Science Advances

AAAS

Supplementary Materials for

A statistical inference approach to reconstruct intercellular interactions in cell migration experiments

Elena Agliari, Pablo J. Sáez, Adriano Barra, Matthieu Piel, Pablo Vargas, Michele Castellana*

*Corresponding author. Email: michele.castellana@curie.fr

This PDF file includes:

Section S1. ME models

S1.1 Maximum-entropy methods with equality constraints

In this Section, we will review maximum-entropy models with equality constraints [33], which rely on the hypothesis that the experimental data provide exact estimates for the correlation and polarization. In particular, here we assume that the nominal cell positions at different times are all different, i.e., all directions of motion $\mathbf{s}_i(t)$ are well defined. Details on the derivation are discussed in Section S1.2, where we present the extension to inequality constraints.

We denote the exact experimental estimate for the correlation and polarization by

$$\langle C(\mathbf{S}) \rangle \;=\; C_{\text{ex}} \tag{S1}$$

$$\langle M(\mathbf{S}) \rangle \;=\; M_{\text{ex}} \tag{S2}$$

where $\langle\rangle$ is the average over a set of temporal snapshots of the population, and we estimate the minimal probability distribution of the velocities with the ME principle given by Eqs. (4), (5) (6)

$$\max_P S[P] \tag{S3}$$

subject to

$$\langle C(\mathbf{S}) \rangle_P = C_{\text{ex}} \tag{S4}$$

$$\langle M(\mathbf{S}) \rangle_P = M_{\text{ex}} \tag{S5}$$

$$\int d\mathbf{S} P(\mathbf{S}) = 1 \tag{S6}$$

where the entropy S is given by Eq. (3) and the observable x by the directions of motion $\mathbf{S}$. In the equations above, $\int d\mathbf{S} \equiv \int d\mathbf{s}_1 \cdots d\mathbf{s}_N$ denotes the integral over the directions $\mathbf{s}_i$ given by Eq. (8), where $\int d\mathbf{s}_i \equiv \int_0^{2\pi} d\theta_i$ and $\langle \cdot \rangle_P \equiv \int d\mathbf{S} P(\mathbf{S}) \cdot$ is the average with respect to the model distribution P. The solution of the optimization problem (S3)-(S6) is given by Eq. (12), where the partition function reads

$$Z = \int d\mathbf{S} \exp \{N[JC(\mathbf{S}) + \boldsymbol{H} \cdot \boldsymbol{M}(\mathbf{S})]\} \tag{S7}$$

The average correlation and polarization can be obtained explicitly for any given J and $\boldsymbol{H}$. Setting

$$\kappa_\pm \equiv \sqrt{\frac{\pm 2J}{N-1}} \tag{S8}$$

for $J > 0$, the result is

$$\langle C(\mathbf{S}) \rangle_P \;=\; -\frac{1}{N-1} + \frac{(N-1)|\boldsymbol{H}|^2}{4NJ^2}$$

$$+ \frac{\displaystyle\int_0^\infty r\,dr\,e^{-NS_+(r)} \left[\frac{\mathrm{I}_1(\sqrt{2J}r)r}{\sqrt{2J}\mathrm{I}_0(\sqrt{2J}r)} \mathrm{I}_0\left(\frac{\sqrt{N-1}}{\kappa_+}r\,|\boldsymbol{H}| \right) - \mathrm{I}_1\left(\frac{\sqrt{N-1}}{\kappa_+}r\,|\boldsymbol{H}| \right) \frac{(N-1)r\,|\boldsymbol{H}|}{N(2J)^{3/2}} \right]}{\displaystyle\int_0^\infty r\,dr\,e^{-NS_+(r)}\mathrm{I}_0\left(\frac{\sqrt{N-1}}{\kappa_+}r\,|\boldsymbol{H}| \right)} \tag{S9}$$

$$\langle M(\mathbf{S}) \rangle_P \;=\; -\frac{(N-1)\boldsymbol{H}}{2NJ} + \frac{\displaystyle\int_0^\infty r\,dr\,e^{-NS_+(r)} \frac{(N-1)r\boldsymbol{H}}{N\sqrt{2J}\,|\boldsymbol{H}|} \mathrm{I}_1\left(\frac{\sqrt{N-1}}{\kappa_+}r\,|\boldsymbol{H}| \right)}{\displaystyle\int_0^\infty r\,dr\,e^{-NS_+(r)}\mathrm{I}_0\left(\frac{\sqrt{N-1}}{\kappa_+}r\,|\boldsymbol{H}| \right)} \tag{S10}$$

where

$$S_+(r) \equiv \frac{N-1}{2N}r^2 - \log[\mathrm{I}_0(\sqrt{N-1}\kappa_+ r)] \tag{S11}$$

and I_n is the modified Bessel function of the first kind [34].

For $J < 0$, we obtain

$$\langle C(\mathbf{S})\rangle_P = -\frac{1}{N-1} + \frac{(N-1)\,|\mathbf{H}|^2}{4NJ^2} \tag{S12}$$

$$+ \frac{\int_0^\infty rdr e^{-NS_-(r)} \left[\frac{J_1(\sqrt{-2J}r)r}{\sqrt{-2J}J_0(\sqrt{-2J}r)} J_0\left(\frac{\sqrt{N-1}}{\kappa_-}r\,|\mathbf{H}|\right) - J_1\left(\frac{\sqrt{N-1}}{\kappa_-}r\,|\mathbf{H}|\right) \frac{(N-1)r\,|\mathbf{H}|}{N(-2J)^{3/2}} \right]}{\int_0^\infty rdr e^{-NS_-(r)}J_0\left(\frac{\sqrt{N-1}}{\kappa_-}r\,|\mathbf{H}|\right)}$$

$$\langle M(\mathbf{S})\rangle_P = -\frac{(N-1)\mathbf{H}}{2NJ} - \frac{\int_0^\infty rdr e^{-NS_-(r)} \frac{(N-1)r\mathbf{H}}{N\sqrt{-2J}\,|\mathbf{H}|} J_1\left(\frac{\sqrt{N-1}}{\kappa_-}r\,|\mathbf{H}|\right)}{\int_0^\infty rdr e^{-NS_-(r)}J_0\left(\frac{\sqrt{N-1}}{\kappa_-}r\,|\mathbf{H}|\right)} \tag{S13}$$

where

$$S_-(r) \equiv \frac{N-1}{2N}r^2 - \log[J_0(\sqrt{N-1}\kappa_-r)] \tag{S14}$$

and J_n is the Bessel function of the first kind.

Finally, for $J = 0$ the velocities behave as independent variables, and the correlation factors out as a product of polarizations:

$$\langle C(\mathbf{S})\rangle_P = |\langle M(\mathbf{S})\rangle_P|^2 \tag{S15}$$

$$\langle M(\mathbf{S})\rangle_P = \frac{I_1(|\mathbf{H}|)\mathbf{H}}{I_0(|\mathbf{H}|)\,|\mathbf{H}|} \tag{S16}$$

S1.2   Maximum-entropy methods with bound constraints

In this Section we discuss maximum-entropy models with bound constraints [21, 22], which apply to a more general scenario where the experimental data do not provide exact values for the correlation and polarization, but their lower and upper bounds

$$C_{\text{ex}}^{\min} \leq \langle C(\mathbf{S})\rangle \leq C_{\text{ex}}^{\max} \tag{S17}$$

$$M_{\text{ex}}^{\min} \leq \langle M(\mathbf{S})\rangle \leq M_{\text{ex}}^{\max} \tag{S18}$$

We construct the minimal probability distribution of the velocities with the maximum-entropy principle

$$\max_P S[P] \tag{S19}$$

subject to

$$C_{\text{ex}}^{\min} \leq \langle C(\mathbf{S})\rangle_P \leq C_{\text{ex}}^{\max} \tag{S20}$$

$$M_{\text{ex}}^{\min} \leq \langle M(\mathbf{S})\rangle_P \leq M_{\text{ex}}^{\max} \tag{S21}$$

$$\int d\mathbf{S}P(\mathbf{S}) = 1 \tag{S22}$$

The optimization problem (S19)-(S22) involves both equality and inequality constraints, and can be solved by means of the KKT conditions [24, 25]. To achieve this, we introduce the auxiliary function

$$L \equiv S[P] - N\Bigg[J_-(C_{\text{ex}}^{\min} - \langle C(\mathbf{S})\rangle_P) + J_+(\langle C(\mathbf{S})\rangle_P - C_{\text{ex}}^{\max}) $$
$$+ \mathbf{H}_- \cdot (M_{\text{ex}}^{\min} - \langle M(\mathbf{S})\rangle_P) + \mathbf{H}_+ \cdot (\langle M(\mathbf{S})\rangle_P - M_{\text{ex}}^{\max}) + \nu\left(\int d\mathbf{S}P(\mathbf{S}) - 1\right) \Bigg] \tag{S23}$$

where $J_\pm$, $H^\pm$ and ν are the Lagrange multipliers related to the inequality constraints (S20) and (S21) and to the equality constraint (S22), respectively. Note that we defined the multipliers by including the additional factor N in such a way that these are of order unity for large N. The KKT conditions are given by the stationarity condition for the auxiliary function

$$0 = \frac{\partial L}{\partial P(\mathbf{S})} = -\log P(\mathbf{S}) - 1 - N[JC(\mathbf{S}) + \mathbf{H} \cdot \mathbf{M}(\mathbf{S}) + \nu] \tag{S24}$$

where we have set

$$J \quad \equiv \quad J_+ - J_- \tag{S25}$$
$$\mathbf{H} \quad \equiv \quad \mathbf{H}_+ - \mathbf{H}_- \tag{S26}$$

In addition, the optimum must satisfy the primal-feasibility conditions (S20), (S21) and (S22), the dual-feasibility conditions

$$\begin{aligned} J_\pm &\geq 0, \\ H_\alpha^\pm &\geq 0, \quad \alpha = x, y \end{aligned} \tag{S27}$$

and the complementary-slackness conditions

$$\begin{aligned} \pm J_\pm[\langle C(\mathbf{S})\rangle_P - C_{\text{ex}}^{\text{max(min)}}] &= 0, \\ \pm H_\alpha^\pm[\langle M_\alpha(\mathbf{S})\rangle_P - M_\alpha^{\text{max(min)}}] &= 0 \end{aligned} \tag{S28}$$

where, in what follows, the identities that contain a $\pm$ sign denote two distinct equations, one for each value of the sign, and the index α runs over the two vector components. By solving Eq. (S24) for P and using the normalization condition (S22), we obtain the explicit form (12) for the distribution of normalized velocities, where the partition function is given by Eq. (S7).

S1.2.1 Partition function

We will now compute the partition function, so as to obtain an explicit expression for the averages $\langle\rangle_P$ that appear in the KKT conditions. First, we rewrite Z as a function of the sum of the directions of motion as

$$Z = \exp\left(-\frac{NJ}{N-1}\right) \int d\mathbf{S} \prod_\alpha \exp\left[\frac{J}{N-1}\left(\sum_i \mathbf{s}_{i\alpha}\right)^2 + H_\alpha \sum_i \mathbf{s}_{i\alpha}\right] \tag{S29}$$

Let us consider the case $J > 0$ first, and rewrite the square in the exponential of Eq. (S29) in terms of a Gaussian integral [35]

$$Z = \exp\left(-\frac{NJ}{N-1}\right) \int \frac{d\boldsymbol{x}}{2\pi} \exp\left[-\frac{1}{2}\left(\boldsymbol{x} - \frac{\mathbf{H}}{\kappa_+}\right)^2\right][f(\boldsymbol{x})]^N \tag{S30}$$

where $\boldsymbol{x} = (x_1, x_2)$, $\kappa_\pm$ is given by Eq. (S8), we set

$$f(\boldsymbol{x}) \equiv \int_0^{2\pi} d\theta \exp[\kappa_+(x_1 \cos\theta + x_2 \sin\theta)] = 2\pi I_0(\kappa_+|\boldsymbol{x}|) \tag{S31}$$

and I_n is the modified Bessel function of the first kind [34]. By using polar coordinates $\boldsymbol{x} = r(\cos\phi, \sin\phi)$ in the integral in Eq. (S30), we obtain

$$
\begin{aligned}
Z &= \frac{N-1}{2\pi} \exp\left(-\frac{NJ}{N-1} - \frac{|\boldsymbol{H}|^2}{2\kappa_+^2}\right) \\
&\quad \times \int_0^\infty r\,dr\,e^{-NS_+(r)} \int_0^{2\pi} d\phi \exp\left[\frac{\sqrt{N-1}}{\kappa_+}r(H_1\cos\phi + H_2\sin\phi)\right] \\
&= (2\pi)^N(N-1)\exp\left(-\frac{NJ}{N-1} - \frac{(N-1)|\boldsymbol{H}|^2}{4J}\right)\int_0^\infty r\,dr\,e^{-NS_+(r)}I_0\left(\frac{\sqrt{N-1}}{\kappa_+}r\,|\boldsymbol{H}|\right) \quad \text{(S32)}
\end{aligned}
$$

where we have defined S_+ according to Eq. (S11), and in the second line we rewrote the integral with respect to ϕ in terms of I_0, cf. Eq. (S31). In Eq. (S32), we have reduced the partition function to a simple, one-dimensional integral that can be evaluated with arbitrary precision. In addition, the average correlation and polarization can be obtained from (S32) by taking its derivatives with respect to the Lagrange multipliers

$$
\langle C(\mathbf{S})\rangle_P = \frac{1}{N}\frac{\partial \log Z}{\partial J} \tag{S33}
$$

$$
\langle \boldsymbol{M}(\mathbf{S})\rangle_P = \frac{1}{N}\frac{\partial \log Z}{\partial \boldsymbol{H}} \tag{S34}
$$

and the result is given by Eqs. (S9) and (S10).

Proceeding along the same lines for $J < 0$, we obtain Eqs. (S12) and (S13), where S_- is given by Eq. (S14), and J_n is the Bessel function of the first kind [34]. Finally, we consider the case $J = 0$, for which the directions $\mathbf{s}_i$ in $P(\mathbf{S})$ behave as independent variables, and the average correlation and polarization are given by Eqs. (S15) and (S16).

Overall, Eqs. (S9), (S10), (S12), (S13), (S15) and (S16) yield the average correlation and polarization for any value of the KKT multipliers, thus allowing us to solve the KKT conditions. In particular, the solution above for the average correlation and polarization is exact for any value of N, thus allowing us to study both small and large cell populations.

S1.2.2 Solution of Karush-Kuhn-Tucker conditions and algorithmic implementation

Hereafter we describe the strategy used to solve the KKT conditions. According to Eq. (S28), the KKT multipliers are nonnegative: It follows that for any pair of maximal and minimal bound constraints, e.g., Eq. (S20), the complementary-slackness conditions (S28) imply three possible cases:

- $J_+ > 0$, $J_- = 0$, thus Eq. (S28) implies $\langle C(\mathbf{S})\rangle_P = C_{\text{ex}}^{\max}$,

- $J_+ = 0$, $J_- > 0$, thus $\langle C(\mathbf{S})\rangle_P = C_{\text{ex}}^{\min}$,

- $J_+ = J_- = 0$,

where the case $J_+ > 0$, $J_- > 0$ would imply $\langle C(\mathbf{S})\rangle_P = C_{\text{ex}}^{\max} = C_{\text{ex}}^{\min}$, and is ruled out because we assume that $C_{\text{ex}}^{\max} > C_{\text{ex}}^{\min}$, and similarly for the other constraints. In the three cases above, each multiplier J_+, J_- may be either equal to zero or positive, and in the latter case the multiplier is determined by the condition $\langle C(\mathbf{S})\rangle_P = C_{\text{ex}}^{\max}$ or $\langle C(\mathbf{S})\rangle_P = C_{\text{ex}}^{\min}$. As a result, each case uniquely determines both multipliers.

In order to solve the full KKT conditions, we considered the three cases above for each of the three pairs of maximal and minimal bound constraints, cf. Eqs. (S20) and (S21), and obtained a total of twenty-seven cases. We solved the KKT conditions for the multipliers in each of these cases as discussed above, and checked whether the solution satisfies the complementary-slackness conditions (S28) and the primal-feasibility conditions (S20) and (S21).

First, the fulfillment of the equality conditions (S28) is assessed as follows. For each nonzero multiplier, there is an equality condition that needs to be satisfied. For example, if $J_+ > 0$, $J_- = 0$ and $\boldsymbol{H}^+ = \boldsymbol{H}^- = \boldsymbol{0}$, the only equality condition is $\langle C(\mathbf{S}) \rangle_P = C_{\text{ex}}^{\text{max}}$. We thus introduce the relative residual associated with this condition

$$\Delta_{\text{eq}} = \left| \frac{\langle C(\mathbf{S}) \rangle_P - C_{\text{ex}}^{\text{max}}}{C_{\text{ex}}^{\text{max}}} \right| \tag{S35}$$

If there are multiple equality conditions, Δ_{eq} is defined as the maximum over the residuals of all equality conditions. Second, the fulfillment of the inequality conditions (S20) and (S21) is assessed by introducing the residual

$$\Delta_{\text{ineq}} = \min \left\{ \frac{\langle C(\mathbf{S}) \rangle_P - C_{\text{ex}}^{\text{min}}}{|C_{\text{ex}}^{\text{min}}|}, \frac{C_{\text{ex}}^{\text{max}} - \langle C(\mathbf{S}) \rangle_P}{|C_{\text{ex}}^{\text{max}}|}, \cdots \right\} \tag{S36}$$

where the first two terms in braces are the relative residuals of Eq. (S20), and $\cdots$ incorporates the analog for Eqs. (S21). If, for a given case under consideration, $\Delta_{\text{eq}} = 0$ and $\Delta_{\text{ineq}} \geq 0$ within numerical precision, then the equality and inequality conditions (S28), (S20) and (S21) are satisfied, and we consider the case as admissible.

Finally, the solution of the ME problem is given by the admissible case with the largest entropy, see Figs. 4, 5 and 6.

Section S2. Tests of ME method with bound constraints

In this Section, we present the Figures relative to the test of the maximum-entropy method with bound constraints for the XY and self-propelled model, discussed in Section 2.3.
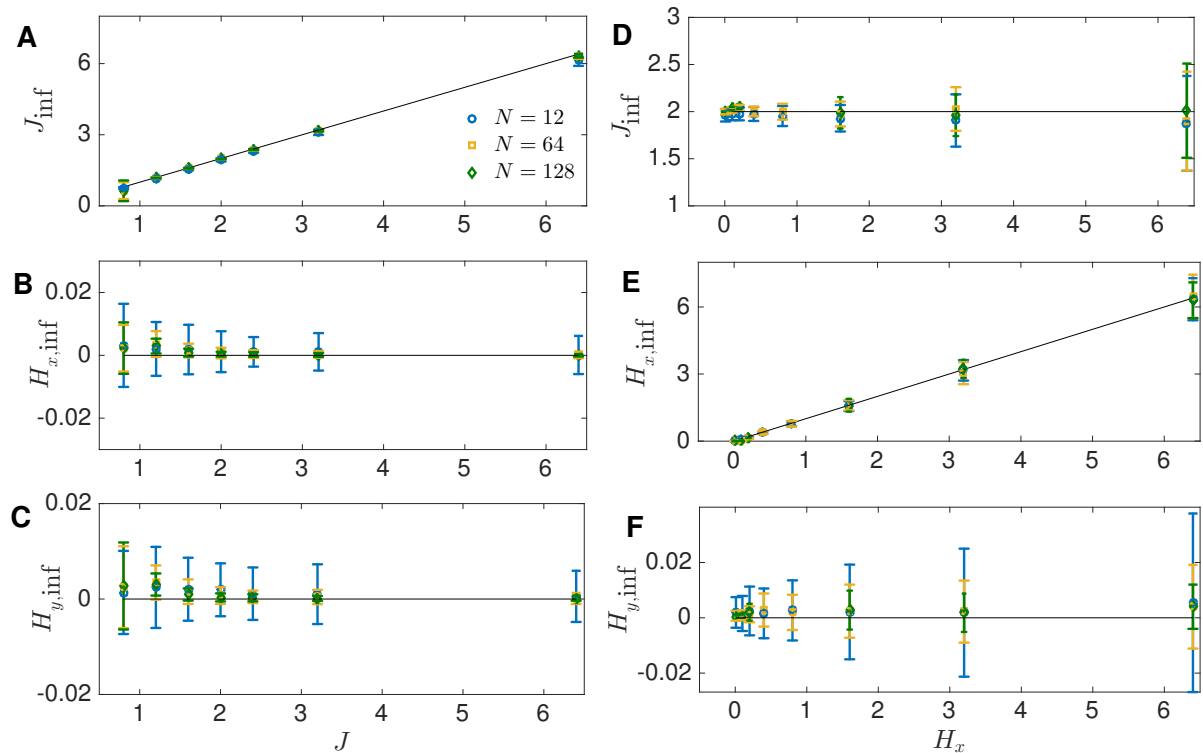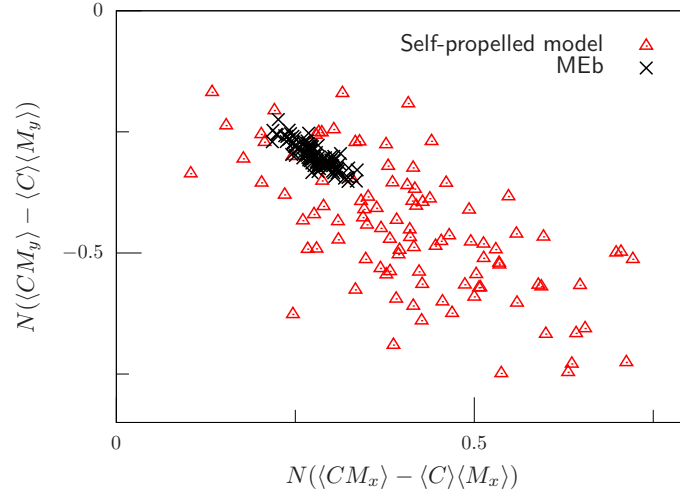
Fig. S1. Test of ME method with bound constraints on synthetic data for the XY model. In (**A**), (**B**) and (**C**) we set $H_x = H_y = 0$ and show the inferred parameters J_{inf}, $\boldsymbol{H}_{\text{inf}}$ as functions of J. In (**D**), (**E**) and (**F**) we set $J = 2$, $H_y = 0$ and show the inferred parameters as function of H_x. The mean and standard deviation of the inferred parameters are represented by circles and error bars, respectively, and solid lines represent the original values of J and $\boldsymbol{H}$ used in the simulation.
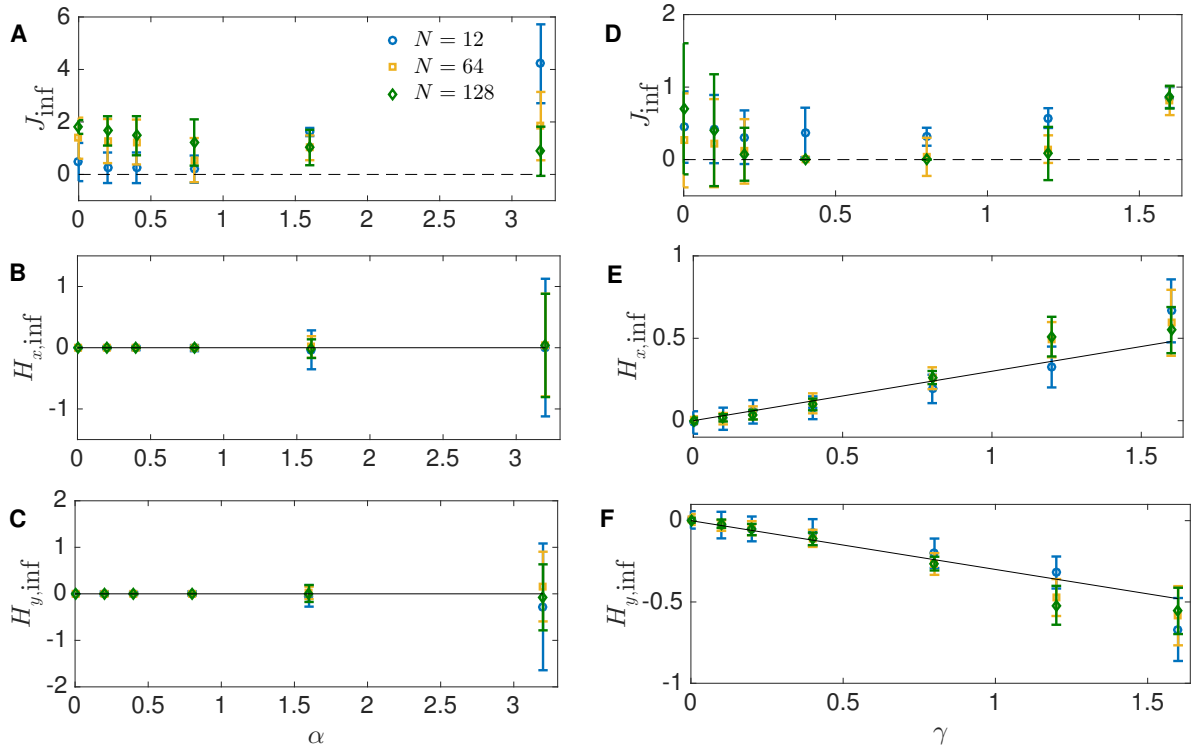
Fig. S2. Test of ME method with bound constraints on synthetic data for the SP model.
Simulations for the self-propelled model were carried on with $T = 100$ samples, and for different choices
of the parameters α and γ, with $\beta = 0.8, T_0 = 50, n_c = 9, r_a = 10, r_b = 1, r_e = 1.5$ and $r_0 = 20$. In addition, we
set $h_x = -h_y = 1/\sqrt{2}$, i.e., the field points in the direction $-\pi/4$. For each parameter configuration, we repeat the
simulation $Q = 100$ times. For each realization we evaluate averages, correlations and the related bounds, which
are then used as input for the maximum-entropy method with bound constraints. We thus obtain Q estimates of
the inferred parameters J_{inf}, $\boldsymbol{H}_{\text{inf}}$ whose mean values (symbols) and standard deviations (error bars) are depicted.
Results obtained with $\gamma = 0$ and varying α are shown in (**A**), (**B**), and (**C**), while those obtained with $\alpha = 1.2$ and
varying γ are shown in (**D**), (**E**) and (**F**). The black solid lines give the best linear fit, i.e., $y = 0$ in (**B**) and (**C**), $y = 0.3\,x$
in (**E**) and (**F**). Notice that, as the size N is varied, the width of the area where particles are placed at the beginning
of the simulation is modified in such a way that the initial particle density remains equal to 10^{-4}.

Fig. S3. Consistency test of ME method with bound constraints. Given a simulation for the self-propelled model, we rounded off the particle coordinates within a pixel size σ, chosen to emulate the positional uncertainty in the experiments. The resulting x and y components of the connected correlation function $\langle CM \rangle - \langle C \rangle \langle M \rangle$ between the motional correlation C and the polarization M are shown (red), where each point corresponds to a different simulation. Because of the uncertainty resulting from the pixel size σ, the connected correlation above varies within a confidence interval, and the values shown have been obtained as the mean between the lowest and upper bound of the interval. The connected correlation above predicted by the maximum-entropy method with bound constraints (black) is also shown, where each point corresponds to a different simulation.

Section S3. Visual comparison between experiments

In this Section, we present the visual comparison between tracks of the wound-healing and dendritic-cell experiment, Fig. S4, discussed in Section 2.4.1.
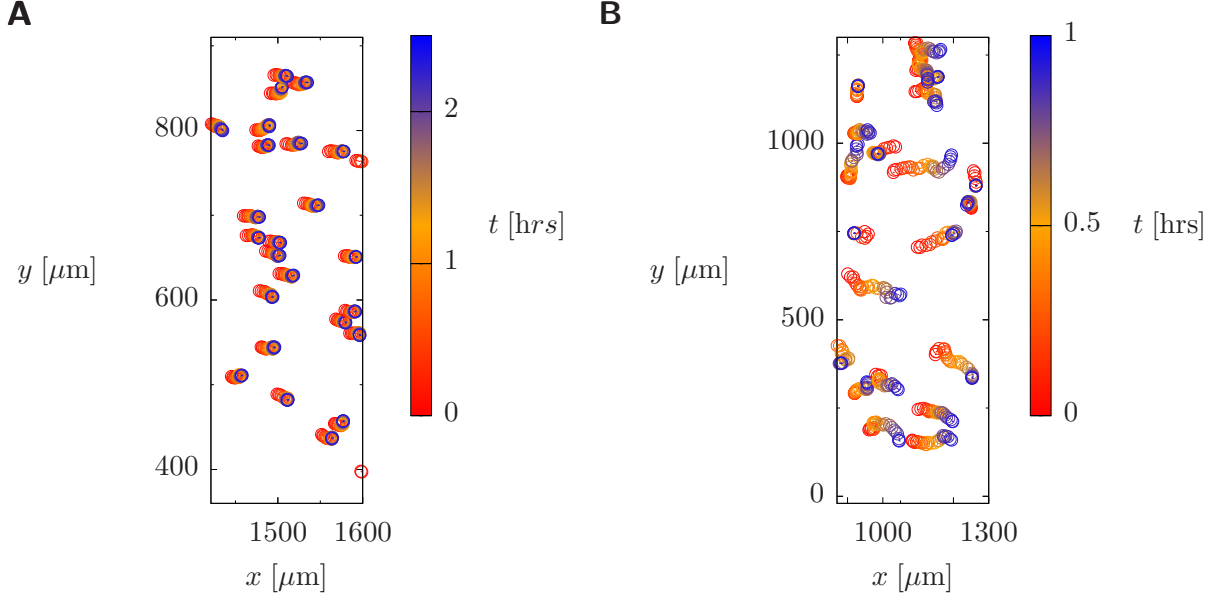
Fig. S4. Visual comparison between tracks of the wound-healing and dendritic cell experiment. (**A**) Tracked cell trajectories for the wound-healing experiment, see Fig. 2B for details. (**B**) Tracked cell trajectories for zone 2 in the dendritic-cell experiment, see Fig. 3B for details. In (**A**), only a fraction of the tracking area of Fig. 2B is shown, so as to obtain an area with the same aspect ratio as in (**B**). Similarly, only a fraction of the tracks is shown in (**A**), so as to display a number of tracks comparable to (**B**).

# Section S4. Analysis of motional data

S4.1 Wound-Healing experiment

Here we focus on the wound-healing experiment, for which the number of cells is $N = 1288$.

First, we notice that the distribution of the related tracks in the observation window is not uniform: Most tracks are found in the right side of the observation area, see Fig. 2A, and cells tend to move along the horizontal direction, see Fig. 2C. This is confirmed by the scatter plots in Fig. S5: panel (A) contains the scatter plot of $v_{x\,i}(t)$ versus $v_{y\,i}(t)$, for all $t = 1, ..., T_i$ and $i = 1, ..., N$; panel (B) contains the scatter plot of $v_i(t) = |\mathbf{v}_i(t)|$ versus $\theta_i(t)$ for all $i = 1, ..., N$ and $t = 1, ..., T_i$. From the former we see that positive velocities along the x direction are more likely to occur than negative ones, and the latter indicates that the density of data points is higher at $\theta \sim 0$.

Further, we investigate the presence of correlations among velocities and angles. To achieve this, we consider the Pearson coefficient, which, for two time sequences $f(i, t)$ and $f(j, t)$ is defined as

$$\tilde{f}_{ij} = \frac{\sum_t [f(i, t) - \langle f_i \rangle][f(j, t) - \langle f_j \rangle]}{\sqrt{\sum_t [f(i, t) - \langle f_i \rangle]^2} \sqrt{\sum_t [f(j, t) - \langle f_j \rangle]^2}} \tag{S37}$$

where $\langle f_i \rangle = \sum_t f(i, t)/T'$ and the sums run until $t = T' = \min[T(i), T(j)]$. Further, in order to assess time auto-correlations, we considered the auto-covariance correlation function, which, for the observable
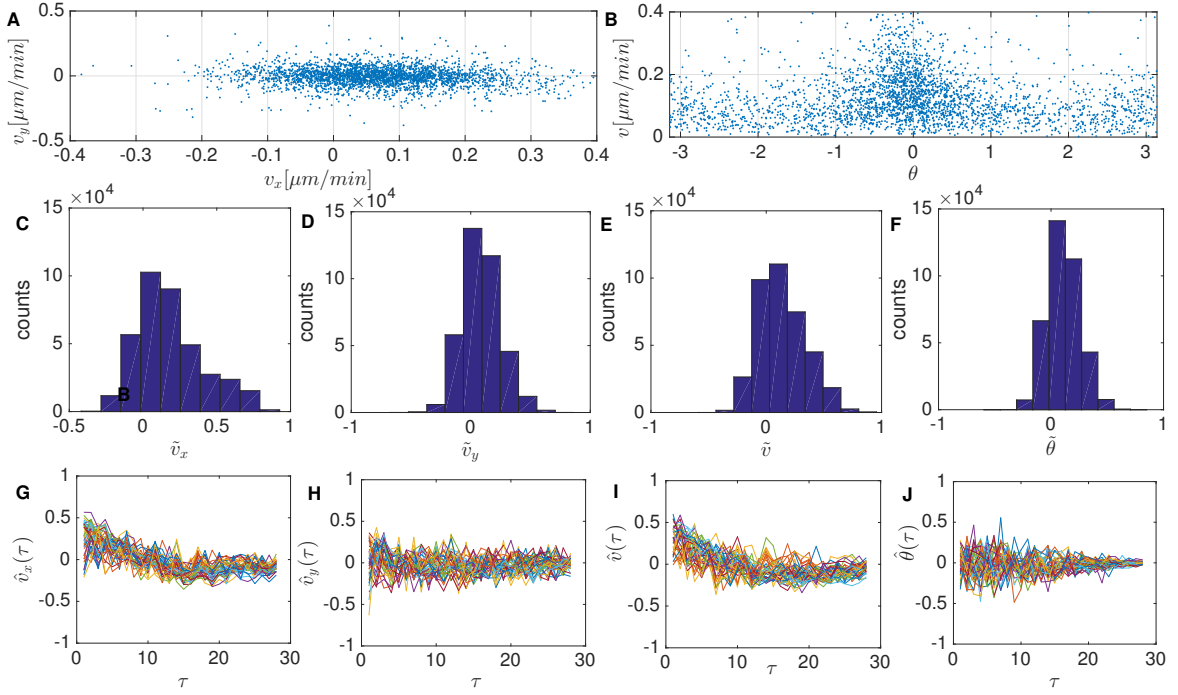
Fig. S5. Statistics of motional data for the wound-healing experiment. First row: scatter plot for the velocity components v_y vs. v_x (**A**) and for the velocity norm v vs. the directional angle θ (**B**). For the sake of clarity, in both panels only a 10% fraction of all available data points is shown. Second row: histograms for the pairwise correlation coefficient between the x-component of the velocities (**C**), between the y-component of the velocities (**D**), between the velocity magnitude (**E**), and between the angles (**F**). Third row: autocorrelation function for the x-component of the velocity (**G**), the y-component of the velocity (**H**), the magnitude of the velocity (**I**), and the angle (**J**). The outcomes of (a 10% fraction of) all cells are shown in different colors to give a visual representation of the variability of the autocorrelation.

$f(i, t)$ is defined as

$$\hat{f}_i(\tau) = \frac{\sum_t [f(i, t + \tau) - \langle f_i \rangle][f(i, t) - \langle f_i \rangle]}{\sum_t [f(i, t) - \langle f_i \rangle]^2} \tag{S38}$$

where the sums run unti $t = T(i) - \tau$. As shown in Fig. S5C-F, pairwise correlations over the whole set of cell couples display a positive skewness for all the observables considered and the related mean ($\overline{}$), median (mdn( )) and standard deviation (std( )) turn out to be $\overline{\tilde{v}_x} = 0.183$, mdn($\tilde{v}_x$) = 0.103, std($\tilde{v}_x$) = 0.225; $\overline{\tilde{v}_y} = 0.088$, mdn($\tilde{v}_y$) = 0.034, std($\tilde{v}_y$) = 0.160; $\overline{\tilde{v}} = 0.139$, mdn($\tilde{v}$) = 0.116, std($\tilde{v}$) = 0.202; $\overline{\tilde{\theta}} = 0.109$, mdn($\tilde{\theta}$) = 0.026 , std($\tilde{\theta}$) = 0.144. Also, the auto-covariance for v_x and v tend to be positive at short times, while the auto-covariance for v_y and θ display basically no dependence on time and fluctuate around zero, see Fig. S5G-J.

Next, we combine the data from all tracks and times, to study the probability distributions of v_x, v_y and v. The related mean, median and standard deviation are $\overline{v_x} = 0.063$, mdn(v_x) = 0.060, std(v_x) = 0.112; $\overline{v_y} = -0.004$, mdn(v_y) = −0.003, std(v_y) = 0.075; $\overline{v} = 0.123$, mdn(v) = 0.107, std(v) = 0.080. The best fits for their histograms are shown in Fig. S6: The best fit for the distribution of v_x is given by a normal distribution $\mathcal{N}(x; \mu, \sigma) = e^{-(x-\mu)^2/(2\sigma^2)}/\sqrt{2\pi\sigma^2}$ with best-fit coefficient $\mu = 0.068 \pm 0.005$ and $\sigma = 0.12 \pm 0.01$
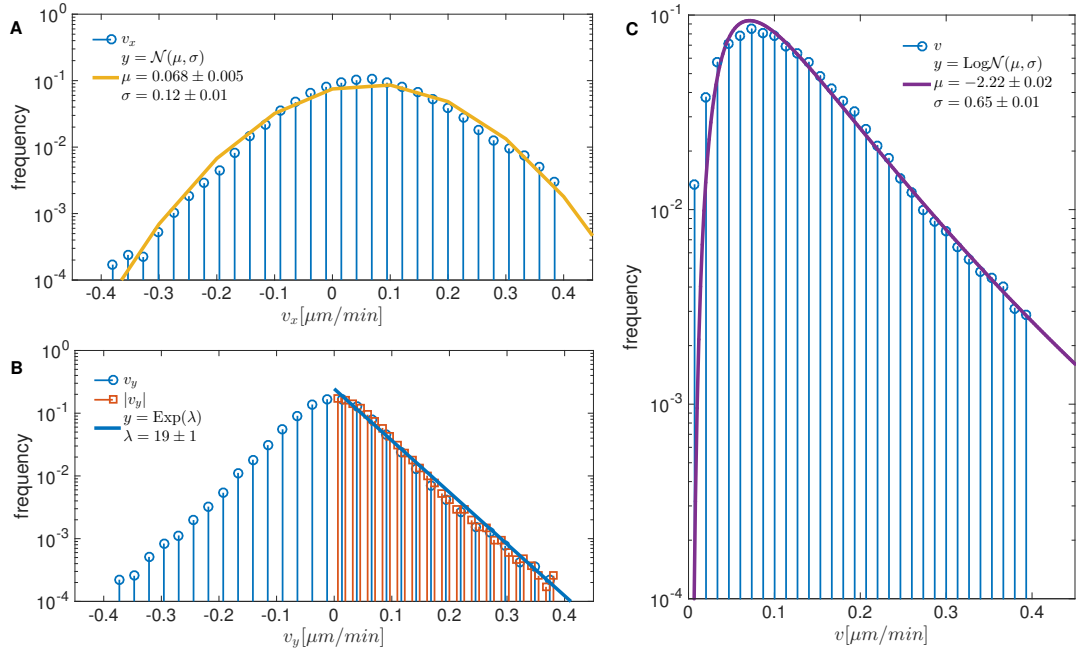
Fig. S6. Empirical distributions for cell velocities in the wound-healing experiment. Density histogram for the x- and y-component of the velocity, panels (**A**) and (**B**), respectively, and for the magnitude v (**C**) in a semi-logarithmic scale plot, with the related fit parameters. To highlight the symmetry of the distributions of v_y, in (**B**) we overlap the histogram for its absolute value $|v_y|$ to the positive branch of the original histogram; in fact, v_y follows an exponential distribution. Finally, v_x is best fitted by a normal centered at a positive value, while the velocity magnitude follows a log-normal distribution.

(panel A), while each branch of the histogram of v_y it is given by an exponential distribution $\text{Exp}(-\lambda)$ with best-fit coefficient $\lambda = 19 \pm 1$ (panel B). The distribution of the modulus of the velocity is log-normal $\text{Log}\mathcal{N}(x; \mu, \sigma) = e^{-(\ln x - \mu)^2/(2\sigma^2)}/\sqrt{2\pi\sigma^2 x^2}$ with best-fit coefficients $\mu = -2.22 \pm 0.02$ and $\sigma = 0.65 \pm 0.01$ (panel C). It is worth stressing the qualitative different distributions of v_x and v_y, and the relatively fat tail of the distribution of v.

S4.2 Dendritic-cell experiment

In what follows we focus on the dendritic-cell experiment, for which we perform the same analyses discussed for the wound-healing experiment.

S4.2.1 Free environment (zone 1)

We consider tracks in zone 1, where the cytokine concentration is negligible. As a consequence, the cellular motion is expected to be unbiased (see Fig. 3). This kind of analysis allows us to get a picture of the migratory abilities of dendritic cells in free motion, i.e., in the absence of an external stimulus.

The overall number of tracks limited to zone 1 is $N_{\text{Low}}^{(1)} = 54$ in the low-density regime and $N_{\text{High}}^{(1)} = 231$

in the high-density regime.

First, in Fig. S7 we assess the existence of correlations between the velocity components: panel (**A**) contains the scatter plot of $v_{x\,i}(t)$ versus $v_{y\,i}(t)$, for all $i = 1, ..., N_{\text{Low}}^{(1)}$ and $t = 1, ..., T_i$; panel (**B**) contains the scatter plot of $v_i(t) = |\mathbf{v}_i(t)|$ versus $\theta_i(t)$ for all $i = 1, ..., N_{\text{Low}}^{(1)}$ and $t = 1, ..., T_i$. In both cases, the data display a uniform distribution, and analogous results are obtained for the high-density case.

As far as pairwise correlations are concerned, Fig. S7C-F shows that $\tilde{v}_x, \tilde{v}_y, \tilde{v}$ and $\tilde{\theta}$ are all pretty symmetrically distributed around zero. The related mean, median and variance are $\overline{\tilde{v}_x} = 0.003$, $\text{mdn}(\tilde{v}_x) = 0.011$, $\text{std}(\tilde{v}_x) = 0.202$; $\overline{\tilde{v}_y} = 0.001$, $\text{mdn}(\tilde{v}_y) = 0.009$, $\text{std}(\tilde{v}_y) = 0.202$; $\overline{\tilde{v}} = 0.151$, $\text{mdn}(\tilde{v}) = 0.121$, $\text{std}(\tilde{v}) = 0.255$; $\overline{\tilde{\theta}} = 0.008$, $\mu(\tilde{\theta}) = 0.008$, $\text{std}(\tilde{\theta}) = 0.207$. Also, the auto-covariance for all these observables (panels G-J) display basically no dependence on time and fluctuate around zero.

These results are robust with respect to the density of dendritic cells—the same conclusions are drawn in both the high- and low-density regimes—and suggest that we can look at tracks as Markov chains where at each time step we extract randomly the two components of $\mathbf{v}$ (and similarly, in a polar system, for v and θ). The indices of central position and dispersion are $\overline{v_x} = 0.030$, $\text{mdn}(v_x) = 0.045$, $\text{std}(v_x) = 1.807$; $\overline{v_y} = 0.046$, $\text{mdn}(v_y) = 0.055$, $\text{std}(v_y) = 1.863$; $\overline{v} = 2.132$, $\text{mdn}(v) = 1.848$, $\text{std}(v) = 1.480$. The probability distributions of velocities are shown in Fig. S8: the best fit for $|v_x|$ (panel A) and $|v_y|$ (panel B) is given by an exponential distribution $\text{Exp}(x; \lambda) = \lambda e^{-\lambda x}$ for which a least-square fit yields compatible values of lambda $\lambda = 0.80 \pm 0.04$ for both components. The analogy of the behavior exhibited by the two components corroborates that the motion is fully isotropic, as evidenced by the polar histogram for θ, shown in Fig. 3C. As for the magnitude of the velocity (panel C), the best fitting function is provided by a Gumbel distribution $\text{Gumbel}(x; \alpha, \beta) = \frac{1}{\beta} e^{-z + e^{-z}}$, where $z = (x - \alpha)/\beta$, with best fitting parameters $\alpha = 1.20 \pm 0.04$ and $\beta = 1.4 \pm 0.1$. Overall, the distributions shown in Fig. S8 are short-tailed, and thus expected to preserve the central-limit-theorem regularity in the large-sample limit.

S4.2.2 Perturbed environment (zone 2)

We consider tracks in zone 2, where the cytokine concentration is high. The overall number of tracks limited to zone 2 is $N_{\text{Low}}^{(3)} = 75$ in the low-density regime and $N_{\text{High}}^{(3)} = 157$ in the high-density regime. As a consequence of the cytokine gradient, the cellular motion is expected to be biased and, in fact, this appears clear even by eye inspection (see Fig. 3). In particular, there is a bias along the x direction, see the scatter plot for the velocity components v_y versus v_x in Fig. S9A. Also, it is clear from the scatter plots v versus θ (panel B) that the longest steps are typically those along the positive horizontal direction ($\theta = 0$). The pairwise correlations (panels C-F) exhibit distributions that are less symmetric than those in zone 1: in particular, $\tilde{v}_x$ and $\tilde{v}$ display a positive skewness and, overall, the indices of central position turn out to be $\overline{\tilde{v}_x} = 0.107$, $\text{mdn}(\tilde{v}_x) = 0.078$, $\text{std}(\tilde{v}_x) = 0.260$; $\overline{\tilde{v}_y} = 0.007$, $\text{mdn}(\tilde{v}_y) = 0.009$, $\text{std}(\tilde{v}_y) = 0.270$; $\overline{\tilde{v}} = 0.217$, $\text{mdn}(\tilde{v}) = 0.208$, $\text{std}(\tilde{v}) = 0.345$; $\overline{\tilde{\theta}} = 0.004$, $\text{mdn}(\tilde{\theta}) = 0.008$, $\text{std}(\tilde{\theta}) = 0.264$. Further, the auto-covariance functions for the velocity components, the velocity magnitude and the direction angle are still fluctuating around zero with poor dependence on time (panels G-J).

These results are robust with respect to cell density, and suggest that the picture of the isotropic random walk is no longer suitable to describe the motion in zone 2.

We now move to the analysis of the velocity distributions. We preliminary obtain $\overline{v_x} = 0.898$, $\text{mdn}(v_x) = 0.550$, $\text{std}(v_x) = 2.066$; $\overline{v_y} = -0.053$, $\text{mdn}(v_y) = -0.013$, $\text{std}(v_y) = 2.031$; $\overline{v} = 2.390$, $\text{mdn}(v) = 1.922$, $\text{std}(v) = 1.868$. The distributions of velocities and of angles in Fig. S10 show that, while v_y is still exponentially distributed with a mean value comparable to the one obtained in the free environment (zone 1), v_x is no longer symmetric around zero. Indeed, the distribution of v_x is still exponential for
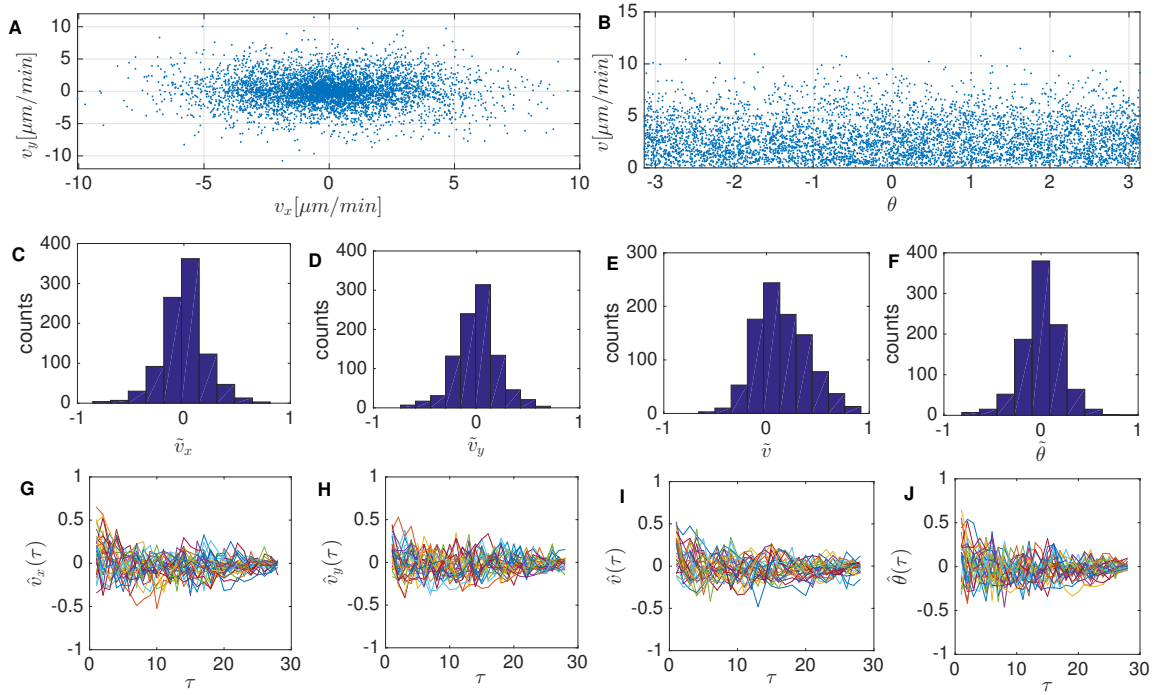
Fig. S7. Statistics of motional data for zone 1 in the dendritic cell experiment. First row: scatter plot for the velocity components v_y vs. v_x (**A**) and for the velocity norm v vs. the directional angle θ (**B**). Second row: histograms for the pairwise correlation coefficient between the x-component of the velocities (**C**), between the y-component of the velocities (**D**), between the velocity magnitude (**E**), and between the angles (**F**). Third row: autocorrelation function for the x-component of the velocity (**G**), the y-component of the velocity (h), the magnitude of the velocity (**I**), and the angle (**J**). The outcomes of all cells (depicted in different colors) are all shown to give a visual representation of the variability of the autocorrelation. For the sake of clarity in panels (**A**), (**B**) (**G**)-(**J**), the data shown in this figure correspond to the low-density case, however, the high-density case yields analogous results.

negative v_x, while it appears to be half normal for positive v_x. More precisely, the density histogram for $v_x > 0$ is fitted by $\mathcal{N}_{1/2}(x;\sigma) = \sqrt{2/(\pi\sigma^2)}e^{-x^2/(2\sigma^2)}$ with best-fit coefficient $\sigma = 2.48 \pm 0.03$. This implies a broader dispersion for step lengths along the x-direction with large, positive steps getting more likely. Moreover, the distribution of the magnitude of the velocity is now best-fitted by a half-normal $\mathcal{N}_{1/2}(x;\sigma)$ with best-fit coefficient $\sigma = 3.0 \pm 0.1$, see Fig. S10E. Thus, even in zone 2 and unlike the wound-healing experiment, the distributions for v_x, v_y and v are well-behaved distributions with short tails. This indicates that in the wound-healing experiment relatively large fluctuations are present.

Section S5. Estimate of positional uncertainty

In this Section, we describe how we estimated the uncertainty in the cell positions resulting from a finite pixel size.

For either experiment, we chose the grayscale image of a single, representative tracked cell or cell
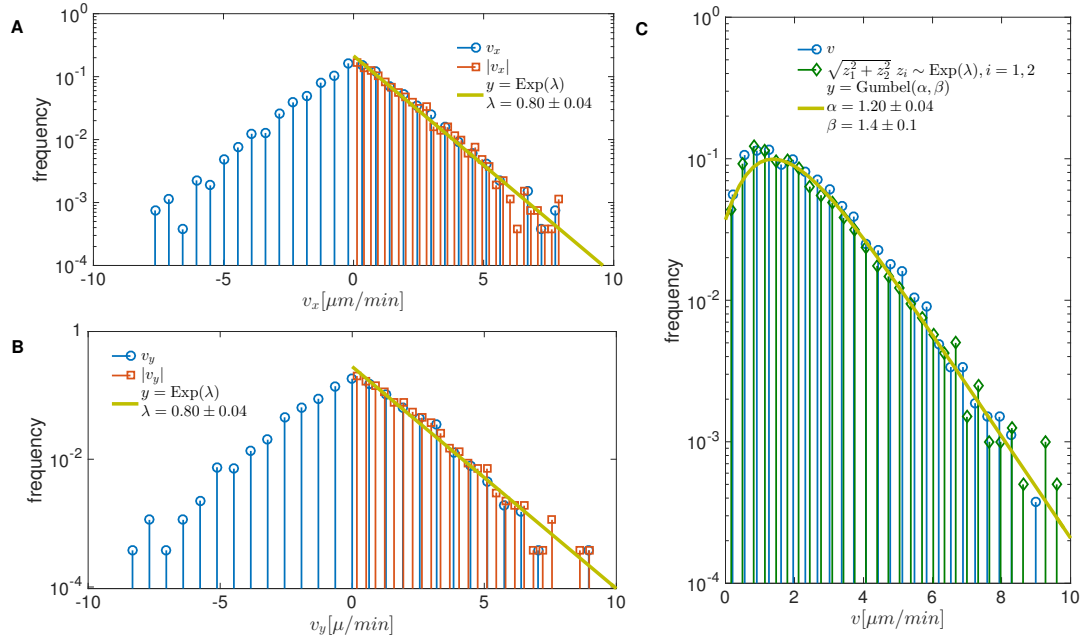
Fig. S8. Empirical distributions for cell velocities for zone 1 in the dendritic cell experiment. Density histogram for the x- and y-component of the velocity, panels (**A**) and (**B**), respectively, and for the magnitude v (**C**) in a semi-logarithmic scale plot, with the related fits (black solid line). To highlight the symmetry of the distributions of v_x and v_y, in panels (**A**) and (**B**) we show that the histogram for their absolute values $|v_x|$ and $|v_y|$ is nicely overlapped with the positive branch of the original histogram. Also, in order to corroborate the claim of an exponential distribution with mean $1/0.40 = 2.5$, in panel (**C**) we also show the histogram for the magnitude z of a two-dimensional vector $\mathbf{z}$ whose components z_1, z_2 are independent and identically distributed from the same exponential distribution $\text{Exp}(0.40)$: remarkably, the density histograms of z and v are nicely overlapped and well fitted by a Gumbel distribution.

nucleus, for the dendritic-cell and wound-healing experiment, respectively, see Fig. S11A. We then resampled the image as follows. For each pixel we consider its color intensity, C, and divide the pixel into four equal subpixels. We draw randomly the intensities c_i, $i = 1, \cdots, 4$ in each subpixel according to a multinomial distribution with event probability $p_i = 1/4$, and a total number of trials equal to $4C$. As a result, the average of the intensities across the subpixels will equal the intensity of the original pixel, i.e., $1/4 \sum_{i=1}^{4} c_i = C$. By repeating this operation for all pixels in the original image, we produce a resampled image, for which we compute the cell center $\mathbf{r}$, see Fig. S11B.

By repeating the procedure above, we obtain a population of resampled images, and the fluctuations of $\mathbf{r}$ across these samples reflect the uncertainty resulting from the finiteness of pixel size in the original image. As a result, we interpret the standard deviation σ of each component of $\mathbf{r}$ across all samples as a rough estimate of the uncertainty on the cell position resulting from the finite pixel size.

For both components, we obtain $\sigma \sim 0.15 \, \mu$m and $\sigma \sim 0.07 \, \mu$m for the cancer-cell and wound-healing experiment, respectively.
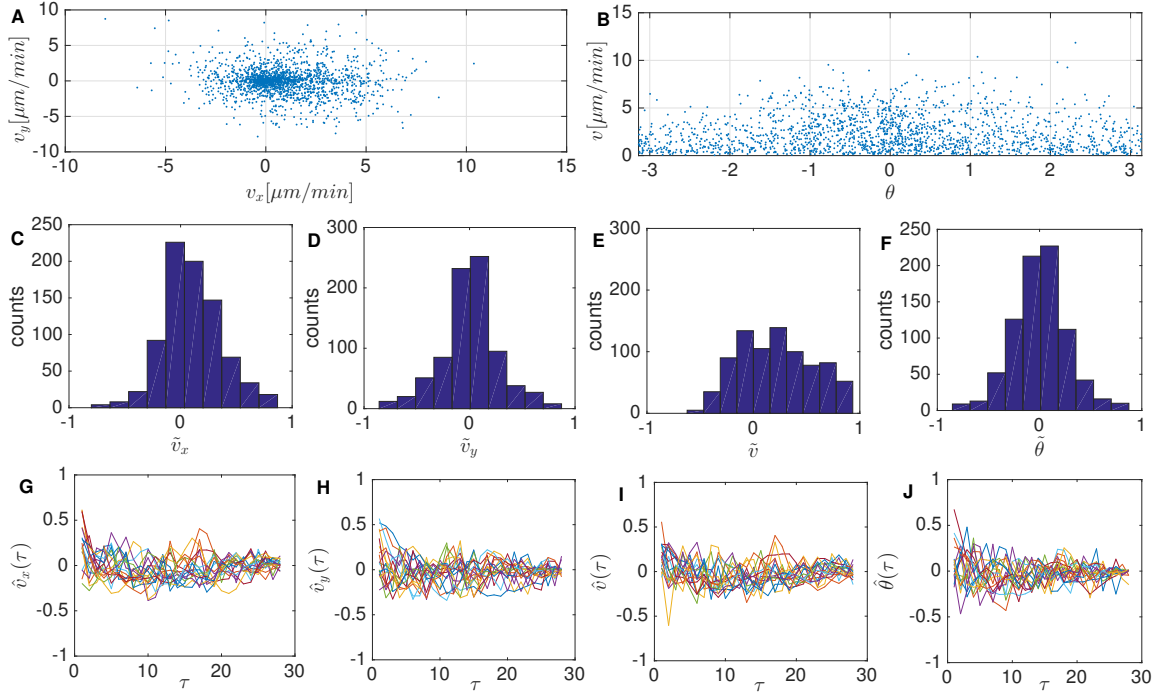
Fig. S9. Statistics of motional data for zone 2 in the dendritic cell experiment. First row: scatter plot for the velocity components v_y vs. v_x (**A**) and for the velocity norm v vs. the directional angle θ (**B**). Second row: histograms for the pairwise correlation coefficient between the x-component of the velocities (**C**), between the y-component of the velocities (**D**), between the velocity magnitude (**E**), and between the angles (**F**). Third row: autocorrelation function for the x-component of the velocity (**G**), the y-component of the velocity (**H**), the magnitude of the velocity (**I**), and the angle (**J**). The outcomes of all cells (depicted in different colors) are all shown to give a visual representation of the variability of the autocorrelation. For the sake of clarity in panels (**A**) and (**B**), the data shown corresponds in this figure to the low-density case, however, the high-density case yields analogous results.

The nominal cell positions are then obtained by rounding off the cell coordinates to σ

$$\boldsymbol{r}_i(t) \rightarrow \sigma \left\lfloor \frac{\boldsymbol{r}_i(t)}{\sigma} \right\rfloor \tag{S39}$$

in such a way that all positions that differ by less than σ are assigned the same nominal position, see Fig. 1.

The cell positions obtained with (S39) are then used to compute the directions of motion according to the velocity definition in Section 2.1 and Eq. (7), where some of these directions are, in general, not defined because of the coarsening scheme described in (S39). The resulting empirical averages are obtained with Eqs. (1), (9) and (10): for example, the average correlation reads

$$\langle C(\mathbf{S}) \rangle_{\text{ex}} = \frac{1}{N_p T} \sum_{i<j=1}^{N} \sum_{t=1}^{T} \mathbf{s}_i(t) \cdot \mathbf{s}_j(t) \tag{S40}$$

We then let all ill-defined directions of motion $\mathbf{s}_i(t)$ which enter in the right-hand side of Eq. (S40) vary between 0 and 2π, and compute the related interval in which $\langle C(\mathbf{S}) \rangle_{\text{ex}}$ varies. The lower and upper
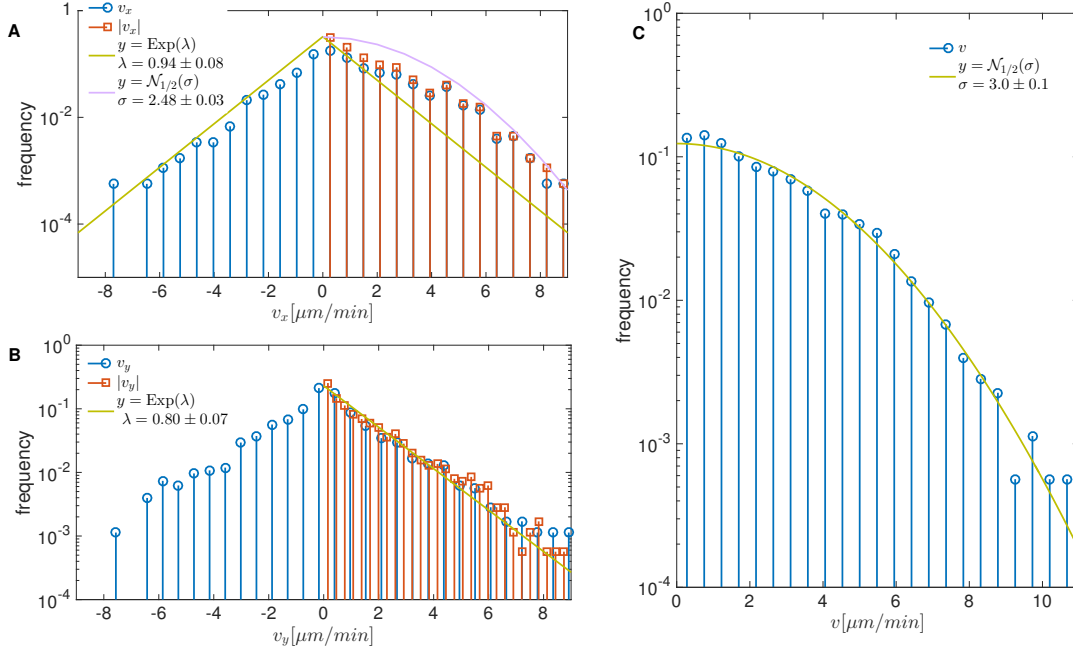
Fig. S10. Empirical distributions for cell velocities for zone 2 in the dendritic cell experiment. Density histogram for the x- and y-component of the velocity, panels (**A**) and (**B**), respectively, and for the magnitude v (**C**) in a semi-logarithmic scale plot, with the related fits (solid black line). To highlight the asymmetry and the symmetry of the distributions of, respectively, v_x and v_y, in panels (**A**) and (**B**) we overlap the histogram for their absolute values $|v_x|$ and $|v_y|$ to the positive branch of the original histogram. The comparison is fine only for v_y which follows an exponential distribution quantitatively consistent with the one obtained for zone 1. As for v_x, the exponential distribution works only for the negative branch, with a mean value slightly smaller than the one obtained in zone 1, while for the positive branch the best fit is given by a half-normal distribution. A half-normal distribution is also found to best-fit the distribution for the velocity magnitude.

bounds of this interval are the values $C_{\text{ex}}^{\min(\max)}$ that we use in the MEb method, cf. Eq. (S17). The same procedure is used to compute the bounds for each component of the polarization M. The resulting numerical values for the lower and upper bounds for the wound-healing and dendritic-cell experiment are shown in Tables S1 and S2, respectively.

A

Original

B
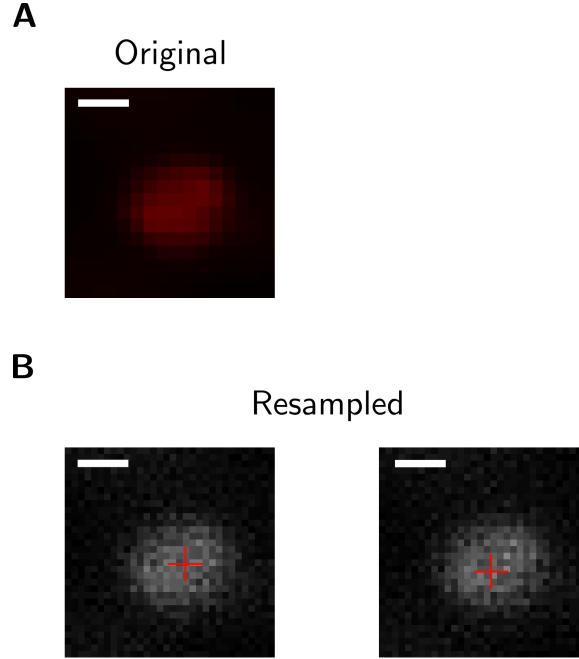
Resampled

Fig. S11. Estimate of the positional error. (**A**) Original image of a representative cell nucleus in the wound-healing experiment, cf. Fig. 3A. (**B**) Two representative images obtained by resampling the image in (**A**), where for each image the cross indicates the cell center (scale bar 5 μm).

Table S1. Confidence intervals for the empirical averages in the wound-healing experiment. Lower and upper bounds for the empirical averages of the correlation C and polarization M.

$\langle C \rangle_{min}$	0.16
$\langle C \rangle_{max}$	0.49
$\langle M \rangle_{min}$	(0.36 , -0.12)
$\langle M \rangle_{max}$	(0.54 , 0.048)

Table S2. Confidence intervals for the empirical averages in the dendritic cell experiment. Results for the high- and low-density case and the high- and low-cytokine region, i.e., zones 1 and 2, respectively, are shown. For each experimental instance, we show the lower and upper bounds for the empirical averages of the correlation C and polarization M.

Density	High		Low	
Zone	1	2	1	2
$\langle C \rangle_{min}$	-0.086	-0.018	-0.14	-0.027
$\langle C \rangle_{max}$	0.15	0.27	0.14	0.28
$\langle M \rangle_{min}$	(-0.12 , 0.0025)	(0.22 , -0.13)	(0.036 , -0.098)	(0.27 , -0.043)
$\langle M \rangle_{max}$	(0.0011 , 0.12)	(0.37 , 0.023)	(0.18 , 0.046)	(0.43 , 0.12)