. Here, $F(x, v)$ corresponds to the deterministic dynamics of the system, and not to a physical force. We employ the partial information to guide our basis selection: indeed, it recovers the intrinsic symmetry of the system, suggesting a symmetrized third order polynomial expansion is a suitable choice (Fig. 2E). Using this expansion, we infer the deterministic flow field of the system (Fig. 2F), which predicts trajectories similar to the experimental ones (Fig. 2D). Importantly, the inferred model is self-consistent: re-inferring from short simulated trajectories yields a similar model (Supplementary Information). Using ULI, we can thus perform inference on small data sets, enabling "single-cell profiling", which could provide a useful tool to characterize cell-to-cell variability [30].

To demonstrate the broad applicability of our approach, we evaluate its performance in the presence of multiplicative noise amplitudes $\sigma_{\mu\nu}(\mathbf{x}, \mathbf{v})$, which occur in a range of complex systems [13,14,31]. ULI accurately recovers the space- and velocity-dependence of both the force and noise field, and the estimators converge to the exact values, even in the presence of measurement errors (Fig. 2G-J). To summarize, we have shown that ULI performs well on short trajectories of non-linear data sets subject to measurement errors, and can accurately infer the spatial structure of multiplicative noise terms.

Collective systems

A major challenge in stochastic inference is the treatment of interacting many-body systems. In recent years, trajectory data on active collective systems, such as collective cell migration [11,12] and animal groups [19–22,32], have become readily available. Previous approaches to such systems frequently focus on the study of correlations [19,33,34] or collision statistics [12,17,32], but no general method for inferring their underlying dynamics has been proposed. The collective behavior of these systems, ranging from disordered swarms [22] to ordered flocking [19], is determined by the interplay of active self-propulsion, cohesive and alignment interactions, and noise. Thus, disentangling these contributions could provide key insights into the physical laws governing active collective systems.

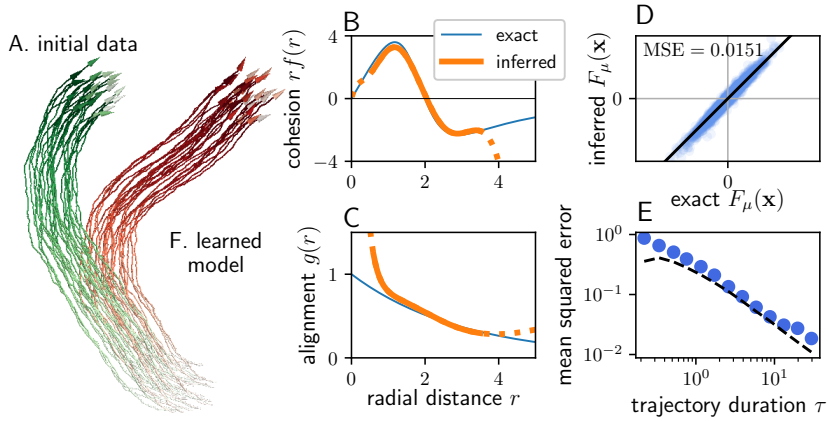


Figure 3: **Interacting flocks. A.** Trajectory (green) of $N = 27$ Viscek-like particles (Eq. 5) in the flocking regime (1000 frames). We perform ULI on this trajectory using a translation-invariant basis of pair interaction and alignment terms, both fitted with $n = 8$ exponential kernels. **B.** Exact (blue) and inferred (orange) cohesion $rf(r)$. Exact form includes short-range repulsion and long-range attraction, $f(r) = \epsilon_0(1-(r/r_0)^3)/((r/r_0)^6+1)$. Dotted inference dependence indicates distances not sampled by the initial data. **C.** Exact and inferred alignment kernel $g(r)$. Exact form: $g(r) = \epsilon_1 \exp(-r/r_1)$. **D.** Inferred *versus* exact components of the force field. **E.** Convergence of the inferred force as a function of trajectory length. Dashed line is the predicted error $\delta\hat{F}^2/\hat{F}^2 \sim N_b/2\hat{I}_b$. **F.** Simulated trajectory (red) employing the inferred force and noise, showing qualitatively similar flocking behavior.

We consider a simple model for the dynamics of a 3D flock with Viscek-style alignment interactions [11,35–37],

$$\dot{\mathbf{v}}_i = \mathbf{p}_i + \sum_{j\neq i}\left[f(r_{ij})\mathbf{r}_{ij} + g(r_{ij})\mathbf{v}_{ij}\right] + \sigma\boldsymbol{\xi}_i \tag{5}$$

where $\mathbf{v}_i = \dot{\mathbf{r}}_i$, $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$, $\mathbf{v}_{ij} = \mathbf{v}_j - \mathbf{v}_i$, and $\mathbf{p}_i = \gamma(v_0^2 - |\mathbf{v}_i|^2)\mathbf{v}_i$ is a self-propulsion force acting along the direction of motion of each particle i. Here, f and g denote the strength of the cohesive and alignment interactions, respectively, as a function of inter-particle distance r_{ij}. This model exhibits a diversity of behaviors, including flocking (Fig. 3A). Intuitively, one might expect that ULI should fail dramatically in such a system: a 3D swarm of N particles has $6N$ degrees of freedom, and "curse of dimensionality" arguments make this problem seem intractable. However, by exploiting the particle exchange symmetry and radial symmetry of the interactions (Supplementary Information), we find that ULI accurately recovers the cohesion and alignment terms (Fig. 3B-C), and captures the full force field (Fig. 3D,E). Furthermore, simulating the inferred model yields trajectories with high similarity to the input data (Fig. 3F). This example illustrates the potential of ULI for inferring complex interactions from trajectories of stochastic many-body systems.

In summary, we demonstrate how to reliably infer the force and noise fields in complex underdamped stochastic systems. We show that the inevitable presence of discreteness and measurement errors result in systematic biases that have so far prohibited accurate inference. To circumvent these problems, we have rigorously derived unbiased estimators, providing an operational framework, Underdamped Langevin Inference, to infer underdamped stochastic dynamics [3]. Our method provides a new avenue to analyzing the dynamics of complex high-dimensional systems, such as assemblies of motile cells [11,12], active swarms [19,21,22,32], as well as non-equilibrium condensed matter systems [23,28,38].

[3] A readily usable PYTHON package to perform Underdamped Langevin Inference is available at https://github.com/ronceray/UnderdampedLangevinInference.

Acknowledgments

We thank Alexandra Fink and Joachim Rädler for generously providing experimental cell trajectories. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project-ID 201269156 - SFB 1032 (Project B12). D.B.B. is supported by a DFG fellowship within the Graduate School of Quantitative Biosciences Munich (QBM) and by the Joachim Herz Stiftung.

References

[1] J. P. Crutchfield and B. McNamara, "Equations of Motion from a Data Series," *Complex Systems*, vol. 1, pp. 417–452, 1987.

[2] B. C. Daniels and I. Nemenman, "Automated adaptive inference of phenomenological dynamical models," *Nature Communications*, vol. 6, pp. 1–8, 2015.

[3] S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Discovering governing equations from data: Sparse identification of nonlinear dynamical systems," *Proc Natl Acad Sci U S A*, vol. 113, no. 15, pp. 3932–3937, 2016.

[4] S. Siegert, R. Friedrich, and J. Peinke, "Analysis of data sets of stochastic systems," *Phys. Lett. A*, vol. 243, pp. 275–280, mar 1998.

[5] M. Ragwitz and H. Kantz, "Indispensable Finite Time Corrections for Fokker-Planck Equations from Time Series Data," *Physical Review Letters*, vol. 87, no. 25, p. 254501, 2001.

[6] M. E. Beheiry, M. Dahan, and J. B. Masson, "InferenceMAP: Mapping of single-molecule dynamics with Bayesian inference," *Nature Methods*, vol. 12, no. 7, pp. 594–595, 2015.

[7] L. Pérez García, J. Donlucas Pérez, G. Volpe, A. V. Arzola, and G. Volpe, "High-performance reconstruction of microscopic force fields from Brownian trajectories," *Nature Communications*, vol. 9, p. 5166, 2018.

[8] A. Frishman and P. Ronceray, "Learning force fields from stochastic trajectories," *Physical Review X*, vol. 10, no. 2, p. 21009, 2020.

[9] D. Selmeczi, S. Mosler, P. H. Hagedorn, N. B. Larsen, and H. Flyvbjerg, "Cell motility as persistent random motion: theories from experiments.," *Biophysical journal*, vol. 89, no. 2, pp. 912–931, 2005.

[10] L. Li, E. C. Cox, and H. Flyvbjerg, "'Dicty dynamics': Dictyostelium motility as persistent random motion," *Physical Biology*, vol. 8, no. 4, p. 046006, 2011.

[11] N. Sepúlveda, L. Petitjean, O. Cochet, E. Grasland-Mongrain, P. Silberzan, and V. Hakim, "Collective Cell Motion in an Epithelial Sheet Can Be Quantitatively Described by a Stochastic Interacting Particle Model," *PLoS Computational Biology*, vol. 9, no. 3, 2013.

[12] J. D'alessandro, A. P. Solon, Y. Hayakawa, C. Anjard, F. Detcheverry, J. P. Rieu, and C. Rivière, "Contact enhancement of locomotion in spreading cell colonies," *Nature Physics*, vol. 13, no. 10, pp. 999–1005, 2017.

[13] 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," *Nature Physics*, vol. 15, no. 6, pp. 595–601, 2019.

[14] G. J. Stephens, B. Johnson-Kerner, W. Bialek, and W. S. Ryu, "Dimensionality and Dynamics in the Behavior of C. elegans," *PLoS Comput Biol*, vol. 4, no. 4, p. e1000028, 2008.

[15] G. J. Stephens, M. Bueno, D. Mesquita, W. S. Ryu, and W. Bialek, "Emergence of long timescales and stereotyped behaviors in Caenorhabditis elegans," *Proc. Natl. Acad. Sci. USA*, vol. 108, pp. 7286–7289, 2011.

[16] J. Gautrais, F. Ginelli, R. Fournier, S. Blanco, M. Soria, H. Chaté, and G. Theraulaz, "Deciphering Interactions in Moving Animal Groups," *PLoS Computational Biology*, vol. 8, no. 9, 2012.

[17] Y. Katz, K. Tunstrøm, C. C. Ioannou, C. Huepe, and I. D. Couzin, "Inferring the structure and dynamics of interactions in schooling fish," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 46, pp. 18720–18725, 2011.

[18] J. Jhawar, R. G. Morris, U. R. Amith-Kumar, M. Danny Raj, T. Rogers, H. Rajendran, and V. Guttal, "Noise-induced schooling of fish," *Nature Physics*, vol. 16, pp. 488–493, 2020.

[19] A. Cavagna, A. Cimarelli, I. Giardina, G. Parisi, R. Santagati, F. Stefanini, and M. Viale, "Scale-free correlations in starling flocks," *Proceedings of the National Academy of Sciences*, vol. 107, no. 26, pp. 11865–11870, 2010.

[20] A. Attanasi, A. Cavagna, L. Del Castello, I. Giardina, T. S. Grigera, A. Jelic, S. Melillo, L. Parisi, O. Pohl, E. Shen, and M. Viale, "Information transfer and behavioural inertia in starling flocks," *Nature Physics*, vol. 10, no. 9, pp. 691–696, 2014.

[21] J. Buhl, D. J. T. Sumpter, I. D. Couzin, J. J. Hale, E. Despland, E. R. Miller, and S. J. Simpson, "From Disorder to Order in Marching Locusts," *Science*, vol. 312, no. 5778, pp. 1402–1406, 2006.

[22] A. Attanasi, A. Cavagna, L. Del Castello, I. Giardina, S. Melillo, L. Parisi, O. Pohl, B. Rossaro, E. Shen, E. Silvestri, and M. Viale, "Collective Behaviour without Collective Order in Wild Swarms of Midges," *PLoS Computational Biology*, vol. 10, no. 7, p. e1003697, 2014.

[23] G. Gogia and J. C. Burton, "Emergent Bistability and Switching in a Nonequilibrium Crystal," *Physical Review Letters*, vol. 119, no. 17, p. 178004, 2017.

[24] A. E. Brown and B. de Bivort, "Ethology as a physical science," *Nature Physics*, vol. 14, pp. 653–657, 2018.

[25] J. N. Pedersen, L. Li, C. Gradinaru, R. H. Austin, E. C. Cox, and H. Flyvbjerg, "How to connect time-lapse recorded trajectories of motile microorganisms with dynamical models in continuous time," *Physical Review E*, vol. 94, no. 6, p. 062401, 2016.

[26] F. Ferretti, V. Chardès, T. Mora, A. M. Walczak, and I. Giardina, "Building General Langevin Models from Discrete Datasets," *Physical Review X*, vol. 10, no. 3, p. 031018, 2020.

[27] B. Lehle and J. Peinke, "Analyzing a stochastic time series obeying a second-order differential equation," *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 91, no. 6, p. 062113, 2015.

[28] K. Kruse and F. Jülicher, "Oscillations in cell biology," *Current Opinion in Cell Biology*, vol. 17, no. 1, pp. 20–26, 2005.

[29] A. Fink, D. B. Brückner, C. Schreiber, P. J. Röttgermann, C. P. Broedersz, and J. O. Rädler, "Area and Geometry Dependence of Cell Migration in Asymmetric Two-State Micropatterns," *Biophysical Journal*, vol. 118, no. 3, pp. 552–564, 2020.

[30] D. B. Brückner, A. Fink, J. O. Rädler, and C. P. Broedersz, "Disentangling the Behavioural Variability of Confined Cell Migration," *J. R. Soc. Interface*, vol. 17, p. 20190689, 2020.

[31] R. Friedrich, J. Peinke, and C. Renner, "How to Quantify Deterministic and Random Influences on the Statistics of the Foreign Exchange Market," *Physical Review Letters*, vol. 84, no. 22, pp. 5224–5227, 2000.

[32] R. Lukeman, Y. X. Li, and L. Edelstein-Keshet, "Inferring individual rules from collective behavior," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 28, pp. 12576–12580, 2010.

[33] A. Cavagna, I. Giardina, and T. S. Grigera, "The physics of flocking: Correlation as a compass from experiments to theory," *Physics Reports*, vol. 728, pp. 1–62, 2017.

[34] W. Bialek, A. Cavagna, I. Giardina, T. Mora, E. Silvestri, M. Viale, and A. M. Walczak, "Statistical mechanics for natural flocks of birds," *Proceedings of the National Academy of Sciences*, vol. 109, no. 13, pp. 4786–4791, 2012.

[35] Tamas Vicsek, A. Czirok, E. Ben-Jacob, I. Cohen, and O. Shochet, "Novel Type of Phase Transition in a System of Self-Driven Particles," *Phys. Rev. Lett.*, vol. 74, no. 6, 1995.

[36] G. Grégoire, H. Chaté, and Y. Tu, "Moving and staying together without a leader," *Physica D: Nonlinear Phenomena*, vol. 181, no. 3-4, pp. 157–170, 2003.

[37] H. Chaté, F. Ginelli, G. Grégoire, F. Peruani, and F. Raynaud, "Modeling collective motion: Variations on the Vicsek model," *European Physical Journal B*, vol. 64, no. 3-4, pp. 451–456, 2008.

[38] M. Baldovin, A. Puglisi, and A. Vulpiani, "Langevin equations from experimental data: The case of rotational diffusion in granular media," *PLoS ONE*, vol. 14, no. 2, pp. 1–16, 2019.

Supplementary Material:

Inferring the dynamics of underdamped stochastic systems

David B. Brückner*, Pierre Ronceray* and Chase P. Broedersz

This Supplemental Material contains a detailed definition of the projection formalism we employ in Underdamped Langevin Inference (ULI) (section 1), derivations of the unbiased estimators for the force and noise fields for discrete data (section 2) and for discrete data with random measurement errors (section 3), a criterion to choose the optimal basis size n_b (section 4), further details on the inference from experimental single cell trajectories (section 5) and detailed information on the models and parameters used for the simulation results shown in Figs. 1-3 in the main text (section 6).

Contents

1 Definition of the projection formalism

We consider d-dimensional processes $\mathbf{x}(t)$, governed by the underdamped Langevin equation

$$\begin{aligned} \dot{x}_\mu &= v_\mu \\ \dot{v}_\mu &= F_\mu(\mathbf{x},\mathbf{v}) + \sigma_{\mu\nu}(\mathbf{x},\mathbf{v})\xi_\nu(t) \end{aligned} \tag{S1}$$

Here, $x_\mu(t)$ are the components of the vector $\mathbf{x}(t)$ and $1 \leqslant \mu \leqslant d$. The term $F_\mu(\mathbf{x},\mathbf{v})$ denotes the μ-component of the force field, and $\sigma_{\mu\nu}(\mathbf{x},\mathbf{v})$ is the noise strength tensor, which is multiplicative and can thus depend on the state of the system given by $\{\mathbf{x},\mathbf{v}\}$. $\xi_\mu(t)$ represents a Gaussian white noise with the properties $\langle \xi_\mu(t)\xi_\nu(t')\rangle = \delta_{\mu\nu}\delta(t-t')$ and $\langle \xi_\mu(t)\rangle = 0$. Our aim is to infer the force field $F_\mu(\mathbf{x},\mathbf{v})$ and the noise amplitude $\sigma_{\mu\nu}(\mathbf{x},\mathbf{v})$ from an observed trajectory of the process. We interpret this stochastic differential equation in the Itô-sense, and thus infer the force field $F_\mu(\mathbf{x},\mathbf{v})$ corresponding to this convention, which may include spurious drift terms due to multiplicative noise amplitudes.

In ULI, we approximate the force field as a linear combination of basis functions $b = \{b_\alpha(\mathbf{x},\mathbf{v})\}$ where the index α runs over all basis functions in the set, $1 \leqslant \alpha \leqslant n_b$. Here and throughout the main text and supplementary material, we employ the Einstein summation convention. Thus, summations over the basis functions, indexed by $\{\alpha,\beta\}$ run from $1...n_b$. Summations over the d-dimensional dynamical quantities such as $x_\mu(t), v_\mu(t), F_\mu(\mathbf{x},\mathbf{v}), \sigma_{\mu\nu}(\mathbf{x},\mathbf{v})$, indexed by $\{\mu,\nu,\rho,\tau,...\}$, run from $1...d$.

To extract the coefficients of the expansions of the force and noise terms, we can project the dynamics onto the space spanned by $b_\alpha(\mathbf{x},\mathbf{v})$ using the steady-state probability distribution $P(\mathbf{x},\mathbf{v})$ as a measure [1]. To do so, we define orthonormalized projectors $c_\alpha(\mathbf{x},\mathbf{v}) = B^{-1/2}_{\alpha\beta} b_\beta(\mathbf{x},\mathbf{v})$, such that

$$\langle c_\alpha c_\beta\rangle = \int c_\alpha(\mathbf{x},\mathbf{v}) c_\beta(\mathbf{x},\mathbf{v}) P(\mathbf{x},\mathbf{v}) \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{v} = \delta_{\alpha\beta} \tag{S2}$$

We then approximate the force field as a linear combination of these basis functions

$$F_\mu(\mathbf{x},\mathbf{v}) \approx F_{\mu\alpha} c_\alpha(\mathbf{x},\mathbf{v}) \tag{S3}$$

Note that if we use a complete set of basis functions, this becomes an exact equality. In any real application however, a truncated set of basis functions must be used, in order to limit the number of parameters (*i.e.* the coefficients $F_{\mu\alpha}$) to be inferred from a trajectory of finite length. The projection coefficients $F_{\mu\alpha}$ are given by

$$F_{\mu\alpha} = \int F_\mu(\mathbf{x},\mathbf{v}) c_\alpha(\mathbf{x},\mathbf{v}) P(\mathbf{x},\mathbf{v}) \mathrm{d}\mathbf{x}\, \mathrm{d}\mathbf{v} \tag{S4}$$

These coefficients thus form a $(d \times n_b)$ matrix, and $F_{\mu\alpha}$ gives the projection coefficient of the μ-component of the force field onto the basis function c_α. Similarly, we expand the noise term

$$\sigma^2_{\mu\nu}(\mathbf{x},\mathbf{v}) \approx \sigma^2_{\mu\nu\alpha} c_\alpha(\mathbf{x},\mathbf{v}) \tag{S5}$$

with the projection coefficients

$$\sigma^2_{\mu\nu\alpha} = \int \sigma^2_{\mu\nu}(\mathbf{x},\mathbf{v}) c_\alpha(\mathbf{x},\mathbf{v}) P(\mathbf{x},\mathbf{v}) \mathrm{d}\mathbf{x}\, \mathrm{d}\mathbf{v} \tag{S6}$$

Note that we expand σ^2 rather than σ because we can only derive estimators for σ^2; since the noise averages to zero, we must take squares of the increments to extract the magnitude of the fluctuations.

In practice, we aim to infer the force and noise fields governing the dynamics of a system from a single trajectory of finite length τ, sampled at a time interval Δt. Thus, the exact probability distribution $P(\mathbf{x},\mathbf{v})$ is unknown, and we cannot enforce the condition Eq. (S2) exactly. Thus, we define *empirical* orthonormalized projectors

$$\hat{c}_\alpha(\mathbf{x},\mathbf{v}) = \hat{B}^{-1/2}_{\alpha\beta} b_\beta(\mathbf{x},\mathbf{v}) \tag{S7}$$

where

$$\hat{B}_{\alpha\beta} = \frac{\Delta t}{\tau} \sum_t b_\alpha(\mathbf{x}(t),\mathbf{v}(t)) b_\beta(\mathbf{x}(t),\mathbf{v}(t)), \tag{S8}$$

such that $\langle \hat{c}_\alpha \hat{c}_\beta \rangle = \delta_{\alpha\beta}$ where $\langle ... \rangle$ refers to a time-average along the observed trajectory.

Our aim in performing inference is to find the terms $F_\mu(\mathbf{x},\mathbf{v})$ and $\sigma^2_{\mu\nu}(\mathbf{x},\mathbf{v})$. Thus, we search for an operational definition of the estimators of the projection coefficients, $\hat{F}_{\mu\alpha}$ and $\widehat{\sigma^2}_{\mu\nu\alpha}$. These estimators consists of increment-constructions projected onto the trajectory-dependent orthonormal basis functions, constructed in such a way that the leading order term in Δt converge to the exact projection coefficients. Due to the Gaussian nature of the stochastic noise, this projection procedure – which is equivalent to a least-square regression of the local estimator – corresponds for the force field to a maximum-likelihood approximation [1].

2 Derivation of the discrete estimators

To derive the leading order bias in the estimators for F and σ, we start by defining the increments of the positions:

$$\Delta x^{(n)}_\mu(t) = x_\mu(t + n\Delta t) - x_\mu(t) \tag{S9}$$

The estimator for the accelerations is then given by a linear combination of these increments:

$$\hat{a}_\mu(t) = \frac{\Delta x^{(2)}_\mu(t) - 2\Delta x^{(1)}_\mu(t)}{\Delta t^2} = \frac{x_\mu(t+2\Delta t) - 2x_\mu(t+\Delta t) + x_\mu(t)}{\Delta t^2} \tag{S10}$$

Note that this is not in general the most natural way to define $\hat{a}_\mu(t)$, as this expression is not centered around t. However, it makes the expression causal: the noise at $t' > t$ is independent of the state at t, thus using forward increments significantly simplifies

the calculations. We will later shift the definition back to a centered one, which will only add higher order terms to our results. Similarly, we define the discrete velocity estimator

$$\hat{v}_\mu(t;\lambda) = \frac{\lambda \Delta x_\mu^{(2)}(t)}{2\Delta t} + \frac{(1-\lambda)\Delta x_\mu^{(1)}(t)}{\Delta t} \tag{S11}$$

Note that we have kept some freedom in how we calculate the velocities from the three points $\{t, t+\Delta t, t+2\Delta t\}$, denoted by the parameter λ. While most previous approaches [2–5] use $\lambda = 0$, we will later show that in the presence of measurement errors, we have to choose $\lambda = 1$ (*i.e.* $\hat{v}$ odd under time reversal around $t + \Delta t$) to obtain an unbiased estimator. For now, we keep it as a variable parameter.

2.1 Itô integrals

Throughout this appendix, we will make use of Itô integrals [6], defined as follows:

$$I_0^{(n)} = \int_t^{t+n\Delta t} \mathrm{d}s = n\Delta t \tag{S12}$$

$$I_{00}^{(n)} = \int_t^{t+n\Delta t} \mathrm{d}s \int_t^s \mathrm{d}s' = (n\Delta t)^2 \tag{S13}$$

$$I_\mu^{(n)} = \int_t^{t+n\Delta t} \mathrm{d}\xi_\mu(s) \tag{S14}$$

$$I_{0\mu}^{(n)} = \int_t^{t+n\Delta t} \mathrm{d}s \int_t^s \mathrm{d}\xi_\nu(s') \tag{S15}$$

$$I_{\mu\nu}^{(n)} = \int_t^{t+n\Delta t} \mathrm{d}\xi_\mu(s) \int_t^s \mathrm{d}\xi_\nu(s') \quad \text{etc.} \tag{S16}$$

Throughout the text, we will frequently make use of the following identity

$$\langle I_{0\mu}^{(n)} I_{0\nu}^{(m)} \rangle = (\Delta t^3)\delta_{\mu\nu} f_{nm} \quad \text{where} \quad f_{nm} = \begin{cases} 1/3 & n = m = 1 \\ 5/6 & n = 1, m = 2 \\ 8/3 & n = m = 2 \\ 4/3 & n = 1, m = 3 \\ 14/3 & n = 2, m = 3 \\ 9 & n = m = 3 \end{cases} \tag{S17}$$

2.2 Force field

A first intuitive guess for the estimator of the force projections $F_{\mu\alpha}$ are the average projections of the acceleration

$$\hat{A}_{\mu\alpha} = \langle \hat{a}_\mu c_\alpha(\mathbf{x}, \hat{\mathbf{v}}) \rangle \tag{S18}$$

and indeed, this quantity has been used as a proxy for $F_{\mu\alpha}$ throughout the literature [2–5]. To rigorously derive the leading order contributions to this quantity in terms of the

dynamical terms F_μ and $\sigma_{\mu\nu}$, we start by expanding the increments

$$\Delta x_\mu^{(n)} = v_\mu I_0^{(n)} + \int_t^{t+n\Delta t} \mathrm{d}s (v_\mu(s) - v_\mu) \tag{S19}$$

$$= v_\mu I_0^{(n)} + \int_t^{t+n\Delta t} \mathrm{d}s \left[\int_t^s \mathrm{d}s' F_\mu(\mathbf{x}(s'), \mathbf{v}(s')) + \int_t^s \mathrm{d}\xi_\nu(s') \sigma_{\mu\nu}(\mathbf{x}(s'), \mathbf{v}(s')) \right] \tag{S20}$$

$$= v_\mu I_0^{(n)} + \sigma_{\mu\nu} I_{0\nu}^{(n)} + F_\mu I_{00}^{(n)} + (\partial_{v_\rho} \sigma_{\mu\nu}) \sigma_{\rho\tau} I_{0\nu\tau}^{(n)} + \mathcal{O}(\Delta t^{5/2}) \tag{S21}$$

where we defined $v_\mu \equiv v_\mu(t)$, $F_\mu \equiv F_\mu(\mathbf{x}(t), \mathbf{v}(t))$, etc. We will use this short-hand notation as well as the Einstein summation convention throughout.

Next, we expand the basis functions $c_\alpha(\mathbf{x}, \hat{\mathbf{v}})$ around the true velocities $\mathbf{v}$:

$$c_\alpha(\mathbf{x}, \hat{\mathbf{v}}) = c_\alpha(\mathbf{x}, \mathbf{v}) + (\partial_{v_\rho} c_\alpha)(\hat{v}_\rho - v_\rho) + \frac{1}{2} (\partial^2_{v_\rho v_\tau} c_\alpha)(\hat{v}_\rho - v_\rho)(\hat{v}_\tau - v_\tau) + \mathcal{O}(\Delta t^{3/2}) \tag{S22}$$

From Eq. (S21), the leading order term of $\hat{v}_\rho - v_\rho$ is given by

$$\hat{v}_\rho - v_\rho = \frac{\lambda}{2\Delta t} \sigma_{\rho\nu} I_{0\nu}^{(1)} + \frac{1-\lambda}{\Delta t} \sigma_{\rho\nu} I_{0\nu}^{(2)} + \mathcal{O}(\Delta t) \tag{S23}$$

Thus, the leading order contribution to the second term of Eq. (S23) is a fluctuating (zero average) term of order $\Delta t^{1/2}$. To evaluate Eq. (S18), we also need the acceleration estimator $\hat{a}_\mu$. Substituting Eq. (S21) into Eq. (S10), we find the leading order terms of the acceleration estimator

$$\hat{a}_\mu = \frac{1}{\Delta t^2} \left[\sigma_{\mu\nu} (I_{0\nu}^{(2)} - 2 I_{0\nu}^{(1)}) + F_\mu \Delta t^2 \right] + \mathcal{O}(\Delta t^{3/2}) \tag{S24}$$

Thus, the leading order contribution to the acceleration is a fluctuating (zero average) term of order $\Delta t^{-1/2}$.

When we evaluate Eq. (S18) by substituting Eq. (S21) and (S23), we obtain

$$\hat{A}_{\mu\alpha} = \langle F_\mu c_\alpha(\mathbf{x}, \mathbf{v}) \rangle + \frac{1+2\lambda}{6} \left\langle (\partial_{v_\rho} c_\alpha(x, v)) \sigma_{\rho\nu} \sigma_{\mu\nu} \right\rangle + \mathcal{O}(\Delta t) \tag{S25}$$

The second term in this expression is an $\mathcal{O}(\Delta t^0)$-bias which means that the acceleration projections do not converge to the projections of the force, even in the limit of infinite sampling rate ($\Delta t \to 0$). This cross-term originates from the product of the fluctuating terms in the basis functions (of order $\Delta t^{1/2}$) and the accelerations (of order $\Delta t^{-1/2}$), which multiplied together give a term of order Δt^0 with non-zero average.

Our expression for the $\mathcal{O}(\Delta t^0)$-bias has several interesting properties:

- As one might expect, it vanishes in the deterministic limit $\sigma \to 0$; it is thus a property of stochastic systems.

- It vanishes for purely positional terms in the force-field, as it depends on the derivative $\partial_{v_\rho} c_\alpha(x, v)$. This makes sense, since it originates from the $\hat{v}$-dependence of the basis functions (Eq. (S23)). As shown by our derivation, it is a consequence of averaging the second derivative of a stochastic signal conditioned on its first derivative.

- A seemingly simple solution to remove the bias would be to set $\lambda = -1/2$. This results in a rather unconventional definition of the discrete velocity estimator,

$$\hat{v}_\mu(t; \lambda = -1/2) = \frac{1}{\Delta t}\left[-\frac{1}{4}x(t + 2\Delta t) + \frac{3}{2}x(t + \Delta t) - \frac{5}{4}x(t)\right] \tag{S26}$$

  for which $\hat{A}_{\mu\alpha}$ is a convergent estimator of $F_{\mu\alpha}$. However, using this definition of $\hat{v}_\mu$ results in large correction terms at the next order in Δt, and thus does not perform well at finite Δt. This estimator would also be strongly biased by measurement errors. For these reasons, we disregard it and turn to the derivation of a better estimator.

For $\lambda \neq -1/2$, the bias does not vanish, and has to be explicitly corrected for. Eq. (S25) allows us to derive an estimator for $F_{\mu\alpha}$ which is unbiased to first order in Δt, *i.e.* which converges to the exact projection coefficients in the limit $\Delta t \to 0$:

$$\hat{F}_{\mu\alpha} = \langle \hat{a}_\mu c_\alpha(\mathbf{x}, \hat{\mathbf{v}})\rangle - \frac{1 + 2\lambda}{6}\left\langle (\partial_{v_\nu} c_\alpha(\mathbf{x}, \hat{\mathbf{v}}))\widehat{\sigma^2_{\mu\nu}}(\mathbf{x}, \hat{\mathbf{v}})\right\rangle \tag{S27}$$

Note that in going from Eq. (S25) to Eq. (S27), we have replaced v and σ by their estimators, as their values are not known. This introduces additional correction terms, but these are of higher order in Δt. Eq. (S27) further implies that the noise term σ^2 has to be inferred before the force field can be inferred. In the presence of measurement errors (section 3), we show below that we must choose $\lambda = 1$, rendering $\hat{v}_\mu$ odd under time reversal around $t + \Delta t$. We therefore use this choice for λ throughout.

Note that Eq. (S27) now conditions the acceleration $\hat{a}_\mu$ (Eq. (S10)) on its first point $x(t)$. In order to make this estimator symmetric, we shift the conditioning $c(\mathbf{x}(t), \hat{\mathbf{v}}(t)) \to c(\mathbf{x}(t + \Delta t), \hat{\mathbf{v}}(t))$. The resulting corrections, due to expanding $c(\mathbf{x}(t + \Delta t), \hat{\mathbf{v}}(t))$ around $\mathbf{x}(t)$, are of higher order in Δt. We can then relabel all time points such that $t \to t - \Delta t$, to arrive at our final formula for the estimator:

$$\hat{F}_{\mu\alpha} = \langle \hat{a}_\mu c_\alpha(\mathbf{x}, \hat{\mathbf{v}})\rangle - \frac{1}{2}\left\langle (\partial_{v_\nu} c_\alpha(\mathbf{x}, \hat{\mathbf{v}}))\widehat{\sigma^2}_{\mu\nu}(\mathbf{x}, \hat{\mathbf{v}})\right\rangle \tag{S28a}$$

$$\mathbf{x} = \mathbf{x}(t) \tag{S28b}$$

$$\hat{\mathbf{v}} = \frac{\mathbf{x}(t + \Delta t) - \mathbf{x}(t - \Delta t)}{2\Delta t} \tag{S28c}$$

$$\hat{\mathbf{a}} = \frac{\mathbf{x}(t + \Delta t) - 2\mathbf{x}(t) + \mathbf{x}(t - \Delta t)}{\Delta t^2} \tag{S28d}$$

2.3 Noise term

To derive an estimator for σ, we derive the leading order contributions to the quantity

$$\Delta t\langle \hat{a}_\mu \hat{a}_\nu \hat{c}_\alpha(\mathbf{x}, \hat{\mathbf{v}})\rangle = \Delta t\langle [\sigma_{\mu\rho} I_{0\rho}^{(2)} - 2\sigma_{\mu\rho} I_{0\rho}^{(1)}][\sigma_{\nu\rho} I_{0\rho}^{(2)} - 2\sigma_{\nu\rho} I_{0\rho}^{(1)}] c_\alpha(\mathbf{x}, \mathbf{v})\rangle + \mathcal{O}(\Delta t) \quad \text{(S29)}$$

$$= \frac{2}{3}\langle \sigma_{\mu\rho}\sigma_{\nu\rho} c_\alpha(\mathbf{x}, \mathbf{v})\rangle + \mathcal{O}(\Delta t) \quad \text{(S30)}$$

where we have used Eqs. (S17), (S21). Here, the somewhat counter-intuitive factor of $2/3$ stems from the expectation values of the Itô-integrals given in Eq. (S17). Thus, an unbiased estimator to first order in Δt for the noise term is

$$\boxed{\widehat{\sigma^2}_{\mu\nu} = \frac{3\Delta t}{2}\langle \hat{a}_\mu \hat{a}_\nu \hat{c}_\alpha(\mathbf{x}, \hat{\mathbf{v}})\rangle} \quad \text{(S31)}$$

2.4 Comparison to the exact formula for a linear damping force

Pedersen et al. [3] calculated the discretization effect for a linear viscous damping force $F(v) = -\gamma v$ (*i.e.* the one-dimensional underdamped Ornstein-Uhlenbeck process; in its discrete form also know as the Persistent Random Walk), to which we can compare our expression for the $\mathcal{O}(\Delta t^0)$-bias. The equation of motion for this process is given by

$$\dot{v} = -\gamma v + \sigma\xi(t) \quad \text{(S32)}$$

In ref. [3], the acceleration projections

$$\langle \hat{a}|\hat{v}\rangle = -\hat{\gamma}\hat{v} \quad \text{(S33)}$$

are considered. Here, $\langle \hat{a}|\hat{v}\rangle$ denotes conditional averaging of $\hat{a}$ with respect to $\hat{v}$, which is equivalent to using a basis of δ-functions, *i.e.* $b_\alpha(v) = \delta(v - v^{(\alpha)})$. Using our definition of the velocity estimator (Eq. (S11)) with $\lambda = 0$, one obtains [3]

$$\hat{\gamma} = \frac{1}{\Delta t}\left[1 - \frac{(1 - e^{-\gamma\Delta t})^2}{2(e^{-\gamma\Delta t} - 1 + \gamma\Delta t)}\right] \approx \frac{2}{3}\gamma - \frac{5}{18}\gamma^2\Delta t + \frac{23}{270}\gamma^3\Delta t^2 + \mathcal{O}(\Delta t^3) \quad \text{(S34)}$$

From Eq. (S25), we expect to find a similar bias, since we are considering a v-dependent component of the force field. To compare Eq. (S34) to our result, we use the basis $b = \{v\}$. Then, the normalised projection coefficient is given by

$$c(v) = \frac{v}{\sqrt{\langle v^2\rangle}} = \frac{\sqrt{2\gamma}}{\sigma} v \quad \text{(S35)}$$

since $\langle v^2\rangle = \sigma^2/2\gamma$ for the Ornstein-Uhlenbeck process. Thus, Eq. (S25) predicts

$$\langle \hat{a}_\mu c_\alpha(\hat{v})\rangle = F_{\mu\alpha} + \frac{(\partial_v c)\sigma^2}{6} + \mathcal{O}(\Delta t) = F_{\mu\alpha} + \frac{\sqrt{2\gamma}\sigma}{6} + \mathcal{O}(\Delta t) \quad \text{(S36)}$$

and therefore

$$\langle \hat{a}_\mu c_\alpha(\hat{v})\rangle c_\alpha(v) = F_\mu + \frac{\gamma}{3} v + \mathcal{O}(\Delta t) = -\frac{2}{3}\gamma v + \mathcal{O}(\Delta t) \quad \text{(S37)}$$

Thus, our approach recovers the leading order correction of the expression derived by Pedersen et al. [3].

3 Derivation of estimators in the presence measurement errors

In any real experiment, the recorded positions are subject to measurement errors, due to, e.g., motion blur or uncorrelated localization errors. Such random measurement errors can be modelled as an uncorrelated noise $\eta_\mu(t)$ (not necessarily Gaussian) acting on the positions $x_\mu(t)$ [3,7], meaning that the signal we actually observe is

$$y_\mu(t) = x_\mu(t) + \eta_\mu(t) \tag{S38}$$

where $\langle \eta_\mu(t)\eta_\nu(t')\rangle = \Lambda_{\mu\nu}\delta(t-t')$.

3.1 Force field

We will now again calculate the leading order contributions to the estimator of the projected accelerations with measurement error (w.m.e.):

$$\hat{A}^{(\text{w.m.e.})}_{\mu\alpha} = \langle \hat{a}^{(\text{w.m.e.})}_\mu c_\alpha(\overline{\mathbf{y}}, \hat{\mathbf{w}})\rangle \tag{S39}$$

Here, $\hat{\mathbf{a}}^{(\text{w.m.e.})}$ and $\hat{\mathbf{w}}$ are the empirical acceleration and velocity derived from the signal subject to measurement error $\mathbf{y}(t)$, respectively. Note that we are no longer conditioning on a single position-like coordinate, but rather the average quantity $\overline{\mathbf{y}}$, which is a linear combination of the three time-points entering the acceleration. This allows us to find a conditioning in terms of $\overline{\mathbf{y}}$ and $\hat{\mathbf{w}}$ such that the leading order terms due to the measurement errors cancel. We thus write

$$\overline{y}_\mu(\beta,\gamma) = \beta y_\mu(t+2\Delta t) + \gamma y_\mu(t) + (1-(\beta+\gamma))y_\mu(t+\Delta t) \tag{S40}$$

The velocity estimator including measurement noise is

$$\hat{w}_\mu = \hat{v}_\mu + \frac{\lambda\Delta\eta^{(2)}_\mu}{2\Delta t} + \frac{(1-\lambda)\Delta\eta^{(1)}_\mu}{\Delta t} := \hat{v}_\mu + \frac{f^{(v)}_\mu(\eta;\lambda)}{\Delta t} \tag{S41}$$

Similarly,

$$\hat{a}^{(\text{w.m.e.})}_\mu = \hat{a}_\mu + \frac{\Delta\eta^{(2)}_\mu - \Delta\eta^{(1)}_\mu}{\Delta t^2} := \hat{a}_\mu + \frac{f^{(a)}_\mu(\eta)}{\Delta t^2} \tag{S42}$$

We assume here that the measurement error η_μ is relatively small compared to the scale of variation of the fitting functions, such that we can expand the basis functions as

$$c_\alpha(\overline{\mathbf{y}}, \hat{\mathbf{w}}) = c_\alpha(\overline{\mathbf{x}}, \hat{\mathbf{v}}) + (\partial_{x_\nu} c_\alpha)\overline{\eta}_\nu(\beta,\gamma) + (\partial_{v_\nu} c_\alpha)\frac{f^{(v)}_\nu(\eta;\lambda)}{\Delta t} + \mathcal{O}(\eta^2) \tag{S43}$$

Combining Eqs. (S43) and (S42), the estimator of the acceleration projection thus reads

$$\hat{A}^{(\text{w.m.e.})}_{\mu\alpha} = \hat{A}_{\mu\alpha} + (\partial_{x_\nu} c_\alpha)\frac{\langle \overline{\eta}_\nu(\beta,\gamma) f^{(a)}_\mu(\eta)\rangle}{\Delta t^2} + (\partial_{v_\nu} c_\alpha)\frac{\langle f^{(v)}_\nu(\eta;\lambda) f^{(a)}_\mu(\eta)\rangle}{\Delta t^3} + \mathcal{O}(\eta^2) \tag{S44}$$

This shows that the leading order contribution to the estimator of the acceleration projection is of order Δt^{-3}, inducing a "dangerous" bias which diverges fast with $\Delta t \to 0$.

Indeed, the standard approach [2–5] is to take $\hat{A}^{(\text{w.m.e.})}_{\mu\alpha}$ with $\lambda = \beta = \gamma = 0$ as a proxy for the force projections, which results in

$$\hat{A}^{(\text{w.m.e.})}_{\mu\alpha} = \hat{A}_{\mu\alpha} - (\partial_{x_\nu} c_\alpha)\frac{2\Lambda_{\mu\nu}}{\Delta t^2} - (\partial_{v_\nu} c_\alpha)\frac{3\Lambda_{\mu\nu}}{\Delta t^3} + \mathcal{O}(\eta^2) \tag{S45}$$

Here we propose to make use of the free parameters λ, β and γ, to find a construction for our estimator such that the divergent cross-terms in Eq. (S44) cancel. Thus, we solve the following equations for $\{\lambda, \beta, \gamma\}$:

$$\langle \overline{\eta}_\nu(\beta,\gamma) f^{(a)}_\mu(\eta)\rangle = [(\beta+\gamma) - 2(1-(\beta+\gamma))]\Lambda_{\mu\nu} \tag{S46}$$

$$\langle f^{(v)}_\nu(\eta;\lambda) f^{(a)}_\mu(\eta)\rangle = [3\lambda - 3]\Lambda_{\mu\nu} \tag{S47}$$

These terms vanish for $\lambda = 1$ and $\beta + \gamma = 2/3$. There is thus a remaining freedom in the choice of β and γ. For simplicity, we choose the symmetric option $\beta = \gamma = 1/3$. We have thus determined optimal 'conditioning variables', *i.e.* the arguments $\overline{\mathbf{y}}$ and $\hat{\mathbf{w}}$ of the basis function $c_\alpha(\overline{\mathbf{y}}, \hat{\mathbf{w}})$, that are constructed in such a way that any measurement error-induced cross-terms cancel.

Thus, an unbiased estimator for the force projections in the presence of measurement errors is

$$\begin{aligned}
\hat{F}^{(\text{w.m.e.})}_{\mu\alpha} &= \langle \hat{a}^{(\text{w.m.e.})}_\mu c_\alpha(\overline{\mathbf{y}}(t), \hat{\mathbf{w}}(t))\rangle \\
&\quad - \frac{1}{2}\left\langle (\partial_{v_\nu} c_\alpha(\overline{\mathbf{y}}(t), \hat{\mathbf{w}}(t)))\widehat{\sigma^2}^{(\text{w.m.e.})}_{\mu\nu}(\overline{\mathbf{y}}(t), \hat{\mathbf{w}}(t))\right\rangle + \mathcal{O}(\Delta t, \eta^2) \qquad \text{(S48a)} \\
\overline{\mathbf{y}} &= \frac{1}{3}(\mathbf{y}(t-\Delta t) + \mathbf{y}(t) + \mathbf{y}(t+\Delta t)) \qquad \text{(S48b)} \\
\hat{\mathbf{w}} &= \frac{\mathbf{y}(t+\Delta t) - \mathbf{y}(t-\Delta t)}{2\Delta t} \qquad \text{(S48c)} \\
\hat{\mathbf{a}}^{(\text{w.m.e.})} &= \frac{\mathbf{y}(t+\Delta t) - 2\mathbf{y}(t) + \mathbf{y}(t-\Delta t)}{\Delta t^2} \qquad \text{(S48d)}
\end{aligned}$$

As before, to infer the force field, we have to first find an estimator for the noise term $\widehat{\sigma^2}^{(\text{w.m.e.})}_{\mu\nu}$ that is not biased due to measurement errors.

3.2 Noise term

To derive an unbiased estimator for the noise amplitude in the presence of measurement errors, we follow a very similar line of thought to the derivation of the measurement error-corrected estimator for the force field. Specifically, using the increments of the process, we derive an estimator constructed such that the bias-terms due to the measurement error $\eta(t)$ vanish. However, in contrast to the force estimator, we now consider an estimator constructed from four points around t, $\{t-\Delta t, t, t+2\Delta t, t+3\Delta t\}$. This gives us three increments, rather than two as before, to construct our estimator. This additional freedom is required to construct an estimator that is not spoilt by measurement errors.

As before, we first start by constructing increments of the form

$$\Delta y_\mu^{(n)} = y_\mu(t+n\Delta t) - y_\mu(t) \tag{S49}$$

but now with $n = \{1,2,3\}$. We will later transform our results to a notation centered around t. Similar to Eq. (S21), we expand these increments, now including the measurement error

$$\begin{aligned}\Delta y_\mu^{(n)} &= \Delta x_\mu^{(n)} + \Delta\eta_\mu^{(n)} \\ &= v_\mu I_0^{(n)} + \sigma_{\mu\nu} I_{0\nu}^{(n)} + \Delta\eta_\mu^{(n)} + F_\mu I_{00}^{(n)} + (\partial_{v_\rho}\sigma_{\mu\nu})\sigma_{\rho\tau} I_{0\nu\tau}^{(n)} + \mathcal{O}(\Delta t^{5/2})\end{aligned} \tag{S50}$$

Since we are aiming to infer the term $\sigma_{\mu\nu} I_{0\nu}^{(n)}$, which has zero average, we need to consider products of the increments (similar to the noise-free version (S31), where $\widehat{\sigma^2} \sim \hat{a}^2$).

$$\Delta_{\mu\nu}^{(n,m)} := \Delta y_\mu^{(n)} \Delta y_\nu^{(m)} \tag{S51}$$

We thus aim to construct an estimator of the form

$$\Delta t^{-3} \left\langle c_\alpha(\tilde{\mathbf{y}}, \check{\mathbf{w}}) \sum_{1\le m\le n\le 3} k_{mn} \Delta_{\mu\nu}^{(n,m)} \right\rangle \stackrel{!}{=} \sigma_{\mu\nu\alpha}^2 + \mathcal{O}(\Delta t, \eta^2) \tag{S52}$$

We therefore need to find the coefficients k_{mn} for the linear combination of increment products and conditioning coordinates $\tilde{\mathbf{y}}$ and $\check{\mathbf{w}}$ such that all dynamical and measurement error cross-terms except for σ^2 cancel out to first order.

We start by expanding the increment products

$$\begin{aligned}\Delta_{\mu\nu}^{(n,m)} &= v_\mu v_\nu (nm\Delta t^2) + \sigma_{\mu\rho}\sigma_{\nu\tau} I_{0\rho}^{(n)} I_{0\tau}^{(m)} + \Delta\eta_\mu^{(n)}\Delta\eta_\nu^{(m)} + (v_\mu F_\nu mn^2 + v_\nu F_\mu m^2 n)\Delta t^3 \\ &\quad + (nv_\mu\sigma_{\nu\rho} I_{0\rho}^{(m)} + mv_\nu\sigma_{\mu\rho} I_{0\rho}^{(n)})\Delta t + (mv_\mu\Delta\eta_\nu^{(n)} + nv_\nu\Delta\eta_\mu^{(m)})\Delta t + \mathcal{O}(\Delta t^{7/2})\end{aligned} \tag{S53}$$

Note that the last two terms in this expansion are zero on average, so one might think that we do not have to include them in the derivation of k_{mn}. This is correct in the case of constant noise. However, in the case of multiplicative noise, these terms correlate with terms in the expansion of the basis function $c_\alpha(\tilde{\mathbf{y}}, \check{\mathbf{w}})$, so we have to consider them.

Deriving the coefficients k_{mn} is essentially a linear algebra problem, so we define the vectors

$$\mathbf{d}_{\mu\nu}^{\mathrm{nm}} = \left(\Delta_{\mu\nu}^{(1,1)}, \Delta_{\mu\nu}^{(2,2)}, \Delta_{\mu\nu}^{(3,3)}, \Delta_{\mu\nu}^{(1,3)}, \Delta_{\mu\nu}^{(2,3)}, \Delta_{\mu\nu}^{(1,2)}\right)^T \tag{S54}$$

$$\mathbf{k}^{\mathrm{nm}} = (k_{11}, k_{22}, k_{33}, k_{13}, k_{23}, k_{12})^T \tag{S55}$$

$$\mathbf{t}_{\mu\nu} = \left(v_\mu v_\nu \Delta t^2, \sigma_{\mu\nu}^2 \Delta t^3, \Lambda\delta_{\mu\nu}, F_\mu v_\nu \Delta t^3, v_\mu \sigma_{\nu\rho} I_{0\rho}^{(1)}, v_\mu \sigma_{\nu\rho} I_{0\rho}^{(2)}, v_\mu \sigma_{\nu\rho} I_{0\rho}^{(3)}\right)^T \tag{S56}$$

Note, that in the definition of $\mathbf{t}_{\mu\nu}$ we have temporarily discarded the symmetry under exchange of μ, ν for simplicity. We will later symmetrize our results to regain this symmetry. Furthermore, we have discarded the last term in Eq. (S53) in our definition of $\mathbf{t}_{\mu\nu}$. We will ignore this term in our derivation of k_{mn} as we can take care of it through our choice of conditioning coordinates $\tilde{\mathbf{y}}$ and $\check{\mathbf{w}}$.

With these definitions, we explicitly evaluate the increment products:

$$\mathbf{d}_{\mu\nu}^{\mathrm{nm}} = \underline{\underline{R}}^T \cdot \mathbf{t}_{\mu\nu} + \mathcal{O}(\Delta t^{7/2}) \tag{S57}$$

where

$$\underline{\underline{R}} = \begin{pmatrix} 1 & 4 & 9 & 3 & 6 & 2 \\ 1/3 & 8/3 & 9 & 4/3 & 14/3 & 5/6 \\ 2 & 2 & 2 & 1 & 1 & 1 \\ 2 & 16 & 54 & 12 & 30 & 6 \\ 2 & 0 & 0 & 3 & 0 & 2 \\ 0 & 4 & 0 & 0 & 3 & 1 \\ 0 & 0 & 6 & 1 & 2 & 0 \end{pmatrix} \tag{S58}$$

Thus, a general estimator for the variable V is given by solving the equation

$$\hat{V}_{\mu\nu} = \mathbf{k}^{\mathrm{nm}} \cdot \mathbf{d}_{\mu\nu}^{\mathrm{nm}} \stackrel{!}{=} \boldsymbol{\ell}_V \cdot \mathbf{t}_{\mu\nu} \tag{S59}$$

for $\mathbf{k}^{\mathrm{nm}}$. In our case the two quantities of interest are σ^2 and Λ, as we may also wish to infer the amplitude of the measurement error from the data. The constraint vectors $\boldsymbol{\ell}_V$ for these quantities are given by

$$\boldsymbol{\ell}_{\sigma^2} = (0, 1, 0, 0, 0, 0)^T \tag{S60}$$

$$\boldsymbol{\ell}_{\Lambda} = (0, 0, 1, 0, 0, 0)^T \tag{S61}$$

So far, we have derived everything in "(nm)-space", for increments as defined in Eq. (S49), which has the key advantage that they are easy to expand. However, for the final form of our estimators, we choose a more natural definition of the increments,

$$\begin{aligned} \Delta y_\mu^{(-)} &= y_\mu(t + \Delta t) - y_\mu(t) \\ \Delta y_\mu^{(0)} &= y_\mu(t + 2\Delta t) - y_\mu(t + \Delta t) \\ \Delta y_\mu^{(+)} &= y_\mu(t + 3\Delta t) - y_\mu(t + 2\Delta t) \end{aligned} \tag{S62}$$

For this "$(+-)$-space", we define, similarly to before,

$$\mathbf{d}_{\mu\nu}^{+-} = \left(\Delta_{\mu\nu}^{(0,0)}, \Delta_{\mu\nu}^{(-,-)}, \Delta_{\mu\nu}^{(+,+)}, \Delta_{\mu\nu}^{(+,-)}, \Delta_{\mu\nu}^{(0,+)}, \Delta_{\mu\nu}^{(0,-)}\right)^T \tag{S63}$$

$$\mathbf{k}^{+-} = (k_{00}, k_{--}, k_{++}, k_{+-}, k_{0+}, k_{0-})^T \tag{S64}$$

We can transform between the two spaces using

$$\mathbf{d}_{\mu\nu}^{\mathrm{nm}} = \underline{\underline{M}}\mathbf{d}_{\mu\nu}^{+-} \tag{S65}$$

with the transformation matrix

$$\underline{\underline{M}} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 2 \\ 1 & 1 & 1 & 2 & 2 & 2 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 & 2 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{S66}$$

Thus, we need to solve the transformed equation

$$\underline{\underline{Q}}\mathbf{k}^{+-} = \boldsymbol{\ell}_V \tag{S67}$$

where $\underline{\underline{Q}} = \underline{\underline{R}}(\underline{\underline{M}}^T)^{-1}$. Finally, we add two additional constraints to the matrix $\underline{\underline{Q}}$ which ensure that the final estimator is symmetric in the increments,

$$\underline{\underline{Q_{\mathrm{sym}}}}\mathbf{k}_{\mathrm{sym}}^{+-} = \boldsymbol{\ell}_V^{\mathrm{sym}} \tag{S68}$$

We can now solve for the coefficients:

$$\mathbf{k}_{\mathrm{sym}}^{+-} = \underline{\underline{Q_{\mathrm{sym}}}}^{\dagger}\boldsymbol{\ell}_V^{\mathrm{sym}} \tag{S69}$$

where $\underline{\underline{Q_{\mathrm{sym}}}}^{\dagger}$ is the Moore-Penrose pseudoinverse of the non-square matrix $\underline{\underline{Q_{\mathrm{sym}}}}$. This yields

$$\mathbf{k}_{\mathrm{sym}}^{+-}(\sigma^2) = \frac{6}{11}(-1, 1, 1, -6, 1, 1)^T \tag{S70}$$

$$\mathbf{k}_{\mathrm{sym}}^{+-}(\Lambda) = \frac{1}{44}(10, 1, 1, 8, -10, -10)^T \tag{S71}$$

With this solution for the coefficients, the estimator for σ^2 is now operational for the case of constant noise. The estimator for Λ is valid equally in the case of multiplicative noise.

As we noted before, in the case of multiplicative noise, we need to adjust the conditioning variables $\tilde{\mathbf{y}}$ and $\check{\mathbf{w}}$ in order to avoid divergent biases due to the last term in Eq. (S53),

similar to the case of the force field. As our estimator is a four-point construct, we also construct the conditioning variables from four points:

$$\tilde{y}_\mu = \sum_{n=0}^{3} a_n y_\mu(t + n\Delta t) = \tilde{x}_\mu + \tilde{\eta}_\mu(\{a_n\}) \tag{S72}$$

$$\check{w}_\mu = \frac{1}{\Delta t}\left[b_1 \Delta y_\mu^{(-)} + b_2 \Delta y_\mu^{(0)} + b_3 \Delta y_\mu^{(+)}\right] = \check{v}_\mu + \frac{g_\mu^{(v)}(\eta, \{b_n\})}{\Delta t} \tag{S73}$$

where $\sum_{n=0}^{3} a_n = \sum_{n=1}^{3} b_n = 1$. Similarly to before (Eq. (S43)), we expand the basis functions

$$c_\alpha(\tilde{\mathbf{y}}, \check{\mathbf{w}}) = c_\alpha(\tilde{\mathbf{x}}, \check{\mathbf{v}}) + (\partial_{x_\nu} c_\alpha)\tilde{\eta}_\nu(\{a_n\}) + (\partial_{v_\nu} c_\alpha)\frac{g_\nu^{(v)}(\eta, \{b_n\})}{\Delta t} + \mathcal{O}(\eta^2) \tag{S74}$$

The remaining bias in our estimator (Eq. (S52)) is due to the last term in Eq. (S53),

$$q_{\mu\nu}^{(m,n)} = (m v_\mu \Delta\eta_\nu^{(n)} + n v_\nu \Delta\eta_\mu^{(m)})\Delta t \tag{S75}$$

We define

$$\mathbf{q}_{\mu\nu}^{\mathrm{nm}} = \left(q_{\mu\nu}^{(1,1)}, q_{\mu\nu}^{(2,2)}, q_{\mu\nu}^{(3,3)}, q_{\mu\nu}^{(1,3)}, q_{\mu\nu}^{(2,3)}, q_{\mu\nu}^{(1,2)}\right)^T \tag{S76}$$

and can thus write

$$\begin{aligned}\Delta t^{-3}\left\langle c_\alpha(\tilde{\mathbf{y}}, \check{\mathbf{w}})\mathbf{k}_{\mathrm{sym}}^{\mathrm{mn}} \cdot \mathbf{d}_{\mu\nu}^{\mathrm{nm}}\right\rangle &= \langle \sigma_{\mu\nu}^2 c_\alpha(\tilde{\mathbf{x}}, \check{\mathbf{v}})\rangle \\ &+ \Delta t^{-3}\left\langle \mathbf{k}_{\mathrm{sym}}^{\mathrm{mn}} \cdot \mathbf{q}_{\mu\nu}^{\mathrm{nm}}\left((\partial_{x_\rho} c_\alpha)\tilde{\eta}_\rho(\{a_n\}) + (\partial_{v_\rho} c_\alpha)\frac{g_\rho^{(v)}(\eta, \{b_n\})}{\Delta t}\right)\right\rangle + \mathcal{O}(\Delta t, \eta^2)\end{aligned} \tag{S77}$$

This shows that the bias terms are of order Δt^{-3} and Δt^{-4}, and thus diverge in the limit $\Delta t \to 0$. We now need to find coefficients $\{a_n\}$ and $\{b_n\}$ such that

$$\left\langle \mathbf{k}_{\mathrm{sym}}^{\mathrm{mn}} \cdot \mathbf{q}_{\mu\nu}^{\mathrm{nm}} \tilde{\eta}_\rho(\{a_n\})\right\rangle = 0 \tag{S78}$$

$$\left\langle \mathbf{k}_{\mathrm{sym}}^{\mathrm{mn}} \cdot \mathbf{q}_{\mu\nu}^{\mathrm{nm}} \frac{g_\rho^{(v)}(\eta, \{b_n\})}{\Delta t}\right\rangle = 0 \tag{S79}$$

We start by explicitly evaluating $\mathbf{q}_{\mu\nu}^{\mathrm{nm}}$:

$$\mathbf{q}_{\mu\nu}^{\mathrm{nm}} = \underline{\underline{E}} \cdot \mathbf{h}_\mu \cdot (v_\nu \Delta t) \tag{S80}$$

where

$$\underline{\underline{E}} = \begin{pmatrix} -2 & 2 & 0 & 0 \\ -4 & 0 & 4 & 0 \\ -6 & 0 & 0 & 6 \\ -5 & 0 & 3 & 2 \\ -4 & 3 & 0 & 1 \\ -3 & 2 & 1 & 0 \end{pmatrix} \tag{S81}$$

and

$$\mathbf{h}_\mu = \left(\eta_\mu(t), \eta_\mu(t+\Delta t), \eta_\mu(t+2\Delta t), \eta_\mu(t+3\Delta t)\right)^T. \tag{S82}$$

We first focus on the conditioning of the configurational (position-like) coordinate, *i.e.* solving Eq. (S78). Defining $\mathbf{a} = (a_0, a_1, a_2, a_3)^T$, Eq. (S78) becomes

$$\left\langle \left(\mathbf{k}^{\mathrm{mn}}_{\mathrm{sym}} \cdot \mathbf{q}^{\mathrm{nm}}_{\mu\nu}\right) \left(\mathbf{a} \cdot \mathbf{h}_\rho\right) \right\rangle = 0 \tag{S83}$$

Evaluating

$$\mathbf{k}^{\mathrm{mn}}_{\mathrm{sym}} \cdot \mathbf{q}^{\mathrm{nm}}_{\mu\nu} = \underline{\underline{E}}^T \cdot \underline{\underline{M}} \cdot \mathbf{k}^{+-}_{\mathrm{sym}}(\sigma^2) \tag{S84}$$

$$= \frac{1}{11}(-30, 36, 42, -48)^T \tag{S85}$$

shows that Eq. (S78) is solved by

$$\mathbf{a} = \frac{1}{4}(1,1,1,1)^T. \tag{S86}$$

Next, we find the conditioning of the velocity coordinate, *i.e.* solving Eq. (S79). Defining $\mathbf{b} = (b_1, b_2, b_3)^T$, Eq. (S79) becomes

$$\left\langle \left(\mathbf{k}^{\mathrm{mn}}_{\mathrm{sym}} \cdot \mathbf{q}^{\mathrm{nm}}_{\mu\nu}\right) \frac{\mathbf{b} \cdot \underline{\underline{F}} \cdot \mathbf{h}_\rho}{\Delta t} \right\rangle = 0 \tag{S87}$$

where

$$\underline{\underline{F}} = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix} \tag{S88}$$

are the coefficients of the measurement error $\mathbf{h}$ in the velocity estimator $\check{\mathbf{w}}$. Evaluating

$$\underline{\underline{F}} \cdot \mathbf{k}^{\mathrm{mn}}_{\mathrm{sym}} \cdot \mathbf{q}^{\mathrm{nm}}_{\mu\nu} = \underline{\underline{F}} \cdot \underline{\underline{E}}^T \cdot \underline{\underline{M}} \cdot \mathbf{k}^{+-}_{\mathrm{sym}}(\sigma^2) \tag{S89}$$

$$= \frac{6}{11}(11, 1, -15)^T \tag{S90}$$

shows that Eq. (S79) is solved by

$$\mathbf{b} = \frac{1}{6}(1,4,1)^T. \tag{S91}$$

Summarizing, an unbiased estimator for the projection coefficients of the multiplicative noise amplitude in the presence of measurement error is given by

$$\widehat{\sigma^2}^{(\mathrm{w.m.e.})}_{\mu\nu\alpha} = \Delta t^{-3} \left\langle c_\alpha(\tilde{\mathbf{y}}, \check{\mathbf{w}}) \mathbf{k}^{+-}_{\mathrm{sym}} \cdot \mathbf{d}^{+-}_{\mu\nu} \right\rangle \tag{S92a}$$

$$\mathbf{k}^{+-}_{\mathrm{sym}} = \frac{6}{11}(-1,1,1,-6,1,1)^T \tag{S92b}$$

$$\tilde{\mathbf{y}} = \frac{1}{4}(\mathbf{y}(t-\Delta t) + \mathbf{y}(t) + \mathbf{y}(t+\Delta t) + \mathbf{y}(t+2\Delta t)) \tag{S92c}$$

$$\check{\mathbf{w}} = \frac{\Delta\mathbf{y}^{(-)} + 4\Delta\mathbf{y}^{(0)} + \Delta\mathbf{y}^{(+)}}{6\Delta t} \tag{S92d}$$

with $\Delta\mathbf{y}^{(+/0/-)}$ as defined by Eq. (S62) and

$$\mathbf{d}_{\mu\nu}^{+-} = \left(\Delta y_\mu^{(0)}\Delta y_\nu^{(0)}, \Delta y_\mu^{(-)}\Delta y_\nu^{(-)}, \Delta y_\mu^{(+)}\Delta y_\nu^{(+)}, \Delta y_\mu^{(+)}\Delta y_\nu^{(-)}, \Delta y_\mu^{(0)}\Delta y_\nu^{(+)}, \Delta y_\mu^{(0)}\Delta y_\nu^{(-)}\right)^T \tag{S93}$$

3.3 Scaling of the inference error with the measurement error amplitude

To determine the critical measurement error amplitude at which the estimators fail, we investigate the scaling of the error curves with the observation time interval Δt for the damped harmonic oscillator. We find that the error curves of the estimator without noise correction (section 2.2) collapse with $\sigma\Delta t^{3/2}$, while the curves of the estimator with noise correction (section 3.1) collapse with Δt.

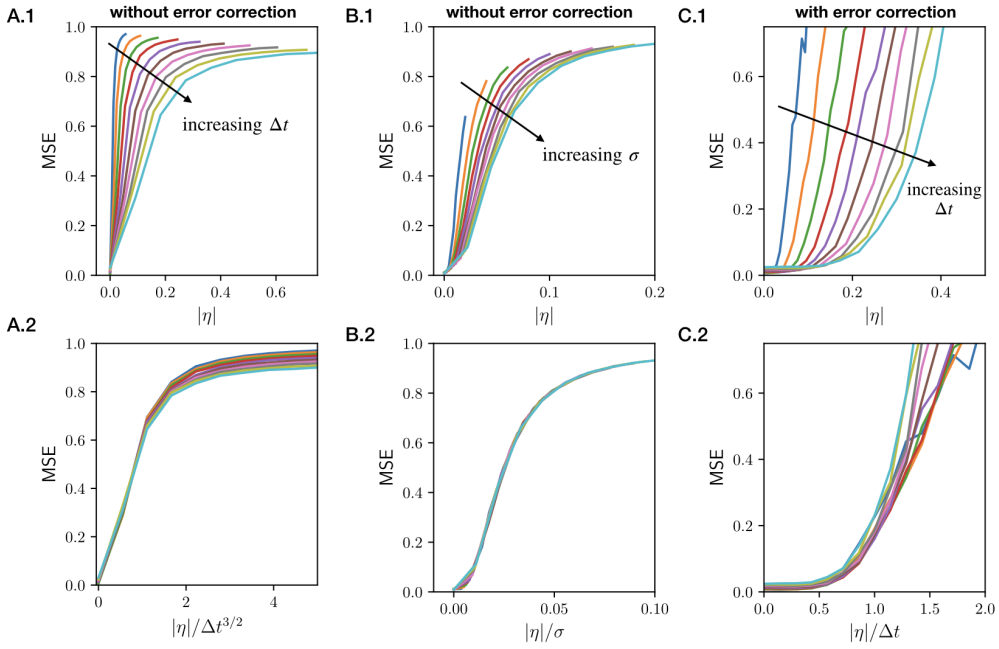


Figure S1: **Error scaling of the force estimator in the presence of measurement error. A.** *Top:* mean-square-error for the estimator without noise correction (section 2.2) as a function of the measurement error amplitude $|\eta|$ for different values of Δt. *Bottom:* Data collapse by dividing by $\Delta t^{3/2}$. **B.** *Top:* same plot as in B, but for different values of σ. *Bottom:* Data collapse by dividing by σ. **C.** *Top:* mean-square-error along the trajectory for the estimator with noise correction (section 3.1) as a function of the measurement error amplitude $|\eta|$ for different values of Δt. *Bottom:* Data collapse by dividing by Δt.

4 Choosing the basis size n_b

We perform inference by projecting the dynamics onto a finite set of n_b basis functions $\{b_\alpha(\mathbf{x}, \mathbf{v})\}$ where the index α runs over all basis functions in the set, $1 \leqslant \alpha \leqslant n_b$. Thus, we need to choose a value for n_b, and this will clearly influence the accuracy of the inference, leading to the question: is there an optimal choice $n_b^{(\mathrm{opt})}$? A criterion to choose $n_b^{(\mathrm{opt})}$ was proposed for overdamped stochastic processes in ref. [1], which is based on the empirical estimate of the information content in the observed trajectory. Here, we show that this criterion similarly applies to underdamped stochastic systems.

The larger n_b, the more accurately it can capture the features of the force field. Thus, the information $\hat{I}_b = \frac{\tau}{2}\hat{\sigma}^{-2}_{\mu\nu}\hat{F}_{\mu\alpha}\hat{F}_{\nu\alpha}$ captured by the force field representation [1] increases with n_b. However, for a finite trajectory, the error in the inferred force field will also increase with n_b. Thus, we expect a trade-off between the inference error and the completeness of the force field projection. To choose n_b^{opt}, we therefore maximize the information $\hat{I}_b$ that can be statistically resolved, by determining the basis size which maximizes $\hat{I}_b - \delta\hat{I}_b$, where $\delta\hat{I}_b \approx \sqrt{2\hat{I}_b + N_b^2/4}$ is the typical error in the inferred information and $N_b = d \times n_b$ is the number of parameters to infer [1].

In Fig. S2, we plot the inference error $\delta\hat{F}^2/\hat{F}^2$ as a function of the number of parameters N_b for the 1D Van der Pol oscillator projected on a basis consisting of Fourier components in x and polynomials in v. For all trajectory lengths, we see the expected behaviour: at small N_b, the error first decreases with increasing N_b, since it is dominated by underfitting. Beyond the optimum, the error increases with N_b, due to the increasing inference error. Clearly, the optimal basis size $n_b^{(\mathrm{opt})}$ increases with the length of the trajectory, as more information on the features of the force field becomes available. We find that the overfitting criterion to maximize $\hat{I}_b - \delta\hat{I}_b$ yields an accurate prediction of the optimal basis size (starred symbols in Fig. S2).

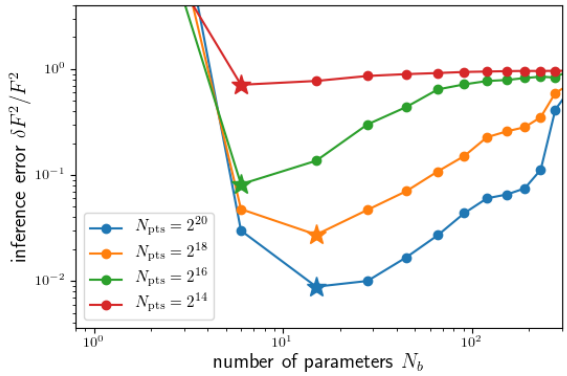


Figure S2: **Quantification of the inference performance as a function of basis size.** Here we study the 1D Van der Pol oscillator $\dot{v} = \mu(1 - x^2)v - x + \sigma\xi(t)$ as an example. The inference error $\delta\hat{F}^2/\hat{F}^2$ is plotted as a function of number of parameters N_b used in the projection basis, consists of Fourier components in x and polynomials in v. Each curve corresponds to the result obtained from a single trajectory with the number of time frames indicated in the legend. Starred symbols indicate the predicted optimal basis size $n_b^{(\rm opt)}$ determined by maximizing $\hat{I}_b - \delta\hat{I}_b$.

5 Inference from experimental single cell trajectories

Here, we discuss the inference from experimental single cell trajectories shown in Fig. 2 of the main text in more detail. Specifically, we show that the experimental trajectories contain enough information to perform Underdamped Langevin Inference, and that the inferred models can be inferred self-consistently. Importantly, here the term $F_\mu(\mathbf{x}, \mathbf{v})$ corresponds to the underlying deterministic dynamics of the system, and not to a physical force. We therefore call it the "deterministic term" of the dynamics. Details on cell culture, experimental protocols and tracking procedures can be found in ref. [5].

5.1 Information content of experimental single cell trajectories

As discussed in the main text, the observed trajectories are limited in length due to the finite life-time of a single cell, up until the point where it divides. The expected mean-squared-error in the inferred flow field projected onto a basis b is given by

$$\delta\hat{F}^2/\hat{F}^2 \sim N_b/2\hat{I}_b, \tag{S94}$$

where N_b is the number of degrees of freedom in the basis b, *i.e.* the number of fit parameters. $\hat{I}_b$ is the empirical estimate of the information content of the trajectory of length τ, given by

$$\hat{I}_b = \frac{\tau}{2}\hat{\sigma}^{-2}_{\mu\nu}\hat{F}_{\mu\alpha}\hat{F}_{\nu\alpha}, \tag{S95}$$

measured in natural information units (1 nat $= 1/\log 2$ bits). We estimate this information by projecting onto a third-order polynomial basis, and find that the average information per trajectory is 94.2 nats (Fig. S3). To perform accurate inference, we need $\hat{I}_b \gg N_b$. In previous work [5, 8], we inferred models averaged over large numbers of cell trajectories using a basis of 30×30 coarse-grained bins, *i.e.* $N_b = 900$. Thus, single-cell inference was not possible with this approach. In contrast, here we use the partial information to guide a principled selection of basis functions, which shows that most of the information is captured by a symmetrised third-order polynomial basis $\{x, v, x^3, x^2v, xv^2, v^3\}$. Thus, we infer $N_b = 6$ parameters and the criterion $\hat{I}_b \gg N_b$ is fulfilled.

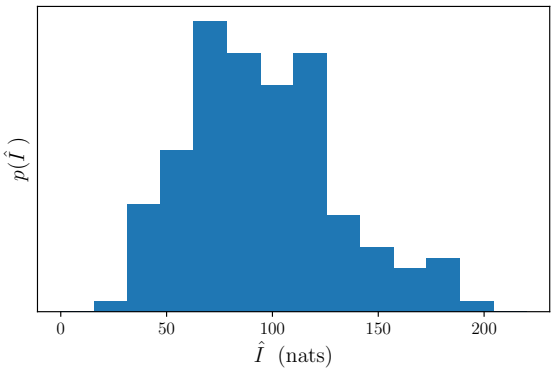


Figure S3: **Information content of single cell trajectories.** Histogram of the information content $\hat{I}_b$ of $N = 149$ single cell trajectories, obtained by projecting onto a third-order polynomial basis. The information is measured in natural information units (1 nat $= 1/\log 2$ bits). The average information per trajectory is 94.2 nats.

5.2 Self-consistency test of the single-cell inference

To test whether the inferred single-cell models are self-consistent, we simulate trajectories based on the inferred dynamics (Fig. S4). These trajectories perform stochastic transitions on a similar time-scale to the experimental trajectories and exhibit similar oscillation loops in the xv-phase space (Fig. S4E,F). To test model stability, we simulate trajectories of the same length as the experimental ones and sample at the same time interval as in experiment ($\Delta t = 10$ min). From these trajectories, we then infer a bootstrapped flow field, which exhibits similar qualitative features as the original flow field inferred from experiments (Fig. S4G). To quantify this, we directly compare the values of the bootstrapped $F(x, v)$ relative to the experimentally inferred $F(x, v)$ along the experimental trajectory (Fig. S4H), which shows strong correlation with a typical mean-squared-error of order 0.3. Thus, ULI with a symmetrised third-order polynomial basis provides robust, self-consistent models for single-cell trajectories.

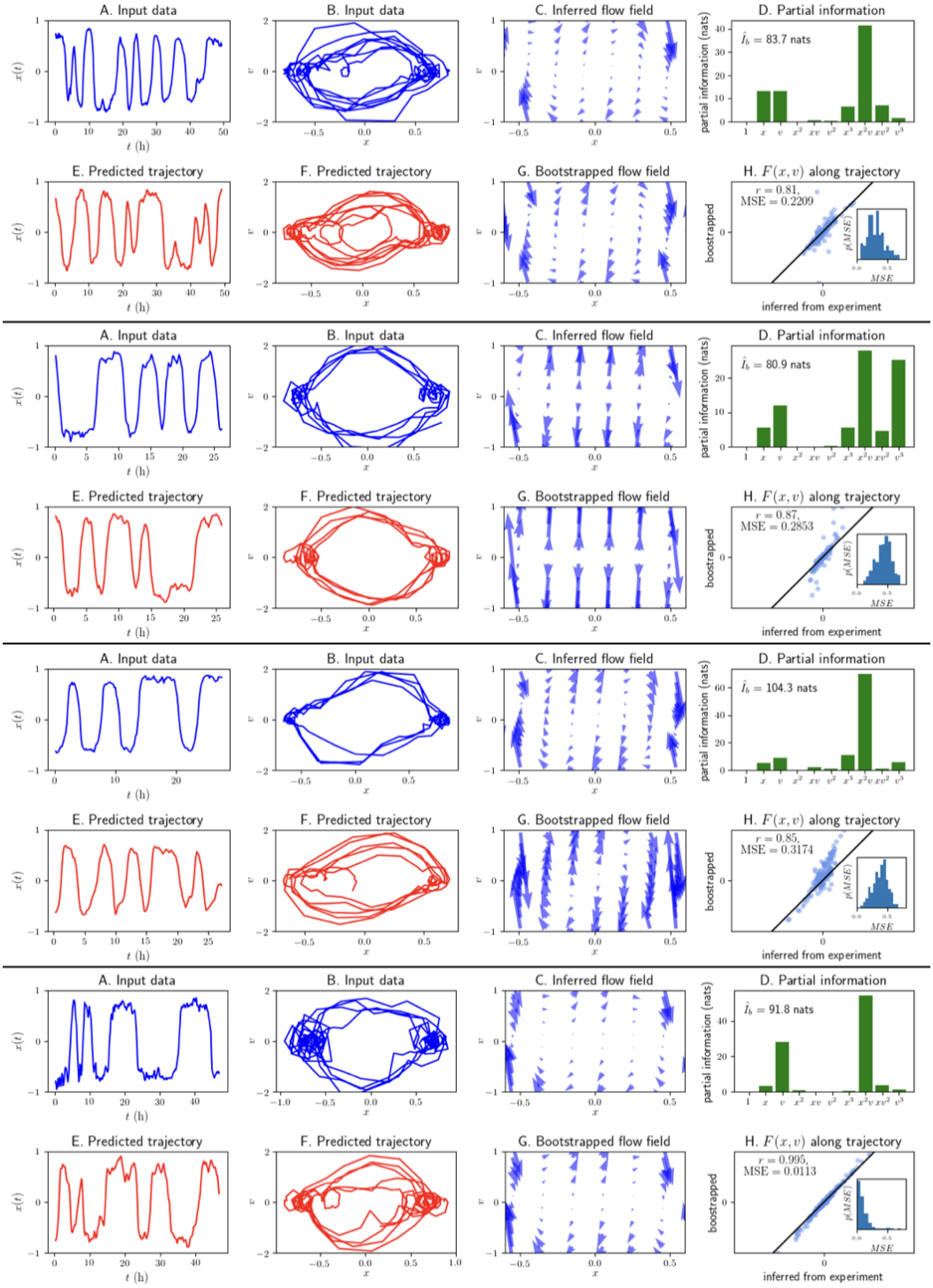

A. Input data
B. Input data
C. Inferred flow field
D. Partial information
$\hat{I}_b = 83.7$ nats
E. Predicted trajectory
F. Predicted trajectory
G. Bootstrapped flow field
H. $F(x, v)$ along trajectory
$r = 0.81$,
MSE $= 0.2209$
$p(MSE)$
MSE
boostrapped
inferred from experiment
partial information (nats)
t (h)
$x(t)$
v
x
A. Input data
B. Input data
C. Inferred flow field
D. Partial information
$\hat{I}_b = 80.9$ nats
E. Predicted trajectory
F. Predicted trajectory
G. Bootstrapped flow field
H. $F(x, v)$ along trajectory
$r = 0.87$,
MSE $= 0.2853$
$p(MSE)$
MSE
boostrapped
inferred from experiment
partial information (nats)
t (h)
$x(t)$
v
x
A. Input data
B. Input data
C. Inferred flow field
D. Partial information
$\hat{I}_b = 104.3$ nats
E. Predicted trajectory
F. Predicted trajectory
G. Bootstrapped flow field
H. $F(x, v)$ along trajectory
$r = 0.85$,
MSE $= 0.3174$
$p(MSE)$
MSE
boostrapped
inferred from experiment
partial information (nats)
t (h)
$x(t)$
v
x
A. Input data
B. Input data
C. Inferred flow field
D. Partial information
$\hat{I}_b = 91.8$ nats
E. Predicted trajectory
F. Predicted trajectory
G. Bootstrapped flow field
H. $F(x, v)$ along trajectory
$r = 0.995$,
MSE $= 0.0113$
$p(MSE)$
MSE
boostrapped
inferred from experiment
partial information (nats)
t (h)
$x(t)$
v
x

Figure S4: **Inferring single-cell models from two-state migration trajectories. A.** Experimentally recorded trajectory of the cell nucleus position, sampled at a time-interval $\Delta t = 10$ min. **B.** xv-plot of the trajectory shown in A. **C.** Flow field inferred from the trajectory in A using ULI with a symmetrised third-order polynomial basis, $\{x, v, x^3, x^2v, xv^2, v^3\}$. **D.** Partial information of the trajectory shown in A, projected onto a third-order polynomial basis. The total estimated information $\hat{I}_b$ of the trajectory is given. **E.** Trajectory simulated using the inferred model, consisting of the deterministic flow field in C and the inferred constant noise amplitude. The process is simulated at a small time-interval and subsequently sampled at the experimental time-interval $\Delta t = 10$ min. **F.** xv-plot of the simulated trajectory shown in E. **G.** Bootstrapped flow field inferred from the simulated trajectory in E using ULI with a symmetrised third-order polynomial basis. **H.** Scatter plot of the deterministic term evaluated at the points visited by the experimental trajectory, comparing the flow field inferred from experiment against the bootstrapped result. The mean-squared-error (MSE) and the Pearson r-coefficient are given. *Inset:* histogram of the mean-squared-error of $N = 300$ bootstrap realizations. The four subfigures correspond to four individual cell trajectories. The top subfigure corresponds to the trajectory shown in Fig. 2 of the main text.

6 Model details and simulation parameters for numerical results

To benchmark ULI, we apply it to several canonical examples of underdamped stochastic processes (Fig. 1-3). To simulate these processes, we employ a simple discretization scheme

$$\mathbf{x}(t + \mathrm{d}t) = \mathbf{x}(t) + \mathbf{v}(t)\mathrm{d}t \tag{S96}$$

$$\mathbf{v}(t + \mathrm{d}t) = \mathbf{v}(t) + \mathbf{F}(\mathbf{x}(t), \mathbf{v}(t))\mathrm{d}t + \sqrt{\mathrm{d}t}\, \underline{\underline{\sigma}}(\mathbf{x}(t), \mathbf{v}(t)) \cdot \boldsymbol{\zeta}(t) \tag{S97}$$

where $\boldsymbol{\zeta}$ is a vector of independent Gaussian random numbers with zero mean and unit variance. We simulate this equation with a small time interval $\mathrm{d}t$ to ensure numerical stability. To generate a realistic experimental position trajectory, we sample the simulated trajectory with a larger interval Δt and add an uncorrelated measurement error to the positions. We use $\mathrm{d}t = \Delta t/20$ throughout. Thus, ULI only has access to the trajectory

$$\{\mathbf{y}(0), \mathbf{y}(\Delta t), \mathbf{y}(2\Delta t), ..., \mathbf{y}(\tau - \Delta t), \mathbf{y}(\tau)\} \quad \text{where} \quad \mathbf{y}(t) = \mathbf{x}(t) + \boldsymbol{\eta}(t) \tag{S98}$$

and $\boldsymbol{\eta}$ is a vector of independent Gaussian random numbers with zero mean and unit variance, such that

$$\eta_\mu(t)\eta_\nu(t') = \Lambda\delta_{\mu\nu}\delta(t - t') \tag{S99}$$

and we define $|\eta| = \sqrt{\Lambda}$. The total duration of a trajectory with N_{steps} observation points given by $\tau = N_{\text{steps}}\Delta t$.

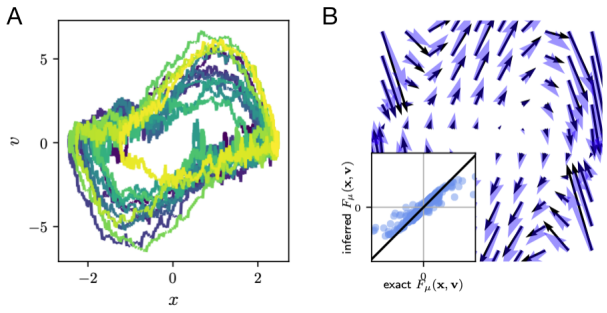


Figure S5: **Inferring Van der Pol dynamics with a Fourier basis. A.** Same trajectory as in Fig. 2A in the main text. **B.** ULI applied to the trajectory in A with basis functions $b = \{\sin(a_x x), \sin(a_v v), \sin(a_x x)\cos(a_v v), \cos(a_x x)\sin(a_v v)\}$. In general, a reasonable choice of the non-linear parameters a_x, a_v is $a_x = 2\pi/L_x, a_v = 2\pi/L_v$, where L_x, L_v are the widths of the sampled phase space in the x and v directions, respectively. From the trajectory in A, we see that $L_x = 6, L_v = 12$ are reasonable choices. *Inset:* inferred components of the force along the trajectory *versus* the exact values.

6.1 Damped harmonic oscillator (Fig. 1)

We simulate the 1D stochastic damped harmonic oscillator,

$$\dot{v} = -\gamma v - kx + \sigma\xi \tag{S100}$$

We use $\gamma = k = \sigma = 1$ in all panels. Furthermore, we use
Fig. 1A-C: $N_{\text{steps}} = 10^3, \Delta t = 0.1, |\eta| = 0$
Fig. 1F-H: $N_{\text{steps}} = 10^3, \Delta t = 0.1, |\eta| = 0.02$
Fig. 1D: $\Delta t = 0.1, |\eta| = 0$
Fig. 1E: $\tau = 10^2, |\eta| = 0$
Fig. 1J: $\Delta t = 0.1, |\eta| = 0.02$
Fig. 1K: $\Delta t = 0.1, N_{\text{steps}} = 10^4$

6.2 Van der Pol oscillator (Fig. 2)

For the Van der Pol oscillator, we use $\kappa = 2, \sigma = 1$ throughout.
In Fig. 2A,B, we simulate the 1D Van der Pol oscillator

$$\dot{v} = \kappa(1 - x^2)v - x + \sigma\xi \tag{S101}$$

with $\Delta t = 0.01, N_{\text{steps}} = 10^4, |\eta| = 0.002$. As shown in Supplementary Fig. S5, we recover the dynamics similarly well using a Fourier basis rather than a polynomial basis in the inference.

In Fig. 2C, we simulate the d-dimensional Van der Pol oscillator $F_\mu(\mathbf{x}, \mathbf{v}) = \kappa_\mu(1 - x_\mu^2)v_\mu - x_\mu$ (no summation over μ, $1 \leqslant \mu \leqslant d$) with $d = 1...6$. Here, we use the same parameters as for the 1D Van der Pol oscillator, and take $\kappa_\mu = 2 \; \forall \; \mu$.

In Fig. 2 G-J, we simulate the 1D Van der Pol oscillator with multiplicative noise

$$\dot{v} = \mu(1 - x^2)v - x + \sigma(x, v)\xi \tag{S102}$$

where $\sigma^2(x, v) = \sigma_0 + \sigma_x x^2 + \sigma_v v^2$. We use $\sigma_0 = 1, \sigma_x = 0.3, \sigma_v = 0.1, \Delta t = 0.01, N_{\text{steps}} = 10^4, |\eta| = 0.002$.

6.3 Interacting flocks (Fig. 3)

The model we simulate is a three-dimensional flock of $N = 27$ aligning self-propelled particles, with "soft Lennard-Jones"-type interactions. The particles are initialized on a $3 \times 3 \times 3$ grid with zero velocity. The force on particle i is given by

$$\mathbf{F}_i = \gamma(v_0^2 - |\mathbf{v}_i|^2)\mathbf{v}_i + \sum_{j \neq i} \left[\epsilon_0 \frac{1 - (r/r_0)^3}{(r_{ij}/r_0)^6 + 1} \mathbf{r}_{ij} + \epsilon_1 \exp(-r_{ij}/r_1)\mathbf{v}_{ij} \right] \tag{S103}$$

where $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$, $\mathbf{v}_{ij} = \mathbf{v}_j - \mathbf{v}_i$, while the noise $\sigma\xi_i(t)$ on each particle is isotropic and uncorrelated with others. We choose the parameters $\gamma = 1$, $v_0 = 1.5$, $\epsilon_0 = 4$, $r_0 = 2$, $\epsilon_1 = 1$, $r_1 = 3$ and $\sigma = 1$, which result in a flocking behavior similar to that of bird flocks. The simulation is performed with a time step $dt = 0.005$. It is run for 2000 steps to reach steady state before recording, then the trajectory consisting in 1000 time points with time interval $\Delta t = 0.02$ is recorded.

For the inference, we employ a translation-invariant basis with single-particle and pair interaction terms that is invariant under particle exchange $i \leftrightarrow j$, such that

$$F_{i,\mu} \approx F^{(1)}_{\mu\alpha} c^{(1)}_\alpha(\mathbf{v}_i) + F^{(2)}_{\mu\alpha} \sum_{j \neq i} c^{(2)}_\alpha(\mathbf{x}_i - \mathbf{x}_j, \mathbf{v_i}, \mathbf{v_j}) \tag{S104}$$

The single particle fitting functions are chosen to be polynomials of order up to 3 in the velocity (20 functions). The pair interactions are chosen to be of two kinds: radial functions $\sum_j k(r_{ij})\mathbf{r}_{ij}$ and velocity alignment functions $\sum_j k(r_{ij})\mathbf{v}_{ij}$. We choose the same set of fitting kernels $k(r)$ for both radial force and alignment, $k_n(r) = \exp(-r/r_n)$ with $r_n = 0.5n$ and $n = 1 \ldots 8$. The outcome of force inference is not very sensitive to this choice; r-dependent Gaussian kernels centered at different radii gives similar results. These result in 8 functions for each component of the vectors $\mathbf{r}_{ij}$ and $\mathbf{v}_{ij}$, hence 48 functions pair interaction functions. There are thus 68 functions in the basis, and thus 204 fit parameters for the force field. Inferring the noise tensor and these fit coefficients, we find that the total information in the trajectory presented in Fig. 3 of the main text is $\hat{I} = 320,000$ nats – more than enough to precisely resolve these parameters. Indeed, we find a mean-squared error on the force of 0.015 along the trajectory; this error could be reduced by adding more functions to the basis, or by using longer trajectories, as shown in the convergence plot in Fig. 3E.

References

[1] A. Frishman and P. Ronceray, "Learning force fields from stochastic trajectories," *Physical Review X*, vol. 10, no. 2, p. 21009, 2020.

[2] G. J. Stephens, B. Johnson-Kerner, W. Bialek, and W. S. Ryu, "Dimensionality and Dynamics in the Behavior of C. elegans," *PLoS Comput Biol*, vol. 4, no. 4, p. e1000028, 2008.

[3] J. N. Pedersen, L. Li, C. Gradinaru, R. H. Austin, E. C. Cox, and H. Flyvbjerg, "How to connect time-lapse recorded trajectories of motile microorganisms with dynamical models in continuous time," *Physical Review E*, vol. 94, no. 6, p. 062401, 2016.

[4] F. Ferretti, V. Chardès, T. Mora, A. M. Walczak, and I. Giardina, "Building general Langevin models from discrete data sets," *Phys. Rev. X*, vol. in press, 2020.

[5] 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," *Nature Physics*, vol. 15, no. 6, pp. 595–601, 2019.

[6] P. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*. Springer, 1992.

[7] C. L. Vestergaard, J. N. Pedersen, K. I. Mortensen, and H. Flyvbjerg, "Estimation of motility parameters from trajectory data: A condensate of our recent results," *European Physical Journal: Special Topics*, vol. 224, no. 7, pp. 1151–1168, 2015.

[8] D. B. Brückner, A. Fink, J. O. Rädler, and C. P. Broedersz, "Disentangling the Behavioural Variability of Confined Cell Migration," *J. R. Soc. Interface*, vol. 17, p. 20190689, 2020.

Chapter 8

Learning cell-cell interactions from pair-wise collisions

This chapter is based on the following publication:

Learning the dynamics of cell-cell interactions in confined cell migration

David B. Brückner, Nicolas Arlt, Alexandra Fink, Pierre Ronceray, Joachim O. Rädler†, Chase P. Broedersz†

† corresponding author

Proceedings of the National Academy of Sciences 118, e2016602118 (2021)

Learning the dynamics of cell-cell interactions in confined cell migration

David B. Brückner[1], Nicolas Arlt[1], Alexandra Fink[2], Pierre Ronceray[3], Joachim O. Rädler[2,†] and Chase P. Broedersz[1,4,†]

[1]Arnold Sommerfeld Center for Theoretical Physics and Center for NanoScience, Department of Physics, Ludwig-Maximilian-University Munich, Theresienstr. 37, D-80333 Munich, Germany, [2]Faculty of Physics and Center for NanoScience, Ludwig-Maximilian-University, Geschwister-Scholl-Platz 1, D-80539 Munich, Germany, [3]Center for the Physics of Biological Function, Princeton University, Princeton, NJ 08544, USA, [4]Department of Physics and Astronomy, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands, [†]corresponding authors

The migratory dynamics of cells in physiological processes, ranging from wound healing to cancer metastasis, rely on contact-mediated cell-cell interactions. These interactions play a key role in shaping the stochastic trajectories of migrating cells. While data-driven physical formalisms for the stochastic migration dynamics of single cells have been developed, such a framework for the behavioral dynamics of interacting cells still remains elusive. Here, we monitor stochastic cell trajectories in a minimal experimental cell collider: a dumbbell-shaped micropattern on which pairs of cells perform repeated cellular collisions. We observe different characteristic behaviors, including cells reversing, following and sliding past each other upon collision. Capitalizing on this large experimental data set of coupled cell trajectories, we infer an interacting stochastic equation of motion that accurately predicts the observed interaction behaviors. Our approach reveals that interacting non-cancerous MCF10A cells can be described by repulsion and friction interactions. In contrast, cancerous MDA-MB-231 cells exhibit attraction and anti-friction interactions, promoting the predominant relative sliding behavior observed for these cells. Based on these experimentally inferred interactions, we show how this framework may generalize to provide a unifying theoretical description of the diverse cellular interaction behaviors of distinct cell types.

Significance — When cells migrate collectively, such as to heal wounds or invade tissue, they coordinate through cell-cell interactions. While much is known about the molecular basis of these interactions, the system-level stochastic dynamics of interacting cell behavior remains poorly understood. Here, we design an experimental 'cell collider', providing a large ensemble of interacting cell trajectories. Based on these trajectories, we infer an interacting equation of motion, which accurately predicts characteristic pairwise collision behaviors of different cell lines, including reversal, following or sliding events. This data-driven approach can be used to quantitatively study how molecular perturbations control cell-cell interactions, and may be extended to larger cell collectives, where the inferred interactions could provide key insights into multi-cellular dynamics.

Collective cellular processes such as morphogenesis, wound healing, and cancer invasion, rely on cells moving and rearranging in a coordinated manner. For example, in epithelial wound healing, cells collectively migrate towards the injury and assemble to close the wound [1–3]. In contrast, in metastasizing tumors, cancer cells migrate outwards in a directed fashion and invade surrounding tissue [4]. At the heart of these emergent collective behaviors lie contact-

mediated cell-cell interactions [3,5–10], which are apparent in two-body collisions of cells [10–13]. These cellular interactions depend on complex molecular mechanisms, including cadherin-dependent pathways and receptor-mediated cell-cell recognition [5,10,11,14–17]. At the cellular scale, this molecular machinery leads to coordinated, functional behaviors of interacting cells [3, 5–10], which are highly variable and may take distinct forms in different biological contexts [10, 18–21].

Achieving a quantitative understanding of the stochastic migratory dynamics of cells at the behavioral level could yield key insights into both the underlying molecular mechanisms [22,23] and the biological functions [10] associated to these behaviors. For non-interacting, single migrating cells, data-driven approaches have revealed quantitative frameworks to describe the behavior of free unconstrained migration [24–26] and confined migration in structured environments [27–29]. However, it is still poorly understood how the migratory dynamics of cells are affected by cell-cell interactions and a quantitative formalism for the emergent behavioral dynamics of interacting cells is still lacking [30]. Indeed, it is unclear whether cellular collision behaviors follow a simple set of interaction rules, and if so, how these rules vary for different types of cells.

The study of interacting cell dynamics is complicated by the complex settings in which they take place, confounding contributions of single-cell behavior, interaction with the local microenvironment, and cell-cell interactions. Thus, simplified assays have been developed where cells are confined by one-dimensional micro-patterned patches [31,32] or tracks [19,20,33,34], microfluidics [35], and suspended fibers [36]. In these systems, cells exhibit characteristic behaviors upon pair-wise collisions, including reversal, sliding and following events. Upon contact, many cell types exhibit a tendency to retract, repolarize and migrate apart - termed Contact Inhibition of Locomotion (CIL) [10,13,37]. Indeed, diverse cell types, including epithelial and neural crest cells, predominantly reverse upon collision [19,33,34]. In contrast, the breakdown of CIL is commonly associated with cancer progression [11,18,19,19,38], and cancerous cells have been observed to move past each other more readily than non-cancerous cells [19]. However, it is unclear how to describe these distinct collision behaviors in terms of physical interactions.

Models for collective cell migration often assume repulsive potentials or alignment terms [9,30, 39–42], but the form of these interactions is not derived directly from experimental data. Such data-driven approaches have been developed for single cell migration [24–29], but have not yet been extended to interacting systems. The search for unifying quantitative descriptions of the dynamics of interacting cell trajectories is further complicated by their intrinsic stochasticity, resulting in highly variable migration and collision behavior [19,33,34,36]. Thus, developing a system-level understanding of cell-cell interactions requires a quantitative data-driven approach to learn the full stochastic dynamics of interacting migrating cells.

Here, we develop a theoretical framework for the dynamics of interacting cells migrating in confining environments, inferred directly from experiments. Specifically, we confine pairs of migrating cells into a minimal 'cell collider': a two-state micropattern consisting of two square adhesive sites connected by a thin bridge. Both non-cancerous (MCF10A) and cancerous (MDA-MB-231) human breast tissue cells frequently migrate across the bridge, giving rise to repeated cellular collisions. In line with prior observations [19], we find that while MCF10A cells predominantly reverse upon collision, MDA-MB-231 cells tend to interchange positions by sliding past each other. To provide a quantitative dynamical framework for these distinct interacting behaviors, we focus on a simplified, low-dimensional representation of these collision dynamics by measuring the trajectories of the cell nuclei. The cell collider experiments yield large data sets of such interacting trajectories, allowing us to infer the stochastic equation of motion governing the two-body dynamics of interacting cells. Our data-driven approach reveals the full

structure of the cellular interactions in terms of the relative position and velocity of the cells. Specifically, the dynamics of MCF10A cells are captured by repulsion and friction interactions. In contrast, MDA-MB-231 cells exhibit novel and surprising dynamics, combining attractive and 'anti-friction' interactions, which have no equivalent in equilibrium systems. This inferred model quantitatively captures the key experimental observations, including the distinct collision phenotypes of both cell lines. Our framework can be generalized to provide a conceptual classification scheme for the system-level dynamics of cell-cell interactions, and is able to capture various previously observed types of cell-cell collision behaviors.

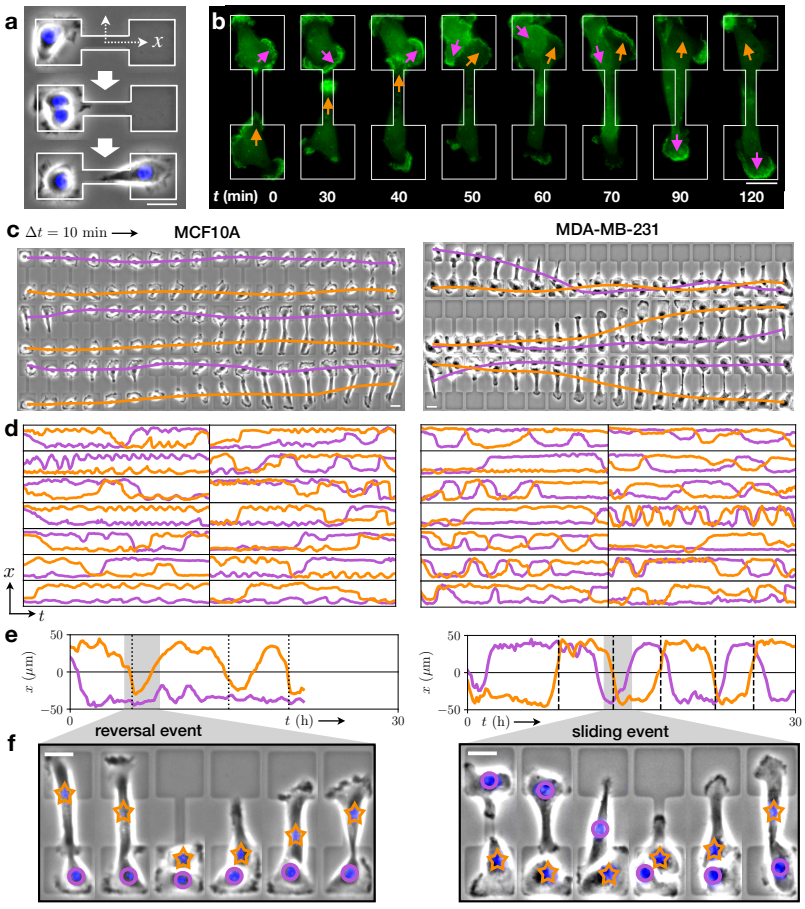


Figure 1: **Stochastic switching dynamics of confined cell pairs. a.** Experimental design: single cells are confined to two-state micropatterns (white outline). We track cell pairs resulting from cell divisions. The stained nucleus is colored in blue. **b.** Time-series of two interacting MDA-MB-231 cells transfected with LifeAct-GFP. Arrows highlight regions of pronounced actin activity, and the arrow color indicates the cell identity. **c.** Brightfield image series with overlaid nuclear trajectories (orange, violet). Images are taken at a time interval $\Delta t = 10$ min. **d.** Sample set of nuclear trajectories $x_{1,2}(t)$ as a function of time, shown for 14 cell pairs. Axes limits are $0 < t < 30$ h and -60 µm $< x <$ 60 µm, with $x = 0$ at the center of the bridge. In total, we tracked 155 MCF10A cell pairs (corresponding to a total trajectory length of 3200 h) and 90 MDA-MB-231 cell pairs (2700 h). **e.** Trajectories of individual cell pairs, with highlighted reversal (dotted lines) and sliding events (dashed lines). **f.** Key stages of the reversal and sliding events, corresponding to the sections highlighted in grey in **e**. Images are shown at 40 min time intervals for MCF10A, and 30 min intervals for MDA-MB-231. Orange stars and violet circles indicate the identities of the cells. In panels **c**-**f**, the left column corresponds to MCF10A cells, and the right column to MDA-MB-231 cells. All scale bars correspond to 25µm.

Two-state micropatterns provide minimal cell collider

To investigate the two-body interaction dynamics of migrating cells, we designed a micropatterned system in which two cells repeatedly collide. The micropattern confines the cells to a fibronectin-coated adhesive region, consisting of a narrow bridge separating two square islands. Outside this dumbbell-shaped region the substrate is passivated with PLL-PEG, to which the cells do not adhere. We first confine single cells to these patterns, as described in previous work [27]. Here, we identify cells which undergo division from which we obtain confined, isolated pairs of daughter cells (Fig. 1a). We employ phase-contrast time-lapse microscopy to study the homotypic interactions of pairs of non-cancerous (MCF10A) and cancerous (MDA-MB-231) human mammary epithelial cells. The confining bridge between the two islands leads to two well-defined configurations, with either both cells on the same island, or on opposite sides of the pattern, between which the system repeatedly switches (Fig. 1c,d and Movies S1-S4). During these switching events, the cells interact with each other. Therefore, our experimental setup offers a simple platform to study the interactions of confined migrating cells in a standardized manner: a minimal 'cell collider'.

Within this cell collider, cells are highly motile and exhibit actin-rich lamellipodia-like protrusions forming at the cell periphery (Fig. 1b, Movie S5). As a simplified, low-dimensional representation of the interaction dynamics, we use the trajectories of the cell nuclei, which reflect the long time-scale interacting behavior of the cells (Fig. 1c). These coupled cell trajectories are highly stochastic. Using this assay, we monitor the stochastic two-body dynamics of hundreds of cells over long periods of time (up to 40h per cell pair) in standardized micro-environments, yielding an unprecedented amount of statistics on cell-cell interactions (Fig. 1d). Importantly, we find that most of the interactive behavior is captured by the x position along the long axis of the pattern (SI Appendix, Section S3). Thus, our cell-collider experiments provide a large data set of low-dimensional trajectories of interacting cells, allowing in-depth statistical analysis of the cellular dynamics.

Cell pairs exhibit mutual exclusion

A key feature of the trajectories for both cell lines is the apparent preference for the configuration in which the cells are on opposite islands (Fig. 1d). Indeed, the positions of the two cells are strongly correlated: the cross-correlation function $\langle x_1(t)x_2(t')\rangle$ exhibits a pronounced negative long-time scale correlation for both cell lines (Fig. 2a). Correspondingly, the joint probability distribution of positions $p(x_1, x_2)$ exhibits prominent peaks where cells occupy opposite sides, and only faint peaks where they are on the same side (Fig. 2b), suggesting two distinct configurations. These configurations are connected by 'paths' in the probability density, along which transitions occur. We find that the cumulative probability $S(t)$ that a configuration switch has not occurred after time t decays more slowly for opposite-side than same-side configurations (Fig. 2c). Taken together, these results indicate that both MCF10A and MDA-MB-231 cells exhibit a mutual exclusion behavior.

MCF10A and MDA-MB-231 cells exhibit distinct collision behavior

While the cells mutually exclude each other, they are also highly migratory and thus frequently transit the constricting bridge. This results in repeated stochastic collision events, providing statistics for how these cells interact during a collision. Following a collision, we observe three

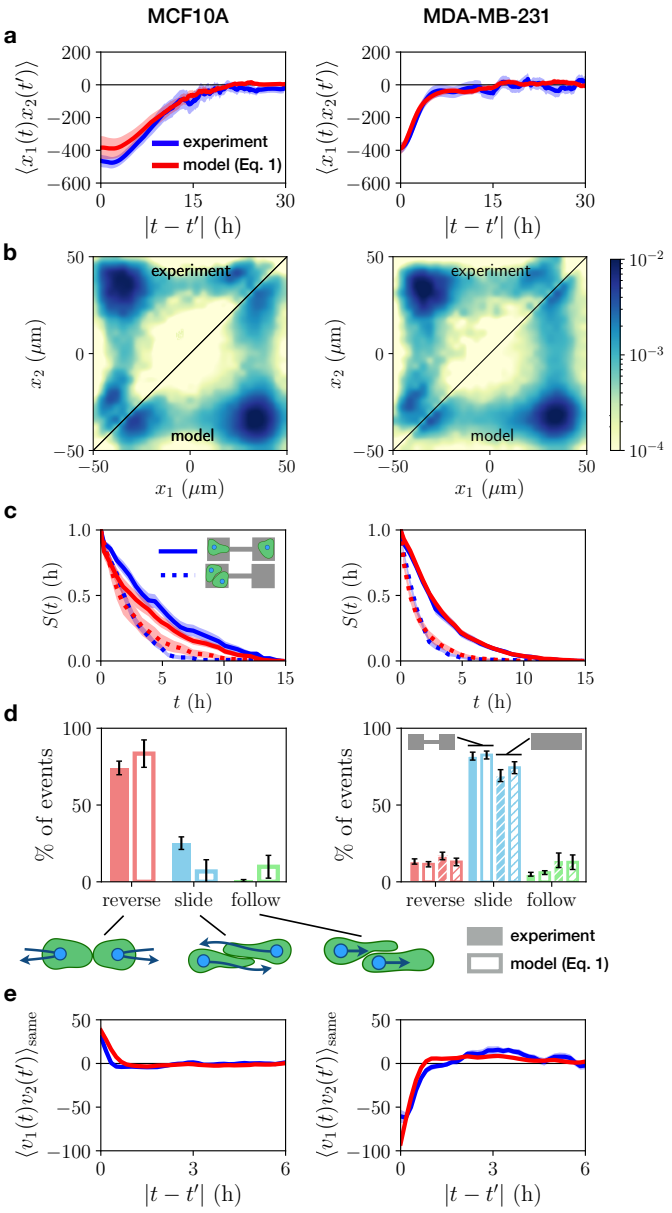


Figure 2: **Statistics of the stochastic interaction dynamics. a.** Cross-correlation function of cell positions $\langle x_1(t)x_2(t')\rangle$. **b.** Joint probability distributions $p(x_1, x_2)$ of cell positions, plotted logarithmically. The top triangle of the symmetric distribution shows the experimental result, the bottom triangle shows the model prediction (for full distributions and linear plots, see SI Appendix, Fig. S12,13). **c.** Probability distribution $S(t)$ giving the probability that a configuration switch has *not* occurred after time t, for the opposite-side configuration (solid) and the same-side configuration (dotted). **d.** Percentages of each of the three types of collision events observed, which are sketched below. For MDA-MB-231 cells, dashed bars correspond to data from cells on micropatterned tracks, with the corresponding model prediction obtained using a single-cell term inferred from single cells on a track, and interaction terms inferred from cell pairs on two-state patterns. **e.** Velocity cross-correlation function $\langle v_1(t)v_2(t')\rangle_{\text{same}}$, calculated for times where the cells occupy the same island. In panels **a** and **c**, experimental data are shown in blue, and model predictions (corresponding to Eqn.(1)) in red. Shaded regions and errorbars denote bootstrap errors (SI Appendix, Section S3).

distinct types of behaviors: reversal events, where the cells turn around upon collision; sliding events, where the cells interchange positions by sliding past each other; and following events where the cells remain in contact and perform a joint transition (Fig. 1e,f, SI Appendix, Section S3). These three behaviors have been previously used as observables of cell-cell interactions in one-dimensional and fibrillar environments [19, 33, 34, 36, 43].

To quantify the interaction behavior of MCF10A and MDA-MB-231 cells, we identify collision events and measure the percentage that result in reversal, sliding or following events (Fig. 2d). Both cell lines exhibit only a small fraction of following events. Remarkably however, we find that collisions of MCF10A cells predominantly result in reversals, while MDA-MB-231 cells typically slide past each other upon collision, in line with observations in other confining geometries [19]. To further explore the generality of this result, we perform additional experiments with MDA-MB-231 cells on micropatterned tracks without constrictions, but the same overall dimensions of the two-state micropatterns (Movies S6, S7). We find that sliding events similarly dominate for MDA-MB-231 cells on this pattern, with similar overall event ratios (Fig. 2d). The different responses to cell-cell contacts are also reflected by the velocity cross-correlation of the two cells when occupying the same side of the two-state micropatterns: $\langle v_1(t)v_2(t')\rangle_{\text{same}}$: MCF10A cells exhibit a positive velocity correlation while MDA-MB-231 cells exhibit a negative velocity correlation (Fig. 2e). Taken together, these findings show that while both cell lines exhibit similar mutual exclusion behavior, there are clear differences in their collision dynamics. This raises a key question: is there an overarching dynamical description which captures both the similarities and differences of these interaction behaviors?

Contact acceleration maps reveal dynamics of cell-cell interactions

Here, we aim to describe the underlying interaction dynamics that capture the full stochastic long time-scale behavior of repeatedly colliding cell pairs. The dynamics of single migrating cells is well described by an equation of motion that is second order in time [24–29], making accelerations the natural quantity to describe cell motility. Specifically, we previously showed that the migration dynamics of single cells in confinement can be described by the average acceleration as a function of cell position x and velocity $v = \mathrm{d}x/\mathrm{d}t$, given by the conditional average $F(x, v) = \langle \dot{v}|x, v\rangle$, where $\dot{v} = \mathrm{d}v/\mathrm{d}t$ [27–29]. To uncover the general structure of the cell-cell interactions in our experiments, we therefore first focus on the observed cellular accelerations upon contact as a function of the distance and relative velocity of the cells. We anticipate contributions from cell-cell interactions to depend on the relative position Δx and relative velocity Δv of the cell pair. Under certain assumptions, which we test in the next section, we can estimate the interactive contribution to cellular accelerations by first subtracting the single-cell contribution $F(x, v)$, and then determining the remaining acceleration as a function of Δx and Δv: $G(\Delta x, \Delta v) = \langle \dot{v} - F(x, v)|\Delta x, \Delta v\rangle$ (see Methods and SI Appendix, Section S3). To further illustrate this approach, we verify that it accurately recovers the functional dependence of simple interactions from simulated trajectories (SI Appendix, Section S3). Thus, we interpret this 'contact acceleration map' as the average acceleration due to the interactions of a cell pair.

Strikingly, we find that MCF10A and MDA-MB-231 cells exhibit qualitatively different contact acceleration maps (Fig. 3a,d). Indeed, for MCF10A cells, the contact acceleration exhibits a clear dependence on the relative position, while MDA-MB-231 cells exhibit accelerations that mainly depend on the relative velocity. We investigate these differences by measuring the 1D-dependence of the contact accelerations as a function of just Δx or Δv. These plots reveal that MCF10A cells exhibit a combination of repulsive accelerations (Fig. 3b) and a weak friction-like component (Fig. 3c). By contrast, MDA-MB-231 cells exhibit contact accelerations with opposite

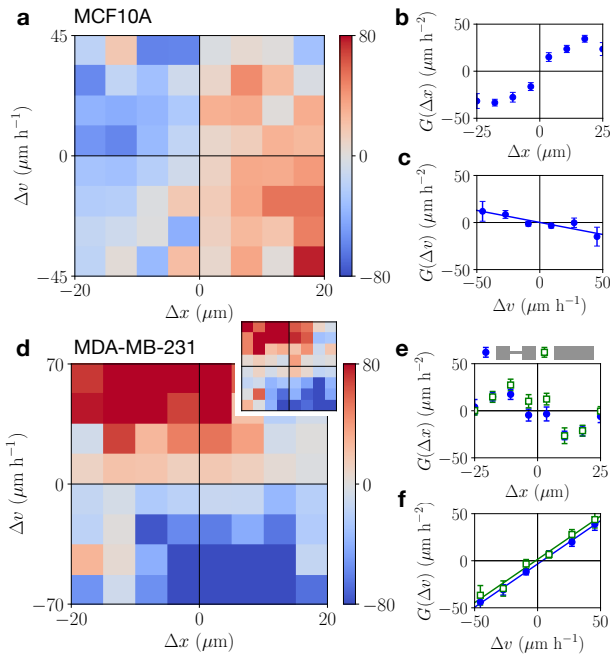


Figure 3: **Contact acceleration maps. a,d.** Contact acceleration maps $G(\Delta x, \Delta v)$, measured in units of μm/h^2. Inset in **d**: Map for MDA-MB-231 cells on micropatterned tracks. **b,e.** Contact accelerations as a function of the cell separation Δx: $G(\Delta x) = \langle \dot{v} - F(x,v)|\Delta x\rangle$. **c,f.** Contact accelerations as a function of the relative velocity of the cells Δv: $G(\Delta v) = \langle \dot{v} - F(x,v)|\Delta v\rangle$. Lines indicate linear fits. Error bars show bootstrap errors. Panels **a-c** show data for MCF10A cells, and panels **d-f** for MDA-MB-231 cells. In panels **e**,**f**, open green symbols correspond to data from experiments on micropatterned tracks.

sign, suggesting an attractive component (Fig. 3e) and an effective linear 'anti-friction' (Fig. 3f). Interestingly, we find that the contact accelerations on micropatterned tracks are qualitatively and quantitatively similar, suggesting that these findings are not very sensitive to the confinement geometry (Inset Fig. 3d). These findings suggest that the contact accelerations of these cells exhibit features that could be described as combinations of cohesive (repulsion/attraction) and frictional terms. This raises the question: are the simple physical interactions suggested by these maps sufficient to describe the complex interaction dynamics of these cell pairs?

Interacting equation of motion captures experimental statistics

To investigate whether the interacting dynamics of MDA-MB-231 and MCF10A cells can be described by the physical interactions implied by the contact acceleration maps, we consider a simple model for cell-cell interactions in confining environments. Motivated by the structure of the contact accelerations, we postulate that the dynamics of the cells can be described by a stochastic equation of motion of the form

$$\frac{\mathrm{d}v}{\mathrm{d}t} = F(x,v) + f(|\Delta x|)\Delta x + \gamma(|\Delta x|)\Delta v + \sigma\eta(t) \tag{1}$$

Here, we assume that the interactions between each cell and the confinement can be described by a term $F(x,v)$, similar to single cell experiments [27]. Furthermore, we assume that the

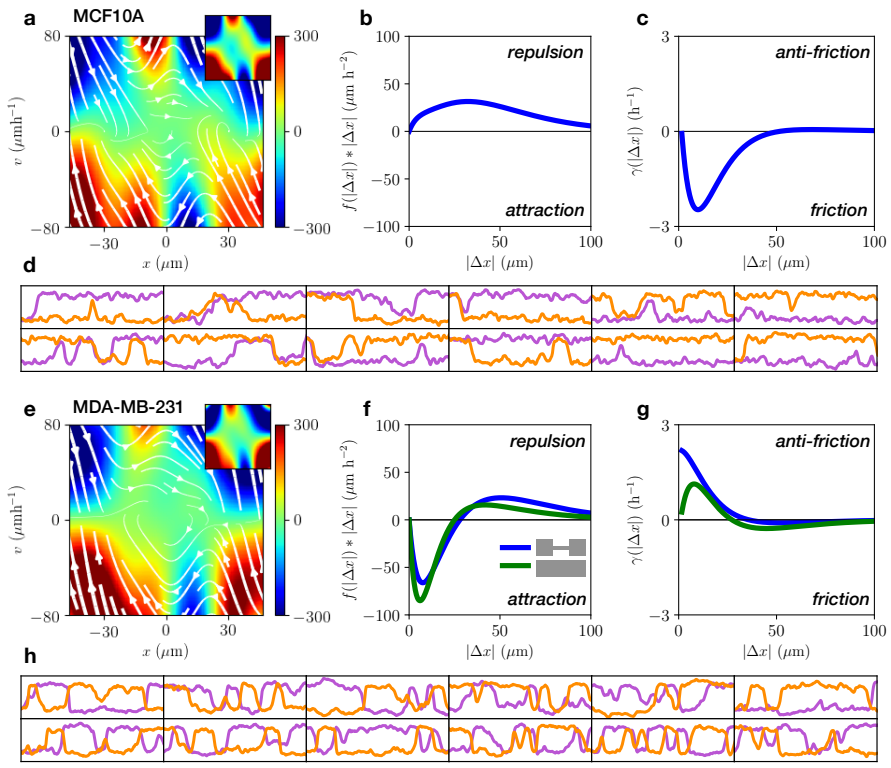


Figure 4: **Equation of motion for interacting cells. a,e.** Single-cell contribution $F(x, v)$ to the interacting dynamics, measured in units of µm/h^2. White lines indicate the flow field given by $(\dot{x}, \dot{v}) = (v, F(x, v))$. Inset: corresponding term inferred from experiments with single cells [27]. **b,f.** Cohesive interaction term $f(|\Delta x|)\Delta x$. Positive values indicate repulsive interactions, while negative values correspond to attraction. **c,g.** Effective frictional interaction term $\gamma(|\Delta x|)$. Here, positive values indicate an effective anti-friction, and negative values an effective frictional interaction. **d,h.** Trajectories obtained from model simulations. Axes limits are 0 ¡ t ¡ 30 h and -60 µm ¡ x ¡ 60 µm. Panels **a-d** show data for MCF10A cells, and panels **e-h** for MDA-MB-231 cells. For MDA-MB-231 cells, green lines show the interactions inferred from cell pairs interacting on micropatterned tracks.

interactions between the two cells can be separately described by two interaction terms: a cohesive term $f(|\Delta x|)\Delta x$, which captures repulsion and attraction; and an effective friction term $\gamma(|\Delta x|)\Delta v$ that may depend on the distance between the cells. The intrinsic stochasticity of the migration dynamics is accounted for by a Gaussian white noise $\eta(t)$, with $\langle \eta(t) \rangle = 0$ and $\langle \eta(t)\eta(t') \rangle = \delta(t - t')$. Note that this equation of motion captures the effective dynamics that describe the cellular accelerations, rather than mechanical forces acting on the cell.

To investigate this model, we first require a systematic approach to infer the systems' stochastic dynamics and delineate single-cell (one-body) and interactive (two-body) contributions to the dynamics. Thus, we employ a rigorous inference method, Underdamped Langevin Inference (ULI) [44], to infer the terms of this equation of motion from the experimentally measured trajectories. In this approach, the inferred terms are completely constrained by the short timescale information in the measured trajectory, i.e. the velocities and accelerations of the cells (see Methods and SI Appendix, Section S4).

Importantly, there is no a priori reason why (1) should provide a reasonable ansatz to correctly capture cell-cell interactions, which could require a more complex description. Thus, we inves-

tigate the predictive power of our model by testing whether it correctly captures experimental statistics that were not used to constrain the terms in (1). Specifically, while the model is learnt on the experimental short time-scale dynamics, we aim to make predictions for long time-scale statistics such as correlation functions. To this end, we simulate stochastic trajectories of interacting cell pairs based on our model (Fig. 4d,h) to make a side-by-side comparison with the experiments. Remarkably, we find that the model performs well in predicting key experimental statistics for both cell lines, including the joint probability distributions (Fig. 2b), the distributions of switching times (Fig. 2c), the cross-correlations of positions and velocity (Fig. 2a,e), as well as the relative fractions of reversal, sliding and following events (Fig. 2d). In contrast, performing the same inference procedure with simpler models than (1), e.g. with only cohesive or friction interactions, shows that simulated trajectories of these models do not capture the observed statistics (SI Appendix, Section S4). To further challenge our approach, we test whether we can use the interactions learnt from experiments on two-state micropatterns to predict the collision behavior in a different confinement geometry. Specifically, we use the single-cell term $F(x, v)$ inferred from single cell data of MDA-MB-231 cells migrating on micropatterned tracks, together with the interactions inferred from cell pair experiments on two-state micropatterns, to predict the collision ratios of cell pairs on tracks. We find that this model accurately predicts the observed event ratios (Fig. 2d), showing that the inferred interactions have predictive power also beyond the data set on which they are learnt.

Remarkably, our inference approach reveals that the inferred single-cell contributions $F(x, v)$ on two-state micropatterns are qualitatively and quantitatively similar to the equivalent term inferred from experiments with single cells for both cell lines [27] (Fig. 4a,e, SI Appendix, Section S4). Also, the inferred noise amplitudes are similar to those inferred from single cell experiments for both cell lines, $\sigma \approx 50\ \mu\mathrm{m/h}^{3/2}$. This suggests that the presence of another cell does not significantly alter the confinement dynamics experienced by one of the cells, and instead manifests in the interaction terms of the equation of motion. Our inference yields the spatial dependence of the cohesion term (Fig. 4b,f) and the effective friction term (Fig. 4c,g). Importantly, the functional dependence of the inferred terms is in accord with our interpretation of the contact acceleration maps (Fig. 3): MCF10A cells exhibit a repulsive cohesive interaction, and a regular effective friction, which reflects that cells slow down as they move past each other. In contrast, MDA-MB-231 cells interact through a predominantly attractive cohesion term, becoming weakly repulsive at long distances, and exhibit effective 'anti-friction'. We infer a similar 'anti-friction' interaction from MDA-MB-231 cell pairs migrating on micropatterned tracks, suggesting that this result is not sensitive to the presence of the constriction (Fig. 4f,g). This anti-friction generates sliding behavior, where cells on average accelerate as they move past each other with increasing relative speed. These results are robust with respect to the details of the inference procedure (SI Appendix, Section S4). Taken together, these findings demonstrate that the dynamics of interacting MCF10A and MDA-MB-231 cells on confining micropatterns are well described by our model ((1)) with distinct types of interactions for the two cell lines.

Interaction behavior space: a theoretical framework for cell-cell interactions

To conceptualize the distinct interactions of MCF10A and MDA-MB-231 cells, we propose an *interaction behavior space*, spanned by the amplitudes of the cohesive and frictional contributions (Fig. 5). Based on our inference, the two cell lines occupy diagonally opposed quadrants in this space. To investigate whether our model ((1)) is able to capture cellular interaction behaviors more broadly, we predict trajectories for various locations within this interaction map. For interactions consisting of repulsion and friction, we find that collisions predominantly result in reversal events, as we have observed for MCF10A cells. In contrast, for positive friction

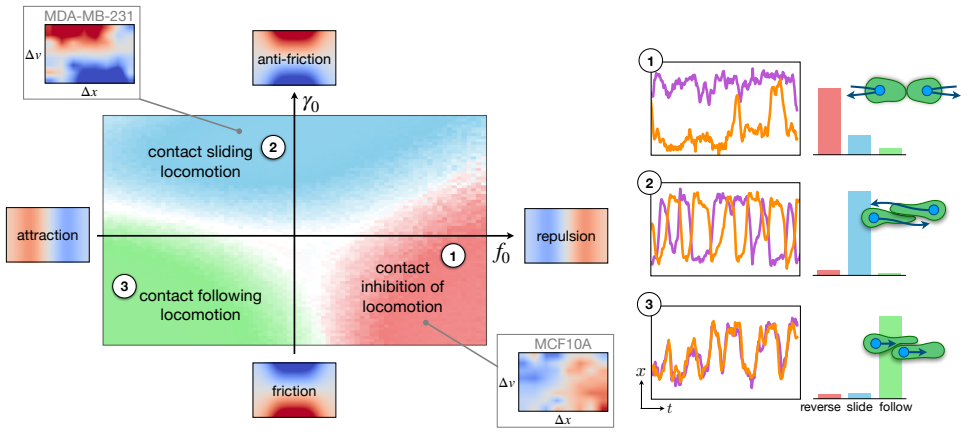


Figure 5: **Interaction behavior space.** We construct an interaction space by varying the amplitude of the cohesive and friction interactions, f_0 and γ_0, respectively. Contact acceleration maps for purely attractive, repulsive, frictional and anti-frictional interactions are indicated on the axes. Based on the inferred short-range interactions, we place MDA-MB-231 and MCF10A cells into diagonally opposed quadrants. Predicted behaviors in the interaction space are obtained by varying the cohesion and friction interactions in our model. Specifically, we simulate a model including the inferred MDA-MB-231 single-cell term $F(x,v)$ together with a cohesive term $f(|\Delta x|) = f_0 g_c(|\Delta x|)$ and an effective friction term $\gamma(|\Delta x|) = \gamma_0 g_f(|\Delta x|)$, for varying f_0 and γ_0. The distance-dependent functions $g_{c,f}$ are positive and monotonically decreasing. These results do not sensitively depend on the specific choice of $F(x,v)$ or $g_{c,f}$ ($g_{c,f} = \exp[(-|\Delta x|/R_0)]$ is used here) (SI Appendix, Section S5). For each parameter combination, reversal, sliding and following events where identified. At each point, the dominant behavior is indicated by the color scheme, and white regions correspond to states where no single behavior contributes more than 50% of events. Numbered insets show sample trajectories from different parts of the interaction map, and the corresponding percentages of reversal (red), sliding (blue), and following events (green).

coefficients, corresponding to effective 'anti-friction', we find that sliding events dominate for all parameter values. This regime thus corresponds to the dynamics we have observed for MDA-MB-231 cells. Finally, attractive interactions with regular friction result in a dominance of following events. The interaction behavior space thus provides an insightful connection between the inferred interaction terms governing the instantaneous dynamics of the system, and the emergent macroscopic, long time-scale collision behavior.

Discussion

In this study, we introduced a conceptual framework for the stochastic behavioral dynamics of interacting cells. To this end, we designed a micropatterned 'cell collider' in which pairs of cells repeatedly collide with each other, providing large amounts of statistics on the long time-scale interactions of migrating cell pairs. A key advantage of this setup is that it yields a large number of collisions under controllable conditions. Moreover, the dynamics of single cells migrating in this confinement is well understood [27], providing a benchmark for the dynamics inferred for interacting cells. We compare the homotypic interaction behavior of the non-malignant MCF10A and the metastatic MDA-MB-231 mammary epithelial cell lines. While phenomenological bottom-up models have been developed to describe cell-cell interactions [30, 32, 43, 45–47], we propose an alternative, top-down approach to learn the interacting stochastic equations of motion governing cell migration from the experimentally observed trajectories. Such an effective model captures the emergent dynamics at the cellular scale which are driven by underlying mechanisms, including the intra-cellular polarity machinery. Our in-

ferred models for interacting cells quantitatively capture the distinct behaviors of the two cell lines. This inference reveals that the dynamics can be decomposed into a one-body motility component, which qualitatively matches that observed in single cell experiments [27], and a two-body interaction term.

The interaction terms we inferred from experiments take qualitatively different forms for the two cell lines: while MCF10A cells exhibit repulsion and effective friction, MDA-MB-231 cells exhibit attraction and a novel and surprising effective 'anti-friction' interaction. At the single-cell level, MDA-MB-231 cells are known to be more invasive than MCF10A cells [48, 49], and express lower levels of the cell-cell adhesion protein E-cadherin [19, 50], possibly underlying the different friction-like interactions we found for these cell lines. These two cell lines also display remarkably different collective behaviors [51–53]: MCF10A cells in 2D epithelial sheets exhibit aligned, directed motion and form compact spheroids in 3D culture, with few invasive branches. In contrast, MDA-MB-231 cells in 2D epithelial sheets exhibit non-aligned, random motion and form invasive, non-contiguous clusters in 3D culture, with significant single-cell dispersion from the cluster. These differences in collective behavior may relate to the distinct types of interactions we have inferred from the two-body dynamics of these cell lines.

Based on the inferred equation of motion, we predict an interaction behavior space to link the interaction terms, which govern the instantaneous stochastic dynamics, to the emergent collision behaviors. The three distinct regimes emerging in our model correspond to specific behaviors observed in experiments for various cell types: predominant reversal behavior on 1D lines has been termed *contact inhibition of locomotion* [33, 34], a common type of cell-cell interaction [5, 8, 10, 13]. By inhibiting intracellular Rho-signalling in neural crest cells, this reversal-dominated behavior could be tuned to following-dominated behavior [34]. Such following behavior has also been identified as an important mechanism in collective migration [12,20,21,54], and was termed *contact following locomotion* [20]. Finally, previous work has shown that reducing the expression levels of E-cadherin enables otherwise reversing cells to mainly slide past each other [19]. For this regime of predominant sliding interactions, we propose the term *contact sliding locomotion*. Based on our interaction behavior space, we find that the 'anti-friction' interactions we identified for MDA-MB-231 cells promote such sliding behavior. The interaction behavior space could thus provide a quantitative classification of distinct modes of interaction that may be achieved through molecular perturbations in experiments [19,34]. On the other end of the scale, the 'anti-friction' interaction type we find here could play a role in collective systems such as the fluidization of tissue in tumor invasion [53,55,56]. The form of the interaction terms we inferred from experiments may thus inform models for collective cell migration [9,30,39–42]. Furthermore, the inference framework we have developed for the dynamics of interacting cell pairs can be extended to infer the dynamics of more complex collective systems, such as small clusters of cells [32, 41, 57], epithelial sheets [42, 58], or 3D organoids [55, 56]. In summary, our model, which we rigorously derive directly from experimental data, could potentially describe the diversity of previously observed cell-cell interaction behaviors in a unifying quantitative framework.

Methods

Sample preparation and cell culture − Fibronectin micropatterns are made by microscale plasma-initiated protein patterning as described previously [27].

MCF10A cells (ATCC) are cultured at 37°C in an atmosphere containing 5% CO_2. The culturing medium DMEM/F-12 including Glutamax (Gibco) is supplemented with 5% Horse Serum (Thermo Fisher), 20 ng/ml hEGF (Sigma), 500ng/ml Hydrocortisone (Sigma), 100ng/ml Cholera Toxin (Sigma) and 10 µg/ml Insulin (Sigma). When passaging cells, the supernatant is aspirated and centrifuged at 300rcf for 8 mins. The adherent cells are washed once with PBS before being detached by a 12-min incubation with Accutase at 37°C. Then the cell solution is re-suspended with culture medium and subsequently centrifuged at 500rcf for 6 mins. Both cell pellets are re-suspended in medium and a fraction is used for further cultivation. For experiments, a drop containing 10,000 cells is added to an ibidi µ-dish (ibidi GmbH) and left to adhere for at least 4h. After that, the medium is exchanged to culture medium without phenol red. 15 nM Hoechst 33342 are added for nuclear staining. Cells are kept in a 5% CO_2-atmosphere and at 37°C during experiments.

MDA-MB-231 cells (DSMZ) are cultured in Minimum Essential Medium (MEM, c.c. pro), containing 10% FBS (Gibco) and 2mM L-Glutamine (c.c. pro). Cells are grown in a 5% CO_2 atmosphere at 37°C. For passaging and experiments, cells are washed once with PBS and trypsinised for 3 min. This cell solution is centrifuged at 1000 rcf for 3 min. The cell pellet is re-suspended in MEM and 10,000 cells are added per µ-dish and left to adhere in the incubator for 4h. The medium is then exchanged to L-15 medium containing L-glutamine (Gibco, supplemented with 10% FCS) and 25 nM Hoechst 33342 (Invitrogen) for staining cell nuclei. Experiments are performed at 37°C without CO_2.

Microscopy and cell tracking − All measurements are performed in time-lapse mode for up to 50 h on an IMIC digital microscope (TILL Photonics) or on a Nikon Eclipse Ti microscope using a 10x objective. The samples are kept in a heated chamber (ibidi GmbH or Okolab) at 37°C throughout the measurements. Images (brightfield and DAPI) are acquired every 10 mins. Trajectories of cell pairs are obtained by selecting cells that undergo division during the experiment. Following division and subsequent re-attachment to the micropattern, we track the trajectories of the cell nuclei. A band pass filter is applied to the images of the nuclei, then images are binarised. The cell trajectories are determined by tracking the binarised images using a Matlab tracking algorithm [59]. For further details, see SI Appendix, Section S2.

Contact acceleration maps − To gain insight in the general structure of the accelerations due to cell-cell interaction, we introduce contact acceleration maps. We estimate single-cell component of the dynamics from the accelerations observed at time-points where the cells are far apart $F(x,v) = \langle \dot{v}_i | x_i, v_i; |\Delta x_{ij}| > \ell \rangle$, where we take the threshold distance $\ell = 25$µm. To obtain the accelerations due to cell-cell contacts, we take the time points where cells are close together and calculate the average acceleration as a function of relative position $\Delta x_{ij} = x_i - x_j$ and velocity $\Delta v_{ij} = v_i - v_j$ of cell i and cell j: $G(\Delta x, \Delta v) \approx \langle \dot{v}_i - F(x_i, v_i) | \Delta x_{ij}, \Delta v_{ij}; |\Delta x_{ij}| < \ell \rangle$. We show that for simple simulated examples, this approach accurately recovers the structure of the interaction terms. For more details, see SI Appendix, Section S3.

Underdamped Langevin Inference − From the short time-scale dynamics of the measured cell trajectories $x(t)$, we infer the second order stochastic differential equation that governs the motion [26, 27, 44, 60]. Specifically, to infer the terms of our model ((1)), we employ Underdamped Langevin Inference [44], a method which is robust with respect to the effects of the discrete sam-

pling of $x(t)$ and the presence of measurement errors. Briefly, we fit the experimentally measured accelerations using a linear combination of basis functions $\{b(x_i, v_i), u(|\Delta x_{ij}|)\Delta x_{ij}, u(|\Delta x_{ij}|)\Delta x_{ij}\}$ using rigorous stochastic estimators [44]. For the single cell terms $b(x_i, v_i)$, we use a combination of polynomials and Fourier modes, while for the interaction kernels $u(|\Delta x_{ij}|)$ we use exponential functions. The inference results do not sensitively depend on the choice of basis functions. For more details, see SI Appendix, Section S4.

References

[1] M. Poujade, E. Grasland-Mongrain, A. Hertzog, J. Jouanneau, P. Chavrier, B. Ladoux, A. Buguin, and P. Silberzan, "Collective migration of an epithelial monolayer in response to a model wound," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 41, pp. 15988–15993, 2007.

[2] B. Stramer, W. Wood, M. J. Galko, M. J. Redd, A. Jacinto, S. M. Parkhurst, and P. Martin, "Live imaging of wound inflammation in Drosophila embryos reveals key roles for small GTPases during in vivo cell migration," *Journal of Cell Biology*, vol. 168, no. 4, pp. 567–573, 2005.

[3] H. Weavers, J. Liepe, A. Sim, W. Wood, P. Martin, and M. P. Stumpf, "Systems Analysis of the Dynamic Inflammatory Response to Tissue Damage Reveals Spatiotemporal Properties of the Wound Attractant Gradient," *Current Biology*, vol. 26, no. 15, pp. 1975–1989, 2016.

[4] P. Friedl and K. Wolf, "Tumour-cell invasion and migration: diversity and escape mechanisms.," *Nature reviews. Cancer*, vol. 3, no. 5, pp. 362–74, 2003.

[5] C. Carmona-Fontaine, H. K. Matthews, S. Kuriyama, M. Moreno, G. A. Dunn, M. Parsons, C. D. Stern, and R. Mayor, "Contact inhibition of locomotion in vivo controls neural crest directional migration," *Nature*, vol. 456, no. 7224, pp. 957–961, 2008.

[6] V. Villar-Cerviño, M. Molano-Mazón, T. Catchpole, M. Valdeolmillos, M. Henkemeyer, L. M. Martínez, V. Borrell, and O. Marín, "Contact Repulsion Controls the Dispersion and Final Distribution of Cajal-Retzius Cells," *Neuron*, vol. 77, no. 3, pp. 457–471, 2013.

[7] E. Theveneau, L. Marchant, S. Kuriyama, M. Gull, B. Moepps, M. Parsons, and R. Mayor, "Collective Chemotaxis Requires Contact-Dependent Cell Polarity," *Developmental Cell*, vol. 19, no. 1, pp. 39–53, 2010.

[8] J. R. Davis, C. Y. Huang, J. Zanet, S. Harrison, E. Rosten, S. Cox, D. Y. Soong, G. A. Dunn, and B. M. Stramer, "Emergence of embryonic pattern through contact inhibition of locomotion," *Development*, vol. 139, no. 24, pp. 4555–4560, 2012.

[9] B. Smeets, R. Alert, J. Pešek, I. Pagonabarraga, H. Ramon, and R. Vincent, "Emergent structures and dynamics of cell colonies by contact inhibition of locomotion," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, no. 51, pp. 14621–14626, 2016.

[10] B. Stramer and R. Mayor, "Mechanisms and in vivo functions of contact inhibition of locomotion," *Nature reviews. Molecular cell biology*, vol. 18, no. 1, pp. 43–55, 2017.

[11] J. W. Astin, J. Batson, S. Kadir, J. Charlet, R. A. Persad, D. Gillatt, J. D. Oxley, and C. D. Nobes, "Competition amongst Eph receptors regulates contact inhibition of locomotion and invasiveness in prostate cancer cells," *Nature Cell Biology*, vol. 12, no. 12, pp. 1194–1204, 2010.

[12] J. M. Teddy and P. M. Kulesa, "In vivo evidence for short- and long-range cell communication in cranial neural crest cells," *Development*, vol. 131, no. 24, pp. 6141–6151, 2004.

[13] M. Abercrombie and J. E. Heaysman, "Observations on the social behaviour of cells in tissue culture. II. "Monolayering" of fibroblasts," *Experimental Cell Research*, vol. 6, no. 2, pp. 293–306, 1954.

[14] J. R. Davis, A. Luchici, F. Mosis, J. Thackery, J. A. Salazar, Y. Mao, G. A. Dunn, T. Betz, M. Miodownik, and B. M. Stramer, "Inter-cellular forces orchestrate contact inhibition of locomotion," *Cell*, vol. 161, no. 2, pp. 361–373, 2015.

[15] R. Moore, E. Theveneau, S. Pozzi, P. Alexandre, J. Richardson, A. Merks, M. Parsons, J. Kashef, C. Linker, and R. Mayor, "Par3 controls neural crest migration by promoting microtubule catastrophe during contact inhibition of locomotion," *Development*, vol. 140, no. 23, pp. 4763–4775, 2013.

[16] H. K. Matthews, L. Marchant, C. Carmona-Fontaine, S. Kuriyama, J. Larraín, M. R. Holt, M. Parsons, and R. Mayor, "Directional migration of neural crest cells in vivo is regulated by Syndecan-4/Rac1 and non-canonical Wnt signaling/RhoA," *Development*, vol. 135, no. 10, pp. 1771–1780, 2008.

[17] S. Kadir, J. W. Astin, L. Tahtamouni, P. Martin, and C. D. Nobes, "Microtubule remodelling is required for the front-rear polarity switch during contact inhibition of locomotion," *Journal of Cell Science*, vol. 124, no. 15, pp. 2642–2653, 2011.

[18] M. Abercrombie, "Contact inhibition and malignancy," *Nature*, vol. 281, no. 5729, pp. 259–262, 1979.

[19] D. F. Milano, N. A. Ngai, S. K. Muthuswamy, and A. R. Asthagiri, "Regulators of Metastasis Modulate the Migratory Response to Cell Contact under Spatial Confinement," *Biophysical Journal*, vol. 110, no. 8, pp. 1886–1895, 2016.

[20] D. Li and Y. L. Wang, "Coordination of cell migration mediated by sitedependent cell-cell contact," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 115, no. 42, pp. 10678–10683, 2018.

[21] M. Hayakawa, T. Hiraiwa, Y. Wada, H. Kuwayama, and T. Shibata, "Polar pattern formation induced by contact following locomotion in a multicellular system," *eLife*, vol. 9, p. e53609, 2020.

[22] P. Maiuri, J. F. Rupprecht, S. Wieser, V. Ruprecht, O. Bénichou, N. Carpi, M. Coppey, S. De Beco, N. Gov, C. P. Heisenberg, C. Lage 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*, vol. 161, no. 2, pp. 374–386, 2015.

[23] I. Lavi, M. Piel, A.-M. Lennon-Duménil, R. Voituriez, and N. S. Gov, "Deterministic patterns in cell motility," *Nature Physics*, vol. 12, no. August, pp. 1146–1152, 2016.

[24] D. Selmeczi, S. Mosler, P. H. Hagedorn, N. B. Larsen, and H. Flyvbjerg, "Cell motility as persistent random motion: theories from experiments.," *Biophysical journal*, vol. 89, no. 2, pp. 912–931, 2005.

[25] L. Li, E. C. Cox, and H. Flyvbjerg, "'Dicty dynamics': Dictyostelium motility as persistent random motion," *Physical Biology*, vol. 8, no. 4, p. 046006, 2011.

[26] J. N. Pedersen, L. Li, C. Gradinaru, R. H. Austin, E. C. Cox, and H. Flyvbjerg, "How to connect time-lapse recorded trajectories of motile microorganisms with dynamical models in continuous time," *Physical Review E*, vol. 94, no. 6, p. 062401, 2016.

[27] 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," *Nature Physics*, vol. 15, no. 6, pp. 595–601, 2019.

[28] D. B. Brückner, A. Fink, J. O. Rädler, and C. P. Broedersz, "Disentangling the Behavioural Variability of Confined Cell Migration," *J. R. Soc. Interface*, vol. 17, p. 20190689, 2020.

[29] A. Fink, D. B. Brückner, C. Schreiber, P. J. Röttgermann, C. P. Broedersz, and J. O. Rädler, "Area and Geometry Dependence of Cell Migration in Asymmetric Two-State Micropatterns," *Biophysical Journal*, vol. 118, no. 3, pp. 552–564, 2020.

[30] R. Alert and X. Trepat, "Physical Models of Collective Cell Migration," *Annual Review of Condensed Matter Physics*, vol. 11, no. 1, pp. 77–101, 2020.

[31] S. Huang, C. P. Brangwynne, K. K. Parker, and D. E. Ingber, "Symmetry-breaking in mammalian cell cohort migration during tissue pattern formation: Role of random-walk persistence," *Cell Motility and the Cytoskeleton*, vol. 61, no. 4, pp. 201–213, 2005.

[32] F. J. Segerer, F. Thüroff, A. Piera Alberola, E. Frey, and J. O. Rädler, "Emergence and persistence of collective cell migration on small circular micropatterns," *Physical Review Letters*, vol. 114, no. 22, p. 228102, 2015.

[33] R. A. Desai, S. B. Gopal, S. Chen, and C. S. Chen, "Contact inhibition of locomotion probabilities drive solitary versus collective cell migration," *Journal of the Royal Society Interface*, vol. 10, no. 88, 2013.

[34] E. Scarpa, A. Roycroft, E. Theveneau, E. Terriac, M. Piel, R. Mayor, E. Scarpa, A. Roycroft, E. Theveneau, E. Terriac, M. Piel, and R. Mayor, "A novel method to study contact inhibition of locomotion using micropatterned substrates," *Biology Open*, vol. 2, pp. 901–906, 2013.

[35] B. Lin, T. Yin, Y. I. Wu, T. Inoue, and A. Levchenko, "Interplay between chemotaxis and contact inhibition of locomotion determines exploratory cell migration," *Nature Communications*, vol. 6, 2015.

[36] J. Singh, B. A. Camley, and A. S. Nain, "Rules of Contact Inhibition of Locomotion for Cells on Suspended Nanofibers," *bioRxiv*, 2020.

[37] R. Mayor and C. Carmona-Fontaine, "Keeping in touch with contact inhibition of locomotion," *Trends in Cell Biology*, vol. 20, no. 6, pp. 319–328, 2010.

[38] M. Abercrombie and J. E. Heaysman, "Invasiveness of Sarcoma Cells," *Nature*, vol. 174, no. 4432, pp. 697–698, 1954.

[39] N. Sepúlveda, L. Petitjean, O. Cochet, E. Grasland-Mongrain, P. Silberzan, and V. Hakim, "Collective Cell Motion in an Epithelial Sheet Can Be Quantitatively Described by a Stochastic Interacting Particle Model," *PLoS Computational Biology*, vol. 9, no. 3, 2013.

[40] M. Basan, J. Elgeti, E. Hannezo, W. J. Rappel, and H. Levine, "Alignment of cellular motility forces with tissue flow as a mechanism for efficient wound healing," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 7, pp. 2452–2459, 2013.

[41] K. Copenhagen, G. Malet-engra, W. Yu, G. Scita, N. Gov, and A. Gopinathan, "Frustration-induced phases in migrating cell clusters," *Science Advances*, vol. 4, p. eaar8483, 2018.

[42] S. Garcia, E. Hannezo, J. Elgeti, J.-F. Joanny, P. Silberzan, and N. S. Gov, "Physics of active jamming during collective cellular motion in a monolayer," *Proceedings of the National Academy of Sciences*, vol. 112, no. 50, pp. 15314–15319, 2015.

[43] D. A. Kulawiak, B. A. Camley, and W. J. Rappel, "Modeling Contact Inhibition of Locomotion of Colliding Cells Migrating on Micropatterned Substrates," *PLoS Computational Biology*, vol. 12, no. 12, 2016.

[44] D. B. Brückner, P. Ronceray, and C. P. Broedersz, "Inferring the dynamics of underdamped stochastic systems," *Physical Review Letters*, vol. 125, no. 5, p. 58103, 2020.

[45] B. A. Camley and W. J. Rappel, "Velocity alignment leads to high persistence in confined cells," *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 89, no. 6, p. 062705, 2014.

[46] J. Löber, F. Ziebert, and I. S. Aranson, "Collisions of deformable cells lead to collective migration," *Scientific Reports*, vol. 5, pp. 1–7, 2015.

[47] S. Vedel, S. Tay, D. M. Johnston, H. Bruus, and S. R. Quake, "Migration of cells in a social context," *Proceedings of the National Academy of Sciences*, vol. 110, no. 1, pp. 129–134, 2013.

[48] M. Mak, C. A. Reinhart-King, and D. Erickson, "Microfabricated physical spatial gradients for investigating cell migration and invasion dynamics," *PLoS ONE*, vol. 6, no. 6, p. e20825, 2011.

[49] C. M. Kraning-Rush, S. P. Carey, M. C. Lampi, and C. A. Reinhart-King, "Microfabricated collagen tracks facilitate single cell metastatic invasion in 3D," *Integrative Biology*, vol. 5, no. 3, pp. 606–616, 2013.

[50] C. L. Sommers, E. W. Thompson, J. A. Torri, R. Kemler, E. P. Gelmann, and S. W. Byers, "Cell adhesion molecule uvomorulin expression in human breast cancer cell lines: relationship to morphology and invasive capacities.," *Cell growth & differentiation*, vol. 2, pp. 365–372, aug 1991.

[51] S. P. Carey, A. Starchenko, A. L. McGregor, and C. A. Reinhart-King, "Leading malignant cells initiate collective epithelial cell invasion in a three-dimensional heterotypic tumor spheroid model," *Clinical and Experimental Metastasis*, vol. 30, no. 5, pp. 615–630, 2013.

[52] R. M. Lee, M. I. Vitolo, W. Losert, and S. S. Martin, "Distinct Roles of Tumor-Associated Mutations in Collective Cell Migration," *bioRxiv*, 2020.

[53] W. Kang, J. Ferruzzi, C.-P. Spatarelu, Y. L. Han, Y. Sharma, S. A. Koehler, J. P. Butler, D. Roblyer, M. H. Zaman, M. Guo, Z. Chen, A. F. Pegoraro, and J. J. Fredberg, "Tumor invasion as non-equilibrium phase separation," *bioRxiv*, p. 2020.04.28.066845, jan 2020.

[54] T. Fujimori, A. Nakajima, N. Shimada, and S. Sawai, "Tissue self-organization based on collective cell migration by contact activation of locomotion and chemotaxis," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 116, no. 10, pp. 4291–4296, 2019.

[55] A. Palamidessi, C. Malinverno, E. Frittoli, S. Corallino, E. Barbieri, S. Sigismund, G. V. Beznoussenko, E. Martini, M. Garre, I. Ferrara, C. Tripodo, F. Ascione, E. A. Cavalcanti-Adam, Q. Li, P. P. Di Fiore, D. Parazzoli, F. Giavazzi, R. Cerbino, and G. Scita, "Unjamming overcomes kinetic and proliferation arrest in terminally differentiated cells and promotes collective motility of carcinoma," *Nature Materials*, vol. 18, pp. 1252–1263, 2019.

[56] Y. L. Han, A. F. Pegoraro, H. Li, K. Li, Y. Yuan, G. Xu, Z. Gu, J. Sun, Y. Hao, S. K. Gupta, Y. Li, W. Tang, H. Kang, L. Teng, J. J. Fredberg, and M. Guo, "Cell swelling, softening and invasion in a three-dimensional breast cancer model," *Nature Physics*, vol. 16, no. 1, pp. 101–108, 2020.

[57] J. D'alessandro, A. P. Solon, Y. Hayakawa, C. Anjard, F. Detcheverry, J. P. Rieu, and C. Rivière, "Contact enhancement of locomotion in spreading cell colonies," *Nature Physics*, vol. 13, no. 10, pp. 999–1005, 2017.

[58] T. E. Angelini, E. Hannezo, X. Trepat, J. J. Fredberg, and D. A. Weitz, "Cell migration driven by cooperative substrate deformation patterns," *Physical Review Letters*, vol. 104, no. 16, pp. 1–4, 2010.

[59] D. Blair and E. Dufresne, "The Matlab Particle Tracking Code Repository," *http://site.physics.georgetown.edu/matlab/*, 2008.

[60] F. Ferretti, V. Chardès, T. Mora, A. M. Walczak, and I. Giardina, "Building General Langevin Models from Discrete Datasets," *Physical Review X*, vol. 10, no. 3, p. 031018, 2020.

Supplementary Material:

Learning the dynamics of cell-cell interactions in confined cell migration

David B. Brückner, Nicolas Arlt, Alexandra Fink, Pierre Ronceray,
Joachim O. Rädler and Chase P. Broedersz

Contents

1 Movie descriptions

Supplementary Movie 1
A pair of MCF10A cells transitioning repeatedly between the square adhesion sites of a two-state micropattern. The cell nucleus is fluorescently labelled to allow automated tracking of cell positions. Scale bar: 25 µm.

Supplementary Movie 2
A pair of MCF10A cells transitioning repeatedly between the square adhesion sites of a two-state micropattern. The cell nucleus is fluorescently labelled to allow automated tracking of cell positions. Scale bar: 25 µm.

Supplementary Movie 3
A pair of MDA-MB-231 cells transitioning repeatedly between the square adhesion sites of a two-state micropattern. The cell nucleus is fluorescently labelled to allow automated tracking of cell positions. Scale bar: 25 µm.

Supplementary Movie 4
A pair of MDA-MB-231 cells transitioning repeatedly between the square adhesion sites of a two-state micropattern. The cell nucleus is fluorescently labelled to allow automated tracking of cell positions. Scale bar: 25 µm.

Supplementary Movie 5
A pair of MDA-MB-231 cells transfected with LifeAct-GFP to visualize actin on a two-state micropattern. The outline of the micropattern is drawn in white as a reference. Actin hotspots are visible at the tip of the lamellipodia. Scale bar: 25 µm.

Supplementary Movie 6
A pair of MDA-MB-231 cells migrating on a rectangular micropattern without constriction with the same overall dimensions as the two-state micropatterns. The cell nucleus is fluorescently labelled to allow automated tracking of cell positions. Scale bar: 25 µm.

Supplementary Movie 7
A pair of MDA-MB-231 cells migrating on a rectangular micropattern without constriction with the same overall dimensions as the two-state micropatterns. The cell nucleus is fluorescently labelled to allow automated tracking of cell positions. Scale bar: 25 µm.

Supplementary Data 1-3
The three data files correspond to the data for MDA-MB-231 cells on two-state micropatterns, MCF10A cells on two-state micropatterns, and MDA-MB-231 cells on micropatterned tracks, respectively. The files contain the x-positions of the cells in µm, ordered consecutively such that rows 1,2 correspond to the two cells in the first cell pair.

2 Further experimental details

2.1 Micropattern design

The adhesive sites of the two-state micropatterns have square dimensions $(36.7 \pm 0.6)^2$ µm^2. The connecting bridge has length (35.3 ± 0.5) µm and width (6.9 ± 0.6) µm. The quoted errors correspond to the deviations in the dimensions of final protein patterns which are due to the intrinsic variance of the manual stamping process and the measurement uncertainty associated with the limited resolution of the brightfield images.

2.2 Cell exclusion criteria

We track the trajectories of a large number of cell pairs confined to two-state micropatterns. Following previous work [1], we apply the following inclusion criteria in our analysis:

1. Two-cell trajectories are obtained by selecting cells which undergo division during the experiment. Tracking begins after both daughter cells have re-attached to the pattern after division. Tracking is terminated when one of the two cells rounds up for division again.
2. Both cells and their protrusions are confined within the borders of the micropattern.
3. Both cells show no abnormalities such as multiple nuclei or the occurrence of cell death or detachment from the substrate.
4. At least one transition occurs during the experiment (i.e. at least one of the two cells performs a transition across the bridge of the micropattern).

2.3 Cell transfections

For live-cell imaging of actin, approximately 10,000 MDA-MB-231 cells are seeded in patterned μ-dishes and left to adhere for 12 h. As a cell culturing medium, MEM with Glutamax (Gibco) supplemented with 10% FCS is used. 500ng LiveAct-GFP mRNA (prepared in-house) is resuspended in OptiMEM (Gibco) to a final volume of 150 µL. This solution is then added to a mix of 1.25 µL Lipofectamine 2000 (Invitrogen) and 123.75 µL OptiMEM, and left to incubate for 20 minutes at room temperature. Subsequently, cells are rinsed once with PBS and the transfection mix is added and left on the cells for at least 5 h, before being replaced by L-15 medium. Cells are imaged every 10 minutes on the Nikon Ti Eclipse microscope using a 60x oil-immersion objective.

2.4 Tracking procedure

Brightfield and fluorescence images of the stained nuclei are acquired every 10 min. A band pass filter is applied to the images of the nuclei, then images are binarised. The cell trajectories are determined by tracking the nuclei using a Matlab tracking algorithm [2]. After application of the tracking algorithm, each trajectory is inspected manually to verify that only two particles have been identified, and that the trajectories exhibit no gaps that consist of more than one missing frame. In the rare case of a single missing frame, we interpolate linearly between the previous and the subsequent coordinate. Furthermore, we verify that the tracking algorithm has correctly identified the identity of the two cells. In the cases where the two cells have been mixed up, this is corrected manually. After tracking, the reference boundaries of the patterns are determined manually by means of the bright-field images, on which the micropatterns are visible.

3 Analysis of the experimental dynamics

3.1 Error analysis

All errorbars throughout the paper correspond to bootstrap errors, as described in refs. [3,4]. Briefly, from our data set of N cell pair trajectories $\{\mathbf{x}_k\}$, where $k = 1...N$, we generate N_{BS} bootstrap realizations by randomly sampling N cell pair trajectories with replacement for each realization. To obtain the error in an observable θ_{expt} measured from the experimental data set, we estimate the value of θ for each bootstrap realization and take the standard deviation of all obtained θs as our estimate for the error in θ_{expt}. To obtain the error in an observable θ_{model} predicted by our model, we perform ULI on each of the N_{BS} bootstrap realizations to obtain N_{BS} bootstrapped models. Then, we simulate a large number of trajectories for each bootstrap model, and estimate θ for each set of trajectories. The standard deviation of these θs is our estimate for the error in θ_{model}.

3.2 Stochastic switching dynamics and cross-correlation functions

Survival probability functions − To quantify the transition dynamics of cell pairs within the two-state micropattern, we define two configurations: one where both cells occupy the same side of the pattern, and one where they occupy opposite sides, with the two sides defined by $x < 0$ and $x > 0$ (Fig. S1a). We obtain similar results if the boundaries are instead defined by $x < -L/2$ and $x > L/2$, where L is the length of the bridge. We define a switching event as a switch from a same-side to an opposite-side configuration, and vice versa. To gain insight into the dynamics of these configuration switches, we calculate the times between switches, which yields the average

dwell times of the same-side configuration $\{\tau_{\text{same}}\}$ and the opposite-side configuration $\{\tau_{\text{opp}}\}$. Thus, we can calculate the survival probability distribution function of each configuration, i.e. the probability that a switch has not occurred after time t, given that the system is initially in state $k = \{\text{same}, \text{opp}\}$:

$$S_k(t) = 1 - \int_0^t p(\tau_k) \mathrm{d}\tau_k \tag{S1}$$

where $p(\tau_k)$ is the probability distribution of τ_k. These distributions show that the same-side configuration is typically occupied for shorter times than the opposite-side configuration (Fig. S1b,d). Both survival probability functions exhibit an initially exponential decay (Fig. S1c,e), but become noisy at long times due to limited sampling.

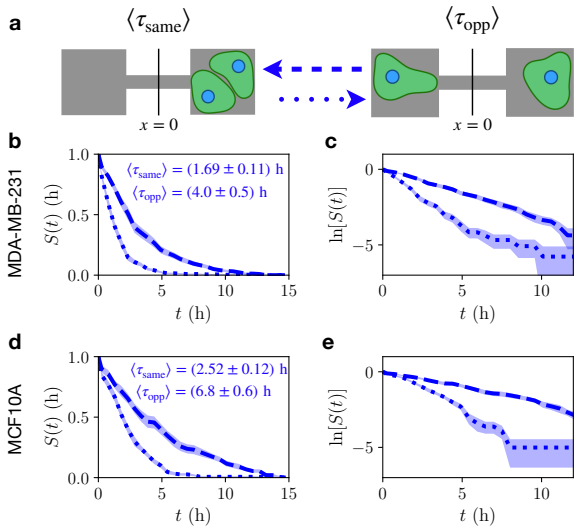


Figure S1: **Stochastic switching dynamics.** **a.** Sketch of the same- and opposite-side configurations. **b.** Survival probability functions of the same- (dotted line) and opposite-side (dashed line) configurations, for MDA-MB-231 cells. The average dwell times of each state are given. **c.** Log-linear plot of panel **b**. **d.** Survival probability functions of the same- (dotted line) and opposite-side (dashed line) configurations, for MCF10A cells. The average dwell times of each state are given. **e.** Log-linear plot of panel **d**. Shaded regions and error intervals for the average dwell times denote bootstrap errors.

Position cross-correlation function – The cross-correlation function of cell positions, defined by:

$$\langle x_1(t)x_2(t')\rangle := \frac{1}{2\sum_{j=1}^{N_{\text{pairs}}} T_j} \sum_{j=1}^{N_{\text{pairs}}} \sum_{t'=1}^{T_j} x_1(t)x_2(t') \tag{S2}$$

where N_{pairs} is the number of tracked cell pairs, and T_j is the total number of time-points in the trajectory of pair j.

Velocity cross-correlation function – To investigate the correlations in migration velocities when the cells are in contact, we calculate the cross-correlation function of cell velocities $v(t) = (x(t) - x(t + \Delta t))/\Delta t$, subject to the constraint that the cells are in the same-side configuration:

$$\langle v_1(t)v_2(t')\rangle_{\text{same}} := \frac{1}{2\sum_{j=1}^{N_{\text{pairs}}} T_j^{\text{same}}} \sum_{j=1}^{N_{\text{pairs}}} \sum_{\{\theta_{\text{same}}\}} v_1(t)v_2(t') \tag{S3}$$

where $\{\theta_{\text{same}}\}$ is the set of time-point combinations (t, t'), where the cells are in the same-side configuration at both time t and t', and T_j^{same} is the total number of such time-point combinations of cell pair j.

3.3 Movement in the second dimension

The two-state micropattern is designed in such a way that most of the behavior occurs in the x-direction. We find that most of the interaction behavior is indeed captured by the x-components of the trajectories: the variance in y-motion is small (Fig. S2a,b), and the joint probability distribution $p(y_1, y_2)$ is peaked around $(0, 0)$ and exhibits no special structure (Fig. S2d,f). Furthermore, the cross-correlation function of y-positions $\langle y_1(t)y_2(t')\rangle$ vanishes (Fig. S2g,h). Thus, we find that the x-component captures most of the behavior displayed by these interacting cells, and thus provides a simple, low-dimensional representation of the system's behavioral dynamics.

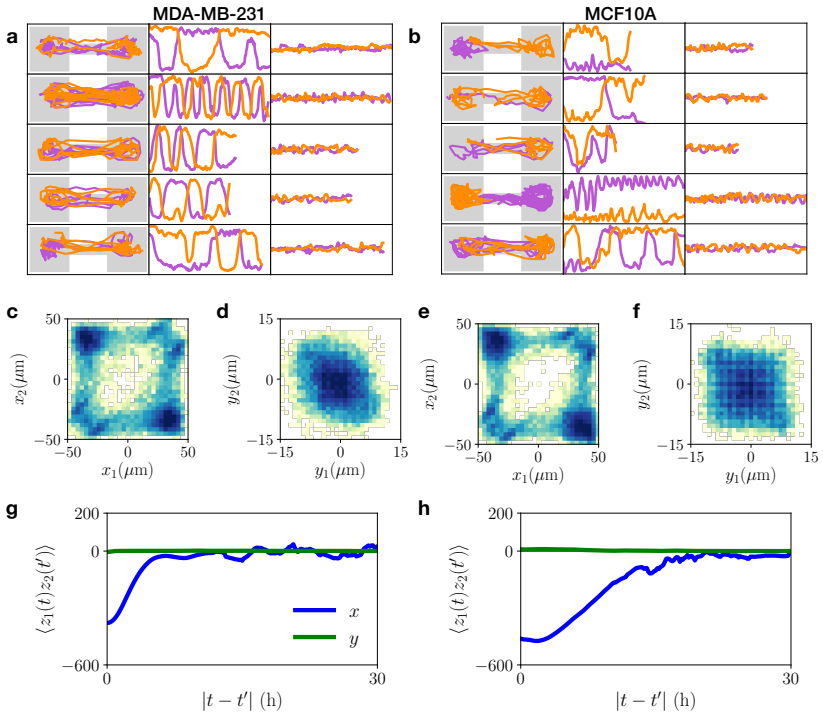


Figure S2: **Motion in 2D. a,b.** Several examples of 2D trajectories. Left: xy-trajectory plotted on top of the micropattern dimension (shown in grey). Axis limits are -50 µm $< x <$ 50 µm and -20 µm $< y <$ 20 µm; $(x = 0, y = 0)$ corresponds to the center of the bridge. Middle: x-trajectory as a function of time t. Axis limits are -50 µm $< x <$ 50 µm and $0 < t <$ 30 h. Right: y-trajectory as a function of time t. Axis limits are -50 µm $< y <$ 50 µm and $0 < t <$ 30 h, to allow direct comparison with the x-trajectory. **c,e.** Joint probability distributions $p(x_1, x_2)$ of the x-positions, plotted logarithmically on the same colour scale as in Fig. 2 in the main text. Here shown without the Gaussian interpolation employed in Fig. 2 in the main text. **d,f.** Joint probability distributions $p(y_1, y_2)$ of the y-positions, plotted logarithmically on the same colour scale as in Fig. 2 in the main text. **g,h.** Position cross-correlation function for x (blue) and y-components (green). For all panels, the left hand side corresponds to MDA-MB-231 cells, while the right hand side corresponds to MCF10A cells.

3.4 Collision events

To investigate the outcomes of cellular collisions, we identified collision events and defined criteria to classify them as reversal, sliding, and following events. Similar classifications have been developed in the literature before [**?**,5–7]. Here, we define a collision as an event where the two cell nuclei move to within a threshold distance Δx_c of each other (Fig. S3a). To classify reversals and sliding events, we are interested in whether the cells have moved past each other as a response to the collision event. Thus, we identify collisions as reversals if the cells do not switch positions shortly after the collision,

and as sliding events if they do switch positions. In practice, we find that we can easily identify these three events by inspecting the trajectories within a fixed time window dT following the collision. Thus we identify reversals if cells do not switch positions after a time interval dT, while for sliding events cells switch positions at least once within dT after the collision. To avoid artefacts where cells enter and exit the threshold distance Δx_c repeatedly in a short time-period, we only identify the first collision event in a time-interval dT. Thus, we assume that the time-scale between subsequent transitions is $\gg dT$. Indeed, we find that subsequent traversals of the threshold distance Δx_c usually do not correspond to new collisions, but occur while the cells still remain in contact. For following events, we are interested in identifying head-tail collisions which result in adhesion and subsequent joint migration [8]. In practice, we find that these events are robustly selected by identifying events where two cells perform the same transition across the micropatterned bridge within a time interval dT.

We choose the threshold distance $\Delta x_c = 20$ µm throughout, i.e. slightly larger than a typical cell radius when the cell occupies the island. For the time interval, we choose $dT = 6\Delta t = 1$ h throughout. This is a reasonable choice as this interval needs to be sufficiently larger than our time resolution $\Delta t = 10$ min, but smaller than the shortest switching time-scale, i.e. $\langle \tau_{\text{same}} \rangle \approx 1.7$ h for the MDA-MB-231 cells. By visual inspection of a large number of trajectories, we find that these values for Δx_c and dT robustly identify reversal, sliding and following events for both cell types (Fig. S3b-f). Furthermore, we find that the measured percentages of events are robust within a wide range of values around those chosen for our analysis (Fig. S3c-f). We achieve this level of robustness due to the clear separation of time-scales between dT and the transition time, and due to our criterion that no two collisions can occur within an interval dT.

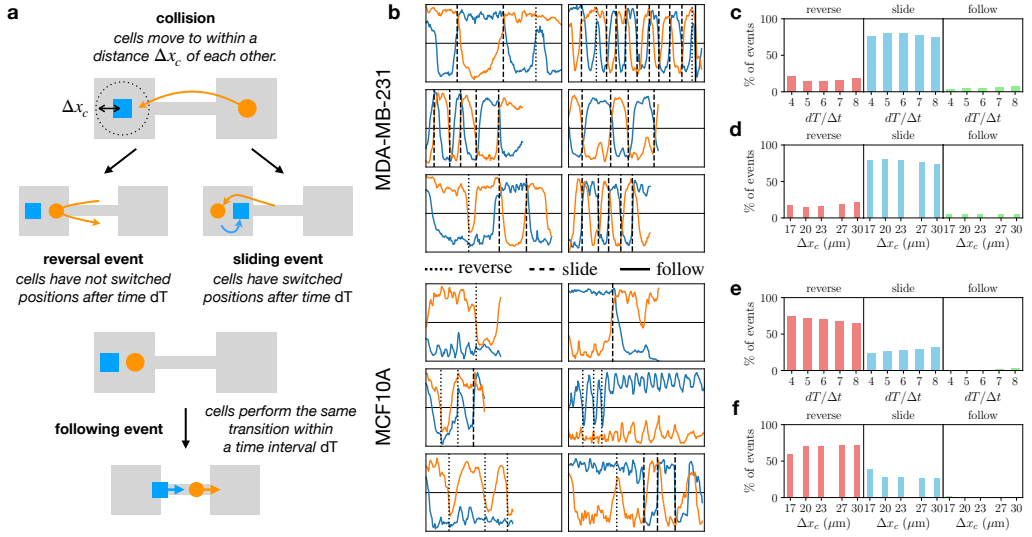


Figure S3: **Definition and identification of collision events. a.** Sketches of the definitions of reversal, sliding and following events. The blue square and orange circle show the nuclear positions of the two cells. **b.** Example trajectories with collision events shown by vertical lines (dotted: reversal, dashed: sliding, solid: following). Top 6 trajectories correspond to MDA-MB-231 cells, bottom 6 trajectories to MCF10A cells. Each trajectory plot shows the x-position of the cell nucleus on the y-axis, with axis limits $-50\ \mu\text{m} < x < 50\ \mu\text{m}$; the x-axis shows the time in the interval $0 < t < 33$ h. Following events occur rarely in our data set; an example can be seen in the top right trajectory. **c,e.** Percentages of reversal, sliding, and following events obtained for different values of the threshold dT, using the standard value $\Delta x_c = 20\ \mu\text{m}$, for MDA-MB-231 and MCF10A cells, respectively. **d,f.** Percentages of reversal, sliding, and following events obtained for different values of the threshold Δx_c, using the standard value $dT = 6\Delta t = 1$ h, for MDA-MB-231 and MCF10A cells, respectively.

3.5 Contact acceleration maps

Construction of CAMs – To obtain insight into the detailed interaction dynamics of the cell pairs, we measure 'contact acceleration maps' (CAMs), which give the interaction component of the acceleration of the cells as a function of the separation $\Delta x_{ij} = x_i - x_j$ of cells i and j, and their relative velocity $\Delta v_{ij} = v_i - v_j$. To measure CAMs, we assume that the dynamics can be split into separate single-cell and interaction parts, resulting in the equation of motion for cell i:

$$\begin{aligned} \dot{x}_i &= v_i \\ \dot{v}_i &= F(x_i, v_i) + G(\Delta x_{ij}, \Delta v_{ij}) + \sigma\eta_i(t) \end{aligned} \tag{S4}$$

where $\langle \eta_i(t) \rangle = 0$, and $\langle \eta_i(t)\eta_j(t') \rangle = \delta_{ij}\delta(t - t')$. As show by the rigorously inferred model using Underdamped Langevin Inference (ULI), the dynamics can indeed be described by such an equation of motion.

For a system of the form given by Eq. (S4), a simple way of obtaining an estimate of the interacting component $G(\Delta x, \Delta v)$ is through a conditional averaging procedure. Specifically, for interactions which decay beyond a threshold distance ℓ, and the one-body term $F(x, v)$ can be estimated as

$$F(x, v) \approx \langle \dot{v}_i \big| x_i, v_i; |\Delta x_{ij}| > \ell \rangle \tag{S5}$$

Next, we estimate the interaction term by calculating

$$G(\Delta x, \Delta v) \approx \langle \dot{v}_i - F(x_i, v_i) \big| \Delta x_{ij}, \Delta v_{ij}; |\Delta x_{ij}| < \ell \rangle \tag{S6}$$

Note that more general binning approaches, e.g. conditioning on (x_1, x_2) or (v_1, v_2), are not able to recover simple dependencies on Δx or Δv, as non-trivial correlations exist in these phase spaces. As we show in the next paragraph, this construction yields a good estimate for the interactions in the type of system we are considering here. To give this approach further credence, we show in section 4.3.2 that our model inferred from ULI, which does not rely on the assumption of short-ranged interactions, captures the experimentally measured CAMs well.

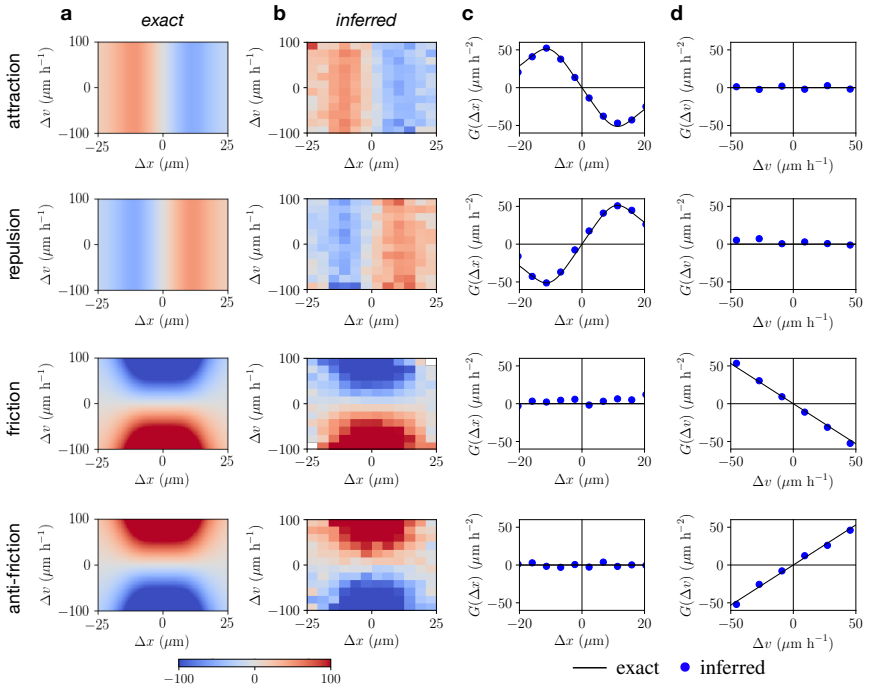


Figure S4: **Testing CAM inference.** For all cases, we used 100 trajectories with 200 time-points each, sampled at a time-interval $\Delta t = 10$ min, which is comparable to the experimental case. We use a cutoff of $\ell = 25$ μm. **a.** Exact interaction terms $G(\Delta x, \Delta v) = f_0 g(|\Delta x|)\Delta x + \gamma_0 g(|\Delta x|)\Delta v$ in units of μm h^{-2}. For attraction/repulsion, we take $f_0 = \pm 6$ h^{-2}, $\gamma_0 = 0$. For friction/anti-friction, we take $f_0 = 0, \gamma_0 = \pm 2$ h^{-1}. **b.** Inferred CAM. **c.** Inferred (blue dots) and exact (black line) Δx-dependence of the interaction. **d.** Inferred (blue dots) and exact (black line) Δv-dependence of the interaction.

Testing on simulated data − First, we test whether CAMs yield a good indicator for the form of interaction terms governing the type of system we are considering here. Specifically, we estimate CAMs from simulated trajectories with similar length and time resolution as in the experiment, obtained by simulating the inferred single-cell term with known, artificial interactions. To obtain a simulated data set with known interactions, which resembles the experimental data set, we use the single-cell term $F_{231}(x_i, v_i)$ and noise amplitude σ_{231} inferred from MDA-MB-231 experiments, with additional cohesion or frictional interactions. Thus, we simulate the model

$$\dot{v}_i = F_{231}(x_i, v_i) + f_0 g(|\Delta x_{ij}|)\Delta x_{ij} + \gamma_0 g(|\Delta x_{ij}|)\Delta v_{ij} + \sigma_{231}\eta_i(t) \tag{S7}$$

with $g(|\Delta x_{ij}|) = 1/[(|\Delta x_{ij}|/R_0)^4 + 1]$, $R_0 = 15$ µm, and we vary the coefficients f_0 and γ_0. Independently of the precise choice of parameters or functional forms, we find that the inferred CAMs recover the known interactions (Fig. S4).

Application to experimental data − We calculate CAMs from the experimental data using a threshold $\ell = 25$ µm. We obtain a confinement term that is qualitatively very similar to that obtained from single cell experiments (Fig. S5) [1]. Next, we obtain the CAMs as shown in Fig. S5, and in Fig. 3 of the main text. The resulting maps are robust with respect to changing the threshold ℓ within a reasonable range. To obtain insight into the type of interactions implied by the measured CAMs, we first compare them to CAMs plotted for simple analytical interactions (Fig. S4). By inspection, we find that the CAM for MDA-MB-231 cells looks very similar to an anti-friction interaction, while that for MCF10A cells looks similar to a repulsive interaction (compare Figs. S5b, S4a). By performing averaging conditioned on only one variable, we show that the Δx and Δv dependencies of the contact accelerations confirm this conclusion (Fig. 3b,e in the main text). In addition, we also find that MDA-MB-231 cells exhibit an attractive component, and MCF10A cells exhibit an additional weak friction component (Fig. 3c,f in the main text). We use these results to guide our rigorous inference approach (section 4.1).

To further test these findings without relying on conditional averaging of the CAMs, we calculate the first moments of the contact accelerations with respect to Δx and Δv. For the cohesive component, we calculate:

$$p_{\Delta x} = \langle [\dot{v}_i - F(x_i, v_i)]\Delta x_{ij} \big| |\Delta x_{ij}| < \ell \rangle \tag{S8}$$

For the case of simple monotonic interactions, the sign of this quantity indicates the type of interaction observed: we expect $p_{\Delta x} < 0$ for simple monotonic attractive interactions, and $p_{\Delta x} > 0$ for repulsive interactions. Thus, this quantity summarizes the sign of the interaction. For the friction component, we calculate:

$$p_{\Delta v} = \langle [\dot{v}_i - F(x_i, v_i)]\Delta v_{ij} \big| |\Delta x_{ij}| < \ell \rangle \tag{S9}$$

which gives $p_{\Delta v} < 0$ for friction interactions, and $p_{\Delta v} > 0$ for anti-friction interactions. Indeed, we find that the dipole moments further support our conclusions based on the CAMs: MDA-MB-231 cells exhibit contact accelerations indicating attraction and anti-friction, while MCF10A cells exhibit repulsion and weak friction.

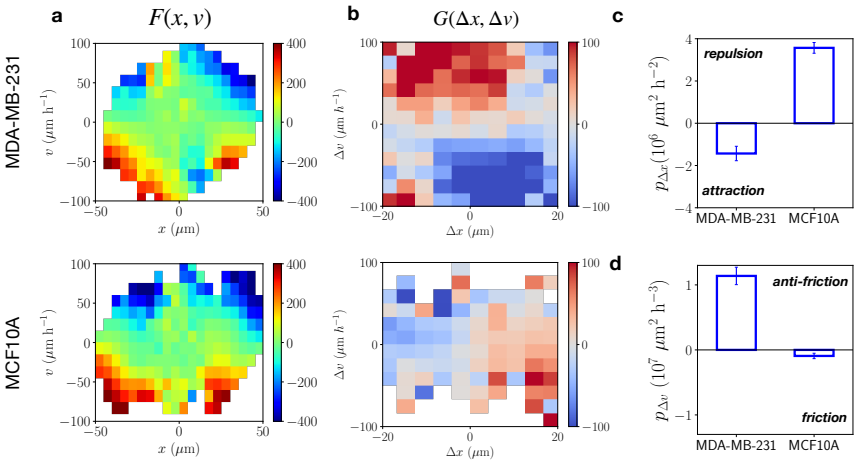


Figure S5: **Calculation of contact acceleration maps and dipole moments. a.** Estimated deterministic single-cell contribution to the dynamics $F(x, v)$. **b.** Contact acceleration map $G(\Delta x, \Delta v)$. In the main text, we show an 8×8 grid, in the range $-45 < \Delta v < 45$ for MCF10A and $-70 < \Delta v < 70$ for MDA-MB-231. For comparison, we here show 10×10 grids in the range $-100 < \Delta v < 100$, which show the same qualitative features. **c.** First moment of the contact accelerations with respect to Δx, $p_{\Delta x}$. **d.** First moment of the contact accelerations with respect to Δv, $p_{\Delta v}$. Error bars in **c**, **d** correspond to bootstrap errors. *Top row:* MDA-MB-231 cells. *Bottom row:* MCF10A cells.

4 Inference method and model selection

4.1 Application of Underdamped Langevin Inference

To infer an interacting stochastic equation of motion for confined migrating cell pairs, we employ a rigorous inference method, Underdamped Langevin Inference (ULI) [9]. In this section, we lay out the details of applying ULI to our system. For further details on the method itself, see ref. [9]. Our inference ansatz is to postulate that the system can be described by the general equation of motion for cell i with position $x_i(t)$, velocity $v_i(t) = \mathrm{d}x_i/\mathrm{d}t$, and acceleration $\dot{v}_i(t) = \mathrm{d}v_i/\mathrm{d}t$:

$$\begin{aligned} \dot{x}_i &= v_i \\ \dot{v}_i &= F(x_i, v_i) + f(|\Delta x_{ij}|)\Delta x_{ij} + \gamma(|\Delta x_{ij}|)\Delta v_{ij} + \sigma\eta_i(t) \end{aligned} \tag{S10}$$

where $\Delta x_{ij} = x_i - x_j$, $\Delta v_{ij} = v_i - v_j$, $\langle \eta_i(t) \rangle = 0$, and $\langle \eta_i(t)\eta_j(t') \rangle = \delta_{ij}\delta(t - t')$.

Using ULI, the stochastic equation of motion of such an interacting system can be reconstructed by projecting the dynamics onto a set of n_b basis functions $\{b_\alpha(x, v)\}_{\alpha=1\ldots n_b}$, which are subjected to an orthonormalization scheme $\hat{c}_\alpha(x, v) = \sum_{\beta=1}^{n_b} \hat{B}^{-1/2}_{\alpha\beta} b_\beta(x, v)$ such that $\langle \hat{c}_\alpha \hat{c}_\beta \rangle = \delta_{\alpha\beta}$. The total deterministic contribution $F^{(\mathrm{total})} = F(x_i, v_i) + f(|\Delta x_{ij}|)\Delta x_{ij} +$

$\gamma(|\Delta x_{ij}|)\Delta v_{ij}$ of the system can then be approximated as a linear combination of these basis functions, $F^{(\text{total})} \approx \sum_{\alpha=1}^{n_b} F_\alpha^{(\text{total})} \hat{c}_\alpha(x,v)$. Using ULI, we estimate the coefficients of this expansion of the deterministic term $\hat{F}_\alpha^{(\text{total})}$ and the noise amplitude $\hat{\sigma}$ using the increments of the observed position trajectories $x_i(t)$.

For interacting systems, we separate single-particle and interaction contributions into separate sets of basis functions. We approximate the cohesion and friction terms $f(|\Delta x_{ij}|)$ and $\gamma(|\Delta x_{ij}|)$ using a set of interaction kernels $\{u_\alpha(|\Delta x_{ij}|)\}$ (see section 4.2). We fit the single-cell term $F(x_i, v_i)$ with a basis consisting of Fourier components in x_i and polynomials in v_i including terms up to third order [1]:

$$F(x_i, v_i) \approx \sum_{n=0}^{N} \sum_{m=0}^{M} [A_{nm} \cos(2\pi n x_i / w) + B_{nm} \sin(2\pi n x_i / w)] v_i^m \tag{S11}$$

where $N = M = 3$ and $w = 100$ µm. As we show in section 4.2, our inference results are not sensitive to the precise choise of basis employed.

A key assumption of our model (Eq. (S10)) is that the noise $\eta_i(t)$ is uncorrelated in time. To self-consistently test this assumption, we calculate the trajectories of the noise increments $\Delta W_i(t) = \int_t^{t+\Delta t} \eta_i(s)\, \mathrm{d}s$. An empirical estimator for $\Delta W_i(t)$ is [1, 10, 11]:

$$\widehat{\Delta W}_i(t) \approx \frac{\Delta t}{\hat{\sigma}} \left[\dot{v}_i(t) - \hat{F}^{(\text{total})}(x_i, v_i) \right] \tag{S12}$$

Thus, we calculate the auto-correlation function of the noise as $\hat{\phi}_{\Delta W} = \langle \widehat{\Delta W}_i(t) \widehat{\Delta W}_i(t') \rangle$. We find that for both cell lines, the noise decays to zero within a single time-step, confirming our white noise assumption. The weak negative correlation at $|t - t'| = \Delta t$ is due to the presence of measurement errors in the positions, as discussed in refs. [1, 12].

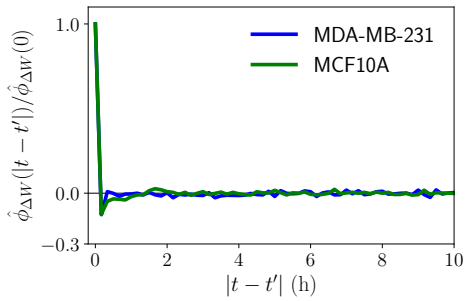


Figure S6: **Inferred noise correlation functions.** The correlation functions are normalized by their value at $|t - t'| = 0$. The blue curve corresponds to MDA-MB-231 cells, the green curve to MCF10A cells.

Three conditions for accurate inference from stochastic underdamped systems are (i) sufficiently long trajectories to constrain the n_b parameters of the fitted model, (ii) a

sufficiently small measurement time interval Δt to resolve the dynamics and (iii) measurement errors on the positions that are smaller than the typical displacement in a single time-step:

(i) Trajectory length – Inference from a finite data set relies on having sufficient information to accurately resolve the features of the underlying dynamical terms of the equation of motion. The information contained in a set of trajectories for a system of the type of Eq. (S10) can be empirically estimated as $\hat{I}_b = \frac{\tau}{2\hat{\sigma}^2} \sum_{\alpha=1}^{n_b} \left(\hat{F}_\alpha^{(\text{total})} \right)^2$, where τ is the total length of the trajectory. The parameters of the expansion can be accurately inferred if $\hat{I}_b \gg n_b$, where $\hat{I}_b$ is given in natural information units (1 nat $= 1/\log 2$ bits) [9]. Here, we employ a basis with $n_b = 34$ parameters (28 parameters for the single cell term and 6 parameters for the interaction kernels). As shown in table 1, our data sets contain enough information to constrain these parameters.

(ii) Discretization – To ensure a sufficiently accurate temporal sampling of the observed signal, we ensured that the measurement time interval Δt should be small enough to resolve the time-scales of the collision dynamics, i.e. the switching time $\langle \tau_{\text{same}} \rangle = (1.69 \pm 0.11)$ h of MDA-MB-231 cells. Our measurement time interval is $\Delta t = 10$ min, and thus sufficiently small to resolve this time-scale. Additionally, the time interval plays an important role as velocities and accelerations are obtained as discrete derivatives from the position trajectories $x_i(t)$. Indeed, even for small Δt, inference from underdamped systems exhibits systematic discretization biases [9, 12, 13]. The leading order term of the bias is removed through the construction of the ULI estimators [9]. We show empirically that higher order biases do not strongly affect our inference results by performing a self-consistency test (see section 4.3.1).

(iii) Measurement error – In any tracking experiment, the observed position trajectories are subject to time-uncorrelated measurement error $m(t)$, such that the observed signal is $y(t) = x(t) + m(t)$, where $x(t)$ is the true signal. ULI yields accurate inference results in the regime $|m| < v\Delta t$, where $v\Delta t$ is the typical displacement in a single time-step. We can evaluate this condition from the data, using the average speed of the cells, and comparing it to the measurement error amplitude inferred from the trajectories [9]. As shown in table 1, this condition is fulfilled for both data sets.

4.2 Robustness with respect to the projection basis

To infer the interaction terms of the dynamics, we approximate the cohesion and friction terms $f(|\Delta x_{ij}|)$ and $\gamma(|\Delta x_{ij}|)$ using a set of interaction kernels $\{u_\alpha(|\Delta x_{ij}|)\}$. Physically, we expect cell-cell interactions to be spatially local. Thus, to ensure accurate inference in the region of interest, we choose kernels which decay at large distances, $u_\alpha(|\Delta x_{ij}| \to \infty) \to 0$. A simple choice for such kernels is a set of exponentials $u_n(|\Delta x_{ij}|) = \exp(-|\Delta x_{ij}|/nr_0)$ with $n = 1...N$. This basis therefore has two hyperparameters that have to be chosen, the number for kernels N and the maximum decay length $r_{\max} = Nr_0$. Alternatively, we also test a basis consisting of Gaussian functions $u_n(|\Delta x_{ij}|) =$

	N_{pairs}	N_{pts}	$\hat{I}_b$ (nats)	$\hat{\sigma}$ ($\mu\text{m}\,\text{h}^{3/2}$)	$\langle\lvert\hat{v}\rvert\rangle\Delta t$ (µm)	$\lvert\hat{m}\rvert$ (µm)
MDA-MB-231	90	15,979	$11,800 \pm 150$	51.4	2.6	1.3
MCF10A	155	19,470	$11,900 \pm 160$	47.9	2.4	1.4

Table 1: **Statistics of the stochastic trajectory data sets for both cell lines**. From left to right, the columns denote: (i) The number of tracked cell pairs. (ii) The total number of recorded time-points. (iii) The empirical estimate of the information content of the data set, obtained by projecting the dynamics onto our standard basis choice (see section 4.2). The error in the inferred information content is estimated as $\delta\hat{I}_b \approx [2\hat{I}_b + n_b^2/4]^{1/2}$ [9]. (iv) The inferred noise amplitude. (v) The typical displacement in a single time-step. (vi) The inferred amplitude of the measurement error, which is in line with previous estimates for single cell migration in the same setup [1].

$\exp(-(|\Delta x_{ij}| - nr_0)^2/2W^2)$ with $n = 1...N$. This basis therefore has three hyperparameters $N, r_{\text{max}} = Nr_0$, and W. While this inference scheme could be supplemented by an additional optimization of the hyperparameters, we find this not to be necessary in this case, as the inferred interaction terms do not sensitively depend on the choice of hyperparameters or basis functions (Figs. S7,S8). Furthermore, the predictive power of the inferred model is not sensitively affected by the choice of basis (Fig. S9). Throughout the main text, we choose an exponential basis with an intermediate value of $N = 3$ functions and a maximum decay length $r_{\text{max}} = 20$ µm (black line in Figs. Figs. S7,S8).

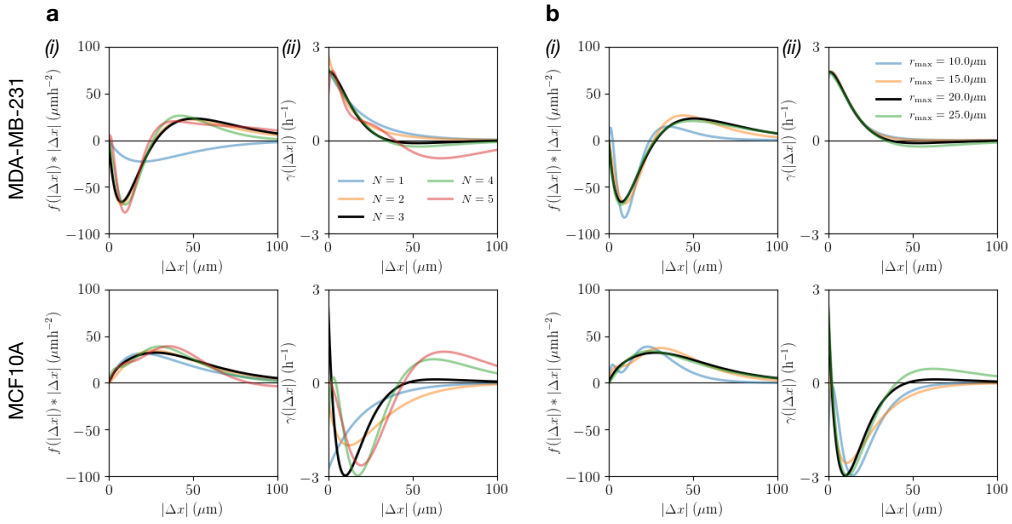


Figure S7: **Inference results for exponential interaction kernels. a**, Varying the number of kernels N, using $r_{\text{max}} = 20$ µm **b**, Varying the maximum decay length r_{max}, using $N = 3$. *(i)*, Cohesive component $f(|\Delta x_{ij}|)|\Delta x_{ij}|$. *(ii)*, Friction kernel $\gamma(|\Delta x_{ij}|)$. *Top row:* MDA-MB-231 cells. *Bottom row:* MCF10A cells. Black line corresponds to the curves shown in Fig. 4 of the main text, using $N = 3$ and $r_{\text{max}} = 20$ µm.

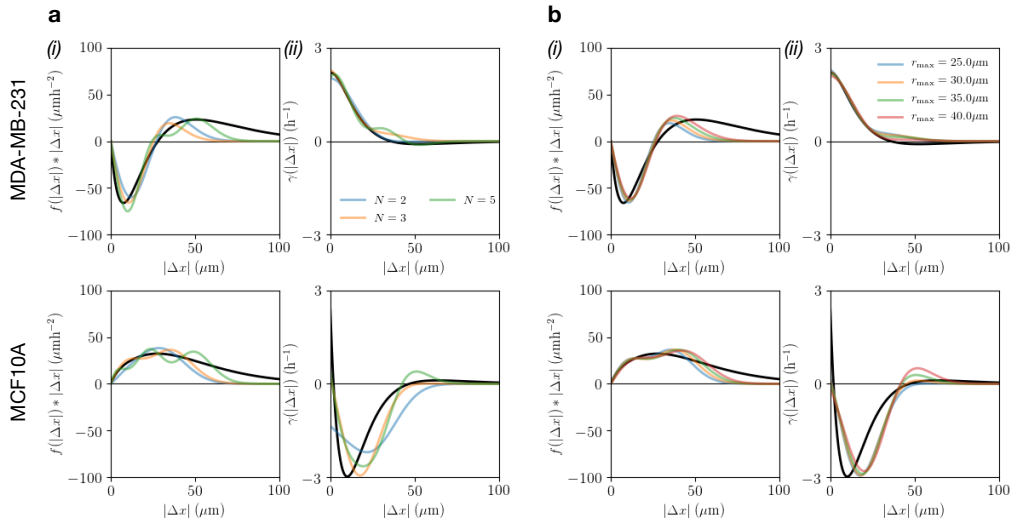


Figure S8: **Inference results for Gaussian interaction kernels. a**, Varying the number of kernels N, using $r_{\max} = 25$ µm and $W = 4$ µm. **b**, Varying the maximum decay length $r_{\max}$, using $N = 3$ and $W = 4$ µm. *(i)*, Cohesive component $f(|\Delta x_{ij}|)|\Delta x_{ij}|$. *(ii)*, Friction kernel $\gamma(|\Delta x_{ij}|)$. *Top row:* MDA-MB-231 cells. *Bottom row:* MCF10A cells. Black line corresponds to the curves shown in Fig. 4 of the main text, using an exponential basis with $N = 3$ and $r_{\max} = 20$ µm.

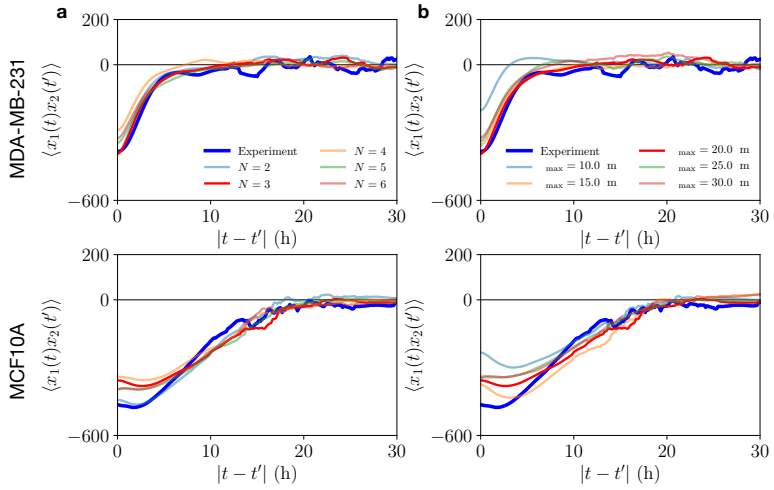


Figure S9: **Predicted position cross-correlation functions for various exponential bases. a**, Varying the number of kernels N. **b**, Varying the maximum decay length $r_{\max}$. *Top row:* MDA-MB-231 cells. *Bottom row:* MCF10A cells.

4.3 Simulations of the inferred model

An important step in performing inference from data is to test the inferred model by simulating stochastic trajectories based on the inferred model terms, and to compare their statistical properties to those observed experimentally. We simulate the dynamics using Verlet integration with a small time step $\mathrm{d}t$. To compare the statistics of these simulated trajectories to those observed experimentally, we sample the simulated trajectories with the same discrete time interval as in experiments, $\Delta t = 10\ \mathrm{min} \gg \mathrm{d}t$.

4.3.1 Self-consistency test

First, we determine whether the inferred model is self-consistent: for a self-consistent inference, re-inferring a model from simulated trajectories should yield the same model. Here, we use the same number of simulated trajectories as experimentally observed trajectories, with a similar trajectory length, and the same sampling interval Δt as in the experiment. We apply this test to the inferred models for MDA-MB-231 and MCF10A cells, and find that the re-inferred model accurately matches the original inferred model (Fig. S10), showing that our inference approach is numerically stable.

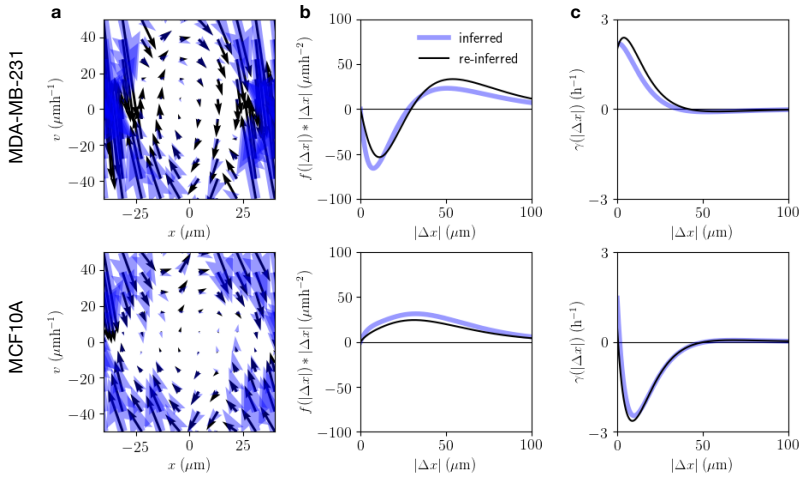


Figure S10: **Stability test of the inferred model.** **a**, Flow field of the confinement term $F(x_i, v_i)$. Blue arrows: inferred from experimental data, black arrows: re-inferred from simulated trajectories. **b**, Cohesive component $f(|\Delta x_{ij}|)|\Delta x_{ij}|$. **c**, Friction kernel $\gamma(|\Delta x_{ij}|)$. *Top row:* MDA-MB-231 cells. *Bottom row:* MCF10A cells.

4.3.2 Testing the predictive power of the model

To test the predictive power of the model, we apply the same analysis routines that were applied to the experimental data to our simulated data (results shown in Fig.

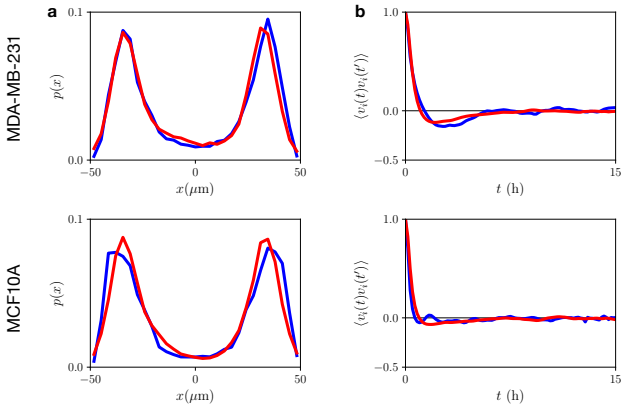


Figure S11: **Experimental and predicted dynamics of the inferred model (Eq. 1 in the main text). a**, Probability distribution of all cell positions $p(x)$ (experiment shown in blue, model predictions in red). **b**, Normalized velocity auto-correlation function $\langle v_i(t)v_i(t')\rangle$. *Top row:* MDA-MB-231 cells. *Bottom row:* MCF10A cells.

2 of the main text). The inferred model is fully constrained by the short time-scale accelerations of the dynamics. Thus, comparing the predicted long time-scale features such as correlation functions to the experimental data provides an independent test of the model. In addition to the statistics shown in the main text, here we show several additional statistics to test experiment-model agreement. To test how accurately the model captures the dynamics at the single-cell level, we plot the distribution of all cell positions $p(x)$, and the velocity auto-correlation function $\langle v_i(t)v_i(t')\rangle$. As shown in Fig. S11a,b, these statistics are well captured by the model.

In our model, we assume that the cell-cell interactions separate into a cohesive contribution $f(|\Delta x|)\Delta x$ and an effective linear friction $\gamma(|\Delta x|)\Delta v$. This choice is motivated by the observation that the Δv-dependent component of the contact acceleration maps is linear (Fig. 3c,f in the main text). We find that the contact acceleration maps predicted by the model are qualitatively very similar to those inferred from experiments (Fig. S12), indicating that this assumption is valid.

Next, we show side-by-side comparisons of the full joint probability distribution of positions $p(x_1, x_2)$ and velocities $p(v_1, v_2)$ (Fig. S13). The experimental distributions $p(x_1, x_2)$ exhibit several features (Fig. S13a). First, there is a clear minimum around $(0, 0)$, corresponding to both cells occupying the connecting bridge. Second, we find peaks where each cell occupies one island, and fainter peaks where both cells occupy the same island. This reflects the mutual exclusion behavior exhibited by these cells. These peaks are connected by horizontal and vertical 'paths', indicating that during transitions, typically, only one cell performs a transition at a time. Finally, we find that the peaks corresponding to both cells occupying the same island are 'split', and exhibit two distinct close-by maxima. Our model captures almost all of these features,

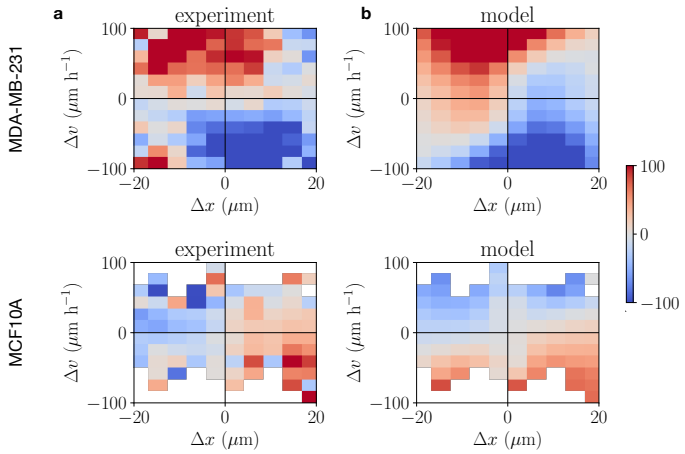


Figure S12: **Experimental and predicted contact acceleration maps. a**, Experimentally measured contact acceleration map. **b**, Contact acceleration map measured from simulation data, plotted for the same region of phase space sampled in the experiment. *Top row:* MDA-MB-231 cells. *Bottom row:* MCF10A cells.

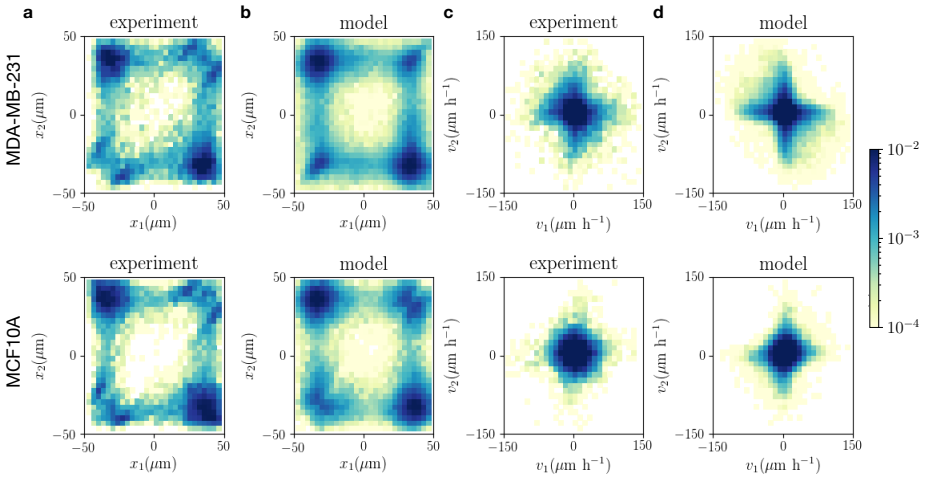


Figure S13: **Experimental and predicted joint probability distributions. a,b.** Experimental and predicted joint probability distribution of positions $p(x_1, x_2)$. **c,d.** Experimental and predicted joint probability distribution of velocities $p(v_1, v_2)$. *Top row:* MDA-MB-231 cells. *Bottom row:* MCF10A cells.

including the relative occupation of the same- and opposite-side configuration, and the path-structure of the map (Fig. S13b). However, the model does not exhibit the same splitting of the same-side probabilities, which may be due to movement in the second dimension (the short axis of the micropattern, y), which is not captured by the model. Our model further captures the structure of the velocity distributions $p(v_1, v_2)$ (Fig. S13c,d).

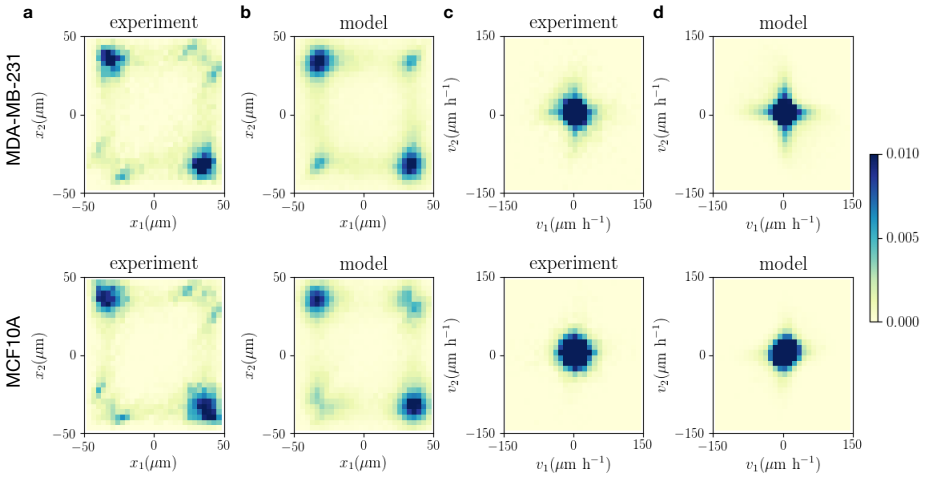


Figure S14: **Experimental and predicted joint probability distributions, plotted on a linear scale.** Same panels as shown in Fig. S13, but with a linear colour scale.

4.3.3 Ruling out simpler models

We arrived at our model (Eq. (S10)) by first excluding simpler alternatives. First, we consider the non-interacting case, consisting only of the single-cell term:

$$\dot{v}_i = F(x_i, v_i) + \sigma\eta_i(t) \tag{S13}$$

As expected, such a model is unable to capture the correlations in the system, and can therefore be ruled out (Fig. S15). This model is still able to capture the distinct minimum in the joint probability density around $(x_1, x_2) = (0, 0)$, suggesting that this feature is due to the single-cell term: due to the confinement very little occupancy is expected near the center of the connecting bridge.

Next, we consider a model including only a cohesive term:

$$\dot{v}_i = F(x_i, v_i) + f(|\Delta x_{ij}|)\Delta x_{ij} + \sigma\eta_i(t) \tag{S14}$$

While this model can approximately capture the dynamics of MCF10A cells, except for the velocity cross-correlation function, it completely fails to describe the MDA-MB-231 statistics (Fig. S16). In fact it predicts that cells are more likely to occupy the same-side configuration, in qualitative disagreement with our experimental observations, likely due to the attractive nature of the cohesive interaction in MDA-MB-231 cells.

Finally, we consider a model including frictional interactions, but no cohesion:

$$\dot{v}_i = F(x_i, v_i) + \gamma(|\Delta x_{ij}|)\Delta v_{ij} + \sigma\eta_i(t) \tag{S15}$$

This model qualitatively fails to account for the MCF10A statistics (Fig. S17): it predicts that cells are more likely to occupy the same-side configuration, likely due to the regular

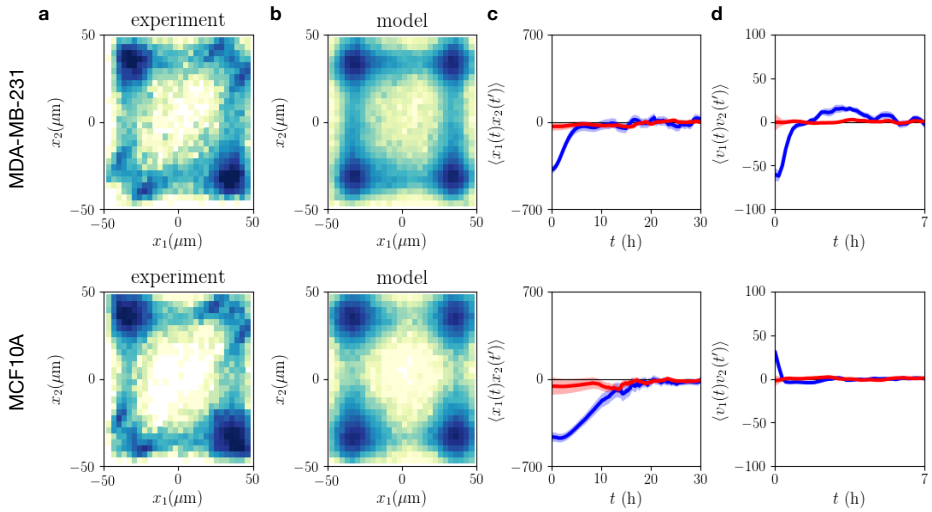


Figure S15: **Experimental and predicted dynamics for an inferred model without interactions. a.** Experimental joint probability distribution $p(x_1, x_2)$. The colour bar corresponds to that shown in Fig. S13. **b.** Model prediction of the joint probability distribution $p(x_1, x_2)$. **c.** Position cross-correlation functions for the experiment (blue) and model prediction (red). **d.** Velocity cross-correlation functions for same-side configurations. *Top row:* MDA-MB-231 cells. *Bottom row:* MCF10A cells.

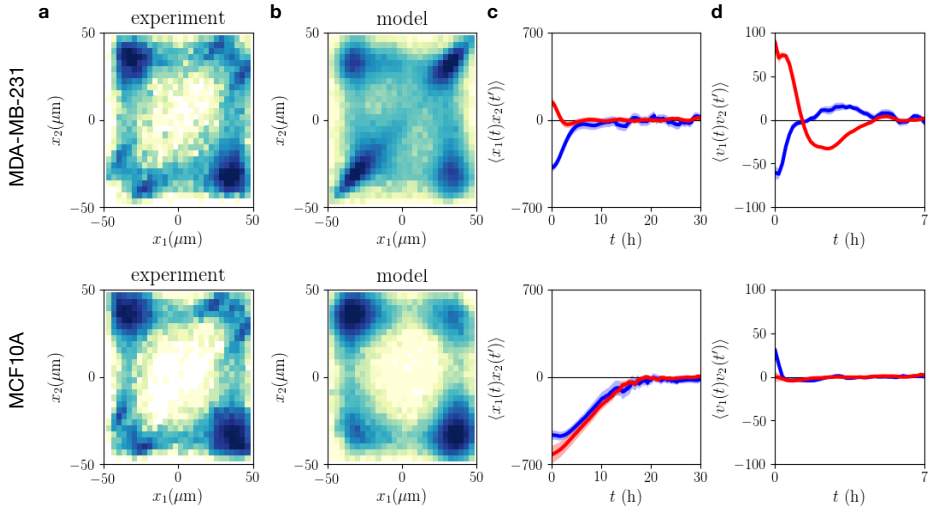


Figure S16: **Experimental and predicted dynamics for an inferred model with only cohesive, but no friction interactions.** See Fig. S15 for captions.

friction between MCF10A cells, which acts to slow cells down when they are close to each other.

In conclusion, we find that the simplest model within the class of models considered here, which can accurately capture the statistics of both MCF10A and MDA-MB-231 cell pairs, requires both cohesive and friction interactions.

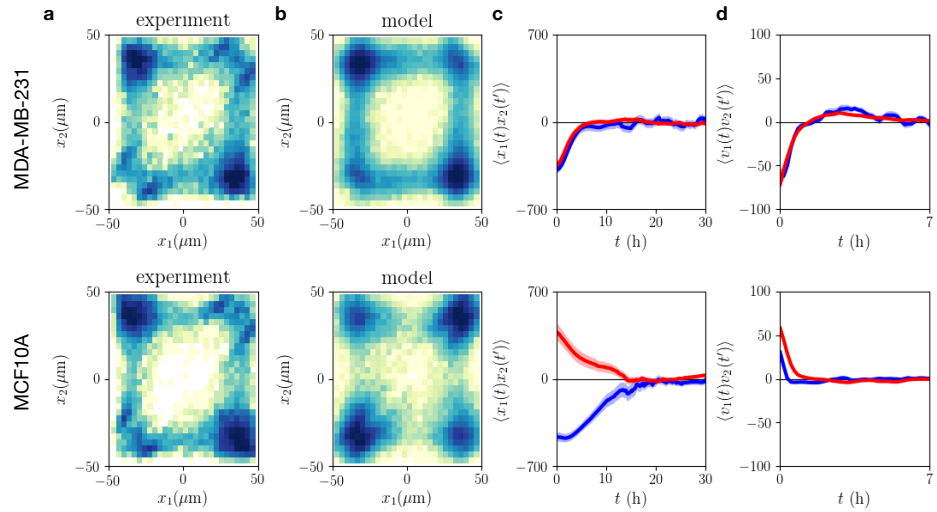


Figure S17: **Experimental and predicted dynamics for an inferred model with only friction, but no cohesive interactions.** See Fig. S15 for captions.

4.4 Separation of single-cell and interaction terms

Here, we directly compare the single-cell term inferred from experiments with interacting cell pairs ($F(x_i, v_i)$ in Eq. (S10)) to the deterministic term inferred from experiments in which only a single cell occupies the pattern [1], denoted $F_{\mathrm{sc}}(x, v)$. In Fig. S18, the terms are compared side by side. Furthermore, we show the deterministic flow field $(\dot{x}, \dot{v}) = (v, F(x, v))$ superimposed for both experiments. These results indicate a remarkable similarity of the inferred terms, indicating that the contributions of single-cell dynamics (corresponding to the internal motility of the cell and its interaction with the local micro-environment placed by the micropattern) are not strongly affected by the presence of another cell.

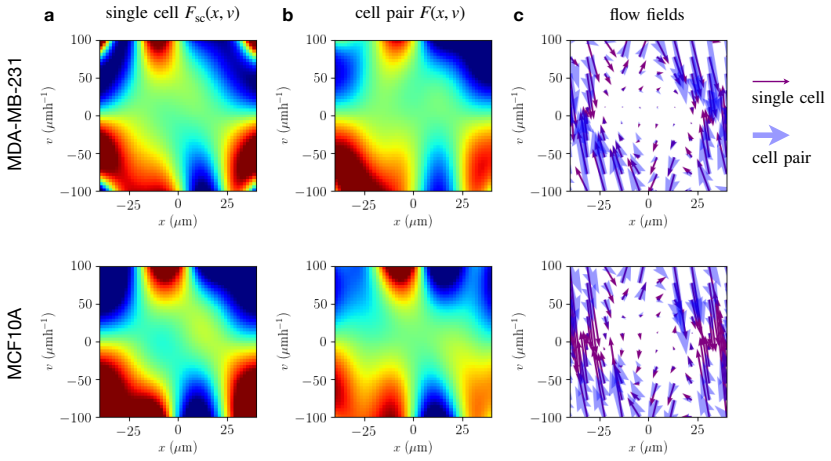


Figure S18: **Disentangling single-cell and interaction contributions. a**, Deterministic term $F_{\mathrm{sc}}(x, v)$ inferred from experiments with single cells confined to two-state micropatterns [1], obtained by applying ULI with the same basis expansion as used for cell pair experiments (Eq. (S11)), without interaction terms. Plotted with the same colour scale as in Fig. 4 in the main text. **b**, Single-cell term $F(x, v)$ inferred from cell pair experiments (as shown in Fig. 4 in the main text). **c**, Direct comparison of the flow fields of both terms. Fat blue arrows: inferred from cell pair data, thin darkviolet arrows: inferred from single-cell experiments. *Top row:* MDA-MB-231 cells. *Bottom row:* MCF10A cells.

4.5 Generalization of the inference approach to higher dimensions and heterotypic interactions

4.5.1 Inference in 2D and 3D multi-cellular systems

The cell-cell interaction inference procedure we have developed here can be readily generalized to 2D and 3D systems, including assemblies of many cells. Specifically, our model (Eq. (S10)) can be generalized to higher dimensions as follows:

$$\begin{aligned}\dot{\mathbf{r}}_i &= \mathbf{v}_i \\ \dot{\mathbf{v}}_i &= \mathbf{F}(\mathbf{x}_i, \mathbf{v}_i) + \sum_{j \neq i} \left[f(r_{ij})\mathbf{r}_{ij} + \gamma(r_{ij})\mathbf{v}_{ij} \right] + \sigma \boldsymbol{\eta}_i(t)\end{aligned} \tag{S16}$$

where $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$, $r_{ij} = |\mathbf{r}_{ij}|$, and $\mathbf{v}_{ij} = \mathbf{v}_i - \mathbf{v}_j$, and the sum goes over all particles $j = 1...N$. The Gaussian white noise in this case has the property $\langle \eta_{i,\mu}(t)\eta_{j,\nu}(t') \rangle = \delta_{\mu\nu}\delta_{ij}\delta(t-t')$, where $\{i, j\}$ are particle indices and $\{\mu, \nu\} = 1...d$, where d is the dimensionality of the system.

Here, the single-cell contribution $\mathbf{F}(\mathbf{x}_i, \mathbf{v}_i)$ reflects the properties of the environment in which cells migrate. In a spatially unstructured system, such as in the case of an epithelial monolayer, there will be no space-dependence, $\mathbf{F}(\mathbf{x}_i, \mathbf{v}_i) \equiv \mathbf{F}(\mathbf{v}_i)$. For example, for a simple persistent random motion model, the single-cell contribution would take the form $\mathbf{F}(\mathbf{v}_i) = -\tau_{\mathrm{p}}^{-1}\mathbf{v}_i$, where τ_{p} is the persistence time of the cell. For the interactions terms in Eq. (S16), we assume radially symmetric interactions, i.e. that f and γ only depend on r_{ij}, similar to the model we have inferred here. In this case, the interactions we have inferred here from confined 1D migration of cells could be directly generalized to a 2D or 3D scenario, by taking $f(r_{ij}) = f(|\Delta x_{ij}|)$ and $\gamma(r_{ij}) = \gamma(|\Delta x_{ij}|)$. The Underdamped Langevin Inference method we have used here for 1D inference can be directly generalized to 2D or 3D data by assuming that interactions are in the radial direction only. Note, however, that this method could also infer non-radial forces that depend also, for instance, on the velocity vector of each of the particles through alignment or avoidance torques. For a demonstration, see ref. [9].

4.5.2 Heterotypic interactions

Our model (Eq. (S10)) could furthermore be generalized to account for heterotypic interactions, where cells of different types interact with one another. In this case, we can no longer assume that the single-cell, interaction, and noise terms are identical for all cells, but they will depend on the cell type of each cell s_i:

$$\begin{aligned}\dot{\mathbf{r}}_i &= \mathbf{v}_i \\ \dot{\mathbf{v}}_i &= \mathbf{F}_{s_i}(\mathbf{x}_i, \mathbf{v}_i) + \sum_{j \neq i} \left[f_{s_i s_j}(r_{ij})\mathbf{r}_{ij} + \gamma_{s_i s_j}(r_{ij})\mathbf{v}_{ij} \right] + \sigma_{s_i} \boldsymbol{\eta}_i(t)\end{aligned} \tag{S17}$$

where for a system with N cells with n different cell types, $i, j = 1...N$, and $s_i \in \{1...n\}$. In this case, for each cell the single-cell and noise contribution is determined by its cell type: $\mathbf{F}_{s_i}$ and σ_{s_i}. The interaction components $f_{s_i s_j}$ and $\gamma_{s_i s_j}$ depend on the cell type of both cells in a pairwise interaction. In the example of two cell types interacting, there are thus four types of interactions: $f_{s_1 s_1}, f_{s_1 s_2}, f_{s_2 s_1}, f_{s_2 s_2}$, and similarly for γ. To infer these interactions, the same set of basis functions $\{c_\alpha(x, v)\}$ can be used for both cell types, but a different set of coefficients will be inferred for each type s_i, i.e. $F_{s_i}^{(\text{total})} \approx \sum_{\alpha=1}^{n_b} F_{\alpha, s_i}^{(\text{total})} \hat{c}_\alpha(x, v)$. Note that if $f_{s_2 s_1} \neq f_{s_1 s_2}$, the cells will exhibit non-reciprocal interactions.

5 Cell-cell interactions on a micropatterned track

To further test the generality of our results, we investigate the dynamics of MDA-MB-231 cell pairs in a different confinement geometry: a short track with the same overall dimensions as the two-state micropatterns, but without a constriction (Fig. S19a). Specifically, this micropattern has average dimensions $((103.4 \pm 0.3)\ \mu\text{m}) \times ((34.8 \pm 0.2)\ \mu\text{m})$. We track a total of 84 cell pairs, resulting in a large data set of coupled cell trajectories (Fig. S19b). By applying the same inference scheme (using the same basis expansion) as for the two-state data, we determine the single cell contribution $F(x, v)$ and the cohesive and frictional interactions (Fig. S19c-e). As before, the single cell contribution is similar to the deterministic contribution inferred from experiments with only a single cell in the pattern, $F_{\text{sc}}(x, v)$ (Inset Fig. S19c; see section 4.4 and ref. [1]). Importantly, we find that the inferred model exhibits the same types of interactions as in the two-state geometry, including a short-range attractive and an anti-friction component. The model performs well on predicting the statistics of the trajectories: the joint probability distribution, the collision statistics, the position probability distribution, and the velocity auto-correlation function are well captured (Fig. S19f-h,j). The cross-correlation of positions deviates slightly, but decays on a similar time-scale (Fig. S19i). This deviation is likely due to a larger freedom to explore the y-dimension of the pattern than in the two-state geometry, which is not accounted for by our 1D model. Taken together, these results demonstrate that our inference procedure can be generalized to other geometries, and that the results of the inferred cell-cell interactions do not sensitively depend on the precise confinement geometry.

We challenge our inference approach further by attempting to predict the collision statistics of interacting cells on a track geometry using only observations from other experiments to constrain our model. Specifically, we combine the single-cell and noise terms inferred from experiments where only a single cell migrates on the track together with the interactive terms inferred from cell pair experiments on two-state micropatterns:

$$\begin{aligned} \dot{x}_i &= v_i \\ \dot{v}_i &= F_{\text{sc}}^{\text{track}}(x_i, v_i) + f^{\text{two-state}}(|\Delta x_{ij}|)\Delta x_{ij} + \gamma^{\text{two-state}}(|\Delta x_{ij}|)\Delta v_{ij} + \sigma_{\text{sc}}^{\text{track}}\eta_i(t) \end{aligned} \tag{S18}$$

Strikingly, this model quantitatively predicts the relative percentage of collision events

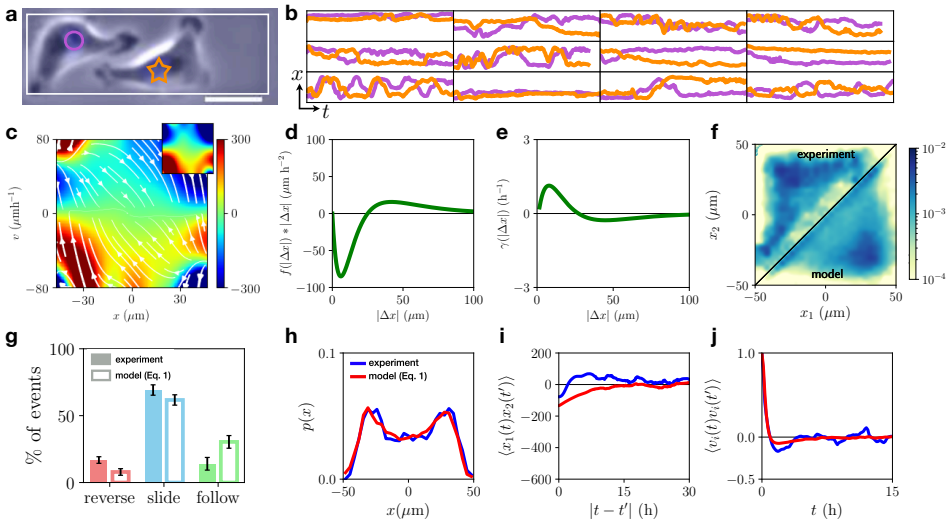


Figure S19: **Cell-cell interaction dynamics on a micropatterned track (MDA-MB-231 cells).** **a**, Microscopy image of an interacting cell pair confined to a micropatterned track. Micropattern outline is shown in white. Scale bar: 25 µm. **b**, Sample set of nuclear trajectories $x_{1,2}(t)$ as a function of time, shown for 12 cell pairs. Axes limits are $0 < t < 30$ h and -60 µm $< x <$ 60 µm, with $x = 0$ at the center of the pattern. A total of 84 cell pairs were tracked. **c**, Single-cell contribution $F(x, v)$ to the interacting dynamics, measured in units of µm/h^2. White lines indicate the flow field given by $(\dot{x}, \dot{v}) = (v, F(x, v))$. Inset: corresponding term inferred from experiments with single cells [1]. **d**, Cohesive interaction term $f(|\Delta x|)\Delta x$. Positive values indicate repulsive interactions, while negative values correspond to attraction. **e**, Effective frictional interaction term $\gamma(|\Delta x|)$. Here, positive values indicate an effective anti-friction, and negative values an effective frictional interaction. **f**, Joint probability distributions $p(x_1, x_2)$ of cell positions, plotted logarithmically. The top triangle of the symmetric distribution shows the experimental result, the bottom triangle shows the model prediction. **g**, Percentages of each of the three types of collision events observed. **h**, Probability distribution of all cell positions $p(x)$ (experiment shown in blue, model predictions in red). **i**, Cross-correlation function of cell positions $\langle x_1(t)x_2(t')\rangle$, plotted on the same scale as the plots in Fig. 2 of the main text. **j**, Normalized velocity auto-correlation function $\langle v_i(t)v_i(t')\rangle$.

observed in experiment, highlighting the potential generalizability of the inferred interactions.

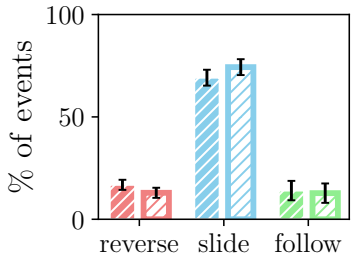


Figure S20: **Predicted collision behaviors using the inferred interactions from two-state micropatterns.** Experimental (solid) and predicted (empty) collision percentages using Eqn. S10. These correspond to the results shown in Fig. 2d of the main text.

6 Construction of the interaction behavior space (Fig. 5)

By inferring a model directly from experimental data, we have shown that the migration of cells in two-state micropatterns is well captured by an equation of motion of the form Eq. (S10). In order to predict behaviors beyond those sampled in our experiments, we extrapolate this description by scanning the interaction space spanned by the prefactors of cohesion and friction. Thus, we simulate trajectories using the equation of motion

$$\dot{v}_i = F(x_i, v_i) + f_0 g_c(|\Delta x_{ij}|)\Delta x_{ij} + \gamma_0 g_f(|\Delta x_{ij}|)\Delta v_{ij} + \sigma\eta_i(t) \tag{S19}$$

Here, we employ the confinement term $F(x_i, v_i)$ and the noise strength σ inferred from MDA-MB-231 experiments. For the inference on our experimental data (section 4), we employed a functional expansion of the interactions using $N = 3$ exponentials, which were required to capture the quantitative details of the correlation functions (Fig. 2 in the main text). The collision behavior is already well captured with a simple expansion consisting a single exponential ($N = 1$). Thus, for simplicity, in the construction of the IBS we use single exponentials. For the result in the main text (Fig. 5), we use $g_c = g_f = e^{-|\Delta x_{ij}|/R_0}$ with $R_0 = 30$ µm.

To construct the IBS, we scan the (f_0, γ_0)-plane to obtain a set of models, each of which can be used to simulate a large set of stochastic trajectories. We apply the same analysis routines to these trajectories as to the experimental data, and identify for each value of (f_0, γ_0) the relative percentage of reversal, sliding and following events (Fig. S21). We find that the interaction space exhibits well-defined behavioral regimes, in which one of the three behaviors dominates. For example, for a combination of repulsive and frictional interactions, we find that reversal events dominate over all other events (bottom right in Fig. S21a). Thus, we can identify for each value of (f_0, γ_0) the collision type that dominates the observed behavior, plotted in Fig. S21d. Using this construction, we can therefore connect the interaction parameters governing the instantaneous short-time dynamics to the emergent long-time behavior. We therefore term this construction *interaction behavior space* (IBS).

Importantly, if we perform the same simulations with the MCF10A single-cell term, we obtain a qualitatively similar IBS (Fig. S22), indicating that our predictions are not sensitive to the details of the single-cell term, but are determined by the interactions. To determine where to approximately place the MCF10A and MDA-MB-231 cells within the IBS, we perform an inference with only a single ($N = 1$) exponential basis function. In accord with our more general inference (using $N = 3$), we find that MCF10A exhibit repulsive and frictional interactions, while MDA-MB-231 cells exhibit attractive and anti-friction interactions (stars in Fig. S22). Furthermore, the qualitative structure of the IBS is not sensitive to the choice of interaction decay length R_0 within a reasonable range, or to using non-exponential kernels (Fig. S23).

These results demonstrate two things: First, a large variety of behaviors can be captured by our model (Eq. (S19)). Second, with new experiments, similar models can be inferred, and the inferred interactions can be placed within the IBS, assuming no additional parameters are required to describe their behavior. Thus, the IBS could provide a way to connect different experimental observations within a single framework.

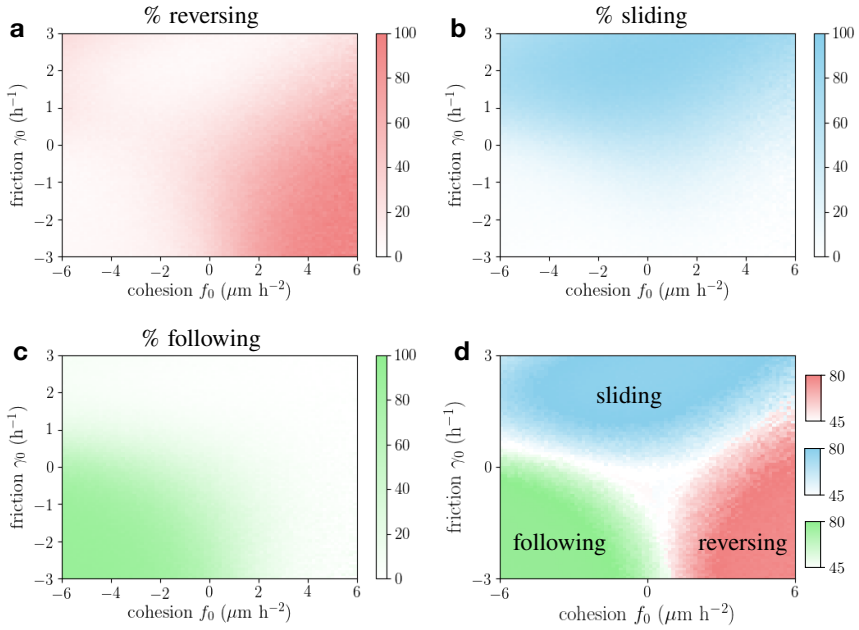


Figure S21: **Identification of behavioral regimes in the interaction behavior space.** **a**, Predicted percentage of reversal events as a function of the cohesion and friction prefactors. Here, the same interaction kernel is used as in Fig. 5 of the main text: $g_{c,f} = e^{-|\Delta x_{ij}|/R_0}$ with $R_0 = 30$ µm. **b**, Predicted percentage of sliding events. **c**, Predicted percentage of following events **d**, To construct the behavioral regimes, we identify the behavior with the maximal percentage in each bin, and plot its percentage in the respective colour scheme. The colour scheme is constructed such that percentages around 50%, where no single behavior contributes the majority of events, are plotted in white.

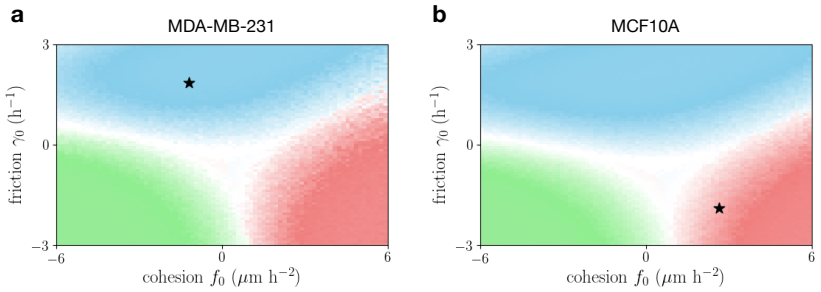


Figure S22: **Robustness of the IBS with respect to the single-cell term $F(x_i, v_i)$.** **a**, Predicted behavioral regimes using the MDA-MB-231 single-cell term. **b**, Predicted behavioral regimes using the MCF10A single-cell term. In both cases, we use $g_{c,f} = e^{-|\Delta x_{ij}|/R_0}$ with $R_0 = 30$ µm. Black stars indicate the inferred values for (f_0, γ_0) for each cell type using an interaction basis consisting only of a single exponential $u_{\text{single}}(|\Delta x_{ij}|) = e^{-|\Delta x_{ij}|/R_0}$ with $R_0 = 30$ µm.

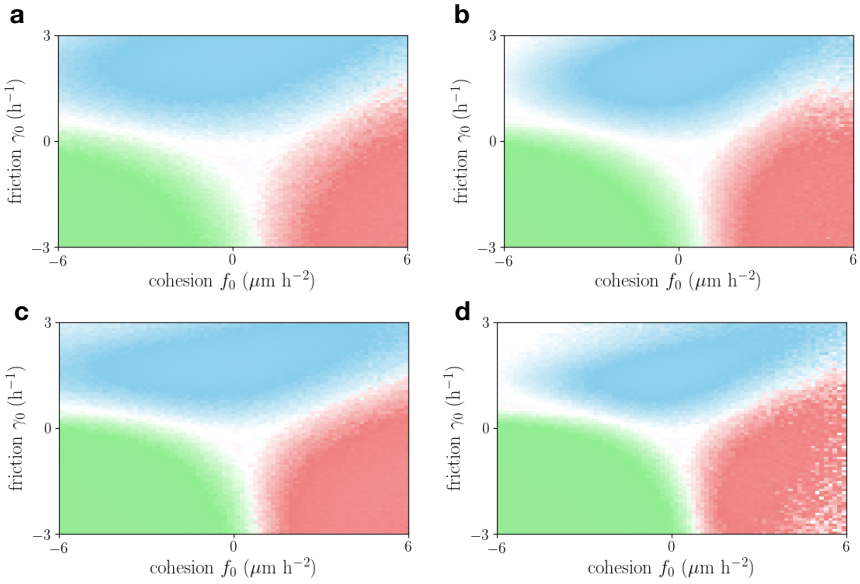


Figure S23: **Robustness of the IBS with respect to the functional form $g_{c,f}(|\Delta x_{ij}|)$ of the interaction kernels.** **a**, Predicted behavioral regimes using $g_{c,f} = e^{-|\Delta x_{ij}|/R_0}$ with $R_0 = 30$ µm (shown in main text). **b**, Predicted behavioral regimes using $g_{c,f} = e^{-|\Delta x_{ij}|/R_0}$ with $R_0 = 40$ µm. **c**, Predicted behavioral regimes using $g_{c,f} = 1/[(|\Delta x_{ij}|/R_0)^4 + 1]$ with $R_0 = 30$ µm. **d**, Predicted behavioral regimes using $g_{c,f} = 1/[(|\Delta x_{ij}|/R_0)^4 + 1]$ with $R_0 = 40$ µm.

References

[1] 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," *Nature Physics*, vol. 15, no. 6, pp. 595–601, 2019.

[2] D. Blair and E. Dufresne, "The Matlab Particle Tracking Code Repository," *http://site.physics.georgetown.edu/matlab/*, 2008.

[3] B. Efron, "Bootstrap Methods: Another Look at the Jackknife," *The Annals of Statistics*, vol. 7, pp. 1–26, jul 1979.

[4] B. Efron and R. Tibshirani, *An Introduction to the Bootstrap*. CRC Press, 1994.

[5] D. F. Milano, N. A. Ngai, S. K. Muthuswamy, and A. R. Asthagiri, "Regulators of Metastasis Modulate the Migratory Response to Cell Contact under Spatial Confinement," *Biophysical Journal*, vol. 110, no. 8, pp. 1886–1895, 2016.

[6] R. A. Desai, S. B. Gopal, S. Chen, and C. S. Chen, "Contact inhibition of locomotion probabilities drive solitary versus collective cell migration," *Journal of the Royal Society Interface*, vol. 10, no. 88, 2013.

[7] J. Singh, B. A. Camley, and A. S. Nain, "Rules of Contact Inhibition of Locomotion for Cells on Suspended Nanofibers," *bioRxiv*, 2020.

[8] D. Li and Y. L. Wang, "Coordination of cell migration mediated by sitedependent cell-cell contact," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 115, no. 42, pp. 10678–10683, 2018.

[9] D. B. Brückner, P. Ronceray, and C. P. Broedersz, "Inferring the dynamics of underdamped stochastic systems," *Physical Review Letters*, vol. 125, no. 5, p. 58103, 2020.

[10] D. Selmeczi, S. Mosler, P. H. Hagedorn, N. B. Larsen, and H. Flyvbjerg, "Cell motility as persistent random motion: theories from experiments.," *Biophysical journal*, vol. 89, no. 2, pp. 912–931, 2005.

[11] G. J. Stephens, B. Johnson-Kerner, W. Bialek, and W. S. Ryu, "Dimensionality and Dynamics in the Behavior of C. elegans," *PLoS Comput Biol*, vol. 4, no. 4, p. e1000028, 2008.

[12] J. N. Pedersen, L. Li, C. Gradinaru, R. H. Austin, E. C. Cox, and H. Flyvbjerg, "How to connect time-lapse recorded trajectories of motile microorganisms with dynamical models in continuous time," *Physical Review E*, vol. 94, no. 6, p. 062401, 2016.

[13] F. Ferretti, V. Chardès, T. Mora, A. M. Walczak, and I. Giardina, "Building General Langevin Models from Discrete Datasets," *Physical Review X*, vol. 10, no. 3, p. 031018, 2020.

Chapter 9

Cell-cell interactions in collective cell spreading

This chapter is based on the following manuscript:

Disentangling cadherin-mediated cell-cell interactions in collective cancer cell migration

Themistoklis Zisis*, David B. Brückner*, Tom Brandstätter, Joseph d'Alessandro, Angelika Vollmar, Chase P. Broedersz†, Stefan Zahler†

* equal contribution
† corresponding author

This manuscript is submitted for publication.

Disentangling cadherin-mediated cell-cell interactions in collective cancer cell migration

Themistoklis Zisis[1]*, David B. Brückner[2]*, Tom Brandstätter[2], Joseph d'Alessandro[3], Angelika M. Vollmar[1], Chase P. Broedersz[2,4,+], Stefan Zahler[1,+]

[1]Ludwig-Maximilians-University Munich, Department of Pharmacy, Center for Drug Research, 81377 Munich, Germany
[2]Arnold Sommerfeld Center for Theoretical Physics and Center for NanoScience, Department of Physics, Ludwig-Maximilians-University Munich
[3]Institut Jacques Monod (IJM), CNRS UMR 7592 and Université de Paris, 75013, Paris, France
[4]Department of Physics and Astronomy, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands

* authors contributed equally
[+] corresponding authors

Cell dispersion from a confined area is fundamental in a number of biological processes, including cancer metastasis. To date, a quantitative understanding of the interplay of single cell motility, cell proliferation, and intercellular contacts remains elusive. In particular, the role of E- and N-Cadherin junctions, central components of intercellular contacts, is still controversial. Combining theoretical modeling with *in vitro* observations, we investigate the collective spreading behavior of colonies of human cancer cells (T24). Inhibition of E- and N-Cadherin junctions decreases colony spreading and average spreading velocities, without affecting the strength of correlations in spreading velocities of neighboring cells. Based on a biophysical simulation model for cell migration, we show that the behavioral changes upon disruption of these junctions can be explained by reduced repulsive excluded volume interactions between cells. This suggests that cadherin-based intercellular contacts sharpen cell boundaries leading to repulsive rather than cohesive interactions between cells, thereby promoting efficient cell spreading during collective migration.

Collective cell migration is central to a number of key physiological processes, including morphogenesis during development [1], as well as immune response [2], wound repair [3] and tissue homeostasis [4] in the developed organism. Aberrant cell migration is associated with several pathologies, such as the spread of malignant cancer cells to previously healthy tissues

during metastasis [5]. The migratory dynamics of cell collectives in these processes are not merely the outcome of many independently moving cells: they are controlled by cell-cell interactions [6, 7]. Specifically, cells form mechanosensitive cell–cell adhesion junctions (adherens junctions) and coordinate their movements by actively interacting with each other [8]. These interactions facilitate a coordination of collective behavior where a colony of cells invades an empty area [9]. However, it remains unclear how different types of cell-cell interactions control such collective spreading behavior.

The trajectories of single migrating cells are well described by stochastic trajectory models, both for cells migrating on 2D surfaces [10-12] and in confining environments [13-15]. Yet, it is challenging to describe the stochastic collective migration of a cancer cell colony, as cell division and cell-cell contacts have to be taken into consideration. Cell-cell contacts lead to a variety of interactions between cells. Firstly, cells exhibit excluded volume (EV) interactions, where an individual cell occupies space and exerts a repelling force on other cells that move within this space [16]. Secondly, many cell types have the tendency to reorient their direction of motion and move apart upon contact, which is referred to as Contact Inhibition of Locomotion (CIL) [17, 18]. In physical stochastic trajectory models, these interactions are frequently incorporated as a combination of *repulsive interactions*, modelling EV, and *velocity interactions* including velocity alignment as well as CIL [19-23]. Conceptually, there is a key difference between these interactions: while repulsive interactions depend on the relative positions of cells, velocity interactions depend on their motion, i.e. their velocities or polarities. However, it remains unclear how changes in cell-cell contacts within a migrating colony influence these distinct types of interactions and the resulting collective migratory behavior.

Intercellular interactions are strongly dependent on Cadherins, highly conserved calcium-dependent transmembrane proteins that constitute the main component of adherens junctions. Type I classical cadherins (including epithelial (E) and neuronal (N) cadherin as well as P-, R- and M-cadherin [24]) form strong cell–cell adhesion by predominantly homotypic interaction between their extracellular domains [25]. The intracellular cadherin domains connect to β- and α-catenins that associate with the actin cytoskeleton to mediate mechanotransduction [26]. Changes in the normal expression levels of the different cadherin types has been associated with carcinogenesis. One of the most studied processes related to several epithelial tumors is the cadherin switch observed during Epithelial-Mesenchymal transition (EMT). EMT involves the loss of epithelial cell polarity and cell-cell adhesion and the gain of migratory and invasive properties, resulting in the predominance of a mesenchymal phenotype [27]. More specifically, there typically is a strong downregulation of E-Cadherin in parallel with an upregulation of N-Cadherin

in EMT. As a result, E-Cadherin adherens junctions disassociate while N-Cadherin junctions establish a relatively weak (compared to E-Cadherin) adherens junction [28].

However, the role of E- or N- Cadherin-mediated intercellular adhesions in cancer cell migration remains controversial. On the one hand, E-Cadherin downregulation has been related to cancer development [29, 30], and it has been shown that the presence of E-Cadherin induces a spreading cell monolayer to retract and form a spheroid aggregate, a process called dewetting [31], suggesting its role as a potent tumor suppressor. On the other hand, a number of studies suggest the opposite effect: E-Cadherin is required for metastasis in multiple models of breast cancer [32], it promotes expansion of bladder carcinoma in situ [33], and is highly present in patients with prostate cancer [34], ovarian cancer [35], and glioblastoma [36]. A similar controversy characterizes the involvement of N-Cadherin in migratory behavior. Although N-cadherin is a marker of EMT and its expression has been associated with the development of multiple cancer types [28], there are studies pointing in the opposite direction. In fact, N-cadherin loss was associated with increased tumor incidence [37] and metastasis [38]. Consequently, a question is yet to be answered: what is the distinct contribution of E- and N-Cadherin junctions to cell-cell interactions and the resulting spreading dynamics of cancer cell colonies?

Here, we aim to investigate this question by combining experimental observations on collectively migrating cells and a minimal physical model of the spreading behavior. We use an epithelial bladder cancerous cell line (T24), which is characterized by high N-Cadherin expression and limited [39] or zero functional levels of E-Cadherin [40, 41]. After initial confinement of a colony of cells to a circular micropattern, the cells are released using chemical tools [42, 43]. We quantify the collective migration by identifying and tracking the entire ensemble of single cell trajectories in each colony. To investigate the effect of cell-cell contacts for the migration, we inhibit E- or N-Cadherin junctions via specific blocking antibodies. In both cases, our analysis reveals that such inhibition leads to a reduced spreading velocity of the cell colonies. To elucidate these dynamics, we develop a minimal active particle model for collective migration, that includes cell proliferation as well as repulsive and CIL interactions. This model shows that inhibiting E- or N-Cadherin has an effect akin to reducing the strength of repulsive cell-cell interactions in the model. In other words, disturbing either of these cadherin junctions decreases the displacement generated when neighboring cells push each other away in order to create space for themselves. Therefore, we show that both E- and N-Cadherins contribute to the maintenance of intercellular contacts that facilitate cell spreading via repulsive interactions, causing cells to move further away from each other. This could be a consequence of cadherins 'sharpening' cellular boundaries, through e.g. shape regulation, changes in interfacial tension, or increased cell-cell recognition

[44]. These observations indicate the important role of cadherins in metastatic events and their potential as cancer treatment targets.

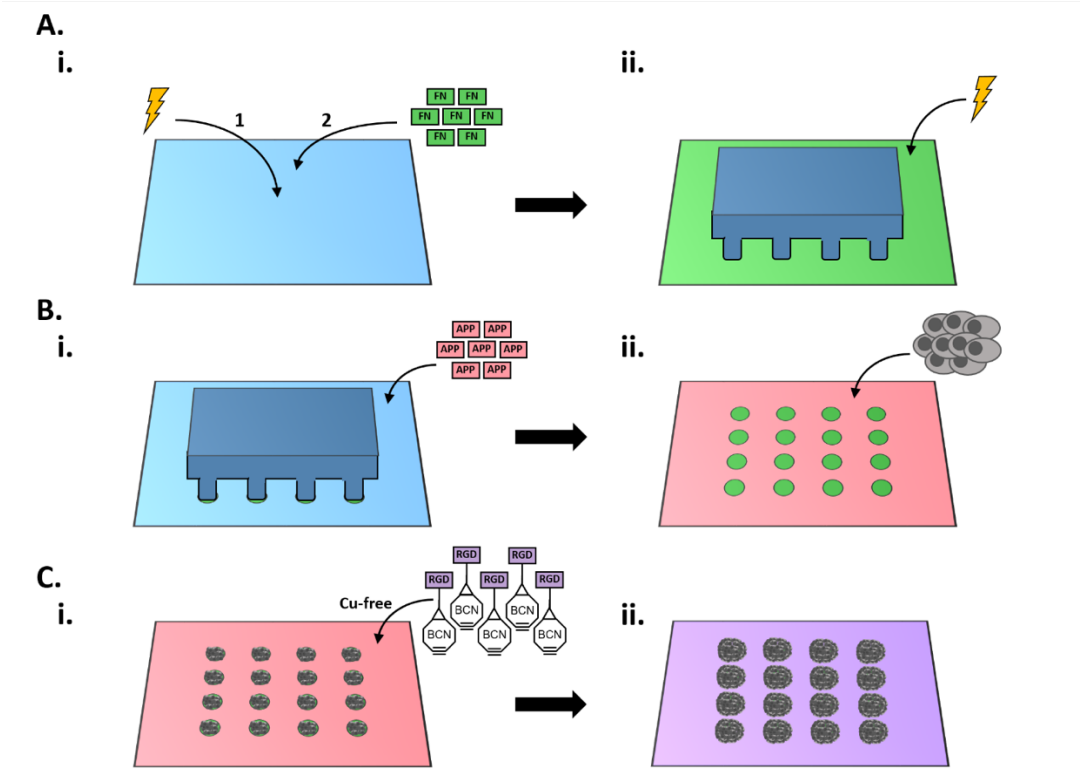


Figure 1 | Schematic representation of the microcontact printing and click-chemistry process. A i) Ibidi's uncoated surface (here one well is represented) undergoes plasma treatment to become reactive, for subsequent attachment of fibronectin (FN). ii) Example of PDMS-square stamp with circular patterns produced with standard microcontact printing techniques (blue). The stamp is placed at the center of the well and the surface is plasma-treated again. The whole surface except for the stamp-protected circular areas loses its fibronectin coating and becomes hydroxylated. B i) With the stamp remaining in place, APP is added next to it and absorbed by the whole surface except for the stamp-protected circular areas (green). B ii) This results in fibronectin-coated circular areas (green) surrounded by an otherwise cell repellent APP surface (red). T24 cancer cells are seeded on the circular areas forming the initial cell population. C i) BCN -RGD peptides are then added and bind to the APP coated surface via click chemistry reaction between the BCN and the azide groups of the APP. C ii) The previously cell-repellent surface is now coated with RGD and thus, highly cell adhesive. The cells are able to expand (migrate) from the circular areas to the rest of the surface.

Release from a micropatterned circular adhesive area leads to collective cell spreading

To generate an experimental setup for tracking collective cell spreading dynamics, we develop a micropatterned platform from which cells can be released in a standardized manner. Specifically, we design a new patterning approach based on a novel sequence of surface plasma treatment, standard microcontact printing, fibronectin coating and click chemistry steps. This process results in the production of circular fibronectin-coated adhesive areas that are surrounded by cell-repellent azido (PLL-g-PEG) (APP)-coated surfaces. These non-adhesive surfaces can then be activated on demand, via a biocompatible click chemistry reaction between the azide groups of the APP on the surface and added BCN-RGD peptides to allow time-controlled cell migration outside the circular areas [45] (see Materials and Methods and Fig. 1). Subsequently, we use T24 urothelial bladder carcinoma epithelial cells which is a well-established malignant cell line [46], widely used in cell migration research [47-50] and in EMT transition [50]. The cells are detectable using fluorescence microscopy imaging via their nuclear H2B-GFP fluorescent tag.

We perform time-lapse fluorescence and bright-field microscopy for the first 24 hours after surface activation. Here, we observe cells increasingly spreading outwards over time, in all directions, covering a large circular area (Fig. 2 A). While the cells form an approximately confluent monolayer, there are occasional gaps within the layer and significant nearest-neighbor rearrangements during the spreading process (Supplementary Movie S1). Thus, to gain access to the dynamics of the entire cell collective, we perform tracking of the fluorescently tagged nuclei as previously described [43], giving access to the full ensemble of cell trajectories in each escaping cluster (Fig. 2 B).

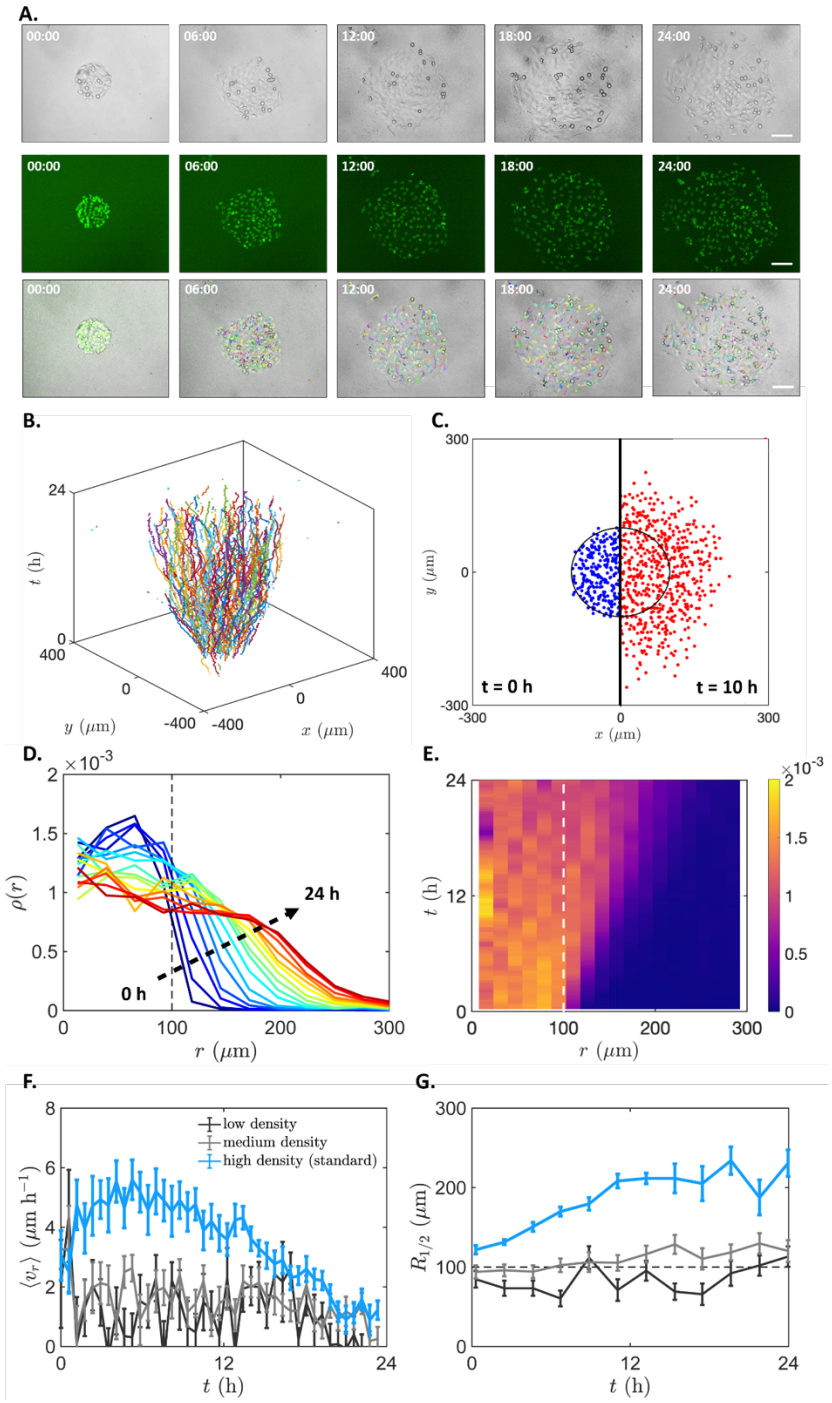


Figure 2 | Cell spreading and evolution of cell density of control (untreated) T24 cells. A) Time-lapse bright-field (upper), fluorescence (middle) microscopy images or overlay with cell tracks (lower) showing the T24 cell migration with 6h intervals from 0h to 24h after surface activation. B) Space diagram of colony spreading up to 24h after surface activation. C) Colony spreading radius of T24 cells at 0h (blue) vs 10h (red) after surface activation. D) Evolution of the density profiles over 24 hours (blue to red) plotted as

the mean of all colonies (n=14). All curves are separated by 1 h intervals. E) Kymograph of the cell density evolution, corresponding to (D). F) Mean radial velocity (u_r) over time (average of all colonies per density). The high-density colonies (blue) exhibited a direct increase in radial velocity, larger than the medium and low-density colonies (gray and black, respectively), peaking around 5h after surface activation. G) Average distance where density has decayed to half of its value in the center of the original confinement (i.e. at r=0). The distance was higher over time for the high density compared to the medium and low-density condition. Error bars: SEM; n_{high}= 12, n_{medium}= 15, n_{low}= 12. Space diagram of colony spreading for the different cell densities and complete density evolution profiles shown as mean of all colonies are provided in Supplementary Fig. S4.

At the single cell level, these trajectories are also highly stochastic, as expected from single cells which perform stochastic persistent random motion on unstructured 2D substrates [12]. As shown by the space-time trajectories of the system, the cells have an overall tendency to escape the cluster, and after a period of 10h, a large fraction of the cells has left the initial confinement (Fig. 2 C). The spreading process is quantified by the evolution of the radial density profile $\rho(r)$ of the cluster (Fig. 2 D, E). Specifically, we calculate the average number of cells per area element as a function of the distance to the center of the initial confinement radius. As a function of time, the density within the confinement initially decreases, due to cells leaving the confinement through random migration. Correspondingly, the density outside the confinement increases. Interestingly, after a period of approximately 10h, the density inside the confinement stabilizes at a constant value. To further quantify the overall spreading, we calculate the average radius at which the density profile has decayed to half its value at the center of the initial confinement $R_{1/2}$ (Fig. 2 G). Finally, we quantify the average radial velocity of the spreading cells as a function of time, which reveals a marked peak at intermediate spreading times (Fig. 2 F). These statistics are helpful to investigate the impact of collective effects. Thus, we analyzed clusters initialized in the same confinement radius, but with lower cell concentrations. At these lower concentrations, less spreading is achieved (Fig. 2 G), and the peak in radial velocity disappears (Fig. 2 F), indicating that the dynamics observed in our experiments are density dependent, and therefore have a distinct collective character. The collective escape behavior is therefore likely determined by a combination of single-cell motility, cell proliferation, and cell-cell interactions.

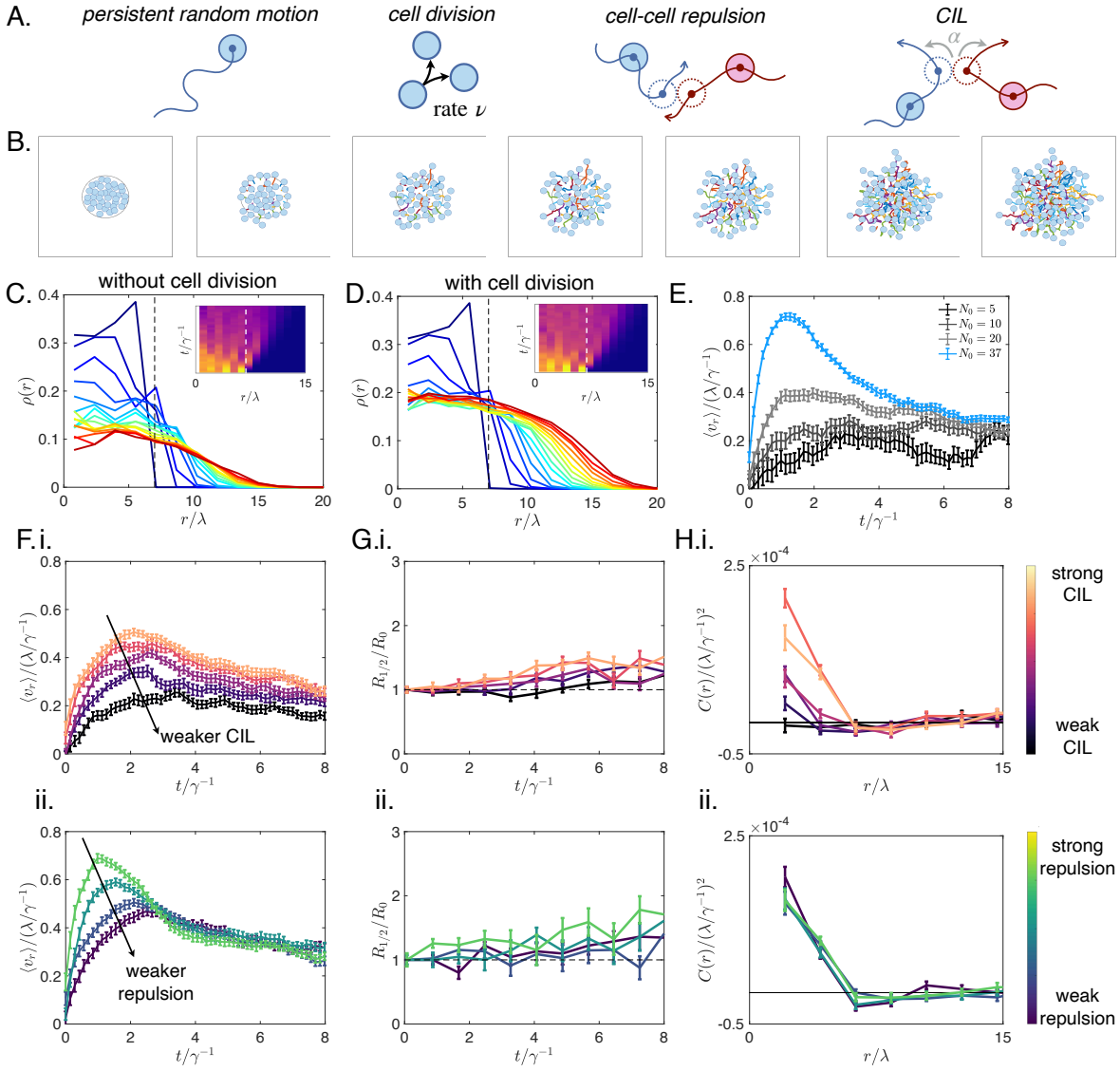


Figure 3 | Computational model for collective cell escape. A) Schematic of the components of our active particle model, from left to right: persistent random motion of individual particles, cell division with constant rate ν, excluded volume interactions, and Contact Inhibition of Locomotion. B) Time-series of a cluster escape simulation. Cell positions are shown as blue circles of radius λ, which is the radius of the repulsive potential. Previous motion of the cells is shown as colored trajectories. C, D) Evolution of the density profile over time (blue to red) plotted as the mean of n=30 colonies. Inset: Kymograph of the cell density evolution. Dashed lines indicate the initial confinement radius. B corresponds to a model without cell division, while C includes cell division. E) Mean radial velocity over time for clusters with different initial density. F) Mean radial velocity over time for clusters with (i) different CIL interaction amplitudes, and (ii) different strengths of cell-cell repulsion interactions. G) Average distance where density has decayed to half of its value in the center of the original confinement (i.e. at r=0). H) Cross-correlation of velocity fluctuations. Error bars: SEM; n=30 for all panels.

Minimal active particle model captures experimental colony spreading

To elucidate the interplay of the various factors affecting the collective migration in our experiments, we develop a minimal active particle model for collective cell migration (Fig. 3 A). In this model, single cells perform persistent random motion, as observed for single cell migration on two-dimensional substrates [10]. We include cell-cell interactions in our model through two distinct contributions [21, 43, 51]: a repulsive component modelling excluded volume (EV) interactions, and Contact Inhibition of Locomotion (CIL) which models the tendency of cells to reorient away from contacts upon collision. We first confine the particles into a circular region of radius *R* and then observe their behavior upon release, exactly like in the experiment (Methods Section and Fig. 3 B).

Interestingly, this model predicts a rapid decay of the density within the initial confinement area over time, as particles perform random motion and are repelled by their neighbors and move outwards (Fig. 3 C). This observation is inconsistent with our experimental data, which showed only a weak decay in the initial confinement area (Fig. 2 D, E). As shown by our cell proliferation estimations, cell division plays an important role on the time-scale of tissue spreading in this system: the number of cells nearly doubles within 10h (Supplementary Fig. S3). We therefore include a basic implementation of cell division in our model, where cells stochastically perform divisions at a constant rate. This model including cell division exhibits a slower decay of density, and an overall density profile that is consistent with our experimental observations (Fig. 2 D, 3 D). This also suggests that divisions play an important role in the experiment by keeping the cell layer close to confluent. This prevents the density from decreasing too quickly, in which case cells would not interact significantly, further supporting the important role of cell proliferation in collective cell spreading phenomena.

Having included cell division, we find that our model captures other key features of the experimentally observed dynamics. Importantly, we observed that the model predicts a peak in the radial velocity (Fig. 3 F i), similar to experiments (Fig. 2 G). This peak in radial velocity on a time-scale of the order of the persistence time of the cells corresponds to the outward diffusive flux expected for a collection of self-propelled particles [43, 52]. Specifically, upon removal of the confinement, cells at the boundary of the cluster are repelled by the bulk of the cluster, leading to a re-orientation of their movement in an outward direction. This causes the initial increase of the average radial velocity, which is followed by a decreasing trend due to the randomization of movement once the cluster has spread significantly. Furthermore, our model reproduces the gradual increase of the spreading radius (Fig. 3 G i), and a positive cross-correlation of velocity fluctuations indicating short-ranged alignment of cell movement (Fig. 3 H i). Finally, our model

correctly predicts a reduction in the radial velocity for lower cell densities, as we observed experimentally (Fig. 3 E). Taken together, these results demonstrate that our cell cluster experiments exhibit the behavior expected for a collection of active particles with interactions. In the experiment, the interactions between cells are known to be controlled by transmembrane proteins, including E- and N-Cadherins [44, 53], whose role we seek to elucidate in the next section.

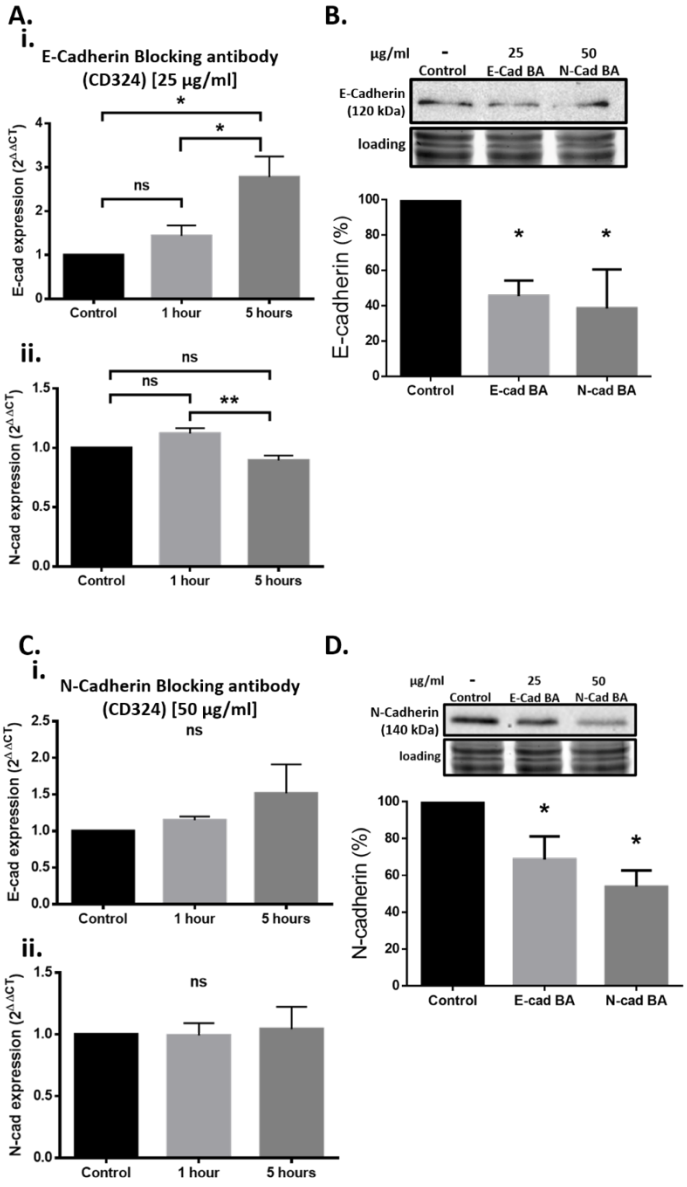


Figure 4 | Effect of blocking antibody treatment on E- and N-Cadherin gene and protein expression. A) Quantitative PCR analysis of (i) E- and (ii) N-Cadherin gene expression in untreated (control) or treated

T24 cells with E- blocking antibody for 1 and 5 hours, respectively. E-Cadherin blocking antibody treatment at the highest concentration tested (25um/ml) resulted in a significant upregulation of the E-Cadherin gene expression after 5 hours compared to control (mean diff. ± SE= 1.780± 0.4290, p=0.0142). Furthermore, the same treatment resulted in a significant downregulation of N-Cadherin gene expression at the same timepoint compared to the 1h timepoint, indicating a cadherin-switching effect (mean diff ± SE = 0.2251 ±0.05014, p=0.0099). B) Quantitative Western Blot analysis of E- cadherin protein levels in untreated (control) or treated T24 cells with 25µg/ml E-Cadherin or 50µg/ml N-Cadherin blocking antibody after 24 hours. Both antibodies significantly reduced the levels of E-Cadherin after 24 hours (Control vs E-CAD BA: mean diff ± SE=54.43±19.40, p=0.03; Control vs N-CAD BA: mean diff ± SE= 61.47±19.40, p=0.03). C) Quantitative PCR analysis of (i) E- and (ii) N-Cadherin gene expression in untreated (control) or treated T24 cells with N- blocking antibody for 1 and 5 hours, respectively. N-Cadherin blocking antibody treatment at the second highest concentration tested (50 µg/ml) resulted in a non-significant upregulation of N- and E-Cadherin gene expression at 5h compared to control (N-Cadherin: mean diff ± SE= 0.0447±0.1676, p=0.9618; E-Cadherin: mean diff ± SE = 0.5150±0.3258, p=0.3232). D) Quantitative Western Blot analysis of N-Cadherin protein levels in untreated T24 cells (control), cells treated with 25µg/ml E-Cadherin blocking antibody and cells treated with 50µg/ml N-Cadherin blocking antibody. Both antibodies significantly reduced the levels of N-Cadherin after 24 hours (Control vs E-CAD BA: mean diff ± SE=31.37±12.50, p=0.02; Control vs N-CAD BA: mean diff ± SE= 53.80±12.50, p=0.02). Untreated cells were used for data normalization. One representative Western blot is shown per condition including a total protein loading control. Whole Western blots are shown in supplementary Figure S4. Statistical analysis was performed using 1-way ANOVA followed by Tukey's multiple comparisons (qPCR) or Sidak's multiple comparisons (WB) test; $p< 0.05$ (*), $p< 0.01$ (**); n = 3 (triplicates).

Effect of blocking antibody treatment on E- and N-Cadherin gene and protein expression

To investigate the role of E- and N-Cadherin adherens junctions in collective cell migration, we inhibit their function using either E- or N-Cadherin blocking antibodies at different concentrations. To assess the effect of E-Cadherin blocking antibody on the different cadherin gene expression levels, we perform qPCR for E- and N- Cadherin genes at 1h and 5h after E-Cadherin blocking antibody treatment at the highest concentration tested (25 µg/ml). The qPCR serves as a short-term indicator of compensatory reactions of the cells upon functional blocking of an adhesion molecule in the crucial 5h time window after activation. This 5h timepoint coincides with the peak spreading velocities in the control condition and is therefore of particular interest. We find a significant upregulation of the E-cadherin gene expression after 5 hours compared to control (Fig. 4 Ai). This increase can be considered as a compensatory mechanism of the cell to normalize its E-Cadherin functionality after the antibody-mediated blockage. Furthermore, the same treatment results in an early slight upregulation followed by significant

downregulation of N-cadherin gene expression at 5h (Fig. 4 Aii). The latter result indicates that the upregulated E-Cadherin blocks the expression of N-cadherin [54, 55] , which may correspond to a known phenomenon called cadherin switching (extensively reviewed by Loh et al. [28]). Moreover, using Western Blot (WB), we evaluate the effect of E- or N-blocking antibody on E-Cadherin protein levels, as WB provides a longer time-scale endpoint image of the blocking effect on the total E-Cadherin levels. Here, we observe a significant downregulation of E-Cadherin at 24h after treatment with the E-Cadherin blocking antibody, verifying the antibody functionality. E-Cadherin is also downregulated after N-Cadherin blocking antibody treatment (Fig. 4 B), which further implies the presence of a cadherin switching effect. Specifically, the N-Cadherin blocking antibody could transiently increase the gene expression of N-Cadherin, as a compensatory mechanism, which in turn could represses E-Cadherin expression. Interestingly, for E-Cadherin in the control (untreated) condition, we detect multiple shorter bands rather than one band of 130-135 kDa which is the normal size of the protein. The observed bands were a size of ~120 kDa, 95 kDa and 55kDa (as shown in Supplementary Fig. S5, respectively). Such deviations from the 135 kDa range, involving predominantly a soluble 80 kDa species [56] [57] [58] as well as 97 kDa [59], 48 kDa [60] and 23 kDa [61] fragments are common in the literature and have been associated with the development of different cancer types [28] [62] [63]. Therefore, as E-Cadherin protein expression is known to be very limited [39] or non-existent [40, 41] in T24 cells, it is probable that the shorter E-Cadherin fragments we see are a result of protein degradation.

We then investigate the effect of N-Cadherin blocking antibody on cadherin gene expression levels, by performing qPCR for E- and N- Cadherin genes 1h and 5h after N-Cadherin blocking antibody treatment at the second highest concentration tested (50 μg/ml). In that case, a slight tendency towards upregulation of E-Cadherin gene expression is observed at 5h compared to control (Fig 4 Ci), while the N-Cadherin expression levels were not significantly different from untreated cells (Fig. 4 Cii). This lack of significance could result from the fact that in T24 cells, the presence of N-Cadherin is much higher compared to E-Cadherin [39] and thus a higher concentration of blocking antibody would be required for a stronger effect. However, we observe the clear long-term influence of E- or N- blocking antibody on N-Cadherin protein levels by WB where we identify a significant downregulation of N-Cadherin protein levels at 24h after E- and N- Cadherin blocking antibody treatment (Fig. 4 D). Therefore, we conclude that treatment with either E- or N- cadherin blocking antibody starts with a transient upregulation in the corresponding cadherin gene expression which in turn leads to activation of the cadherin switching mechanism that results in the downregulation of the opposite cadherin. This result is further verified by the WB results, where E- or N- cadherin protein levels are significantly downregulated when cells are treated with opposite blocking antibody over the long-term 24h

timepoint. With regards to the WB-detected N-Cadherin bands in the untreated condition, a clear band at the expected size (140kDa) is always observed, suggesting that there was no apparent degradation or soluble form as was the case for E-Cadherin. This is not surprising, as N-Cadherin is the predominant and fully functional cadherin in the T24 cell line [39, 64, 65].

In summary, these findings verify that (i) there is a low gene and protein expression of functional (membrane-bound) E-Cadherin in our T24 cells (Supplementary Fig. S5 C, D) and that (ii) besides the direct blocking effect, there is an 'off target' blocking effect, where the continuous overexpression of the cadherin being directly blocked leads to a downregulation of the opposite cadherin due to cadherin switching.

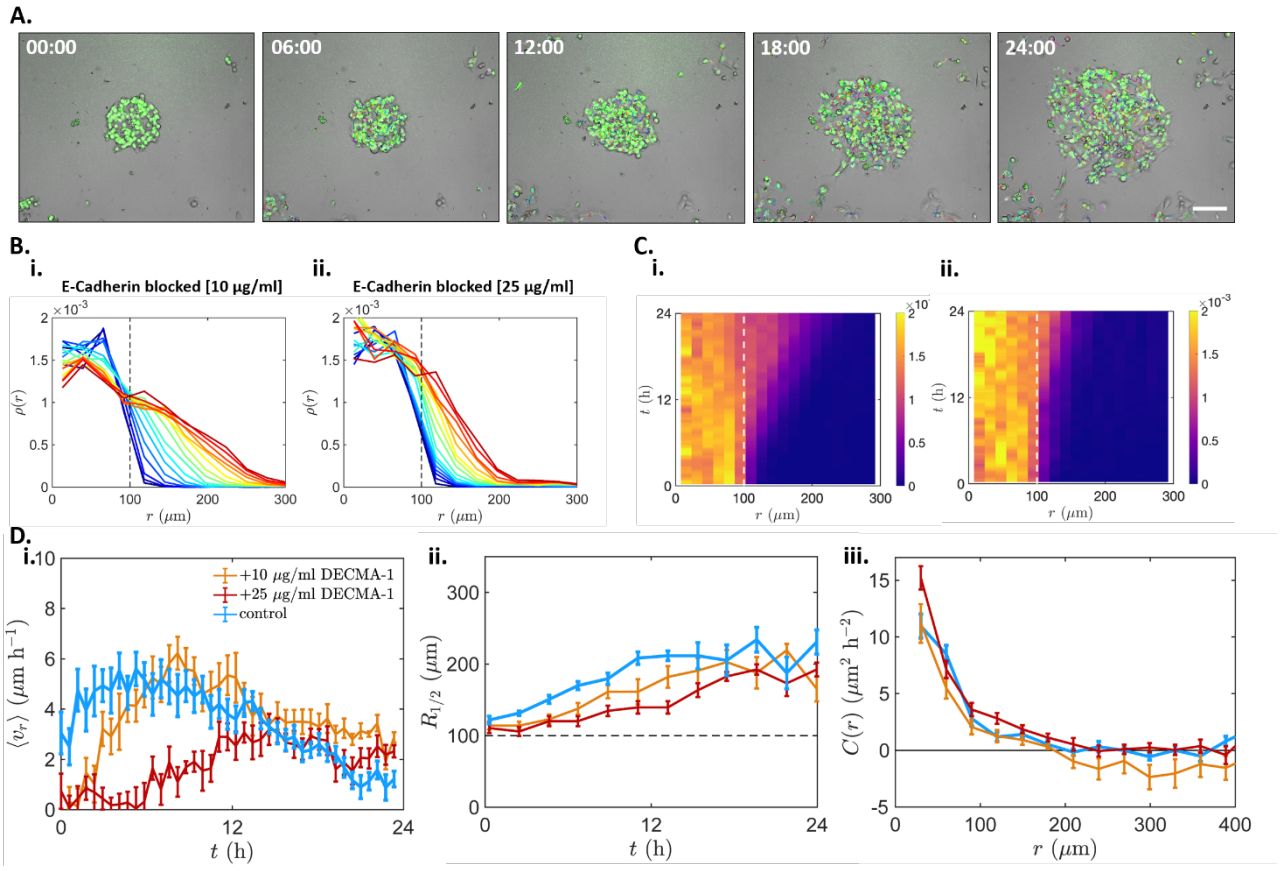


Figure 5 | Evolution of cell density profile, radial velocities and average distance of T24 cells treated with increasing concentrations of E-Cadherin blocking antibody. A) Time-lapse overlay of bright-field and fluorescence microscopy images with cell tracks of the 25 µg/ml E-Cadherin blocking, showing the T24 cell migration with 6h intervals from 0h to 24h after surface activation. B) Space diagram of colony spreading up to 24h after surface activation. B) Evolution of the density profiles over 24 hours (blue to red) plotted as the mean of all colonies per condition for T24 cells treated with (i) 10 µg/ml or (ii) 25 µg/ml E-Cadherin

blocking antibody. All curves are separated by 1 h intervals. C) Kymographs of the cell density evolution, for T24 cells treated with (i) 10μg/ml and (ii) 25 μg/ml E-Cadherin blocking antibody. D) i) Mean radial velocity (u_r) over time (average of all colonies per condition). The control condition exhibited a direct increase in radial velocity, peaking around 5h after surface activation (blue). 10 μg/ml E-Cadherin blocking antibody slowed down this increase in radial speed, which peaked at 8h (orange). The highest concentration of blocking antibody (25 μg/ml) resulted in even lower radial velocity that did not reach the initial peaks exhibited in the other conditions (red). ii) Average distance where density has decayed to half of its value in the center of the original confinement (i.e. at r=0). The distance was the highest over time in the control condition and decreased with increasing concentrations of E-Cadherin blocking antibody. iii) Cross correlation of velocity fluctuations showing no significant differences between conditions. Error bars: SEM; $n_{control}$= 12, n_{10ECAD}= 13, n_{25ECAD}= 8.

Disrupting E- and N-Cadherin junctions decreases speed of collective spreading

Having quantified the E- and N-Cadherin expression upon different levels of E- or N-Cadherin blocking, we move on to analyzing the collective migration behavior in these conditions. First, we find that a low concentration of E-Cadherin blocking antibody (10 μg/ml) does not significantly affect migration behaviour such as the colony spreading represented by density profiles and radial velocities of the cells (Fig. 5 A, Bi, Ci, Di, Dii and Supplementary Movie S4). However, blocking E-Cadherin at a higher concentration of antibody (25 μg/ml) reduces the average spreading of the colonies (Fig. 5 Bii, Cii, Dii) as well as the average radial velocity of the cells (Fig. 5 Di). Similarly, increasing concentrations of N-cadherin blocking antibody leads to reduced average colony spreading and radial velocities, with the highest one (100 μg/ml) having the strongest effect (Fig. 6 A, B, C, Di, Dii and Supplementary Movie S5). In contrast, we find that the average velocity of single migrating cells in experiments with sparsely seeded cells is not significantly affected by the addition of either blocking antibody, for the whole duration of the experiment (Supplementary Fig S1). Furthermore, the proliferation of cells is similar across all conditions (Supplementary Fig S3). These observations suggest that the change in spreading behaviour upon Cadherin blocking is not mediated by changes in the behaviour of single cells or their proliferation, but is mainly caused by the reduction in cell-cell interactions and is thereby a collective effect.

To identify a possible change in cell-cell interactions due to cadherin blocking, we calculate the cross-correlation functions of velocity fluctuations between pairs of cells, which quantifies how similar cellular velocities are as a function of their distance from one another (Methods Section and Fig. 5 Diii and 6 Diii). As expected, in the control condition, we find that cells tend to align their direction of motion with neighbouring cells, but exhibit no correlations at long distances.

Unexpectedly, however, we find that all observed experimental conditions have a similar cross-correlation function. This indicates that while we expect a change in cell-cell interactions to be responsible for the change in spreading behavior, this change does not directly affect the degree of velocity alignment, quantified through the velocity cross-correlation. In a following section, we will turn to a theoretical model for a possible explanation of these observations.

To summarize our experimental findings, we find that by partially blocking either E- or N-cadherin adherens junctions, the collective spreading behaviour of initially confined clusters of T24 cells becomes less efficient. This suggests that cell-cell contacts are important for coordinated migration, possibly by promoting cell-cell interactions. This result is in agreement with earlier reports showing that preventing cells from forming stable cell-cell contacts resulted in uncoordinated and random cell movement [66], leading to significantly lower migration velocities [67]. In contrast to other studies observing no E-Cadherin expression in T24 cells, we detect its presence (120 kDa) among other fragmented species of the protein. Furthermore, we show that as a type III carcinogenic line, T24 cells exhibit an increased N-Cadherin vs E-Cadherin expression ratio (3/1 as shown in Fig. S 5 C, D), characteristic for EMT [28]. Interestingly, we find that the limited E-Cadherin expression is still important for the efficiency of the collective migration, as is the more predominantly expressed N-Cadherin. Therefore, the interplay between E- and N-Cadherin in T24 cells points to a crucial balance in cell-cell contacts that seems to be important for collective migration. In the next section, we use our minimal active particle model to elucidate the nature of these interactions and how they influence the cell spreading behavior.

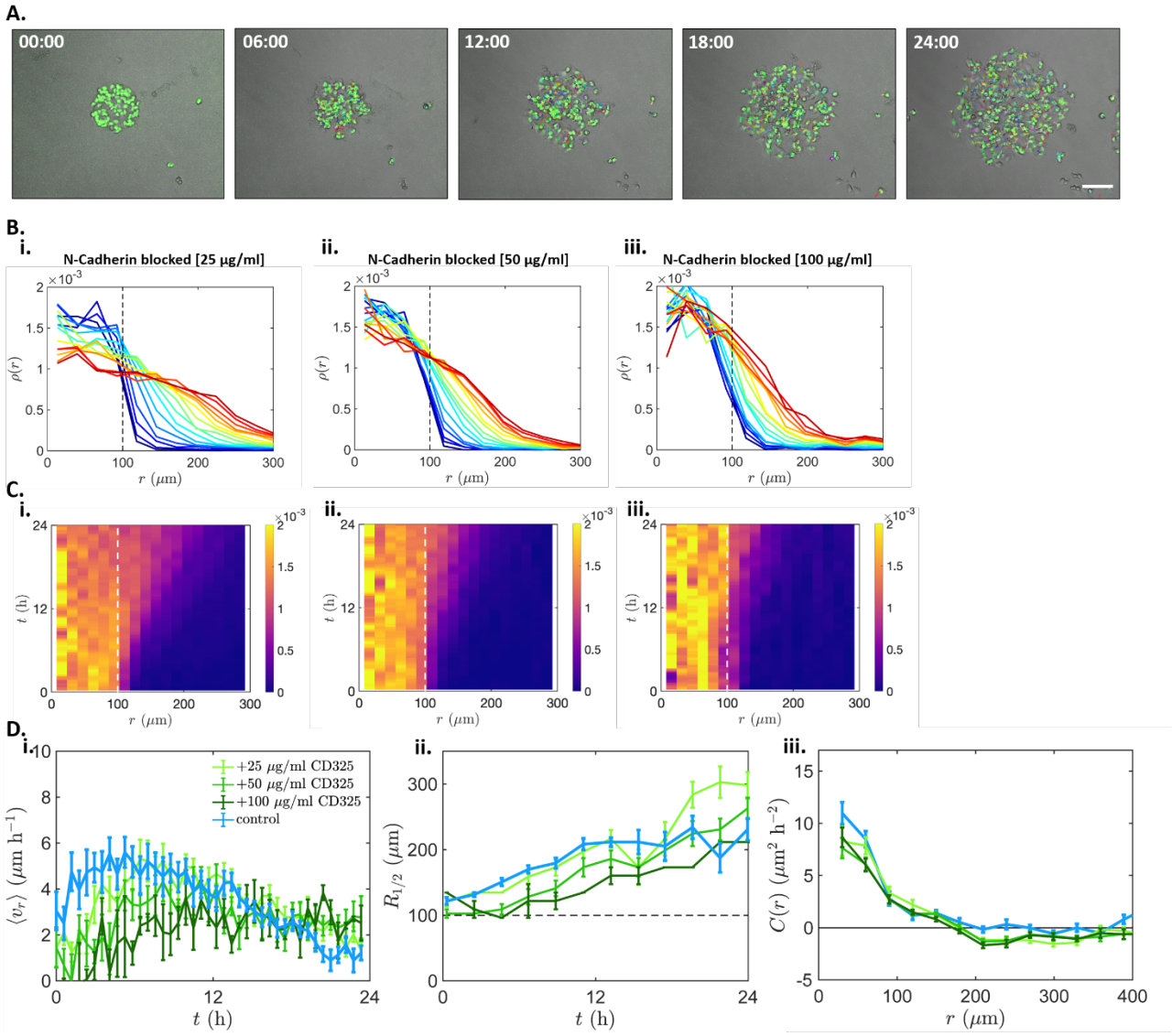


Figure 6 | Evolution of cell density profile, radial velocities and average distance of T24 cells treated with increasing concentrations of N-Cadherin blocking antibody. A) Time-lapse overlay of bright-field and fluorescence microscopy images with cell tracks of the 100 µg/ml N-Cadherin blocking, showing the T24 cell migration with 6h intervals from 0h to 24h after surface activation. B) Evolution of the density profiles over 24 hours (blue to red) plotted as the mean of all colonies per condition for T24 cells treated with (i) 25 µg/ml, (ii) 50 µg/ml or (iii) 100 µg/ml N-Cadherin blocking antibody. All curves are separated by 1 h intervals. C) Kymographs of the cell density evolution, for T24 cells treated with (i) 25 µg/ml, (ii) 50 µg/ml or (iii) 100 µg/ml N-Cadherin blocking antibody. D) i) Mean radial velocity (u_r) over time (average of all colonies per condition). The control condition exhibited a direct increase in radial velocity, peaking around 5h after surface activation (blue). Increasing concentrations of N-Cadherin blocking antibody reduced this increase in radial speed, with the highest reduction observed in the 100µg/ml treated cells (dark green). ii) Average distance where density has decayed to half of its value in the center of the original

confinement (i.e. at r=0). The distance was the highest in the control condition and decreased with increasing concentrations of N-Cadherin blocking antibody up to 11h. After this timepoint, the 25 µg/ml N-Cadherin blocking antibody treated colonies surpassed the control ones. iii) Cross correlation of velocity fluctuations showing no significant differences between conditions. Error bars: SEM; $n_{control}$= 12, n_{25NCAD}= 8, n_{50NCAD}= 6 $n_{100NCAD}$= 3.

Varying cell-cell interactions in a minimal active particle model captures the effects of Cadherin blocking

To investigate how changes in cell-cell interactions affect the spreading behavior in our model, we first vary the strength of contact inhibition of locomotion (CIL). We implement CIL as an angular repulsion that acts as a torque on cells undergoing a contact, with strength α, similar to previous work [43] (see Fig. 3 A and Methods Section). We find that decreasing α, corresponding to weaker CIL, leads to a reduction in radial velocity, spreading, and cross-correlations (Fig. 3 F i, G i, H i). Thus, while the first two findings are in line with the changes in behavior upon cadherin inhibition in the experiment, the change in cross-correlation is not observed in the experiment. In contrast, reducing the strength of the repulsive interactions between particles leads to a reduction of the radial velocity peak and the overall spreading, while keeping the cross-correlations constant (Fig. 3 F ii, G ii, H ii) - similar to what we observed experimentally upon blocking E- or N-Cadherin-mediated intercellular contacts (Fig. 5 B ii, D and 6 B ii-iii, D). These results are robust over a wide range of parameters in the model (Supplementary Figs. 8-11). These observations suggest that disrupting cell-cell junctions through E and N-Cadherin blocking has an effect akin to reducing excluded volume interactions between cells.

The reduced spreading for weaker CIL and weaker repulsive interactions can be understood intuitively. Firstly, CIL interactions ensure that cells at the cluster boundary do not cross paths, leading to outward alignment of their velocities. In fact, in this setup, CIL has an effect very similar to velocity alignment interactions: an alternative model with velocity alignment instead of CIL produces very similar results (Supplementary Fig. 7), highlighting the similarity of these two interaction types in this setup. Secondly, repulsion ensures that boundary cells are repelled by the bulk of the cluster, which further rectifies their motion into a radially outward direction. Thus, both stronger CIL and stronger repulsive interactions lead to faster, more efficient spreading dynamics (Fig. 3 F, G).

However, we can distinguish the two types of interaction through the cross-correlation of cell velocities: this quantity serves as a good indicator for changes in CIL-behavior. Specifically, changing repulsive interactions has no significant effect on the correlation function, since it is a

position-dependent interaction (Fig. 3 H ii). In contrast, CIL is a velocity-dependent interaction, and its strength therefore controls the magnitude of the velocity cross-correlations (Fig. 3 H i).

Taken together, these results show that cell-cell interactions are key drivers of tissue spreading in this setup, and that disrupting cell-cell junctions through E- and N-Cadherin blocking has an effect akin to reducing repulsive interactions between cells. Therefore, the congruity between our experimental and modeling data suggests that both E- and N-Cadherin-mediated intercellular contacts create repulsive events via excluded volume interactions that are critical for the efficient cell spreading during collective migration. This effect could be due to cadherins 'sharpening' cell boundaries by for example regulating cell shape, improving cell-cell recognition, or increasing interfacial tension. Indeed, both E- and N-cadherin have been shown to determine inter-cellular interfacial tension in the developing epithelium [44, 68, 69]. These results are also in qualitative agreement with previous work where the interactions of colliding pairs of cells were inferred directly from observed trajectories [51]. Specifically, it was shown that the cancerous MDA-MB-231 cell line exhibits less repulsive interactions than the non-malignant MCF10A cell line, which is known exhibit higher E-cadherin expression than MDA-MB-231 cells [70, 71]. Our work therefore further supports the important role of cadherin-mediated cell-cell interactions, and elucidates their role in collective cell migration.

This study provides new insight into the role of different cadherin junctions in the dynamics of collective cancer cell migration. In our setup, we reveal that blocking E- or N-Cadherin in collectively migrating T24 cancer cells significantly reduces their spreading efficiency. The observed phenomenology is well captured by a biophysical model of stochastically migrating cells. Our model shows that cell proliferation as well as the excluded volume and Contact Inhibition of Locomotion interactions between cells drive tissue spreading in our setup. Our combined experimental and theoretical results further indicate that disrupting E- and N-Cadherin-mediated intercellular contacts leads to a decrease in repulsive cell-cell interactions, which in turn reduces the spreading efficiency of the cell collective. Therefore, from a biomedical point of view, this study underscores the importance of E- and N-Cadherins as potential pharmacological targets in metastatic cancer research. Furthermore, our experimental setup design could be adapted for future research in the field, such as studying the impact of mechanical cell-cell communication on cell spreading on mechanically compliant substrates [72-74], or chemotactic cell spreading in external gradients [75, 76].

Methods

T24 cell culture transfection with H2B-GFP plasmid for nucleus labeling.
H2B–GFP expression vectors, were obtained from Addgene (#11680). T24 cells exponentially growing in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum (FCS) were transfected with 2.5 μg of the H2B–GFP expression vector carrying a G418 resistance as selection marker, using an Amaxa R-Kit (Program I-013) under constant humidity at 37°C and 5% CO2. 24h after the transfection, cells were treated with G418 (A1720, Sigma-Aldrich) to an end concentration of 0.8mg/ml in 2ml well-plates and then further cultivated in T25 flasks and later on in T75 flasks with the same concentration of G418 (0.8mg/ml). After two rounds of additional cell sorting by flow cytometry the GFP+ cells at passage 30 were frozen in a nitrogen tank at a concentration of 1x 10^6 cells/ml.

For all collective migration experiments, T24 cells were pre-grown as monolayers and diluted down to the desired concentrations in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum (FCS), 10.000 U/ml penicillin/streptomycin and 0.8 mg/ml antibiotic G418 under constant humidity at 37°C and 5% CO_2.

Microcontact printing for circular pattern generation
8-well uncoated μ-Slides (ibidi, Martinsried, Germany) underwent 3 min of oxygen plasma treatment (Plasma cleaner typ ''ZEPTO,'' Diener electronic, Ebhausen, Germany) at 0.3 mbar for activation (generation of OH-hydroxyl bonds). Then, 250 μl/well of 0.05 mg/ml fibronectin (R&D Systems, US) solution in MilliQ were added to the now highly reactive surface and incubated at room temp for 2 hours. After washing 2 times with 500 μl of milliQ H_2O the surface was allowed to dry. Following that, we used standard microcontact printing techniques to create PDMS stamps with circular patterns. We placed one stamp at the center of each well and plasma treated the surface one last time at the same conditions as before. This step removes all fibronectin from the surface except the areas that are protected by the stamp, so all the unprotected areas on the surface become hydroxylated and highly reactive again. Without removing the stamps, we added a 7 μl drop of 1mg/ml PLL(20)-g[3.5]- PEG-N_3(3) (APP) (Susos AG, Switzerland) solution in MilliQ right next to each stamp allowing surface tension to absorb the liquid underneath the stamp. We let the above condition settle for 45 min. We gently removed the stamp and washed 2 times with 500 μl of MilliQ. Now the circular areas contain fibronectin and are highly cell-adhesive while the surrounding areas are initially cell repellent. At this point, T24 cells were trypsinized after reaching confluency, diluted to the desired density (70.000 cell/ml) in the aforementioned DMEM-based medium and 250 μl of this cell suspension were added in each well and allowed to settle overnight at 37 °C. The next day, the cell medium was replaced with 200 μl of fresh medium and the slide was placed under the microscope. Finally, 10 μl of 100μM BCN-cRGDfk (Synaffix, Netherlands) in PBS were added in the medium of each well to a final concentration of 20 μM. The BCN groups formed a link with the Azide groups of the APP-covered, cell-repellent areas around the colonies. This resulted in the binding of RGD on the surface, thereby rendering the surrounding areas cell adhesive and initiating cell migration.

Blocking antibody treatment of T24 circular colonies

For the blocking antibody treatment experiments, we followed the exact same cell preparation protocol as above with the addition of the following steps: On the next day, after the first washing step, 200 µl of 5mM EGTA solution were added in each well for 30 min. This step was performed in order to break the existing cadherin junctions and allow the blocking antibodies (anti N-Cadherin antibody: LEAF™ Purified anti-human CD325, #350804, Biolegend, USA; anti E-Cadherin antibody: CD324 #16-3249-82, Invitrogen, USA) to bind to their respective epitopes. Following that, the wells were washed two times with 200 µl of fresh cell medium. Subsequently, 200 µl of the appropriate E- (10 or 25µg/ml) or N- (25, 50 or 100µg/ml). Cadherin blocking antibody solution in cell medium were added in each well. Cells were incubated additionally for 30 min and then the slide was placed under the microscope. Finally, 10 µl of 100µM BCN-cRGDfk (Synaffix, Netherlands) in PBS were added in the medium of each well to a final concentration of 20 µM.

Cell imaging

Live cell imaging was performed using the T24 seeded 8-well fibronectin/APP patterned slides with an Eclipse Ti inverted microscope (Nikon, Dusseldorf, Germany) with a 4x/10x phase contrast objective and a CCD camera ([DS-Qi1Mc] Nikon, Dusseldorf, Germany). The slides were inserted into a 37 °C heating and incubation system that was flushed with actively mixed 5% CO_2 at a rate of 10 l/h, and the humidity was kept at 80% to prevent dehydration. The cells were imaged in bright-field and the fluorescence of the nuclei was detected at a 488 nm wavelength using the integrated fluorescence LED. Time-lapse video microscopy was performed with a time interval of 5 min between images over 24 h.

Tracking of single cell trajectories

The positions of individual cells were detected as previously described [43] using custom-made ImageJ macros implementing the 'Find Maxima' built-in function. The individual trajectories were then reconstructed using a squared-displacement minimization algorithm (http://site.physics.georgetown.edu/matlab) and data analysis was performed via custom-made Matlab programs.

qPCR

T24 cells were lysed for mRNA isolation. Briefly, "Buffer RLT, Lysis Buffer" (RNeasy® Mini Kit (250) PCR lab) was mixed with DTT 2M at a ratio of 50:1. After medium aspiration and ice-cold PBS rinsing, ice-cold lysis buffer was added and the lysates were stored at -80 °C. For the mRNA, isolation the RNeasy® Mini Kit (250) (QIAGEN, Hilden, Germany) was used according to the modified manufacturer's instructions. 2 µl of the mRNA samples was used directly for mRNA concentration determination using a Nanodrop® Spectrophotometer (PEQLAB Biotechnologie, Erlangen, Germany) with absorption at 260 nm (specific for mRNA) while impurities were determined at 280 nm. For the reverse transcription of mRNA to cDNA, 2X RT master mix was prepared containing: 10% TaqMan RT Puffer-10x, 0,04% dNTPs, 10% random

hexamers, 5% Reverse Transcriptase, 21%RNAase free water, 50% H2O+ RNA 2.5µg. For the quantitative PCR the following primers were obtained from metabion GmbH: E-Cadh_1_F (MM125, 5´TGG GCC AGG AAA TCA CAT CC3´), E-Cadh_1_R (MM126, 5´GGC ACC AGT GTC CGG ATT AA3´); N-Cadh_2_F (MM133, 5´CCT TTC AAA CAC AGC CAC GG3´), N-Cadh_2_R (MM134, 5´TGT TTG GGT CGG TCT GGA TG3´). We used 2 µl of the acquired cDNA in each well of the MicroAmp® Fast Optical 96-Well Reaction Plate or 2 µl of autoclaved Millipore H2O for the no-template controls (NTCs), respectively. 10.5 µl of PCR master mix containing 6.25 µl of PowerUPTM SYBR® Green Master Mix, 3.75 µl of autoclaved Millipore H2O, 0.25 µl of forward primer and 0.25 µl of reverse primer were added to each probe well and the qPCR was performed in a QuantStudio™ 3 Real-Time PCR system (ThermoFisher). Data were normalized to the housekeeping gene GAPDH. The analysis was carried out with the ΔΔCT method as previously described [77], using the ThermoFisher cloud and threshold cycle was set to > 9-15 and ≤ 30 to allow acceptable detection for best reproducibility.

Western Blots

Cells were harvested and lysed in RIPA lysis buffer containing a protease inhibitor mix (Roche #4693159001). Lysates were centrifuged at 10,000 x g for 10 min and 4 °C. Protein amounts were assessed by Bradford assay, and an equal amount of protein was separated by SDS-PAGE and transferred to nitrocellulose membranes (Hybond-ECLTM, Amersham Bioscience). Membranes were incubated with blocking buffer containing 5% BSA and 0.1% Tween 20 in PBS for 1h at room temperature, followed by 3x 5 min. rinsing with PBS-T. After that, membranes were incubated with rabbit anti-ECAD (24E10) monoclonal Ab (1:500; #3195, Cell Signaling Technology, Inc. USA) or rabbit anti-NCAD (D4R1H) XP® monoclonal Ab (1:500; #13116, Cell Signaling Technology Inc. USA) at 4°C overnight. Membranes were washed again with PBS-T 3 times for 5 min. Secondary antibody (HRP-Goat-Anti-Rabbit 1:1000; #111-035-144, Dianova, Germany) were used for 2h incubation at room temperature and subsequently conjugated with horseradish peroxidase and freshly prepared ECL solution (protected from light), which contained 2.5 mM luminol (detailed description of ECL solution preparation in table 1). Conjugated proteins were detected by the ChemiDoc™ Touch Imaging System (Bio-Rad, USA) and quantified by ImageLab software (Bio-Rad, USA). For quantification protein amount was normalized to total protein-loading, detected by 2,2,2-trichloroethanol activation as described previously [77] [78].

#	Reagent	Volume	Stock
1	distilled Water	4500µl	
2	Tris-Base pH 8.5	500µl	
3	p-coumaric acid	22µl	15mg/ml in DMSO (Aliquots at -20°C)
4	luminol	50µl	44mg/ml in DMSO (Aliquots at -20°C)
5	H2O2 30%	3µl	

Table 1 | Western Blot Solution Reagents

Cross-correlation functions of velocity fluctuations

To investigate the interactions of cells in the experiment, we calculate the spatial velocity cross-correlation function

$$C(r) = \frac{\sum_{ij} v_i \cdot v_j \delta(r - r_{ij})}{\sum_{ij} \delta(r - r_{ij})}$$

where v_i is the two-dimensional velocity vector of cell *i* and $\delta(r - r_{ij})$ is the Dirac delta-function. This function measures how 'similar' the velocities (magnitude and direction) of cells at distance *r* from one another are on average. Using discrete bins as an approximation for the delta-function for finite data, we obtain expected results both for experimental and simulated data.

The complete velocity field is composed of the collective outward motion, a dilatational mode, and additional velocity fluctuations due to interactions between the cells. Following previous work [79], we calculate these fluctuations by obtaining the scalar dilatation Λ as a function of time, by optimizing the quantity $\sum_i [x_i(t+T) - \Lambda x_i(t)]^2$. The fluctuation velocities are then giving by $u_i = [x_i(t+T) - \Lambda x_i(t)]/T$. Note that here, we use a time-interval $T = 15\Delta t$ which is larger than the time-resolution of the experiment. This allows us to average out the short-time scale noise fluctuations of the cellular velocities, and instead focusses on longer time-scale process relevant to the spreading dynamics. We test this approach in our simulations, and find that it accurately detects the presence of velocity-dependent interactions, such as CIL (Supplementary Figure 6).

Computational modeling

To provide a minimal computational model for the escape process, we implement a simple active particle model for collective cell migration. Similar to previous works [21, 22, 43, 80-82], we describe the motion of the cells using stochastic equations of motion with interactions. Specifically, we use the equation of motion

$$\frac{d\vec{v_i}}{dt} = -\gamma \vec{v_i} + \vec{F}_{\mathrm{rep}}(|r_i - r_j|) + \vec{F}^i{}_{\mathrm{CIL}} + \sigma \vec{\eta}_i(t)$$

where $\vec{\eta}_i(t)$ represents a Gaussian white noise with $\langle \vec{\eta}_i(t) \rangle = 0$ and $\langle \vec{\eta}_i(t) \vec{\eta}_j(t') \rangle = \delta(t - t')\delta_{ij}$. The model furthermore includes a persistence term $-\gamma v$, where γ^{-1} is the persistence time of the cells. The repulsive interactions are implemented as the repulsive part of a quadratic potential

$$\vec{F}_{\mathrm{rep}} = -\varepsilon(2\lambda - r_{ij}) \frac{\vec{r}_{ij}}{r_{ij}}$$

where λ represents the radius of the cells, and ε is the strength of the interaction.

The contact inhibition of locomotion (CIL) interaction $\vec{F}_{\mathrm{CIL}}$ is implemented in the form of a rotation of the velocity vector away from the distance vector $\vec{r}_{ij} = \vec{r}_j - \vec{r}_i$ to nearest neighbours, which are defined by being within an interaction range of radius 2.5λ, and being on collision course with cell *i*, i.e. $\vec{v}_i \cdot \vec{r}_{ij} > 0$. The angular displacement only depends on the velocity direction, a constant acceleration α and the

number as well as the positions of nearest neighbours: For each nearest neighbour the direction of the axis of rotation is found such that the rotation will be away from the nearest neighbours. All directions of the axes of rotations of all nearest neighbours are added up and multiplied by the acceleration α. Specifically, we use

$$\vec{F}^{i}{}_{\mathrm{CIL}} = \alpha \sum_{j} s_{ij} \left(\hat{e}_z \times \frac{\vec{v}_i}{|\vec{v}_i|} \right)$$

where

$$s_{ij} = \begin{cases} -\mathrm{sign}\left(\frac{\vec{v}_i}{|\vec{v}_i|} \times \vec{r}_{ij} \right) \cdot \hat{e_z}, & |\vec{r}_{ij}| < 2.5\lambda \text{ and } \vec{v}_i \cdot \vec{r}_{ij} > 0 \\ 0, & \text{otherwise} \end{cases}$$

In simulations where velocity alignment rather than CIL is used (Supplementary Figure 8) , we replace $\vec{F}_{\mathrm{CIL}}$ by an alignment interaction $\beta \vec{V}_i / |\vec{V}_i|$ with strength β, which is implemented as a constant acceleration in the direction of the average velocity $\vec{V}_i = \langle \vec{\mathrm{v}}_{\mathrm{j}} \rangle_{j \in NN_i}$ of nearest neighbours within an interaction range of radius 2.5λ.

Finally, cell division is implemented with a constant probability $\nu \mathrm{d}t$ of dividing, provided there is sufficient space for the appearance of new cells. In a division event, a cell produces a daughter cell in its direct neighborhood with an initial velocity pointing away from the mother cell.

The simulation is performed in non-dimensional units such that $\gamma^{-1} = \lambda = 1$. We use the parameters $\sigma^2 = 2$, $\nu = 0.1$, and vary ε between 0.1 and 40, and α between 0 and 12. We initialize $N = 37$ particles within the initial confinement radius R. The stochastic trajectories of the model are then simulated by step-wise Euler updates with a time-step of $dt = 10^{-3}$. We first perform a pre-equilibration run with a confinement potential at $r = R$, modelling the initial confinement phase. At $t = 0$, we remove the boundary by setting the confinement potential to zero, leading to the escape of the simulated cluster.

Statistical evaluation

For statistical analysis of the data one-way ANOVA followed by Dunnett's multiple comparisons test was performed using GraphPad Prism version 8.0.0 for Windows, (GraphPad Software, San Diego, California USA, www.graphpad.com). n.s.= not significant, * p < 0.05, ** p < 0.01.

Author Contributions

T.Z., D.B.B., C.P.B. and S.Z. designed the study. T.Z. performed all experiments. J.A. contributed tracking software. D.B.B. and T.Z. analyzed data. D.B.B. and T.B. developed the theoretical model. T.Z. and D.B.B. wrote the paper with input from all authors.

Acknowledgements

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project-ID 201269156 - SFB 1032 (Projects B8 and B12). D.B.B. is supported in part by a DFG fellowship within the Graduate School of Quantitative Biosciences Munich (QBM) and by the Joachim Herz Stiftung.

References

1. Friedl, P. and D. Gilmour, *Collective cell migration in morphogenesis, regeneration and cancer.* Nature Reviews Molecular Cell Biology, 2009. **10**(7): p. 445-457.
2. Rey-Barroso, J., et al., *Switching between individual and collective motility in B lymphocytes is controlled by cell-matrix adhesion and inter-cellular interactions.* Scientific Reports, 2018. **8**(1): p. 5800.
3. Grada, A., et al., *Research Techniques Made Simple: Analysis of Collective Cell Migration Using the Wound Healing Assay.* Journal of Investigative Dermatology, 2017. **137**(2): p. e11-e16.
4. Wood, J.M. and M.F. Olson, *Collective migration: spatial tension relief.* Curr Biol, 2012. **22**(4): p. R125-7.
5. Trepat, X., Z. Chen, and K. Jacobson, *Cell migration.* Compr Physiol, 2012. **2**(4): p. 2369-92.
6. De Pascalis, C. and S. Etienne-Manneville, *Single and collective cell migration: the mechanics of adhesions.* Molecular biology of the cell, 2017. **28**(14): p. 1833-1846.
7. George, M., F. Bullo, and O. Campàs, *Connecting individual to collective cell migration.* Scientific Reports, 2017. **7**(1): p. 9720.
8. Ladoux, B. and R.M. Mège, *Mechanobiology of collective cell behaviours.* Nat Rev Mol Cell Biol, 2017. **18**(12): p. 743-757.
9. Etienne-Manneville, S., *Adherens junctions during cell migration.* Subcell Biochem, 2012. **60**: p. 225-49.
10. Gail, M.H. and C.W. Boone, *The locomotion of mouse fibroblasts in tissue culture.* Biophys J, 1970. **10**(10): p. 980-93.
11. Metzner, C., et al., *Superstatistical analysis and modelling of heterogeneous random walks.* Nature Communications, 2015. **6**(1): p. 7516.
12. Selmeczi, D., et al., *Cell Motility as Persistent Random Motion: Theories from Experiments.* Biophysical Journal, 2005. **89**(2): p. 912-931.

13. Brückner, D.B., et al., *Stochastic nonlinear dynamics of confined cell migration in two-state systems.* Nature Physics, 2019. **15**(6): p. 595-601.
14. Fink, A., et al., *Area and Geometry Dependence of Cell Migration in Asymmetric Two-State Micropatterns.* Biophysical Journal, 2020. **118**(3): p. 552-564.
15. Brückner, D.B., et al., *Disentangling the behavioural variability of confined cell migration.* Journal of The Royal Society Interface, 2020. **17**(163): p. 20190689.
16. Dyson, L. and R.E. Baker, *The importance of volume exclusion in modelling cellular migration.* J Math Biol, 2015. **71**(3): p. 691-711.
17. Abercrombie, M. and J.E.M. Heaysman, *Observations on the social behaviour of cells in tissue culture: II. "Monolayering" of fibroblasts.* Experimental Cell Research, 1954. **6**(2): p. 293-306.
18. Carmona-Fontaine, C., et al., *Contact inhibition of locomotion in vivo controls neural crest directional migration.* Nature, 2008. **456**(7224): p. 957-961.
19. Camley, B.A. and W.-J. Rappel, *Physical models of collective cell motility: from cell to tissue.* Journal of physics D: Applied physics, 2017. **50**(11): p. 113002.
20. Giniūnaitė, R., et al., *Modelling collective cell migration: neural crest as a model paradigm.* Journal of Mathematical Biology, 2020. **80**(1): p. 481-504.
21. Alert, R. and X. Trepat, *Physical Models of Collective Cell Migration.* Annual Review of Condensed Matter Physics, 2020. **11**(1): p. 77-101.
22. Brückner DB, e.a., *Learning the dynamics of cell-cell interactions in confined cell migration. .* Arxiv:2008.03978, 2020.
23. Schnyder, S.K., J.J. Molina, and R. Yamamoto, *Control of cell colony growth by contact inhibition.* Scientific Reports, 2020. **10**(1): p. 6713.
24. Paulson, A., et al., *Regulation of cadherin expression in nervous system development.* Cell adhesion & migration, 2014. **8**: p. 19-28.
25. Halbleib, J.M. and W.J. Nelson, *Cadherins in development: cell adhesion, sorting, and tissue morphogenesis.* Genes Dev, 2006. **20**(23): p. 3199-214.
26. Shapiro, L. and W.I. Weis, *Structure and biochemistry of cadherins and catenins.* Cold Spring Harbor perspectives in biology, 2009. **1**(3): p. a003053-a003053.
27. Campbell, K. and J. Casanova, *A common framework for EMT and collective cell migration.* Development, 2016. **143**(23): p. 4291.
28. Loh, C.-Y., et al., *The E-Cadherin and N-Cadherin Switch in Epithelial-to-Mesenchymal Transition: Signaling, Therapeutic Implications, and Challenges.* Cells, 2019. **8**(10): p. 1118.
29. Yu, W., et al., *Cadherin Signaling in Cancer: Its Functions and Role as a Therapeutic Target.* Frontiers in Oncology, 2019. **9**(989).
30. Wheelock, M.J., et al., *Cadherin switching.* Journal of Cell Science, 2008. **121**(6): p. 727.
31. Pérez-González, C., et al., *Active wetting of epithelial tissues.* Nature Physics, 2019. **15**(1): p. 79-88.
32. Padmanaban, V., et al., *E-cadherin is required for metastasis in multiple models of breast cancer.* Nature, 2019. **573**(7774): p. 439-444.

33. Bindels, E.M.J., et al., *E-Cadherin Promotes Intraepithelial Expansion of Bladder Carcinoma Cells in an in Vitro Model of Carcinoma in Situ.* Cancer Research, 2000. **60**(1): p. 177.
34. Putzke, A.P., et al., *Metastatic progression of prostate cancer and e-cadherin regulation by zeb1 and SRC family kinases.* Am J Pathol, 2011. **179**(1): p. 400-10.
35. Reddy, P., et al., *Formation of E-cadherin-mediated cell-cell adhesion activates AKT and mitogen activated protein kinase via phosphatidylinositol 3 kinase and ligand-independent activation of epidermal growth factor receptor in ovarian cancer cells.* Mol Endocrinol, 2005. **19**(10): p. 2564-78.
36. Lewis-Tuffin, L.J., et al., *Misregulated E-cadherin expression associated with an aggressive brain tumor phenotype.* PloS one, 2010. **5**(10): p. e13665-e13665.
37. Su, Y., et al., *N-cadherin functions as a growth suppressor in a model of K-ras-induced PanIN.* Oncogene, 2016. **35**(25): p. 3335-41.
38. Lammens, T., et al., *N-cadherin in neuroblastoma disease: expression and clinical significance.* PLoS One, 2012. **7**(2): p. e31206.
39. Elie-Caille, C., et al., *Molecular and nanoscale evaluation of N-cadherin expression in invasive bladder cancer cells under control conditions or GW501516 exposure.* Molecular and Cellular Biochemistry, 2020. **471**.
40. Lascombe, I., et al., *N-Cadherin as a Novel Prognostic Marker of Progression in Superficial Urothelial Tumors.* Clinical Cancer Research, 2006. **12**(9): p. 2780.
41. Rebel, J.M., et al., *E-cadherin expression determines the mode of replacement of normal urothelium by human bladder carcinoma cells.* Cancer Res, 1994. **54**(20): p. 5488-92.
42. van Dongen, S.F.M., et al., *Triggering Cell Adhesion, Migration or Shape Change with a Dynamic Surface Coating.* Advanced Materials, 2013. **25**(12): p. 1687-1691.
43. d'Alessandro, J., et al., *Contact enhancement of locomotion in spreading cell colonies.* Nature Physics, 2017. **13**(10): p. 999-1005.
44. Maître, J.L. and C.P. Heisenberg, *Three functions of cadherins in cell adhesion.* Curr Biol, 2013. **23**(14): p. R626-33.
45. Manova, R., T.A. van Beek, and H. Zuilhof, *Surface functionalization by strain-promoted alkyne-azide click reactions.* Angew Chem Int Ed Engl, 2011. **50**(24): p. 5428-30.
46. Bubeník, J., et al., *Established cell line of urinary bladder carcinoma (T24) containing tumour-specific antigen.* International Journal of Cancer, 1973. **11**(3): p. 765-773.
47. Lang, K., et al., *Signal processing in migrating T24 human bladder carcinoma cells: role of the autocrine interleukin-8 loop.* Int J Cancer, 2002. **99**(5): p. 673-80.
48. Jin, H., et al., *Divergent behaviors and underlying mechanisms of cell migration and invasion in non-metastatic T24 and its metastatic derivative T24T bladder cancer cell lines.* Oncotarget, 2015. **6**(1): p. 522-536.
49. Luo, K.-W., et al., *EGCG inhibited bladder cancer T24 and 5637 cell proliferation and migration via PI3K/AKT pathway.* Oncotarget, 2018. **9**(15): p. 12261-12272.

50. McBeth, L., et al., *Glucocorticoid receptor beta increases migration of human bladder cancer cells.* Oncotarget, 2016. **7**(19): p. 27313-24.
51. Brückner, D.B., et al., *Learning the dynamics of cell–cell interactions in confined cell migration.* Proceedings of the National Academy of Sciences, 2021. **118**(7): p. e2016602118.
52. Cates, M.E. and J. Tailleur, *When are active Brownian particles and run-and-tumble particles equivalent? Consequences for motility-induced phase separation.* EPL (Europhysics Letters), 2013. **101**: p. 20010.
53. GM., C., *Cell-Cell Interactions.*, in *The Cell: A Molecular Approach*. 2000: Sunderland (MA).
54. Nieman, M.T., et al., *N-Cadherin Promotes Motility in Human Breast Cancer Cells Regardless of Their E-Cadherin Expression.* Journal of Cell Biology, 1999. **147**(3): p. 631-644.
55. Kuphal, S. and A.K. Bosserhoff, *Influence of the cytoplasmic domain of E-cadherin on endogenous N-cadherin expression in malignant melanoma.* Oncogene, 2006. **25**(2): p. 248-259.
56. Reckamp, K.L., et al., *Tumor response to combination celecoxib and erlotinib therapy in non-small cell lung cancer is associated with a low baseline matrix metalloproteinase-9 and a decline in serum-soluble E-cadherin.* J Thorac Oncol, 2008. **3**(2): p. 117-24.
57. Oh, D.Y., et al., *Phase II trial of dacomitinib in patients with HER2-positive gastric cancer.* Gastric Cancer, 2016. **19**(4): p. 1095-1103.
58. Bodnar, L., et al., *Role of WNT/β-Catenin Pathway as Potential Prognostic and Predictive Factors in Renal Cell Cancer Patients Treated With Everolimus in the Second and Subsequent Lines.* Clin Genitourin Cancer, 2018. **16**(4): p. 257-265.
59. Kuefer, R., et al., *The role of an 80 kDa fragment of E-cadherin in the metastatic progression of prostate cancer.* Clin Cancer Res, 2003. **9**(17): p. 6447-52.
60. Schmeiser, K. and R.J. Grand, *The fate of E- and P-cadherin during the early stages of apoptosis.* Cell Death Differ, 1999. **6**(4): p. 377-86.
61. Trillsch, F., et al., *E-Cadherin fragments as potential mediators for peritoneal metastasis in advanced epithelial ovarian cancer.* Br J Cancer, 2016. **114**(2): p. 213-20.
62. Tang, M.K.S., et al., *Soluble E-cadherin promotes tumor angiogenesis and localizes to exosome surface.* Nat Commun, 2018. **9**(1): p. 2270.
63. Grabowska, M.M. and M.L. Day, *Soluble E-cadherin: more than a symptom of disease.* Front Biosci (Landmark Ed), 2012. **17**: p. 1948-64.
64. Mialhe, A., et al., *Expression of E-, P-, n-cadherins and catenins in human bladder carcinoma cell lines.* J Urol, 2000. **164**(3 Pt 1): p. 826-35.
65. Elie-Caille, C., et al., *Molecular and nanoscale evaluation of N-cadherin expression in invasive bladder cancer cells under control conditions or GW501516 exposure.* Molecular and Cellular Biochemistry, 2020. **471**(1): p. 113-127.
66. Theveneau, E. and R. Mayor, *Integrating chemotaxis and contact-inhibition during collective cell migration: Small GTPases at work.* Small GTPases, 2010. **1**(2): p. 113-117.
67. Vedula, S.R.K., et al., *Emerging modes of collective cell migration induced by geometrical constraints.* Proceedings of the National Academy of Sciences, 2012. **109**(32): p. 12974.

68. Käfer, J., et al., *Cell adhesion and cortex contractility determine cell patterning in the Drosophila retina.* Proceedings of the National Academy of Sciences, 2007. **104**(47): p. 18549.
69. Hayashi, T. and R.W. Carthew, *Surface mechanics mediate pattern formation in the developing retina.* Nature, 2004. **431**(7009): p. 647-652.
70. Milano, D.F., et al., *Regulators of Metastasis Modulate the Migratory Response to Cell Contact under Spatial Confinement.* Biophys J, 2016. **110**(8): p. 1886-1895.
71. Sommers, C.L., et al., *Cell adhesion molecule uvomorulin expression in human breast cancer cell lines: relationship to morphology and invasive capacities.* Cell growth & differentiation : the molecular biology journal of the American Association for Cancer Research, 1991. **2**(8): p. 365-372.
72. Dietrich, M., et al., *Guiding 3D cell migration in deformed synthetic hydrogel microstructures.* Soft Matter, 2018. **14**(15): p. 2816-2826.
73. Reinhart-King, C.A., M. Dembo, and D.A. Hammer, *Cell-cell mechanical communication through compliant substrates.* Biophys J, 2008. **95**(12): p. 6044-51.
74. Humphries, D.L., J.A. Grogan, and E.A. Gaffney, *Mechanical Cell–Cell Communication in Fibrous Networks: The Importance of Network Geometry.* Bulletin of Mathematical Biology, 2017. **79**(3): p. 498-524.
75. Majumdar, R., M. Sixt, and C.A. Parent, *New paradigms in the establishment and maintenance of gradients during directed cell migration.* Current opinion in cell biology, 2014. **30**: p. 33-40.
76. Camley, B.A., *Collective gradient sensing and chemotaxis: modeling and recent developments.* J Phys Condens Matter, 2018. **30**(22): p. 223001.
77. Fleige, S., et al., *Comparison of relative mRNA quantification models and the impact of RNA integrity in quantitative real-time RT-PCR.* Biotechnol Lett, 2006. **28**(19): p. 1601-13.
78. Chopra, A., W.G. Willmore, and K.K. Biggar, *Protein quantification and visualization via ultraviolet-dependent labeling with 2,2,2-trichloroethanol.* Scientific Reports, 2019. **9**(1): p. 13923.
79. Attanasi, A., et al., *Collective Behaviour without Collective Order in Wild Swarms of Midges.* PLOS Computational Biology, 2014. **10**(7): p. e1003697.
80. Basan, M., et al., *Alignment of cellular motility forces with tissue flow as a mechanism for efficient wound healing.* Proceedings of the National Academy of Sciences, 2013. **110**(7): p. 2452.
81. Garcia, S., et al., *Physics of active jamming during collective cellular motion in a monolayer.* Proceedings of the National Academy of Sciences, 2015. **112**(50): p. 15314.
82. Sepúlveda, N., et al., *Collective Cell Motion in an Epithelial Sheet Can Be Quantitatively Described by a Stochastic Interacting Particle Model.* PLOS Computational Biology, 2013. **9**(3): p. e1002944.

Supplementary Information for

E- and N-Cadherin-mediated intercellular contacts enhance collective spreading of migrating cancer cells

Themistoklis Zisis[1]*, David B. Brückner[2]*, Tom Brandstätter[2], Joseph d'Alessandro[3], Angelika M. Vollmar[1], Chase P. Broedersz[2,4,+], Stefan Zahler[1,+]

[1]*Ludwig-Maximilians-University Munich, Department of Pharmacy, Center for Drug Research, 81377 Munich, Germany*

[2]*Arnold Sommerfeld Center for Theoretical Physics and Center for NanoScience, Department of Physics, Ludwig-Maximilians-University Munich*

[3]*Institut Jacques Monod (IJM), CNRS UMR 7592 and Université de Paris, 75013, Paris, France*

[4]*Department of Physics and Astronomy, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands*

* authors contributed equally

[+] corresponding authors

Movie Descriptions

Supplementary Movie S1: Video of bright-field microscopy imaging showing the control (untreated) T24 cell migration from 0h to 24h after surface activation.

Supplementary Movie S2: Video of fluorescence microscopy imaging showing the control (untreated) T24 cell migration from 0h to 24h after surface activation.

Supplementary Movie S3: Overlay video of bright-field and fluorescence microscopy imaging with cell tracks showing the control (untreated) T24 cell migration from 0h to 24h after surface activation.

Supplementary Movie S4: Overlay video of bright-field and fluorescence microscopy imaging with cell tracks showing the T24 cell migration upon 25 µg/ml E-Cadherin blocking, from 0h to 24h after surface activation.

Supplementary Movie S5: Overlay video of bright-field and fluorescence microscopy imaging with cell tracks showing the T24 cell migration upon 100 µg/ml N-Cadherin blocking, from 0h to 24h after surface activation.

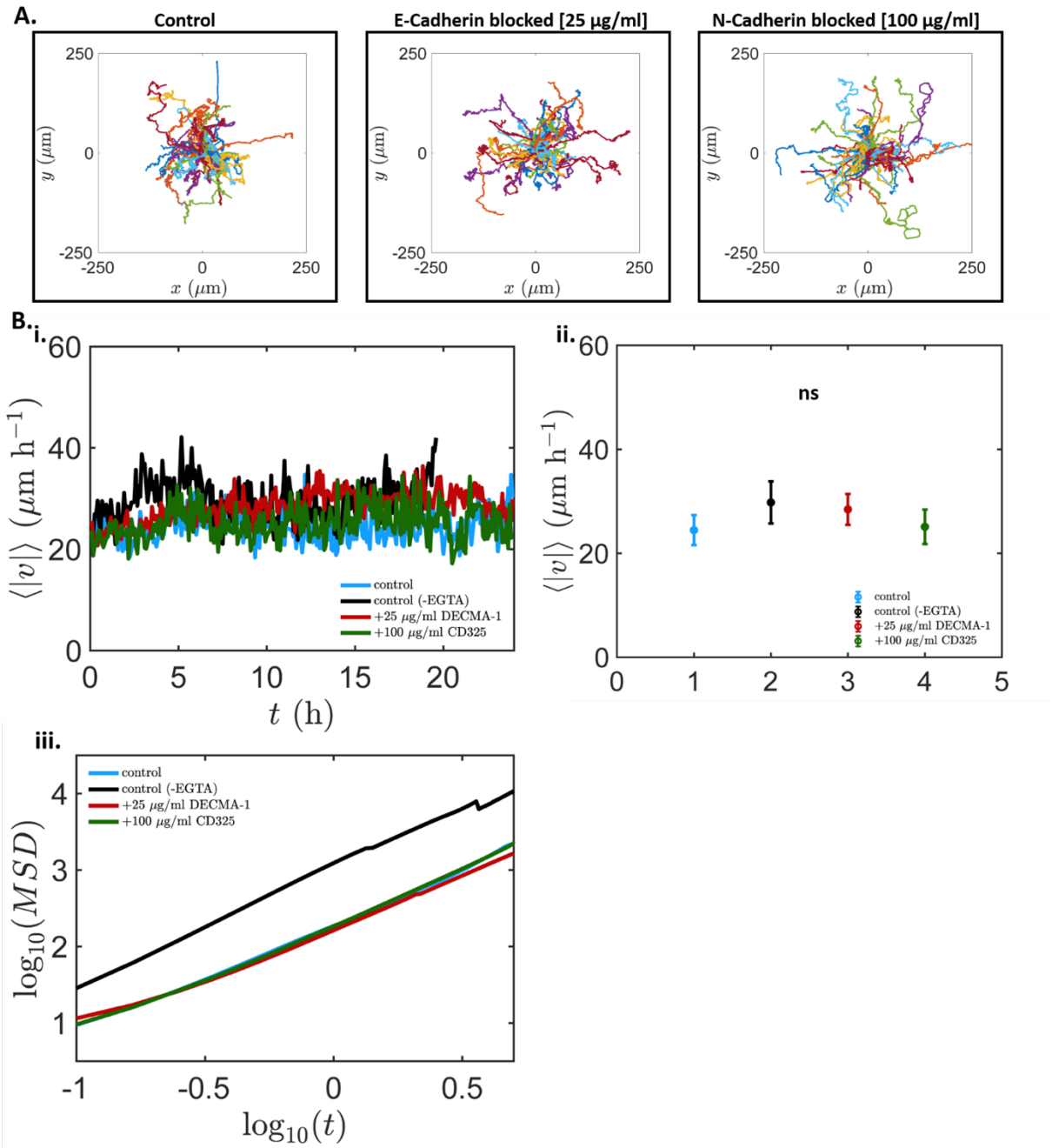


Figure S1 | Single cell trajectories, radial velocities and SEM for single T24 cell migration in the different blocking conditions. A) Single T24 cell trajectories in the control condition and at the highest blocking antibody concentration for each cadherin type. B) i) Mean radial velocity (u_r) over time and (ii) corresponding SEM graph showing no significant differences (1-way ANOVA, $p>0.05$) between averaged cell velocities of single cells for every condition. Average velocity of single cells was stable and was not affected by the addition of the different antibodies or EGTA pre-treatment. iii) Mean square displacement (MSD) plot showing all conditions having a 1.3 curve gradient.

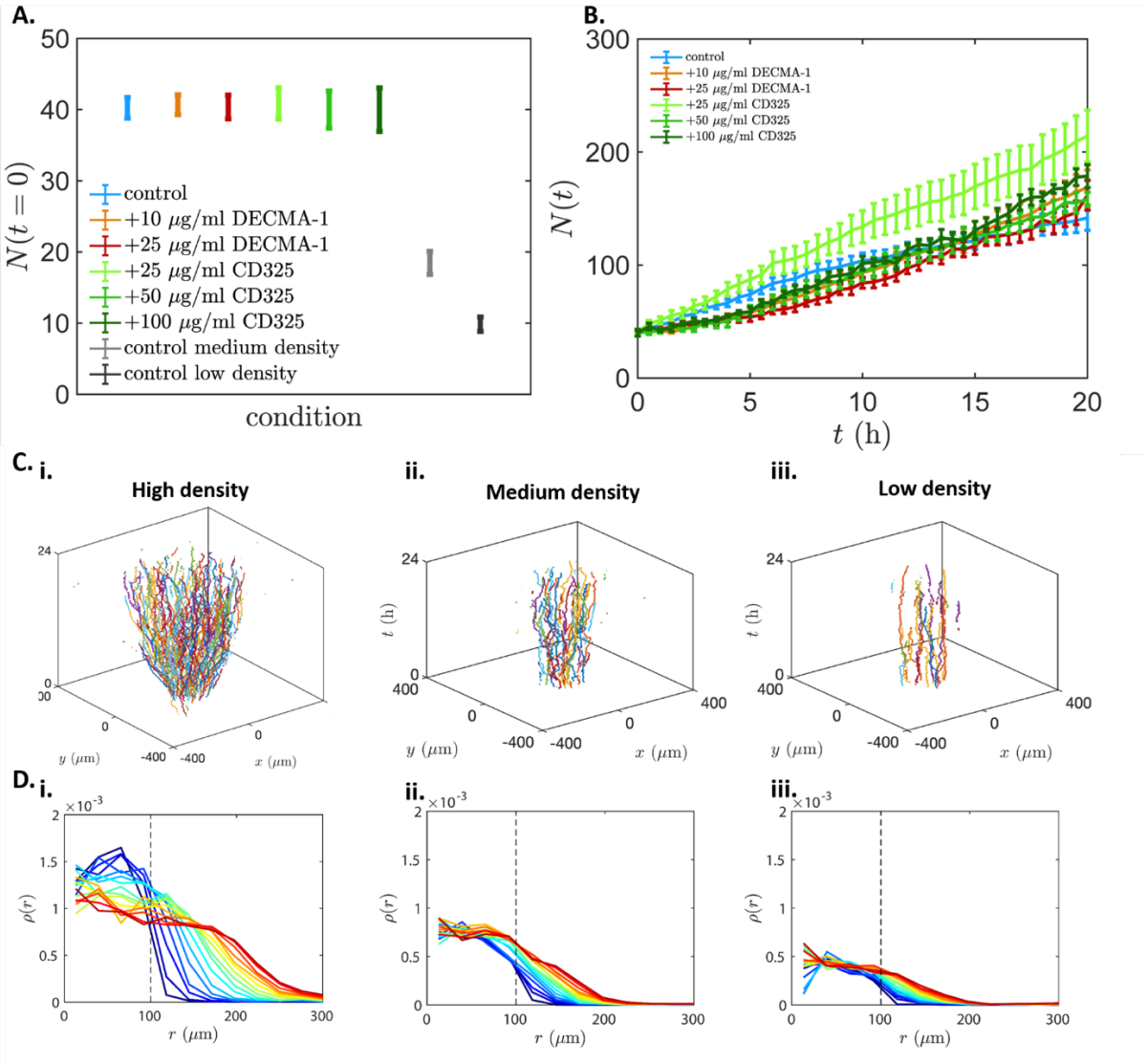


Figure S2 | Initial number of cells of each colony for every condition and T24 cell proliferation in the different blocking conditions, followed by space diagrams and evolution of density profiles for colonies with different cell densities. A) Number of cells (t=0) for each colony per condition, color-coded according to the colony's cell density. For every condition we ensured constant average initial cell density with an average cell number of ~40, except medium and low cell density control conditions. B) Cell proliferation shown as the average total number of cells of all colonies for each blocking condition. In all conditions, except the 25 µg/ml N-Cadherin blocking and the combination blocking, the proliferation rate was not affected by treatment with blocking antibodies. C) Space diagram of (i) a high cell density, (ii) a medium cell density and (iii) a low cell density colony spreading up to 24h after surface activation. D) Evolution of the density profiles of (i) high cell density, (ii) medium cell density and (iii) low cell density colonies over time (blue to red) plotted as the mean of all colonies (n_{high}= 12, n_{medium}= 15, n_{low}= 12). All curves are separated by 1.5 h intervals.

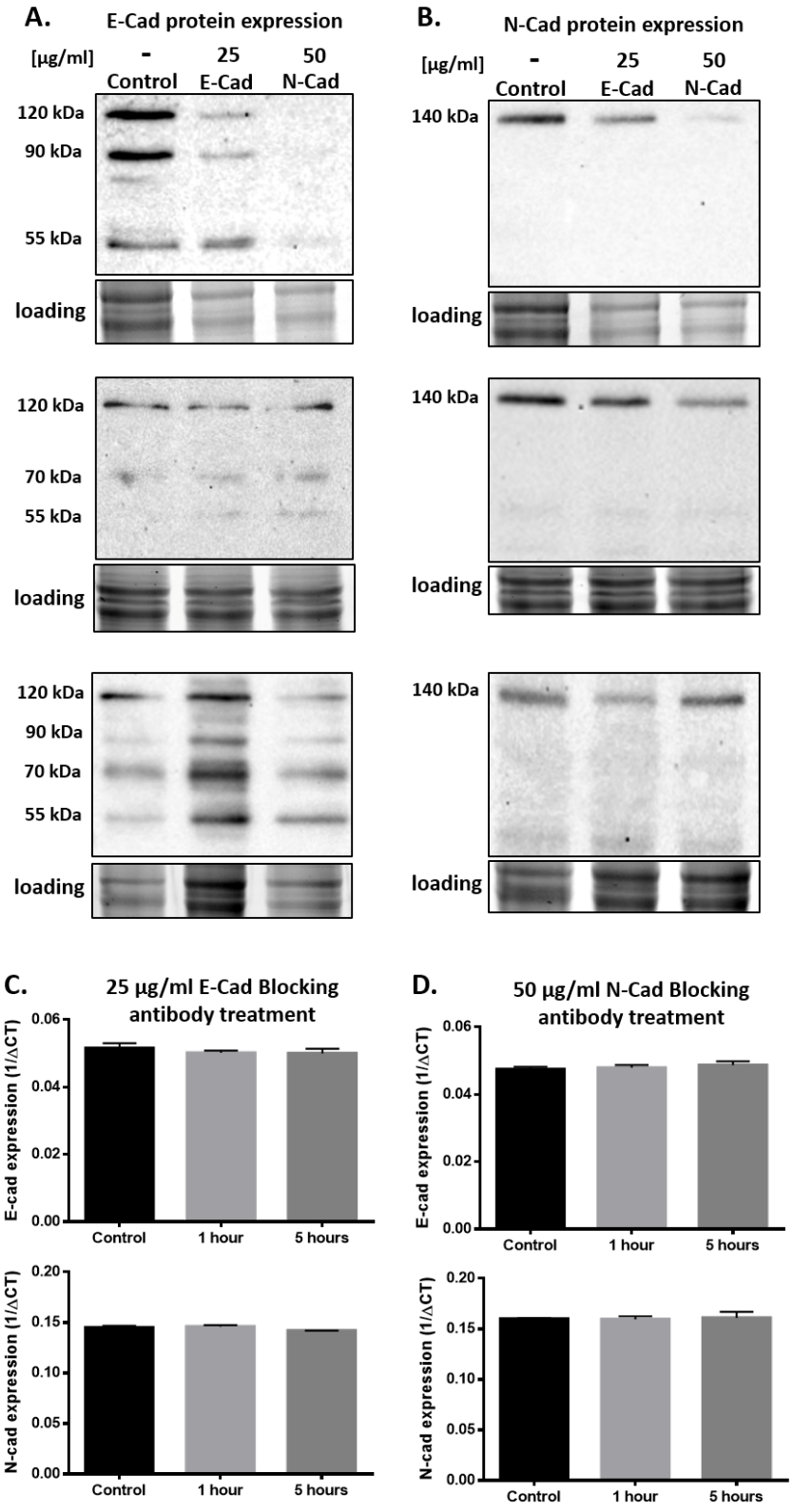

A.
E-Cad protein expression
[µg/ml]
-
25
50
Control
E-Cad
N-Cad
120 kDa
90 kDa
55 kDa
loading
120 kDa
70 kDa
55 kDa
loading
120 kDa
90 kDa
70 kDa
55 kDa
loading
B.
N-Cad protein expression
[µg/ml]
-
25
50
Control
E-Cad
N-Cad
140 kDa
loading
140 kDa
loading
140 kDa
loading
C.
25 µg/ml E-Cad Blocking antibody treatment
E-cad expression (1/ΔCT)
0.06
0.04
0.02
0.00
Control
1 hour
5 hours
N-cad expression (1/ΔCT)
0.20
0.15
0.10
0.05
0.00
Control
1 hour
5 hours
D.
50 µg/ml N-Cad Blocking antibody treatment
E-cad expression (1/ΔCT)
0.06
0.04
0.02
0.00
Control
1 hour
5 hours
N-cad expression (1/ΔCT)
0.20
0.15
0.10
0.05
0.00
Control
1 hour
5 hours

Figure S3 | Complete Western blot triplicates and qPCR results. A) E- cadherin protein levels in untreated (control) or treated T24 cells with 25μg/ml E-Cadherin or 50μg/ml N-Cadherin blocking antibody after 24 hours. WB triplicates show the different band sizes occurring and the accompanying protein loading, with the 120kDa being the functional protein observed in all three WBs. B) N- cadherin protein levels in untreated (control) or treated T24 cells with 25μg/ml E-Cadherin or 50μg/ml N-Cadherin blocking antibody after 24 hours. WB triplicates show the 140kDa band size occurring and the accompanying protein loading. In all cases, to calculate protein expression levels all band intensities were calibrated according to control using the loading band intensities. C) No significant differences in gene expression levels of E- and N-Cadherin at 1h and 5h upon 25μg/ml E-Cadherin blocking antibody treatment as determined by qPCR (E-Cadherin: 1-way ANOVA $F=0.5435$, $p=0.6068$; N-Cadherin: 1-way ANOVA $F= 3.519$, $p=0.0974$). D) No significant differences in gene expression levels of E- and N-Cadherin at 1h and 5h upon 50μg/ml N-Cadherin blocking antibody treatment as determined by qPCR (E-Cadherin: 1-way ANOVA $F= 0.4286$, $p=0.6699$; N-Cadherin: 1-way ANOVA $F= 0.04005$, $p=0.9610$). In all cases the levels of N-Cadherin expression are ~3-fold higher than the levels of E-Cadherin expression.

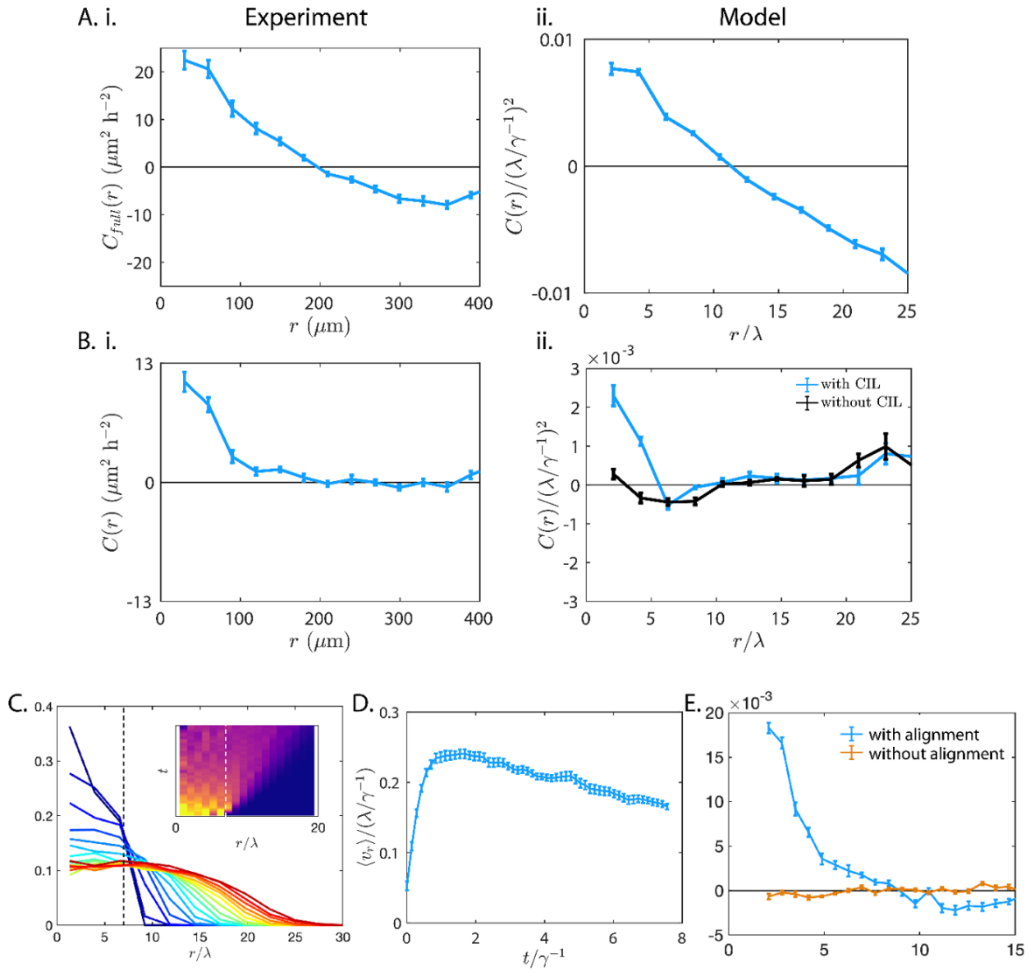


Figure S4 | Calculation of the spatial fluctuation velocity cross-correlation function and comparison to a model with velocity alignment interactions. A) Correlation function of the full velocities for (i) experiment and (ii) model with CIL. As expected, the correlation function is initially positive, corresponding to neighboring cells on average moving in the same direction. After a distance on the order of the initial confinement radius, the function turns negative, corresponding to cells on opposite ends of the cluster moving on average in opposite directions. B) Correlation function of the velocity fluctuations, where the overall dilatation of the cluster is subtracted. For the model, corresponding curves for a simulation with CIL (blue), and without CIL interactions (black) is shown. As expected, in both cases, the negative part of the correlation due to the overall dilatation of the cluster disappears and only simulations with CIL exhibit significant fluctuation velocity correlations. C) Evolution of the density profile over time (blue to red). Inset: Kymograph of the cell density evolution. Dashed lines indicate the initial confinement radius. D) Mean radial velocity as a function of time. E) Cross-correlation of velocity fluctuations, for a model with and without velocity alignment interactions. Error bars: SEM; n=30 for all panels.

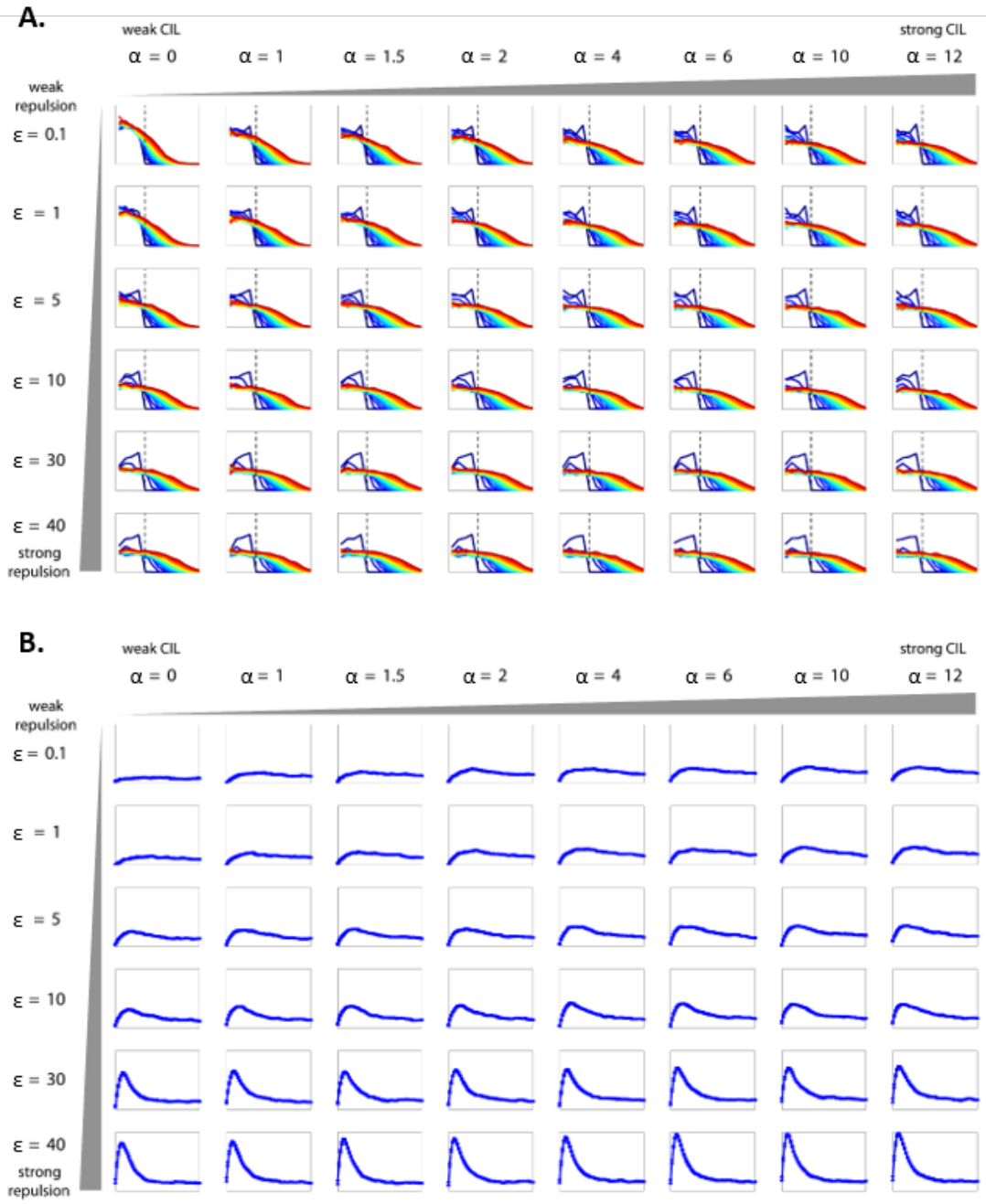


Figure S5 | Full parameter sweep: density profiles and radial velocity curves. A) Evolution of the density profile over time (blue to red) for all parameter combinations of repulsion strength ε and CIL interaction amplitude α, averaged over n=30 clusters per condition. The profiles exhibit further spreading for larger repulsions and larger CIL amplitudes. B) Mean radial velocity as a function of time for all parameter combinations of repulsion strength ε and CIL interaction amplitude α. We generally observe larger radial velocity peaks for larger repulsions and larger CIL amplitudes. Error bars: SEM; n=30 for all panels.

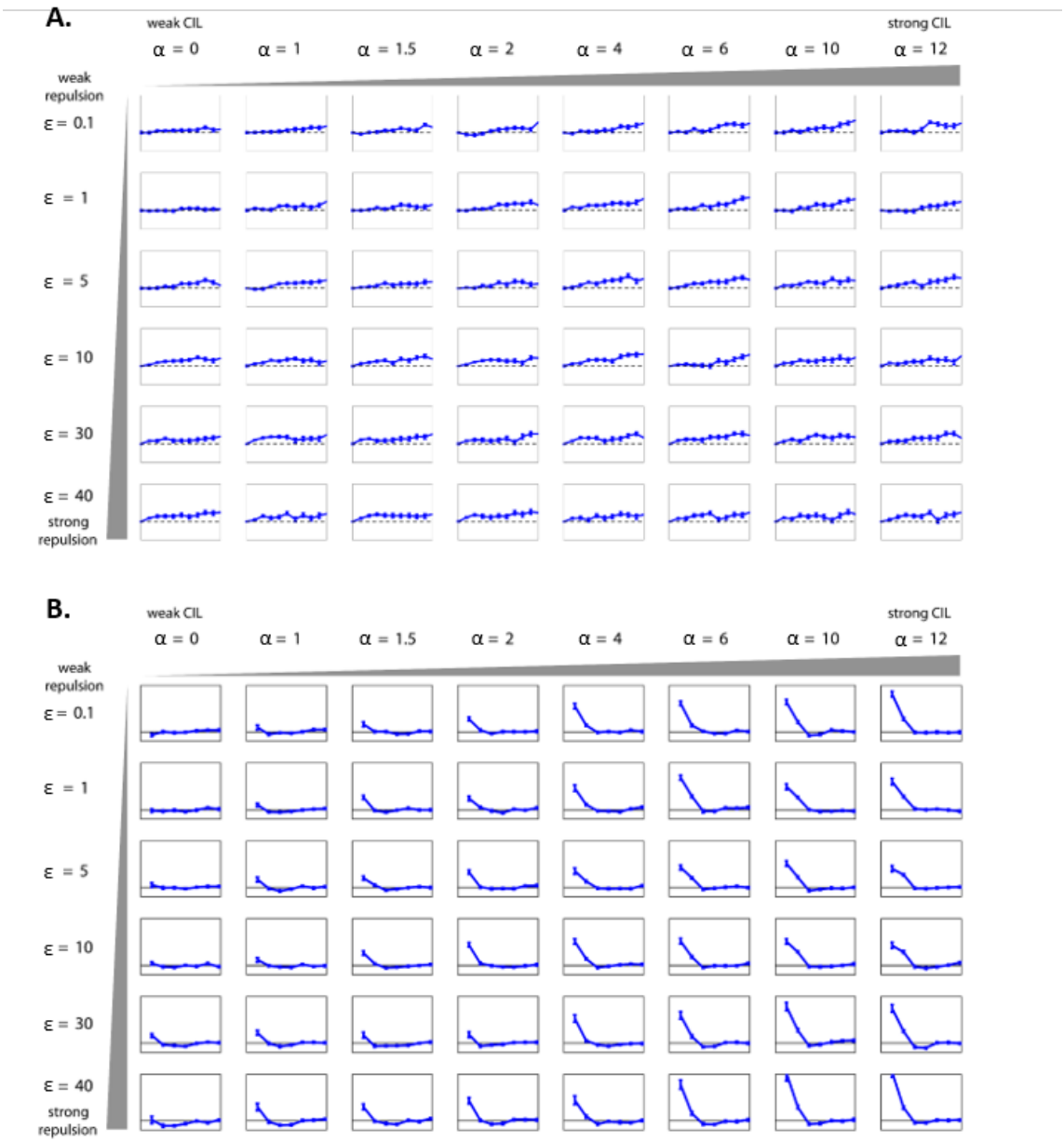


Figure S6 | Full parameter sweep: spreading radius and velocity fluctuation cross-correlation function. A) Average distance where density has decayed to half of its value in the center of the original confinement (i.e. at r=0) for all parameter combinations of repulsion strength ε and CIL interaction amplitude α. We observe larger spreading radii for larger repulsions and larger CIL amplitudes. Error bars: SEM; n=30 for all panels. B) Cross-correlation of velocity fluctuations for all parameter combinations of repulsion strength ε and CIL interaction amplitude α. We generally observe larger radial velocity peaks for larger repulsions and larger CIL amplitudes. Error bars: SEM; n=30 for all panels.

Acknowledgements

The work presented in this thesis is the product of an immense collaborative effort, and it is a great pleasure to conclude it by thanking everyone who I had the privilege to work and interact with throughout my PhD. While science is objective, I strongly believe that scientific research itself is much more personal than it is often portrayed. Research is done by people − people with tastes, preferences, and their own way of doing things. The work in this thesis has therefore been shaped by the people around me as much as by myself, and much of the credit for the ideas and concepts presented here goes to them.

First of all, I would like to express my gratitude to my PhD advisor Chase Broedersz. Chase, I would have to turn this section into a chapter to list everything I would like to thank you for. I am deeply grateful for your immense support over the years, for giving me the freedom to explore in all kinds of directions, and for trusting in my visions and ideas. From you I have learnt how to speak, how to write and how to think. Most importantly, you have taught me that science relies on collaboration, and how to turn collaborations into something that is not just productive and intellectually stimulating, but also a lot of fun. Thank you for making me part of the amazing research environment you have created right from day one.

I am also very thankful to Joachim Rädler. Joachim, you have been the best co-advisor I could have ever asked for. Thank you for your immense generosity in including me in your research endeavours. It was truly inspiring to work with you and learn from your ideas and intuitions. I would also like to thank you for all the energy and committment you have spent on making SFB1032 what it is − an amazing environment to start a scientific career.

The title of this thesis already gives away to whom I owe immense gratitude: a data-driven approach is not much fun without data. Alexandra, working with you has been a pleasure throughout the years, and I could not imagine what my PhD would have been like without you. Thank you for our fun brainstorming sessions at the Victorian Tea House, for patiently answering all of my confused questions in the lab, and for making me feel welcome in joining your project, which was already very much an ongoing effort at the time.

Pierre, my stay with you in Princeton was among the most intense, productive, inspiring and fun part of my studies (including the most hilarious bike crash of my life). It was (is) an honour to call myself 'your protegé', and I am looking forward to future adventures with you, be it on the blackboard, in the mountains, or hopefully some time on the ski slopes! Thank you for being an amazing mentor and collaborator. Thank you also to Ricard, Marianne, Andreas, and Kamesh for making me feel welcome at the Lewis-Sigler-Institute, and for lots of nice discussions over lunch.

A PhD is nothing without a group, and I was part of the nicest, most fun group I could have imagined. Joris, Federica, Federico, Grześ, Jacqueline, Hugo, Mareike, Johannes × 2, Benedikt, Karsten, Srikanth, Matthew, Nico, Tom, Timo, Steven, Lucas, Janni, Hamed, and Estelle − thank you all for creating an amazing working environment. Joris, thank you for being a great friend along the way and for all our adventures − from camping in Morocco and Death Valley to scaling the ruins of Chernobyl, hitchhiking across the Alps, and skiing in the Rockies − and to many more to come. Fede and Grześ, la pierdolatsch, our time in a shared office was the most fun period of my PhD, and I definitely missed you guys after moving to the luxurious, well-climatised North Side of the building.

I would like to especially thank the Master students I had the privilege to work with. Your enthusiasm, energy and dedication have contributed immensely to the work presented in this thesis, and have also motivated and inspired myself to keep going. Hugo, thanks for always being up for a surprise − including a spontaneous spin. Mareike, discussing and working with you has been a joy, and I think we both learnt a lot from your project. Hanging out with you outside of work was just as much fun, from Studentenstadt parties, Amsterdam bars to Oktoberfest. Matthew, I often felt that I am learning more from you about cell migration than the other way round. Your fearlessness in trying to understand real biology, use machine learning and perform big data analysis has inspired me beyond our own collaboration. Nico, thank you for your dedication and inspiredness which helped turn a cute pet project into a major research direction. Figuring out this problem together with you was a lot of fun. Timo, supervising a student doing actual lab work was super exciting, and your dedication and perseverance has really advanced our whole operation by a big step. Steven, it was great to watch someone else tackle inference problems with my (and Pierre's) baby ULI. I also really enjoyed all of the journal club meetings you led, where you put things into context with the literature even better than the rest of us. Johannes, our collaboration has been great from the start, and I think the whole active contraction field has reached a new level with your simulations. Tom, working with you has been an inspiration, and I was so happy to see how you took over Ming's data from me, and found way cooler results than I could have imagined. Without your extraordinary committment this might have easily gone unnoticed. Johannes and Tom, I am really excited to see you both take over the various cell migration projects and I am looking forward to see what you will make of it, and how these projects will evolve with your fresh perspectives.

I am also grateful to Erwin Frey and his whole group, which was deeply entangled with our group during my time in Munich. Erwin, thank you for creating a motivating, positive, and extremely inspiring environment in your Chair. Thank you also for being part of my thesis advisory committee and for many inspiring discussions about my work following various group seminars. Emmanuel, Patrick, Philipp, Leschi, and Moritz, it's been great hanging out with you in and outside the office, especially on several awesome surfing trips. Fridtjof, Manon, Isabella, and Andriy thank you for many inspiring discussions and for a fun week stuck in Denver.

I would like to thank the whole Rädler group for making all our ongoing joint projects and collaborations so much fun. Christoph, Sophia, Georg, Theresa, Stefan and Eva, our regular meetings have always been fun and inspiring, and I can't wait to see what will follow in the

future. I would also like to thank Themis Zisis and Stefan Zahler for a great collaboration on collective cell migration.

Special thanks also go out to Ming Guo and his group. Ming, thank you for welcoming me at MIT, and for letting me be part of your research endeavours. My stay in Boston was a super fun part of my PhD and laid the basis for an ongoing collaboration that has been extremely stimulating. Yu Long, Jiawei and Wenhui, thank you for making me feel part of your group and lots of fun lunches in and around MIT!

I am very happy to have participated in the graduate school 'Quantitative Biosciences Munich' (QBM) and I want to thank all the people that put so much effort into it, in particular Ulrike Gaul, Erwin Frey, Filiz Civril and Markus Hohle. Along similar lines, I am very grateful to Marilena Pinto for running SFB1032 and for organizing the many seminars, workshops and retreats. I would also like to thank the Joachim Herz Stiftung for supporting me with a generous 'Add-on Fellowship' which has enabled an extraordinary number of conference visits and international research stays.

Special thanks also go out to Henrik Flyvbjerg, a fierce reviewer who has made our paper [1] so much better, and a great discussion partner both during your visit to Munich, as well as in follow up discussions by email. The introduction section 'Accelerations without force and mass' is very much a product of these discussions. Similarly, I would like to thank Ulrich Schwarz for important comments on the same paper, and for writing a wonderful perspective article about our work [203]. Furthermore, I would like to thank Edouard Hannezo, Ned Wingreen, Joshua Shaevitz, Greg Stephens, Andrea Cavagna, Joseph d'Alessandro, Antonio Costa, Ricard Alert, David Lubensky, and Kevin Bassler for inspiring scientific discussions that resonated with me for some time, and made important contributions to this thesis.

Thank you also to George Craig, Dieter Braun, and Jan Lipfert for being part of my thesis committee and taking the time to review this thesis.

I would like to thank my parents for always encouraging me to follow my passions and supporting me every possible way. I also thank my Munich-QBM squad Lina, Kimbu, Joris, David, Zhenya and Alex; as well as Davide, Olaf, Joeri, Rapha, Philipp and Max for being an amazing group of friends. Moreover, I want to thank my 'Berlin'-friends Anand, Benjo, Makan, Prell, Ali and Jonathan for many inspiring discussions and several amazing hut trips. Thank you also to my former Cambridge flatmates James, Moritz, and Tom, without whom I would be a different person and a different scientist.

Most importantly, I would like to thank Jessi for all your love, support, lots of proofreading and everything else.