

# A Rho-GTPase based model explains spontaneous collective migration of neural crest cell clusters

Brian Merchant<sup>1</sup>, Leah Edelstein-Keshet<sup>1</sup> and James J. Feng<sup>1,2\*</sup>

<sup>1</sup> Department of Mathematics, University of British Columbia, Vancouver, BC V6T 1Z2, Canada

<sup>2</sup> Department of Chemical and Biological Engineering, University of British Columbia, Vancouver, BC V6T 1Z3, Canada

## Abstract

We propose a model to explain the spontaneous collective migration of neural crest cells in the absence of an external gradient of chemoattractants. The model is based on the dynamical interaction between Rac1 and RhoA that is known to regulate the polarization, contact inhibition and co-attraction of neural crest cells. Coupling the reaction-diffusion equations for active and inactive Rac1 and RhoA on the cell membrane with a mechanical model for the overdamped motion of membrane vertices, we show that co-attraction and contact inhibition cooperate to produce persistence of polarity in a cluster of neural crest cells by suppressing the random onset of Rac1 hotspots that may mature into new protrusion fronts. This produces persistent directional migration of cell clusters in corridors. Our model confirms a prior hypothesis that co-attraction and contact inhibition are key to spontaneous collective migration, and provides an explanation of their cooperative working mechanism in terms of Rho GTPase signaling. The model shows that the spontaneous migration is more robust for larger clusters, and is most efficient in a corridor of optimal confinement.

**Keywords:** contact inhibition of locomotion, co-attraction, persistence of polarity, Rac1, RhoA, symmetry-breaking, cell mechanics

---

\*Corresponding author. E-mail: james.feng@ubc.ca

## 1 Introduction

During vertebrate embryogenesis, neural crest cells (NCCs) delaminate from the neural plate, become highly migratory through an epithelial-to-mesenchymal transition, and then travel long distances to target locations where they differentiate into a wide range of cell types. To coordinate this long range migration, NCCs must integrate information from a variety of external signals, including chemoattractants [1–4]. The process is highly complex, with apparently different mechanisms among different species and between cranial and trunk streams of NCC [5–7], and thus many unanswered questions remain. Surprisingly, a body of evidence suggests that NCCs may migrate *spontaneously* as a group in the absence of chemoattractants. *In vitro* experiments, with NCCs plated at the end of a fibronectin corridor, find that the NCCs are able to migrate down the corridor with high persistence, in the absence of any directional information from an external chemoattractant [8]. Spontaneous collective motion has also been documented for clusters of bovine capillary endothelial cells [9] and epithelial sheets of Madin-Darby canine kidney (MDCK) cells [10] in confined geometries.

*In vivo*, before the budding of endodermal pouches in zebrafish, NCCs require the presence of chemoattractant Sdf1, sensed through filopodia, to direct their collective migration. However, after the budding of endodermal pouches, NCCs are able to migrate efficiently even when their filopodia are strongly antagonized by the F-actin depolymerizing drug Latrunculin B [11]. The pouches apparently provide sufficient physical confinement as guidance. Other *in vivo* experiments have shown that groups of enteric NCCs transplanted to their target location are able to migrate in the reverse direction, suggesting the absence of an external guiding chemoattractant gradient [12]. Instead, a spontaneous symmetry-breaking appears to determine the direction of migration.

The current hypothesis for this spontaneous collective directional migration is that it emerges from two intercellular interactions between NCCs: contact inhibition of locomotion (CIL) and co-attraction (COA). CIL describes the tendency of NCCs to move away from each other upon contact, a process mediated by N-cadherins and the non-canonical Wnt signalling pathway [13,14]. COA describes how NCCs attract each other through the autocrine production of the short-ranged chemoattractant C3a and its receptor C3aR [8]. Two recent computational models [15,16] have sought to demonstrate how spontaneous migration of a group of NCCs can arise from the simultaneous action of CIL and COA. The model of Woods *et al.* [15] treats the cells as particles that move according to Newton’s second law of motion and interact according to rules that represent COA and CIL. Later, the cellular Potts model of Szabó *et al.* [16] avoids the inertia-based second-order dynamics of [15], and reproduces CIL and COA through lattice-based rules that govern the preferential direction of motion for the entire cell. While these models successfully reproduce the spontaneous persistent migration of cell clusters, they leave a fundamental question unanswered: how do COA and CIL, which have no inherent anisotropy, cause a cluster of cells to develop a collective polarity and persistent movement in a certain direction? More specifically for the essentially one-dimensional (1D) corridors tested in these models, how do rules of COA and CIL produce a symmetry-breaking between left- and rightward migration? These questions have motivated the

present study.

We started out by building a simple model that produced CIL and COA (see Fig. S1 and Movie 1 in the online Supporting Information<sup>1</sup>). For the current purpose, suffices it to say that in this model, CIL and COA alone failed to reproduce spontaneous migration that persisted in direction. In the corridor geometry of Szabó *et al.* [16], a cluster of model cells exhibited CIL and COA and remained cohesive. But its centroid merely executed a random walk to the left and the right. There was no symmetry-breaking and no persistent directional migration.

This led us to reexamine the earlier models for additional mechanisms that may have helped CIL and COA to produce persistent directionality in collective migration. In [15], one such candidate mechanism is the inertia of the particles. In reality, cell migration is non-inertial and dominated by overdamped dynamics. The cellular Potts model [16, 17] requires the polarization of a cell to be biased in favor of the most recent displacement. An additional rule explicitly enhances persistence of polarization in the presence of neighbors “to achieve realistic persistence of free cells and cells in clusters”. From the above, we hypothesize that aside from CIL and COA, these models required a third ingredient—persistence of polarity (POP) from one time step to the next—in order to reproduce spontaneous directional cell migration in clusters. This seems similar to numerous active-particle models that rely on particle-level rules of alignment to produce symmetry-breaking and pattern formation in collective motion [18, 19].

Short-term persistence of directional motion occurs naturally as, for one, the remodeling of the cytoskeleton takes tens of seconds. Nevertheless, invoking postulated rules of POP seems unsatisfactory to us as an explanation for spontaneous collective migration of NCCs over the timescale of hours. This issue partly motivated our work. Instead of a rule-based implementation of CIL and COA as found in earlier models [15, 16], we wished to base both on known biochemical pathways of key regulators such as Rac1 and RhoA [8, 20]. Furthermore, we asked whether and how POP could result from such biochemical pathways, and, if so, how it would interact with CIL and COA chemically or mechanically to produce spontaneous collective migration of NCCs.

In this paper, we present a two-dimensional (2D) model for spontaneous directional migration of clusters of NCCs based on the biochemistry and transport of molecular signals. On this more fundamental level, we first show that CIL and COA arise naturally from the underlying reaction and diffusion of Rho GTPases. Second and more importantly, we demonstrate that CIL and COA, acting on a randomization scheme for cell polarity, produce persistence of cell polarity, and consequently spontaneous collective migration of NCCs. This provides a plausible biological basis for POP from the biochemistry of Rac1 modulation by CIL and COA. As it turns out, POP is not an additional rule to be posed alongside CIL and COA. It is in fact a natural outcome of CIL and COA, as well as a conduit through which these two fundamental mechanisms give rise to the observed spontaneous collective migration.

---

<sup>1</sup>The Supplement Information is attached at the end of this PDF, and movies are available online at <https://doi.org/10.1016/j.ydbio.2018.01.013> and on the bioRxiv: <https://doi.org/10.1101/181743>

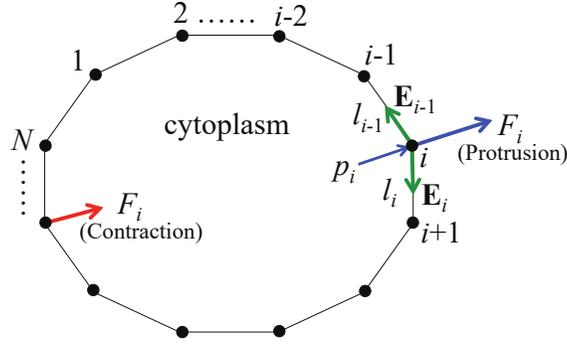


Figure 1: A model cell represented by a polygon of elastic edges. The vertices bear active and inactive levels of Rho GTPases and move according to the various forces acting on them. The edge between vertices  $i$  and  $i + 1$  has length  $l_i$  and carries an elastic tension of  $\mathbf{E}_i$ . The normal forces include pressure  $p_i$  and a protrusion or contraction force  $F_i$ .

## 2 Methods

Our model draws from a growing literature on computing cell migration from signaling proteins in the interior and on the membrane of cells [21, 22, e.g.]. Our general conceptualization of the NCC collective migration is as follows. Polarization and protrusion of individual cells are governed by Rho GTPases on the membrane [23–25], subject to turnover between the membrane-bound and cytoplasmic forms of the signaling proteins. In many cell types, Rac1 promotes F-actin assembly in lamellipodia at the protrusive front of the cell, while RhoA enhances myosin-induced cell contraction at the rear [26, 27]. For NCCs, in particular, Rac1 and RhoA have been identified as the key proteins modulating CIL and COA [8, 20]. Therefore, our model only accounts for Rac1 and RhoA and omits other GTPases such as Cdc42 and various downstream regulators. The level of active Rac1 and RhoA on the membrane determines the protrusive and contractile forces on the membrane and in turn the deformation and movement of the cell. Cell-cell interaction occurs through modulating each other’s Rac-Rho biochemistry. For example, cell-cell contact inhibits Rac1 and elevates RhoA at the site of contact in both cells. Thus, the cell protrusions retract and the cells move apart. Finally, the Rac-Rho dynamics is subject to a random noise so as to produce the tortuous trajectory commonly seen for single migrating cells.

To implement the above ideas, we represent each cell by a polygon of  $N$  vertices connected by elastic edges (Fig. 1). These edges represent the membrane-cortex complex [28], similar to the vertex models widely used for epithelial morphogenesis [29, 30]. Rac1 and RhoA levels are defined on the vertices as well as in the cytoplasm. A kind of “mesh refinement” has been tested, and balancing accuracy with computational cost, we have chosen  $N = 16$  for all simulations in the rest of the paper. For details see Fig. S2 in the Supporting Information (SI).

We imagine the cells being spread on a substrate, but do not explicitly account for focal adhesions. Each vertex on the cell membrane is subject to 2D forces acting in the plane: a pressure force from the cytoplasm enclosed by the cell membrane, cortical tension in the membrane segments, and a protrusion or contraction force determined by the Rac1 and RhoA levels on the vertex. As a

result, the vertices move, without inertia, at a speed determined by the resultant force and a friction factor. The biochemical and mechanical components of the model are intimately coupled. For clarity of narration, however, we will describe each in turn.

## 2.1 Biochemistry of Rac1 and RhoA

For both Rac1 and RhoA, the model tracks three forms of the signaling protein: the active membrane-bound form, the inactive membrane-bound form, and the inactive cytosolic form. For Rac1, we normalize the amount of these forms by the total amount of Rac1 in the cell (details in SI), and denote them as  $R^a$ ,  $R^i$  and  $R^c$ , respectively. Similarly, we denote the normalized amounts of RhoA by  $\rho^a$ ,  $\rho^i$  and  $\rho^c$ . Note that the membrane-bound forms are defined on the cell boundary vertices, and may exhibit spatial distributions. In fact, cell polarization will be marked by spatially segregated distributions of active Rac1 and RhoA. The cytosolic levels are functions of time but not space. Since the bulk diffusion of Rho GTPases inside the cytosol is much faster than on the membrane, we assume the cytosol to be well mixed [21, 31, 32]. The total amounts of Rac1 and RhoA are each conserved.

The biochemistry of Rac1 and RhoA can be represented by a set of reaction-diffusion equations. We assume that on the membrane, the active and inactive forms of each protein interconvert with activation and inactivation rates, denoted by  $K^+$  and  $K^-$  for Rac1 and  $\kappa^+$  and  $\kappa^-$  for RhoA. Only the inactive form of protein may dissociate from the membrane to diffuse within the cytosol. The membrane association and dissociation rates are denoted by  $M^+$  and  $M^-$  for Rac1, and by  $\mu^+$  and  $\mu^-$  for RhoA. Using Fick’s law to compute the 1D diffusion flux  $J$  of the active and inactive Rac on the membrane, we discretize the reaction-diffusion equations as follows (see SI for details):

$$J_i^X = -D \left( \frac{R_{i+1}^X/L_i - R_i^X/L_i}{l_i} \right) \quad (1)$$

$$\frac{dR_i^a}{dt} = K^+ R_i^i - K^- R_i^a + (J_{i-1}^a - J_i^a), \quad (2)$$

$$\frac{dR_i^i}{dt} = -K^+ R_i^i + K^- R_i^a + (J_{i-1}^i - J_i^i) + \frac{M^+ R^c}{N} - M^- R_i^i, \quad (3)$$

$$\frac{dR^c}{dt} = \sum_{i=1}^N \left( -\frac{M^+ R^c}{N} + M^- R_i^i \right), \quad (4)$$

where the subscript  $i$  indicates the  $i^{\text{th}}$  vertex on the cell membrane,  $l_i$  is the edge length between vertex  $i$  and  $i + 1$ ,  $L_i$  is the average of  $l_i$  and  $l_{i-1}$ ,  $D$  is the diffusivity on the membrane, and  $N$  is the total number of vertices.  $J_i$  approximates the diffusive flux from vertex  $i$  to vertex  $i + 1$ , its superscript  $X$  being  $a$  or  $i$  for the active and inactive forms of Rac.

RhoA obeys similar kinetic equations:

$$j_i^X = -D \left( \frac{\rho_{i+1}^X/L_i - \rho_i^X/L_i}{l_i} \right) \quad (5)$$

$$\frac{d\rho_i^a}{dt} = \kappa^+ \rho_i^i - \kappa^- \rho_i^a + (j_{i-1}^a - j_i^a), \quad (6)$$

$$\frac{d\rho_i^i}{dt} = -\kappa^+ \rho_i^i + \kappa^- \rho_i^a + (j_{i-1}^i - j_i^i) + \frac{\mu^+ \rho^c}{N} - \mu^- \rho_i^i, \quad (7)$$

$$\frac{d\rho^c}{dt} = \sum_{i=1}^N \left( -\frac{\mu^+ \rho^c}{N} + \mu^- \rho_i^i \right), \quad (8)$$

The equations imply conservation of the total amount of each Rho GTPase species.

These biochemical interactions are central to the model’s ability to reproduce cell polarization, stochastic repolarization (i.e. random changes in migration direction), CIL and COA, all of which are encoded in nonlinear expressions we use for the Rho GTPase activation and inactivation rates  $K^\pm$  and  $\kappa^\pm$  (See SI for algebraic details). We design these rate functions according to biological observations and prior modeling in the literature. In the following we briefly discuss the modeling of each of these effects.

- *Polarization.* To capture cell polarity, the activation rates of Rac1 and RhoA each reflect the species’ autocatalytic capacity, while their de-activation rates reflect mutual inhibition on the cell membrane [32]. Following [31, 32], we represent the Rac and Rho auto-activation and mutual inhibition through Hill functions, which allow spontaneous polarization of cells with Rac1 and RhoA peaking on opposite sides of a cell. This polarity is the precursor of cell motility.
- *Stochastic repolarization.* Similar to other migratory cells [33], NCCs intrinsically produce random Rac1-mediated protrusions [13] that can out-compete existing protrusive fronts and change the cell’s existing polarized morphology. To capture this, roughly every  $T_r$  minutes we randomly select a subset of vertices and up-regulate the Rac1 activation rate on them (see p. 2 of SI for more details). Aside from the reaction and diffusion, competition between protrusions is also mediated by a negative feedback of the membrane-cortex tension on Rac1 activation. This is based on observations that as a new protrusion raises the membrane-cortex tension globally over a migratory cell, the elevated tension not only resists actin protrusions mechanically, but also through biochemical signaling that inhibits Rac1 [34] or the SCAR/WAVE2 complex downstream of Rac1 [35]. Thus, hotspots of Rac1 activity compete with each other on the membrane, allowing an upstart to replace an existing protrusion on occasion.
- *Contact inhibition of locomotion.* Contact between two NCCs is known to activate the non-canonical Wnt signalling pathway, which results in the down-regulation of Rac1 and the up-regulation of RhoA [20], leading to CIL. This is effected in the kinetic equations by CIL factors that elevate the Rac1 inactivation rate  $K^-$  and the RhoA activation rate  $\kappa^+$  on any

vertex that has come into contact with a neighboring cell. As a result, the polarity of two cells approaching each other is modified, with the protrusions retracting and the cells shrinking from each other.

- *Co-attraction.* Previous work has demonstrated that binding between the NCC autocrine C3a and its receptor C3aR leads to up-regulation of Rac1 [8]. This enhances protrusion toward each other between neighbors and produces COA. Since C3a diffuses through the extracellular matrix at much faster timescales than NCC migration, we need not model the diffusion of C3a explicitly, but rather assume a steady state exponential distribution of C3a surrounding each NCC [15]. In our model, we realize COA by up-regulating the Rac1 activation rate on any vertex of a cell that is sufficiently close to a neighboring cell to “sense” the C3a distribution of the latter (see SI for details).

## 2.2 Mechanics of cell deformation and motility

The cell being represented as a polygon (Fig. 1), its shape and movement is specified by the position  $\mathbf{r}_i$  of each vertex and its speed  $\frac{d\mathbf{r}_i}{dt}$ . Thus, we write

$$\eta \frac{d\mathbf{r}_i}{dt} = \mathbf{E}_{i-1} + \mathbf{E}_i + (p_i + F_i)\mathbf{n}_i, \quad (9)$$

where  $\eta$  is a friction factor due to the viscous hinderance to cell movement,  $\mathbf{E}_i$  is the elastic tension along the edge between vertices  $i$  and  $i + 1$ ,  $p$  is the cytoplasmic pressure,  $F$  is the protrusion force due to actin filaments on the membrane, and  $\mathbf{n}$  is the unit outward normal vector at vertex  $i$  (Fig. 1). The elastic tension  $\mathbf{E}$  is proportional to the strain in each edge, relative to an undeformed length of  $l_0$ , with an elastic modulus  $\lambda$ . The cytoplasmic pressure is such that a reduction in the cell area is resisted by an outward normal force:

$$p = K_c \left[ 1 - \frac{A^c(t)}{A_0^c} \right], \quad (10)$$

where  $A^c(t)$  is the cell area at time  $t$ ,  $A_0^c$  is its resting area, and  $K_c$  is the cytoplasmic stiffness. Expansion of the cell area is constrained by the membrane elasticity  $\lambda$ . As our model is 2D, conserving the cell area is the counterpart of conserving cell volume in 3D. We may think of  $A$  as the cell area viewed from above if we take the height of the cell to be constant.

The active protrusion or contraction force is determined by the activated levels of Rho GTPase at the vertex. If the active Rac1 is greater than active RhoA, the active force is protrusive (positive). Otherwise it is contractile (negative):

$$F_i = \begin{cases} K_R \min(R_i^a - \rho_i^a, 0.05), & R_i^a > \rho_i^a \\ -K_\rho \min(\rho_i^a - R_i^a, 0.05), & R_i^a \leq \rho_i^a \end{cases}, \quad (11)$$

where  $K_R$  and  $K_\rho$  are constants governing the magnitude of the Rac1 and RhoA forces, respectively.

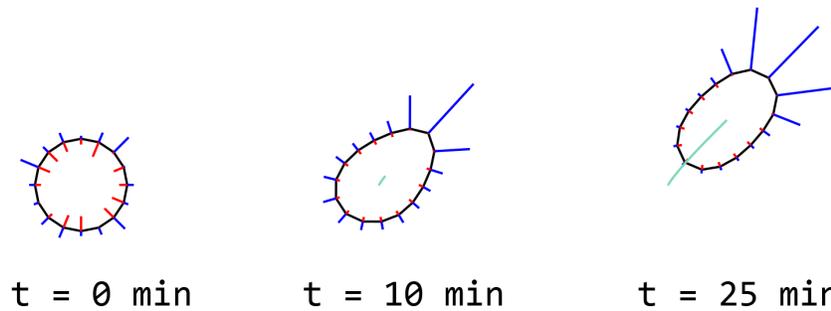


Figure 2: A single cell (diameter  $d = 40 \mu\text{m}$ ) polarizes and migrates. The lengths of line segments pointing in or out are proportional to the local levels of active GTPases (outwards blue lines for Rac1, inwards red lines for RhoA). The distance of migration is indicated by the trace of the cells centroid.

In both cases, the force is capped at a constant maximum magnitude when the difference in the GTPase levels exceeds 0.05. The rationale for this functional form is explained in SI.

The model has a list of geometric, physical and kinetic parameters. These are tabulated in Section 3 of SI, along with the sources from which we have determined their values for use in the following simulations. Moreover, we have studied the sensitivity of the model to the kinetic parameters, and demonstrated the robustness of the results over wide ranges (see Section 4 of SI).

### 3 Results

#### 3.1 Single-cell behavior: polarization and random walk

The first and most basic prediction of the model is that a single cell spontaneously polarizes and migrates (Fig. 2; also see Movie 2 in SI). As initial condition for  $t = 0$ , we divide the total amount of Rac1 into 10% membrane-bound active, 10% membrane-bound inactive and 80% cytosolic. The membrane-bound active and inactive Rac1 is each randomly distributed among the membrane nodes. A similar scheme is followed for RhoA. Unless explicitly stated otherwise, such a random initial condition is used in all subsequent simulations, including multiple realizations from different random initial conditions for gathering statistical information.

The polarization is realized via the wave-pinning mechanism described by Edelstein-Keshet and coworkers [31, 36]. In essence, the nonlinear autocatalysis of Rac1 or RhoA and their mutual inhibition (Fig. S3) imply possible coexistence of multiple solutions in various parameter regimes, including those that correspond to a low- or high-activity state. These appear at opposite ends of the cell membrane, separated by a wavefront that propagates with the continual uptake of proteins. Eventually the propagation stops with a pinned wavefront when membrane uptake ceases after the finite cytosolic pool of the proteins is sufficiently depleted. This yields a polarized state. The mutual inhibition between Rac1 and RhoA leads to complementary distributions of these two species, with one end of the cell featuring high Rac1 and low RhoA and the opposite end low Rac1 and high

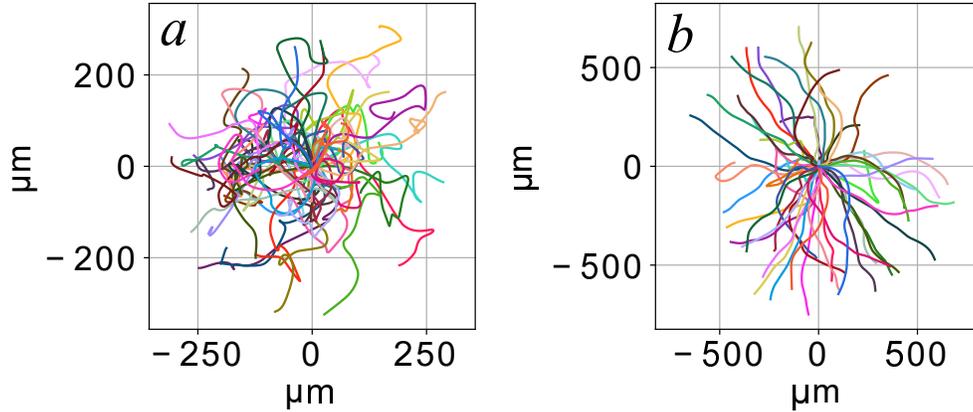


Figure 3: (a) Fifty single-cell trajectories show considerable tortuosity due to the periodic repolarization scheme in our model. Each trajectory starts from the origin with a randomly assigned initial Rac and Rho distribution and lasts 4 hours. The persistence ratio is  $R_p = 0.4 \pm 0.17$  and the persistence time  $T_p = 26 \pm 24$  min. (b) Increased persistence in cell migration when the Rac1 baseline activation rate ( $K_b^+$ ) is reduced to 40% of its usual level.  $R_p = 0.85 \pm 0.17$ ,  $T_p = 725 \pm 1140$  min.

RhoA. The former becomes a protruding front and the latter a retracting rear. Thus the cell moves forward. Note that cell-substrate adhesion is not explicitly accounted for, nor are stress fibers. The steady-state migration speed has been tuned to the experimentally observed  $3 \mu\text{m min}^{-1}$  by adjusting the coefficient  $K_\rho$  for the contractile force (Eq. 11) relative to the protrusive force and friction (Eq. 9). See Table 1 of SI for details.

The model also captures random repolarization and changes in the direction of motion. As a result, the trajectory of the single cell resembles that of a random walker (Fig. 3a). As explained in Subsection 2.1, the random change in polarization is realized in our model by periodically up-regulating the Rac activation rate  $K^+$  at randomly selected vertices. Thus, these vertices become potential hotspots of Rac1 activity. They compete with existing Rac1 peaks through cortical-tension-based inhibition [34] and the polarizing nature of Rac1-RhoA chemical dynamics [31]. If a new hotspot supersedes an existing one and becomes a new protrusion, the cell repolarizes and changes its direction of migration (see Movie 2).

We have included random repolarization in our model since such changes in migratory direction are a common feature of single-cell migration for several cell types [37,38]. As will be discussed below, it is also essential for capturing the spontaneous collective migration of a cluster of cells. *In vitro*, mammary epithelial cells alternate between two modes of migration: a highly directional phase of cell motion, and a re-orientation phase in which the cell produces new leading edges and sharply changes its direction of motion [39]. This corresponds closely with our model predictions. Theveneau *et al.* [40] reported similar bimodal behaviour of a single neural crest cell (see their Supplementary Information, Movie S5), and each phase lasts between 15 and 20 min. In our model prediction, the directional motion and repolarization each takes about 20 minutes. These periods

are determined by the kinetic rates and the parameter  $T_r$ , the time period between each new choice of random up-regulation of Rac1.

One way to quantify the random changes in migratory direction is through the persistence ratio  $R_p$ , defined as the end-to-end distance of a cell's trajectory divided by the path length of the tortuous trajectory [16, 39, 41]. Since the cell essentially executes a random walk, the end-to-end distance increases with time as  $\sqrt{t}$  while the contour length as  $t$ . Thus,  $R_p \sim 1/\sqrt{t}$  will in generally decrease with  $t$ , and is meaningful only if measured for a specific time period. Figure 3(a) depicts multiple trajectories of a single cell generated from different random initial conditions in the model. The average persistence ratio is  $R_p = 0.4$  over a period of 4 hours. For single neural crest cells, *in vitro* experimental data [13, 42, 43] suggest  $R_p$  values ranging from 0.1 to 0.4 over unspecified period of time, making comparison difficult.

A more general quantitative measure of the randomness of single-cell motility is the persistence time  $T_p$  [41], defined as the characteristic time for the decay of the directional autocorrelation [44]. Unlike the persistence ratio  $R_p$ ,  $T_p$  is independent of the observation period and thus more convenient for comparisons. The average  $T_p$  is around 26 minutes in our simulation (Fig. 3a). We did not find any experimental data of  $T_p$  for NCCs, but human mammary epithelial cells have a persistence time of  $T_p \sim 10$  min [39] while fibroblasts exhibit a range of persistence times between 15 to 120 minutes [45, 46].

Pankov *et al.*'s experiment [33] showed that inhibiting Rac1 GEFs (activators) using a drug led to increased cell persistence, because cells produced weaker Rac1 hotspots to compete with established ones. Our model has captured this feature. Figure 3(b) shows cell trajectories predicted by the model using the same randomization scheme, but with the baseline rate of Rac1 activation reduced to 40% of its normal value to model the effect of a drug that globally inhibits Rac1 activation. This drastically increases cell persistence from 0.4 to 0.85, and the persistence time from 26 to 725 min. The cell almost moves in straight lines in this case, since random spikes in Rac1 activation are too weak to override the cell's existing polarity. This feature will be relevant to our modeling of persistence of polarity as discussed in Subsection 3.4 below.

### 3.2 Contact inhibition of locomotion

Scarpa *et al.* [43] carried out *in vitro* experiments to study CIL between two cells placed on a fibronectin strip. Thus, the cells are confined to effectively 1D movement in a corridor, the area outside being non-adhesive and prohibitive for the cells. This experiment has been simulated by a phase-field model [47]. To test our models ability to recapitulate CIL, we place two cells in a corridor with boundaries designed to be repulsive to the cell via the same biochemical pathways as underlie CIL. That is, the boundary will up-regulate the RhoA activation rate as well as the Rac1 inactivation rate (Eqs. S4, S9 in SI). As an initial perturbation, we raise the Rac1 level on the parts of the cells facing each other. This gives them a polarity such that they initially move toward each other. We also turn off COA in this subsection for a cleaner manifestation of CIL. The sequence of the simulation is illustrated in Fig. 4 and Movie 3(a).

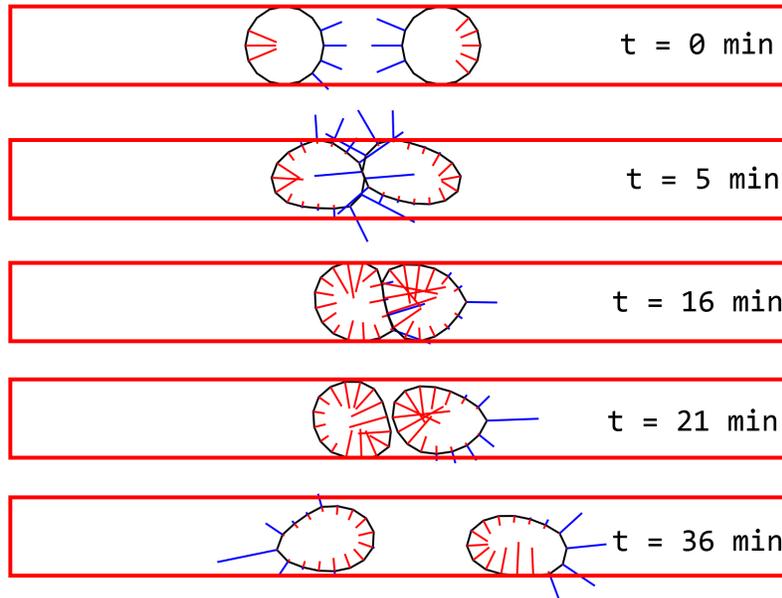


Figure 4: A typical CIL interaction in a corridor of one-cell-diameter width. Contact between the cells causes local activation of RhoA, which in turn reverses the cell polarization due to Rac1-RhoA chemical dynamics. See also Movie 3(a) in SI.

Upon contact, the vertices at the cell’s front receive a boost in their rates of RhoA activation and Rac1 inactivation. This quickly extinguishes the Rac-rich protrusion fronts (at 16 min) and subsequently turns them into Rho-rich retraction fronts (21 min). Owing to reaction-diffusion and global conservation of the GTPases (e.g., Eqs. 2–4), there is now an abundance of inactive Rac1 on the cell membrane and a dearth of active RhoA away from the contact area. Besides, the retraction relieves the membrane tension and its inhibition of active Rac1. These favorable conditions promote the appearance and growth of tentative Rac hotspots on the far end of each cell (21 min). In time, these hotspots develop into new protrusion fronts pointing away from the regions of cell-cell contact, and the two cells move apart (36 min). The sequence of contact-inhibition agrees closely with the experimental observations of Scarpa *et al.* [43] (cf. their Fig. 3c,d and Movie 1).

In the above experiment, COA is not turned off. While most of the cell-cell collisions result in CIL-induced separation, Scarpa *et al.* have also noted cells colliding and adhering to each other, and even moving past each other with little interaction. If we activate COA in our simulation, the two cells collide and separate for a short time before approaching each other again (Movie 3b). This typically results in the two staying close and moving together, resembling the adhesion in the experiment. The temporal evolution of the centroid-centroid separation agrees well with that of Scarpa *et al.* [43]; see comparison in Fig. S4 of SI.

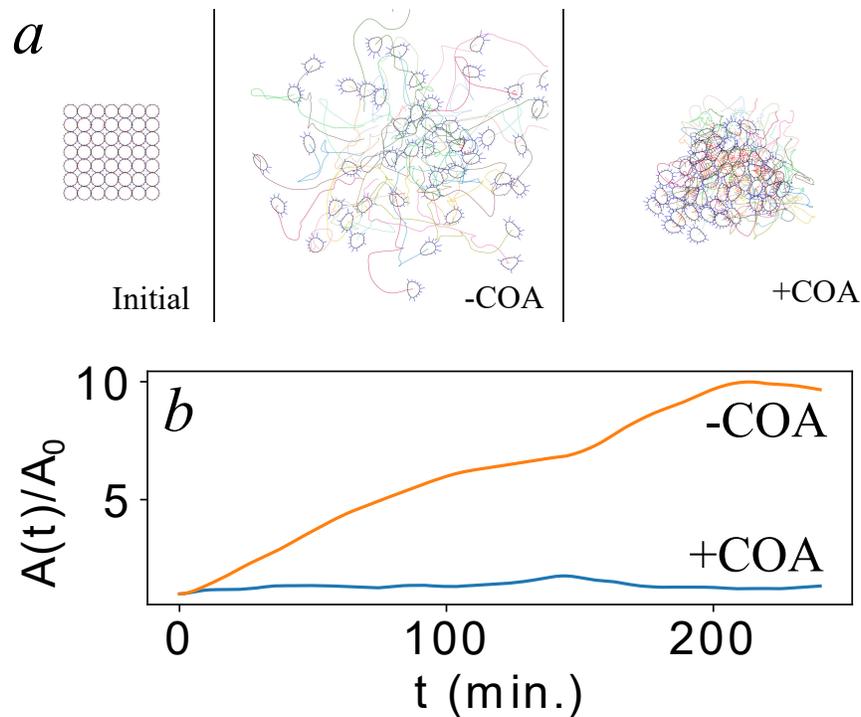


Figure 5: The effect of COA for a cluster of 49 cells. (a) Starting from a square initial array, 49 cells scatter in 4 hours with COA turned off (-COA) but stays in a cluster with COA turned on (+COA). (b) Temporal evolution of the area of the cell cluster  $A(t)$  normalized by the initial area  $A_0$ , with and without COA.

### 3.3 Co-attraction

To demonstrate the effect of COA, we arrange 49 cells in a 7 by 7 initial square (Fig. 5a), each bearing a random initial distribution of Rac1 and RhoA as described above for a single cell. The cells spontaneously polarize and move, as demonstrated in Fig. 2 for single cells. Meanwhile they interact through CIL. Here we compare two simulations, with COA turned on and off. See also Movies 4(a,b) in the SI.

With COA turned off, we see that the cells scatter as a result of their random depolarization and CIL (Fig. 5a). The total area of the cluster  $A$ , defined as the area of the convex hull of the cell centroids, increases roughly linearly in time (Fig. 5b). This resembles the scattering of Brownian particles. With COA turned on, however, the cluster retains its integrity in time (Fig. 5a). Its total area fluctuates around a mean value:  $A/A_0 = 1.5 \pm 0.18$  (Fig. 5b). This illustrates how COA, modeled on the activation of Rac1 through the C3a/C3aR ligand-receptor pathway, helps keep NCCs clustered and thereby allows for continued intercellular contact and interaction. This will be shown next to be essential for a group’s spontaneous collective migration.

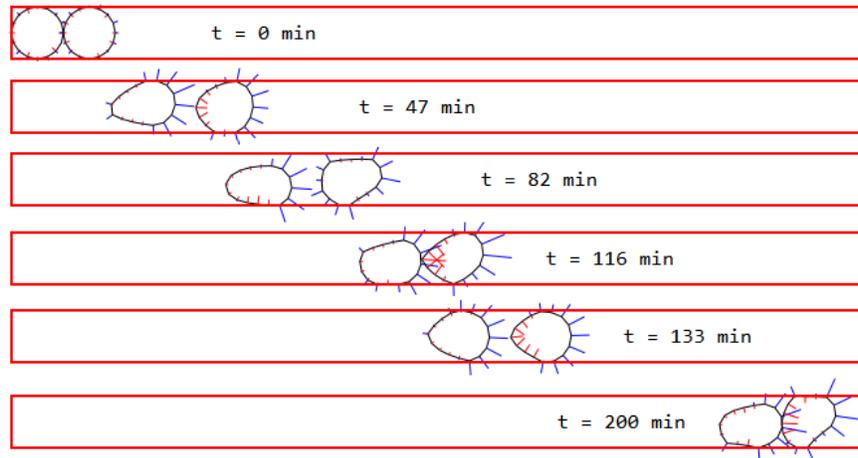


Figure 6: Persistence of polarity exhibited by two cells migrating within a 1-cell-diameter wide corridor. Highly persistent group motion develops due to the trailing cell constantly neutralizing any new protrusions on the rear end of the leading cell through CIL, and the leading cell reinforcing forward protrusions on the trailing cell through the effect of COA.

### 3.4 Persistence of polarization

As explained in the Introduction, analyzing prior models and experiments has led us to believe that a degree of persistence of polarity (POP) is necessary for spontaneous collective migration of neural crest cells. To illustrate this persistence and to explore its origin, we will first analyze the two-cell migration simulation depicted in Fig. 6 and Movie 5.

At the start of the simulation, the two cells are placed next to the left end of the corridor, whose boundaries confine the cells via CIL as explained in connection to Fig. 4. This configuration mimics that of *in vitro* experiments [8], and models the regular initial “migratory corridors” for NCCs close to the neural tube, where they first undergo epithelial-to-mesenchymal transition. The asymmetry in initial conditions reflects how at the start of their migration the mesenchymal NCCs abut against the epithelial NCCs [48]. Because of the initial asymmetry and confinement in the geometry, the two cells develop protrusion fronts toward the right, and the cells start to move to the right ( $t = 47$  min). What is somewhat surprising is that this motion is sustained for the entire duration of the simulation (10 hours) across some 40 cell diameters, the initial portion being shown in Fig. 6. This is in contrast to the behavior of a single cell, which, after the initial geometry-induced asymmetric motion away from the left end, essentially adopts a 1D random walk in the corridor with no persistent net movement in either direction. The persistent migration is reminiscent of the paracrine-mediated streaming of tumor cells [49]. If the cells are placed in the middle of the corridor, they typically break the symmetry by chance and develop a collective migration to the left or right.

Comparing the persistent two-cell motion with the random walk of the single cell, we first note that COA keeps the two cells within close proximity through the entire duration. This ensures the continual interaction between the two through CIL and COA. The effect of CIL on the leading cell

is such that it never develops a viable new protrusion in the rear. Any such hotspot, as appears at 82 min, is quickly extinguished by CIL as the budding protrusion comes into contact with the trailing cell (116 min). If such contact weakens the protrusion front of the trailing cell, this effect is short-lived as a slow-down of the trailer will end the contact. The nascent Rac1 hotspot on the rear of the trailing cell, visible at 133 min, cannot compete with the forward protrusion that is reinforced by COA. Thus, the pair continues its directional migration, and similar cycles of interaction repeat in time (e.g. 200 min). Throughout the entire migration, forward protrusions are long-lived and dominant on both cells, while rear protrusions are ephemeral and immaterial. This is clearly seen from the distribution of the life-time of protrusions in different directions (Fig. 7). The persistence ratio for the two-cell migration is  $R_p = 0.98$  over 10 hours, while for a single cell  $R_p = 0.23$  over the same period. Also the persistence time is  $T_p = 443$  min for the pair and only 20 min for the single cell. Therefore, COA and CIL act together to suppress the bursts of Rac1 up-regulation on the cell membrane, which would have produced random repolarization on a single cell. Persistence of polarity (POP) arises naturally from the cooperation between COA and CIL, and perpetuates the initial asymmetric motion of the cells endowed by the geometric confinement. This is the most important insight gained from this study.

This insight answers the main questions that had motivated our model. If CIL and COA are postulated as ad hoc rules on the *supracellular* scale, they are insufficient for spontaneous collective migration. From this we have hypothesized that POP is a necessary third ingredient. Interestingly, having built our model from the underlying GTPase dynamics on the *intracellular* scale, we find that POP emerges from the collaboration between COA and CIL, and need not be added separately after all. What has been added to our model, and missing from prior rule-based models [15, 16], is the Rac-Rho biochemistry that allows a description and rational explanation of random repolarization. By stunting new Rac1 hotspots and suppressing this random repolarization, COA and CIL gives rise to POP. In an earlier version of our model, we realized repolarization by periodically erasing the existing polarization of a cell and imposing a random new Rac1 and RhoA distribution. This did not allow Rac1 suppression by CIL and COA, and failed to yield POP (Fig. S1 and Movie 1).

The idea that Rac1 suppression promotes persistence in single-cell migration is well established in the experimental literature. Pankov *et al.* [33] showed that human fibroblasts with Rac1 activity suppressed by RNA interference or the Rac GEF inhibitor NSC 23766 had increased cell persistence. This was due to a decrease in the strength and number of random Rac1-mediated protrusions forming on the cell periphery that could otherwise become dominant and re-orient the cell's polarity. Later, Bass *et al.* [50] observed that the binding of membrane protein Syndecan-4 with fibronectin suppressed Rac1 to produce highly persistent fibroblast migration. Syndecan-4-null fibroblasts migrated randomly with delocalized Rac1 activity. Subsequently, Matthews *et al.* [51] showed that the same Syndecan-4 mediates increased persistence of migration in NCCs. As noted in connection to Fig. 3(b), reducing the Rac activation rate globally does markedly increase the persistence in the motility of a single cell in our model. Qualitatively, this corresponds to Rac suppression by drugs [33] or Syndecan-4 [50, 51]. In our simulation of a pair of interacting NCCs (Fig. 6), COA

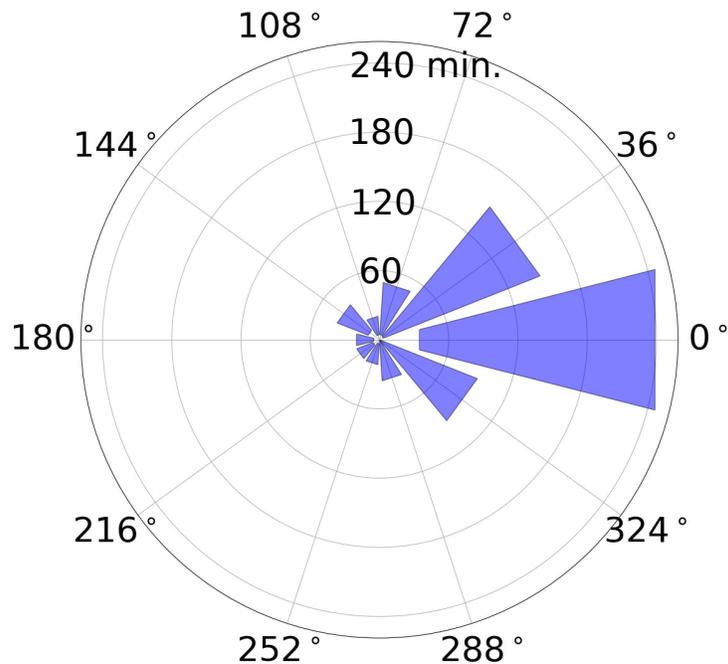


Figure 7: Distribution of the lifetime of protrusions shows a strong bias in favor of forward protrusion down the corridor. We define a protrusion as a vertex on which the active Rac1 is at a level of at least 25% of the maximum among all vertices, and exceeds the active RhoA. Dividing  $2\pi$  into 10 intervals, we collect the lifetime of protrusions oriented in each angular interval on the two cells of Fig. 6 throughout the simulation. The sector centered in each interval is a Tukey box marking the first and third quartiles of the spread.

ensures continual action of CIL, which reduces Rac1 activation. Thus, the simultaneous action of COA and CIL amounts to a pathway to POP that resembles the GEF inhibitor or Syndecan-4 in single-cell migration. Turning off either COA or CIL compromises POP in our model (Fig. S5.)

Finally, we should note that POP is stochastic in nature and is not foolproof. If, by chance, new Rac1 hotspots appear simultaneously on the rear end of both cells, they may overcome the forward polarity and cause the pair to reverse course. Then POP will be lost. Whether this occurs depends on the initial Rac1 and RhoA distributions, and on the number of cells in the cluster. We will discuss the robustness of POP at greater length next.

### 3.5 Spontaneous collective migration of larger clusters

The key features of the persistent migration of two cells carry over to the collective migration of clusters comprising more cells. Figure 8 and Movie 6 demonstrate the spontaneous collective migration of 49 cells down a corridor. The cluster migrates persistently as a whole for about 20 cell diameters during a 10-hour period. Although the cluster area fluctuates, COA keeps the coherence of the group so that the mean area does not increase over time.

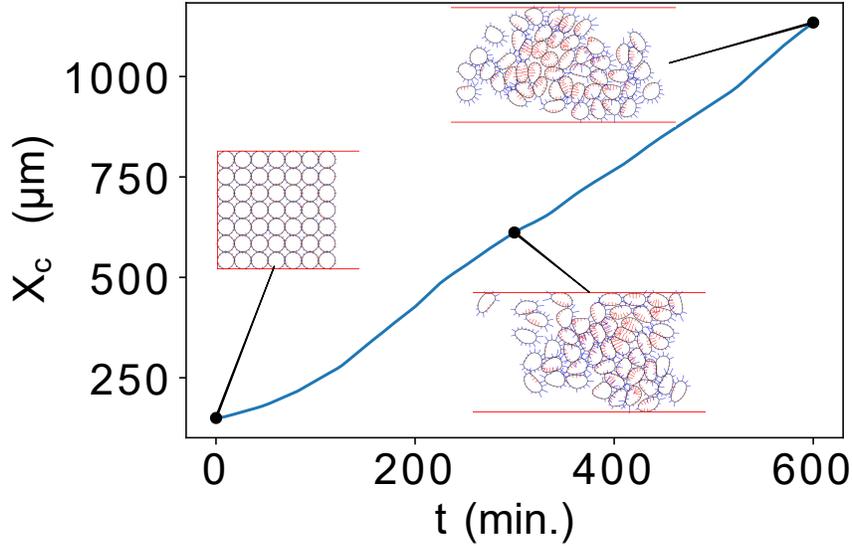


Figure 8: Spontaneous collective migration of a 49-cell cluster down a corridor of width  $w = 7$ .  $X_c$  denotes the cluster centroid location and the snapshots show that the cluster retains its integrity while migrating collectively. See Movie 6 in SI for a more detailed view of the migration.

To quantify the cluster size effect on the spontaneous collective migration, we have to distinguish it from the confinement effect of the corridor. With a larger number of cells inside the same corridor, the cluster will be more crowded and this amounts to an effectively stronger confinement. As an approximate way of separating the two effects, we adopt the following scheme in the next two subsections. In this subsection, we keep the width of the corridor  $w$ , scaled by the cell diameter  $d$ , as  $w = \sqrt{n}$ ,  $n$  being the total number of cells. Thus, we always have a square initial array of cells enclosed on three sides by walls, as in Fig. 8,  $t = 0$ . This gives us, in a sense, “equal confinement” among the different  $n$ . In the next subsection, we will vary  $w$  for a fixed  $n$  to examine the effect of confinement.

Another complication is that in our model for COA, the overall level of C3a is the sum of all such signals emitted from each cell. Thus, a larger number of cells raise the level of C3a around them. As a result, we do not need as strong a COA intensity per cell for larger  $n$ . By modulating the maximum COA strength  $M_{\text{COA}}$  (See Eq. S3 of SI for definition), we ensure that with increasing  $n$ , the average separation  $S$  between neighboring cells remains roughly the same. This is illustrated in Fig. 9(a).  $S$  is calculated by taking a Delaunay triangulation of the cell centroids and averaging all its edges, and then time-averaging this separation over the duration of the simulation. For each  $n$ , we have run 20 independent realizations of the simulation starting from random initial Rac1 and RhoA distributions. The error bars indicate the spread among the 20 runs. Going from  $n = 4$  to 49, we have reduced  $M_{\text{COA}}$  from 24 to 8. These values are tabulated in SI.

Figure 9(b–e) shows that the intensity of persistent collective migration increases with increasing  $n$ , rapidly for small clusters ( $n = 4$  to 9) and mildly for the larger ones. Take the persistence time  $T_p$  in panel (c) for example. For  $n = 4$ ,  $T_p$  averages among all runs to about 120 min. For  $n = 9$ ,

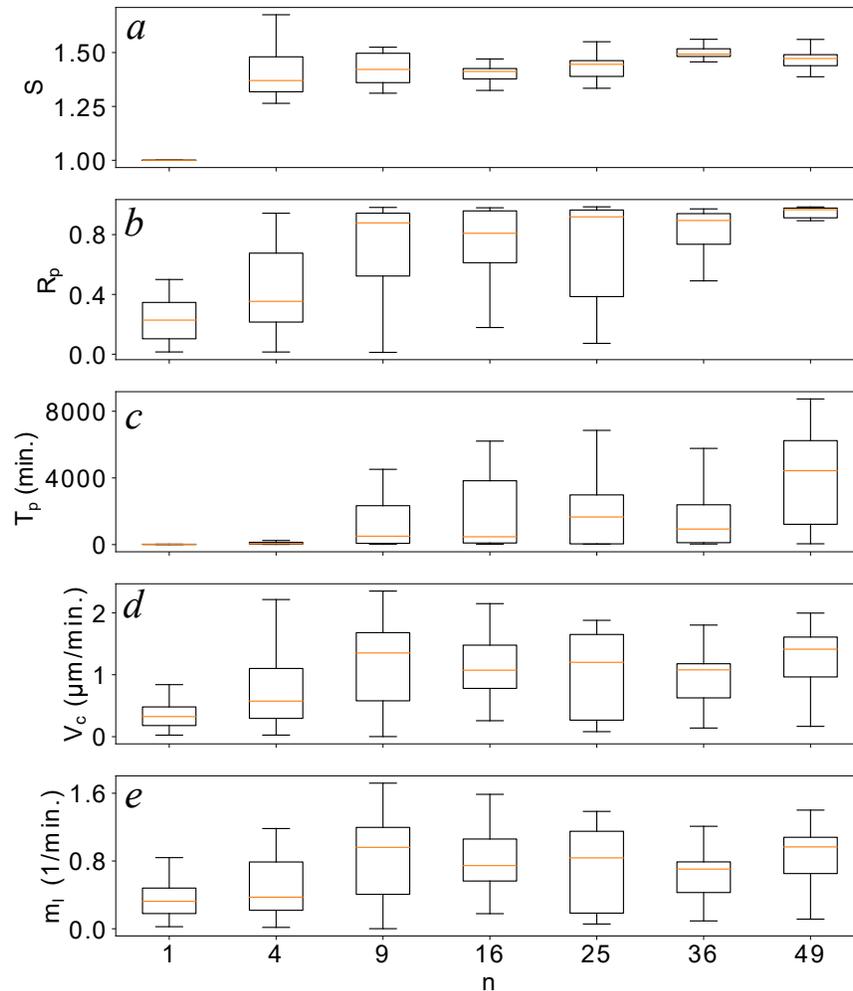


Figure 9: Effect of the cluster size  $n$  on key features of spontaneous collective migration. For each  $n$  value, the Tukey box-whisker represents 20 independent runs from different random initial GTPase distributions. The top and bottom of the box mark the first and third quartiles of the data, with the median indicated by the bar inside. The whiskers mark data within 1.5 times the interquartile range from the edges of the box. (a) As  $n$  increases from 4 to 49, we have decreased the strength of COA to maintain roughly the same nearest neighbour separation  $S$  throughout the simulation. (b) The group persistence ratio  $R_p$  over 10 hours, computed from the trajectory of the centroid of the cluster. (c) The persistence time  $T_p$  of the centroid of the cluster. (d) The average migration speed  $V_c$  of the centroid of the cluster. (e) The migration intensity  $m_I = V_c/S$ .

this has increased to 1627 min. Further increasing  $n$  lengthens  $T_p$  relatively modestly. The same trend is seen for the persistence ratio  $R_p$  (panel b) and cluster migration speed  $V_c$  (panel d). The bottom panel (e) depicts the “collective migration intensity”  $m_I = V_c/S$ , which also increases with  $n$ . This ratio gives us a characteristic frequency at which the cells would pass a fixed location in the corridor. It is similar to the “transport ratio” defined for the percentage of cells passing a certain position in the corridor [16]. But the migration intensity does not depend on choosing an arbitrary “observation post”, and thus provides a more universal measure of the efficiency of

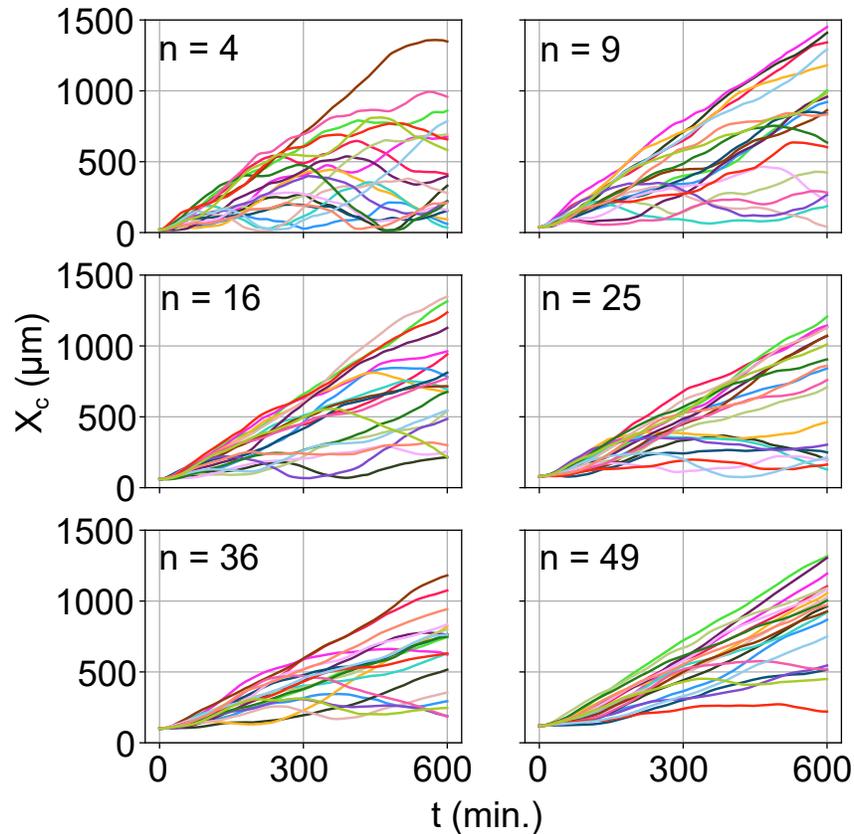


Figure 10: The trajectories of the cluster centroid  $X_c(t)$  for various group sizes. The cell diameter  $d = 40 \mu\text{m}$ . Each plot includes 20 runs starting from different random initial Rac1 and RhoA distributions under otherwise identical conditions. Some of the trajectories suffer from reversals, but the collective migration becomes more robust with increasing  $n$ .

collective migration.

The explanation for Fig. 9 lies in the fact that for small clusters, POP has a greater probability of failure, when the cluster reverses course as a whole. For larger clusters, the chance of POP failure is much reduced. Figure 10 illustrates the stochastic nature of POP fallibility. Take  $n = 9$  for example. Among the 20 trajectories, the 10 that have traveled the farthest have never suffered a reversal, i.e.  $V_c > 0$  for the entire duration. For the rest, reversal may first set in as early as  $t = 60$  min or as late as 520 min. The reversal is invariably accompanied by a large portion of the cells simultaneously developing rear-facing Rac1 hotspots that survive and mature into protrusions. Fig. S6 and Movie 7 in the SI illustrate this process in detail for a 4-cell cluster.

With increasing cluster size  $n$ , the probability of POP failure decreases, and the reversal tends to occur at later times if at all. As a convenient way to quantify the POP failure rate, we count the percentage of trajectories that end with a downward slope at the end of the 10-hour simulation. For  $n = 4$ , 50% of the 20 trajectories fail by this criterion. That percentage declines to 35% for  $n = 9$ , 25% for  $n = 16$ , 15% for  $n = 25$ , 20% for  $n = 36$ , and 10% for  $n = 49$ . Even though 20 realizations may not be a large enough sample size, the available results show a clear trend of more

robust POP with increasing  $n$ . The failure of POP requires a large portion of the group to develop rear protrusions *simultaneously*. If only one or a few cells acquire rear hotspots among many, they will be quick extinguished by their neighbors through CIL. This explains why the probability of POP failure declines with increasing  $n$ . After all, as POP arises from cell-cell interactions, it should be more robust in larger clusters that provide more interactions.

Finally, we can make some quantitative comparisons between our model prediction and experimental data. For NCCs migrating *in vivo* in *Xenopus laevis* embryos, Szabó *et al.* [16] reported a collective migration speed of  $V_c \approx 1.5 \mu\text{m}/\text{min}$ . *In vitro*, the NCCs migrate at  $V_c \approx 1 \mu\text{m}/\text{min}$  down a corridor coated with fibronectin and confined by borders that are rich in versican, an extracellular protein that repels NCCs. With the model parameters used (tabulated in SI), our model predicts an average speed  $V_c = 1.12 \mu\text{m}/\text{min}$  for the larger clusters ( $n \geq 9$ ) (Fig. 9d), comparable with the experimental data. Note that in our model, we have adjusted the contractile force factor  $K_\rho$  (Eq. 11) to produce the experimental speed of  $3 \mu\text{m}/\text{min}$  for a single cell in the rectilinear phase of motility. The agreement in cluster migration speed  $V_c$  is not the result of parameter fitting. We can also compare the persistence of the migrating clusters with experimental measurements. Szabó *et al.* [16] documented a persistence ratio  $R_p \approx 0.85$  over a time period of 4 hours *in vivo*, and  $R_p \approx 0.87$  *in vitro*. In Fig. 9(b), we have predicted an average persistence of  $R_p = 0.77$  for the larger clusters ( $n \geq 9$ ) over 10 hours. A caveat about comparison with experiment is that the model has a large number of parameters, not all of which can be ascertained directly from experimental measurements. The SI summarizes these parameter values and the model’s sensitivity to them.

### 3.6 Confinement effect

In the corridor-based studies of Mayor *et al.* [8, 15, 16, 43] as well as in our simulations, confinement of the boundaries plays two important roles: to limit the migration to effectively one dimension, and to provide the initial geometric asymmetry that ensures persistent migration later. Because of similar confinement effects *in vivo*, e.g. by versican-rich boundaries delimiting the migration of *Xenopus* cephalic NCCs, Szabó *et al.* [16] have investigated the role of confinement in NCC collective migration. Now we can examine confinement effects in our model as well.

For a fixed cluster size,  $n = 16$ , we have tested the efficiency of spontaneous migration in corridors of varying width (Fig. 11). At the start, the cells are arranged in a roughly rectangular shape that is bounded by walls on three sides at the left end of the corridor (Fig. S7). In Fig. 11(a),  $w = 1$  appears to be a special case with much larger average neighbor separation than in the wider corridors. An examination of the cell trajectories shows that the 16 cells typically break up into two or more smaller groups that tend to drift apart (Fig. S8). This highly confined situation only allows longitudinal cell-cell interactions, which proves insufficient to maintain the single-file cluster as a whole. For this reason  $w = 1$  cannot be compared with the wider corridors in Fig. 11.

If we disregard the data for  $w = 1$ , then the most important feature of Fig. 11 is the existence of an optimal confinement, at  $w = \sqrt{n} = 4$  in this case, that produces the most efficient collective migration. This can be appreciated from each of the panels of the plot. For example, the average

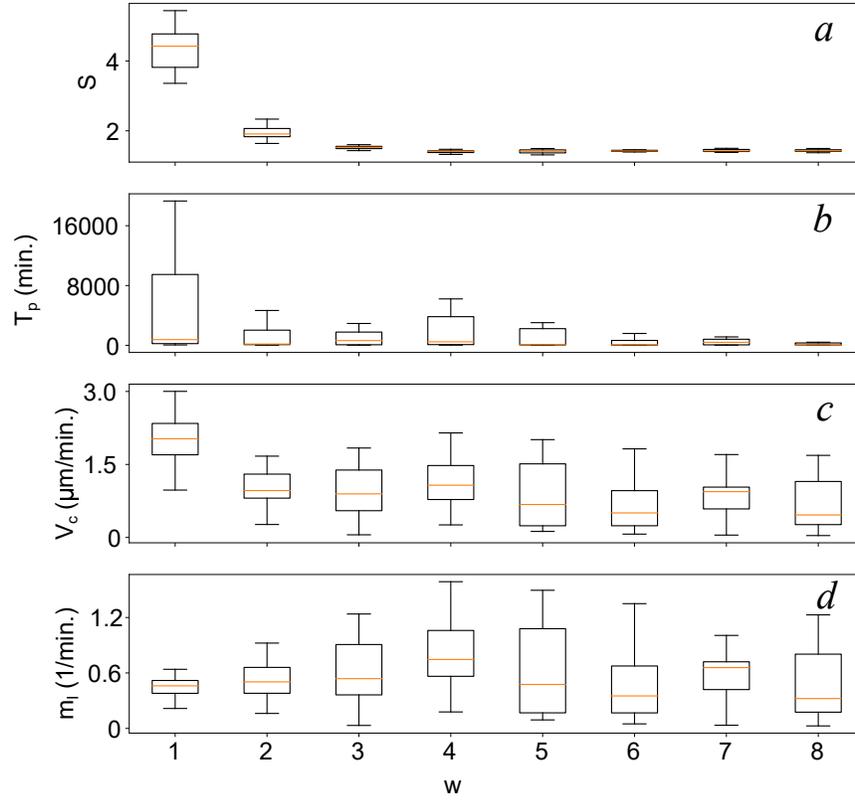


Figure 11: Confinement effect on the spontaneous collective migration of a cluster of 16 cells in a corridor of width  $w$ . In each panel the Tukey boxes represent 20 runs as in Fig. 9. (a) The average neighbour separation  $S$ . (b) The persistence time  $T_p$  of the centroid of the cluster. (c) The average migration speed  $V_c$  of the centroid of the cluster. (d) The migration intensity  $m_I = V_c/S$ .

cell separation  $S$  is smallest for  $w = 4$ . In narrower corridors with  $w < \sqrt{n}$ , the strong confinement forces the cluster into an elongated array, and hampers the coordination between its front and rear via COA. Figure 12 illustrates this for  $w = 2$ . For  $w > \sqrt{n}$ , on the other hand, there are empty spaces between the top and bottom walls and the cluster (Fig. 12,  $w = 8$ ). The weak confinement allows the cluster to spread along the  $y$ -direction, thus also enlarging  $S$ . Similarly, the cluster persistence time  $T_p$ , cluster migration speed  $V_c$  and migration intensity  $m_I$  all attain maximum values at the optimal width. Generally, over-confinement in narrow corridors hampers COA and weakens coordination throughout the elongated cluster. Under-confinement in wide corridors allows the cluster to meander in a 2D plane instead of migrating down a 1D corridor. The width  $w = \sqrt{n}$  produces a roundish cluster shape that fits snugly between the walls. Hence the attainment of optimal cluster migration efficiency. The phenomenon of optimal confinement has also been confirmed for larger clusters of  $n = 25$  and  $36$  (Fig. S9).

The concept of optimal corridor confinement was first proposed by Szabó *et al.* [16] based on a cellular Potts model. Besides, they collected *in vivo* data on NCC streams of different widths migrating in zebrafish and *Xenopus* embryos, and demonstrated that their cluster size and stream width conform to the idea of an optimal confinement. Our model confirms the existence of optimal

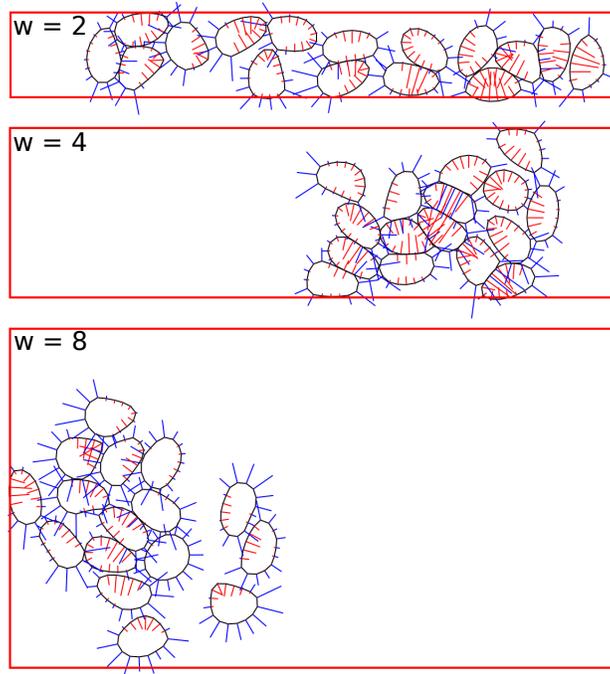


Figure 12: Snapshots of the migration of 16 cells in corridors of increasing width,  $w = 2$ , 4 and 8 at  $t = 225$  min. Note the effect of over-confinement and under-confinement on the cluster configuration in the top and bottom panel, respectively. The mid-panel represents optimal confinement.

confinement, and the optimal width identified here,  $w = \sqrt{n}$ , is consistent with both the *in vivo* and *in silico* data of Szabó *et al.* [16]. Furthermore, we have offered an explanation for the optimal confinement in terms of COA and CIL. Confinement tends to hinder COA’s ability to maintain coherence of the cluster, but it accentuates the role of CIL in keeping the cluster on a straight 1D path. An optimal confinement offers a balance between COA and CIL, both necessary ingredients for spontaneous collective migration. Although optimal confinement at  $w = \sqrt{n}$  prevails for all cluster sizes tested, the effect is generally weak (cf. Fig. 11), and may be easily overwhelmed by other mechanisms such as chemotaxis. It will be interesting to see future studies of this phenomenon.

We have also investigated the effect of initial cluster configuration by deviating from that of Fig. S7. For a fixed corridor width, say  $w = 4$ , we test initial configurations with the cells arranged into rectangular arrays of different width smaller than  $w$  positioned midway between the top and bottom walls, as well as initial configurations with the cells placed at random, non-overlapping positions (Fig. S10). The general trend is that the intensity of migration decreases for more elongated initial configurations, and that the random cluster has poorer migration intensity than the regular array (Fig. S11). As explained in the caption to Fig. S11, these observations are consistent with the understanding of confinement gained from Fig. 11.

## 4 Discussion

Our goal is to provide an explanation of spontaneous collective migration of neural crest cells in terms of key GTPases that regulate cell polarization and protrusion. From a broad perspective, we can summarize this work by the following three points.

First, we have adopted a modeling approach based on the biochemical signalling pathways as an alternative to the existing paradigms of continuum and agent-based modeling. Continuum models seek a “mean-field” description of the average orientation and movement of many cells without resolving the scales of individual cells [52–54]. However, it is challenging to represent intercellular interaction within the continuum formulation. Agent-based models address this concern by assigning rules on individual agents that recapitulate known cell-level behaviour [9, 15, 16, 55]. While helpful when little is known about the underlying biology, these models tend simply to give back the behaviour designed for. The biochemically-based model uses a finer level of resolution. Moreover, it seeks to connect cell-level behavior to intracellular signaling pathways that have been established experimentally. Thus, it can be more general and even simpler than models based on postulated rules. For example, our model predicts cell behaviours ranging from polarization to spontaneous collective migration from a few well-established principles governing GTPases [24] and cell mechanics [56]. We should note that in recent years, biochemically-based modeling of cell mechanics has gained increasing interest and currency [57–59]. This is evidently motivated and enabled by growing experimental knowledge of the role of chemical signaling in complex behaviours such as collective migration [60–62].

Second, we have identified the persistence of polarity, or POP, as an essential factor in enabling spontaneous collective migration of cell clusters. Prior to this work, the extensive experimental studies of Mayor and coworkers had identified CIL and COA as key to spontaneous collective migration. Examining their recent models [15, 16], however, we have come to suspect that POP, an incidental feature of these models, may in fact be essential for the appearance of spontaneous collective migration. To test this hypothesis, we built a simpler model that recapitulates CIL and COA but with POP intentionally suppressed by erasing the cell polarity periodically and replacing it by a random “initial distribution” of the Rho GTPases. As illustrated in Fig. S1 and Movie 1, this model does not exhibit spontaneous collective migration in a corridor. Instead, the centroid of the cell cluster executes a 1D random walk once the cluster has moved away from the end of the corridor, where it enjoyed an initial directional migration thanks to the asymmetric initial condition.

Third, we have demonstrated that POP emerges naturally from the simultaneous action of COA and CIL. Thus, we have not only validated the existing proposal that COA and CIL cause spontaneous collective migration [8, 13], but also placed it on the concrete biological basis of Rac-Rho signaling. COA maintains the integrity of a cell cluster and ensures continual proximity and CIL interaction among neighbors. The two cooperate to suppress Rac1 and perpetuate an initial polarity and direction of migration that may have arisen from asymmetric boundary conditions and confinement. This is consistent with experimental observations that Rac1 suppression by drugs or

Syndecan-4 enhances persistence in the polarity and motility in fibroblast and NCCs [33, 50, 51]. Thus, the model offers an explanation for the origin of POP; it is an emergent behavior based on COA and CIL rather than a separate mechanism to be postulated in addition to these two. This hypothesis can be tested by experiments that use drugs to partially suppress various proteins (or their activation levels) in the Rho GTPase signaling pathway [63, 64]. The resulting impairment of collective migration can be compared with model predictions of how COA, CIL and POP are affected by modulating the suitable Rac1 and RhoA rate constants.

It is interesting to put the current model in the context of directional motion and symmetry-breaking in vastly different systems, e.g. the active swimming of bacteria and flocking of insects or birds [19]. While alignment and directionality in macroscopic flocks rely on visual cues or pressure waves, we have shown that in the NCC context, it emerges from COA and CIL. Generally, one could view our POP as a device for alignment similar to the previously proposed directional persistence of random walk [9], persistence due to particle inertia [15], “persistence decay” for cell polarity [16] and collision-based alignment [65].

## 5 Conclusion

This paper presents a biochemistry-based model for the curious phenomenon of spontaneous collective migration of neural crest cells (NCCs). This represents an alternative to the paradigm of rule- or agent-based modeling. Starting from several well-established signaling pathways—autocatalytic activation and mutual inhibition of Rac1 and RhoA, Rac1 inhibition by contact and by membrane tension, and co-attraction from C3a/C3aR binding—as well as a standard vertex-based model for cell mechanics, we have predicted the following results:

- A single NCC polarizes with a protruding front featuring elevated Rac1 and a contracting rear featuring elevated RhoA. This, together with a randomized repolarization scheme, reproduces the tortuous trajectory of single-cell migration observed experimentally.
- In cell-cell encounters, the model reproduces the well-known contact inhibition of locomotion (CIL). The contact elevates the inactivation rate of Rac1 and the activation rate of RhoA, thereby neutralizing the protrusion front and turning it into a retracting rear.
- The model recapitulates the effect of co-attraction (COA) in preventing a cluster of cells from scattering due to CIL. COA thus provides an aggregative effect that balances the dispersal effect of CIL to keep the integrity of cell clusters.
- In a confining corridor, the simultaneous action of COA and CIL produces a persistence of cell polarity (POP), such that cell clusters will continue to migrate over great lengths in the direction favored by the initial configuration. The collective migration is more robust for larger clusters. For the model parameters used, the predicted cluster migration speed and persistence agree with experiments.

- There exists an optimal confinement that produces maximum persistence and speed of collective migration. For a cluster of  $n$  cells, the optimal corridor width is  $\sqrt{n}$  times the single cell diameter. This is consistent with prior *in vivo* and *in silico* data.

These results confirm the current hypothesis that the spontaneous collective migration of NCCs is due to CIL and COA, and provide a more complete picture by elucidating the biological basis for POP.

We must emphasize the limitations of this work as well. Collective migration of NCCs is a highly complex process, with multiple potential mechanisms at work for different species and for cranial and trunk NCCs [5–7]. This study is restricted to the relatively simple *in vitro* scenario of NCCs spontaneously migrating down a corridor in the absence of chemoattractants. Our model does not preassign “leader” or “trailblazer” cells [2], and as such is likely to be relevant to cranial NCCs that appear to possess a homogeneous migratory capability [6]. Keeping in mind the complexity of *in vivo* NCC migration, one may adapt the present theoretical framework to explore certain aspects of that phenomenon. A first step in this direction will be to add a chemokine such as Sdf1 and study how it modifies the CIL-COA-POP mechanism during chemotaxis [3, 40].

On a more technical level, the model incorporates numerous simplifications. For example, the geometric setup of the corridor is modeled after the *in vitro* experiments [8, 16], and is highly simplified in comparison to the migration paths *in vivo*. Besides, cell-cell adhesion is not explicitly accounted for. Arguably, the role of N-cadherins has been subsumed into the CIL pathway [40]. But cell-cell adhesion may also directly supplement COA in maintaining the integrity of the cluster. Similarly, cell-substrate or cell-ECM adhesion is neglected, as is the actomyosin responsible for force generation. In reality, protrusion and migration are enabled by the actomyosin contractile apparatus that is anchored onto the substrate or ECM. The model treats this in a phenomenological way via the protrusion/contraction force parameters (Eq. 11) and the frictional factor (Eq. 9). Finally, COA is modeled by assigning a steady-state C3a distribution to each cell, with *ad hoc* neighbor screening. Moreover, C3a signals from neighboring cells are summed in a linearly additive manner. It would be more realistic to model the C3a field using reaction-diffusion equations with C3a-C3R binding on cell membranes, and to model the response to C3a-C3aR binding by a nonlinear relationship with saturation. For lack of experimental data, the model has adopted the simplest reasonable form. The validity of these simplifications remains to be verified by carefully designed experiments that isolate the factors in question.

**Acknowledgement:** The authors acknowledge financial support by the Natural Sciences and Engineering Research Council of Canada. We also thank Philip Maini and Roberto Mayor for discussions and comments on various aspects of the project.

## References

- [1] R. McLennan, L. Dyson, K. W. Prather, J. A. Morrison, R. E. Baker, P. K. Maini, P. M. Kulesa, Multiscale mechanisms of cell migration during development: theory and experiment, *Development* 139 (2012) 2935–2944.
- [2] R. McLennan, L. J. Schumacher, J. A. Morrison, J. M. Teddy, D. A. Ridenour, A. C. Box, C. L. Semerad, H. Li, W. McDowell, D. Kay, P. K. Maini, R. E. Baker, P. M. Kulesa, VEGF signals induce trailblazer cell identity that drives neural crest migration, *Developmental biology* 407 (2015) 12–25.
- [3] A. Shellard, R. Mayor, Chemotaxis during neural crest migration, *Semin. Cell Dev. Biol.* 55 (2016) 111–118.
- [4] A. Szabó, R. Mayor, Modelling collective cell migration of neural crest, *Curr. Opin. Cell Biol.* 42 (2016) 22–28.
- [5] P. M. Kulesa, R. McLennan, Neural crest migration: trailblazing ahead, *F1000Prime Rep.* 7 (2015) 02.
- [6] J. Richardson, A. Gauert, L. B. Montecinos, L. Fanlo, Z. M. Alhashem, R. Assar, E. Marti, A. Kabla, S. Hartel, C. Linker, Leader cells define directionality of trunk, but not cranial, neural crest cell migration, *Cell Rep.* 15 (9) (2016) 2076 – 2088.
- [7] M. A. Genuth, C. D. Allen, T. Mikawa, O. Weiner, Chick cranial neural crest cells migrate by progressively refining the polarity of their protrusions, *Dev. Biol.* 444 (2018) S252–S261.
- [8] C. Carmona-Fontaine, E. Theveneau, A. Tzekou, M. Tada, M. Woods, K. M. Page, M. Parsons, J. D. Lambris, R. Mayor, Complement fragment c3a controls mutual cell attraction during collective cell migration, *Dev. Cell* 21 (6) (2011) 1026–1037.
- [9] S. Huang, C. Brangwynne, K. Parker, D. Ingber, Symmetry-breaking in mammalian cell cohort migration during tissue pattern formation: Role of random-walk persistence, *Cytoskeleton* 61 (4) (2005) 201–213.
- [10] S. R. K. Vedula, M. C. Leong, T. L. Lai, P. Hersen, A. J. Kabla, C. T. Lim, B. Ladoux, Emerging modes of collective cell migration induced by geometrical constraints, *Proc. Natl. Acad. Sci. U.S.A.* 109 (32) (2012) 12974–12979.
- [11] E. F. Boer, E. D. Howell, T. F. Schilling, C. A. Jette, R. A. Stewart, Fascin1-dependent filopodia are required for directional migration of a subset of neural crest cells, *PLoS Genet.* 11 (1) (2015) e1004946.
- [12] A. J. Burns, J.-M. M. Delalande, N. M. Le Douarin, In ovo transplantation of enteric nervous system precursors from vagal to sacral neural crest results in extensive hindgut colonisation, *Development* 129 (12) (2002) 2785–2796.

- [13] C. Carmona-Fontaine, H. K. Matthews, S. Kuriyama, M. Moreno, G. A. Dunn, M. Parsons, C. D. Stern, R. Mayor, Contact inhibition of locomotion in vivo controls neural crest directional migration, *Nature* 456 (7224) (2008) 957–961.
- [14] R. Mayor, E. Theveneau, The role of the non-canonical wnt–planar cell polarity pathway in neural crest migration, *Biochem. J* 457 (1) (2014) 19–26.
- [15] M. L. Woods, C. Carmona-Fontaine, C. P. Barnes, I. D. Couzin, R. Mayor, K. M. Page, Directional collective cell migration emerges as a property of cell interactions, *PloS One* 9 (9) (2014) e104969.
- [16] A. Szabó, M. Melchionda, G. Nastasi, M. L. Woods, S. Campo, R. Perris, R. Mayor, In vivo confinement promotes collective migration of neural crest cells, *J. Cell Biol.* 213 (5) (2016) 543–555.
- [17] A. Szabó, R. Ünnepp, E. Méhes, W. O. Twal, W. S. Argraves, Y. Cao, A. Czirók, Collective cell motion in endothelial monolayers, *Phys. Biol.* 7 (2010) 046007.
- [18] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, O. Shochet, Novel type of phase transition in a system of self-driven particles, *Phys. Rev. Lett.* 75 (1995) 1226–1229.
- [19] M. C. Marchetti, J. F. Joanny, S. Ramaswamy, T. B. Liverpool, J. Prost, M. Rao, R. A. Simha, Hydrodynamics of soft active matter, *Rev. Mod. Phys.* 85 (2013) 1143–1189.
- [20] A. Roycroft, R. Mayor, Molecular basis of contact inhibition of locomotion, *Cell. Mol. Life Sci.* 73 (6) (2016) 1119–1130.
- [21] B. Vanderlei, J. J. Feng, L. Edelstein-Keshet, A computational model of cell polarization and motility coupling mechanics and biochemistry, *Multiscale Model. Simul.* 9 (4) (2011) 1420–1443.
- [22] M. P. Neilson, J. A. Mackenzie, S. D. Webb, R. H. Insall, Modelling cell movement and chemotaxis using pseudopod-based feedback, *SIAM J. Sci. Comput.* 33 (3) (2011) 1035–1057.
- [23] M. Raftopoulou, A. Hall, Cell migration: Rho GTPases lead the way, *Dev. Biol.* 265 (1) (2004) 23–32.
- [24] A. B. Jaffe, A. Hall, Rho GTPases: biochemistry and biology, *Annu. Rev. Cell Dev. Biol.* 21 (2005) 247–269.
- [25] M. M. Zegers, P. Friedl, Rho GTPases in collective cell migration, *Small GTPases* 5 (3) (2014) e983869.
- [26] A. J. Ridley, Rho GTPase signalling in cell migration, *Curr. Opin. Cell Biol.* 36 (2015) 103–112.
- [27] R. Mayor, S. Etienne-Manneville, The front and rear of collective cell migration, *Nat. Rev. Mol. Cell Biol.* 17 (2) (2016) 97–109.

- [28] D. V. Köster, S. Mayor, Cortical actin and the plasma membrane: inextricably intertwined, *Curr. Opin. Cell Biol.* **38** (2016) 81–89.
- [29] A. G. Fletcher, M. Osterfield, R. E. Baker, S. Y. Shvartsman, Vertex models of epithelial morphogenesis, *Biophys. J.* **106** (2014) 2291–2304.
- [30] H. Lan, Q. Wang, R. Fernandez-Gonzalez, J. J. Feng, A biomechanical model for cell polarization and intercalation during *Drosophila* germband extension, *Phys. Biol.* **12** (2015) 056011.
- [31] Y. Mori, A. Jilkin, L. Edelstein-Keshet, Wave-pinning and cell polarity from a bistable reaction-diffusion system, *Biophys. J.* **94** (9) (2008) 3684–3697.
- [32] B. Huang, M. Lu, M. K. Jolly, I. Tsarfaty, J. Onuchic, E. Ben-Jacob, The three-way switch operation of Rac1/RhoA GTPase-based circuit controlling amoeboid-hybrid-mesenchymal transition, *Sci. Rep.* **4** (2014) 6449.
- [33] R. Pankov, Y. Endo, S. Even-Ram, M. Araki, K. Clark, E. Cukierman, K. Matsumoto, K. M. Yamada, A Rac switch regulates random versus directionally persistent cell migration, *J. Cell Biol.* **170** (5) (2005) 793–802.
- [34] A. R. Houk, A. Jilkin, C. O. Mejean, R. Boltyskiy, E. R. Dufresne, S. B. Angenent, S. J. Altschuler, L. F. Wu, O. D. Weiner, Membrane tension maintains cell polarity by confining signals to the leading edge during neutrophil migration, *Cell* **148** (1) (2012) 175–188.
- [35] A. Diz-Muñoz, K. Thurley, S. Chintamen, S. J. Altschuler, L. F. Wu, D. A. Fletcher, O. D. Weiner, Membrane tension acts through pld2 and mtorc2 to limit actin network assembly during neutrophil migration, *PLoS Biol.* **14** (6) (2016) e1002474.
- [36] W. R. Holmes, L. Edelstein-Keshet, Analysis of a minimal Rho-GTPase circuit regulating cell shape, *Phys. Biol.* **13** (4) (2016) 046001.
- [37] R. J. Petrie, A. D. Doyle, K. M. Yamada, Random versus directionally persistent cell migration, *Nat. Rev. Mol. Cell Biol.* **10** (8) (2009) 538–549.
- [38] M. Krause, A. Gautreau, Steering cell migration: lamellipodium dynamics and the regulation of directional persistence, *Nat. Rev. Mol. Cell Biol.* **15** (9) (2014) 577–590.
- [39] A. A. Potdar, J. Lu, J. Jeon, A. M. Weaver, P. T. Cummings, Bimodal analysis of mammary epithelial cell migration in two dimensions, *Ann. Biomed. Eng.* **37** (1) (2009) 230–245.
- [40] E. Theveneau, L. Marchant, S. Kuriyama, M. Gull, B. Moepps, M. Parsons, R. Mayor, Collective chemotaxis requires contact-dependent cell polarity, *Dev. Cell* **19** (1) (2010) 39–53.
- [41] R. Gorelik, A. Gautreau, Quantitative and unbiased analysis of directional persistence in cell migration, *Nat. Protoc.* **9** (8) (2014) 1931–1943.

- [42] R. Moore, E. Theveneau, S. Pozzi, P. Alexandre, J. Richardson, A. Merks, M. Parsons, J. Kashef, C. Linker, R. Mayor, Par3 controls neural crest migration by promoting microtubule catastrophe during contact inhibition of locomotion, *Development* 140 (23) (2013) 4763–4775.
- [43] E. Scarpa, A. Roycroft, E. Theveneau, E. Terriac, M. Piel, R. Mayor, A novel method to study contact inhibition of locomotion using micropatterned substrates, *Biol. Open* 2 (9) (2013) 901–906.
- [44] S. Vedel, S. Tay, D. M. Johnston, H. Bruus, S. R. Quake, Migration of cells in a social context, *Proc. Natl. Acad. Sci. U.S.A.* 110 (1) (2013) 129–134.
- [45] E. A. Novikova, M. Raab, D. E. Discher, C. Storm, Persistence-driven durotaxis: Generic, directed motility in rigidity gradients, *Phys. Rev. Lett.* 118 (7) (2017) 078103.
- [46] M. F. Ware, A. Wells, D. A. Lauffenburger, Epidermal growth factor alters fibroblast migration speed and directional persistence reciprocally and in a matrix-dependent manner, *J. Cell Sci.* 111 (16) (1998) 2423–2432.
- [47] D. A. Kulawiak, B. A. Camley, W.-J. Rappel, Modeling contact inhibition of locomotion of colliding cells migrating on micropatterned substrates, *PLoS Comput. Biol.* 12 (2016) e1005239.
- [48] E. Scarpa, A. Szab, A. Bibonne, E. Theveneau, M. Parsons, R. Mayor, Cadherin switch during EMT in neural crest cells leads to contact inhibition of locomotion via repolarization of forces, *Developmental Cell* 34 (4) (2015) 421 – 434.
- [49] E. T. Roussos, M. Balsamo, S. K. Alford, J. B. Wyckoff, B. Gligorijevic, Y. Wang, M. Pozzuto, R. Stobezki, S. Goswami, J. E. Segall, D. A. Lauffenburger, A. R. Bresnick, F. B. Gertler, J. S. Condeelis, Mena invasive (MenaINV) promotes multicellular streaming motility and transendothelial migration in a mouse model of breast cancer, *J. Cell Sci.* 124 (13) (2011) 2120–2131.
- [50] M. D. Bass, K. A. Roach, M. R. Morgan, Z. Mostafavi-Pour, T. Schoen, T. Muramatsu, U. Mayer, C. Ballestrem, J. P. Spatz, M. J. Humphries, Syndecan-4–dependent Rac1 regulation determines directional migration in response to the extracellular matrix, *J. Cell Biol.* 177 (3) (2007) 527–538.
- [51] H. K. Matthews, L. Marchant, C. Carmona-Fontaine, S. Kuriyama, J. Larraín, M. R. Holt, M. Parsons, R. Mayor, Directional migration of neural crest cells in vivo is regulated by Syndecan-4/Rac1 and non-canonical Wnt signaling/RhoA, *Development* 135 (10) (2008) 1771–1780.
- [52] J. Löber, F. Ziebert, I. S. Aranson, Collisions of deformable cells lead to collective migration, *Sci. Rep.* 5 (2015) 9172.

- [53] S. Najem, M. Grant, Phase-field model for collective cell migration, *Phys. Rev. E* **93** (2016) 052405.
- [54] W. Marth, S. Praetorius, A. Voigt, A mechanism for cell motility by active polar gels, *J. R. Soc. Interface* **12** (107) (2015) 20150161.
- [55] B. Smeets, R. Alert, J. Pešek, I. Pagonabarraga, H. Ramon, R. Vincent, Emergent structures and dynamics of cell colonies by contact inhibition of locomotion, *Proc. Natl. Acad. Sci. U.S.A.* **113** (51) (2016) 14621–14626.
- [56] T. Wu, J. J. Feng, Modeling the mechanosensitivity of neutrophils passing through a narrow channel, *Biophys. J.* **109** (11) (2015) 2235–2245.
- [57] B. A. Camley, J. Zimmermann, H. Levine, W.-J. Rappel, Collective signal processing in cluster chemotaxis: Roles of adaptation, amplification, and co-attraction in collective guidance, *PLoS Comput. Biol.* **12** (7) (2016) e1005008.
- [58] J. Varennes, B. Han, A. Mugler, Collective chemotaxis through noisy multicellular gradient sensing, *Biophys. J.* **111** (3) (2016) 640–649.
- [59] J. Delile, M. Herrmann, N. Peyri ras, R. Doursat, A cell-based computational model of early embryogenesis coupling mechanical behaviour and gene regulation, *Nat. Commun.* **8** (2017) 13929.
- [60] F. Fagotto, The cellular basis of tissue separation, *Development* **141** (17) (2014) 3303–3318.
- [61] C. Collins, W. J. Nelson, Running with neighbors: coordinating cell migration and cell–cell adhesion, *Curr. Opin. Cell Biol.* **36** (2015) 62–70.
- [62] T. Das, K. Safferling, S. Rausch, N. Grabe, H. Boehm, J. P. Spatz, A molecular mechanotransduction pathway regulates collective migration of epithelial cells, *Nat. Cell Biol.* **17** (3) (2015) 276–287.
- [63] J. Park, W. R. Holmes, S. H. Lee, H.-N. Kim, D.-H. Kim, M. K. Kwak, C. J. Wang, L. Edelstein-Keshet, A. Levchenko, Mechanochemical feedback underlies coexistence of qualitatively distinct cell polarity patterns within diverse cell populations, *Proc. Natl. Acad. Sci. U.S.A.* **114** (28) (2017) E5750–E5759.
- [64] W. R. Holmes, J. Park, A. Levchenko, L. Edelstein-Keshet, A mathematical model coupling polarity signaling to cell adhesion explains diverse cell migration patterns, *PLoS Comput. Biol.* **13** (5) (2017) e1005524.
- [65] L. Coburn, L. Cerone, C. Torney, I. D. Couzin, Z. Neufeld, Tactile interactions lead to coherent motion and enhanced chemotaxis of migrating cells, *Phys. Biol.* **10** (4) (2013) 046002.

# Supporting Information for “A Rho-GTPase based model explains spontaneous collective migration of neural crest cell clusters”

Brian Merchant<sup>1</sup>, Leah Edelstein-Keshet<sup>1</sup> and James J. Feng<sup>1,2\*</sup>

<sup>1</sup> Department of Mathematics, University of British Columbia, Vancouver, BC V6T 1Z2, Canada

<sup>2</sup> Department of Chemical and Biological Engineering, University of British Columbia, Vancouver, BC V6T 1Z3, Canada

## 1 Kinetic model: diffusion fluxes and kinetic rates

The polarization of the cell is controlled by the autocatalysis of Rac1 and RhoA, as well as their mutual inhibition [1,2]. As discussed in the main text, the model tracks fractions of Rac1 and RhoA in their membrane-bound active and inactive states, as well as in the cytosol. The total amount of Rac1 is  $\mathcal{R}_T = \sum \mathcal{R}_i^a + \sum \mathcal{R}_i^i + \mathcal{R}^c$ , the sum of all membrane-bound active ( $\mathcal{R}_i^a$ ) and inactive ( $\mathcal{R}_i^i$ ) Rac1 over all vertices ( $i = 1, \dots, N$ ), plus the amount uniformly dissolved in the cytosol ( $\mathcal{R}^c$ ). Then at each vertex on the membrane, we define the fractions  $R_i^a = \mathcal{R}_i^a / \mathcal{R}_T$  and  $R_i^i = \mathcal{R}_i^i / \mathcal{R}_T$ . For the cytosolic Rac1, the fraction is  $R^c = \mathcal{R}^c / \mathcal{R}_T$ . RhoA fraction are defined similarly, using the symbol  $\rho$  in place of  $R$ . These fractions are the dimensionless variables used in the kinetic model. To convert them into amount of the proteins, we need the total amount of Rac1  $\mathcal{R}_T$  and RhoA  $\rho_T$  in a cell. These values are estimated from the literature and given in Table 2 below.

In order to write reaction-diffusion equations (RDEs) for Rac1 and RhoA, we need to formulate the boundary diffusion fluxes for each. This requires converting the fractional amount on a vertex, say  $R_i^a$ , into a local concentration. As the total amount  $\mathcal{R}_T$  is constant, we can use the “fractional concentration” instead of the actual concentration in the RDEs. In our notation, the edge between vertices  $i$  and  $i + 1$  has a length  $l_i$ . Thus, vertex  $i$  is sandwiched by two edges of length  $l_{i-1}$  and  $l_i$ . As the amount of a GTPase, say  $R_i$ , is defined at each vertex, we calculate its “concentration” at the given vertex by spreading  $R_i$  over half of the two neighboring edges length  $L_i = (l_i + l_{i-1})/2$ . Now the diffusive flux from vertex  $i$  toward  $i + 1$  can be obtained from Fick’s law:

$$J_i = -D \frac{R_{i+1}/L_{i+1} - R_i/L_i}{l_i}. \quad (\text{S1})$$

---

\*Corresponding author. E-mail: james.feng@ubc.ca

The reaction-diffusion equations can thus be discretized using the net flux ( $J_{i-1} - J_i$ ) at this vertex, as done in the main article (Eqs. 1–8). The fluxes can be written for all membrane-bound Rac1 and RhoA species in the same way. Recall that in the above,  $l_i$  and  $L_i$  may vary in time as the edges stretch and contract elastically.

Let us now turn to the reaction rates  $K^\pm$  and  $\rho^\pm$  in the reaction-diffusion equations in the main article (Eqs. 2, 3, 6, 7). It is through these rates that the nonlinear effect allowing for cell polarization enters, as do signals due to CIL, COA, tension mediated Rac1 inhibition, and stochastic modulation of repolarization. These interactions are illustrated in the wiring diagram of Fig. S3.

- The Rac1 activation rate  $K^+(i, t)$ , at vertex  $i$  and time  $t$ , is the sum of a baseline rate  $K_b^+$  and an auto-activation rate  $K_A^+$ , each modulated by additional effects:

$$K^+(i, t) = (x_r(i, t) + x_a(i, t))K_b^+ + \frac{[R_i^a/L_i(t)]^3}{C_H^3 + [R_i^a/L_i(t)]^3}K_A^+, \quad (\text{S2})$$

where  $x_r$  is the randomization factor and  $x_a$  is the co-attraction (COA) factor. Their time-dependence is understood but not always written out explicitly in the following. The auto-activation of Rac1 is represented by a Hill function with a half-max constant  $C_H$  [3]. The randomization factor allows the cell to periodically nucleate new Rac1 hotspots which may potentially mature and out-compete the existing protrusive front. Roughly every  $T_r$  minutes (sampled from a normal distribution with a mean of  $T_r$  and a standard deviation of  $0.1T_r$ ), we randomly select 25% of the vertices on a cell and set  $x_r = 12$  on them for the duration until the next random selection, at which point  $x_r$  is put to 1 on those vertices that are not chosen again for the new cycle of upregulation. This produces the random Rac1 hotspots. Numerical experimentation shows that the model is not sensitive to the percentage 25%. It produces qualitatively the same outcome when the percentage changes from 6.25% to 50%, although for very low percentages a larger  $x_r$  value is needed. For all simulations we have used  $T_r = 20$  min.

The COA factor  $x_a(i)$  is calculated as follows. For cell  $a$ , we calculate the distance  $s_{ij}$  between its vertex  $i$  and vertex  $j$  on another cell  $b$ . Following Woods *et al.* [4], we assume an exponential decay of the C3a signal emitted from  $j$  such that vertex  $i$  on cell  $a$  receives a COA signal

$$\chi(i, j) = \frac{2^{-s_{ij}/l_c} M_{\text{COA}}}{(1 + y_{ij})^2}, \quad (\text{S3})$$

where  $l_c = 110 \mu\text{m}$  is the half-decay length [4] and  $M_{\text{COA}}$  is a dimensionless maximum COA strength. The denominator of Eq. (S3) is a “screening effect” due to cells intervening between vertices  $i$  and  $j$ ,  $y_{ij}$  being the number of times that the line connecting  $i$  and  $j$  passes through another cell other than  $a$  and  $b$ . Finally,  $x_a(i) = \sum_b \sum_{j=1}^N \chi(i, j)$  is computed by summing  $\chi(i, j)$  over all vertices of all the other cells.

As the COA factor  $x_a$  sums up influence from all other cells, it tends to be greater for larger clusters. To normalize the COA effect, we have tuned the maximum COA strength  $M_{\text{COA}}$  to

achieve a mean cell-cell distance (between centroids) of 1.4–1.5 times its initial value. Among the simulations reported in the main article,  $M_{\text{COA}} = 24$  for the two-cell migration of Fig. 6. In subsection 3.5, the following values have been used for different cluster sizes:

$n$	4	9	16	25	36	49
$M_{\text{COA}}$	24.0	16.0	14.0	12.0	9.0	8.0

The above values correspond to  $N = 16$  vertices on each cell. Should the number of vertices  $N$  change,  $M_{\text{COA}}$  is scaled by  $16/N$  so as to maintain the same level of COA effect regardless of the artificial level of membrane discretization.

- The Rac1 inactivation rate  $K^-(i, t)$  is the sum of a baseline rate  $K_b^-$  and a RhoA-mediated mutual inhibition rate  $K_{MI}^-$ , each modulated by additional effects:

$$K^-(i, t) = [x_s(t) + x_{\text{CIL}}(i, t) + x_{\text{CIL}}^b(i, t)]K_b^- + \frac{[\rho_i^a/L_i(t)]^3}{C_H^3 + [\rho_i^a/L_i(t)]^3}K_{MI}^-, \quad (\text{S4})$$

where  $x_s$  is the tension-mediated Rac1 inhibition factor,  $x_{\text{CIL}}(i)$  is the CIL signal at vertex  $i$ , and  $x_{\text{CIL}}^b(i)$  is a CIL-like signal for contact between the cell and a corridor boundary if one exists. The RhoA inhibition of Rac1 is reflected by the factor in front of  $K_{MI}^-$ , which is a Hill function of the local active RhoA with a half-max constant  $C_H$ .

The tension-mediated Rac1 inhibition [5] is modelled as a Hill function of the strain of the entire cell circumference  $s$ , as the tension inhibition likely saturates at large tension or strain:

$$s = \frac{1}{Nl_0} \sum_{i=1}^N l_i - 1, \quad (\text{S5})$$

$$x_s = \max\left(M_s \frac{s^3}{s_H^3 + s^3}, 0\right), \quad (\text{S6})$$

where  $l_0$  is the length of each edge of the undeformed cell,  $l_i$  is that of a deformed edge,  $s_H = 0.1$  is the half-max constant and  $M_s = 40$  is the maximum magnitude of strain mediated Rac1 inactivation. These values are chosen to produce a suitable level of tension-inhibition of Rac1. If the entire membrane is under compression,  $x_s$  is set to 0.

The CIL factor  $x_{\text{CIL}}(i)$  is activated if the straight-line distance  $d_i$  between vertex  $i$  and the closest edge or vertex on a neighboring cell falls below a threshold  $d_m$ . We define a truncated linear ‘‘CIL influence’’

$$x_c(i) = \begin{cases} 0, & d_i \geq d_m \\ M_{\text{CIL}}(1 - \frac{d_i}{d_m}), & d_i < d_m \end{cases}, \quad (\text{S7})$$

where  $M_{\text{CIL}}$  is the maximum magnitude of the CIL signal. Then  $x_{\text{CIL}}(i)$  is calculated from

the average of the 3 neighboring vertices:

$$x_{\text{CIL}} = \frac{x_c(i-1) + x_c(i) + x_c(i+1)}{3}. \quad (\text{S8})$$

We have chosen  $d_m = 0.5 \mu\text{m}$  and  $M_{\text{CIL}} = 60$  for our membrane of  $N = 16$  vertices. If the number of vertices  $N$  should change,  $M_{\text{CIL}}$  changes as  $60 \times 16/N$  so as to maintain the same level of CIL effect regardless of the artificial level of membrane discretization. The boundary CIL factor  $x_{\text{CIL}}^b$  is calculated similarly.

- The RhoA activation rate  $\kappa^+$  consists of a baseline rate  $\kappa_b^+$  modulated by CIL and boundary-based CIL, and an auto-activation term modelled by a Hill function:

$$\kappa^+(i, t) = [x_{\text{CIL}}(i, t) + x_{\text{CIL}}^b(i, t)]\kappa_b^+ + \frac{[\rho_i^a/L_i(t)]^3}{C_H^3 + [\rho_i^a/L_i(t)]^3}\kappa_A^+, \quad (\text{S9})$$

where  $x_{\text{CIL}}$  and  $x_{\text{CIL}}^b$  are the same as above, and  $C_H$  is the half-max constant for RhoA auto-activation.

- The RhoA inactivation rate  $\kappa^-$  consists of a baseline rate  $\kappa_b^-$  and a Rac1-mediated mutual inhibition effect modelled by a Hill function:

$$\kappa^-(i, t) = \kappa_b^- + \frac{[R_i^a/L_i(t)]^3}{C_H^3 + [R_i^a/L_i(t)]^3}\kappa_{MI}^-, \quad (\text{S10})$$

$C_H$  being the half-max constant in the Hill function.

- The membrane association and dissociation rates for Rac1 and RhoA,  $M^\pm$  and  $\mu^\pm$ , are constants given in Table 2. Since they multiply onto the local amount of the proteins to give their local rate of change, these constants do not vary with the edge length or  $N$ .

## 2 Protrusion and contraction forces

Our model omits the detailed pathways leading from polarization of GTPases to the production of protrusion and contraction forces. Instead, we express these forces as truncated linear functions of the active Rac1 and RhoA levels at each vertex in Eq. (11) of the main article, re-written below:

$$F_i = \begin{cases} K_R \min(R_i^a - \rho_i^a, 0.05), & R_i^a > \rho_i^a \\ -K_\rho \min(\rho_i^a - R_i^a, 0.05), & R_i^a \leq \rho_i^a \end{cases}.$$

Models for Rac-Rho competition, including our own, typically have regimes of bistability where either Rho or Rac dominate. We have chosen here to link the force of protrusion/contraction directly to that Rac-Rho balance, for simplicity.

The reason for implementing a cap to the linear dependence is as follows. Prass *et al.* [6] measured the protrusion forces at the forefront of lamellipodia by placing an atomic force microscopy (AFM) cantilever in the path of a migrating fish keratocyte. They obtained a stalling force around 1 nN, corresponding to about 100 polymerizing actin filaments pushing on every 1  $\mu\text{m}$  of the leading edge. The upper bound on this force may be due to a variety of constraints, e.g. limited space for a certain number of actin barbed ends and finite supply of any of the proteins regulating actin polymerization. Without further information, we have implemented an upper bound on  $R^a - \rho^a$  in our highly simplified expression above.

The threshold of 0.05 is inspired by the modeling of Jilkin *et al.* [3]. These authors inferred from experimental sources that a migrating cell should have about 40% of its GTPases in the membrane-bound activated state. In our notation, this corresponds to 2.5% on each of the  $N = 16$  vertices of our model membrane. Considering that only about half of the vertices will be inside the Rac1-rich protrusion front, the level of  $R^a$  on such a vertex should be on the order of 5%. This characteristic level is used, absent more pertinent data, as the threshold at which the upper bound for the forces is set.

Aside from the “active forces” due to Rac1 and RhoA, we have also implemented a cytoplasmic pressure  $p$  (see Eq. 10 in the main article) to help conserve the cell area. The parameter  $K_c$  is related to the bulk modulus of the cell. As we have found few relevant data in the literature, we have used a value of  $K_c = 625$  nN to keep the cell area change at about 5% on average. As the pressure force  $p$  is applied onto each vertex, the coefficient  $K_c$  should be varied according to  $625 \times 16/N$  nN if a different  $N$  is used.

### 3 Parameter Estimation

We divide the model parameters roughly into two categories: geometric and physical parameters, and kinetic rate constants. Each group is tabulated below, along with sources for the adopted values as well as additional explanations given in footnotes.

Symbol	Description	Value	Sources
$d$	cell diameter	$40 \mu\text{m}$	[7]
$\eta$	viscous friction factor	$100 \text{ nN s } \mu\text{m}^{-1}$	[8–10] <sup>1</sup>
$\lambda$	stiffness of cortex	$80 \text{ nN } \mu\text{m}^{-1}$	[11, 12] <sup>2</sup>
$D$	membrane diffusivity	$0.25 \mu\text{m}^2 \text{ s}^{-1}$	[3, 13]
$K_R$	protrusive force constant	$\frac{0.3Nl_0}{0.4} \times 2000 \text{ nN } \mu\text{m}^{-1}$	[6] <sup>3</sup>
$K_\rho$	contractile force constant	$0.2K_R$	[6] <sup>4</sup>

Table 1: Geometric and physical parameters used in our model.

<sup>1</sup>Ref. [8] measured the viscosities of various embryonic tissue, and Refs. [9, 10] adapted these to 2D viscosities of  $100\text{--}1000 \text{ nN s } \mu\text{m}^{-1}$ .

<sup>2</sup>We have converted the measured stiffness from 3D to 2D by multiplying it by a characteristic length of  $10 \mu\text{m}$ .

<sup>3</sup>The stalling force on actin filaments in lamellipodia is around  $10 \text{ kPa}$  [6]. Considering a lamellipodial height of  $200 \text{ nm}$ , we estimate a maximum protrusion force of  $2000 \text{ nN}$  per  $\mu\text{m}$  of the membrane in our 2D model. Assuming a fraction (0.3) of the maximum force is achieved at a typical 40% whole-cell Rac1 activation level [3], we estimate the  $K_R$  coefficient for each of the  $N$  membrane vertices as the value shown.  $l_0 = d \sin(\pi/N)$  is the length of each of the  $N$  edges of the cell membrane, in units of  $\mu\text{m}$ , and  $N = 16$  is used for all the simulations except in Fig. S2.

<sup>4</sup>The contractile force constant  $K_\rho$  should be lower than the protrusive force constant  $K_R$  on account of adhesive frictions on the substrate. We have chosen the factor 0.2 such that protrusion/contraction forces on a single cell, coupled with the friction factor  $\eta$ , produce the correct single-cell speed.

Symbol	Description	Value	Sources
$C_H$	half-max constant in Hill function	$\frac{0.4}{Nl_0} \mu\text{m}^{-1}$	[3] <sup>5</sup>
$M^-, \mu^-$	Rac1, RhoA membrane dissociation rate	$0.15 \text{ s}^{-1}$	[14] <sup>6</sup>
$M^+, \mu^+$	Rac1, RhoA membrane association rate	$0.02 \text{ s}^{-1}$	[14]
$K_b^+$	baseline Rac1 activation rate	$2.4 \times 10^{-3} \text{ s}^{-1}$	[3, 15, 16] <sup>7</sup>
$\kappa_b^+$	baseline RhoA activation rate	$2.8 \times 10^{-3} \text{ s}^{-1}$	[3, 15, 16] <sup>7</sup>
$K_a^+$	maximum Rac1 auto-activation rate	$5 \times 10^{-2} \text{ s}^{-1}$	[3, 15, 16] <sup>7</sup>
$\kappa_a^+$	maximum RhoA auto-activation rate	$3.9 \times 10^{-2} \text{ s}^{-1}$	[3, 15, 16] <sup>7</sup>
$K_b^-$	baseline Rac1 deactivation rate	$8 \times 10^{-4} \text{ s}^{-1}$	[3, 15, 16] <sup>8</sup>
$\kappa_b^-$	baseline RhoA deactivation rate	$6 \times 10^{-3} \text{ s}^{-1}$	[3, 15, 16] <sup>9</sup>
$K_{\text{MI}}^-$	rate of Rac1 inhibition by RhoA	$0.4 \text{ s}^{-1}$	[3, 15, 16] <sup>8</sup>
$\kappa_{\text{MI}}^-$	rate of RhoA inhibition by Rac1	$4 \times 10^{-2} \text{ s}^{-1}$	[3, 15, 16] <sup>9</sup>
$\mathcal{R}_T$	total amount of Rac1 in a cell	$2.5 \times 10^6$	[3, 17] <sup>10</sup>
$\rho_T$	total amount of RhoA in a cell	$1 \times 10^6$	[3, 17] <sup>10</sup>

Table 2: Biochemical parameters used in our model.

<sup>5</sup>We assume half-max in the Hill function at a characteristic steady-state whole-cell activity of 40% [3]. As the membrane has  $N$  segments of length  $l_0 = d \sin(\pi/N)$ , the half-max concentration is computed as given. For lack of data that would distinguish the various auto-activations and mutual inhibitions, we have used the same  $C_H$  for all the Hill functions involved.

<sup>6</sup>This is the measured Rho GTPase-GDI binding rate. We assume that Rho GTPases will only dissociate from the membrane in their inactive form, and their subsequent binding with cytosolic GDIs is rapid [13, 18].

<sup>7</sup>Using purified proteins, Ref. [15] measured an intrinsic activation rate of  $1.5 \times 10^{-4} \text{ s}^{-1}$  for both Rac1 and RhoA, but reported that GEF stimulation could accelerate activation 5 to 1000 times. Using the lifespan of membrane-bound Rho GTPases, Refs. [3, 16] arrived at estimations comparable to that used here.

<sup>8</sup>Using purified proteins, Ref. [15] measured an intrinsic Rac1 deactivation rate of  $1.8 \times 10^{-4} \text{ s}^{-1}$ , but reported that GAP stimulation could accelerate deactivation 5 to 4000 times. Refs. [3, 16] used values comparable to ours.

<sup>9</sup>Using purified proteins, Ref. [15] measured an intrinsic RhoA deactivation rate of  $3.5 \times 10^{-4} \text{ s}^{-1}$ , but reported that GAP stimulation could accelerate deactivation 5 to 4000 times. Refs. [3, 16] used values comparable to ours.

<sup>10</sup>Our Rho GTPases are represented as fractions of the total Rac1  $\mathcal{R}_T$  and total RhoA  $\rho_T$ . These total numbers of proteins are required to convert our dimensionless results to actual concentrations. Ref. [3] used values for COS-1 cells, which are similar in size to *Xenopus* NCC cells (see Fig. 2 in [17]). In estimating the protein copy numbers, we have adopted an approximate Rho GTPase molecular weight of 21 kDa  $\approx 3.5 \times 10^{-11} \text{ ng}$  [3]. For CHO epithelial cells, Ref. [19] reported Rac1 and RhoA amounts that are higher by a factor of about 2.

## 4 Sensitivity to kinetic parameters

Given the large number of kinetic parameters in Table 2, and the wide range of possible variations due to GAP stimulation [15], one wonders how sensitive the outcome of the model is to the choice of these parameters. These kinetic rates are not easily measurable, but the overall timescales and steady-state activity levels (relative to baseline) can sometimes be inferred. Note that the mechanical parameters are not subject to the same degree of variability.

Instead of testing each kinetic parameter individually, we can categorize them based on their functions in the model. The desired outcome of the model, persistent collective migration, depends on three factors: cell polarization, COA and CIL. All three are governed by the activation and deactivation rate constants for Rac1 and RhoA:  $K_b^+$ ,  $K_A^+$ ,  $K_b^-$ ,  $K_{MI}^-$ ,  $\kappa_b^+$ ,  $\kappa_A^+$ ,  $\kappa_b^-$ ,  $\kappa_{MI}^-$ . Prior modeling [2, e.g.] has shown that the spontaneous cell polarization hinges on a balance between the linear baseline rates, such as  $K_b^+$  of Eq. (S2) and  $\kappa_b^+$  of Eq. (S9), on the one hand, and the nonlinear auto-catalytic or mutual inhibitive rates, such as  $K_A^+$  and  $\kappa_A^+$ , on the other. In a simple-minded argument, raising or lowering these rates simultaneously by the same factor would not change the outcome of polarization but merely change the time scale for the process. This turns out to be more or less the case in our model, albeit with minor complications due to parameters and effects that do not scale linearly with these kinetic rates. For example, the membrane diffusivity  $D$  and membrane-cytosolic transport coefficients  $M^\pm$  and  $\mu^\pm$  are kept constant and do not vary in proportion. More importantly, the COA and CIL effects, which modulate the baseline kinetic rates (see Eqs. S2, S4, S9), depend nonlinearly on the location and geometry of cells, and couple the cell mechanics with the biochemistry. In view of the above, we have adopted the following protocol for testing parameter sensitivity. First, we designate the parameter set of Table 2 as the “standard set”. Then we multiply the activation and deactivation rate constants by a factor of 0.1 to produce the “hypo-set”, and by 10 to produce the “hyper-set”. We then compare the model predictions for these three sets of parameters in a test case of 4 cells migrating in a corridor of width  $w = 1$ . At the start the 4 cells are placed at the end of the corridor, in contact with each other.

The main observation is that cell polarization appears largely intact for all three sets. Persistence of polarity and spontaneous collective migration in the corridor also remain qualitatively the same (Fig. S12). CIL is produced by using the standard  $M_{CIL} = 60$  in all three sets. However, we have increased  $M_{COA}$  from the standard value of 24 to 48 for the hypo-set, and reduced it to 18 for the hyper-set to maintain the same average cell separation. Going from the hypo-set to the standard set and then to the hyper-set, we see that faster kinetic rates produce more consistent collective migration. Nevertheless the various qualitative features remain the same. Therefore, Fig. S12 suggests that the model prediction of persistent collective migration is robust over at least two orders of magnitude in the kinetic rates for Rac1 and RhoA activation and deactivation.

We have also explored the working range for  $M_{COA}$  and  $M_{CIL}$  (Fig. S13). With  $M_{COA}$  fixed at its standard value of 24, increasing the CIL strength  $M_{CIL}$  generally increases the cell-cell separation  $S$ , but this effect saturates quickly (Fig. S13a). Meanwhile, the other properties of interest,  $V_c$  and  $m_I$ , are hardly affected. Putting  $M_{CIL}$  to zero abolishes POP, as expected. But for the range

tested,  $30 \leq M_{\text{CIL}} \leq 210$ , the results are insensitive to  $M_{\text{CIL}}$ . As  $M_{\text{CIL}}$  affects the rate at which cells repolarize away from each other or from a wall upon collision, we conclude that as long as this happens sufficiently fast (for  $M_{\text{CIL}}$  down at least to 30),  $M_{\text{CIL}}$  has little effect on the model predictions.

With  $M_{\text{CIL}} = 60$  fixed, increasing the COA strength  $M_{\text{COA}}$  generally reduces cell separation and promotes collective migration (Fig. S13*b*). Even at  $M_{\text{COA}} = 8$ , a third of the standard value, collective migration occurs with reasonable reliability. With stronger COA, the cell cluster is better able to resist the perturbations due to random repolarization of individual cells. Thus, both  $V_c$  and  $m_I$  tend to increase with  $M_{\text{COA}}$ . However, as  $M_{\text{COA}}$  goes beyond 24, its effect on  $V_c$  and  $m_I$  saturates. In fact, for the largest values tested, the average migration efficiency tends to drop, with increasing occurrence of the cluster migration being arrested. In such cases the cells attract each other so strongly as to severely restrict the excursion of the cells at the outer edges. COA overwhelms CIL and spontaneous collective migration is lost. Taken together, the results of Fig. S13 show robustness of the model over a wide range (roughly one order of magnitude) of the COA and CIL parameters.

## 5 Supplemental Figures

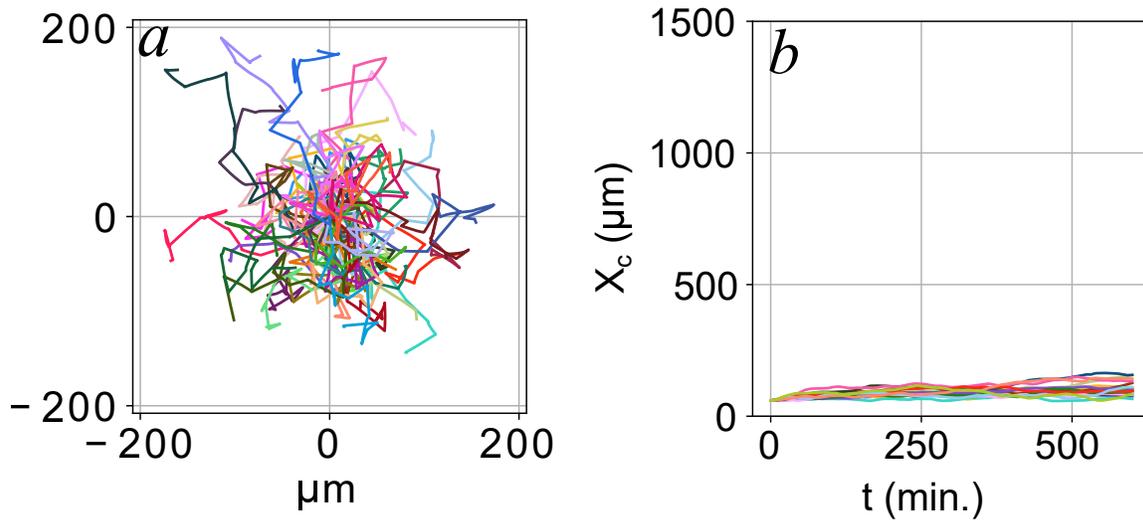


Figure S1: Model with a “wipeout” random repolarization scheme fails to produce persistence of polarity and directional migration (see also Movie 1). This model produces random repolarization by periodically erasing the existing polarization of a cell and imposing a random new Rac1 and RhoA distribution on the membrane roughly every  $T_r$  minutes. This precludes Rac1 modulation as a factor in the repolarization. Although the model preserves CIL and COA, they can no longer produce persistent protrusive fronts over a long period of time, nor spontaneous collective migration. (a) For a single cell, the wipeout scheme produces similar random repolarization to that of Fig. 3(a) in the main article. Here 50 realizations are shown, with average persistence ratio  $R_p = 0.29 \pm 0.14$  over 4 hours, and persistence time  $T_p = 13 \pm 6$  min. (b) However, a cluster of 16 cells fail to show persistent collective migration under otherwise identical conditions to Fig. 10 ( $n = 16$ ) in the main article.

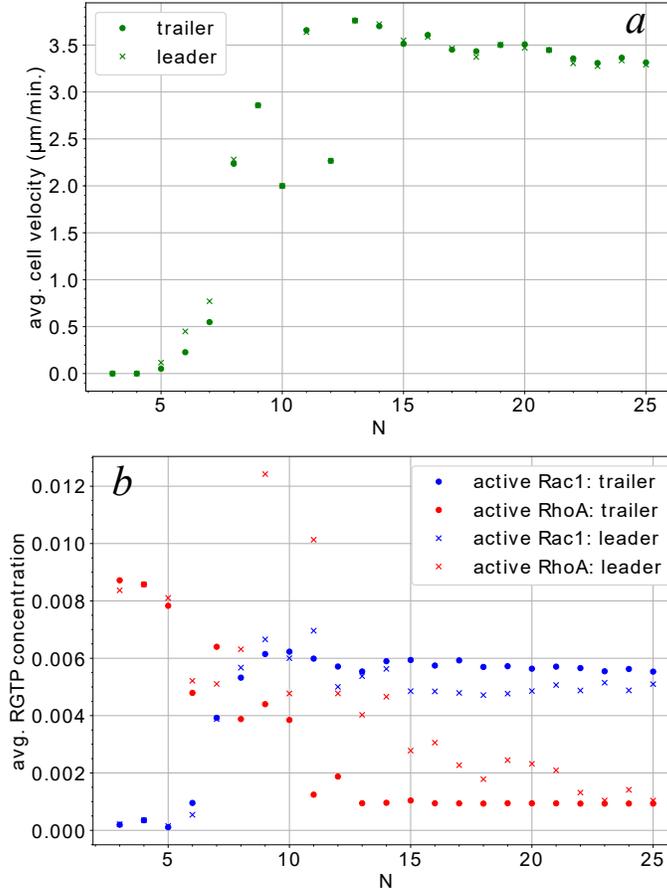


Figure S2: Test of convergence. We refine the resolution of the membrane by increasing  $N$  up to 25 for two cells migrating down a corridor whose width equals the cell diameter. To match conditions closely between different  $N$ , we disable random repolarization and impose a common initial condition with the entire amounts of membrane-bound active and inactive Rac1 (10% each) assigned at the rightmost vertex and the same amounts of membrane-bound active and inactive RhoA assigned at the leftmost vertex. The simulations last 4 hours. With increasing  $N$ , the maximum COA and CIL strengths,  $M_{\text{COA}}$  and  $M_{\text{CIL}}$ , along with the pressure coefficient  $K_c$ , are scaled according to  $N^{-1}$  so as to maintain the same overall physical levels of these effects. (a) Cell speed has stabilized by  $N = 15$ . We average the cell speeds over the latter half of the simulation when a quasi-steady state has been achieved. (b) The maximum nodal concentration of active Rac1 and RhoA averaged over the latter half of the simulation. The concentration of all species has stabilized by  $N = 15$  except for the active RhoA of the leader. Its persistent variation is because the number of cell-cell contacts is a sensitive quantity to capture. As  $N$  changes, we often see one more or fewer contact by the end of the simulation. Since cell contacts cause RhoA peaks in the leader by CIL, this produces the lingering variations. We have chosen  $N = 16$  for all simulations reported in the main paper as a compromise between accuracy and computational cost. Single-cell simulations in prior models (e.g. [20]) have enjoyed finer resolutions. But for the cell clusters studied here, with up to 49 cells, a higher  $N$  incurs prohibitive cost for computing CIL and COA as our current algorithm requires pairwise interactions among vertices on different cells. Although better algorithms may reduce the cost, given the objectives of the current model, we opted for a compromise that tracks each cell with less resolution while capturing details of the cell-cell interactions in the cluster.

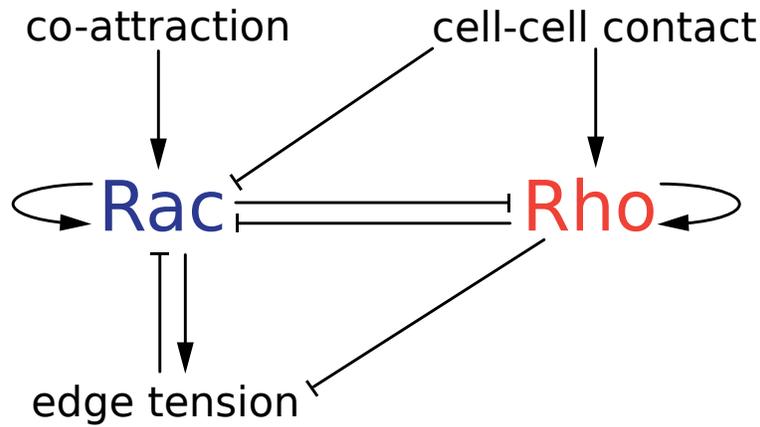


Figure S3: Wiring diagram for the autocatalysis and mutual inhibition of Rac1 and RhoA. These GTPases are also affected by edge tension, co-attraction and contact inhibition.

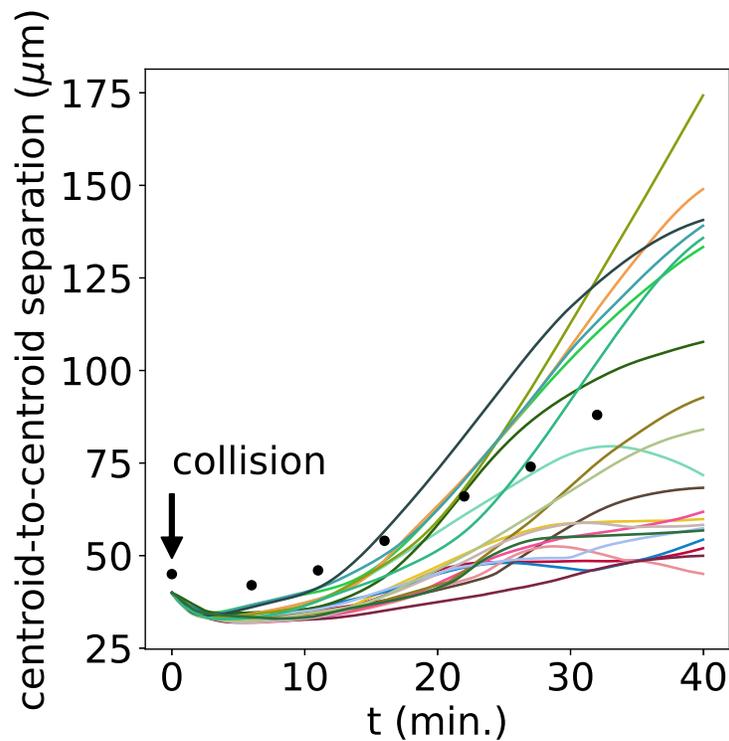


Figure S4: The temporal evolution of the separation between two cells colliding in a corridor with COA turned on, under conditions that are otherwise identical to those of Fig. 4 in the main text. Twenty runs are plotted in comparison with experimental data (black circles) extracted from Fig. 3g of Scarpa *et al.* [21]. For different runs, collision occurs at different times, and we have shifted them, as well as the experimental data, so collision is at  $t = 0$ . Averaged among the 20 runs, the separation is about 75  $\mu\text{m}$  at 30 min after collision, slightly below the experimental separation of about 80  $\mu\text{m}$ .

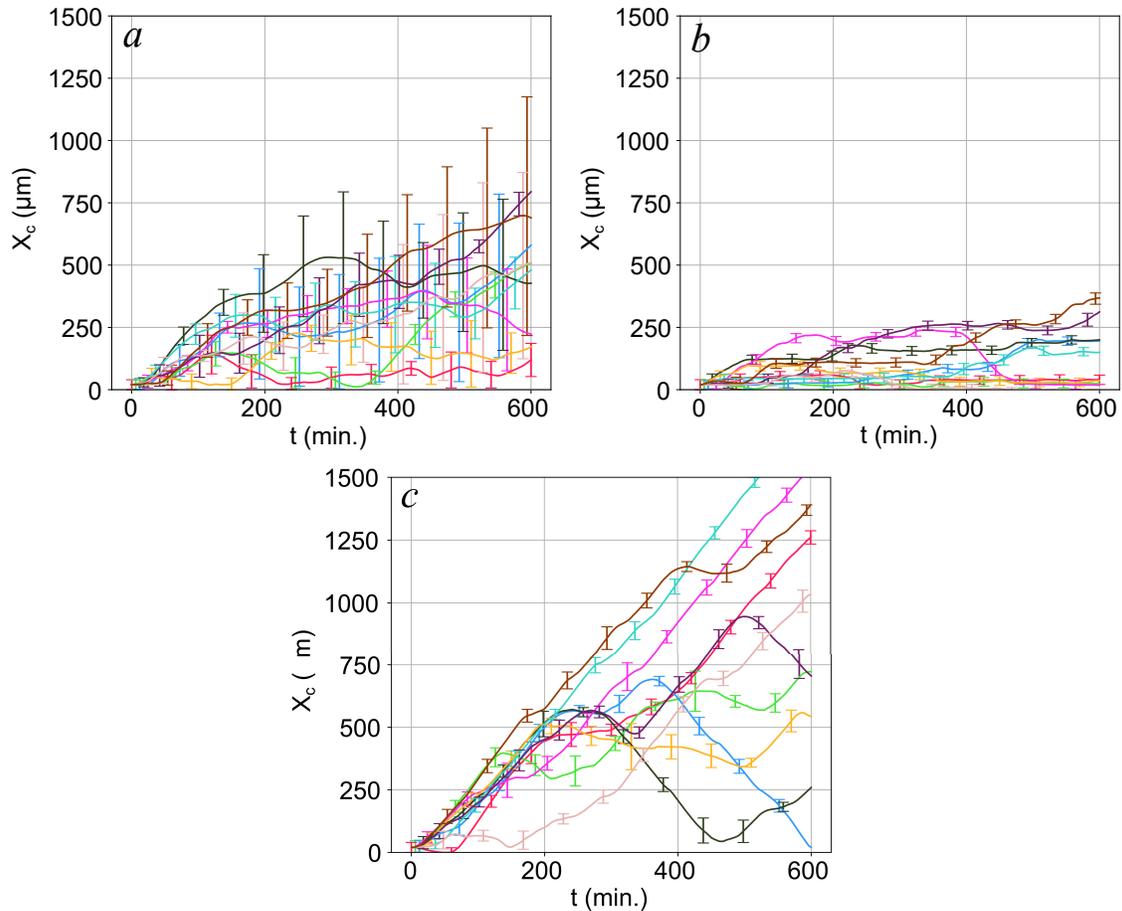


Figure S5: Co-attraction (COA) and contact-inhibition of locomotion (CIL) must operate simultaneously to produce persistence of polarity (POP). This is illustrated by turning off COA or CIL for the two-cell simulation of Fig. 6 in the main article. (a) COA is turned off but CIL is kept on. The position of the centroid of the two cells is shown as a function of time for 10 runs starting from random initial Rac1 and RhoA distributions, with the ends of the error bars denoting the centroids of the two cells. The two cells separate within the corridor, resulting in poor collective migration. (b) CIL is turned off but COA is kept on. The cells stay as a close doublet and do not retract from each other for lack of CIL. But the pair fail to migrate with persistent directionality, and remain mostly near their initial location. (c) With both COA and CIL, the pair show much greater efficiency in spontaneous collective migration. The reversal of some of the trajectories is discussed in association with Fig. 10 in the main article.

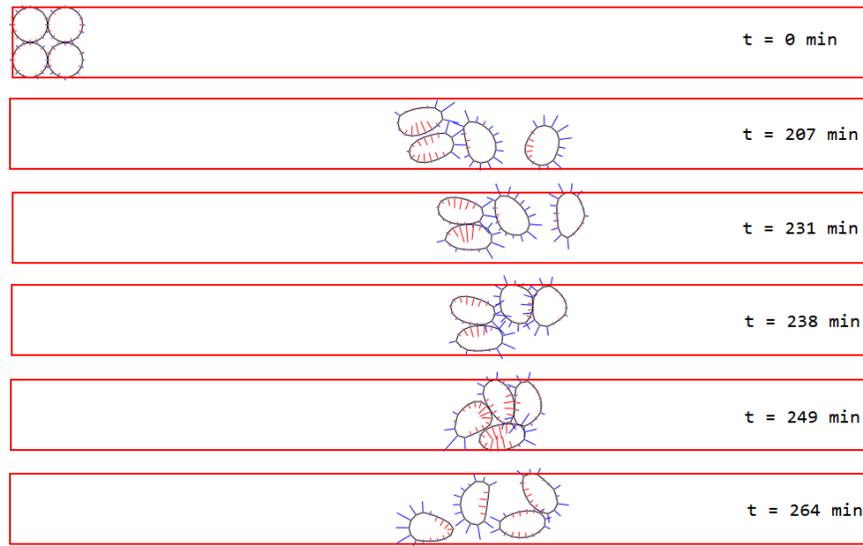


Figure S6: Illustration of the failure of POP: a 4-cell cluster that reverses course when two cells simultaneously develop rear-facing Rac1 hotspots. Having started from the initial configuration of  $t = 0$  min, the cluster migrates persistently for some time, as illustrated by the snapshot of  $t = 207$  min. At  $t = 231$  min, by chance, the two leading cells simultaneously developed Rac1 hotspots in their rear, which cause them to slow and bump into the other two cells at  $t = 238$  min. This in turn reverses the polarity of the latter ( $t = 249$  min), and by  $t = 264$  min, the whole group has reversed course. Upon loss of the persistent migration toward the right, the group centroid essentially executes a random walk, and never regains persistent collective migration. See also Movie 7.

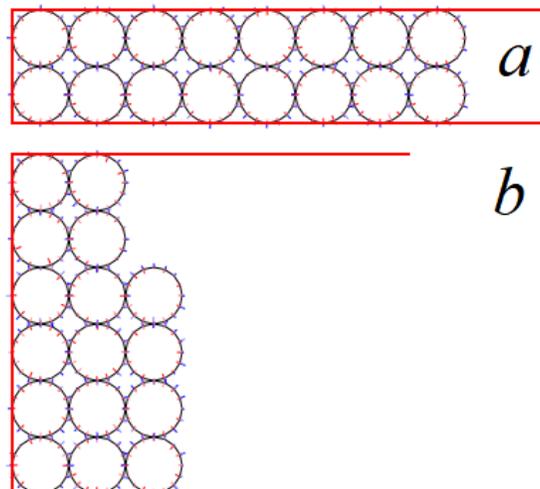


Figure S7: Initial configurations for testing the effects of confinement, corresponding to the simulations in Fig. 11 of the main article. A cluster of  $n = 16$  cells are fitted into corridors of different width  $w$  (measured in cell diameters), with (a)  $w = 2$  and (b)  $w = 6$  being shown as examples.

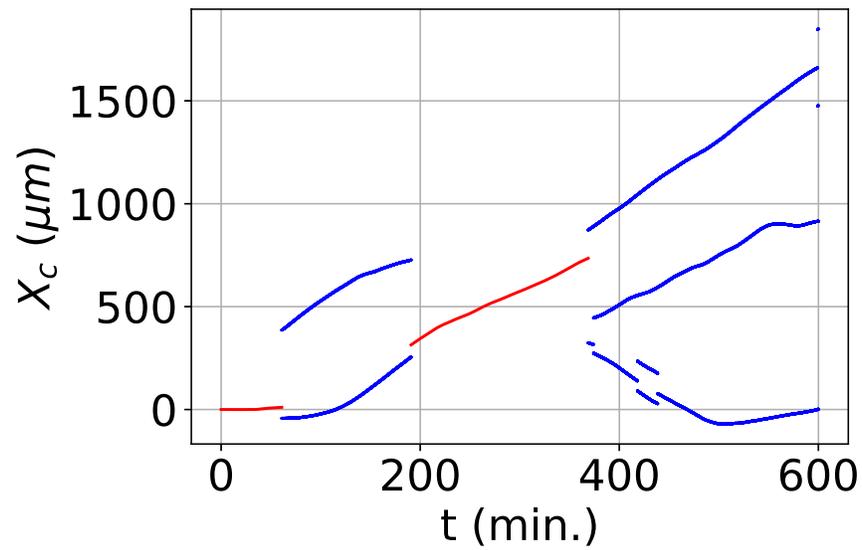


Figure S8: In a narrow corridor of width  $w = 1$ , a single-file cluster tends to break up into smaller groups, each migrating as a small cluster and possibly merging again after some time. Here we illustrate this behavior by plotting the trajectory of the centroid of a 16-cell cluster and the centroids of its sub-clusters as the cells migrate down the corridor. A cluster is considered split if the separation between the two subgroups (measured between cell membranes) exceeds two cell diameters.

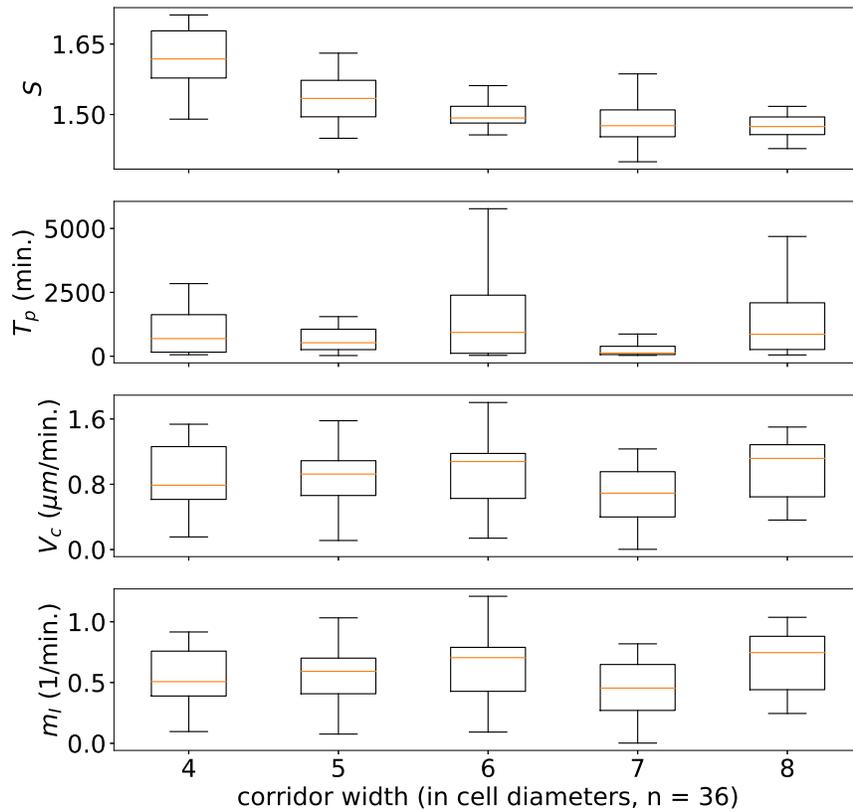


Figure S9: Optimal confinement at  $w = 6$  for a cluster of  $n = 36$  cells. The conditions for the simulations and definition of the quantities plotted are the same as in Fig. 11 of the main text.

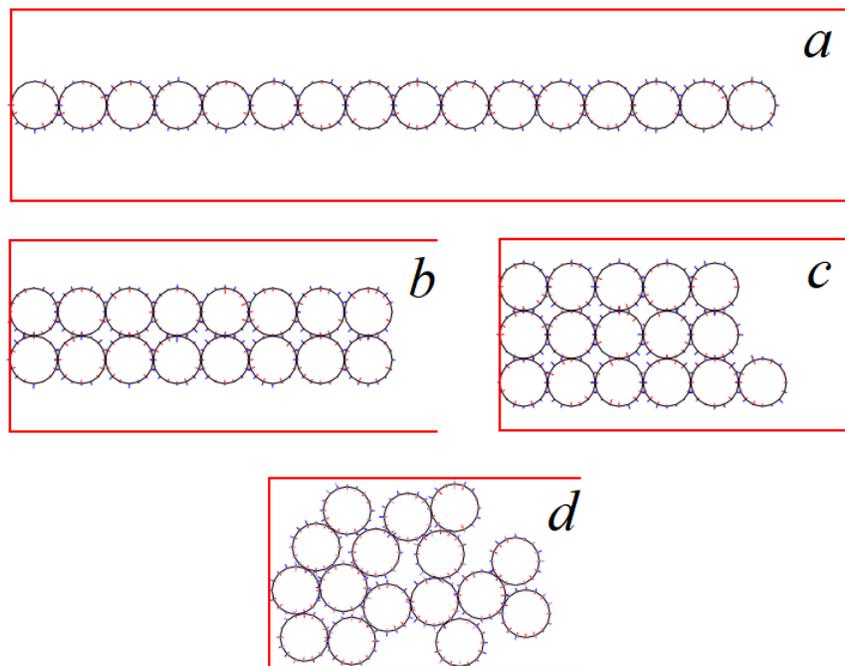


Figure S10: Initial configurations of a 16-cell cluster in a  $w = 4$  corridor. (a) A single-file configuration; (b) a two-row configuration; (c) a three-row configuration; (d) a random initial configuration, generated by placing the first cell next to the vertical corridor boundary, and adding neighbours next to existing cells in random directions while avoiding overlap.

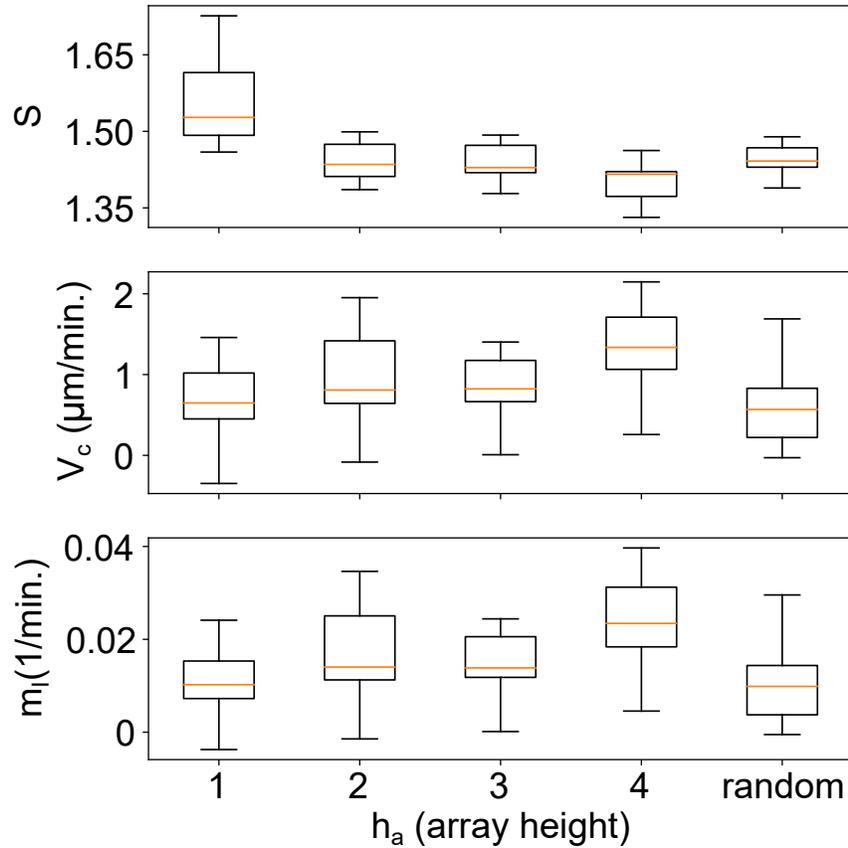


Figure S11: Effect of initial cluster configuration on collective migration of 16 cells in a  $w = 4$  corridor. The three panels show the average cell separation  $S$ , centroid speed  $V_c$  and migration intensity  $m_I$  as functions of the initial configuration,  $h_a$  being the initial array height for the regular configurations (see Fig. S10). The initially more elongated arrays tend to have lower migration intensity. Two factors may have contributed to this trend. First, more elongated arrays have fewer cells in contact with the left wall at the start, and thus receive a weaker initial boost to start the rightward migration relative to the arrangements in Fig. S7. This is the initial bias that POP will maintain over a longer time scale. Second, because of the initial gaps above and below the cells, the lack of CIL with the top and bottom walls weakens the lateral interaction among the cells, at least during an initial transient before the cluster attains a “natural aspect ratio” in time. The random clusters have the lowest mean migration intensity due to weak initial interaction with the vertical boundary. Overall, the effect of initial configuration is consistent with our understanding of the confinement effect.

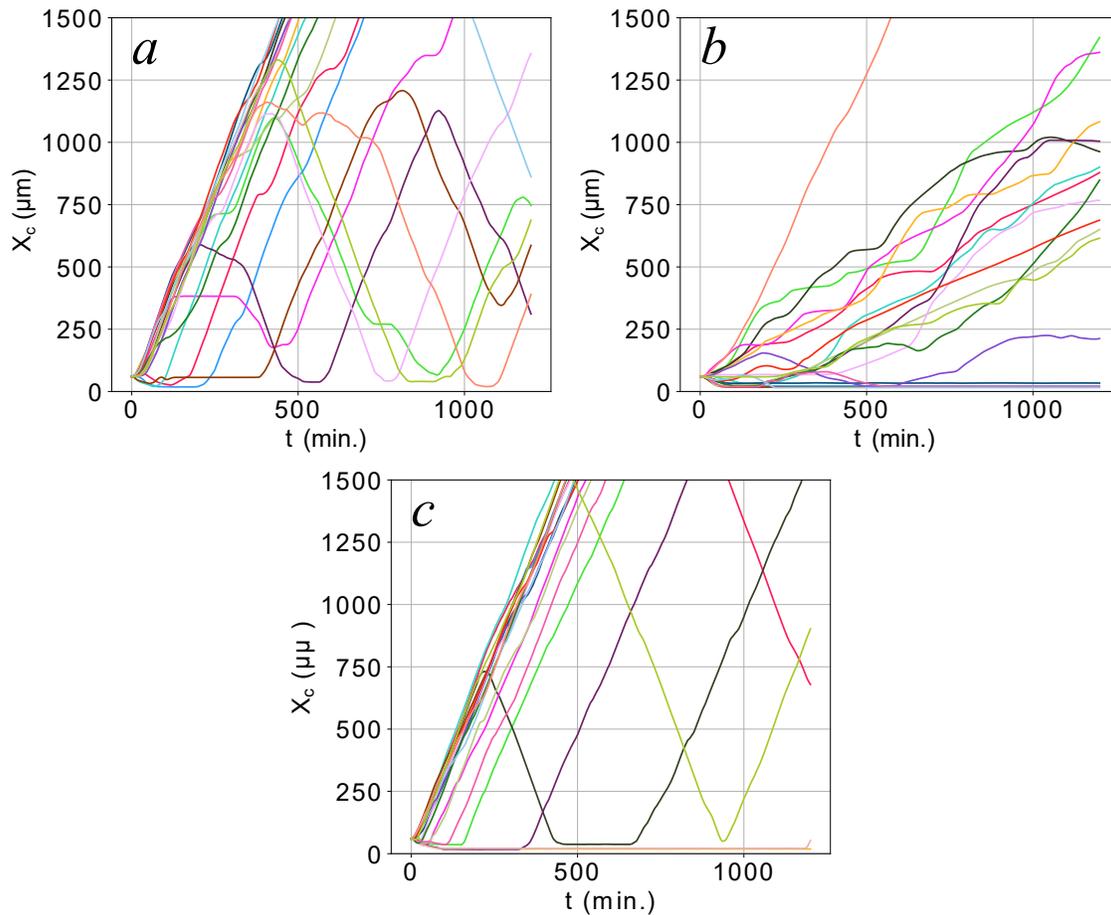


Figure S12: Parameter sensitivity study for a group of 4 cells migrating in a  $w = 1$  corridor. (a) Twenty trajectories under the standard set of conditions of Table 2. (b) The “hypo-set” with the Rac1 and RhoA activation and deactivation rates reduced by a factor of 10 from the standard set of parameters, and the parameter  $M_{COA}$  multiplied by a factor of 2. (c) The “hyper-set” with the Rac1 and RhoA activation and deactivation rates elevated by a factor of 10 from the standard set of parameters, and the parameter  $M_{COA}$  multiplied by a factor of 0.75.

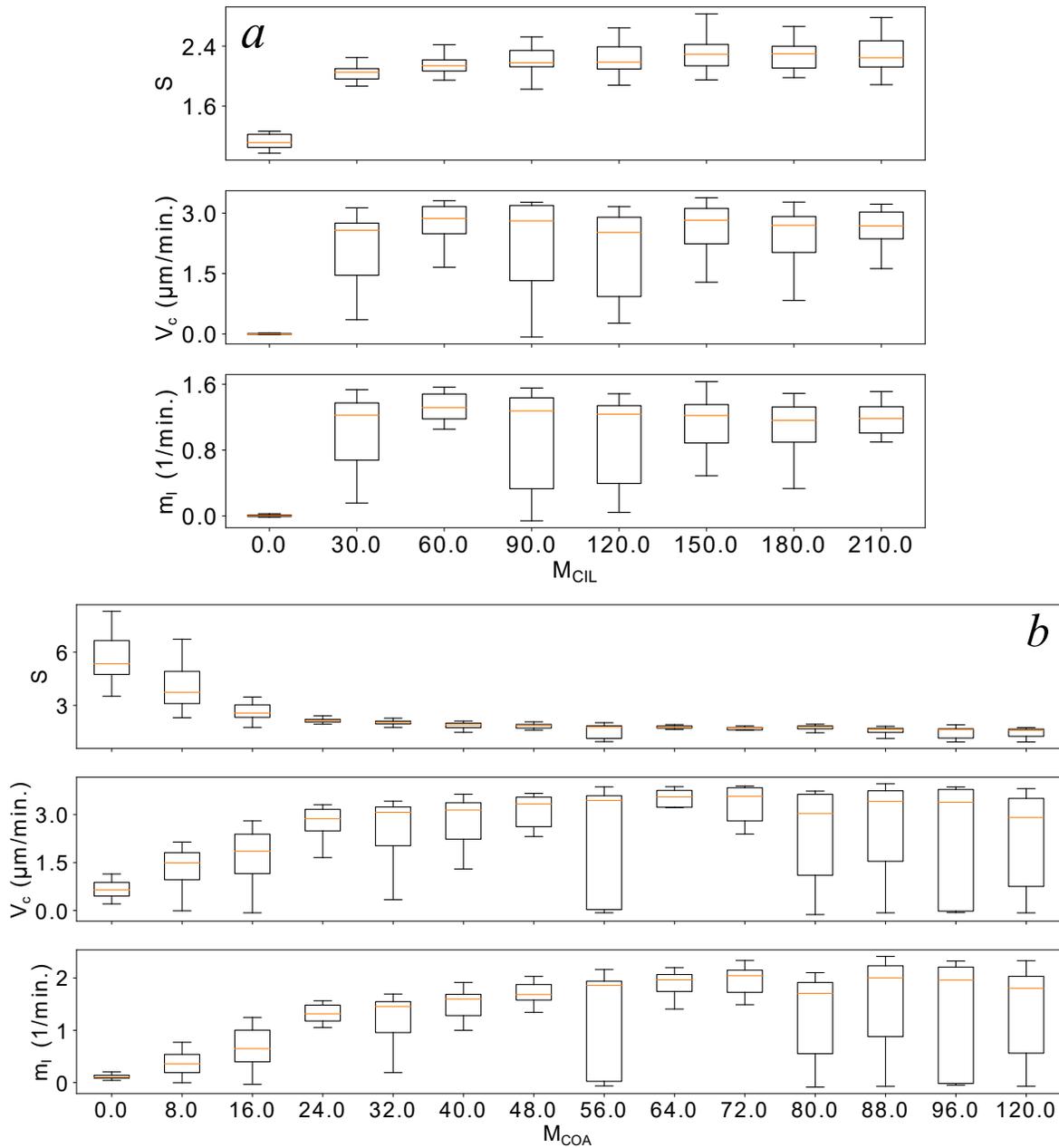


Figure S13: Sensitivity of collective migration of a group of 4 cells in a  $w = 1$  corridor to  $M_{CIL}$  and  $M_{COA}$ . The standard set of parameters has  $M_{CIL} = 60$  and  $M_{COA} = 24$ . (a) The average cell separation  $S$ , centroid speed  $V_c$  and migration intensity  $m_I$  as functions of  $M_{CIL}$ . (b)  $S$ ,  $V_c$  and  $m_I$  as functions of  $M_{COA}$ . All other parameters are fixed at their standard values.

## References

- [1] Y. Mori, A. Jilkinе, L. Edelstein-Keshet, Wave-pinning and cell polarity from a bistable reaction-diffusion system, *Biophys. J.* 94 (9) (2008) 3684–3697.
- [2] W. R. Holmes, L. Edelstein-Keshet, Analysis of a minimal Rho-GTPase circuit regulating cell shape, *Phys. Biol.* 13 (4) (2016) 046001.
- [3] A. Jilkinе, A. F. Marée, L. Edelstein-Keshet, Mathematical model for spatial segregation of the Rho-family GTPases based on inhibitory crosstalk, *Bull. Math. Biol.* 69 (6) (2007) 1943–1978.
- [4] M. L. Woods, C. Carmona-Fontaine, C. P. Barnes, I. D. Couzin, R. Mayor, K. M. Page, Directional collective cell migration emerges as a property of cell interactions, *PloS One* 9 (9) (2014) e104969.
- [5] A. R. Houk, A. Jilkinе, C. O. Mejean, R. Boltyanskiy, E. R. Dufresne, S. B. Angenent, S. J. Altschuler, L. F. Wu, O. D. Weiner, Membrane tension maintains cell polarity by confining signals to the leading edge during neutrophil migration, *Cell* 148 (1) (2012) 175–188.
- [6] M. Prass, K. Jacobson, A. Mogilner, M. Radmacher, Direct measurement of the lamellipodial protrusive force in a migrating cell, *J. Cell Biol.* 174 (6) (2006) 767–772.
- [7] E. Theveneau, L. Marchant, S. Kuriyama, M. Gull, B. Moepps, M. Parsons, R. Mayor, Collective chemotaxis requires contact-dependent cell polarity, *Dev. Cell* 19 (1) (2010) 39–53.
- [8] G. Forgacs, R. A. Foty, Y. Shafrir, M. S. Steinberg, Viscoelastic properties of living embryonic tissues: a quantitative study, *Biophys. J.* 74 (5) (1998) 2227–2234.
- [9] T. Bittig, O. Wartlick, A. Kicheva, M. González-Gaitán, F. Jülicher, Dynamics of anisotropic tissue growth, *New J. Phys.* 10 (6) (2008) 063001.
- [10] J. Solon, A. Kaya-Copur, J. Colombelli, D. Brunner, Pulsed forces timed by a ratchet-like mechanism drive directed tissue movement during dorsal closure, *Cell* 137 (7) (2009) 1331–1342.
- [11] T. G. Kuznetsova, M. N. Starodubtseva, N. I. Yegorenkov, S. A. Chizhik, R. I. Zhdanov, Atomic force microscopy probing of cell elasticity, *Micron* 38 (8) (2007) 824–833.
- [12] J. Solon, I. Levental, K. Sengupta, P. C. Georges, P. A. Janmey, Fibroblast adaptation and stiffness matching to soft elastic substrates, *Biophys. J.* 93 (12) (2007) 4453–4461.
- [13] K. Moissoglu, B. M. Slepchenko, N. Meller, A. F. Horwitz, M. A. Schwartz, In vivo dynamics of Rac-membrane interactions, *Mol. Biol. Cell* 17 (6) (2006) 2770–2779.
- [14] J. L. Johnson, J. W. Erickson, R. A. Cerione, New insights into how the Rho guanine nucleotide dissociation inhibitor regulates the interaction of Cdc42 with membranes, *J. Biol. Chem.* 284 (35) (2009) 23860–23871.

- [15] B. Huang, M. Lu, M. K. Jolly, I. Tsarfaty, J. Onuchic, E. Ben-Jacob, The three-way switch operation of Rac1/RhoA GTPase-based circuit controlling amoeboid-hybrid-mesenchymal transition, *Sci. Rep.* 4 (2014) 6449.
- [16] W. R. Holmes, B. Lin, A. Levchenko, L. Edelstein-Keshet, Modelling cell polarization driven by synthetic spatially graded Rac activation, *PLoS Comput. Biol.* 8 (6) (2012) e1002366.
- [17] D. Michaelson, J. Silletti, G. Murphy, P. D’Eustachio, M. Rush, M. R. Philips, Differential localization of Rho GTPases in live cells regulation by hypervariable regions and RhoGDI binding, *J. Cell Biol.* 152 (1) (2001) 111–126.
- [18] S. Das, T. Yin, Q. Yang, J. Zhang, Y. I. Wu, J. Yu, Single-molecule tracking of small GTPase Rac1 uncovers spatial regulation of membrane translocation and mechanism for polarized signaling, *Proc. Natl. Acad. Sci. U.S.A.* 112 (3) (2015) E267–E276.
- [19] R. Krall, J. Sun, K. J. Pederson, J. T. Barbieri, In vivo Rho GTPase-activating protein activity of *pseudomonas aeruginosa* cytotoxin ExoS, *Infect. Immun.* 70 (1) (2002) 360–367.
- [20] M. P. Neilson, J. A. Mackenzie, S. D. Webb, R. H. Insall, Modelling cell movement and chemotaxis using pseudopod-based feedback, *SIAM J. Sci. Comput.* 33 (3) (2011) 1035–1057.
- [21] E. Scarpa, A. Roycroft, E. Theveneau, E. Terriac, M. Piel, R. Mayor, A novel method to study contact inhibition of locomotion using micropatterned substrates, *Biol. Open* 2 (9) (2013) 901–906.