RESEARCH ARTICLE

Tracking collective cell motion by topological data analysis

Luis L. Bonilla[1,2]*, **Ana Carpio**[2,3], **Carolina Trenado**[1]

1 G. Millán Institute for Fluid Dynamics, Nanoscience & Industrial Mathematics, and Department of Mathematics, Universidad Carlos III de Madrid, Leganés, Spain, **2** Courant Institute of Mathematical Sciences, New York University, New York, United States of America, **3** Departamento de Matemática Aplicada, Universidad Complutense de Madrid, Madrid, Spain

* bonilla@ing.uc3m.es

Abstract

By modifying and calibrating an active vertex model to experiments, we have simulated numerically a confluent cellular monolayer spreading on an empty space and the collision of two monolayers of different cells in an antagonistic migration assay. Cells are subject to inertial forces and to active forces that try to align their velocities with those of neighboring ones. In agreement with experiments in the literature, the spreading test exhibits formation of fingers in the moving interfaces, there appear swirls in the velocity field, and the polar order parameter and the correlation and swirl lengths increase with time. Numerical simulations show that cells inside the tissue have smaller area than those at the interface, which has been observed in recent experiments. In the antagonistic migration assay, a population of fluidlike Ras cells invades a population of wild type solidlike cells having shape parameters above and below the geometric critical value, respectively. Cell mixing or segregation depends on the junction tensions between different cells. We reproduce the experimentally observed antagonistic migration assays by assuming that a fraction of cells favor mixing, the others segregation, and that these cells are randomly distributed in space. To characterize and compare the structure of interfaces between cell types or of interfaces of spreading cellular monolayers in an automatic manner, we apply topological data analysis to experimental data and to results of our numerical simulations. We use time series of data generated by numerical simulations to automatically group, track and classify the advancing interfaces of cellular aggregates by means of bottleneck or Wasserstein distances of persistent homologies. These techniques of topological data analysis are scalable and could be used in studies involving large amounts of data. Besides applications to wound healing and metastatic cancer, these studies are relevant for tissue engineering, biological effects of materials, tissue and organ regeneration.

Author summary

Confluent motion of cells in tissues plays a crucial role in wound healing, tissue repair, development, morphogenesis and in numerous pathological processes such as tumor

invasion and metastatic cancer. For such complex processes, controlled experiments help clarifying the roles of chemical, mechanical and biological cues. Among them, spreading of cellular tissues on an empty space and antagonistic migration assays between cancerous and normal cells are quite revealing. The interfaces between confluent cellular aggregates uncover properties thereof when a combination of modeling, numerical simulation and data analysis is used. Here we have modified an active vertex model with a dynamics that includes inertia, friction and active forces that tend to align cells based on interaction with its immediate neighborhood. Selecting appropriately junction tensions among cells and using the SAMoS software, we have succeed in simulating assays of cellular tissue spreading on an empty space and the invasion of healthy tissue by cancerous one. We have introduced topological data analysis to characterize, track and compare in an automatic manner the interfaces of the tissue both in numerical simulations and from experimental data of normal and Ras modified precancerous Human Embryonic Kidney cells. We find good agreement when normal cells are solidlike and modified cells are liquidlike according to their shape parameters. In addition, cell variability means that a fraction of randomly distributed cells favor mixing, the others segregation. Topological data analysis techniques are scalable and could be used in studies involving large amounts of data. Besides applications to wound healing and metastatic cancer, these studies are relevant in ascertaining how the biophysical features of materials may affect tissue and organ regeneration.

Introduction

Confluent motion of epithelial cell monolayers [1–28] is crucial in many biological processes, such as morphogenesis [3, 26], biological pattern formation [9, 23], biological aggregation and swarming [17, 21], tissue repair [6, 18, 19], development [4], and tumor invasion and metastasis [1–3, 28]. It serves as a relatively simple paradigm for collective motion of cells that retain their cell-cell junctions as they move on a two dimensional (2D) substrate. Confluent cellular motion can be tracked and visualized in experiments. Velocity and stress fields can be obtained by particle imaging velocimetry (PIV), time resolved cellular motion is observed using time-lapse imaging and fluorescence microscopy, traction microscopy allows to measure the forces that cells exert on the substrate as they move [5, 6, 15]. Collective cell migration also poses challenging questions in soft and active matter physics, as it may exhibit fluid, solid or glass behavior with interesting flocking and jamming/unjamming transitions [12, 20, 29–35]. Interesting dynamics occurs as an epithelial cell aggregate advances through an empty space, as in wound healing, or it collides and encroaches a different tissue, as in cancer invasion. Advancing cellular fronts may display wave phenomena [15, 36], grow fingers [16, 37, 38], or break-down and interpenetration against an oppositely moving front [22, 27]. Different aspects of these phenomena have been studied by models ranging from macroscopic continuum mechanics to detailed subcellular agent models [25, 29, 37, 39, 40].

Here we combine particle dynamics [16] with the active vertex model (AVM) [39] to provide a cellular dynamics perspective on monolayers colliding in antagonistic migration assays (AMA) [22, 27] or on monolayers spreading over an empty space [11, 16, 37, 38]. The resulting model describes the collective migration dynamics of a large number of cells and implements exchanges of neighboring cells automatically (T1 transitions) [39]. In contrast to the usual overdamped dynamics in the AVM, the dynamics of the cell centers is underdamped. The underdamped AVM incorporates internal dissipation of cells through a friction parameter, a

Vicsek-like velocity alignment of neighboring cells [30, 41, 42], noise and and active forces that may include cell polarity. We calibrate its parameters so that the simulations agree with experiments. Parameters for collective migration to an empty space are calibrated for Madin-Darby canine kidney (MDCK) cells [10, 16]. In AMA between MDCK cells, Ras modified cells collapse and are pushed backward by normal cells, which detect the former by an ephrin related mechanism [22]. The repulsive interactions between the two cell types drives cell segregation, produce sharp borders [24], and may generate deformation waves at the interface between the two cell types that propagate across the monolayers [36]. In agreement with experiments in the literature, simulations of spreading test with our model exhibit formation of fingers in the moving interfaces, there appear swirls in the velocity field, and the polar order parameter and the correlation and swirl lengths increase with time, all of which has been observed in experiments [10, 11, 16, 37, 38, 43]. Our model is quite flexible, which gives it some advantages when describing behavior across different scales. Compared with particle models with underdamped dynamics, our model does not require introducing leader cells [16] to account for fingering instabilities. Compared to continuum models [38], stochasticity enables our model to reproduce the observed spatial autocorrelation of the velocity [11]. Simulations of our model show that cells in a finger of a moving interface may exhibit fast irregular oscillations in their velocity (periods of about one hour). This has been reported in early experiments [43]. Our underdamped dynamics also predicts that cells inside an aggregate spreading onto an empty space have smaller area than those at the tissue interface. This prediction has been corroborated by experiments [44]. Simulating the AVM with overdamped dynamics, we observe the opposite: cells at the interface and fingers have smaller are than cells inside the tissue [39].

In AMA with Human Embryonic Kidney (HEC) cell assemblies, precancerous Ras modified cells displace normal cells [27]. These latter experiments have been interpreted using continuum mechanics in a simple biophysical model through phenomenological couplings [38], without recourse to biochemical signaling mechanisms and without clear relations to cellular processes. In this paper, we consider wild type (wt) HEC cells to be solidlike whereas invading Ras cells are fluidlike and push the former backward. Experiments show that wt HEC cells keep their shape and area quite unchanged whereas Ras HEC cells may change shape and undergo larger deformations. This enforces our characterization of wt and Ras HEC cells as solidlike and fluidlike, respectively. As time elapses, there are cell exchanges and islands of one cell type form inside the tissue of the other cells, which characterizes a flocking liquid state [32, 34, 40]. In AMA with MDCK cells, the roles are inverted: Ras cells are solidlike and wt cells are fluidlike. The precise form of the separating interface among monolayers of different cell type depends on cell parameters governing segregation vs aggregation of these cells. We characterize it by topological data analysis (TDA). A measure of cellular diversity in the junction tensions produces islands of one type of cells inside the monolayer of the other cells, which is reflected in TDA of simulations and experiments. Cell cohesion given by the underdamped AVM, the cell alignment rule and the active noise force produce fingers in interfaces during assays of cell invasions of empty spaces rendering unnecessary to assume a different phenotype for lead cells [16]. Recapitulating, our model explains a wide variety of experiments on confluent motion of cellular aggregates onto free space (wound healing) and invasion of one aggregate by another (antagonistic assays, cancer). It does this by choosing judiciously physical parameters such as cellular junction tension, adhesion and those in active forces. Fine tuning of parameters may require a deeper study of experimental data. One promising area where our results are very relevant is the study of the biophysical features of materials as they affect tissue and organ regeneration (materiobiology, tissue engineering) [45].

Recent experiments have connected metastasis in colorectal cancer to wound healing and tumor invasion of tissue using appropriate molecular markers [28]. Thus, our description of spreading of cellular tissue and antagonistic migration assays using our modified active vertex model might be relevant for metastatic cancer. In particular, we have shown the role of cellular junction tensions in cell invasion, agglomeration and segregation. Promising mechanisms include Notch signaling pathways [46] and models of the epithelial-mesenchymal transition and cancer stem cell formation [47, 48]. Incorporating these cellular mechanisms to our vertex model may pave the way to future progress in this area, much as incorporating the Notch signaling pathway to cellular Potts models helps understanding many aspects of angiogenesis [49]. Understanding precise biochemical mechanisms influencing cell-cell contact and confluent cellular tissue may help develop therapies for metastatic cancers [48].

When studying spreading and collisions between cellular aggregates, the interfaces become rough and can shed and absorb groups of cells. It is important to be able to track automatically these changes for long time numerical simulations and experiments generate large data sets that is hard to visualize and follow. For the first time in studies of confluent motion of cellular aggregates, we use topological data analysis of time series generated by numerical simulations to automatically group, track and classify the advancing interfaces of cellular aggregates. Topological changes in the interfaces are reflected in barcodes and persistence diagrams of clusters and holes that change with scales (filtration parameters) [50, 51] and themselves evolve in time. We track and study these changes by means of bottleneck or Wasserstein distances [51, 52]. Measuring these changes with time in data available from experiments and comparing with data from numerical simulations allows characterizing milestones in confluent motion of aggregates and the important time scales involved. In this work, we use techniques of topological data analysis with some data taken from experiments and a modest amount of data from numerical simulations so as to explain techniques and results in a clear manner. However, our techniques are scalable and could be used in studies involving large amounts of data, as, for example, those generated to characterize zebrafish patterns by combining machine learning and topological data analysis [53].

The paper is organized as follows. The Section Mathematical Model describes the models we simulate. The numerical values of the parameters are calibrated so as to reproduce experimental observations of collective cell migration in two different cases: an aggregate spreading to an empty space and the collision of two different cellular monolayers in antagonistic migration assays. The Results and discussion section contains the numerical simulations, the characterization of the structure of advancing and interpenetrating cell fronts by means of topological data analysis, and our conclusions. The Methods section provides additional background on topological data analysis for the readers' ease of use, and it details our study of evolving interfaces of a spreading aggregate by taking slices of cells near the front.

Mathematical model

We modify an active vertex model (AVM) [39] and simulate it by adapting the SAMoS software [54]. The AVM combines the Vertex Model (VM) for confluent epithelial tissues [29, 55] with active matter dynamics [39]. Sometimes what we call AVM following Ref. [39] is called *an active self-propelled Voronoi model* [40]. Let us describe first the VM, then the AVM and our modification of its dynamics.

Vertex model

The VM assumes that all cells in the epithelium are roughly the same height and thus that the entire system can be well approximated as a two-dimensional sheet. The conformation of the

tissue in the VM is computed as a configuration that simultaneously optimizes area and perimeter of all cells. Two neighboring cells share a single edge, which is a straight line. Three junction lines typically meet at a vertex, although vertices with a higher number of contacts are also possible. The model tissue is therefore a mesh consisting of polygons (i.e., cells), edges (i.e., cell junctions), and vertices (i.e., meeting points of three or more cells). Each configuration of the mesh has the following associated energy

$$E_{\mathrm{VM}} = \sum_{i=1}^{N}\left[\frac{K_i}{2}\left(A_i - A_i^0\right)^2 + \frac{\Gamma_i}{2}P_i^2\right] + \sum_{\langle\mu,\nu\rangle}\Lambda_{\mu\nu}\,l_{\mu\nu}. \tag{1}$$

Here N is the total number of cells, A_i is the area of the cell i, A_i^0 is its reference area, and K_i is the area modulus, i.e., a constant with units of energy per area squared measuring how hard it is to change the area of the cell. P_i is the cell perimeter and Γ_i (with units of energy per length squared) is the perimeter modulus that determines how hard it is to change perimeter P_i. $l_{\mu\nu}$ is the length of the junction between vertices μ and ν, and $\Lambda_{\mu\nu}$ is the tension of that junction (with units of energy per length). The sum in the last term is over all pairs of vertices that share a junction. Note that the model allows for different cells to have different area and perimeter moduli as well as reference areas, allowing for modelling of tissues containing different cell types. The cell area and perimeter can be written in terms of vertex coordinates. Thus, vertex positions together with their connectivities uniquely determine the energy of the epithelial sheet. The main assumption of the VM is that the tissue will always be in a configuration which minimizes the total energy in Eq (1). To implement the VM, we determine the positions of vertices that minimize E_{VM} for a given set of parameters K_i, Γ_i, and $\Lambda_{\mu\nu}$. Cell rearrangements are modelled by introducing moves that change appropriately the connectivity among cells.

While the moduli K_i and Γ_i are positive, $\Lambda_{\mu\nu} < 0$. When the cell i shares junctions only with others of the same type, $\Sigma_{\langle\mu,\nu\rangle}\Lambda_{\mu\nu}l_{\mu\nu} = \Lambda_{\mu\nu}\Sigma_{\langle\mu,\nu\rangle}l_{\mu\nu} = \Lambda_{\mu\nu}P_i$, and this term can be put together with the perimeter term, thereby yielding $\frac{\Gamma_i}{2}\left(P_i - P_i^0\right)^2$ plus an unimportant constant, provided $P_i^0 = -\Lambda_{\mu\nu}/\Gamma_i > 0$. Thus the junction tension $\Lambda_{\mu\nu}$ determines the target perimeter of a type of cell. Let us assume that there are two cell types, 1 and 2, with moduli K_j, Γ_j, $j = 1, 2$, Λ_{11}, Λ_{22}, Λ_{12}, and target areas and perimeters A_j^0, P_j^0, $j = 1, 2$, respectively. We can complete squares and drop additive constants, thereby obtaining

$$E_{\mathrm{VM}} = \sum_{j=1}^{2}\sum_{i_j=1}^{N_j}\left[\frac{K_j}{2}\left(A_{i_j} - A_j^0\right)^2 + \frac{\Gamma_j}{2}\left(P_{i_j} - P_j^0\right)^2\right] + (2\Lambda_{12} - \Lambda_{11} - \Lambda_{22})\sum_{\langle\mu,\nu\rangle}l_{\mu\nu}, \tag{2}$$

in which $N_1 + N_2 = N$.

Clearly, $\Lambda_{12} < (\Lambda_{11} + \Lambda_{22})/2$ implies that energy is minimized when the number of junctions between both types of cells increases. Cells of different types therefore tend to mix. Conversely, when $\Lambda_{12} > (\Lambda_{11} + \Lambda_{22})/2$ cells of different type segregate, as suppressing junctions between cells of different type minimizes energy. There is also a competition between the two first terms in Eq (2) to minimize energy. Assume $\Lambda_{12} = (\Lambda_{11} + \Lambda_{22})/2$ and therefore the last term in Eq (2) vanishes. The shape index $p_j^0 = P_j^0/\sqrt{A_j^0} = |\Lambda_{jj}|/(\Gamma_j\sqrt{A_j^0})$ controls the ratio of the type j cell perimeter to its area. For the VM, the value $p^{0*} = 3.812$ (which corresponds to pentagons) separates solidlike and fluidlike behavior of the tissue [31]. For $p^0 < p^{0*}$, cortical tension is prevalent over cell-cell adhesion, cells do not exchange neighbors and the monolayer is solidlike. For $p^0 > p^{0*}$, cell-cell adhesion dominates, neighbor exchanges occur, and the cellular tissue behaves like a fluid [31].

Active vertex model

To introduce dynamics in the VM, we have to go from polygon vertices $\mathbf{r}_\mu$ to polygon centers that represent cells, $\mathbf{r}_i$, consider these centers as particles and introduce dynamics for them [39]. In this, the AVM is similar to the self-propelled Voronoi model [31]. The core assumption of the AVM is that the tissue configurations that optimize the energy in Eq (1) correspond to the Voronoi tessellations of the plane with polygons as cells and cell centers acting as Voronoi seeds. Given a Voronoi tessellation, we consider its dual Delaunay triangulation, comprising Voronoi seeds and the edges joining them (triangles), which have the property that no seed is inside the circumcircle of any triangle; see Fig 1. From a Voronoi tesselation it is straightforward to obtain the dual Delaunay triangulation and vice versa. However, working with Delaunay triangulations has an advantage: they retain their nature when triangle vertices move by flipping edges conveniently [39], whereas Voronoi tessellations do not. The latter
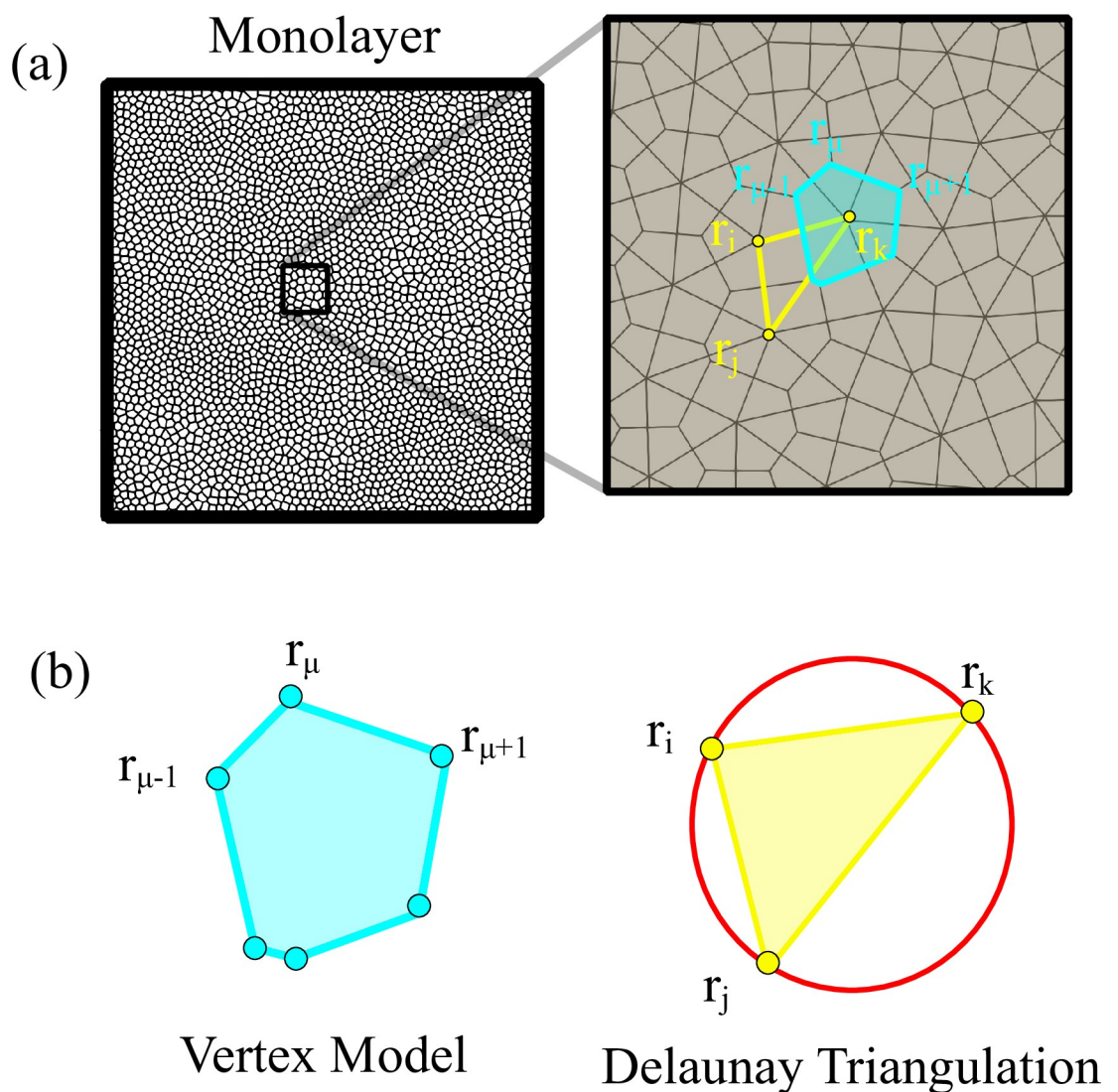

Fig 1. Voronoi tessellation and Delaunay triangulation. (a) Here $\mathbf{r}_\mu$ are vertices of polygons in the Voronoi tessselation and $\mathbf{r}_i$ are centers of polygons that are vertices of Delaunay triangles. Here the zoom of a monolayer shows (b) the Voronoi tesselation and the Delaunay triangulation.

have to be reset after motion of polygon vertices. In the AVM, the area A_i in Eq (1) of the cell i is the area of the associated Voronoi polygon, Ω_i, given by the following discrete version of Green's formula:

$$A_i = \frac{1}{2}\sum_{\mu \in \Omega_i}(\mathbf{r}_\mu \times \mathbf{r}_{\mu+1}) \cdot \mathbf{N}_i,\tag{3}$$

where $\mathbf{r}_\mu$ is the position of vertex μ, and $\mathbf{N}_i$ is a unit vector perpendicular to the surface of the polygon. For the 2D tissue $\mathbf{N}_i$ is directed along the z axis and therefore does not depend on the position of the vertices. The sum in Eq (3) is over all vertices of the Voronoi cell and we close the loop with $\mu + 1 = 1$ when μ equals the total number of vertices in the cell, N_{Ω_i}. The cell perimeter is

$$P_i = \frac{1}{2}\sum_{\mu \in \Omega_i}|\mathbf{r}_\mu - \mathbf{r}_{\mu+1}|.\tag{4}$$

The relation between the vertices $\mathbf{r}_\mu$ of the Voronoi polygons (i.e., cells) and the vertices $\mathbf{r}_i$ of the Delaunay triangles (seeds of the Voronoi polygons, i.e., cell centers) is

$$\mathbf{r}_\mu = \frac{\lambda_1\mathbf{r}_i + \lambda_2\mathbf{r}_j + \lambda_3\mathbf{r}_k}{\lambda_1 + \lambda_2 + \lambda_3}.\tag{5}$$

Here $\mathbf{r}_i$, $\mathbf{r}_j$ and $\mathbf{r}_k$ are position vectors of the corners of the triangle and λ_i, $i = 1, 2, 3$, are the barycentric coordinates; cf. Fig 1, and Ref. [39] for details. The forces $\mathbf{F}_i = -\nabla_{\mathbf{r}_i}[E_{\mathrm{VM}} + V_{\mathrm{soft}}(|\mathbf{r}_i - \mathbf{r}_j|)]$ are [39]

$$\begin{aligned}\mathbf{F}_i = &-\sum_{k=1}^{N}\frac{K_k}{2}\left(A_k - A_k^0\right)\sum_{v\in\Omega_k}[\mathbf{r}_{v+1,v-1}\times\mathbf{N}_k]^T\left[\frac{\partial\mathbf{r}_v}{\partial\mathbf{r}_i}\right]\\ &-\sum_{k=1}^{N}\Gamma_k P_k\sum_{v\in\Omega_k}(\hat{\mathbf{r}}_{v,v-1}-\hat{\mathbf{r}}_{v+1,v})^T\left[\frac{\partial\mathbf{r}_v}{\partial\mathbf{r}_i}\right]\\ &-\sum_{k=1}^{N}\sum_{v\in\Omega_k}[\Lambda_{v-1,v}\hat{\mathbf{r}}_{v,v-1}-\Lambda_{v,v+1}\hat{\mathbf{r}}_{v+1,v}]^T\left[\frac{\partial\mathbf{r}_v}{\partial\mathbf{r}_i}\right]\\ &+k\sum_{\langle j,i\rangle}(2a - |\mathbf{r}_i - \mathbf{r}_j|)\frac{\mathbf{r}_i - \mathbf{r}_j}{|\mathbf{r}_i - \mathbf{r}_j|}\,\Theta(2a - |\mathbf{r}_i - \mathbf{r}_j|).\end{aligned}\tag{6}$$

Here $\left[\frac{\partial\mathbf{r}_v}{\partial\mathbf{r}_i}\right]$ is the 3×3 Jacobian matrix connecting coordinates of cell centres with coordinates of the dual Voronoi tessellation, and the non-commutative row-matrix product $[\cdot]^T[\cdot]$ is a 3×1 column vector. $\Theta(x) = 1$ if $x > 0$, else $\Theta(x) = 0$, is the Heaviside unit step function. We have included a range repulsive force of short range a that avoids self intersections of the triangulation for very obtuse triangles [39].

In the AVM, the usual dynamics for the cell centers is a gradient flow of the energy in Eq (1), that is overdamped dynamics with forces $\mathbf{F}_i$ given by Eq (6), plus active forces $f_a\,\mathbf{n}_i$, and stochastic forces $\mathbf{v}_i$ [39]

$$\gamma\,\dot{\mathbf{r}}_i = f_a\mathbf{n}_i + \mathbf{F}_i + \mathbf{v}_i,\quad \gamma_r\dot{\theta}_i = \boldsymbol{\tau}_i \cdot \mathbf{N}_i + v_i^r(t),\tag{7}$$

where $\dot{\mathbf{r}}_i = d\mathbf{r}_i/dt$, $\boldsymbol{\tau}_i$ is the torque acting on the polarity $\mathbf{n}_i = (\cos\theta_i, \sin\theta_i)$, $\mathbf{N}_i$ is the local normal to the cell surface (the unit length vector in the z-direction), γ_r is the orientational friction, and $v_i^r(t)$ is a zero mean Gaussian white noise responsible for orientational randomness, such

that $\langle v_i^r(t)v_j^r(t')\rangle = 2D_r\delta_{ij}\delta(t-t')$. Terms aligning cell velocity or shape to polarity or terms aligning the polarity of different cells can be included in the energy of Eq (1) [39]. A particularly simple dynamics follows from $f_a = v_0$ (constant active force), $\boldsymbol{v}_i = \boldsymbol{\tau}_i = \mathbf{0}$ [31]. The AVM describes naturally cell motion and accounts for patterns of the confluent tissue observed on multiple scales, from cell sizes to much larger distances. Furthermore, cell contacts are generated dynamically from the positions of cell centers.

Dynamics including velocity alignment and inertia

In this work, we shall modify the AVM dynamics. Instead of Eq (7), we shall use the particle dynamics of Ref. [16] but with different forces between particles. As discussed in Ref. [56], trajectories of motile cells can be explained by assuming that their acceleration is a certain functional of velocity. Despite the mass of the cell being so small that inertia is negligible compared with typical forces exerted on the cell, the formula for acceleration resembles Newton's second law [56]. In this formula, a linear damping term represents dissipative processes coming from friction with substrate, with other cells, or rupture of adhesion bonds. Active memory terms, which are linear in the velocity, may propel single cells and account for the observed non-monotonic velocity autocorrelation [56]. When considering cellular tissue, Sepúlveda *et al* model cells as actively motile particles and replace the memory terms by Vicsek-like alignment "forces" [41], and interparticle and random "forces" [16]. Thus, the acceleration in these models is a consequence of the collective motion of cells and the interaction with the environment and it does not follow from Newton's second law with a mass given by that of a single cell. However, we will continue denoting by *forces* (per unit mass) the terms comprising the acceleration [16]. In contrast to Eq (7), the cells in Ref. [16] are not self-propelled, so that they can stop their motion and start moving again if there are missing cells in their neighborhood and the active force is zero:

$$\dot{\mathbf{r}}_i = \mathbf{v}_i, \quad \dot{\mathbf{v}}_i = -\alpha\mathbf{v}_i + \sum_{\langle j,i\rangle}\left[\frac{\beta}{n_i}(\mathbf{v}_j - \mathbf{v}_i) + \mathbf{f}_{ij}\right] + \boldsymbol{\varphi}_i + \sigma_0\boldsymbol{\eta}_i(t), \quad \tau\dot{\boldsymbol{\eta}}_i = -\boldsymbol{\eta}_i + \boldsymbol{\xi}_i(t). \tag{8}$$

Here, the sum is over the nearest neighbors of the vertex i of the Delaunay triangulation, n_i is the number of these neighbors, the friction coefficient α comes from internal cell friction or adhesion to the substrate or other cells. The term containing the coefficient β tries to synchronize the velocity of the nearest neighbor cells that of the ith cell and it is similar to the Vicsek model [30, 41, 42, 57]. $\mathbf{f}_{ij}$ is the force per unit mass exerted by cell j on cell i. In our simulations we use $\sum_{\langle j, i\rangle}\mathbf{f}_{ij} = \mathbf{F}_i/m_i$, where $\mathbf{F}_i$ is given by Eq (6), and not by an interparticle potential as in Ref. [16]. m_i is a reference mass, for example $m_i = \gamma\gamma_r^2/D_r$. The active forces are $\boldsymbol{\varphi}_i + \sigma_0\boldsymbol{\eta}_i(t)$. In Ref. [16], $\boldsymbol{\varphi}_i = 0$ and $\boldsymbol{\eta}_i(t)$ is a zero mean Ornstein-Uhlenbeck noise, representing a stochastic force with nonzero correlation time τ. $\boldsymbol{\xi}_i(t)$ is a zero-mean delta-correlated Gaussian white noise. For spreading tests, we have used the numerical values of the parameters indicated in Table 1. For antagonistic migration assays, we have used the numerical values collected in Table 2. As we shall see in the description of the numerical simulations, the dynamics given by Eq (8) with our choice of forces allows us to reproduce many features observed in experiments.

Table 1. Parameters corresponding to the experiments with MDCK cells in Ref. [16].

α	β	τ	σ_0	K	Γ	Λ	λ	l_0	ζ
h^{-1}	h^{-1}	h	$\frac{\mu m}{h^{3/2}}$	-	-	-	-	-	-
0.534	41.36	0.56	95	1	0.1	-1	0.1	0	0.5

Table 2. Two sets of parameters corresponding to the experiments with HEK cells in Ref. [27].

α	β	τ	σ_0	K_j	Γ_j	Fig #
h^{-1}	h^{-1}	h	$\frac{\mu m}{h^{3/2}}$	-	-	-
0.0602	13.85	1.66	55.88	1	1	8
0.42	0.602	1.66	13.97	1	1	9, 10

Boundaries

The cells at the boundary between a cellular monolayer and an empty space, or between tissues, are special. They may form actin cables, thereby having a line tension and a bending stiffness [39]:

$$E_{\text{lt}} = \frac{1}{2} \sum_{\langle i,j \rangle} \lambda_{ij} (l_{ij} - l_0)^2, \tag{9}$$

$$E_{\text{bend}} = \frac{1}{2} \sum_i \zeta_i (\theta_i - \pi)^2, \qquad \cos \theta_i = \frac{\mathbf{r}_{ji} \cdot \mathbf{r}_{ki}}{|\mathbf{r}_{ji}| \, |\mathbf{r}_{ki}|}. \tag{10}$$

Here the modulus λ_{ij} is the line tension of the edge connecting vertices i and j, $l_{ij} = |\mathbf{r}_{ij}|$ ($\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$) is the edge length (of preferred magnitude l_0), ζ_i is the bending stiffness of angle θ_i at the boundary particle i, and $\mathbf{r}_j$ and $\mathbf{r}_k$ are the positions of boundary particles to the left and to the right of particle i. The line tension energy of Eq (9) tries to keep boundary edges at a length l_0 whereas the bending energy of Eq (10) tries to keep the boundary line flat. The sums in these formulas are over boundary particles only and we assume that each boundary cell has exactly two boundary neighbors [39].

Initial condition

A random configuration of the particles comprising a confluent cell monolayer is usually different from those configurations observed in experiments. Thus, we have to carry out an initialization stage until the particle configuration is compatible with their observed velocity distributions. For spread tests, we proceed as follows. We set a square box of size 1 mm^2 area, $N \approx 4000$ particles (comparable to the number of cells in the experiments), the packing ratio and the particle mean velocity. Then, we numerically solve Eqs (1) and (8) with forces $\boldsymbol{\varphi}_i = \mathbf{0}$ and $\sum_{\langle j, i \rangle} \mathbf{f}_{ij} = \mathbf{F}_i/m_i$, $\mathbf{F}_i$ given by Eq (6), until the velocity probability density functions (PDFs) of the experiments are fitted. The parameters adjusted to the experimental data at early time (30 min after stencil removal) are listed in Table 1. We stop the initialization stage when the distribution of mean distances between particles is close to the initial distribution as observed in experiments and displayed in Fig 2. From this simulation, we obtain the particle positions $\mathbf{r}_i$ and solve the underdamped AVM with forces given by Eq (6) and initial random directions for the particle velocities. As we can see in Fig 3, the velocity field obtained from the simulations, Fig 3(b), is very similar to that measured by PIV analysis [16], Fig 3(a).

For AMA, we choose a random configuration having the same number of wt and Ras cells separated by a vertical straight line and we set a known velocity distribution from experiments [27]. This represents the situation of the two monolayers when they first make contact. See details in the next section.
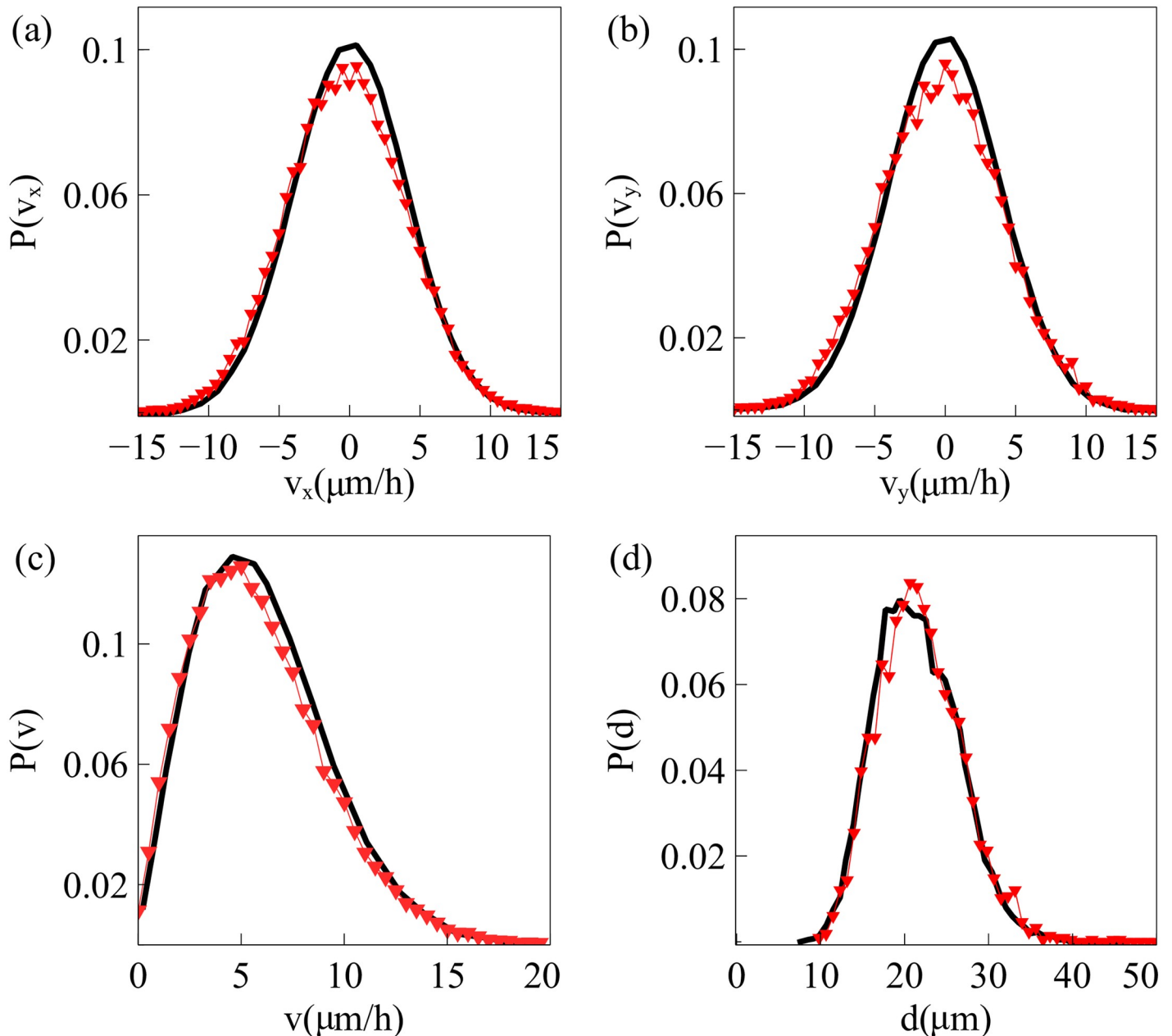
Fig 2. Probability distribution function (PDF) for particle velocities: (a) v_x, (b) v_y, (c) $v = |v|$; and (d) mean distance d between neighboring particles; after the initialization procedure (red triangles) as compared to the experimentally observed PDF (black line) [16]. Parameter values are those in Table 1.

https://doi.org/10.1371/journal.pcbi.1008407.g002

Results and discussion

We have simulated two different tissue configurations: (a) a cellular monolayer spreads over an empty space, and (b) two monolayers comprising wild type and modified cells collide. In each case, the simulations are compared to relevant experimental observations.

Numerical simulations for the spreading configuration

Inspired by wound healing phenomena and experiments on tissue scratching, we are interested in the movement of an epithelium which encroaches on a virgin substrate. The
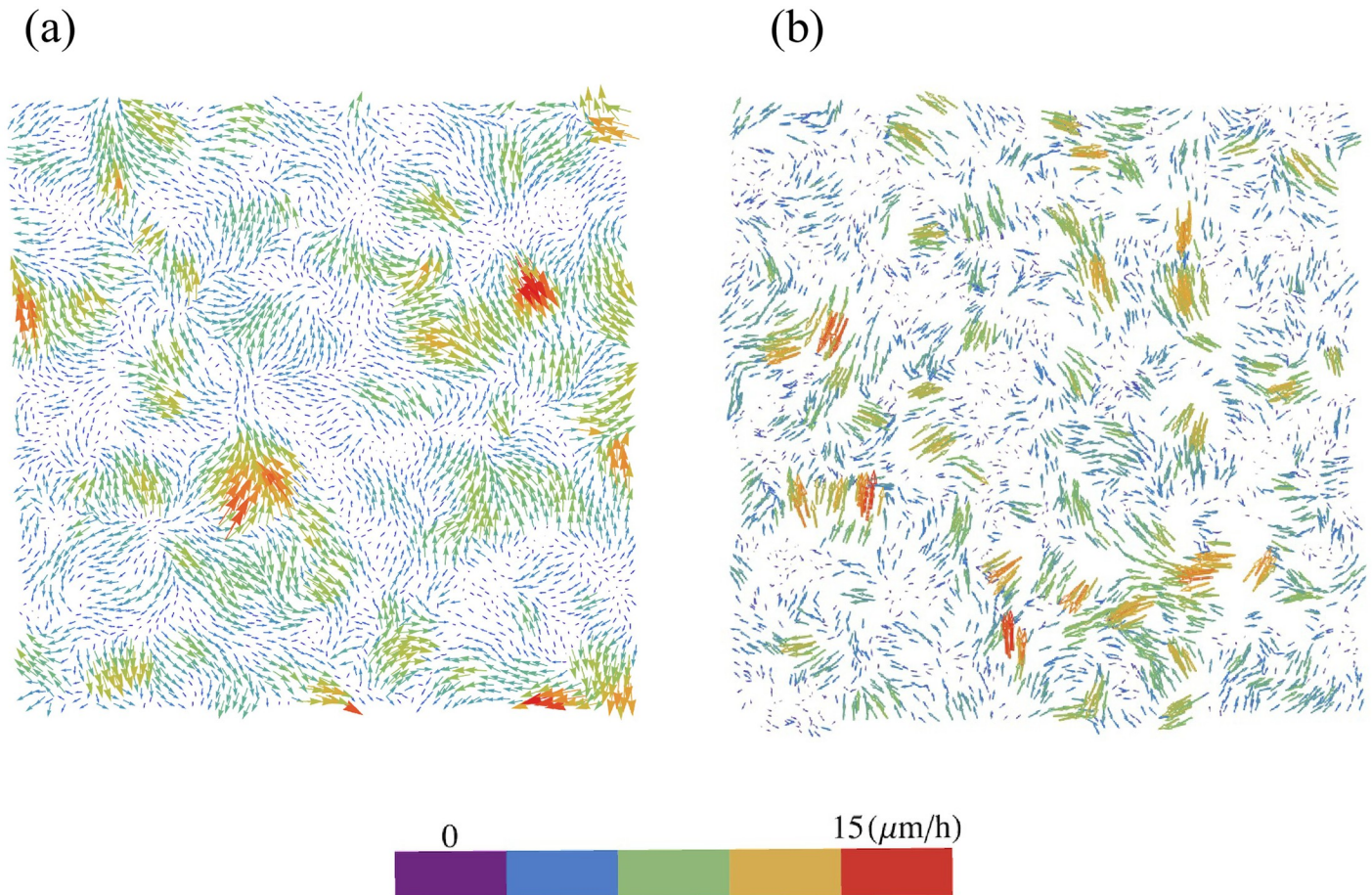
(a)

(b)

Fig 3. Velocity field obtained from (a) experiments [16], (b) simulations after the initialization procedure. Parameter values are those in Table 1.

experimental protocol consists of microfabricated stencils whose removal increases the motility of the epithelium. In our simulations, we consider a narrow strip configuration as that in Fig 4(a), which is similar to those in Ref. [43]. We adapt the SAMoS code [54] to simulate the AVM with dynamics given by Eqs (8) and (6), in which $\varphi_i = 0$. Parameter values are those in Table 1. Cells migrate on the surface maintaining their junctions with their neighbors, which is enforced by the term proportional to β in Eq (8). During healing, noisy forcing in Eq (8) makes some cells to move faster that the others while keeping their contacts. This is the origin of the fingers or instabilities of the interface with the cell free space, which are illustrated by Fig 4(b), see S1 Video for the complete time evolution. In addition, cells on the interface, or close to it, may grow beyond the target area A^0 in Eq (1). As they do so, each cell has a probability to divide into two daughter cells, which equals $r_d(A - A^0)dt$. Here dt is the time step and r_d is the division rate. We have normalized the target area to $A^0 = \pi$, $dt = 0.05$, $r_d = 0.01$, and we check whether the cell divides with ten times the above probability every 10 time steps that we observe $A > A^0$. With these parameters, there is some cell division near the interface of the confluent layer and the empty space.

After two hours of stencil removal, the PIV recorded from the experiments reveals the complex movements that can appear inside the bulk of the tissue, cf. Fig 2A of Ref. [10]. Cells do not move independently and their velocities are correlated. The presence of these cellular
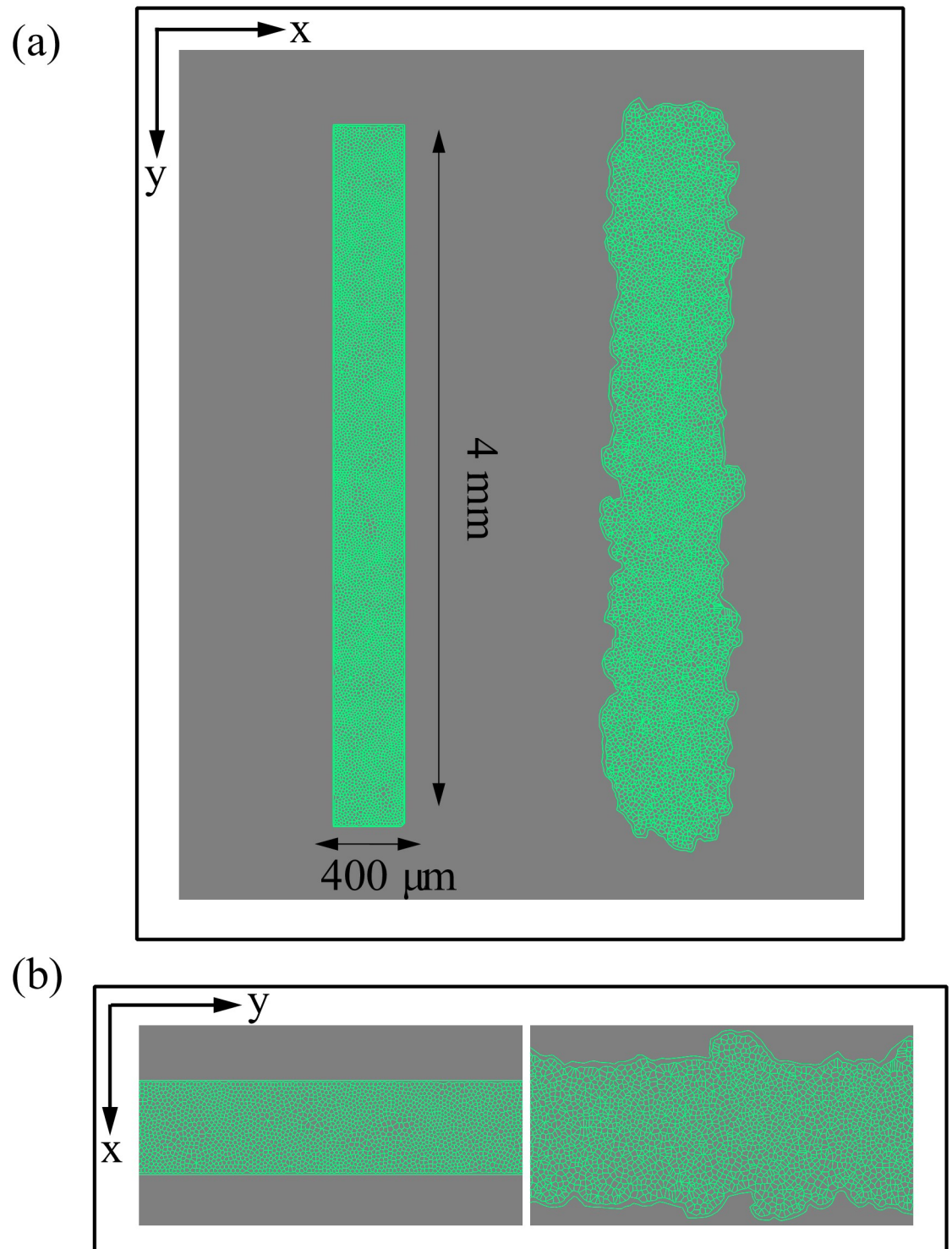
Fig 4. Initial configuration and configuration after 20 h of stencil removal showing the formation of fingers according to the numerical simulation of the model. (a) Full view, (b) zoom. Initial box size is 1.6 mm^2, $P^0 = 10$, $A^0 = \pi$, and shape index $p^0 = 5.65$. Parameter values are those in Table 1. See S1 Video.

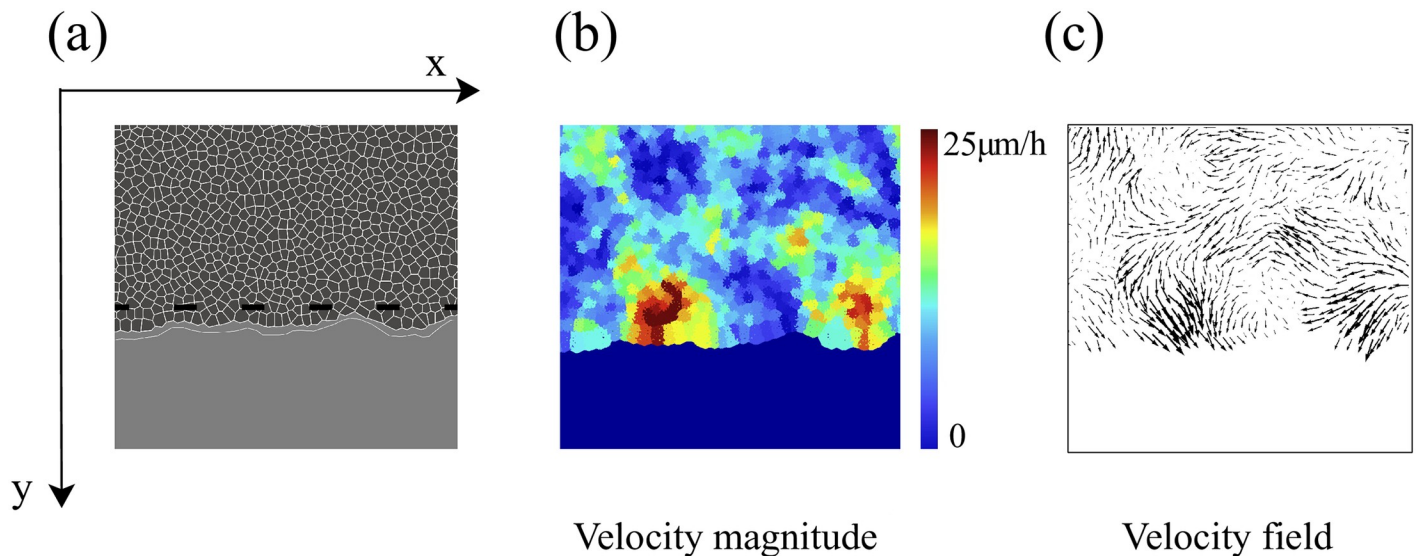https://doi.org/10.1371/journal.pcbi.1008407.g004

Fig 5. Cell velocity field after 2 h of stencil removal in an invasion configuration calculated from simulations of the model with the parameters of Fig 4 and Table 1. (a) Phase contrast visualizing cells, (b) profile of cell speed (modulus of velocity), (c) velocity field. These panels should be compared with those obtained from experimental MDCK data in Fig 2A of Ref. [10].

flows shows the existence of motion inside the monolayer [11]. Similar to experiments, our simulations in Fig 5 show that incipient fingers appear in areas of high speed; see also S2 Video. In our simulations, we find these areas without having to postulate the existence of special leader cells. Having calculated numerically the velocity field, we can quantify the orientational motion inside the epithelium. Take for example, the configuration after 35 h of stencil removal shown in Fig 6. In addition to the velocity field and the speed (modulus of the
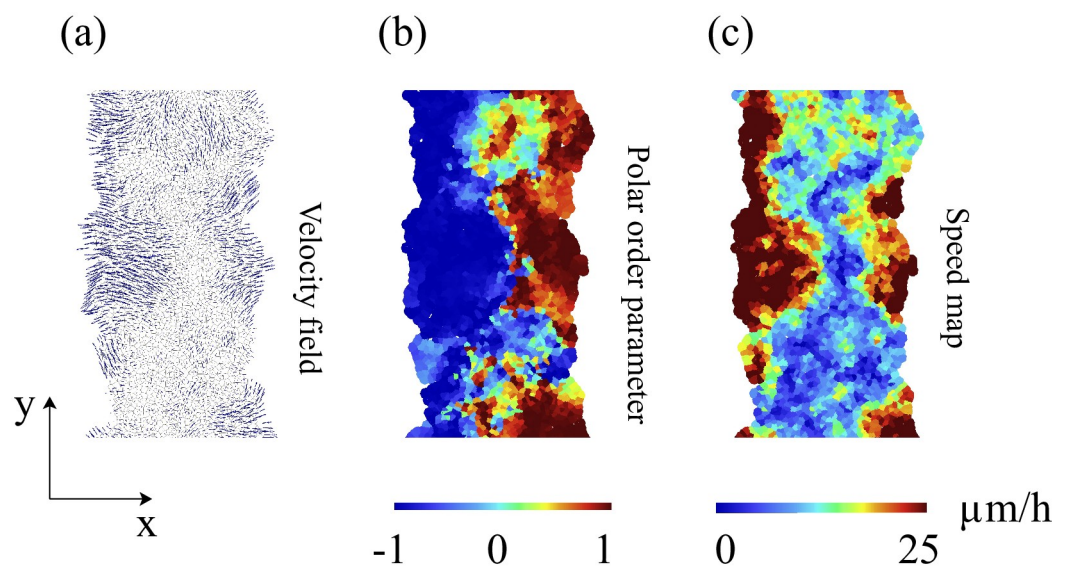

Fig 6. (a) Numerically simulated cell velocity field, (b) local polar order parameter $\cos\vartheta_i$, and (c) speed ($|v|$) map after 35 h of stencil removal in an invasion configuration for a 400 μm wide strip. Parameter values as in Fig 5. See S2 Video.
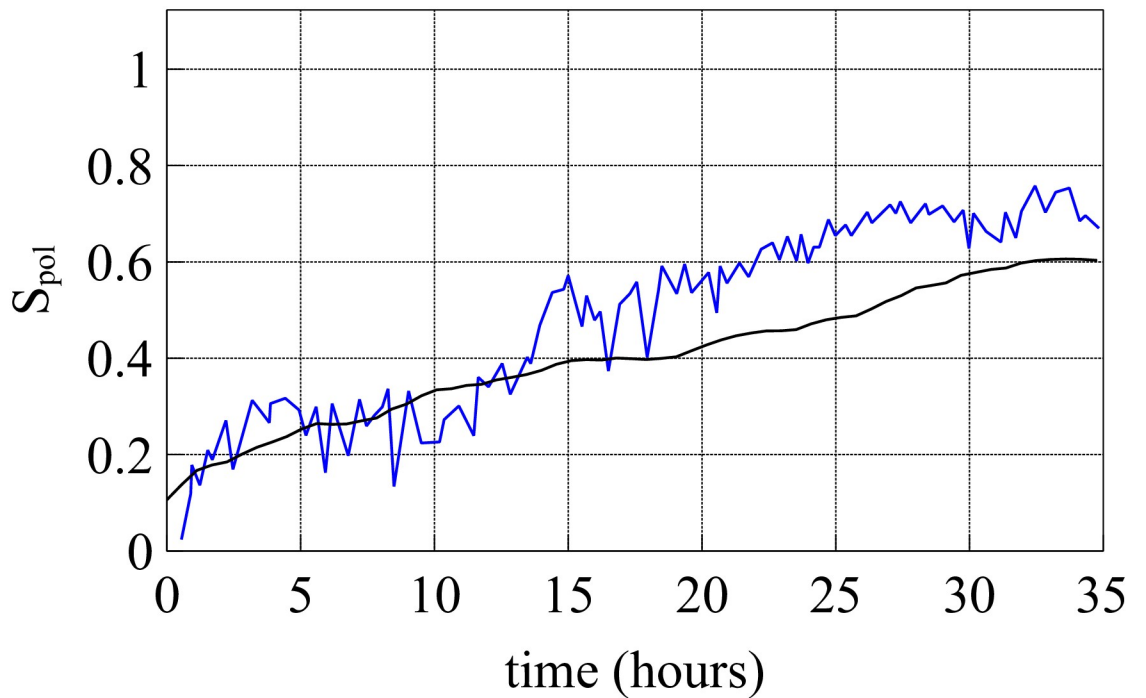
Fig 7. Evolution of the polar order parameter $S_{pol}(t)$ corresponding to Fig 5. Here $t = 0$ corresponds to 1.5 h after stencil removal [10]. An average over 5 simulations exhibits the same trend as measurements reported in Ref. [10] (jagged line).

velocity vector) map, we have depicted a density map of the polar order parameter S_{pol},

$$S_{pol} = \frac{1}{N} \sum_{i=1}^{N} \cos \vartheta_i, \quad \cos \vartheta_i = \frac{v_x(i)}{\sqrt{v_x(i)^2 + v_y(i)^2}}. \quad (11)$$

Here ϑ_i is the angle that the velocity vector of the ith cell forms with the outer normal to the strip (the x axis in Fig 4). Fig 6(b) depicts the density plot of the cellular polar order parameter, $\cos\vartheta_i$, after 35 h of stencil removal (similar to experimental data reported in Fig 92 of Ref. [43]). Fig 7 shows that an ensemble average of the polar order parameter (over 5 realizations, smooth line) increases with time and follows the same trend as the measurements reported in Ref. [10] (jagged line). At early times, S_{pol} in Fig 7 does not exhibit a particular trend. The angles are distributed homogeneously and are not located in specific areas. After a while, the cells start orient themselves perpendicular to the strip, specially at the edges of the tissue, as shown in Fig 6. This effect occurs in strips of width larger than 300 μm. On shorter strips, their two sides are no longer independent and the appearance of a finger changes the motion of the whole strip.

The fronts of advancing cells in Figs 5 and 6 clearly show the formation of fingers. The AVM keeps cells together while the term proportional to β in Eq (8) induces a common average direction in their motion. This effect becomes stronger the larger β is, which promotes and enforces finger formation. Thus, unlike the particle model of Ref. [16], we do not need a longer range attractive potential interaction between cells. We do not need to distinguish leader cells to trigger finger formation [16] because advancing cells at the forefront of the monolayer pull those behind them. A comparison of our simulation results in Figs 5 and 6 to the experiments reported in Refs. [10, 43] shows that the appearance and size of the cell velocity field are
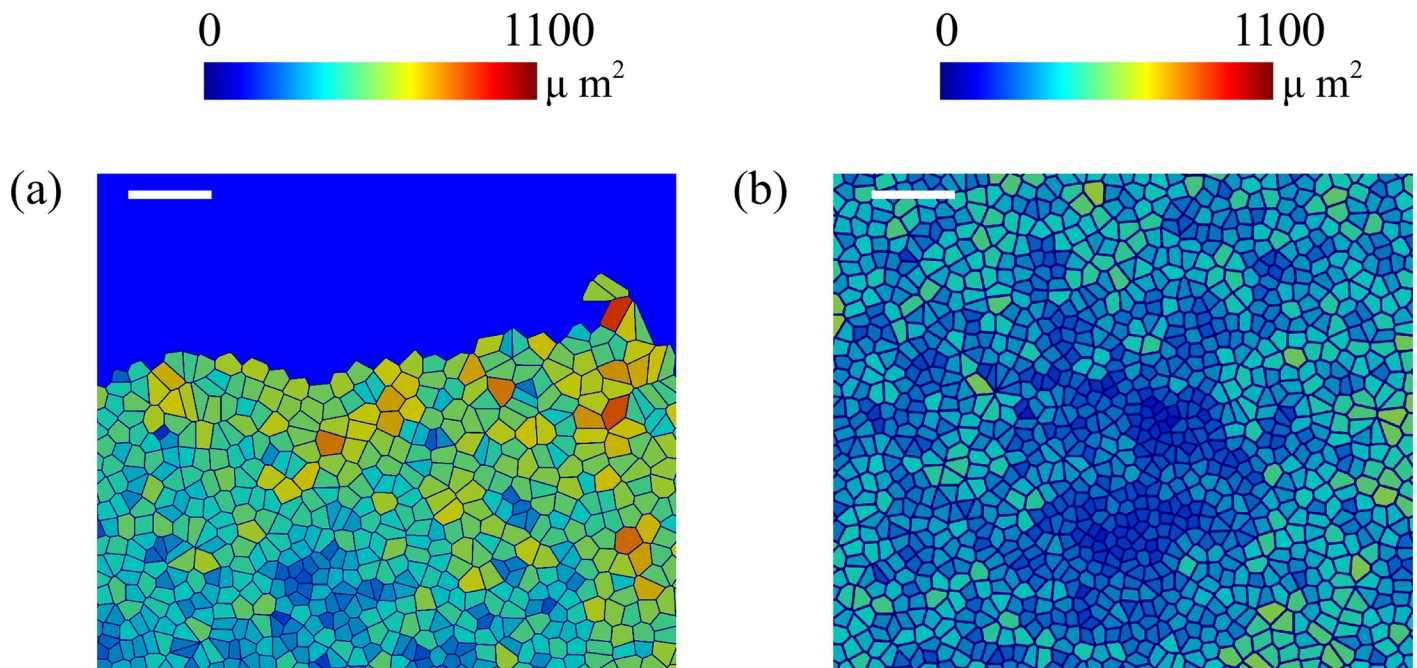
Fig 8. Areas of cells during a simulation of a spreading configuration: (a) Area of cells near the interface, (b) area of cells far from the interface. Our simulations exhibit the same trend as measurements reported in Ref. [44]. The bar length in both panels is 100 μm.

https://doi.org/10.1371/journal.pcbi.1008407.g008

reproduced qualitatively. Our simulations show that the area and velocity of cells both increase as their distance to the boundary of the cellular tissue decreases. Fig 8 shows that cells near the interface in a spreading configuration have larger areas than cells far from the interface. This is particularly noticeable in the fingers: the cells in them are faster and have a larger area than the cells elsewhere. The cells far from the tissue border are compressed and have smaller area than boundary ones. This prediction of the underdamped AVM with dynamics as in Eq (8) has been observed in experiments; see Fig 4 of Ref. [44]. In experiments, the area of finger cells reaches larger values than in the simulations, which is related to the fact that we use a fixed target area for all cells and the cell area cannot depart arbitrarily far from target in the AVM. We have also simulated the AVM with the overdamped dynamics of Eq (7) and with the same boundary and initial conditions. In this case, after fingers are formed as in Fig 6 of Ref. [39], the interior cells far from the interface have larger area than cells at the boundary and in the fingers. This is also shown in Fig 8 of Ref. [39]. However, this behavior is contrary to experimental observations [44].

Our numerical simulations of spreading configurations show that the cells inside a finger move faster than those at other portions of the interface. We have observed that the average velocity of finger cells may oscillate irregularly about some average value with a short period of about one hour. Fig 9 shows the average velocity of 9 finger cells during a 7 hour time interval. The velocity of a single cell in the finger oscillates somewhat more irregularly in a similar fashion. For much longer time intervals, the average velocity may experience an overall upward trend. The average velocity of boundary cells in flat regions also oscillates with time but it does not show a definite behavior over long time intervals: it may even display a downward trend. In experiments, the velocity of cells leading interfacial fingers has also been observed to oscillate rapidly and irregularly with periods of about one hour or less, which is similar to the findings based on numerical simulations of our model; see Fig 101A of Ref. [43]. Some models based on continuum mechanics predict longer periods of tens of hours [43].
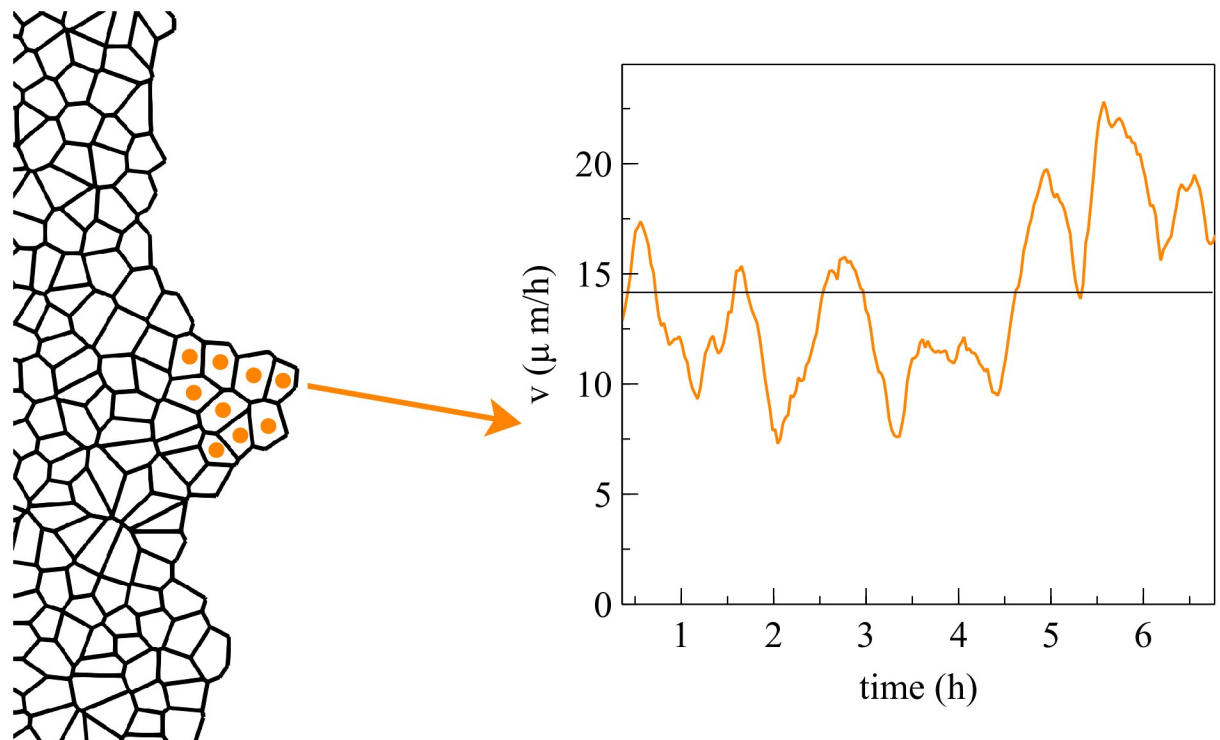
Fig 9. Average velocity of the marked cells during finger expansion. The velocity of each cell oscillates in a similar but somewhat more irregular manner (not shown).

The velocity field in Fig 6(a) exhibits swirl patterns [11]. To characterize them, we have depicted in Fig 10(a) the correlation function for the x-component of the velocity field:

$$I(|\mathbf{r}|, t) = \frac{\langle v_x^*(\mathbf{r}', t)\, v_x^*(\mathbf{r}' + \mathbf{r}, t)\rangle_{\mathbf{r}'}}{\sqrt{\langle v_x^*(\mathbf{r}', t)^2\rangle_{\mathbf{r}'} \langle v_x^*(\mathbf{r}' + \mathbf{r}, t)^2\rangle_{\mathbf{r}'}}}, \quad v_x^*(\mathbf{r}, t) = v_x(\mathbf{r}, t) - \langle v_x(\mathbf{r}, t)\rangle_{\mathbf{r}}. \tag{12}$$

Here the averages are spatial averages over $\mathbf{r}'$ and also ensemble averages over simulations with different initial conditions. Fig 10(b) depicts the correlation length defined by the first zero of the correlation function and the swirl size defined by its first local minimum. Empty and blue squares correspond to values given by different simulations. The best fits to straight lines are also shown and compared to a similar line for Angelini *et al*'s experimental data [11]. Clearly correlation length and swirl size increase with time, indicating that cells feel each other on increasingly larger regions as time elapses. This has been observed in other experiments and simulations [10, 27]. The correlation lengths given by our simulations agree quite well with values reported in the literature for similar observation times [10, 11, 27].

Numerical simulations for the collision configuration

Recently, Moitrier *et al* have reported confrontation assays between antagonistically migrating cell sheets [27]. In their experiment, the two confluent cellular monolayers (wild type and modified Ras HEK cells) advance toward an intermediate empty space, collide and the Ras monolayer displaces the wt one. The experiment shows that the velocities of the cells decay exponentially fast the farther they are from the advancing fronts [27]. If $x = L(t)$ is the position of the monolayer front, the velocity of the cells at position $x < L$ is $V^{\text{wt}} \exp[(x - L)/\lambda^{\text{wt}}]$ for the
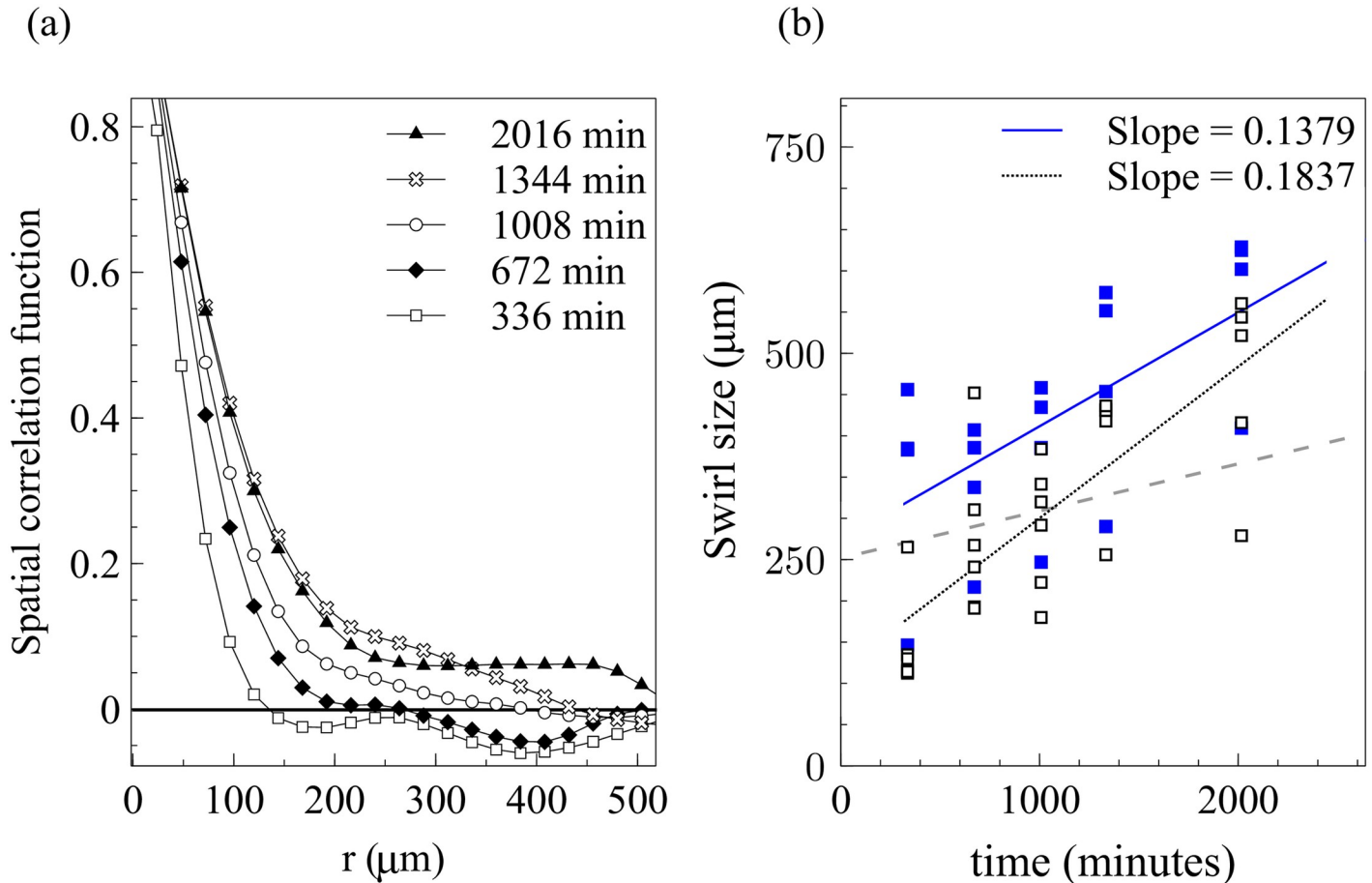
(a)

(b)

Fig 10. (a) Spatial correlation function $I(r, t)$ corresponding to Fig 6(a) for different times. (b) Correlation length given by the first zero of $I(r)$ (empty squares) and swirl size given by the first local minimum of $I(r)$ (blue squares). Dashed line from swirl sizes in Ref. [11].

wt and $-V^{\mathrm{Ras}} \exp[-(x - L)/\lambda^{\mathrm{Ras}}]$ for the Ras cells at $x > L$. After the collision, these velocity functions remain the same but now V^{wt} and V^{Ras} acquire a common and lower value $-V^{\mathrm{interface}}$. Moitrier *et al* interpret their experiments by comparing with simple solutions of a 1D continuum model [27]. In our simulations, we use the SAMoS code to simulate the under-damped AVM cellular model with dynamics given by Eqs (8) and (2). The invading Ras cells (magenta) move to the left whereas the wt cells (green) are pushed backward because they experience aversion to mixing with Ras cells. We model this situation by adding a negative active force $\varphi_i^{\mathrm{Ras}} = a^{\mathrm{Ras}} \exp\left[-[x - L(0)]/\lambda^{\mathrm{Ras}}\right]$ to Ras cells in Eq (8) for $x > L(0)$ (not included in Ref. [16]), whereas wt and Ras cells do not experience an active force if $x < L(0)$. We use $\lambda^{\mathrm{Ras}} = 410 \mu m$, $a^{\mathrm{Ras}} = 9 \mu m/h^2$, $L(0) = 0$. The active force φ^{Ras} keeps Ras cells moving to the left and pushing wt ones. Therefore we no longer need the synchronization force proportional to β to keep cells moving in the same direction. Fig 11 shows finger formation for the active force φ^{Ras} and for $\beta = 13.85 \, h^{-1}$, which is smaller than the value in Table 1. Other parameters are as indicated in Table 2.

Our underdamped AVM uses more features of wt and Ras cells obtained from the experiments than kept by continuum models. The latter lose features at distances close to the cell size. Continuum models fit friction, viscosity and strength of active forces for the two cell populations to explain how Ras cells invade the wt monolayer [27].
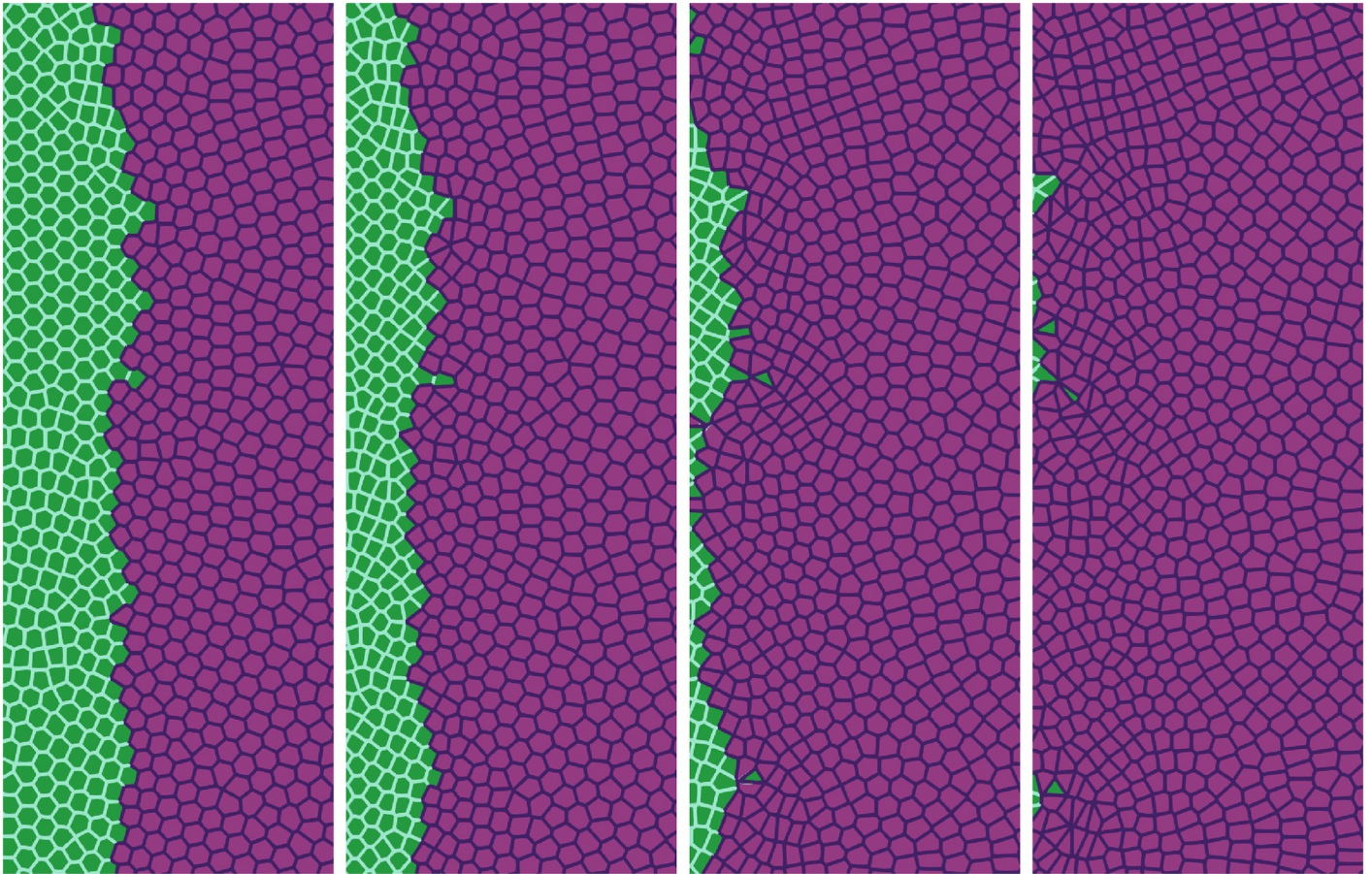
Fig 11. Simulation of the antagonistic migration assay: One population advances pushing back the other. Junction tensions are $\Lambda_{11} = -6.2$, $\Lambda_{22} = -6.8$, which yield shape indices 3.50 (green cells) and 3.84 (magenta cells), respectively. Other parameters are listed in the first row of Table 2, and $\Lambda_{12} = -7.0 < \frac{1}{2}(\Lambda_{11} + \Lambda_{22})$ correspond to weak population mixing. Snapshots are taken at times 2 h, 6.5 h, 13 h, 20 h. See S3 Video.

The AVM allows us to study tissues that behave differently. In our simulations, 5000 cells are split into two populations with different properties specified by the junction tensions Λ_{ij}, $j = 1, 2$, which affect each pair of cell-cell contacts. The simulations producing Figs 11, 12 and 13 have open boundaries because we have focused on the interface between populations. We have fixed $K = \Gamma = 1$ and $-6.8 = \Lambda_{22} < \Lambda_{11} = -6.2$, which produce shape parameters p^0 of 3.50 (green cells) and 3.84 (magenta cells), below and above the transition value $p^{0*} = 3.812$, respectively. Thus, Ras magenta cells are fluidlike (supercritical shape index) and their strain energy density is smaller than that of the solidlike wt cells. This is consistent with the observation that wt cells have larger mean traction force amplitudes than Ras cells [27]. Our aim is to analyze the effect of Λ_{12} on the AMA. Both monolayers occupy the right and left portions of a 4.4 mm wide, 3.1 mm tall box. In Figs 11–13, we show a 1 mm × 2.5 mm region.

In our simulations, we start from having the cell populations separated by a straight vertical interface at $L(0) = 0$. The active force $\boldsymbol{\varphi}$ pushes Ras cells with $x > L(0)$ to the left, whereas $\boldsymbol{\varphi} = \mathbf{0}$ for any cell to the left of $x = L(0)$. The junction tension Λ_{12} in Fig 11 ($\Lambda_{12} = -7.0$) and in the two left panels of Fig 12 ($\Lambda_{12} = -7.5$) favors population mixing. Ras (magenta) cells push wt (green) cells backwards at a velocity close to the observed $V^{\text{interface}}$, meanwhile creating a rugged interface between cell populations. As time elapses, fingers and some isolated islands (lagging wt in the Ras assembly and advancing Ras islands in the receding wt assembly) appear.
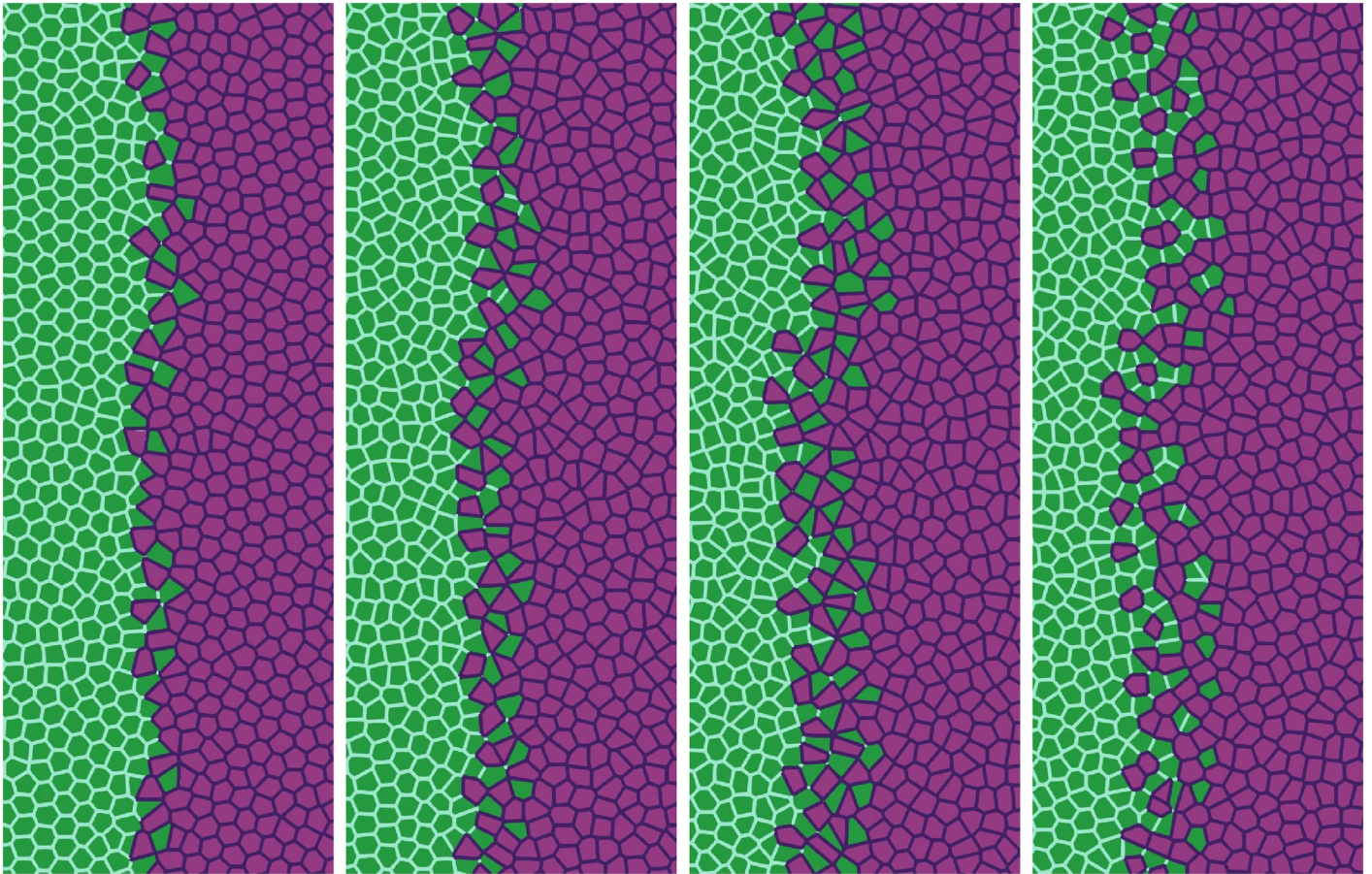
Fig 12. Simulation of the antagonistic migration assay: Creation of a extremely rugged interface. First and second snapshots: $\Lambda_{12} = -7.5 < \frac{1}{2}(\Lambda_{11} + \Lambda_{22})$ (population mixing); third and fourth snapshots: $\Lambda_{12} = -6.0 > \frac{1}{2}(\Lambda_{11} + \Lambda_{22})$ (population segregation). Other parameters are listed in the second row of Table 2 whereas times are as in Fig 11. See S4 Video.

https://doi.org/10.1371/journal.pcbi.1008407.g012

These effects are more pronounced the smaller Λ_{12} is, as shown by comparison of Figs 11 and 12. It is possible to create some realistic mixing of the populations by changing the junction tension Λ_{12} with time. The first two snapshots in Fig 12 have $\Lambda_{12} = -7.5 < \frac{1}{2}(\Lambda_{11} + \Lambda_{22})$, which favors population mixing. Then the interface between cell populations becomes very rugged and there appear islands of one cell type inside a layer of the other type. The third and fourth snapshots in Fig 12 have been obtained with $\Lambda_{12} = -6.0 > \frac{1}{2}(\Lambda_{11} + \Lambda_{22})$ that favors population segregation. The interface becomes smoother and the islands shrink and tend to disappear.

We have also focused on the effects of cellular alignment. There are two terms in Eq (8) that try to synchronize cell velocities: the term proportional to β and the active force $\boldsymbol{\varphi}$, which pushes the Ras cells to the left. Although the values of β used to draw Figs 11–13 are smaller than that in Table 1, different β still make a difference in the behavior during tissue collision, specially in the Ras population. Fig 11 exhibits global polar migration because its β value is larger than that in Figs 12 and 13, but types of cells are not mixed despite having a favorable value $\Gamma_{12} = -7.0$. The smaller value of β in Figs 12 and 13 creates a weaker polar alignment than that in Fig 11. The different patterns observed in these figures illustrate that cell alignment
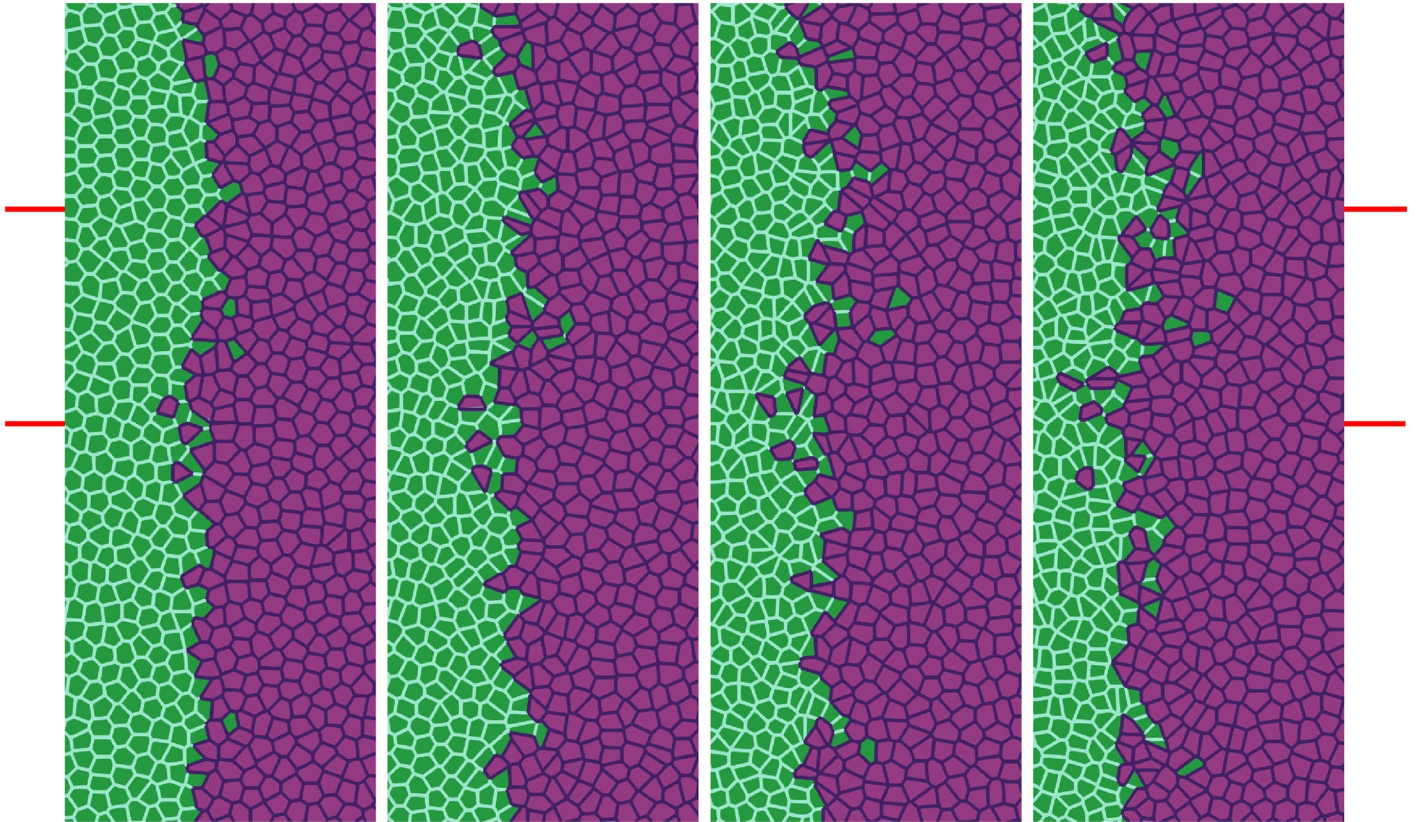
Fig 13. Simulation of the antagonistic migration assay: One population advances pushing back the other, while the interfaces mix, creating scattered islands of cells of different type. Parameters and times are as in Fig 12, except that one fifth of the overall population (randomly placed green and magenta cells) have $\Lambda_{12} = -7.5$ (population mixing) and the other four fifths have $\Lambda_{12} = -6.0$ (population segregation). The marked region has similar size to that reported in experiments [27] and will be used in our TDA studies. See S5 Video.

affects importantly the shape and configuration of the interface. S3–S5 Videos compare the time dynamics of these three sets of simulations.

While the rightmost panel of Fig 12 is similar to some of the experimental data [27], we can obtain a similar formation of islands and fingers by assuming that Λ_{12} is randomly distributed among cells. In particular, we assume that one fifth of magenta and green cells have $\Lambda_{12} = -7.5$, which favors mixing of populations, while the remaining ones have $\Lambda_{12} = -6.0$ and favor population segregation. The result is depicted in Fig 13, which exhibits behavior similar to experimental observations [27], compare also to Fig 14. The topological data analyses of the next section characterize the geometry of the interface between cell types in antagonistic migration assays.

Formation of islands and topological data analysis

Experiments and numerical simulations of cell monolayers produce time series of images that make it possible to identify the structure of interfaces and to compare their time evolution. It is quite cumbersome to process manually these time series. Here we use Topological Data Analysis (TDA) as a computational tool to process automatically time series of images. We next illustrate how to use TDA for this purpose and how to interpret the obtained results. We focus on specific parts of selected snapshots of images from experiments and then on time series of images from numerical simulations. While we have few images of interfaces from experiments,
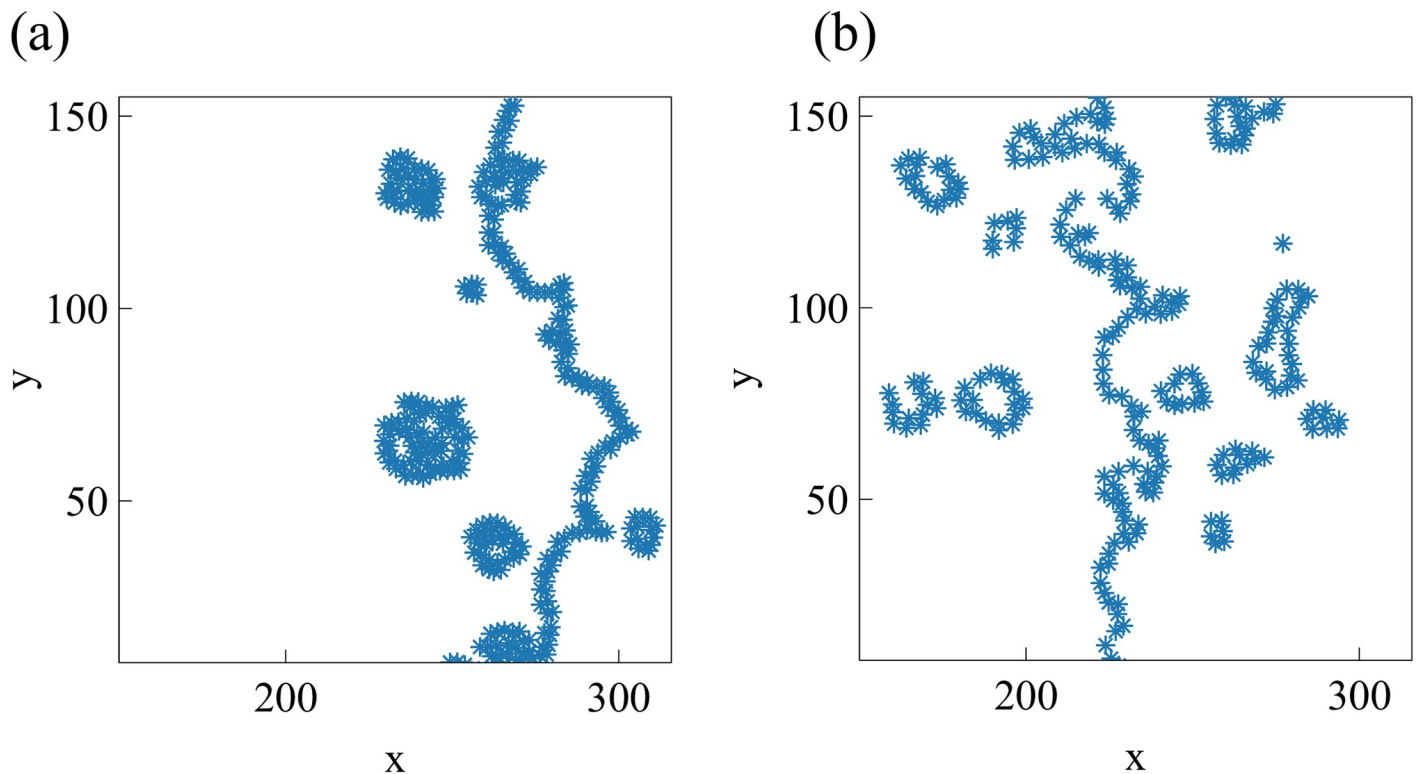
Fig 14. Structure of the interface between colliding layers corresponding to two snapshot sof the collision of two confluent cellular monolayers in Moitrier *et al*'s experiment [27]. These profiles correspond to (a) the third and (b) the fourth panels (counting from the left) in the cover of Soft Matter corresponding to Ref. [27].

we can generate arbitrarily many from numerical simulations. Having many images, the automatic TDA tool enables us to describe in detail the topological changes of the interfaces and to implement hierarchical clustering strategies, thereby classifying the evolving interface structures.

Fig 14 shows the interfaces between two colliding confluent cellular monolayers in an AMA [27]. In this experiment, magenta Ras cells make green wild type cells move back, cf. third and fourth snapshots in the cover of Soft Matter, vol. **15** [27]. The interface between the two cell populations is rather rough, it exhibits fingers, and there are islands or pockets of green cells left behind by the advance of the magenta front. To quantify these phenomena in an automatic way, we proceed as follows. Using Matlab, we transform the images in matrices of ones (green) and zeros (magenta). Then we extract the positions of green/magenta interfaces, represented by the point clouds shown in Fig 14, and process them using TDA. We pursue a similar strategy for images extracted from numerical simulations of our underdamped AVM, which yields a more complete picture of the evolution of interfaces.

Persistent homology. A finite set of data points may be considered a sampling from the underlying topological space. Homology distinguishes topological spaces (e.g., annulus, sphere, torus, or more complicated surface or manifold) by quantifying their connected components, topological circles, trapped volumes, and so forth. Persistent homology characterizes the topological features of clouds of point data or particles at different spatial resolutions [58]. Highly persistent features span a wide range of spatial scales. Persistent features are more likely to represent true features of the data/pattern under study than to constitute artifacts of sampling, noise, or parameter choice [50]. To find the persistent homology of a cloud of point

data/set of particles, we must first view them as a simplicial complex C. Roughly speaking, a simplicial complex is defined by a set of vertices (points or particles) and collections of k-simplices. The latter are the convex hulls of subsets with $k + 1$ vertices, comprising also faces; see the Methods section for precise definitions. Defining a distance function on the underlying space (the euclidean distance, for instance), we can generate a filtration of the simplicial complex, which is a nested sequence of increasingly bigger subsets. More precisely, a *filtration* of a simplicial complex C is a family of subcomplexes $\{C(r) \mid r \in \mathbb{R}\}$ of C such that $C(r) \subset C(r')$ whenever $r \leq r'$. The filtration value of a simplex $S \in C$ is the smallest r such that $S \in C(r)$. The motivation for studying the homology of simplicial complexes is the observation that two shapes can be distinguished by comparing their holes. For $k \in \mathbb{N}$, the Betti number b_k counts the number of k-dimensional holes. A k-dimensional Betti interval $[r_b, r_d)$ represents a k-dimensional hole that is created at the filtration value r_b, exists for $r_b \leq r < r_d$ and disappears at value r_d. We are interested in Betti intervals that persist for a large filtration range: They describe how the homology of $C(r)$ changes with r.

How do we construct a filtration? The Vietoris-Rips filtration $VR(X, r)$ [50, 58], which we will use here, is constructed as follows:

- The set of vertices X is the cloud of points under study.

- Given vertices x_1 and x_2, the edge $[x_1, x_2]$ is included in $VR(X, r)$ if the distance $d(x_1, x_2) \leq r$.

- If all the edges of a higher dimensional simplex are included in $VR(X, r)$, the simplex belongs to $VR(X, r)$.

A default choice for the distance d to study homology of 2D particle configurations is the Euclidean metric. Fig 15 displays two simplexes of a Vietoris-Rips filtration for the point cloud in Fig 14(a). Notice the appearance and disappearance of holes and isolated
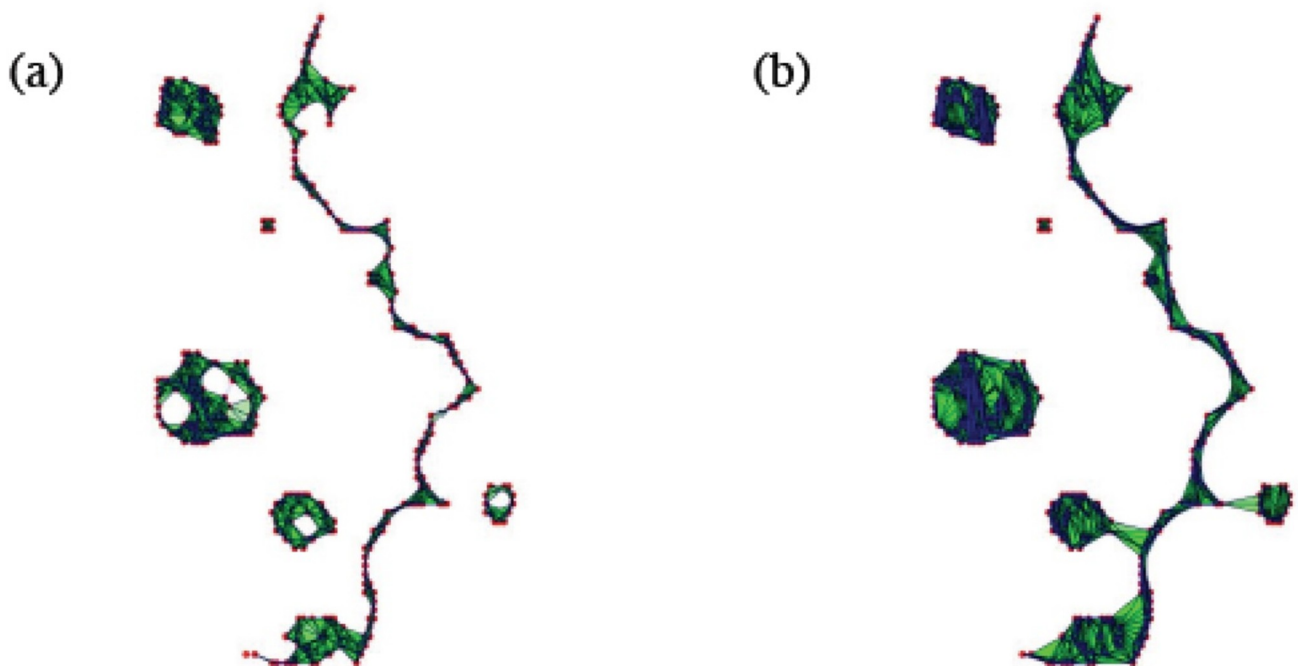

Fig 15. Visualization of the complexes $VR(X, r)$ for the point cloud depicted in Fig 14(a) when (a) $r = 6$ and (b) $r = 10$. For large enough r all the components merge in a single one. Holes appear and disappear as new connections are created, reflecting the overall point cloud arrangement.

components as the threshold distance r to connect points increases. This filtration is governed by three parameters:

- The maximum dimension d_{max}. This is the maximum dimension of the simplices to be constructed. The persistent homology (characterized by its Betti numbers) can be computed up to dimension $d_{max} - 1$. In this case $d_{max} = 2$, we consider points (0-simplices), edges (1-simplices), and triangles (2-simplices).

- The maximum filtration value r_{max} and the number of divisions N. These values define the filtered simplicial complexes to be constructed, for $r \in \left\{ 0, \frac{r_{max}}{N-1}, \frac{2r_{max}}{N-1}, \ldots, \frac{(N-2)r_{max}}{N-1}, r_{max} \right\}$.

Notice that for a set of P points, the full simplicial complex will have about $2^P - 1$ simplices in it. Therefore, d_{max} and r_{max} are usually slowly increased to get information without reaching computational limits. The computation is not too sensitive to the specific value of N. When r_{max} is greater than the diameter of the point cloud, all possible edges form and join all the points in one simplex.

For the readers' ease of use, we include more detailed definitions and intuitive examples in the Methods section. In the next two sections, we apply TDA to experimental and numerical images.

TDA of experiments. Let us consider the snapshots depicted in Fig 14. Fig 15 processes the earlier snapshot depicted in Fig 14(a), in which the green and magenta monolayers have made contact and started interpenetrating each other. Ras cells (magenta) are pushing back wt cells (green) towards the left. As they do so, there are islands of wt cells inside the Ras monolayer. How does TDA capture these features? After constructing the Vietoris-Rips filtration, there are two commonly employed graphical representations that visualize the persistent homology of a point cloud: barcodes and persistence diagrams [59].

Barcodes of a homology H_k depict Betti intervals $[r_b, r_d]$ for k-holes ($k > 0$) or connected components ($k = 0$) as the filtration parameter r varies. The homology class H_0 comprises the points forming the green/magenta interfaces. As the size filtration parameter r increases from zero, there appear edges joining these points, thereby forming clusters as illustrated by Fig 15 for specific values of r and indicated by the barcodes in Fig 16(a) for the selected range of r. The class H_1 further distinguishes compact components of the interface that are detached from the main part of the interface and form topological cycles, cf. the corresponding barcode in Fig 16(a). These components are islands of one cell type (phase) inside the bulk of the other phase.

Persistence diagrams represent the Betti intervals by points in a birth-death plane (see the Methods section for precise definitions). The x axis represents the filtration value r at which components/holes are created. The y axis represents the filtration value r at which they disappear. Those points less close to the diagonal (green) tend to mark robust underlying geometrical features. Fig 16(b) depicts the persistence diagram corresponding to Fig 14(a). Red circles mark connected components of the interface between cell monolayers and the magnitude of the filtration parameter r at which they disappear. As the filtration parameter increases, points comprising the main front merge rapidly in one component that absorbs neighboring clusters. They correspond to blocks of bars in the H_0 panel of Fig 16(a) that start at the lowest value of r. Blue asterisks represent the appearance (horizontal axis) and disappearance (vertical axis) of holes inside such clusters. The first column of asterisks represents the ten bars in the H_1 panel of Fig 16(a) that start at the same value of r and form four groups of bars, which end at about the same value of r. The remaining bars and asterisks are similarly related. They represent the new holes that form as the clusters merge, which gives an idea of the relative arrangement thereof. Relatively narrow barcodes produce points in the persistence diagram that are close packed.
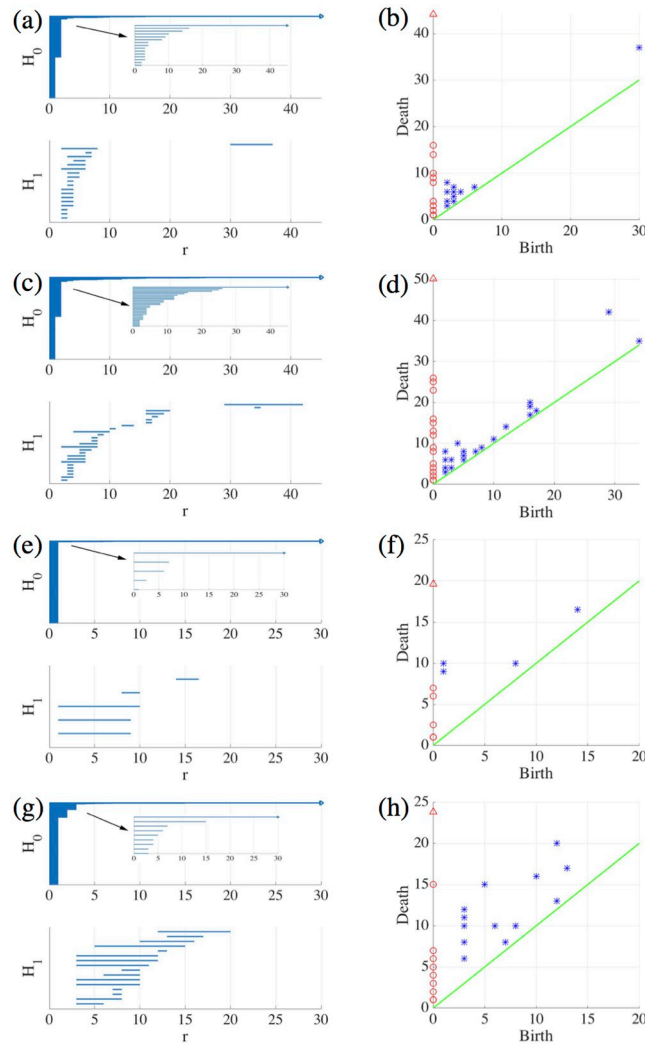
Fig 16. Barcodes (left) and persistence diagrams (right) for the homologies H_0 (circles) and H_1 (asterisks) of the interfaces separating cell types in images from experiments and numerical simulations. We use Vietoris-Rips filtrations with parameters N and r_{max}. (a)-(b) TDA from Fig 14(a) (experiments) with $N = 45$, $r_{max} = 45$; (c)-(d) TDA from Fig 14(b) (experiments) with $N = 45$, $r_{max} = 45$; (e)-(f) TDA from the leftmost panel in Fig 13 (numerical simulations) with $N = 60$, $r_{max} = 30$; (g)-(h) TDA from the rightmost panel in Fig 13 (numerical simulations) with $N = 30$, $r_{max} = 30$. Points in the persistence diagrams mark the beginning (birth) and end (death) of a bar (homology class) in the barcode. Triangles represent a component with infinite persistence. The green line is the diagonal.

Fig 16(c) and 16(d) display the barcodes and persistence diagram corresponding to Fig 14(b). Compared to the earlier snapshot of Fig 14(a) and its TDA in Fig 16(a) and 16(b), there are more islands of each phase in the bulk of the other: the invasion of Ras cells leaves pockets of wt cells inside their midst. The main interface has become more meandering and exhibits more fingers than in the earlier snapshot. As a consequence, the number of clusters or interface components is larger than at the earlier time. Similarly, there are more topological cycles, which reflects the larger number of islands of one cell type in the midst of the other cell type. Barcodes and persistence diagram are more spread out. This is further quantified by the Betti numbers b_j that count the number of elements in H_j, for $j = 0$ (clusters) and for $j = 1$ (holes), as depicted in Fig 17(a) and 17(b) for the snapshots shown in Fig 14. The trends are similar in the simulations, as shown in Fig 17(c) and 17(d).
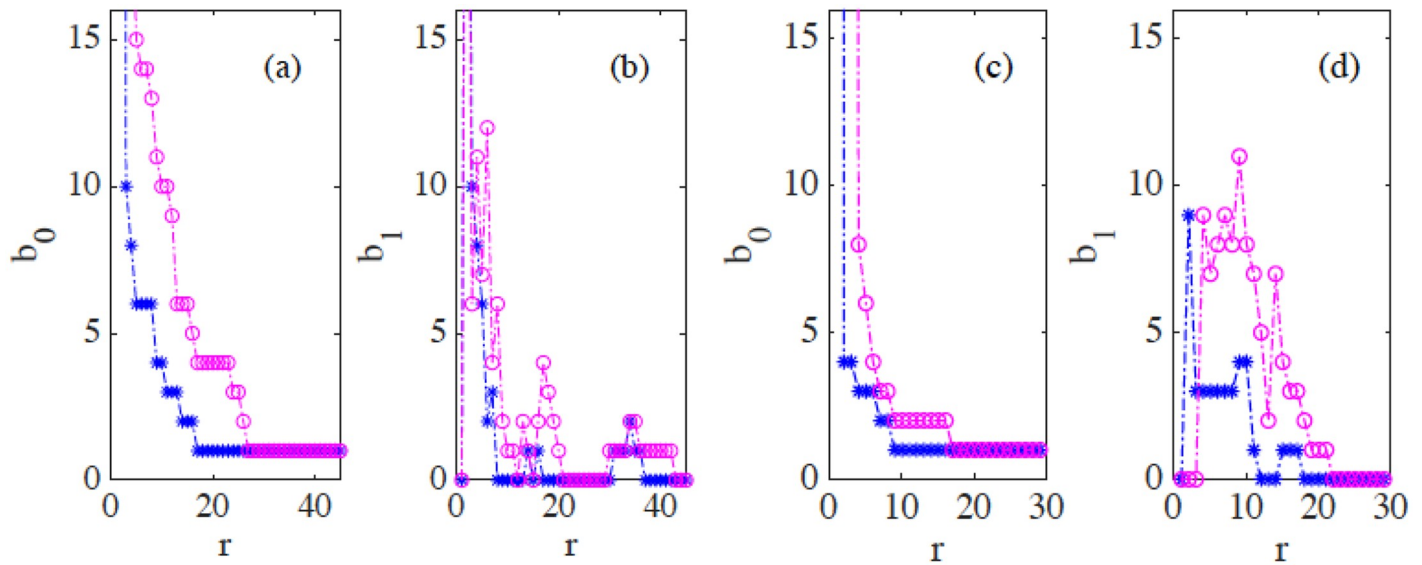
Fig 17. (a)-(b) Betti numbers versus filtration parameter diagrams for Fig 14(a) (blue asterisks, from S1 Data) and 14(b) (magenta circles, later time in the AMA experiment, from S2 Data) show that the number of clusters and holes in the interface between aggregates increases with time. (c)-(d) Same for the numerical simulations considered in Fig 16(e)–16(h) corresponding to the leftmost (S3 Data) and rightmost (S4 Data) panels in Fig 13. As a result of island formation and motion, which increases with time, Panels (a) and (c) show that the number of components decreases more slowly with r for the later time. The peaks in Panels (b) and (d) are similar for r below 20. In both cases, the interfaces formed at the later time display larger numbers of holes with larger sizes as a result of island formation. The additional peaks in Panel (b) near $r = 40$ correspond to islands that have already penetrated further inside the other cell population in the experiment.

https://doi.org/10.1371/journal.pcbi.1008407.g017

How do we characterize quantitatively variations in the persistence diagrams of point clouds? We have to introduce distances between diagrams to measure their differences. In the next section, we explain the bottleneck and Wasserstein distances using time series of numerical simulations, out from which we have generated much more complete data sets than available from experiments [27].

TDA of numerical simulations. As indicated in the previous section, to observe island formation, we have to tune the (negative) junction tensions when simulating antagonistic migration assays. In particular, $\Lambda_{12} < \frac{1}{2}(\Lambda_{11} + \Lambda_{22})$ facilitates mixing of wt and modified cell populations whereas $\Lambda_{12} > \frac{1}{2}(\Lambda_{11} + \Lambda_{22})$ produces population segregation. In Fig 12, Λ_{12} switches from population mixing to segregation after the two first snapshots. Then the pockets of green cells left behind by the advance of the interface shrink and start disappearing, as shown in the third and fourth snapshots of Fig 12. If mixing is weaker, as in Fig 11, the interface forms pronounced fingers, there are less islands and we do not need to change the junction tensions with time. In Fig 13, Λ_{12} randomly takes on a mixing value for one fifth of Ras and wt cells and on a segregation value for the others. The results of changing interface and island formation are qualitatively similar to those observed in experiments.

Let us now interpret the evolution shown in the panels of Fig 13 using TDA (see also S5 Video, out from which we have extracted 12 snapshots). Figs 16(e)–16(h) and 17(c) and 17(d) show the barcodes, persistence diagrams and Betti numbers for the marked sections of the leftmost and rightmost panels in Fig 13. As before, we represent the interfaces by point clouds. At $r = 0$, each point of the interface is a component. For the more regular interface of the leftmost panel in Fig 13, increasing r produces point components appearing as the short H_0 bars in Fig 16(e). These bars end at similar filtration values and appear as a single red circle in the persistence diagram of Fig 16(f). The main three islands correspond to the three intermediate bars in the inset of Fig 16(e), which disappear at larger filtration values. The lowest circle in

Fig 16(f) represents the point components, the three intermediate ones represent the islands in the barcode and their sizes. All clusters finally merge in the main front represented by the arrow on top of the vertical axis in Fig 16(f). Analysis of H_1 confirms that the intermediate circles/bars are round islands and not strings. Each component corresponds to a cycle represented by the three largest H_1 bars in Fig 16(e) and the two first asterisks in Fig 16(f), one of which represents the two bars of similar length. The two shortest bars represent holes formed as components merge during the filtration process and correspond to the two asterisks closer to the diagonal in Fig 16(f).

Fig 16(g) and 16(h) correspond to the more meandering interface of the rightmost panel in Fig 13. There are more points in the cloud representing the interface, whose irregularity results in different extinction values of r for the associate H_0 bars. The main seven islands correspond to the intermediate bars in the inset of Fig 16(g), and their extinction values in the persistence diagram give an idea of the distance to the main front or to another island. The fact that they are islands (enclosed by a boundary) is inferred from the H_1 bars in Fig 16(g). They correspond to the seven bars that appeared first, which are also represented by the first column five asterisks in Fig 16(h) having smaller r. Two of the asterisks correspond to two islands of similar size length each, which have bars of similar size. The length of the bars in the barcode or the distance of the asterisks from the diagonal in the persistence diagram give an idea of the island size. Additional H_1 bars represent holes created during the filtration process as components merge and give an idea of the relative arrangement of the islands or of the fingers in the main front. They are represented by the additional asterisks in Fig 16(h). The Betti numbers in Fig 17(c) and 17(d) show a larger number of island and holes as time increases from the leftmost snapshot in Fig 13 to the rightmost one. Compared to the TDA of experiments in Fig 16(a)–16(d), there are no gaps between bars and asterisks appearing for large r in Fig 16(e)–16(h). The reason is that the distance of islands to the main front is smaller for the simulation than for the experiment.

We have applied TDA to a time series of 12 snapshots (extracted from S5 Video, which visualizes the evolution of the numerically simulated interface in an AMA). Fig 16(e)–16(h) correspond to snapshots 2 and 10. For each of them, we calculate barcodes as explained above. To quantify the variations in the barcode patterns, we introduce distances between persistence diagrams that are stable against random perturbations [51, 52]. Given two persistence diagrams X and Y, their *bottleneck distance* is defined as

$$W_\infty(X, Y) = \inf_{\varphi:X \to Y} \sup_{x \in X} \|x - \varphi(x)\|_\infty. \tag{13}$$

Here φ ranges over all bijections between the persistence diagrams (taking the diagonal into account, see the Methods section) and $\|x\|_\infty = \max_i\{|x_i|\}$ is the usual L_∞-norm over the points x of the persistence diagram X. The bottleneck distance is an example of the more general *Wasserstein distance* between persistence diagrams:

$$W_{q,p}(X, Y) = \left[\inf_{\varphi:X \to Y} \sum_{x \in X} \|x - \varphi(x)\|_p^q \right]^{1/q}, \quad W_q(X, Y) = W_{q,\infty}(X, Y). \tag{14}$$

We have $W_\infty(X, Y) = W_{\infty,\infty}(X, Y)$. Fig 18(a) represents the matrix of bottleneck distances between the persistence diagrams of the 12 frames in S5 Video. These distances are stable in the sense that a small perturbation in the input filtration leads to a small perturbation of its persistence diagram in the bottleneck distances [51] (q-Wasserstein distances share that property too). Efficient algorithms to compute these distances are discussed in [52]. Techniques enabling us to input topological features into deep neural networks and learn task-optimal
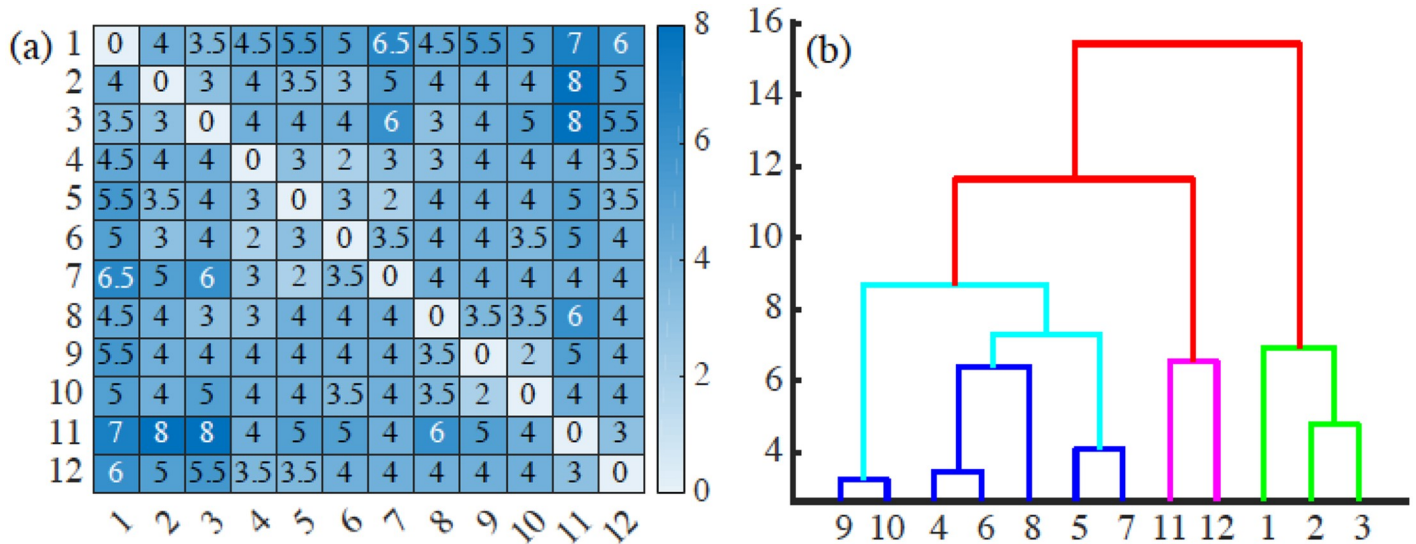
Fig 18. (a) Bottleneck distance matrix for the interfaces between cells populations appearing in the 12 snapshots forming S5 Video and (b) associated dendrogram illustrating how the interfaces between cell populations can be grouped in clusters. Interfaces in frames 1-3, 4-10, 11-12 can be grouped together, and the last two groups are closer to each other than to the initial frames. These groupings reflect similarities between frames as they succeed one another, and the disruptions between frames reflect significant topological changes of the interfaces (e.g., detachment and reattachment of islands). S5–S16 Data have been used to draw this figure.

https://doi.org/10.1371/journal.pcbi.1008407.g018

representations during training are proposed in [60]. We could use a neural network approach if we wanted to relate them to a specific pattern, but that is not the case here.

Once we have a matrix of distances, we resort to unsupervised clustering methods to classify the frames in similar blocks. This classification automatically extracts the similitudes and changes between interfaces at different times of the AMA. Fig 18(b) displays a dendrogram obtained by agglomerative hierarchical clustering using Ward's method [61] and the bottleneck distance. A dendrogram consists of U-shaped lines that connect data points (which, in this case, are the interfaces in each frame) in a hierarchical tree. The height of each U represents the distance between the interfaces that are connected by it. A dendrogram is not a single set of clusters, but a multilevel hierarchy. For a given dendrogram, we identify the natural cluster divisions relying on the inconsistency coefficient [62]. The latter compares the height of a link in a cluster hierarchy with the average height of links below it: larger inconsistency coefficients mark natural divisions [62]. By defining a cutoff value for the inconsistency coefficient, we automatically detect clusters. A 0.9 cutoff in the inconsistency coefficient detects 3 clusters, corresponding to times {1, 2, 3}, {4, 5, 6, 7, 8, 9, 10}, and {11, 12} in S5 Video. Thus, we distinguish the initial period (interfaces close to the original connected one, times {1, 2, 3}), the intermediate period (a phase in which a few islands form and advance, times {4, 5, 6, 7, 8, 9, 10}), and the final period (severe disruption with the abrupt formation of several islands, times {11, 12}). In this way, we gain insight on the time evolution: how fast the interfaces experience significative changes, and when abrupt changes do occur and mark the onset of a new cluster of frames. The chosen cutoff 0.9 is a critical value: Increasing it, we find only one cluster. Lowering the cutoff, we detect 5 clusters of frames, corresponding to times {1, 2, 3}, {4, 6, 8}, {5, 7}, {9, 10}, {11, 12}. With respect to the 0.9 value, the intermediate cluster splits into other 3, reflecting detachment, reattachment and slow progression of islands.

We can also obtain clusters by setting distance cutoffs, i.e., a height in the dendrogram of Fig 18(b). The height of a link between clusters represents the distance between them, which is called cophenetic distance. It is possible to calculate the correlation between the cophenetic

distance and the distances in the matrix of Fig 18(a) (the cophenetic correlation [63]). For our simulations, Ward's hierarchical clustering provides the largest cophenetic correlation when compared with other clustering approaches, such as single-linkage [53]. Thus, the Ward dendrogram gives the most faithful representation of the distance matrix, which is why we use it. Depending on the chosen cutoff height in Fig 18(b), we obtain one, two, three, four or five clusters. The sets of three or five clusters are the same as before (this does not necessarily occur in general). Alternatively, we can use the K-means algorithm to group the interfaces in clusters, selecting an optimal number of clusters by silhouette or elbow type criteria [64, 65]. In our case, K-means with 3 or 5 clusters produces the ones already obtained (this does not necessarily occur in general). We have illustrated TDA with a relatively short time series, but it is clear that it could be used for automatic detection of topological changes in much longer time series, or to quantify how close the interfaces obtained from different simulations or experiments are. In the Methods section, we have used TDA with 1-Wasserstein distance to interpret the evolution of tissue interfaces in the numerically simulated spreading assay of Fig 4.

Conclusions

We have modeled how epithelial cell aggregates advance through empty spaces (wound healing, tissue spreading) and collisions between aggregates (tumoral invasion) using an active vertex model with dynamics for cell centers that includes collective tissue forces [39], and velocity alignment and inertia [16]. The active vertex model implements exchanges of neighboring cells automatically (T1 transitions) and uses the SAMoS software. Compared with particle models with underdamped dynamics, our model accounts for fingering instabilities in spreading tissue without having to introduce leader cells [16]. Compared to continuum models [38], stochasticity enables our model to reproduce the observed fast irregular oscillation of cell velocities in fingers [43] and the spatial autocorrelation of the velocity [11]. Our underdamped AVM predicts that cells at the interface and the fingers have larger area than those well inside the tissue, which has been corroborated by recent experiments [44]. We also observe in numerical simulations of tissue spreading that the velocity of the fastest cells in a finger may oscillate with a short period in a range between 30 minutes to about one hour. A similar short period oscillation has been observed in experiments; cf Fig 101A in L. Petitjean's PhD thesis [43]. Thus, for spreading tissue, detailed comparison to experimental data provides a quantitatively accurate description of cell motion (speed, velocity correlation function and polar order parameter). For antagonistic migration assays, we have reproduced collisions in which one cell population pushes back another whereas both populations mix forming different types of interfaces. The key element to model mixing is to keep different junction parameters for the two colliding tissues: the invading cells are liquid like whereas the receding tissue comprises solid like cells. In addition, a fraction of cells favor mixing, the others segregation, and that these cells are randomly distributed in space. Thus characterized, numerical simulations produce outcomes similar to those observed in experiments [27]. Compared to particle models, ours includes active vertex forces between cells that keep them together preventing gaps and keeping track of cellular compression, enlargement and changes of area. To characterize automatically the dynamics of islands and the rugged interface between aggregates, we have introduced topological data analyses of experiments and time series from numerical simulations. In collisions between aggregates, the interface between wt and Ras cell populations roughens and islands appear. The persistence diagrams of Homology classes 0 (clusters) and 1 (cycles) spread out and the number of these classes given by the corresponding Betti numbers increases. Using time series of data generated by numerical simulations, we have explained how to cluster interfaces using distance matrices based on the bottleneck distance between their

persistence diagrams, which are stable to perturbations in the process. Despite the amount of data from experiments being limited, disruptive events such as island and cluster formation can be automatically captured by topological data analyses of numerical simulations and contrasted with experiments. Similarly, the Wasserstein distance between images enables us to track and classify automatically the evolving shapes of interfaces between cell populations by using time series from experimental or numerical studies. These techniques of topological data analysis are scalable and could be used in studies involving large amounts of data whenever available.

Our results allow to extract parameter values and to determine *biologically relevant* physical mechanisms for characterizing confluent motion of cellular aggregates, as described above. In particular: (i) cells at the interface are larger, inform the aggregate motion and are influenced by it, without needing leader cells to form fingers at the interface; and (ii) in colliding cellular aggregates, the solid or liquid like character of the cells (as determined by their junction parameters) decides the way the invasion goes. These aspects of our model are important in ascertaining how the biophysical features of materials influence tissue/organ regeneration [45]. Our work provides researchers in the field with useful tools to gain biological insight, to devise and to interpret data from experiments. To enhance the value of our results, e.g, for studies of metastatic cancer, future works may add cellular mechanisms such as Notch signaling dynamics [46], models of epithelial/mesenchymal transition and cancer stem cell formation [47, 48] to our vertex model. This venue has been successfully followed in studies of angiogenesis [49].

Methods

Topological data analysis

For the reader's ease of use, this section makes more precise some definitions and includes simple examples to provide an intuitive idea of the meaning of persistent homology features. It is structured as follows. First, we give more precise definitions of persistent homology concepts. Second, we present applications to simpler synthetic data, for an easier visual interpretation of the results in the main text when interfaces are formed by many connected components. Finally, we discuss how to extract information on front roughness from numerical or experimental data, when interfaces define a single component.

Basic definitions of persistent homology and examples. As said before, a finite set of data points may be considered a sampling from the underlying topological space. Data structure can be investigated by creating connections between proximate data points, varying the scale over which these connections are made, and looking for features that persist across scales [59]. Homology distinguishes topological spaces (e.g., annulus, sphere, torus, or more complicated surface or manifold) by quantifying their connected components, topological circles, trapped volumes, and so forth. Persistent homology describes how the homology of a nested family of simplicial complexes changes with respect to a defining parameter. What is a simplicial complex S? To define it, we need three elements [66]:

- A set of points X in a space of dimension D.

- Sets of k-simplices $[v_0, v_1, \ldots, v_k]$ with vertices $v_i \in S$, $i = 0, 1, \ldots, k$, for each $k \geq 1$. A k-simplex is a k-dimensional polytope which is the convex hull of its $k + 1$ vertices:

$$[v_0, v_1, \ldots, v_k] = \left\{ \theta_0 v_0 + \theta_1 v_1 + \ldots + \theta_k v_k \mid \sum_{i=0}^{k} \theta_i = 1, \ \theta_i \geq 0, \ i = 0, 1, \ldots, k. \right\}$$

The vertices must be affinely independent, i.e., the difference vectors $v_1 - v_0, \ldots v_k - v_0$ must be linearly independent. The k-simplex is oriented so that an odd permutation of the points in $[v_0, \ldots, v_k]$ reverses its sign.

- A k-simplex has $k + 1$ faces, each constructed by deleting one of the vertices. The faces must satisfy the following property: If $[v_0, v_1, \ldots, v_k]$ belongs to the simplicial complex S, then all its faces must also be in the simplicial complex S. This can be made more precise. The set of all k-simplices in S is a vector space C_k. The boundary of a k-simplex is the union of all its $(k - 1)$-subsimplices. For each $k \geq 1$, the boundary map $\partial_k: C_k \to C_{k-1}$ is the linear transformation defined by

$$\partial_k([v_0, \ldots, v_k]) = \sum_{j=0}^{k} (-1)^j [v_0, \ldots, \hat{v}_j, \ldots, v_k], \tag{15}$$

where $[v_0, \ldots, \hat{v}_j, \ldots, v_k]$ is the $(k - 1)$-simplex obtained by removing the vertex $\hat{v}_j$ from $[v_0, \ldots, v_k]$.

The motivation for studying the homology of simplicial complexes is the observation that two shapes can be distinguished by comparing their topological features. A disk is not a circle because the disk is solid, while the circle has a hole. Similarly, a circle is not a sphere, because the sphere encloses a two dimensional hole, whereas the circle encloses a one dimensional hole. To distinguish topological features, we need several definitions. Boundary operators connect the vector spaces C_k into a *chain complex* $\ldots \to C_{k+1} \to C_k \to C_{k-1} \to \ldots \to C_0 \to 0$. The kernel and image of boundary operators determine k-cycles $Z_k = \mathrm{Ker}\{\partial_k: C_k \to C_{k-1}\}$ and k-boundaries $B_k = \mathrm{Im}\{\partial_{k+1}: C_{k+1} \to C_k$, respectively. Since a boundary has no boundary [66], B_k is a subspace of Z_k. Thus, C_k is the vector space of all k-chains in the simplicial complex S_r, Z_k is the subspace of C_k consisting of k-chains that are also k-cycles, and B_k is the subspace of Z_k consisting of k-cycles that are also k-boundaries. We say that two k-cycles are homologous (equivalent) if they differ by a k-boundary. This equivalence relation splits Z_k in equivalence classes denoted by $[z]$ if $z \in Z_k$. The kth homology of S_r is the quotient set $H_k = Z_k/B_k$ comprising all equivalent k-cycles. The dimension of H_k, $b_k = \dim Z_k - \dim B_k$, is the kth *Betti number*. In terms of the topological characteristics, b_k is the number of independent holes of dimension k. For instance, b_0 is the number of connected components, b_1 is the number of topological circles, b_2 is the number of trapped volumes, and so on. The topology of a simplicial complex may be described by the sequence of Betti numbers, $\mathbf{b} = (b_0, b_1, \ldots)$. For instance, a topological circle has $\mathbf{b} = (1, 1, 0, \ldots)$, a topological torus has $\mathbf{b} = (1, 2, 1, 0, \ldots)$, and a topological sphere has $\mathbf{b} = (1, 0, 1, 0, \ldots,)$. Betti numbers are a topological invariant, meaning that topologically equivalent spaces have the same Betti number.

The *Vietoris-Rips filtration* (VRF) constructed in the main text provides an example. For each value of a scale proximity parameter $r > 0$ and a given set of points X, we form a simplicial complex $VR(X, r) = S_r$ by finding all k-simplices such that all pairwise distances between their points are smaller than r. The simplicial complex S_r comprises finitely many simplices such that (i) every nonempty subset of a simplex in S_r is also in S_r, and (ii) two k-simplices in S_r are either disjoint or intersect in a lower dimensional simplex. Clearly, if $r_1 \leq r_2$, then $S_{r_1} \subset S_{r_2}$. In S_r, 0-simplices are the data points, 1-simplices are edges, connections between two data points, 2-simplices are triangles formed by joining 3 data points through their edges, 3-simplices are tetrahedra, and we obtain more complicated structures for higher dimensional simplices. Fig 19 shows a simple example of Vietoris-Rips complex, its barcodes for H_0 and H_1, and Betti numbers for different values of the proximity parameter r. Fig 19(d)–19(g) visualize the filtration process: For a grid of values of the filtration distance parameter r we depict balls
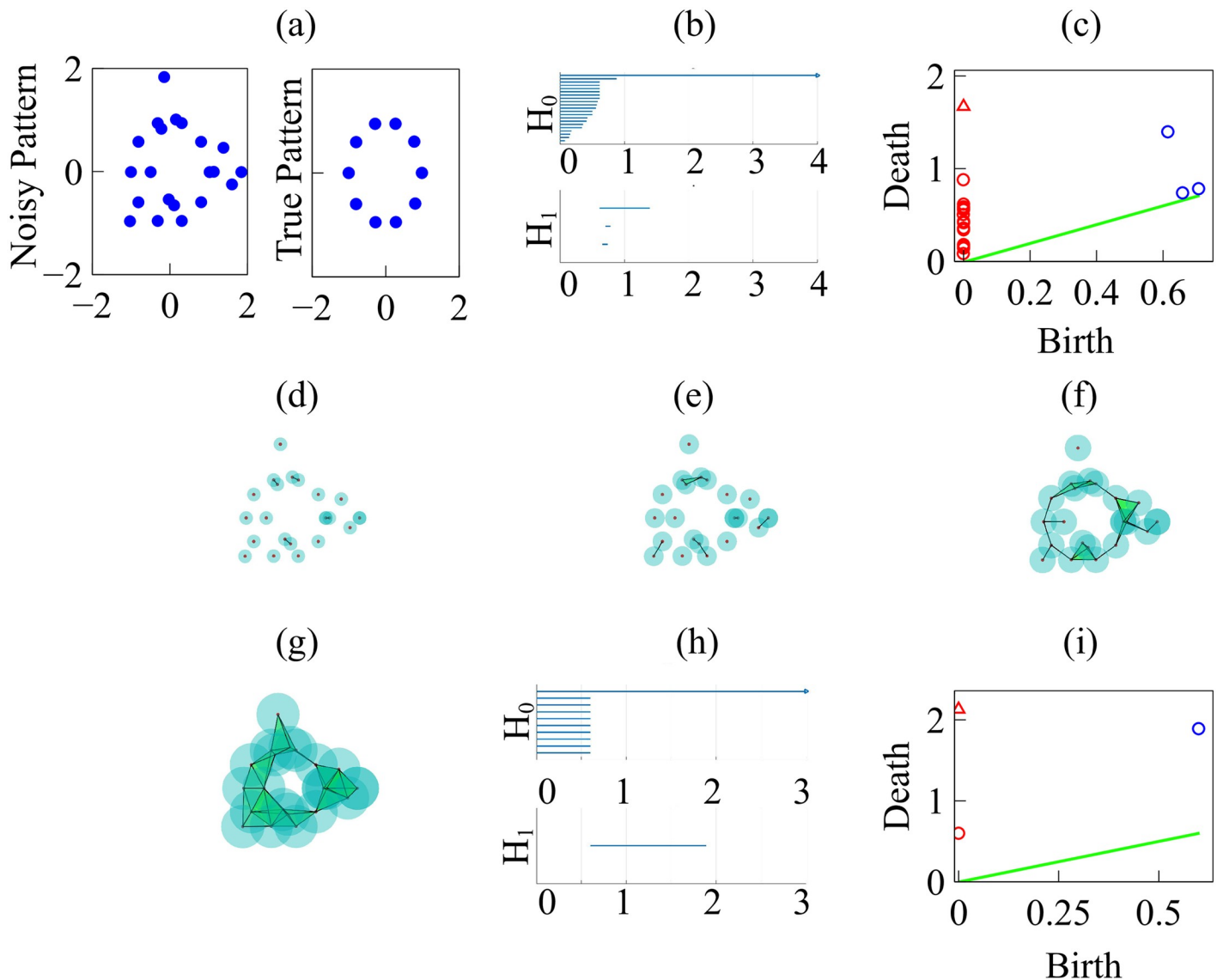
Fig 19. Persistent homology for data on a circle. (a) Noisy versus true circle data. (b) Barcodes for the Betti numbers b_0 (H_0) and b_1 (H_1) for the noisy data. (c) Persistence diagram for the noisy data with $r_{max} = 4$ and $N = 100$. (d)-(g) Vietoris-Rips simplicial complexes formed from the noisy data increasing the filtration parameter r. (h) Barcodes for the Betti numbers b_0 and b_1 for clean data on the circle. (i) Persistence diagram for clean data on the circle, with $r_{max} = 3$ and $N = 100$.

https://doi.org/10.1371/journal.pcbi.1008407.g019

centered at points with radius r and count the components formed. Topological features that persist on wide intervals of r characterize the simplicial complexes of the dataset. To visualize persistent homology, we plot the *barcodes* and *persistence diagrams*. The barcode of a homology H_k depicts each class by its corresponding *Betti intervals* (r_b, r_d). Initially we have one per point, represented as a bar in the top panel of Fig 19(b). As components merge, the number of bars diminishes. For Fig 19(f) we have two components, represented by the two top bars in panel H_0 of Fig 19(b). For Fig 19(g), we have one component represented by the top bar. The arrow means that this component persists for larger r values. Similarly, the largest bar in panel H_1 of Fig 19(b) represents the dominant hole, observed in Fig 19(f)–19(g). This bar, and hole, correspond to the circle in Fig 19(c) placed furthest from the diagonal. The two small bars correspond to the circles in Fig 19(c) which are closest to the diagonal. As seen in Fig 19(f), they

form and disappear as components merge during the filtration process. In *persistence diagrams*, for the selected equally spaced grid of values of r, we represent each bar in the barcode by a point (r_b, r_d) in the Cartesian plane. A point (x, y) of the persistence diagram with multiplicity m represents m features that all appear for the first time at scale x and disappear at scale y. The height of a point over the diagonal, $(y - x)$, gives the length of the corresponding bar in the barcode and is called the *persistence* of the feature. In addition to the off-diagonal points, the persistence diagram also contains each diagonal point, (x, x), counted with infinite multiplicity. These additional points are needed for stability (discussed below) and make the cardinality of every persistence diagram infinite, even if the number of off-diagonal points is finite. Points near the diagonal are inferred to be noise while points further from the diagonal are considered topological signal [59]. Coloring differently different homologies (H_0, H_1, etc) we can accumulate plenty of topological information in one 2D persistence diagram. For example,

- b_0 gives the number of components for a filtration value r, thereby providing the number of clusters in H_0 for that r. We can use this information and the knowledge of the distance, to find out which points belong to which cluster. All the points of a cluster are connected in a simplex. Thus, we have a clustering strategy.

- Similarly, b_1 gives the number of 1-dimensional holes in H_1 for a given value of the proximity parameter r. These holes may be inherent to the shape of a cluster or appear when basic clusters connect to more distant ones. See Figs 20 (one island), 21 (two islands) and 22 (seven islands). Thus, H_1 contains information on both the structure of basic clusters and their relative arrangements. The accompanying barcodes may characterize interfaces or data sets.
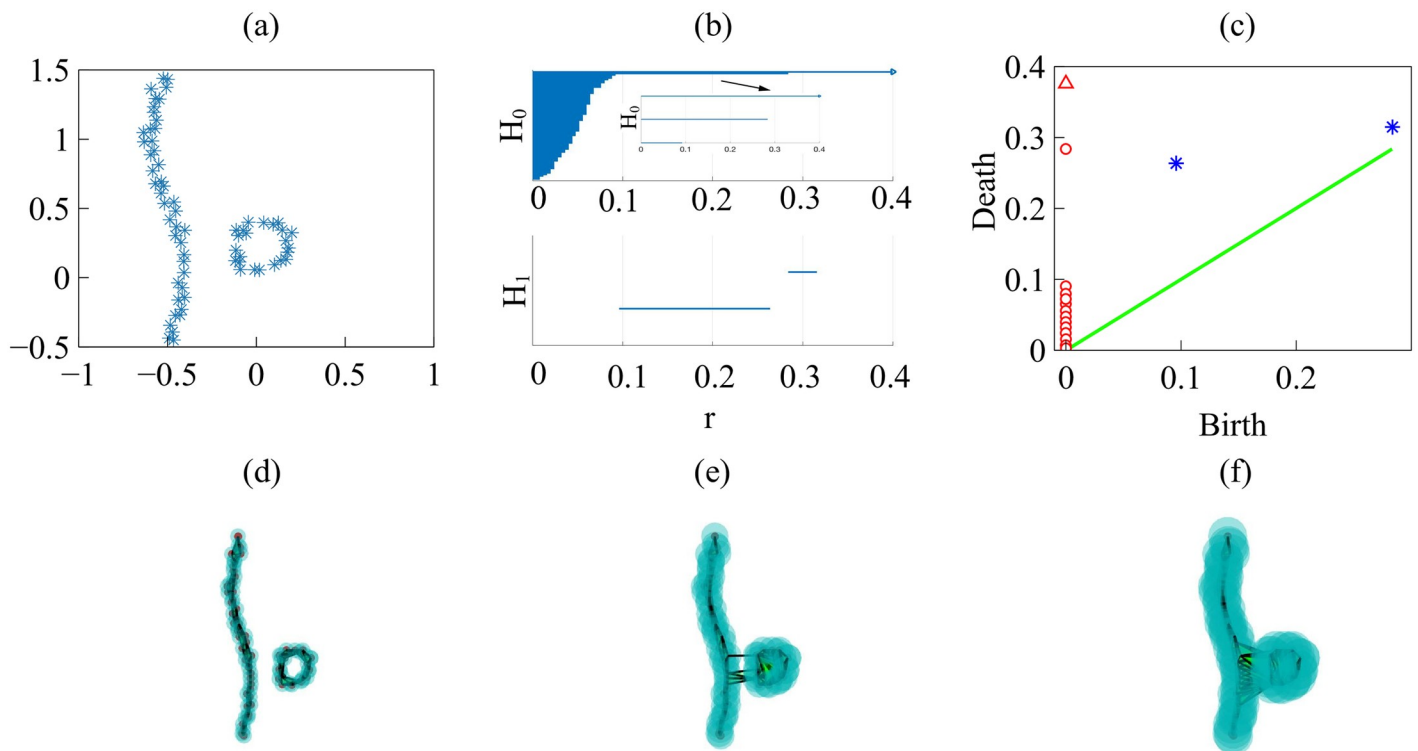

Fig 20. Persistent homology for the border of two colliding populations. We should stress that these examples use schematic figures, with clear fronts, not results from experiments or simulations (which have larger noise, and less clear features). Thus, the persistent homology of schematic figures is clearer and easier to interpret. (a) Interface separating the two populations. (b) Barcodes for the Betti numbers b_0 (H_0) and b_1 (H_1), and (c) Persistence diagram with $r_{max} = 0.4$ and $N = 100$. (d)-(f) Vietoris-Rips simplicial complexes formed increasing the filtration parameter r. S17 Data have been used to draw this figure.
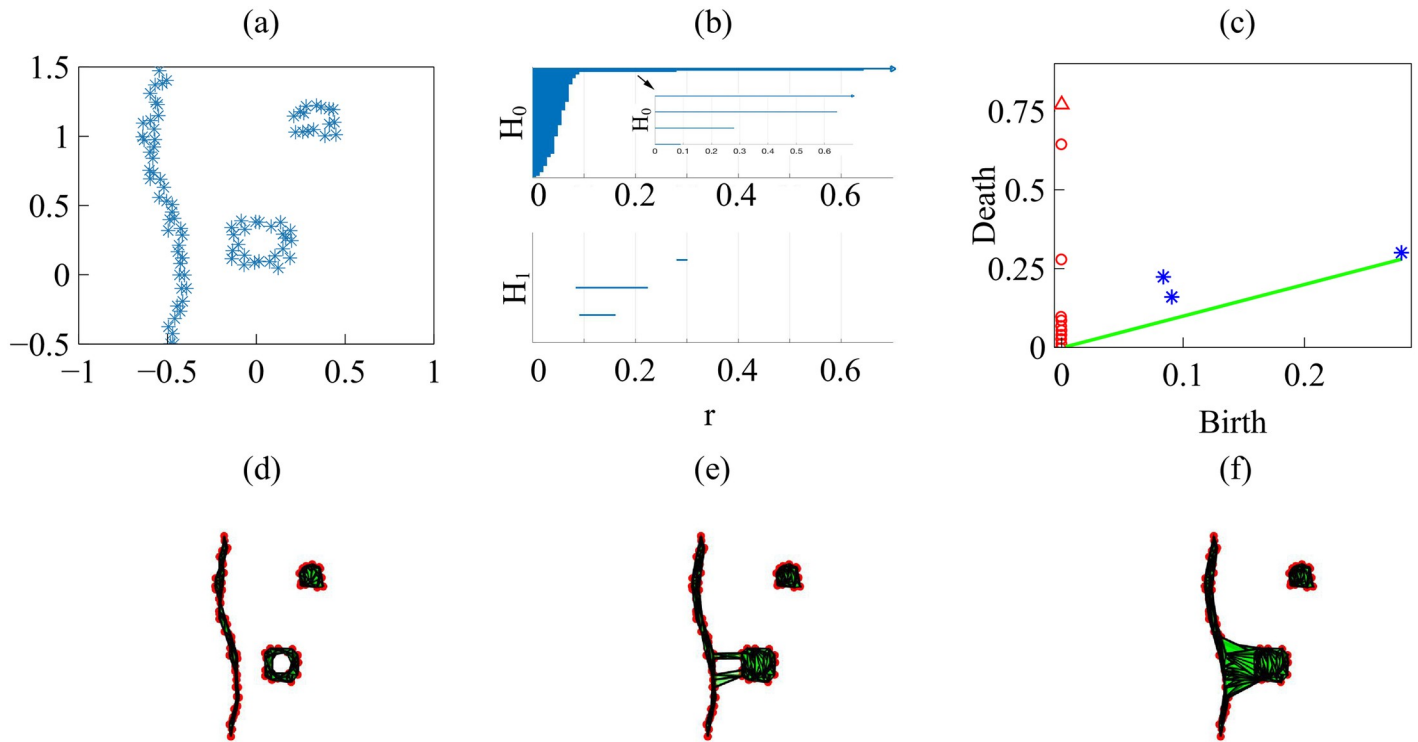
Fig 21. Same as in Fig 20 except that there are one interface and two islands. Parameter values: $r_{max} = 0.7$ and $N = 100$. S18 Data have been used to draw this figure.

https://doi.org/10.1371/journal.pcbi.1008407.g021

Next, we consider the homology of point clouds defining an interface between two populations, see Figs 20–22.

Fig 20(d)–20(f) visualize the filtration process: For a grid of values of the filtration distance parameter r, we depict balls centered at points with radius r and count the components (clusters) formed. Initially we have as many clusters as points. As the filtration parameter r increases, clusters merge and their number is reduced to two: the continuous front and the detached island in Fig 20(d), which are represented by the two top bars in H_0 of Fig 20(b) and circles in Fig 20(c). Fig 20(e) exhibits only one cluster represented by the top bar in Fig 20(b),
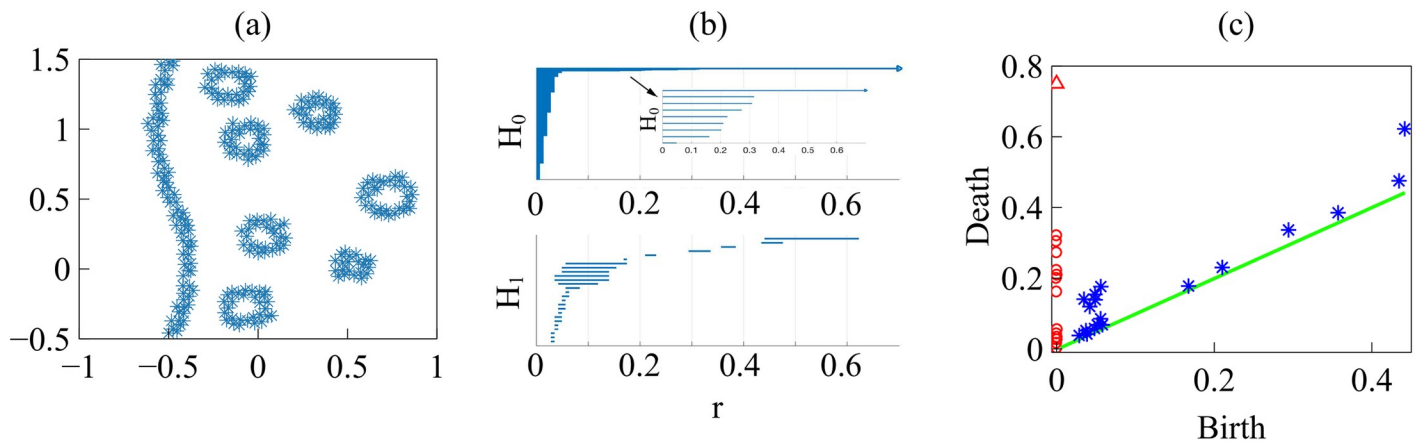

Fig 22. Same as Fig 20 except that there are one interface and seven islands. Parameter values: $r_{max} = 0.7$ and $N = 100$. S19 Data have been used to draw this figure.

https://doi.org/10.1371/journal.pcbi.1008407.g022

which persists for larger values of r and is represented by an arrow. Similarly, the largest bar in H_1 of Fig 20(b) represents the dominant hole defined by the island border in Fig 20(d). This bar, and island, corresponds to the asterisk in Fig 20(c) placed furthest from the diagonal. The small bar corresponds to the asterisk in Fig 20(c) which is closest to the diagonal. As seen in Fig 20(e), the small bar forms and disappears as holes form when clusters merge during the filtration process.

Figs 21 and 22 can be similarly interpreted. Fig 21(d)–21(f) visualize the filtration process. Initially we have one cluster per point, represented as a bar in the top panel of Fig 21(b). As the filtration parameter r increases, clusters merge and the number of bars diminishes. Fig 21(d) exhibits three components, represented by the three top bars in H_0 of Fig 21(b) and circles in Fig 21(c). There are two clusters in Fig 21(e) represented by the two top bars in Fig 21(b). Similarly, the two largest bars in H_1 of Fig 21(b) represent the two dominant holes defined by the island borders in Fig 21(d). These bars, and islands, correspond to the two asterisks in Fig 21(c) furthest from the diagonal. The small bar corresponds to the circle in Fig 21(c) which is closest to the diagonal. As seen in Fig 21(e), it forms and disappears as holes form when clusters merge during the filtration process.

Fig 22 exhibits an increased number of islands and it is described in the same manner. The main eight components correspond to one interface and seven islands. They are represented by the top bars in H_0 of Fig 22(b) and circles in Fig 22(c). Seven bars represent the seven dominant islands, associated to the seven largest bars in H_1 of Fig 22(b) and asterisks in Fig 22(c). The remaining bars represent gaps in the simplicial structure formed during the filtration process. The largest one represent a late hole appearing due to the fact that some islands are far from the main interface front, and corresponds to the final asterisk distant from the diagonal.

Tracking moving interfaces by tracking slices. When the interface is simply connected, homology studies of the boundary points, or all the population points in the plane, hardly give information on its roughness. Instead we may study the evolution of the H_0 homology of slices $x = c$ as c varies, see Figs 23 and 24. We choose as r_{max} the degree of roughness we want to capture, the 'scale' at which we wish to 'resolve'. Consider the two dimensional region occupied by cells (magenta patch) in Fig 24(a). We build a square mesh of step smaller than r_{max} and consider the points that are inside the occupied region. We define a matrix on the mesh $M(i, c)$, equal to one at points inside the patch, and zero outside. For each slice $x = c$, the points at which $M(i, c) = 1$ define a point cloud, we evaluate the zero homology of that cloud using as maximum filtration value r_{max}. The variation of the Betti number $\mathrm{b}_0(r_{max}, c)$ with c measures how irregular the front is: the larger b_0 is, the rougher the interface. As shown in Fig 24 corresponding to the spread assay of Fig 4(a), the rougher lower front has a larger maximal Betti number b_0 than the upper front; cf Fig 24(c) versus Fig 24(b). However, the bigger and more persistent fingers of the upper front cause b_0 to decay more slowly than the corresponding Betti number of the lower front. The fingers of the latter are smaller in size, cf Fig 24(a). Fig 23 visualizes the idea on fragments of this configuration. This complementary slice by slice study gives qualitative information on the shape. This strategy would allow to study the time evolution of 2D interfaces comparing the information obtained for each time, and comparing the effect of different controlling parameters on each population, as done in the main text for mixing interfaces.

Fig 25 quantifies differences between configurations by means of distances between computational images [67]. To compare images and shapes, we first have to define measures $\rho^j(x)$, $x \in \Omega$, over grids of images j ($j = 0, 1, 2, \ldots$). Given two images, $j = 0$ and 1, with measures $\rho^0(x)$ and $\rho^1(x)$, we define their *Wasserstein-1 distance* as a particular type of optimal transport
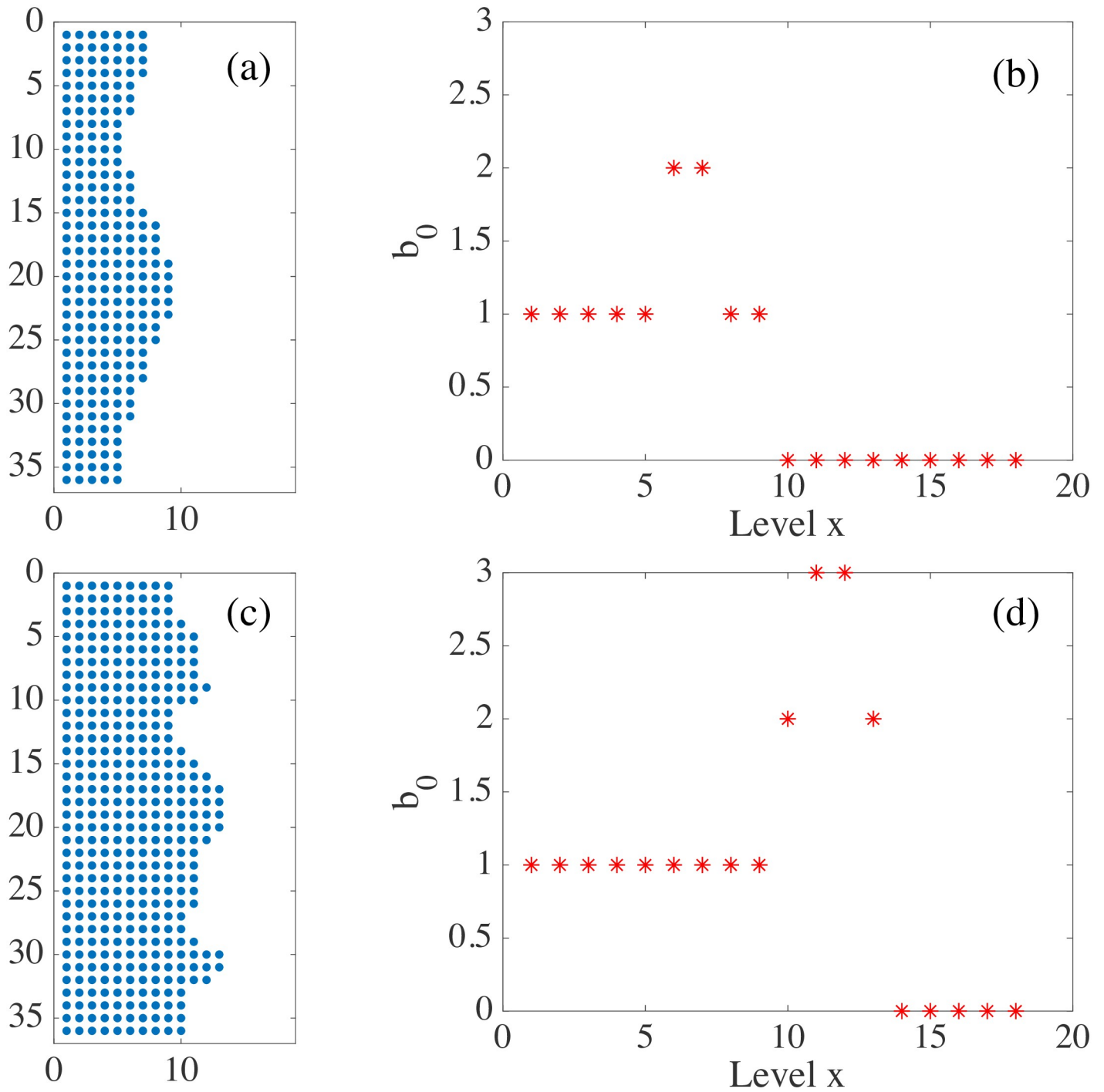
Fig 23. Interfaces of different roughness and Betti numbers $b_0(r_{max})$ for the slices $x = c$ of the displayed point clouds, as c increases. We see how the variation in b_0 gives an idea of the interface roughness, at the scale $r_{max} = 1$ in this case. S20 and S21 Data have been used to draw this figure.
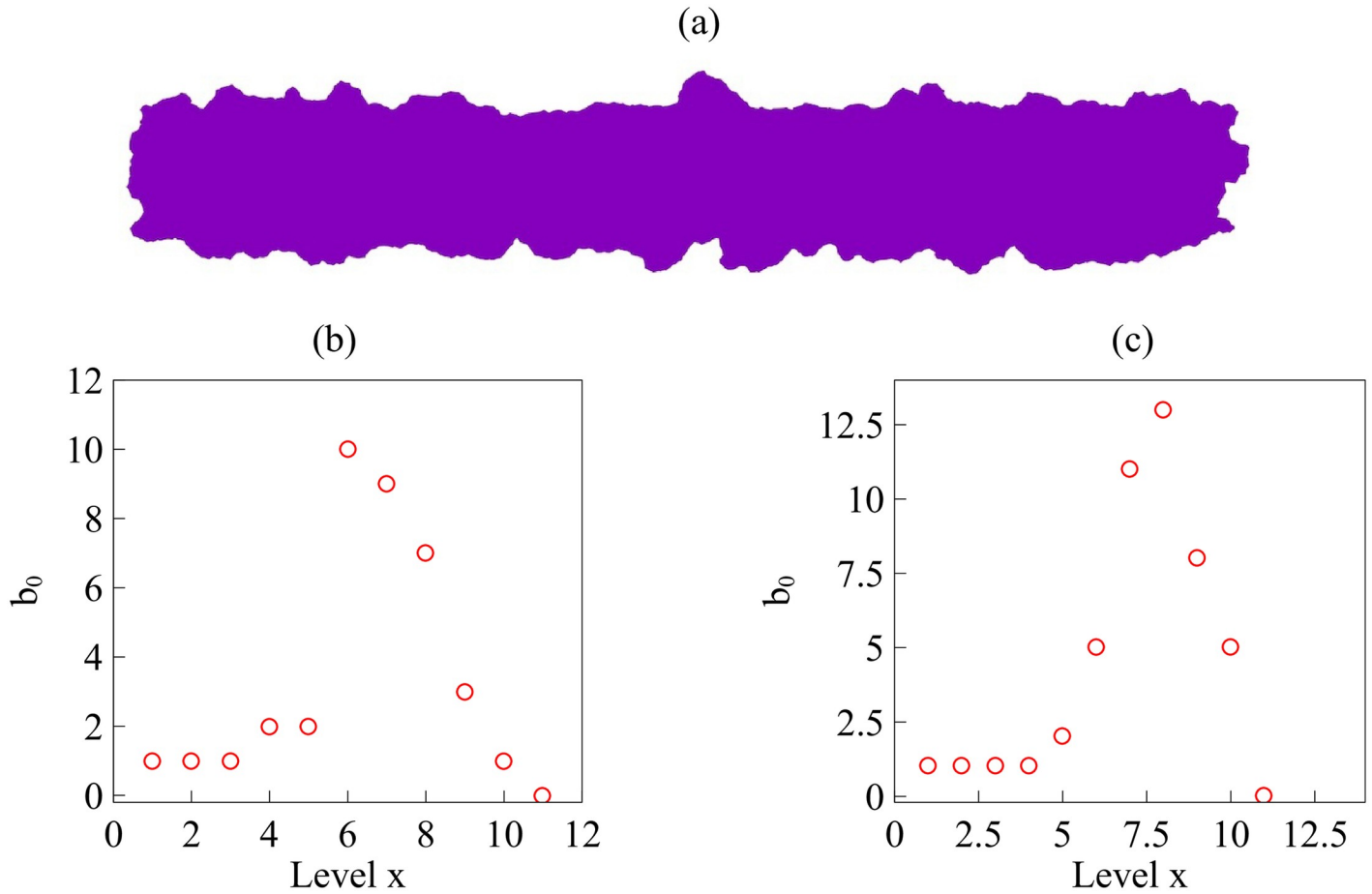
https://doi.org/10.1371/journal.pcbi.1008407.g023

Fig 24. For the expanding strip in (a), we get the Betti numbers b_0 for the upper front (b), and for the lower one (c). The lower front is rougher (larger b_0) but its fingers are shorter (b_0 decays faster to one and zero), whereas the upper front has a dominant persistent finger. We have set r_{max} = 1 mm in the scale of the image Fig 4. S22 and S23 Data have been used to draw this figure.

https://doi.org/10.1371/journal.pcbi.1008407.g024

distance [67],

$$W_{1,p}(\rho^0, \rho^1) = \inf_{\pi:\Omega\times\Omega\to[0,\infty)} \int_{x,y\in\Omega} \|x - y\|_p\, \pi(x, y)\, dxdy, \quad 1 \le p \le \infty,$$

$$\text{subject to} \int_{y\in\Omega} \pi(x,y)\, dy = \rho^0(x)\ (x \in \Omega), \quad \int_{x\in\Omega} \pi(x,y)\, dx = \rho^1(y)\ (y \in \Omega), \quad \pi(x,y) \ge 0\ (x, y \in \Omega).$$

Here $\pi(x, y)$ are probability measures over $\Omega \times \Omega$ whose marginals are $\rho^0(x)$ and $\rho^1(y)$. For images simply composed of points, which is the case of persistence diagrams, ρ^0 and ρ^1 are point measures and this definition recovers Eq (14). Fig 25(e)–25(g) use $W_{1,\infty}$ distances between the four snapshots depicted in Fig 25(a)–25(d) to construct dendrograms for agglomerative hierarchical clustering following Ward's method [61]. In Fig 25(e), we consider the four snapshots of Fig 25(a)–25(d): the smoother first two snapshots are clustered together and so are the last two rougher snapshots. Fig 25(f) does the same considering separately the left and right interfaces in each of the snapshots of Fig 25(a) to 25(d). In Fig 25(g), we cluster the $b_0(r_{max})$ profiles obtained for the four right and left fronts. Dendrograms enable us to do this. Enforcing cutoffs on the inconsistency coefficients we obtain a natural division in three
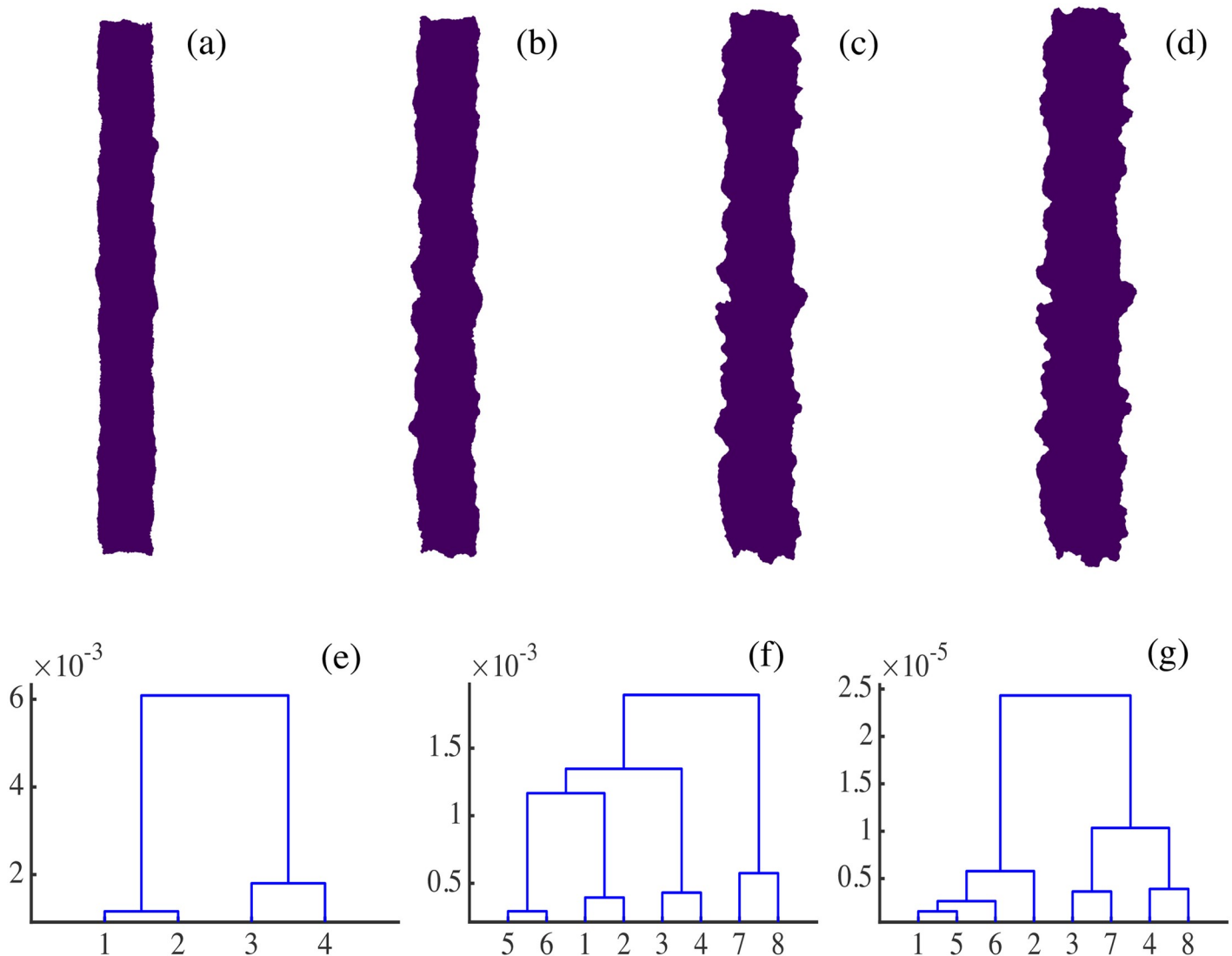
Fig 25. (a)-(d) Consecutive snapshots of the evolution of the spreading configuration in Fig 4. (e)-(g) Dendrograms for hierarchical clustering constructed using Wasserstein distances $W_{1,\infty}$ between the four snapshots. We distinguish between overall snapshots (numbered 1 to 4) in Panel (e), and half snapshots representing right (numbered 1 to 4) and left (numbered 5 to 8) moving interfaces in Panels (f) and (g). (e) The smoother overall snapshots 1 and 2 (corresponding to Panels (a) and (b)) are clustered together and, likewise, the rougher overall snapshots 3 and 4 of Panels (c) and (d). (f) The Wasserstein distance clusters together successive pairs of left and right fronts. (g) Dendrograms using the Betti number profiles $b_0(r_{max})$ (analogous to those in Fig 24) calculated for the left and right interfaces of each snapshot. Note that the interfaces of the first two snapshots are clustered together because they have similar roughness levels.

https://doi.org/10.1371/journal.pcbi.1008407.g025

clusters {1, 2, 5, 6}, {3, 7} and {4, 8}. Fronts of similar roughness are clustered together. The same clusters are obtained enforcing cutoffs on distances, that is, cutting the dendrogram at a height that defines three clusters or by K-means with three clusters. Thus, the results are robust. Cluster analysis for Fig 25(f) shows a higher variability depending on the method employed. Enforcing cutoffs on the inconsistency coefficients we obtain a natural division in four clusters {1, 2}, {3, 4}, {5, 6} and {7, 8}. If we seek for a smaller number of clusters through distance cutoffs or K-means, the results vary. Other clustering methods, such as single linkage, yield lower cophenetic correlation coefficients, which means that Ward's clustering represents these data slightly better.

Supporting information

S1 Video. Spreading test corresponding to Fig 4 in the main text.
(AVI)

S2 Video. Evolution of the velocity field in the spreading test. Color indicates velocity modulus (speed).
(AVI)

S3 Video. Video corresponding to Fig 11 in the main text.
(AVI)

S4 Video. Video corresponding to Fig 12 in the main text.
(AVI)

S5 Video. Video corresponding to Fig 13 in the main text.
(AVI)

S1 Data. Data for TDA of the interfaces extracted from Fig 14(a) and used in Fig 17(a).
(DAT)

S2 Data. Data for TDA of the interfaces extracted from Fig 14(b) and used in Fig 17(b).
(DAT)

S3 Data. Data for TDA of the interfaces corresponding to the leftmost panel of Fig 13 and used in Fig 17(c).
(DAT)

S4 Data. Data for TDA of the interfaces corresponding to the rightmost panel of Fig 13 and used in Fig 17(d).
(DAT)

S5 Data. Data for TDA of the bottleneck distance and dendrograms for clustering for frame 1 of S5 Video shown in Fig 18.
(DAT)

S6 Data. Data for TDA of the bottleneck distance and dendrograms for clustering for frame 2 of S5 Video shown in Fig 18.
(DAT)

S7 Data. Data for TDA of the bottleneck distance and dendrograms for clustering for frame 3 of S5 Video shown in Fig 18.
(DAT)

S8 Data. Data for TDA of the bottleneck distance and dendrograms for clustering for frame 4 of S5 Video shown in Fig 18.
(DAT)

S9 Data. Data for TDA of the bottleneck distance and dendrograms for clustering for frame 5 of S5 Video shown in Fig 18.
(DAT)

S10 Data. Data for TDA of the bottleneck distance and dendrograms for clustering for frame 6 of S5 Video shown in Fig 18.
(DAT)

S11 Data. Data for TDA of the bottleneck distance and dendrograms for clustering for frame 7 of S5 Video shown in Fig 18.
(DAT)

S12 Data. Data for TDA of the bottleneck distance and dendrograms for clustering for frame 8 of S5 Video shown in Fig 18.
(DAT)

S13 Data. Data for TDA of the bottleneck distance and dendrograms for clustering for frame 9 of S5 Video shown in Fig 18.
(DAT)

S14 Data. Data for TDA of the bottleneck distance and dendrograms for clustering for frame 10 of S5 Video shown in Fig 18.
(DAT)

S15 Data. Data for TDA of the bottleneck distance and dendrograms for clustering for frame 11 of S5 Video shown in Fig 18.
(DAT)

S16 Data. Data for TDA of the bottleneck distance and dendrograms for clustering for frame 12 of S5 Video shown in Fig 18.
(DAT)

S17 Data. Data for TDA of the fronts with islands corresponding to Fig 20.
(DAT)

S18 Data. Data for TDA of the fronts with islands corresponding to Fig 21.
(DAT)

S19 Data. Data for TDA of the fronts with islands corresponding to Fig 22.
(DAT)

S20 Data. Data for TDA of the compact front studies corresponding to Fig 23(a).
(DAT)

S21 Data. Data for TDA of the compact front studies corresponding to Fig 23(c).
(DAT)

S22 Data. Data for TDA of the finger spread studies corresponding to Fig 24(a).
(DAT)

S23 Data. Data for TDA of the finger spread studies corresponding to Fig 24(b).
(DAT)

Acknowledgments

Author Contributions

Conceptualization: Luis L. Bonilla, Ana Carpio, Carolina Trenado.

Formal analysis: Luis L. Bonilla, Ana Carpio, Carolina Trenado.

Funding acquisition: Luis L. Bonilla, Ana Carpio.

Investigation: Luis L. Bonilla, Ana Carpio, Carolina Trenado.

Methodology: Luis L. Bonilla, Ana Carpio, Carolina Trenado.

Project administration: Luis L. Bonilla, Ana Carpio.

Software: Ana Carpio, Carolina Trenado.

Supervision: Luis L. Bonilla, Ana Carpio.

Validation: Luis L. Bonilla, Ana Carpio, Carolina Trenado.

Visualization: Ana Carpio, Carolina Trenado.

Writing – original draft: Luis L. Bonilla.

Writing – review & editing: Luis L. Bonilla, Ana Carpio, Carolina Trenado.

References

1. Friedl P, Noble PB, Walton PA, Laird DW, Chauvin PJ, Tabah RJ, et al. Migration of coordinated cell clusters in mesenchymal and epithelial cancer explants in vitro. Cancer Research. 1995; 55:4557–4560. PMID: 7553628

2. Friedl P, Wolf K. Tumour-cell invasion and migration: diversity and escape mechanisms. Nature Cancer. 2003; 3:362–374. https://doi.org/10.1038/nrc1075 PMID: 12724734

3. Friedl P, Gilmour PD. Collective cell migration in morphogenesis, regeneration and cancer. Nature Reviews Molecular Cell Biology. 2009; 10:445–457. https://doi.org/10.1038/nrm2720 PMID: 19546857

4. Weijer CJ. Collective cell migration in development. Journal of Cell Science. 2009; 122(18):3215–3223. https://doi.org/10.1242/jcs.036517 PMID: 19726631

5. du Roure O, Saez A, Buguin A, Austin RH, Chavrier P, Silberzan P, et al. Force mapping in epithelial cell migration. Proceedings of the National Academy of Sciences. 2005; 102:2390–2395. https://doi.org/10.1073/pnas.0408482102 PMID: 15695588

6. Poujade M, Grasland-Mongrain E, Hertzog A, Jouanneau J, Chavrier P, Ladoux B, et al. Collective migration of an epithelial monolayer in response to a model wound. Proceedings of the National Academy of Sciences. 2007; 104:15988–15993. https://doi.org/10.1073/pnas.0705062104 PMID: 17905871

7. Trepat X, Wasserman MR, Angelini TE, Millet E, Weitz DA, Butler JP, et al. Physical forces during collective cell migration. Nature Physics. 2009; 5:426–430. https://doi.org/10.1038/nphys1269

8. Rørth P. Collective cell migration. Annual Review of Cell and Developmental Biology. 2009; 25:407–429. https://doi.org/10.1146/annurev.cellbio.042308.113231 PMID: 19575657

9. Cates ME, Marenduzzo D, Pagonabarraga I, Tailleur J. Arrested phase separation in reproducing bacteria creates a generic route to pattern formation. Proceedings of the National Academy of Sciences. 2010; 107:11715–11720. https://doi.org/10.1073/pnas.1001994107 PMID: 20498087

10. Petitjean L, Reffay M, Grasland-Mongrain E, Poujade M, Ladoux B, Buguin A, et al. Velocity fields in a collectively migrating epithelium. Biophysical Journal. 2010; 98:1790–1800. https://doi.org/10.1016/j.bpj.2010.01.030 PMID: 20441742

11. Angelini T, Hannezo E, Trepat X, Fredberg JJ, Weitz DA. Cell migration driven by cooperative substrate deformation patterns. Physical Review Letters. 2010; 104:168104. https://doi.org/10.1103/PhysRevLett.104.168104 PMID: 20482085

12. Angelini T, Hannezo E, Trepat X, Marquez M, Fredberg JJ, Weitz DA. Glass-like dynamics of collective cell migration. Proceedings of the National Academy of Sciences. 2011; 108:4714–4719. https://doi.org/10.1073/pnas.1010059108 PMID: 21321233

13. Trepat X, Fredberg JJ. Plithotaxis and emergent dynamics in collective cellular migration. Trends in Cell Biology. 2011; 21(11):638–646. https://doi.org/10.1016/j.tcb.2011.06.006 PMID: 21784638

14. Rørth P. Fellow travellers: emergent properties of collective cell migration. EMBO Reports. 2012; 13:984–991. https://doi.org/10.1038/embor.2012.149 PMID: 23059978

15. Serra-Picamal X, Conte V, Vicent R, Anon E, Tambe DT, Bazellieres E, et al. Mechanical waves during tissue expansion. Nature Physics. 2012; 8:628–634. https://doi.org/10.1038/nphys2355

16. Sepúlveda N, Petitjean L, Cochet O, Grasland-Mongrain E, Silberzan P, Hakim V. Collective cell motion in an epithelial sheet can be quantitatively described by a stochastic interacting particle model. PLoS Computational Biology. 2013; 9:e1002944. https://doi.org/10.1371/journal.pcbi.1002944 PMID: 23505356

17. Bernoff AJ, Topaz CM. Nonlocal aggregation models: A primer of swarm equilibria. SIAM Review. 2013; 55(4):709–747. https://doi.org/10.1137/130925669

18. Brugués A, Anon E, Conte V, Veldhuis JH, Gupta M, Colombelli J, et al. Forces driving epithelial wound healing. Nature Physics. 2014; 10:683–690. https://doi.org/10.1038/nphys3040 PMID: 27340423

19. Ravasio A, Cheddadi I, Chen T, Pereira T, Ong HT, Bertocchi C, et al. Gap geometry dictates epithelial closure efficiency. Nature Communications. 2015; 6:7683. https://doi.org/10.1038/ncomms8683 PMID: 26158873

20. Park J-A, Kim J-H, Bi D, Mitchel JA, Qazvini NT, Tantisira K, et al. Unjamming and cell shape in the asthmatic airway epithelium. Nature Materials. 2015; 14:1040–1048. https://doi.org/10.1038/nmat4357 PMID: 26237129

21. Bernoff AJ, Topaz CM. Biological aggregation driven by social and environmental factors: A nonlocal model and its degenerate Cahn-Hilliard approximation. SIAM Journal of Applied Dynamical Systems. 2016; 15(3):1528–1562. https://doi.org/10.1137/15M1031151

22. Porazinski S, de Navascués J, Yako Y, Hill W, Jones MR, Maddison R, et al. EphA2 drives the segregation of Ras-transformed epithelial cells from normal neighbors. Current Biology. 2016; 26:3220–3229. https://doi.org/10.1016/j.cub.2016.09.037 PMID: 27839970

23. Smeets B, Alert R, Pesek J, Pagonabarraga I, Ramon H, Vincent R. Emergent structures and dynamics of cell colonies by contact inhibition of locomotion. Proceedings of the National Academy of Sciences. 2016; 113(51):14621–14626. https://doi.org/10.1073/pnas.1521151113 PMID: 27930287

24. Taylor HB, Khuong A, Wu Z, Xu Q, Morley R, Gregory L, et al. Cell segregation and border sharpening by Eph receptor-ephrin-mediated heterotypic repulsion. Journal of the Royal Society Interface. 2017; 14:20170338. https://doi.org/10.1098/rsif.2017.0338 PMID: 28747399

25. Hakim V, Silberzan P. Collective cell migration: a physics perspective. Reports on Progress in Physics. 2017; 80:076601. https://doi.org/10.1088/1361-6633/aa65ef PMID: 28282028

26. Volkening A, Sandstede B. Iridophores as a source of robustness in zebrafish stripes and variability in Danio patterns. Nature Communications. 2018; 9:3231. https://doi.org/10.1038/s41467-018-05629-z PMID: 30104716

27. Moitrier S, Blanch-Mercader C, Garcia S, Sliogeryte K, Martin T. Collective stresses drive competition between monolayers of normal and Ras-transformed cells. Soft Matter. 2019; 15:537–545. https://doi.org/10.1039/C8SM01523F PMID: 30516225

28. Ganesh K, Basnet H, Kaygusuz Y, Laughney AM, He L, Sharma R, et al. L1CAM defines the regenerative origin of metastasis-initiating cells in colorectal cancer. Nature Cancer. 2020; 1:28–45. https://doi.org/10.1038/s43018-019-0006-x PMID: 32656539

29. Marchetti MC, Joanny JF, Ramaswamy S, Liverpool TB, Prost J, Rao M, et al. Hydrodynamics of soft active matter. Reviews of Modern Physics. 2013; 85:1143–1189. https://doi.org/10.1103/RevModPhys.85.1143

30. Méhes E, Vicsek T. Collective motion of cells: from experiments to models. Integrative Biology. 2014; 6:831–854. https://doi.org/10.1039/C4IB00115J PMID: 25056221

31. Bi D, Yang X, Marchetti MC, Manning L. Motility-driven glass and jamming transitions in biological tissues. Physical Review X. 2016; 6:021011. https://doi.org/10.1103/PhysRevX.6.021011 PMID: 28966874

32. Malinverno C, Corallino S, Giavazzi F, Bergert M, Li Q, Leoni M, et al. Endocytic reawakening of motility in jammed epithelia. Nature Materals. 2017; 16:587–596. https://doi.org/10.1038/nmat4848 PMID: 28135264

33. Giavazzi, Malinverno FC, Corallino S, Ginelli F, Scita G, Cerbino R. Giant fluctuations and structural effects in a flocking epithelium. Journal of Physics D: Applied Physics. 2017; 50:384003. https://doi.org/10.1088/1361-6463/aa7f8e

34. Giavazzi F, Paoluzzi M, Macchi M, Bi D, Scita G, Manning L, et al. Flocking transition in confluent tissues. Soft Matter. 2018; 14:3471–3477. https://doi.org/10.1039/c8sm00126j PMID: 29693694

35. Palamidessi A, Malinverno C, Frittoli E, Corallino S, Barbieri E, Sigismund S, et al. Unjamming overcomes kinetic and proliferation arrest in terminally differentiated cells and promotes collective motility of

carcinoma. Nature Materials. 2019; 18:1252–1263. https://doi.org/10.1038/s41563-019-0425-1 PMID: 31332337

36. Rodríguez-Franco P, Brugués A, Marín-Llauradó A, Conte V, Solanas G, Batlle E, et al. Long-lived force patterns and deformation waves at repulsive epithelial boundaries. Nature Materials. 2017; 16:1029–1037. https://doi.org/10.1038/nmat4972 PMID: 28892054

37. Ouaknin GY, Bar-Yoseph PZ. Stochastic collective movement of cells and fingering morphology: no maverick cells. Biophysical Journal. 2009; 97:1811–1821. https://doi.org/10.1016/j.bpj.2009.05.064 PMID: 19804711

38. Alert R, Blanc-Mercader C, Casademunt J. Active Fingering instability in tissue spreading. Physical Review Letters. 2019; 122:088104. https://doi.org/10.1103/PhysRevLett.122.088104 PMID: 30932560

39. Barton DL, Henkes S, Weijer CJ, Sknepnek R. Active vertex model for cell-resolution description of epi-thelial tissue mechanics. PLoS Computational Biology. 2017; 13:e1005569. https://doi.org/10.1371/journal.pcbi.1005569 PMID: 28665934

40. Alert R, Trepat X. Physical Models of Collective Cell Migration. Annual Review of Condensed Matter Physics. 2020; 11:77–101. https://doi.org/10.1146/annurev-conmatphys-031218-013516

41. Vicsek T, Czirók A, Ben-Jacob E, Cohen I, Shochet O. Novel type of phase transition in a system of self-driven particles. Physical Review Letters. 1995; 75:1226–1229. https://doi.org/10.1103/PhysRevLett.75.1226 PMID: 10060237

42. Vicsek T, Zafeiris A. Collective motion. Physics Reports. 2012; 517:71–140. https://doi.org/10.1371/journal.pone.0031711 PMID: 22359617

43. Petitjean L. Réponse active d'un épithélium à une stimulation mécanique. PhD Thesis, University Paris VI—Pierre et Marie Curie. 2011.

44. Lv J-Q, Chen P-C, Góźdź WT, Li B. Mechanical adaptions of collective cells nearby free tissue bound-aries. Journal of Biomechanics. 2020; 104:109763. https://doi.org/10.1016/j.jbiomech.2020.109763 PMID: 32224050

45. Li Y, Xiao Y, Liu C. The Horizon of Materiobiology: A Perspective on Material-Guided Cell Behaviors and Tissue Engineering. Chemical Reviews. 2017; 117:4376–4421. https://doi.org/10.1021/acs.chemrev.6b00654 PMID: 28221776

46. Boareto M, Jolly MJ, Ben-Jacob E, Onuchic JN. Jagged mediates differences in normal and tumor angiogenesis by affecting tip-stalk fate decision. Proceedings of the National Academy of Sciences. 2015; 112:E3836–E3844. https://doi.org/10.1073/pnas.1511814112 PMID: 26153421

47. Bocci F, Jolly MK, Tripathi SC, Aguilar M, Hanash SM, Levine H, et al. Numb prevents a complete epi-thelial-mesenchymal transition by modulating Notch signalling. Journal of the Royal Society Interface. 2017; 14:20170512. https://doi.org/10.1098/rsif.2017.0512 PMID: 29187638

48. Bocci F, Jolly MK, George JT, Levine H, Onuchic JN. A mechanism-based computational model to cap-ture the interconnections among epithelial-mesenchymal transition, cancer stem cells and Notch-Jag-ged signaling. Oncotarget 2018; 9(52):29906–29920. https://doi.org/10.18632/oncotarget.25692 PMID: 30042822

49. Vega R, Carretero M, Travasso RDM, Bonilla LL. Notch signaling and taxis mechanims regulate early stage angiogenesis: A mathematical and computational model. PLoS Computational Biology. 2020; 16 (1):e1006919. https://doi.org/10.1371/journal.pcbi.1006919 PMID: 31986145

50. Carlsson G. Topology and data. American Mathematical Society Bulletin. 2009; 46(2):255–308. https://doi.org/10.1090/S0273-0979-09-01249-X

51. Edelsbrunner H, Harer HJ. Computational Topology: An Introduction. Providence: American Mathe-matical Society; 2010.

52. Kerber M, Morozov D,Nigmetov A. Geometry helps to compare persistence diagrams. ACM Journal of Experimental Algorithmics. 2017; 22(1):1.4.

53. McGuirl MR, Volkening A, Sandstede B. Topological data analysis of zebrafish patterns. Proceedings of the National Academy of Sciences. 2020; 117(10):5113–5124. https://doi.org/10.1073/pnas.1917763117 PMID: 32098851

54. Soft Active Matter on Surfaces (SAMoS). https://github.com/sknepneklab/SAMoS.

55. Honda H, Eguchi G. How much does the cell boundary contract in a monolayered cell sheet? Journal of Theoretical Biology. 1980; 84(3):575–588.

56. Selmeczi D, Mosler S, Hagerdorn PH, Larsen NB, Flyvbjerg H. Cell motility as persistent random motion: Theories from experiments. Biophysical Journal. 2005; 89:912–931. https://doi.org/10.1529/biophysj.105.061150 PMID: 15951372

57. Bonilla LL, Trenado C. Contrarian compulsions produce exotic time-dependent flocking of active particles. Physical Review E. 2019; 99:012612. https://doi.org/10.1103/PhysRevE.99.012612 PMID: 30780289

58. Zomorodian A, Carlsson G. Computing persistent homology. Discrete and Computational Geometry. 2002; 33:249–274. https://doi.org/10.1007/s00454-004-1146-y

59. Topaz C, Ziegelmeier L, Halverson T. Topological data analysis of biological aggregation Models. PLoS ONE 2015; 10(5):e0126383. https://doi.org/10.1371/journal.pone.0126383 PMID: 25970184

60. Hofer C, Kwitt C, Niethammer M, Uhl A. Deep learning with topological signatures. Advances in Neural Information Processing Systems. 2017; 30 (NIPS 2017).

61. Ward JH. Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association. 1963; 58(301):236–244. https://doi.org/10.1080/01621459.1963.10500845

62. Kovacheva T. A hierarchical clustering approach to find groups of objects, Proceedings of the IV Congress of Mathematicians, Macedonia; 2008. pp 359-373. http://math.bilten.smm.com.mk/source/IV%20kongres/b5-1-opt.pdf

63. Sokal RR, Rohlf FJ. The comparison of dendrograms by objective methods, Taxon. 1962; 11:33–40.

64. MacQueen J. Some methods for classification and analysis of multivariate observations. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics. Berkeley: University of California Press; 1967. pp 281–297. https://projecteuclid.org/euclid.bsmsp/1200512992

65. Kaufman L, Rousseeuw PJ. Finding groups in data: An introduction to cluster analysis. Hoboken: Wiley-Interscience; 1990.

66. Ghrist R. Elementary Applied Topology. Create Space Independent Publishing Platform; 2014.

67. Liu J, Yin W, Li W, Chow YT. Multilevel optimal transport: a fast approximation of Wasserstein-1 distances, University of California at Los Angeles CAM Report 18-54; 2018.