Biophysical*Journal*
Article

Biophysical Society

Shannon entropy for time-varying persistence of cell migration

Yanping Liu,[1] Yang Jiao,[2,3] Qihui Fan,[4] Yu Zheng,[3] Guoqiang Li,[1] Jingru Yao,[1] Gao Wang,[1] Silong Lou,[5] Guo Chen,[1] Jianwei Shuai,[6,7,*] and Liyu Liu[1,*]
[1]Chongqing Key Laboratory of Soft Condensed Matter Physics and Smart Materials, College of Physics, Chongqing University, Chongqing, China; [2]Materials Science and Engineering and [3]Department of Physics, Arizona State University, Tempe, Arizona; [4]Beijing National Laboratory for Condensed Matter Physics and CAS Key Laboratory of Soft Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing, China; [5]Department of Neurosurgery, Chongqing University Cancer Hospital, Chongqing, China; [6]Department of Physics and Fujian Provincial Key Laboratory for Soft Functional Materials Research, Xiamen University, Xiamen 361005, China; and [7]Wenzhou Institute, University of Chinese Academy of Sciences, Wenzhou 325000, China

ABSTRACT Cell migration, which can be significantly affected by intracellular signaling pathways and extracellular matrix, plays a crucial role in many physiological and pathological processes. Cell migration is typically modeled as a persistent random walk, which depends on two critical motility parameters, i.e., migration speed and persistence time. It is generally very challenging to efficiently and accurately quantify the migration dynamics from noisy experimental data. Here, we introduce the normalized Shannon entropy (SE) based on the FPS of cellular velocity autocovariance function to quantify migration dynamics. The SE introduced here possesses a similar physical interpretation as the Gibbs entropy for thermal systems in that SE naturally reflects the degree of order or randomness of cellular migration, attaining the maximal value of unity for purely diffusive migration (i.e., SE = 1 for the most "random" dynamics) and the minimal value of 0 for purely ballistic dynamics (i.e., SE = 0 for the most "ordered" dynamics). We also find that SE is strongly correlated with the migration persistence but is less sensitive to the migration speed. Moreover, we introduce the time-varying SE based on the WPS of cellular dynamics and demonstrate its superior utility to characterize the time-dependent persistence of cell migration, which typically results from complex and time-varying intra- or extracellular mechanisms. We employ our approach to analyze experimental data of in vitro cell migration regulated by distinct intracellular and extracellular mechanisms, exhibiting a rich spectrum of dynamic characteristics. Our analysis indicates that the SE and wavelet transform (i.e., SE-based approach) offers a simple and efficient tool to quantify cell migration dynamics in complex microenvironment.

SIGNIFICANCE We introduce SE to quantify complex cellular migration dynamics, which enables us to extract the time-varying dynamics and persistence of cell migration. We also demonstrate the superior utility of the method in reflecting distinct intra- and extracellular mechanisms regulating cell migration.

INTRODUCTION

Cell migration (1) is a ubiquitous and crucial phenomenon that is found in many physiological processes, including neural system development (2), wound healing (3,4), and immunological responses (5). Eukaryotic cell migration (6) is a complex behavior involving various cellular and subcellular level events and is strictly regulated by intracellular signaling pathways (ICSPs) (7,8) and extracellular matrix (ECM) (9–13). Many human diseases are associated with ill-regulated cell migration, among which cancer metastasis is the most representative and fatal case (14,15).

To get insights into the cell migration in ECM, a number of in vitro experiments attempting to mimic various aspects of realistic in vivo environments have been designed and carried out in recent years. For instance, stiffer substrate usually increases the persistence of cell migration, whereas the soft substrate typically leads to apparently more random motions (16). This unusual behavior, which depends on the substrate stiffness, is typically referred to as "durotaxis," a mechanism that can regulate many pathological processes when combining with mechanical strains (17). Besides the substrate properties, it has been shown that nanoscale

---

Submitted November 23, 2020, and accepted for publication April 26, 2021.

*Correspondence: jianweishuai@xmu.edu.cn or lyliu@cqu.edu.cn

Yanping Liu and Yang Jiao contributed equally to this work.

Editor: Dimitrios Vavylonis.
https://doi.org/10.1016/j.bpj.2021.04.026

topographic features in ECM, coupled with the effective stiffness of cells, can guide persistent migration, which is called "topotaxis" (18). In addition, it was found that the heterogeneous topology of ECM can boost cell invasion into a three-dimensional (3D) funnel-like matrigel interface in a microfabricated biochip (19). Aligned fibers can also facilitate the migration of MDA-MB-231 breast cancer cells into rigid matrigel in a constructed collagen I-matrigel microenvironment (20).

To phenomenologically investigate and quantify the rich spectrum of cellular migratory behaviors, many motility models have been developed (21). For instance, amoeba perform a special random walk, which can increase the probability of finding a target in complex microenvironment (22). Similarly, the CD8 (+) T cell adopts a strategy known as the generalized Lévy walk, which also contributes to finding rare targets (23). Among these models, the persistent random walk (PRW) (24–26) is one of the most representative and commonly used models, which incorporates the memory of cell to the past velocities in additional to the standard Brownian motion (27) and can be derived from the following Langevin equation (28):

$$\frac{d\vec{v}}{dt} = -\frac{\vec{v}}{P} + \frac{S}{\sqrt{P}} \times \tilde{w}, \tag{1}$$

where $\vec{v}$ is the instantaneous velocity of the cell, P is the persistence time, S is the averaged migration speed and $\tilde{w}$ is the random vector derived from a Wiener process (29). Here, both of P and S are typically referred to as the "motility parameters" and together determine the overall cellular migratory capability and dynamical behaviors.

In studying cell behaviors in complex microenvironment, how to efficiently and accurately quantify cellular migratory dynamics has become a crucial issue, and significant efforts have been devoted to deriving effective estimators that can characterize cell migration behaviors (30,31). For instance, the diffusion coefficient associated with individual cells can be obtained from a time-lapse recorded trajectory (32,33). In addition to the diffusion coefficient, direction autocorrelation and other parameters are typically computed to analyze two-dimensional (2D) cell migration (e.g., via the DiPer package) (34). Moreover, motility parameters (e.g., persistence time and migration speed) can be derived from the fittings to mean-square displacement (34–36), velocity autocovariance function (VAC) (25), or FPS (FPS) (35).

Different from the aforementioned cases in which the motility parameters and cellular and environmental properties do not change with time, cell migration can be significantly affected by many factors (1) and thus exhibits time-varying migratory capability and behavior. For example, the time-dependent parameters were found and extracted when cells migrated on tissue culture-treated polystyrene surfaces and untreated polystyrene plates (37). In such cases, the time-dependent motility parameters (e.g., persistence) can be derived from time-lapse recorded trajectories with a Bayesian method (38) to better describe the nature of the migratory dynamics.

Although the aforementioned different types of motility parameters and their generalization to the time-dependent cases have been very successful in describing particular cellular migration dynamics, there are two major issues that need to be addressed: 1) the motility parameters are typically determined in an ad hoc fashion, by assuming that the associated physical model is sufficient to capture the underlying cellular behavior. This will fail and lead to inaccurate results in the cases in which novel cell behaviors are not described by the prescribed physical models (21); 2) the calculation of the numerical values of the motility parameters mainly involves fittings (e.g., R^2 analysis) based on noisy experimental data, which inevitably leads to numerical uncertainties that are generally difficult to quantify. Therefore, it is highly desirable to introduce model-independent metrics that can capture the intrinsic dynamics of migratory behaviors and are robust against noises.

In this work, we introduce the normalized Shannon entropy (SE) based on the FPS of cellular velocity autocovariance function to quantify migration dynamics. The SE introduced here possesses the similar physical interpretation as the Gibbs entropy for thermal systems in that SE naturally reflects the degree of order or randomness of cellular migration, attaining the maximal value of unity for purely diffusive migration (i.e., SE $=$ 1 for the most "random" dynamics) and the minimal value of 0 for purely ballistic dynamics (i.e., SE $=$ 0 for the most "ordered" dynamics). Moreover, we introduce the time-varying SE based on the WPS of cellular dynamics and demonstrate its superior utility to characterize the time-dependent persistence of cell migration, which typically results from complex and time-varying intra- or extracellular mechanisms.

To establish the framework and verify its accuracy, we first employ the time-varying PRW model to simulate cell migration with time-varying characteristics to generate synthetic testing data and compute the FPS of migratory velocities. Second, we analyze the effects of individual parameters on FPS and introduce normalized SE to characterize the persistence of cell migration, inspired by changes in power spectra. We find that SE is strongly correlated with the migration persistence but is less sensitive to the migration speed. Then, we perform the wavelet transform (WT) of migration velocities to obtain WPS, which is employed to derive the time-varying SE that can reflect the time-varying persistence. Finally, one indicator has been further defined to estimate the overall persistence of a given cell population.

We subsequently employ the SE to analyze experimental data of in vitro cell migration regulated by distinct intracellular and extracellular mechanisms exhibiting a rich spectrum of dynamic characteristics and persistence. These include capturing distinct migration persistence regulated by Arpin protein, the transition of cellular migration dynamics in confined microenvironment, and strongly correlated migration dynamics due to remodeled collagen fiber bundles. Our analysis indicates that the SE-based approach offers an efficient tool to estimate the directional persistence of migration, which may reflect the real-time effects of ICSPs and the ECM to some extent.

METHODS

Cell migration with time-varying characteristics

Usually, cell migration exhibits time-dependent characteristics because of the effects of many factors (1), among which the ECM is the most representative one. Here, the "time-dependent" means that cell migration capability (e.g., the persistence time P and migration speed S, cf. Eq. 1) will vary with time and cannot be described well by a Langevin equation with constant parameters P and S (28). Thus, it is necessary to generalize the Langevin equation with time-varying parameters, i.e., P(t) and S(t), and both of them together quantify cell migration capability. Further, the PRW model derived from the Langevin equation with constant parameters will be generalized to the time-varying persistent random walk (TPRW) model; see detail simulations in next subsection.

Considering the time-varying characteristics of cell migration, we first construct a function P(t) based on a linear variation (16,38), which is of the following form:

$$\mathrm{P(t)} \;=\; \mathrm{K_P} \times \mathrm{t} + \mathrm{P_0}, \tag{2}$$

where $\mathrm{K_P}$ is the changing rate of persistence time and $\mathrm{P_0}$ is initial value at $\mathrm{t} = 0$ min. The case of $\mathrm{K_P} = 0$ denotes that cell migration capability is a constant function of time and corresponds to the PRW model. It was recently reported that there is a correlation between the persistence time P and migration speed S, which can be well fitted by a simple exponential curve before saturating at larger speed (39), as follows:

$$\mathrm{P} \;=\; \mathrm{A}e^{\lambda \mathrm{S}}, \tag{3}$$

where A and λ are constant and are mainly determined by cell type and ECM. Thus, migration speed S(t) can be derived when the constants in Eq. 3 are determined. Note that we always set the trivial constant $\mathrm{A} = 1$ in this work.

Numerical simulations of cell migration based on the TPRW model

Using a similar procedure for the PRW model (31,40), cell migration described by the TPRW model can be performed in computer simulations. We will mainly focus on 2D migrations in the ensuing discussions, and the generalization to 3D is straightforward. In particular, the 2D position for one cell at each time step can be readily computed using the following functions:

$$\mathrm{x(t+\Delta t)} \;=\; \mathrm{x(t)} + \Delta \mathrm{x(t,\Delta t)} \tag{4}$$

and

$$\mathrm{y(t+\Delta t)} \;=\; \mathrm{y(t)} + \Delta \mathrm{y(t,\Delta t)}, \tag{5}$$

where Δx and Δy are displacements for a given time interval Δt, and they are written as

$$\Delta \mathrm{x(t,\Delta t)} \;=\; \alpha(\mathrm{t}) \times \Delta \mathrm{x(t-\Delta t,\Delta t)} + \mathrm{F(t)} \times \tilde{\mathrm{W}} \tag{6}$$

and

$$\Delta \mathrm{y(t,\Delta t)} \;=\; \alpha(\mathrm{t}) \times \Delta \mathrm{y(t-\Delta t,\Delta t)} + \mathrm{F(t)} \times \tilde{\mathrm{W}}, \tag{7}$$

where $\alpha(\mathrm{t}) = 1 - \Delta \mathrm{t/P(t)}$ shows the memory of one cell of the past velocities and $\mathrm{F(t)} = [\mathrm{S(t)}^2 \times \Delta \mathrm{t}^3/\mathrm{P(t)}]^{1/2}$ quantifies the amplitude of Gaussian white noise $\tilde{\mathrm{W}}$, which is also called intrinsic noise in cell dynamics. On the one hand, it is evident that the parameter α approaches to 1 and the F does to 0 when P tends to infinity for a given time step Δt, which indicates that cell will migrate along a fixed direction without turning. Similarly, the α approaches to 0 and the F does to $\mathrm{S(t)} \times \Delta \mathrm{t}$ when P tends to Δt, which means that the direction of cell migration cannot be predicted, corresponding to the normal Brownian motion. On the other hand, the α always lies in the closed interval [0, 1]; thus, the displacements will gradually decrease to 0 with time lapsing without taking into account the contribution of F. Thus, we can view the α and F as slow-down and speed-up factors, respectively.

After generating cell migration trajectories, we further add positioning errors to the simulated data to mimic the effects of experimental observations, as follows:

$$\widehat{\mathrm{x}}(\mathrm{t}) \;=\; \mathrm{x(t)} + \sigma_{\mathrm{pos}} \times \tilde{\mathrm{W}} \tag{8}$$

and

$$\widehat{\mathrm{y}}(\mathrm{t}) \;=\; \mathrm{y(t)} + \sigma_{\mathrm{pos}} \times \tilde{\mathrm{W}}, \tag{9}$$

where σ_{pos} is the positioning error and assigned as 0.01 μm (25) in this work.

FPS of cell migration velocities

To get insights into cell migration dynamics, we define a set of parameters $\mathrm{P_0} = 8.0$ min, $\mathrm{K_P} = 0$, $\mathrm{S_0} = 0.5$ μm/min ($\mathrm{A} = 1$ and $\lambda = 4.16$), based on Eqs. 2 and 3. Note that here the motility parameters are nonvarying with time; thus, the cell migration underlying these parameters can be described by the PRW model. Next, we follow the procedure above (cf. Eqs. 4, 5, 6, 7, 8, and 9) to simulate 200 cell migration trajectories and each trajectory contains 4800 + 1 (N + 1) frames, corresponding to a total time of T = 960 min, one of which is plotted in Fig. 1 *a*. Further, the velocity components on the *x* and *y* axes are also obtained from cell positions, based on the displacements for time step $\Delta \mathrm{t} = 0.2$ min, as shown in Fig. 1 *b*. It seems like that the velocity components keep stable with time lapsing.

In general, one can compute a variety of physical quantities from the trajectory data to characterize the cell migration dynamics, e.g., mean-square displacement, VAC, and the FPS. The motility parameters can then be derived and estimated from the fittings to the calculated quantities with the corresponding theoretical formula (34,35,40). Finally, the resulting parameters can be analyzed by statistical methods to extract the desirable properties and characteristics. In this work, we follow these procedures but without numerical fittings to quantify the persistence of cell migration. We introduce the VAC for individual cell migration velocities, which is given as

$$\mathrm{VAC}(\mathrm{t_j}) \;=\; \left\langle \vec{\mathrm{v}}_{\mathrm{i}} \times \vec{\mathrm{v}}_{\mathrm{i+j}} \right\rangle \;=\; \frac{1}{\mathrm{N-j-1}} \sum_{\mathrm{g=1}}^{\mathrm{N-j}} \left(\vec{\mathrm{v}}_{\mathrm{g}} - \frac{1}{\mathrm{N-j}} \sum_{\mathrm{h}}^{\mathrm{N-j}} \vec{\mathrm{v}}_{\mathrm{h}} \right) \times \left(\vec{\mathrm{v}}_{\mathrm{g+j}} - \frac{1}{\mathrm{N-j}} \sum_{\mathrm{h=j+1}}^{\mathrm{N}} \vec{\mathrm{v}}_{\mathrm{h}} \right), \tag{10}$$

where $\mathrm{t_j} = \mathrm{j} \times \Delta \mathrm{t}$, N is the total number of migration velocities per trajectory, and lowercase letters (i, j, g, and h) are time indexes. The resulting VAC for individual trajectories is shown in Fig. 1 *c*. It is obvious that the VAC follows a nonlinear decay in linear-log axes, which is the consequence

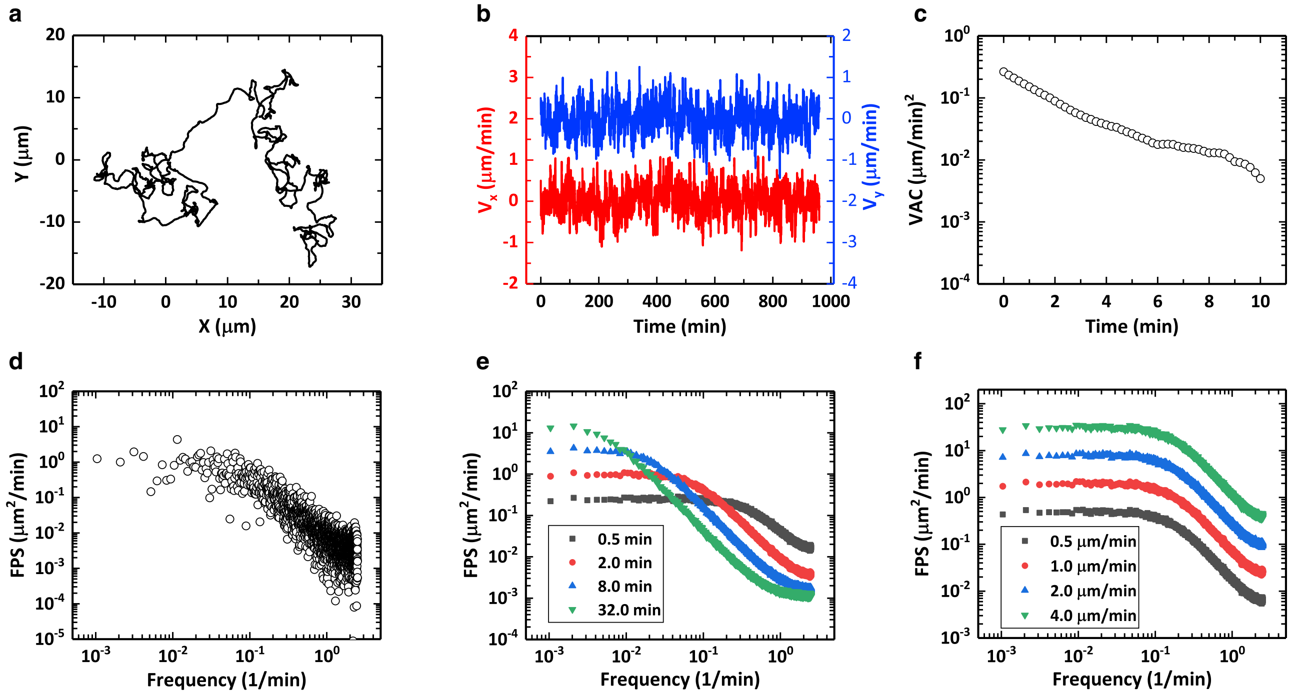

FIGURE 1 FPS analysis of cell migration velocities. (*a*) Individual cell migration trajectories simulated by TPRW model. (*b*) Velocity components on *x* and *y* axes. The red line corresponds to components on the *x* axis, and the blue corresponds to those on the *y* axis. (*c*) Velocity autocovariance function (VAC) in linear-log axes for individual cells. (*d*) FPS (FPS) in log-log axes for individual cells corresponding to the VAC in (*c*). (*e*) The effects of persistence time P on Fourier power spectra. (*f*) The effects of migration speed S on Fourier power spectra. To see this figure in color, go online.

of finite migration data. It is well known that the VAC of an Ornstein-Uhlenbeck (OU) process (41) obeys a linear decay in linear-log axes (i.e., single exponential decay in linear-linear axes); thus, the nonlinearity in Fig. 1 *c* will transition to linearity when considering a large amount of cell-trajectory data.

Although the VAC is a classical and widely used approach to analyze the properties of cell migration, especially the persistence, it still has limitations. For instance, it could not return reliable errors on the fitted motility parameters because of the correlations between velocities in time domain (35). To eliminate the correlations, a novel quantity, FPS, has been introduced based on the Wiener-Khinchin theorem (42,43), which states that "the power spectrum of any generalized stationary random process is the Fourier transform of its autocovariance function." According to this theorem, the Fourier transform of migration velocities is given as follows:

$$\vec{v}_k = \Delta t \sum_{j=1}^{N} e^{i2\pi f_k t_j} \times \vec{v}_j = \Delta t \sum_{j=1}^{N} e^{i2\pi kj/N} \times \vec{v}_j, \quad (11)$$

and similarly, the Fourier transform of velocity autocovariance function is given as

$$\begin{aligned} \mathrm{VA\widehat{C}}(f_k) &= \Delta t \sum_{j=1}^{N} e^{i2\pi f_k t_j} \times \mathrm{VAC}(t_j) \\ &= \Delta t \sum_{j=1}^{N} e^{i2\pi kj/N} \times \mathrm{VAC}(t_j) \end{aligned} \quad (12)$$

finally, the FPS is

$$\mathrm{FPS}(f_k) = \left\langle \left| \vec{v}_k \right|^2 \right\rangle \Big/ T = \frac{(\Delta t)^2}{T} \sum_{j_1=1}^{N} \sum_{j_2=1}^{N} e^{i2\pi f_k (t_{j1}-t_{j2})} \times \left\langle \vec{v}_{j1} \times \vec{v}_{j2} \right\rangle = \Delta t \sum_{j} e^{i2\pi f_k t_j} \times \mathrm{VAC}(t_j) = \mathrm{VA\widehat{C}}(f_k), \quad (13)$$

where $f_k = k/T$ is Fourier frequency, k is frequency index, $T = N \times \Delta t$ is total time for recording individual trajectories, j_1 and j_2 are time indexes, and the symbol "" denotes Fourier transformation. Thus, the corresponding FPS can be easily computed and then plotted in log-log axes, as shown in Fig. 1 *d*.

Next, we continue to analyze the effects of persistence time P and migration speed S on FPS separately; thus, several sets of motility parameters have been defined based on Eqs. 2 and 3, as given in Table 1. There, all the motility parameters are initial values at t = 0 min, and K_P equals 0. The left part in the table aims to study the effects of persistence time on FPS, and the right aims to study the effects of migration speed.

The corresponding results are shown in Fig. 1, *e* and *f*. It is evident that the FPS behave differently in the four cases of persistence times P (see Fig. 1 *e*), i.e., the horizontal region (e.g., 0.001–0.2/min for the code P1) gradually gets narrower as the decay region becomes wider (e.g., 0.2–2.5/min for the code P1), and the horizontal region corresponds to greater power spectral values as the persistence time P increases. Thus, the FPS will transition from the horizontal curve to a sharp decaying curve, which corresponds to the transition from pure diffusive motion to ballistic motion (44), i.e., the power spectral values in a given frequency domain will change from the uniform to nonuniform. In contrast, Fig. 1 *f* indicates that the increasing of S only increases the amplitude of FPS instead of changing

TABLE 1 Prescribed motility parameters in TPRW model

	Increasing persistence time P (min)				Increasing migration speed S (μm/min)		
Code	P_0	λ	S_0	Code	P_0	λ	S_0
P1	0.5	−1.39	0.5	S1	1.0	0	0.5
P2	2.0	1.39	0.5	S2	1.0	0	1.0
P3	8.0	4.16	0.5	S3	1.0	0	2.0
P4	32.0	6.93	0.5	S4	1.0	0	4.0

the decay rate. When the motility parameters are different from the counterparts for other cases, the corresponding FPS will exhibit more complex changes, which can be directly quantified by the Lorentzian power spectrum (31,35).

SE based on FPS

Inspired by the changes in FPS (see Fig. 1, *e* and *f*), we introduce SE to analyze the information encoded in FPS, especially the persistence. Entropy is an extensively used concept in thermodynamics, which is typically used to describe the degree of disorder or randomness in the states of molecules. It was not until 1948 that it was introduced to describe the "uncertainty" in information source by C. E. Shannon (45). Therefore, it is also referred to as "SE" when related to information theory.

Suppose there is a set of possible events with probabilities of occurrence $p_1, p_2, \ldots, p_n$; the SE is given as

$$H = -\sum_{i=1}^{n} p_i \times \log2(p_i), \tag{14}$$

where n is the total number of the events and p_i represents the probability of each event. Note that the SE is more commonly denoted by the capital letter "H" in the literature. The H not only measures how much "choice" is involved in the selection of the event but also how uncertain the outcome can be. To illustrate the relationship between the probabilities of events and H, we consider an example stochastic event with two outcomes associated with probabilities p and q, respectively, whose H is then given by

$$H = -[p \times \log2(p) + q \times \log2(q)], \tag{15}$$

where p lies in the closed interval [0, 1], and $q = 1 - p$. The result is plotted in Fig. 2 *a*, which shows that the H first increases and then decreases as the probability p increases. In addition, it is symmetrical at about $p = 0.5$, which means that the H reaches a maximum (one bit) when the probabilities of events are the same, and we are completely uncertain of the outcome. Inversely, when $p = 0$ or 1, H reaches a minimum (0 bit); thus, we are completely certain of the outcome. Otherwise, H lies in the open interval (0, 1).

Inspired by the analysis above for the case of two possibilities, we first normalize the individual power spectral values in Fig. 1 *d* by dividing the sum of individual power spectral values in the entire frequency domain, as shown in Fig. 2 *b*. The normalized spectral values associated with different frequencies can be viewed as probabilities associated of specific dynamic modes in Eq. 14, in which each value represents the probability of occurrence for each frequency (i.e., specific dynamic modes). Although the pattern of the normalized result is identical with that of FPS, it possesses completely different physical interpretations.

Following the procedure above, we calculate the H for every cell using Eq. 14. Further, we normalize all Shannon entropies (H) by dividing the maximal H corresponding to the case with same occurrence probabilities for all events, which excludes the effects of the total number of frequencies on H. After the normalization of all H, all normalized H are dimensionless and locate in the closed interval [0, 1] (see Fig. 2 *c*), which makes it convenient to compare with other cases. Here, we employ the abbreviation "SE" to denote the normalized H for avoiding confusion in this work. Fig. 2 *c* clearly shows that all SE values fluctuate around 0.6 with small derivation (standard deviation (SD): 0.016), which reflects the similarities between cells in a given cell population. Moreover, the distribution of all SEs in Fig. 2 *d* vividly exhibits the averaged SE (0.603) for this cell population in a more straightforward manner. Because the cell population here is modeled based on the same motility parameters, there are no big differences between all SEs. Similarly, we compute the ensemble-averaged SE for each case with 200 cells listed in Table 1. The detail results are shown in Fig. 2, *e* and *f*. In Fig. 2 *e*, the ensemble-averaged SE significantly decreases when the persistence time P increases, which indicates that the more persistent the cell migration, the smaller the corresponding SE (see Fig. S1 for the details about statistical significance test). However, the ensemble-averaged SEs in Fig. 2 *f* are identical with each other, which illustrates that migration speed S will not significantly affect the SE. Note that data are represented as mean ± SEM, where the SEM is "standard error of the sample mean" and estimated using $SD/(n_{sample})^{1/2}$, which quantifies the difference between the sample mean (estimated based on available data) and the overall mean (ground truth) and thus is a better metric here compared to SD, used to measure the degree of dispersion of the data (Fig. 2, *e* and *f*).

Because cellular migration with larger persistence time P can be considered as more ordered (as the migration dynamics of a later time is more strongly correlated with the state at earlier times), which is also associated with smaller SE, it is reasonable to consider that the SE naturally offers a metric for quantifying the degree of order or randomness of the migration dynamics, in a similar sense as to how Gibbs entropy quantifies the order or randomness of a thermal system. To further illustrate the physical meaning of SE, we show in the Supporting materials and methods that SE attains the maximal value of unity for purely diffusive migration (i.e., SE = 1 for the most "random" dynamics) and the minimal value of 0 for purely ballistic dynamics (i.e., SE = 0 for the most "ordered" dynamics) (see Figs. S4 and S5). With this physical interpretation, it is always easy to understand why SE is less sensitive to migration speed S, i.e., the change of S only speeds up or slows down the migration but does not affect the degree of randomness of migration dynamics. On the other hand, changing the persistence time P can significantly affect the correlation of migration dynamics at different time points, thus strongly influencing the SE.

Time-varying motility parameters incorporating cellular heterogeneity

Different from the cases with constant motility parameters, the function P(t) is constructed again, based on two Gaussian distributions, i.e., $N_1(2.0, 0.1^2)$ and $N_2(20.0, 1.0^2)$. The persistence times at initial P(0) and final P(T) moments are determined by the distributions N_1 and N_2, respectively. The persistence time still obeys a linear function (cf. Eq. 2). Furthermore, the migration speed can be computed using Eq. 3 when both of A and λ are equal to 1. So far, the motility parameters for a cell population have been constructed (see Fig. S2 for more details), which exhibit three characteristics: 1) both of them vary with time and the former changes linearly, whereas the latter exhibits nonlinear behavior; 2) both of them increase with time and together reflect the "enhance" of migration capability; and 3) the motility parameters for each cell are different from the corresponding parameters of another cells, reflecting the heterogeneities among the cell population.

Fig. 3 *a* shows how the motility parameters change over time for one cell, and the corresponding trajectory is plotted in Fig. 3 *b*. The velocity components in Fig. 3 *c* exhibit smaller random fluctuations with time lapsing when compared with those in Fig. 1 *b*, which may reflect the enhanced persistence of cell migration to some extent.

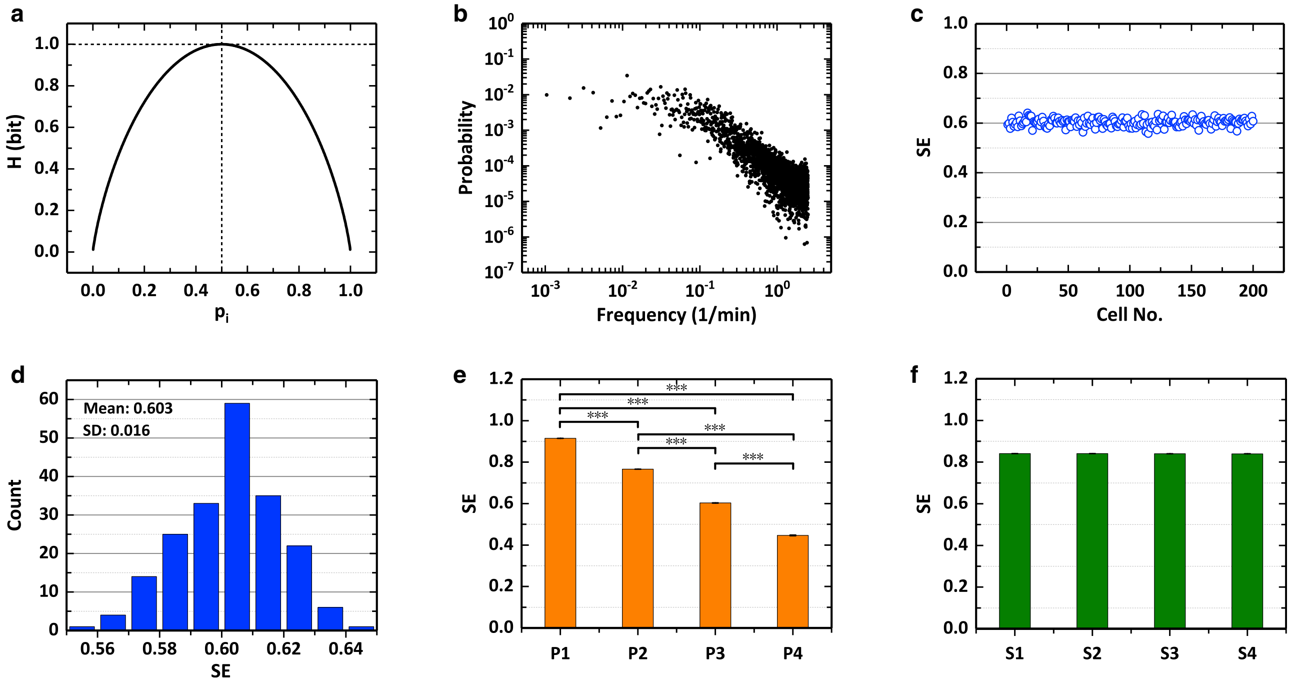

FIGURE 2 SE characterizing the persistence of cell migration. (*a*) SE (H) as a function of probability in the case of two possibilities. (*b*) The probability as a function of frequency obtained from the normalization to FPS in Fig. 1 *d*. (*c*) Shannon entropies (SEs) for a given cell population. (*d*) Distribution of Shannon entropies (SEs) for a given cell population with mean and SD. (*e*) Shannon entropies (SEs) in the cases of four different persistence times. Data are mean ± SEM (standard error of the sample mean); the cell number for each group is 200; ***$p < 0.001$, Kruskal-Wallis test. (*f*) Shannon entropies (SEs) in the cases of four different migration speeds. Data are mean ± SEM; the cell number for each group is 200; $p > 0.05$, analysis of variance. To see this figure in color, go online.

Wavelet power spectrum of cell migration velocities

Because cell migration is significantly affected by ICSPs and ECM, it is necessary to study the time-varying characteristics encoded in cell migration trajectories. Although the SE based on FPS used above is able to capture the intensity of persistence for each cell, it is incapable of obtaining time-dependent information (46). To address this issue for a time series, researchers developed two approaches, i.e., windowed Fourier transform (WFT) and WT.

The WFT is an analytical tool to extract time-frequency information from a time series, which performs the Fourier transform on a sliding segment of a constant time interval from a time series. Here, the segment can be windowed with an arbitrary function, e.g., a boxcar or a Gaussian window. Although the WFT shows the ability of extracting time-frequency information, it still cannot avoid several deficiencies, e.g., 1) it takes much effort to determine the most appropriate window size, 2) the aliasing of high and low frequency components may not fall within the frequency range of the window, and 3) the frequencies corresponding to the segment must be analyzed at each time step, regardless of the window size or the dominant frequencies. These deficiencies make the WFT inaccurate and inefficient under certain situations, as discussed by Kaiser et al. (47), Torrence et al. (48), and Daubechies (49).

Different from the WFT, the window size of WT can vary with frequency, which allows one to analyze the time-frequency characteristics of a time series (48–50). The WT was originally employed by Morlet et al. to analyze seismic signals in the early 1980s (51,52), and later formalized by Goupillaud et al. (53,54). Because of the better performance in studying the nonstationary and infinitely correlated processes, the WT has become an influential tool. For instance, the wavelet coefficients of fractional Brownian motion are stationary and uncorrelated (55). Note that the WT here includes discrete WT and continuous WT; the latter is employed in this work. For a given time series v_j, the continuous WT (48) is performed by computing the convolution of $v_{j'}$ with a scaled and the translated version of the wavelet function ψ_0^*, which is written as

$$W_j(s) = \sum_{j'=0}^{N-1} v_{j'} \times \psi_0^*\left[\frac{(j'-j)\times\Delta t}{s}\right], \tag{16}$$

where the asterisk "*" represents the complex conjugate and s is the wavelet scale relating to Fourier frequency. Moreover, the Morlet wavelet is employed here as the wavelet function, which consists of a plane wave modulated by a Gaussian

$$\psi_0(\eta) = \pi^{-1/4} \times e^{i\omega_0\eta} \times e^{-\eta^2/2}, \tag{17}$$

where $\omega_0 = 6$ is a nondimensional frequency and satisfies the admissibility condition (56). Note that the Morlet used is a complex function; thus, the final results are also complex. Further, we have access readily to the corresponding real part and imaginary part, both of them together are used to compute the wavelet power spectrum (WPS) (48,57), i.e., $|w_j(s)|^2$.

So far, we can analyze the characteristics of cell migration in time-frequency domain following the procedure above, as seen in Fig. 3 *d*. The horizontal axis corresponds to the total time for recording cell trajectory, and the vertical axis corresponds to the Fourier frequency. The legend in the right of the plot shows the power spectral values represented by different colors. Here, the power spectral values along the frequency axis are called the local WPS, and the "local" denotes each moment of the total time. It is evident that the local WPS in the interval (0–0.5/min) are greater than those in other ranges (0.5–2.5/min), which roughly exhibit the persistence or

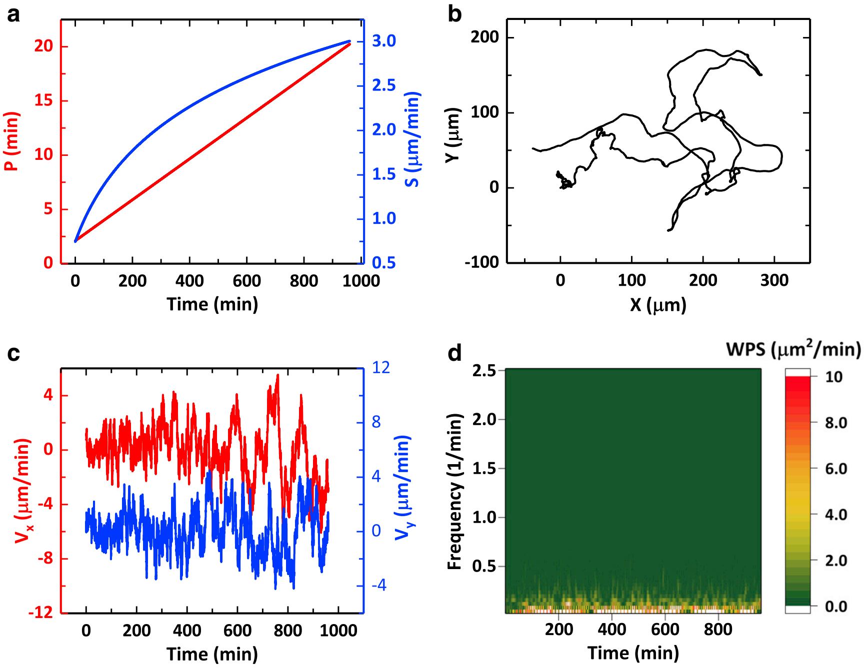

FIGURE 3 WPS of cell migration velocities. (*a*) Motility parameters as functions of time. The red line indicates the persistence time P obeying Eq. 2, and the blue line indicates the migration speed S derived from Eq. 3. (*b*) Individual cell migration trajectories simulated by the TPRW model based on the motility parameters plotted in (*a*). (*c*) Velocity components on *x* and *y* axes. The red line denotes the components on *x* axis, and the blue denotes those on *y* axis. (*d*) WPS of individual cell migration velocities. The different colors denote the power spectral values in the time-frequency domain. To see this figure in color, go online.

correlations of cell migration velocities. When averaging the local WPS along the time axis, we obtain time-averaged power spectral values for all cells, which are also called global WPS. Torrence et al. (48) validated that the local WPS is identical with the FPS of an OU process, and the global WPS tends to approximate the FPS; thus, it is reasonable to compute SE for all cells at every moment based on the local WPS.

Actually, edge effects will occur in the beginning and end of the WPS because of the finite-length time series, which is also called “cone of influence” (COI) (48). Because the COI will affect the true information, we exclude the region affected by COI in this work and the remaining region (see Fig. 3 *d*) exhibits the true characteristics of cell migration.

The time-varying SE based on WPS

Based on the WPS in Fig. 3 *d*, we normalize the local WPS to obtain probabilities and compute the Shannon entropy (SE) at each moment. Furthermore, the time-varying SEs can be obtained for a given cell population with 200 cells that are stacked along the vertical axis, as shown by the SE heatmap in Fig. 4 *a*. From the overall view, the time-varying SEs gradually decrease with time lapsing, which indicates that all persistence for this cell population gradually increases, and this tendency is consistent with that of P in Fig. 3 *a*.

When averaging the SE heatmap of this cell population along the time axis, we obtain the time-averaged SE for each cell, which quantifies the average persistence in the process of one cell migration. The corresponding distribution is exhibited in Fig. 4 *b*, with mean (0.404) showing overall characteristics for this cell population, which can be easily used to compare with other cell populations. Similarly, we obtain the cell-averaged SE by averaging the SE heatmap of this cell population along the vertical axis, as shown in Fig. 4 *c*. Different from the time-varying SE of individual cells, the cell-averaged SE looks smoother (SD: 0.032) because the fluctuations caused by intrinsic noises have been averaged. Moreover, the gradual decrease of the cell-averaged SE illustrates that the persistence of cell migration gradually increases, which conforms to the tendency of P in Fig. 3 *a*. Thus, the cell-averaged SE is better to reflect the overall changes against time in persistence of cell population. Although this case exhibits different dynamical characteristics from conventional motility models, its statistical profiling is consistent with that for in vitro cell experiments, as shown in Fig. S7.

For validating the efficiency of SE in analyzing cell migration, we simulate 200 cell migration trajectories again with the same procedure as the enhance case (see Fig. S2 for detail illustration). The only difference is that both of the initial and final persistence times are obtained from the Gaussian distribution $N_3(11.0, 0.1^2)$, and this case is referred to as “stable,” meaning that the persistence of cell migration almost does not change with time. As seen in Fig. 4 *d*, it is clear that the SE heatmap of this cell population looks more homogeneous, the time-averaged SE in Fig. 4 *e* has a smaller mean (0.389), and the cell-averaged SE in Fig. 4 *f* exhibits a more stable tendency (SD: 0.010) when compared with the results of the enhanced case (see Fig. S7 for more statistical profiling).

Furthermore, we replace the linearly varying P(t) with a randomly varying P′(t), in which the persistence time P(t) at every moment is sampled randomly from a given Gaussian distribution. The results indicate clearly that the SE decreases significantly when the persistence times, i.e., the mean values of three Gaussian distributions, increase, which also validates that the SE correlates negatively with persistence time (see Fig. S3 for detail results).

RESULTS AND DISCUSSION

SE measures the directional persistence of migration regulated by Arpin protein

With the physical meaning and information encoded in SE having been clearly understood from the study of the simulated trajectory data in the previous sections, we now apply an SE-based approach to analyze the migratory dynamics regulated by ICSPs. In this subsection, all the experimental data are obtained from the work by Dang et al. (8), which reported a new protein, Arpin. The new Arpin inhibits the Arp2/3 complex and antagonizes an intrinsic positive

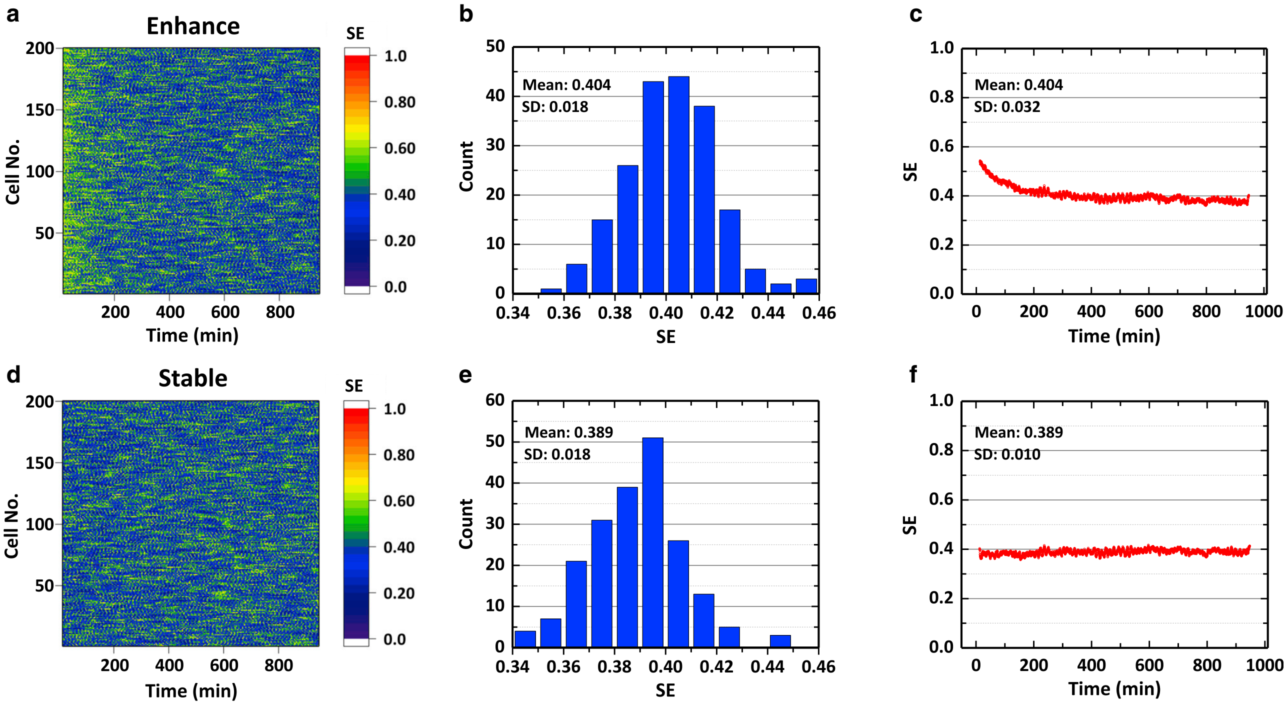

FIGURE 4 The time-varying Shannon entropies (SE heatmap) for enhance (*a–c*) and stable (*d–f*) cases. (*a*) The time-dependent Shannon entropies (SEs) for a cell population with 200 cells. The different colors denote the values of SE. (*b*) Distribution of the time-averaged SE for a cell population. (*c*) The cell-averaged SE as a function of time. (*d–f*) The captions are the same as those for (*a*)–(*c*). To see this figure in color, go online.

feedback loop sustaining lamellipodial protrusion, which promotes turning during migration, i.e., the inhibitory Arpin decreases the directional persistence of migration. Dang et al. (8) analyzed the migratory dynamics of four different motility types regulated by Arpin protein, i.e., the mammary carcinoma cell line MDA-MB-231 in 2D (MDA in 2D), in 3D (MDA in 3D), the motile amoeba *Dictyostelium discoideum*, and fish keratocytes. In this work, we analyze the four motility types above using the SE-based approach; the results for *D. discoideum* are shown in Figs. 5 and 6, and the other results are exhibited in Fig. S12.

Fig. 5 shows the comparison of results based on SE analysis for Arpin-knockout amoeba, wild-type amoeba, and rescued by green fluorescent protein (GFP)-Arpin expression in knockout amoeba. In terms of qualitative analysis, the SE heatmap for rescued amoeba (see Fig. 5 *g*) contains an excess of larger values of SE, whereas the Arpin-knockout amoeba (Fig. 5 *a*) contains a larger number of smaller values. For quantitative analysis, the mean (0.701) of time-averaged SE for rescued amoeba (Fig. 5 *h*) is maximal, wild-type (0.687) is smaller (Fig. 5 *e*), and Arpin-knockout (0.660) is minimal (Fig. 5 *b*). Furthermore, the cell-averaged SEs show the time-dependent characteristics, e.g., the abrupt increase (peak) in the interval 10–11 min for Arpin-knockout amoeba (Fig. 5 *c*), which may be due to mechanisms including observation noises and physical constrains, etc. In addition, the cell-averaged SEs also show the overall fluctuations (corresponding to fluctuations in migration persistence), e.g., the SEs for rescued amoeba exhibit the smallest fluctuations (SD: 0.016) with time lapsing (Fig. 5 *i*), wild-type is the next (SD: 0.022) (Fig. 5 *f*), and those for Arpin-knockout exhibit the largest fluctuations (SD: 0.041) (Fig. 5 *c*).

The above analysis of *D. discoideum* indicates that Arpin content is apparently positively correlated with SE based on the migration trajectories, i.e., Arpin depletion increases directional persistence of migration. To further verify this observation in a rigorous manner, we perform statistical significance tests for three groups of *D. discoideum* (see Fig. S1 for detail procedure and criteria about statistical significance test). Fig. 6 shows ensemble-averaged SEs and indicates evidently that Arpin-knockout amoeba possesses the minimal SE (0.660), which implies greater directional persistence of migration and significantly differs from those for wild-type and rescued by GFP-Arpin expression in knockout amoeba. In addition, there is no significant difference between the wild-type and rescued amoeba, which illustrates that the directional persistence is more than fully rescued by GFP-Arpin expression in knockout amoeba. These results are consistent fully with those reported by Dang et al. (8), which not only illustrate that Arpin depletion does increase the persistence of migration but also illustrate the utility of an SE-based approach in quantifying the persistence of migration (see Fig. S12 for analysis of other three motility types).

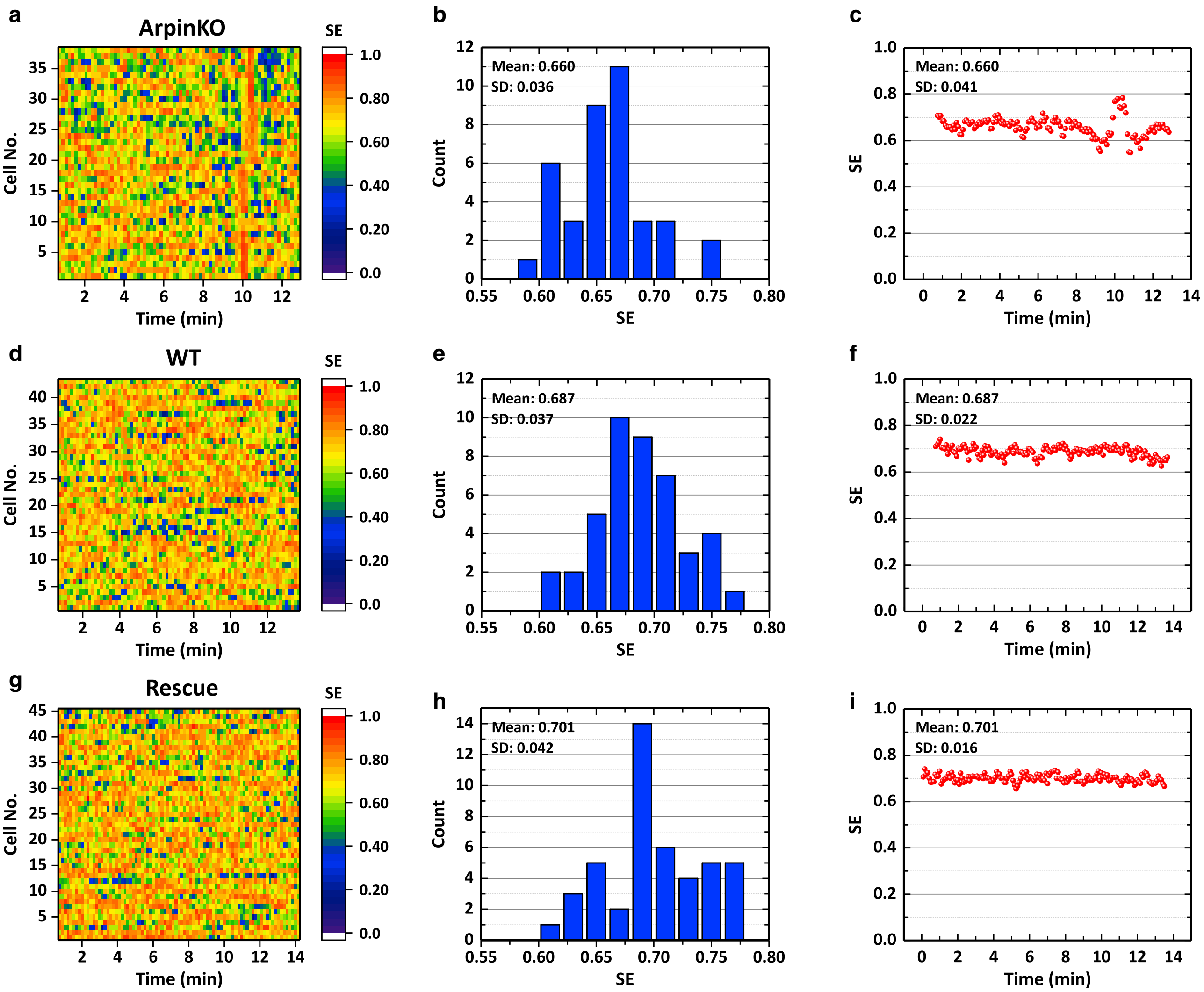

FIGURE 5 SE of the automatically tracked *D. discoideum*. The captions are the same as those in Fig. 4, *a–c*, but the results correspond to Arpin-knockout (ArpinKO) (*a–c*), wild-type (WT) (*d–f*), and rescued by GFP-Arpin expression in knockout amoeba (Rescue) (*g–i*), respectively. Here, the time interval between frames is 5 s and the number of trajectories is 38, 43, and 45, respectively. All the experimental data are obtained from (8). To see this figure in color, go online.

SE detects the transition of cell migration dynamics in confined microenvironment

In this subsection, we continue to investigate transition between two types of cell migration modes resulted from confined physical environments using SE. We will show that the resulting SEs as function of time clearly indicates distinct migration modes as the cell crawls through very confined channels and gets into less confined environment. Firstly, we use the experimental data obtained from (58), in which the authors constructed a microstructural channel array and measured primary breast cancer cell migration through a one-dimensional array of sequentially channels and chambers (see Fig. 7 *a*). To record the cells' positions, the nuclei are stained with Hoechst and marked in red. The trajectory of cell 4 is plotted in Fig. 7 *b*, which contains two sections marked by green and red, respectively. The enlarged red section clearly shows that the cell "flutters" with a smaller migration capability when compared with the green section (see Fig. S6 for similar simulated trajectory).

To quantify this difference, we calculate the time-varying SE and the time-averaged values (marked by *black lines*), as seen in Fig. 7 *c*. The time-averaged SEs exhibit significant differences, i.e., the cell first possesses smaller SE (mean: 0.618) with greater fluctuations and then possesses greater SE (0.794) with smaller fluctuations (transition position, 1210 min). The transition in time-varying SE clearly indicates that the cell migrates with greater persistence initially, then the persistence gradually decreases to a smaller level, which behaves like the transition from ballistic to diffusive motion (44) (a similar transition is shown in Fig. S6).

Following the same procedure, we analyze a single MDA-MB-231 cell migrating on a two-state pattern, as indicated

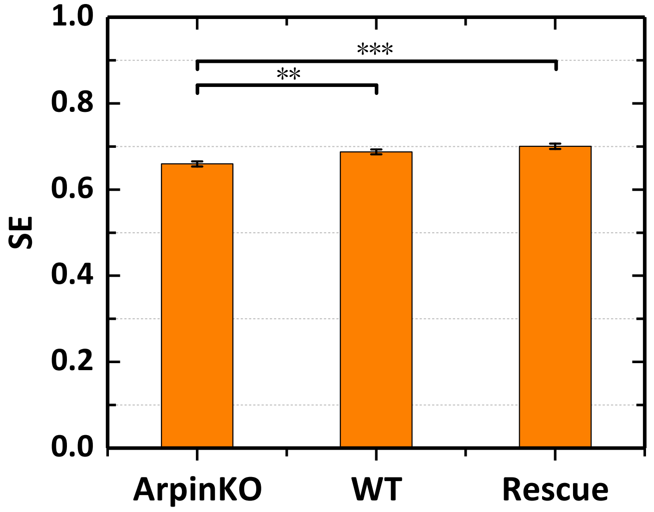

FIGURE 6 Arpin depletion increases directional persistence of migration in *D. discoideum*. Arpin-knockout amoeba possesses the minimal SE and Arpin expression can rescue the directional persistence in Arpin-knockout amoeba. Data are mean ± SEM; the number of trajectories is 38, 43, and 45, respectively; $^{**}p < 0.01$, $^{***}p < 0.001$, analysis of variance. To see this figure in color, go online.

in Fig. 7 *d*. The two-state patterns are coated by fibronectin and consists of two square islands (37 μm × 37 μm) that are connected by bridges. The corresponding data are obtained from (59), and more details are given in (60,61). The trajectory contains three sections, which are denoted by red, green, and blue, respectively (see Fig. 7 *e*). The SE as a function of time clearly indicates three migration modes: 1) the first mode (*red*) indicates SE almost keeps stable (mean: 0.790) with smaller fluctuations (transition point, 1340 min); 2) the second mode (*green*) possesses a smaller SE (mean: 0.541) with greater fluctuations (transition point, 1940 min); and 3) the third mode (*blue*) shows the gradually decreasing SE (mean: 0.655) with gradually increasing fluctuations, as seen in Fig. 7 *f*. These three modes can be explained perfectly by the colored trajectory and correspond, respectively to three migration states in this two-state pattern: 1) the first mode demonstrates that the cell migrates in one island and possess smaller persistence as it has more space to move around, 2) the second mode shows that the cell is migrating through the bridge and possess greater persistence because of stronger physical constraints and confinement, and 3) the third mode shows that the cell migrates from island to bridge and the confinement is enhanced gradually. The results in Fig. 7 further illustrate SE is efficient in detecting the transition of cell migration dynamics.

SE reveals strongly correlated cell migrations regulated by remodeled collagen fiber bundles

To further demonstrate the utility of SE, we perform in vitro experiments to obtain migration dynamics of MCF-10A mammary epithelial cells on top of 3D collagen gel (~2 mm thickness) and on a 2D petri dish (solid polystyrene substrate, as control group) and compute the Shannon entropy (SE) to analyze these data sets. In particular, MCF-10A cells marked with GFP were obtained from China Infrastructure of Cell Line Resource. The culture medium of MCF-10A-GFP is Dulbecco's modified Eagle's medium-F12 (Corning, Corning, NY) supplemented with 5% horse serum (Gibco, Gaithersburg, MD), 1% penicillin/streptomycin (Corning), 20 ng/mL human EGF (Gibco), 10 μg/mL insulin (Roche Diagnostics, Basel, Switzerland), 100 ng/mL cholera toxin (Sigma-Aldrich, St. Louis, MO), and 0.5 μg/mL hydrocortisone (Sigma-Aldrich). Type I collagen extracted from rat tail tendon (Corning) was diluted and the PH neutralized to ~7.2, then the collagen solution was spread on the substrate of petri dish and incubated in 37°C for 30 min until it polymerized into a 3D matrix with a thickness of around 2 mm. The final collagen concentration was 2 mg/mL for the tests. The cell suspension covered the matrix and stayed in the cell incubator overnight before imaging. For cell migration test, 0.5 μL of cell suspension with different concentrations of cells was dropped on top of collagen gel or a solid petri dish, then incubated for 2 h before imaging. Time-lapse images were obtained using both a confocal laser scanning microscope with a 25× water immersion objective and an automatic inverted fluorescent microscope (Nikon Ti-E, Tokyo, Japan) with a 10× objective. Both microscopes are equipped with an on-stage cell-culture incubator to provide a constant temperature of 37°C with humidity 5% CO_2.

We obtain in vitro migration trajectories of MCF-10A cells on 3D collagen I hydrogel with a collagen concentration 2 mg/mL and thickness of ~2 mm by randomly distributing the MCF-10A cells on collagen-based ECM with a low cell density of 10^4 cells/cm^2. Furthermore, we carry out control experiments by obtaining migration trajectories of MCF-10A cells on a 2D petri dish by initially randomly distributing the cells on solid polystyrene substrate with a number density of 10^4 cells/cm^2. We record the migration trajectories for the two cases above with a sampling time of 2 min for every frame.

As shown in Fig. 8 *a1–a4*, we observe highly correlated pairwise migration dynamics for cells on the collagen gel, when the initial distance between the cells is roughly within ~80–120 μm. In particular, the two cells within this influence region tend to rapidly move toward each other, following a ballistic-like path (see Fig. 8 *b*). To better quantify the correlated migration dynamics, we compute the time-varying SEs for both cells exhibiting the correlation behaviors (see Fig. 8 *c*). Subsequently, Pearson's correlation coefficient is calculated (r = 0.813) based on the SEs against time, which indicates the high correlation between the pairwise cells. The detailed illustrations of Pearson's correlation coefficient are given in the

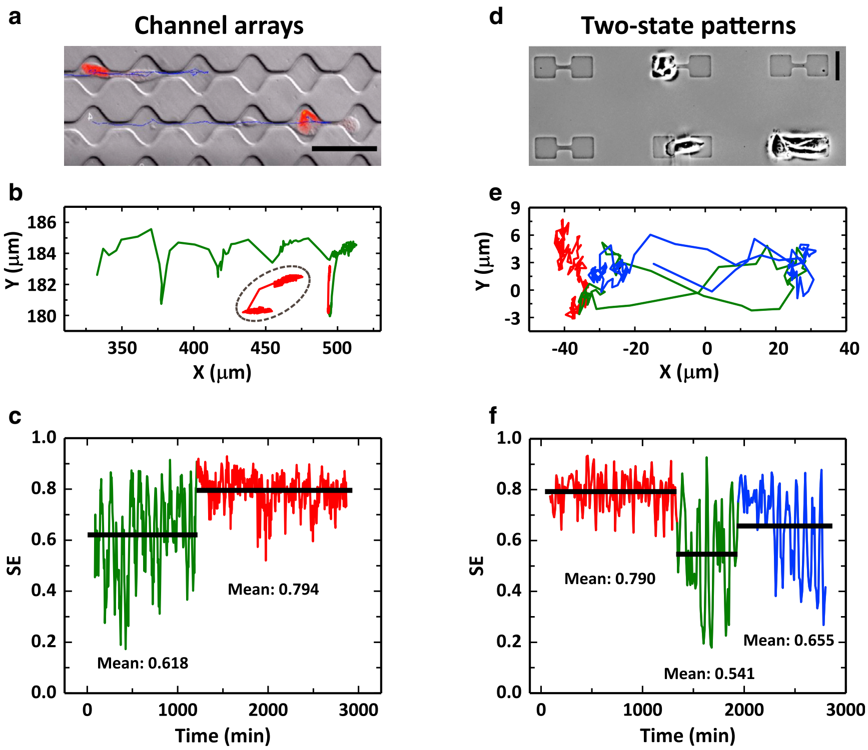

FIGURE 7 SE detects transition of confined cell migration dynamics. (*a*) Primary breast cancer cell migration through microstructure with channels and chambers (channel array). Cell nuclei are stained with Hoechst and shown in red. The scale bar (*horizontal black line*) is 50 μm. Note that this image is flipped 180° left and right (see "flipped 4" for reference). (*b*) Individual cell migration trajectories containing two sections denoted by green and red, respectively. The inset is an enlarged view of the red section. The sample time interval between frames is 5 min. Image (*a*) and data (*b*) are reproduced from (38) and (58) with permission, respectively. (*c*) SE as a function of time. The black lines represent time-averaged SEs, respectively for these sections. (*d*–*f*) The captions are the same as those for (*a*)–(*c*), but the data correspond to a single MDA-MB-231 cell migrating on a two-state pattern. The sample time interval between frames is 10 min and the scale bar (*vertical black line*) is 50 μm. Image (*d*) and data (*e*) courtesy of David B. Brückner (59). To see this figure in color, go online.

Supporting materials and methods; here, we simply note that a perfect correlation corresponds to r = 1 (−1) and totally uncorrelated migration is associated with r = 0. Furthermore, we compute averages of the absolute values of Pearson's correlation coefficients for 14 randomly selected cell pairs migrating on collagen gel, and the averaged coefficient is 0.407, as shown in Fig. 8 *d*. In contrast, the Pearson's correlation coefficient for cell pairs on a 2D petri dish is much lower (0.219), indicating less significant correlations among these cells. In addition, we also validate that the high collagen concentration will contribute to enhancing the directional persistence of cell migration (62), whereas cell density shows no significant effects on the directional persistence (see Figs. S8–S11 for detail analysis).

In this case, we clearly see the utility of the SE-based approach in quantitatively characterizing and distinguishing distinct migration dynamics for cells in different microenvironment. Our subsequent analysis of the time-lapsing confocal imaging data (see Fig. 9) revealing both the migrating cells and the collagen fibers shows that correlated migrations are mainly due to the dynamically reorganized collagen bundles between the two migrating cells. Recent studies (10,11,63–65) suggested that the remodeling of collagen fibers is mainly due to the active tensile forces generated by the migrating cells, and the bundles bridging the two cells typically carry tensile forces that, in turn, regulate the cell migration and lead to the observed strongly correlated migration.

CONCLUSIONS

In this work, we introduce the time-varying Shannon entropy (SE) based on the WPS obtained by performing WT of migration velocities and demonstrate its superior utility to characterize the persistence of cell migration. We first introduce a TPRW model and further construct the functions regarding motility parameters when considering the effects of ICSPs or ECM to simulate cell migration. On the basis of cell migration trajectories, we compute the VAC and FPS to quantitatively investigate the characteristics of cell migration in frequency domain. To study the effects of individual motility parameters on the shape of FPS, we define eight sets of parameters. Results show the persistence time changes the decay rate of FPS, and the migration speed changes the amplitude. Moreover, inspired by the changes in Fourier power spectra, we then introduce SE that can be obtained from the probabilities corresponding to power spectral values to quantify the persistence of cell migration, i.e., the smaller the SE, the closer the cell migration to a ballistic motion.

Because of the transitory nature of cell migration capability in heterogeneous environment, WT is employed to analyze cell migration velocities and derive WPS, which exhibits the time-frequency characteristics of cell migration. Because the local WPS along the frequency axis is identical with the FPS of an OU process, we further compute the SE for each local WPS and naturally obtain the time-varying SE, which accurately reflects the time-varying persistence.

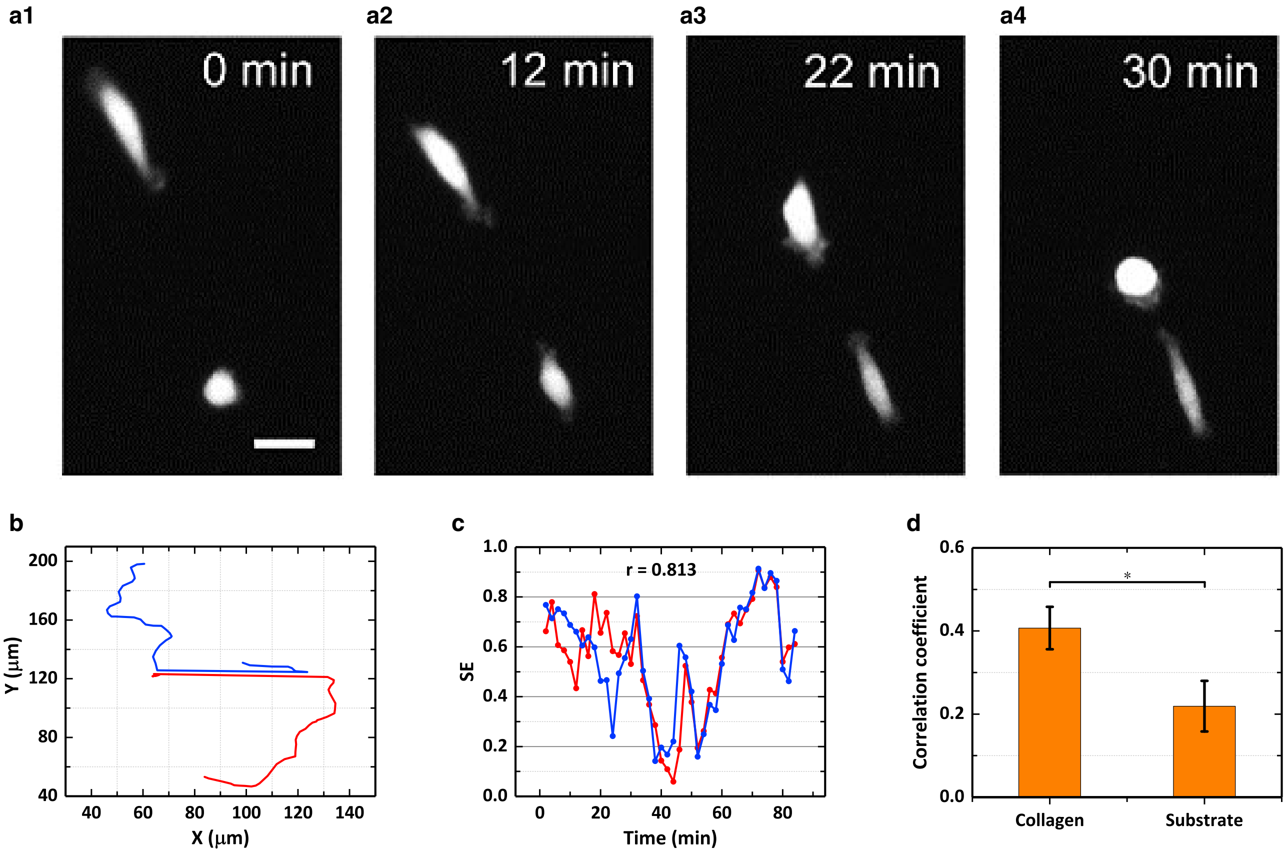

FIGURE 8 Analysis based on SE between pairs of MCF-10A cells migrating on collagen gel and 2D petri dish. (*a1–a4*) Time-lapsing images indicating a typical pair of cells migration on collagen gel. (*b*) Migration trajectories of the pairwise cells. (*c*) Time-dependent Shannon entropies of the pairwise cells. Pearson's correlation coefficient is 0.813. (*d*) Averaged Pearson's correlation coefficients for cells pairs migrating on collagen gel (randomly selected 14 pairs of cells) and 2D petri dish (randomly selected 14 pairs of cells). Data are mean ± SEM; the number of coefficients is 14 and 14, respectively; $^*p < 0.05$, Wilcoxon rank sum test. The scale bar represents 30 μm. To see this figure in color, go online.

To illustrate the utility and efficiency of our approach, we analyze trajectory data of in vitro cell migration regulated by distinct ICSPs and ECM, exhibiting a rich spectrum of dynamic characteristics and persistence. In particular, our results based on the experimental data of *D. discoideum* indicate that SE can sensitively quantify the directional persistence of migration regulated by the Arpin protein, and it can also be as a metric to detect the transition of cell migration modes in a confined microenvironment. In addition, our analysis based on the migration trajectories of MCF-10A cells on a collagen layer and solid 2D substrate indicates that SE can very sensitively capture and quantify the strongly correlated migration dynamics between cells on collagen gel and distinguish the migration dynamics of cells on a petri dish. Although SE alone does not directly offer insights on physical mechanisms, SE-based quantifications certainly aid the subsequent analysis to reveal the mechanism leading to the correlation migration. Therefore, we conclude by remarking that the SE not only efficiently quantifies the directional persistence of migration but also effectively captures the time-varying persistence of cell migration, which can also reflect the real-time effects of ICSPs or ECM to some extent.

The data that support the results within this work are available from the corresponding authors upon request.

SUPPORTING MATERIAL

Supporting material can be found online at https://doi.org/10.1016/j.bpj.2021.04.026.

AUTHOR CONTRIBUTIONS

Y.L. conceived the initial idea, derived the equations, and performed the calculations. Y.L. and Y.J. performed the analysis and developed the framework. Q.F., Y.Z., and G.L. designed the experiments. Q.F. and G.W. performed the experiments. S.L. and G.C. provided ideas and suggestions for the work. Y.L., Y.J., and L.L. wrote the manuscript. L.L. and J.S. supervised the project.

ACKNOWLEDGMENTS

We thank Ben Fabry, Christoph Mark, Claus Metzner, and Lena Lautscham for providing experimental data concerning channel arrays and David B. Brückner, Alexandra Fink, Chase Broedersz, and Joachim Rädler for

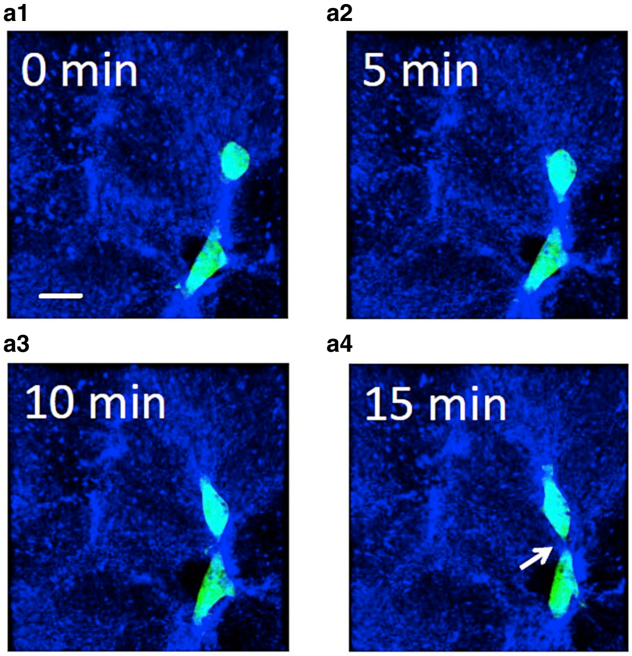

FIGURE 9 Time-lapsing confocal imaging data revealing strongly correlated pairwise cellular migration regulated by the remodeled ECM fiber bundles. The cells are shown in green, and the fiber bundles bridging the migrating cell pairs are shown in light blue. The scale bar represents 25 μm. Recent studies suggest that the fiber bundle bridge was formed because of mechanical remodeling of the collagen gel by the active tensile forces generated by the migrating cells (11,63). To see this figure in color, go online.

providing experimental data concerning two-state micropatterns. Y.J. thanks Arizona State University for support during his sabbatical leave.

This research was supported by the National Natural Science Foundation of China (grant nos. 11974066, 11674043, 11675134, and 11874310), the Fundamental Research Funds for the Central Universities (grant no. 2019CDYGYB007), and the Natural Science Foundation of Chongqing, China (grant nos. cstc2019jcyj-msxmX0477 and cstc2018jcyjA3679).

REFERENCES

1. Vicente-Manzanares, M., and A. R. Horwitz. 2011. Cell migration: an overview. *Methods Mol. Biol.* 769:1–24.
2. Kulesa, P., D. L. Ellies, and P. A. Trainor. 2004. Comparative analysis of neural crest cell death, migration, and function during vertebrate embryogenesis. *Dev. Dyn.* 229:14–29.
3. Tremel, A., A. Cai, …, A. J. O'Connor. 2009. Cell migration and proliferation during monolayer formation and wound healing. *Chem. Eng. Sci.* 64:247–253.
4. Sharma, G.-D., J. He, and H. E. P. Bazan. 2003. p38 and ERK1/2 coordinate cellular migration and proliferation in epithelial wound healing: evidence of cross-talk activation between MAP kinase cascades. *J. Biol. Chem.* 278:21989–21997.
5. Cyster, J. G. 2003. Lymphoid organ development and cell migration. *Immunol. Rev.* 195:5–14.
6. Friedl, P., and S. Alexander. 2011. Cancer invasion and the microenvironment: plasticity and reciprocity. *Cell.* 147:992–1009.
7. Gorelik, R., and A. Gautreau. 2015. The Arp2/3 inhibitory protein arpin induces cell turning by pausing cell migration. *Cytoskeleton (Hoboken).* 72:362–371.
8. Dang, I., R. Gorelik, …, A. Gautreau. 2013. Inhibitory signalling to the Arp2/3 complex steers cell migration. *Nature.* 503:281–284.
9. Polacheck, W. J., I. K. Zervantonakis, and R. D. Kamm. 2013. Tumor cell migration in complex microenvironments. *Cell. Mol. Life Sci.* 70:1335–1356.
10. Nan, H., L. Liang, …, Y. Jiao. 2018. Realizations of highly heterogeneous collagen networks via stochastic reconstruction for micromechanical analysis of tumor cell invasion. *Phys. Rev. E.* 97:033311.
11. Zheng, Y., H. Nan, …, Y. Jiao. 2019. Modeling cell migration regulated by cell extracellular-matrix micromechanical coupling. *Phys. Rev. E.* 100:043303.
12. Kim, J., Y. Zheng, …, B. Sun. 2020. Geometric dependence of 3D collective cancer invasion. *Biophys. J.* 118:1177–1182.
13. Charras, G., and E. Sahai. 2014. Physical influences of the extracellular environment on cell migration. *Nat. Rev. Mol. Cell Biol.* 15:813–824.
14. Gupta, G. P., and J. Massagué. 2006. Cancer metastasis: building a framework. *Cell.* 127:679–695.
15. Jemal, A., R. Siegel, …, E. Ward. 2010. Cancer statistics, 2010. *CA Cancer J. Clin.* 60:277–300.
16. Novikova, E. A., M. Raab, …, C. Storm. 2017. Persistence-driven durotaxis: generic, directed motility in rigidity gradients. *Phys. Rev. Lett.* 118:078103.
17. Lo, C. M., H. B. Wang, …, Y. L. Wang. 2000. Cell movement is guided by the rigidity of the substrate. *Biophys. J.* 79:144–152.
18. Park, J., D. H. Kim, …, A. Levchenko. 2016. Directed migration of cancer cells guided by the graded texture of the underlying matrix. *Nat. Mater.* 15:792–801.
19. Zhu, J., L. Liang, …, L. Liu; U.S.-China Physical Sciences-Oncology Alliance. 2015. Enhanced invasion of metastatic cancer cells via extracellular matrix interface. *PLoS One.* 10:e0118058.
20. Han, W., S. Chen, …, L. Liu. 2016. Oriented collagen fibers direct tumor cell intravasation. *Proc. Natl. Acad. Sci. USA.* 113:11208–11213.
21. Codling, E. A., M. J. Plank, and S. Benhamou. 2008. Random walk models in biology. *J. R. Soc. Interface.* 5:813–834.
22. Li, L., S. F. Nørrelykke, and E. C. Cox. 2008. Persistent cell motion in the absence of external signals: a search strategy for eukaryotic cells. *PLoS One.* 3:e2093.
23. Harris, T. H., E. J. Banigan, …, C. A. Hunter. 2012. Generalized Lévy walks and the role of chemokines in migration of effector CD8+ T cells. *Nature.* 486:545–548.
24. Weiss, G. H. 2002. Some applications of persistent random walks and the telegrapher's equation. *Physica A: Statistical Mechanics and Its Applications.* 311:381–410.
25. Li, L., E. C. Cox, and H. Flyvbjerg. 2011. 'Dicty dynamics': Dictyostelium motility as persistent random motion. *Phys. Biol.* 8:046006.
26. Sadjadi, Z., M. R. Shaebani, …, L. Santen. 2015. Persistent-random-walk approach to anomalous transport of self-propelled particles. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 91:062715.
27. Lemons, D. S., and A. Gythiel. 1997. Paul Langevin's 1908 paper "On the theory of Brownian motion". *Am. J. Phys.* 65:1079–1081.
28. Schienbein, M., and H. Gruler. 1993. Langevin equation, Fokker-Planck equation and cell migration. *Bull. Math. Biol.* 55:585–608.
29. Stokes, C. L., D. A. Lauffenburger, and S. K. Williams. 1991. Migration of individual microvessel endothelial cells: stochastic model and parameter measurement. *J. Cell Sci.* 99:419–430.
30. Liu, Y.-P., X.-C. Zhang, …, J.-W. Shuai. 2017. Derivation of persistent time for anisotropic migration of cells. *Chin. Phys. B.* 26:128707.
31. Liu, Y.-P., X. Li, …, J.-W. Shuai. 2020. Motile parameters of cell migration in anisotropic environment derived by speed power spectrum fitting with double exponential decay. *Front. Phys.* 15:13602.
32. Vestergaard, C. L., J. N. Pedersen, …, H. Flyvbjerg. 2015. Estimation of motility parameters from trajectory data. *Eur. Phys. J. Spec. Top.* 224:1151–1168.

33. Vestergaard, C. L., P. C. Blainey, and H. Flyvbjerg. 2014. Optimal estimation of diffusion coefficients from single-particle trajectories. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 89:022726.

34. Gorelik, R., and A. Gautreau. 2014. Quantitative and unbiased analysis of directional persistence in cell migration. *Nat. Protoc.* 9:1931–1943.

35. Pedersen, J. N., L. Li, …, H. Flyvbjerg. 2016. How to connect time-lapse recorded trajectories of motile microorganisms with dynamical models in continuous time. *Phys. Rev. E.* 94:062401.

36. Wu, P.-H., A. Giri, and D. Wirtz. 2015. Statistical analysis of cell migration in 3D using the anisotropic persistent random walk model. *Nat. Protoc.* 10:517–527.

37. Bergman, A. J., and K. Zygourakis. 1999. Migration of lymphocytes on fibronectin-coated surfaces: temporal evolution of migratory parameters. *Biomaterials.* 20:2235–2244.

38. Metzner, C., C. Mark, …, B. Fabry. 2015. Superstatistical analysis and modelling of heterogeneous random walks. *Nat. Commun.* 6:7516.

39. Maiuri, P., J. F. Rupprecht, …, R. Voituriez. 2015. Actin flows mediate a universal coupling between cell speed and cell persistence. *Cell.* 161:374–386.

40. Wu, P.-H., A. Giri, …, D. Wirtz. 2014. Three-dimensional cell migration does not follow a random walk. *Proc. Natl. Acad. Sci. USA.* 111:3949–3954.

41. Uhlenbeck, G. E., and L. S. Ornstein. 1930. On the theory of the Brownian motion. *Phys. Rev.* 36:0823–0841.

42. Cohen, L. 1998. Generalization of the Wiener-Khinchin theorem. *IEEE Signal Process. Lett.* 5:292–294.

43. Leibovich, N., A. Dechant, …, E. Barkai. 2016. Aging Wiener-Khinchin theorem and critical exponents of 1/fβ noise. *Phys. Rev. E.* 94:052130.

44. Huang, R. X., I. Chavez, …, E. L. Florin. 2011. Direct observation of the full transition from ballistic to diffusive Brownian motion in a liquid. *Nat. Phys.* 7:576–580.

45. Shannon, C. E. 1948. A mathematical theory of communication. *Bell Syst. Tech. J.* 27:379–423.

46. Kumar, P., and E. Foufoula-Georgiou. 1997. Wavelet analysis for geophysical applications. *Rev. Geophys.* 35:385–412.

47. Hudgins, L. H. 1995. A Friendly Guide to Wavelets. Gerald Kaiser. Birkhaüser, Boston, 1994. 300 pp. $34.50 hc ISBN 0-8176-3711-7. *Physics Today.* 48:57.

48. Torrence, C., and G. P. Compo. 1998. A practical guide to wavelet analysis. *Bull. Am. Meteorol. Soc.* 79:61–78.

49. Daubechies, I. 1992. Ten lectures on wavelets. *Comput. Phys.* 93:1671.

50. Lau, K. M., and H. Weng. 1995. Climate signal detection using wavelet transform: how to make a time series sing. *Bull. Am. Meteorol. Soc.* 76:2391–2402.

51. Morlet, J., G. Arens, …, D. Giard. 1982. Wave-propagation and sampling theory. 1. Complex signal and scattering in multilayered media. *Geophysics.* 47:203–221.

52. Morlet, J., G. Arens, …, D. Giard. 1982. Wave-propagation and sampling theory. 2. Sampling theory and complex waves. *Geophysics.* 47:222–236.

53. Goupillaud, P., A. Grossmann, and J. Morlet. 1984. Cycle-octave and related transforms in seismic signal analysis. *Geoexploration.* 23:85–102.

54. Grossmann, A., and J. Morlet. 1984. Decomposition of hardy functions into square integrable wavelets of constant shape. *SIAM J. Math. Anal.* 15:723–736.

55. Flandrin, P. 1992. Wavelet analysis and synthesis of fractional brownian-motion. *IEEE Trans. Inf. Theory.* 38:910–917.

56. Farge, M. 1992. Wavelet transforms and their applications to turbulence. *Annu. Rev. Fluid Mech.* 24:395–457.

57. Shyu, H. C., and Y. S. Sun. 2002. Construction of a Morlet wavelet power spectrum. *Multidimens. Syst. Signal Process.* 13:101–111.

58. Lautscham, L. A., C. Kämmerer, …, B. Fabry. 2015. Migration in confined 3D environments is determined by a combination of adhesiveness, nuclear volume, contractility, and cell stiffness. *Biophys. J.* 109:900–913.

59. Bruckner, D. B., A. Fink, …, C. P. Broedersz. 2019. Stochastic nonlinear dynamics of confined cell migration in two-state systems. *Nat. Phys.* 15:595–601.

60. Brückner, D. B., A. Fink, …, C. P. Broedersz. 2020. Disentangling the behavioural variability of confined cell migration. *J. R. Soc. Interface.* 17:0689.

61. Fink, A., D. B. Brückner, …, J. O. Rädler. 2020. Area and geometry dependence of cell migration in asymmetric two-state micropatterns. *Biophys. J.* 118:552–564.

62. Riching, K. M., B. L. Cox, …, P. J. Keely. 2014. 3D collagen alignment limits protrusions to enhance breast cancer cell persistence. *Biophys. J.* 107:2546–2558.

63. Fan, Q., Y. Zheng, …, F. Ye. 2021. Dynamically re-organized collagen fiber bundles transmit mechanical signals and induce strongly correlated cell migration and self-organization. *Angew. Chem. Int.Engl*, Published online February 2, 2021.

64. Liang, L., C. Jones, …, Y. Jiao. 2016. Heterogeneous force network in 3D cellularized collagen networks. *Phys. Biol.* 13:066001.

65. Nan, H., Y. Zheng, …, Y. Jiao. 2019. Absorbing-active transition in a multi-cellular system regulated by a dynamic force network. *Soft Matter.* 15:6938–6945.