Biophysical Journal
Article

Biophysical Society

An entropy-based approach for assessing the directional persistence of cell migration

Yanping Liu,[1,2] Yang Jiao,[3,4] Xinwei Li,[1,2] Guoqiang Li,[5] Wei Wang,[1,2] Zhichao Liu,[1,2] Dui Qin,[1,2] Lisha Zhong,[6] Liyu Liu,[7] Jianwei Shuai,[8,9,10,*] and Zhangyong Li[1,2,*]

[1]Chongqing Key Laboratory of Big Data for Bio Intelligence, Chongqing University of Posts and Telecommunications, Chongqing, China; [2]Department of Biomedical Engineering, Chongqing University of Posts and Telecommunications, Chongqing, China; [3]Materials Science and Engineering, Arizona State University, Tempe, Arizona; [4]Department of Physics, Arizona State University, Tempe, Arizona; [5]Chongqing Key Laboratory of Environmental Materials and Remediation Technologies, College of Chemistry and Environmental Engineering, Chongqing University of Arts and Sciences, Chongqing, China; [6]School of Medical Information and Engineering, Southwest Medical University, Luzhou, China; [7]Chongqing Key Laboratory of Soft Condensed Matter Physics and Smart Materials, College of Physics, Chongqing University, Chongqing, China; [8]Department of Physics, Xiamen University, Xiamen, China; [9]Fujian Provincial Key Laboratory for Soft Functional Materials Research, Xiamen University, Xiamen, China; and [10]Wenzhou Institute, University of Chinese Academy of Sciences, Wenzhou, China

ABSTRACT Cell migration, which is primarily characterized by directional persistence, is essential for the development of normal tissues and organs, as well as for numerous pathological processes. However, there is a lack of simple and efficient tools to analyze the systematic properties of persistence based on cellular trajectory data. Here, we present a novel approach, the entropy of angular distribution , which combines cellular turning dynamics and Shannon entropy to explore the statistical and time-varying properties of persistence that strongly correlate with cellular migration modes. Our results reveal the changes in the persistence of multiple cell lines that are tightly regulated by both intra- and extracellular cues, including Arpin protein, collagen gel/substrate, and physical constraints. Significantly, some previously unreported distinctive details of persistence have also been captured, helping to elucidate how directional persistence is distributed and evolves in different cell populations. The analysis suggests that the entropy of angular distribution-based approach provides a powerful metric for evaluating directional persistence and enables us to better understand the relationships between cellular behaviors and multiscale cues, which also provides some insights into the migration dynamics of cell populations, such as collective cell invasion.

SIGNIFICANCE We present a novel approach that combines cellular turning dynamics with Shannon entropy to explore the directional persistence of cell migration, which helps to accurately analyze the statistical and time-varying characteristics of persistence and elucidate the importance of intra/extracellular cues in regulating migration dynamics. We also further validate the superior utility of the approach in revealing the migration modes of multiple cell lines.

INTRODUCTION

Cell migration is an essential function that affects numerous physiological processes (1–3), including wound healing (4,5), morphogenesis (6,7), immunological responses (8), and organ formation (9). However, the onset of dysregulated migration is commonly associated with human diseases, such as cancer invasion and metastasis (10,11).

In general, cell migration in complex microenvironments is mainly governed by intracellular molecular pathways (12–14) and extracellular environmental cues (15–17) and thus exhibits unique migration behaviors (18,19) and collective phenomena (20,21); for example, chemotaxis regulated by gradients of diffusible factors (22,23) and durotaxis in response to gradients of mechanical stiffness (24,25). More specifically, increased random migration is observed in the zebrafish mib1ta52b mutant, suggesting that at least Mib1 helps to enhance persistent cell migration by regulating the Ctnnd1-Rac1 pathway (26). Similarly, persistent migration is primarily controlled by the microtubule motor Kif1C through the stabilization of an extended cell rear (27), and olfactory neural stem cells with a complex persistence phenotype may lead to an altered brain in schizophrenia (28). In addition, collagen fibers, as major components of the extracellular matrix (ECM), could be

---

Submitted October 10, 2023, and accepted for publication February 13, 2024.

*Correspondence: jianweishuai@xmu.edu.cn or lizy@cqupt.edu.cn

Editor: Padmini Rangamani.

https://doi.org/10.1016/j.bpj.2024.02.010

remodeled by external mechanical stimuli (29-31) and further facilitate cell aggregation and invasion into rigid Matrigel, providing further insight into cancer intravasation (32). Due to the nonlinear mechanical properties of collagen fiber, it can also transmit signals from one cell to another and directly serve as a communication medium (33). As reported in our previous work (30,34), MCF-10A cells can migrate toward each other when the distance between them is less than 120 μm on top of a three-dimensional (3D) collagen gel and show a strongly correlated movement. Significantly, aligned collagen fibers contribute to directed cell migration (29,35,36) and could predict recurrence after ductal carcinoma in situ (37–39). Similarly, physical constraints in the ECM can also force cells to remodel and adopt specialized migration modes (40). For example, invasive MDA-MB-231 cells exhibit a limit cycle in a two-state micropattern, whereas MCF-10A cells exhibit excitable bistable dynamics (41). Cells reversing, following, and sliding past each other have also been observed in this micropattern (42), manifesting the existence of individual differences and helping to explain the diverse roles of cells in collective motility.

To accurately describe the directional persistence (43) of cell migration in complex environments, some novel models have been developed. For example, amoebae perform a random zigzag motion to increase the chances of finding a target (44). Similarly, CD8+ T cell motility in the brain follows a generalized Lévy walk, allowing cells to find rare targets more efficiently (45,46). In addition, a two-state model is constructed that contains persistent and random components and possesses superiority in fitting the cell speed distribution (47,48). Among these models, the persistent random walk (PRW) model is one of the most representative, which incorporates cellular memory of past velocities and could be derived from the Langevin equation (49,50). Subsequently, various models have been proposed based on the PRW model, such as an anisotropic one to explain the effects of anisotropic ECM (51,52) and a time-varying one to describe the local features of the ECM (34,53).

Along with quantitative models, some methods are applied to characterize the features of cell migration. For example, mean squared displacement (MSD) is introduced to quantify the ability of cells to migrate, which is habitually characterized by two motility parameters: persistence time P and migration speed S (48). Moreover, the slope of the MSD on the *log-log* axis could be used to determine how similar the migration is to ballistic/Brownian motion or how strong the persistence is (52,54). In addition to the MSD, some measures, such as the velocity autocovariance function (VAC) (55), the Fourier power spectrum (FPS) (53,56), and the ratio of displacement to trajectory length (48), can be used to estimate the ensemble-averaged characteristics of the parameters, whereas other measures, such as the DiPer program (48), Bayesian inference (57), and wavelet-based approach (53), allow us to uncover the time-dependent characteristics. However, these measures have three main shortcomings: 1) the estimation of motility parameters is usually performed in an ad hoc manner, assuming that the associated dynamical model is adequate to analyze the cellular behavior. However, this approach will fail and result in inaccurate outcomes in instances where novel cell behaviors are not accounted for by the established models (51,58). 2) The derivation of motility parameters primarily involves complicated computations, including the inference and fitting (R^2 analysis) based on experimental data with inherent noise, which inevitably leads to numerical uncertainties that are typically challenging to quantify (52,53,57,59). 3) most of the measures are easily biased, especially by cell speed and the correlation with time (48,56,60). Therefore, measuring directional persistence during migration in real time in a simple and effective manner becomes a significant problem.

Here, we develop a novel approach based on cellular turning dynamics and Shannon entropy to accurately characterize the directional persistence of cell migration, which is rigorously regulated by intra/extracellular cues. The entropy here has a physical interpretation similar to that of Gibbs entropy for thermal systems. It reflects the degree of randomness or order in cell migration dynamics, with a value of 1 representing the most random dynamics (e.g., purely diffusive migration) and 0 representing the most ordered dynamics (e.g., purely ballistic dynamics). Subsequently, we analyzed three types of experimental cell data regarding Arpin protein, collagen gel/substrate, and physical constraints and found that the approach is capable of quantifying the changes in persistence caused by these multiscale cues, especially the nuanced but significant differences that other methods have failed to capture. By analyzing the statistical and time-varying properties of directional persistence, we gain an understanding of how cellular migration modes correlate with these multiscale cues and how directional persistence is distributed and evolves across different cell populations. Therefore, this approach allows us to systematically study the characteristics of directional persistence and further elucidate the dynamic mechanisms underlying cell motility, such as cancer metastasis.

MATERIALS AND METHODS

In vitro cell migration data

In this article, we analyzed three types of cell migration trajectory data affected by the cues at different scales, i.e., *Dictyostelium discoideum* migration regulated by the Arpin protein ($\sim 10^{-2}$ μm (61)), MCF-10A cells migrating on top of a 3D collagen gel (~ 10 μm (62)) and a two-dimensional (2D) substrate, and MDA-MB-231 cells migrating through a microstructured array ($\sim 5 \times 10^2$ μm (63)). The *D. discoideum* data were taken with permission from published work (64), and the MDA-MB-231 cell data were courtesy of Claus Metzner (57,63). Finally, the MCF-10A cell data were generated in our laboratory. See work (30,34) for more details on the experiments.

Cell migration simulated by biophysical model

PRW model

To systematically study the characteristics of cell migration, we introduce a classical biophysical model, the PRW. With this model, the position of a cell at each time step can be easily calculated using the following equations:

$$x(t+\Delta t) = x(t) + \Delta x(t, \Delta t), \tag{1}$$

$$y(t+\Delta t) = y(t) + \Delta y(t, \Delta t), \tag{2}$$

where Δt is the time step, x and y are the coordinates of the cell in the 2D plane, and Δx and Δy are the cell displacements in the time step of Δt. Here, the displacements are further written as

$$\Delta x(t, \Delta t) = \alpha \bullet \Delta x(t - \Delta t, \Delta t) + F \bullet W, \tag{3}$$

$$\Delta y(t, \Delta t) = \alpha \bullet \Delta y(t - \Delta t, \Delta t) + F \bullet W, \tag{4}$$

where the displacements on the left side are the consequences of those on the right side in time step of Δt, $\alpha = 1 - \Delta t / P$ indicates the memory of the cell to past velocities, and $F = \sqrt{S^2 \bullet \Delta t^3 / P}$ quantifies the amplitude of the white noise $W \sim \mathcal{N}(0,1)$. In particular, the case where α approaches 1 and F approaches 0 when $P \gg \Delta t$ corresponds to ballistic motion, whereas the case where α approaches 0 and F approaches $S \bullet \Delta t$ when $P \sim \Delta t$ corresponds to Brownian motion.

To account for the effects of experimental measures, we also further add positioning errors to the simulated trajectory data, as given by

$$\widehat{x}(t) = x(t) + \sigma_{pos} \bullet W, \tag{5}$$

$$\widehat{y}(t) = y(t) + \sigma_{pos} \bullet W, \tag{6}$$

where σ_{pos} is set to 0.01 μm (60). In all computer simulations, we set $\Delta t = 0.2$ min, $S = 0.3$ μm/min, and total time $T = 960$ min.

Time-varying PRW

Since cell migration is generally regulated by intra/extracellular cues, including Arpin protein, biochemical factors, and geometric structures, the trajectory data always possess time-dependent characteristics, such as the transition from Brownian to ballistic motion. Inspired by the time-dependent characteristics, we modify the PRW model and obtain a time-varying PRW (TPRW) model. The details of the simulations are as follows: 1) determining the functions, i.e., $P(t)$ and $S(t)$, of the motility parameters versus time based on the time-varying characteristics of the experimental migration trajectories, such as continuously increasing, decreasing, or first increasing and then decreasing, etc.; 2) calculating the time-varying factors, i.e., $\alpha(t)$ and $F(t)$, based on the equations $\alpha(t) = 1 - \Delta t / P(t)$ and $F(t) = \sqrt{S(t)^2 \bullet \Delta t^3 / P(t)}$, respectively; 3) performing the simulations according to the same procedures as those for the PRW model, especially updating the values of the motility parameters according to their functions at each time step. In this work, we designed four classes of persistence changes based on the different changing features of the migration trajectories in the Supporting Material, i.e., the constant, the linear, the abrupt, and the nonlinear persistence changes in Fig. 4, to explore the effects of directional persistence on migration dynamics.

Conventional evaluations of cell migration dynamics

Statistical analysis of cell migration velocity

For experimental or simulated trajectory data $\vec{r}_i$, the migration velocity $\vec{v}_i$ can be computed from the displacement at each time step, namely $\vec{v}_i = (\vec{r}_i - \vec{r}_{i-1}) / \Delta t$. Subsequently, we analyze two aspects of these velocities: 1) velocity vs. time for estimating the stationarity, and 2) the distribution of squared velocities for learning the statistical features.

Correlation of angular displacement with velocity

Based on the migration velocities, we first calculate the angular displacement (angle) between any two consecutive velocity vectors using the following equation:

$$\theta = cos^{-1}\left(\frac{\vec{v}_i \bullet \vec{v}_{i+1}}{\left|\vec{v}_i\right| \bullet \left|\vec{v}_{i+1}\right|}\right). \tag{7}$$

Second, we map the θ into the close interval of $[-\pi, \pi]$ by referring to the deflection (left or right) of the second vector relative to the first vector. Finally, we record the second vector and plot the velocity vs. angle profile to reveal the correlation between angular displacement and velocity.

MSD evaluation

In addition to the statistical analysis of the migration velocities, we also calculate the MSD based on the trajectory data $\vec{r}_i$ using the following equation:

$$MSD(\tau) = \frac{1}{N-\tau+1}\sum_{i=0}^{N-\tau}\left(\vec{r}_{i+\tau} - \vec{r}_i\right)^2, \tag{8}$$

where N is the total number of cell positions for individual trajectories, and τ is the lag for time. By averaging the MSDs for all trajectories, we obtain an ensemble-averaged MSD profile that quantifies the ability of cells to migrate. Notably, one could extract the slope α of the MSD profile on the *log-log* axis and compare it to $\alpha = 1$ (or 2) to estimate how similar the migration behavior is to Brownian (or ballistic) motion.

VAC evaluation

Similarly, according to the migration velocities above, the VAC is also computed by

$$VAC(\tau) = \langle \vec{v}_i \bullet \vec{v}_{i+\tau} \rangle = \frac{1}{N^* - \tau - 1}$$

$$\sum_{k=1}^{N^*-\tau}\left(\vec{v}_k - \frac{1}{N^*-\tau}\sum_{l=1}^{N^*-\tau}\vec{v}_l\right)\left(\vec{v}_{k+\tau} - \frac{1}{N^*-\tau}\sum_{l=\tau+1}^{N^*-\tau}\vec{v}_l\right), \tag{9}$$

where $N^* = N - 1$ is the total number of the migration velocities. With the VAC evaluation, one could not only directly estimate the degree of correlation (or persistence) but also determine the migration modes encoded by the linearity of the VAC profile. If the VAC profile on the *log-lin* axis could be well fitted by a linear line, it illustrates that it follows a single-exponential decay, and the migration behavior satisfies the PRW model to some extent; otherwise, a double-exponential decay may be at least superior, and the migration behavior may not be isotropic but anisotropic or heterogeneous (65).

FPS evaluation

It has been reported that the above two evaluations are strongly correlated with time, so they may not provide more accurate results. Therefore, another evaluation, FPS, is proposed to transform the VAC from the time domain to the frequency domain by Fourier transform according to the

Wiener-Khinchin theorem (66), which states that the power spectrum of any generalized stationary random process is the Fourier transform of its autocovariance function. The theorem requires that the time series (i.e., velocities) are stable with time, and thus the stable migration velocities generated by the PRW model can be computed directly via this theorem. However, for the unstable velocity series generated by the TPRW model, we will compute the FPS profile according to the definition of the power spectrum, i.e., the squared magnitude of a given Fourier series coefficient.

3D distribution of angular displacements

Based on the migration velocity generated by the biophysical models above, we first calculate the angle between any two consecutive velocity vectors (i.e., $\tau = 1$) and further map it to the interval of $-\pi \sim \pi$. Second, we divide the above interval into n bins and count the number of angles in each bin for a frequency diagram. Then, we normalize the diagram and subsequently obtain a 2D probability density function (PDF). Finally, we further compute PDFs when $\tau = 2, 3, \cdots$ and get a 3D PDF regarding angular displacements and time lag (i.e., $\tau * \Delta t$).

Entropy of angular distribution

According to the definition of directional persistence, it is evident that the correlation between any two velocities will decrease as the τ increases; therefore, the shape of the PDF will transition from an uneven "peak" to a uniform "plain." To characterize this change in shape quantitatively, we further introduce the Shannon entropy with the following form:

$$H = -\sum_{i=1}^{n} p_i \bullet log2\,(p_i), \tag{10}$$

where n is the total number of events (equivalent to the bins) and p_i are the probabilities of the random events occurring. Note that the H is maximal $log\,2(n)$ when p_i are equal to each other, i.e., $p_i = 1/n$, meaning that we are not completely sure of the result. Furthermore, we substitute the p_i with a single PDF for a given τ and thus obtain a value of H. After performing the same calculations for all τ, we obtain the relationship between H and τ. Since the H is strongly correlated with the bin size n, we normalize all H by dividing the maximum H to eliminate this correlation. As a consequence, all H are rescaled to the close interval of [0, 1], among which $H = 1$ corresponds to an absolutely uniform PDF (i.e., the most random state, or purely random dynamics) and $H = 0$ corresponds to an absolutely uneven PDF (i.e., the most ordered state, or purely ballistic dynamics). We use the term entropy of angular distribution (EAD) to refer to the normalized H in this paper to avoid confusion.

Correlation between EAD, time lag, and persistence

To better illustrate the correlation between EAD, time lag, and persistence, we further simulate cell migration trajectories using the PRW model with different prescribed persistence, then compute the EAD at each time lag following the above procedures, and finally obtain EAD profiles with respect to time lag and the prescribed persistence. Based on these profiles, we mainly perform two types of analyses: the effects of time lag and persistence on the evolution of EAD profiles.

Joint parameter distribution

Although we can clarify the correlation between EAD, time lag, and persistence based on the datasets generated by the PRW/TPRW models, it is actually infeasible because we have no prior knowledge of the persistence encoded in the experimental data. With this consideration, we propose a joint parameter with two components to more accurately characterize the properties of persistence. The first parameter is the time lag (i.e., TL_1) when EAD $= 1$, which indicates the time it takes for an uneven PDF to become uniform. In contrast, the second parameter is defined as the EAD (i.e., EAD_1) when $\tau = 1$, which indicates the degree of the unevenness of the PDF at the first time lag. Thus, we can combine these two parameters (TL_1, EAD_1) to jointly quantify the directional persistence of cell migration. With the joint parameter's distribution (JPD), we can analyze the distribution of directional persistence for a cell population and reveal the differences between multiple cell populations. Note that the value of TL_1 is determined when the EAD is theoretically equal to 1, but most likely the EAD could be less than 1 due to the effects of errors, noise, and so on. That is, the EAD is still less than 1 when it becomes stable. Thus, we will use the stable EAD to replace the theoretical 1 above when determining the value of TL_1. See the experimental sections below for real examples.

Time-varying EAD

In addition to the statistical profiles of the directional persistence, it is also essential to reveal the time-dependent features of the directional persistence due to the local effects of the ECM. Next, we compute the time-varying EAD by drawing on the sliding-window algorithm employed in the windowed Fourier transform. First, we determine the window length L_w for a given angle series obtained from velocity series and compute the first EAD based on the first window ($1 \sim L_w$). Second, we slide the window one step forward and calculate the second EAD based on the second window ($2 \sim L_w + 1$). Then, we iteratively perform the sliding and computation. Finally, we obtain an EAD profile with time changes. Note that the performance of the windowed Fourier transform is significantly correlated with the length of the window; i.e., a short window leads to inaccurate results, whereas a long window leads to poor temporal resolution. In this study, we determine the window length according to the condition that the angles contained in a window can form a distribution with an obvious peak. Although the shortcomings still exist, the approach allows us to explore the main time-dependent features of directional persistence, especially the significant transition. Therefore, the algorithm for the time-varying EAD makes sense when utilized to analyze the time-varying features of migration data.

Correlation analysis

The correlation coefficient reflects the degree of correlation between two variables. If the data are continuous numerical variables, and both satisfy normality (or have obvious single peaks), the Pearson coefficient is preferred, and, if the data do not satisfy the normality after bijective transformation, the Spearman or Kendall coefficients are optional.

RESULTS

Conventional evaluations of cell migration simulated by the TPRW model

After performing simulations using the TPRW model, we obtain trajectory data of cell migration that are mainly dominated by the prescribed persistence time. Fig. 1 *A* exhibits a representative trajectory of individual cells, and it seems that the trajectory has different directionality (or straightness) by referring to the compactness (or density) of the points, i.e., high tortuosity (dense) on the lower-left and low tortuosity (sparse) on the upper-right. The compactness may denote the changing trend of persistence from

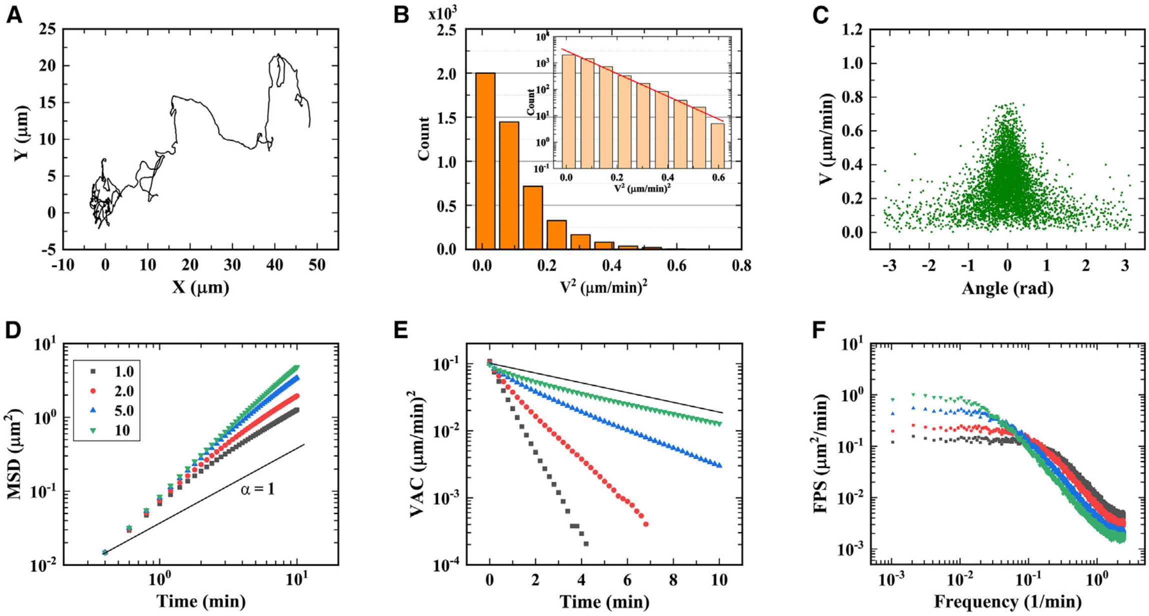

FIGURE 1 Conventional evaluations of cell migration simulated by the TPRW model. (*A*) A representative migration trajectory with a time-varying persistence linearly ranging from 0.5 to 10 min. (*B*) Distribution of the squared velocities. The inset shows the same results on the *log-lin* axis. (*C*) Velocity as a function of angular displacement. The results in (*B and C*) are derived from the data in (*A*). (*D*) MSD profiles for the four classes of prescribed persistence. They increase linearly from 0.5 to 0.5, 1.0, 2.0, and 5.0 min. (*E*) VAC profiles. (*F*) FPS profiles. The sample size (i.e., the number of simulated trajectories) is $n = 100$ for each ensemble-averaged profile in (*D*)–(*F*). The three straight lines in (*B*), (*D*), and (*E*) are auxiliary for enabling us to better understand the changes. To see this figure in color, go online.

weak to strong, which is also consistent with the gradually increasing persistence prescribed in the model (i.e., 0.5–10 min). Subsequently, the squared velocities are calculated, and the corresponding histogram is plotted in Fig. 1 *B*. Evidently, the counts gradually decrease from $\sim 2 \times 10^3$ to 0, and this progress follows an exponential decay as presented by the red line in the inset. Moreover, the migration velocities are also correlated with the angular displacements, as exhibited in Fig. 1 *C*. The large velocities are mainly concentrated around 0 rad, indicating that the faster migrating cells prefer to migrate in the same direction as before, whereas the slower migrating cells prefer to turn. Thus, large velocities are coupled with stronger persistence to some extent, consistent with the previous report (67).

In the following, we first calculate MSDs for trajectory data governed by four classes of persistence parameters that increase linearly from start to end, i.e., [start, end] = [0.5, 0.5], [0.5, 1.0], [0.5, 2.0], and [0.5, 5.0]. These parameter designs are mainly inspired by real situations (see Supporting Material), e.g., a cell migrating toward a nutrient source or through a structure with gradually increasing constraints. The results indicate that all MSDs grow with time, and the differences between them become more and more pronounced (Fig. 1 *D*). Specifically, the MSD for the first class grows with a slope greater than 1 (i.e., $\alpha > 1$) when time is less than ~2 min and the α decreases to a value close to 1 (i.e., $\alpha \sim 1$) when time is greater than ~2 min. As the end of persistence increases to 5.0, both types of α increase and gradually approach 2.

Similarly, the VACs also behave differently as the end of persistence increases (Fig. 1 *E*). For example, the VAC for the first class decreases as time gradually increases, and this trend could be well fitted by a straight line on the *log-lin* axis, indicating that the VAC follows a single-exponential decay. In contrast, the VAC for the fourth class is nonlinear, meaning that it cannot be well fitted by a single-exponential decay but at least by a double-exponential function.

Finally, we further calculate all FPS profiles for the classes (Fig. 1 *F*), which essentially shows the distribution of the migration modes concerning frequency. Among them, the FPSs in the low-frequency interval (less than ~0.08/min) correspond to correlation (or behavioral mode) on the long timescale, whereas the counterparts of high frequency correspond to randomness (or white noise) on the short timescale. Thus, the migration behavior has a higher correlation when the FPSs in the low-frequency interval are much larger than those in the high-frequency interval. As shown in Fig. 1 *F*, the differences gradually become more apparent as the end of persistence increases.

EAD as an indicator of directional persistence

To vividly display the 3D PDFs of angular displacements, especially the changes in the shape of the PDF, we randomly select a migration trajectory from the data simulated by the PRW model with a persistence of 10 min, and we further compute the PDF at each time lag ranging from 0.2 to 40 min. Then, all the PDFs are stacked along the time-lag

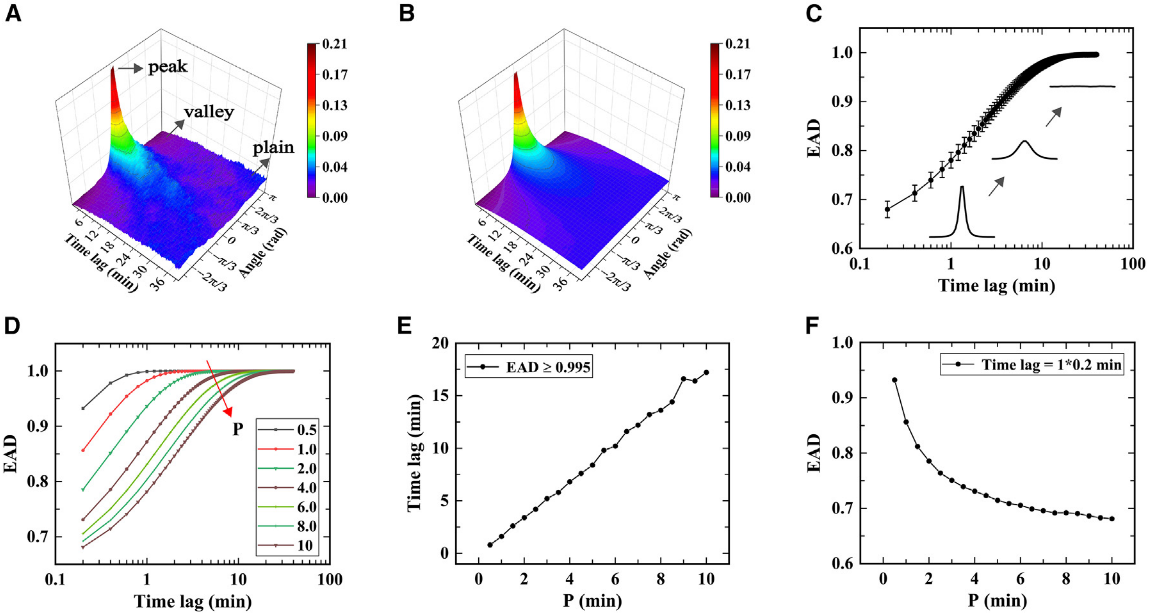

FIGURE 2 The EAD approach quantifies the directional persistence of cell migration. (*A*) 3D PDF of angular displacements based on individual trajectories simulated by the PRW model with a persistence of 10 min. (*B*) Ensemble-averaged 3D PDF of angular displacements. (*C*) EAD as a function of time lag. The result corresponds to the PDF in (*B*) and is presented as mean ± SD. (*D*) The effect of the prescribed persistence on the EAD profiles. The persistence parameter increases from 0.5 to 10 min, as indicated by the red arrow. (*E*) Relationship between time lag (TL_1) and prescribed persistence when EAD ≥ 0.995. (*F*) Relationship between EAD (EAD_1) and prescribed persistence when time lag $= 0.2$ min (i.e., $\tau = 1$). The bin size is 36 for the calculation of the PDF, and the sample size is $n = 100$ for the ensemble-averaged PDF. To see this figure in color, go online.

axis to form the final 3D distribution, as shown in Fig. 2 *A*. It is clear that, when the time lag is small ($<\sim$ 12 min), the PDF possesses larger probabilities around 0 rad and smaller around $\pm\pi$ rad, forming a trend of higher in the middle (peak) and lower at both ends (valley). As the time lag increases, the peaks become lower and the valleys become higher. Eventually, this change causes the peaks and valleys to be basically equal in height, forming a plain as a whole. To further study the ensemble characteristics of the 3D PDF for a cell population, we average the 3D PDFs for all trajectories in that population and obtain the ensemble-averaged 3D PDF (Fig. 2 *B*). In general, the ensemble-averaged PDF is highly similar to the PDF in Fig. 2 *A*, preliminarily indicating that the cells in this population follow similar migration dynamics.

Next, we compute the EAD along the time-lag axis of each PDF (e.g., Fig. 2 *A*) and obtain EAD profiles for all trajectories as well as the ensemble-averaged EAD profile (Fig. 2 *C*). The results indicate that the averaged EAD has the smallest value (0.680) when the time lag $= 0.2$ min (i.e., $EAD_1 = 0.680$) and gradually increases until the EAD reaches 0.995 (i.e., $TL_1 = 24$ min). Note that the value of 0.995 is artificially determined, which does not mean that the EAD will stop increasing when EAD ≥ 0.995, just that it will increase slowly. According to our previous understanding of the correlations encoded in VAC and combined with the 3D PDF, it is easy to deduce that the small EAD means strong correlation (i.e., persistence), whereas the large EAD means weak correlation (i.e., randomness). Thus, the averaged EAD in Fig. 2 *C* shows the transition from the state with strong persistence to the state with random property. Moreover, the error bars also gradually decrease, which we believe is caused by the high sensitivity of the persistence state and the low sensitivity of the random state to noise.

To further validate the above conclusion, we again simulate cell migration using the PRW model with different persistence parameters and obtain the corresponding EAD profiles, as shown in Fig. 2 *D*. As the persistence parameter increases from 0.5 to 10 min, the EAD profiles gradually move to the lower-right region of the panel (see the red arrow), and cause the joint parameter (TL_1, EAD_1) to behave differently; e.g., the TL_1 increases linearly (Fig. 2 *E*) and the EAD_1 decreases nonlinearly (Fig. 2 *F*).

Statistical characteristics of directional persistence for cell populations

Besides the application of the joint parameter in characterizing the directional persistence of individual trajectories, it also allows us to reveal the statistical characteristics of the directional persistence for cell populations, as shown in Fig. 3. It is evident that there are three clusters (or peaks) in Fig. 3 *A*, and their characteristics are as follows: 1) the first cluster is mainly located at (1.860, 0.856), i.e., the joint parameter (TL_1, EAD_1) of the center of the peak, and possesses a smaller region; 2) the second cluster is located at (3.698, 0.780) and possesses a region similar in size to that of the first cluster; and 3) the third cluster is located

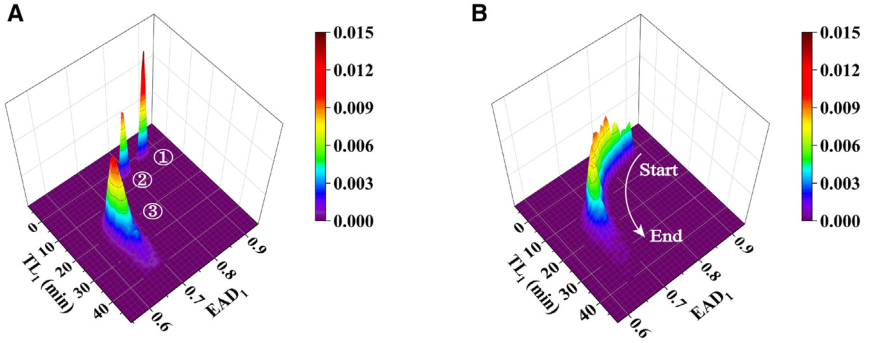

FIGURE 3 Normalized 3D JPDs of TL_1 and EAD_1. The two distributions show the discrete (*A*) and continuous (*B*) changes in persistence for two cell populations simulated by the TPRW model. The former corresponds to the prescribed three clusters of persistence, i.e., $\mathcal{N}_1 \sim (1.0, 0.002)$, $\mathcal{N}_2 \sim (2.0, 0.002)$, and $\mathcal{N}_3 \sim (10.0, 0.002)$, and their sample sizes are 5×100, 5×100, and 90×100 for better visualization. However, the latter is calculated based on 100 distributions (i.e., $\mathcal{N}_n$, $n = 1$–100) with the means ranging from 1.0 to 10.0 min, and each distribution corresponds to a sample size of 100. To see this figure in color, go online.

at (17.483, 0.685) and possesses a larger region. Here, the TL_1 of the three clusters gradually increases from 1.860 to 17.483, and the EAD_1 gradually decreases from 0.856 to 0.685, illustrating that the third cluster corresponds to stronger persistence, followed by the second and the first. The above results are also consistent with the three types of Gaussian distributions of persistence prescribed in the PRW model, which are $\mathcal{N}_1 \sim (1.0, 0.002)$, $\mathcal{N}_2 \sim (2.0, 0.002)$, and $\mathcal{N}_3 \sim (10.0, 0.002)$, respectively.

Unlike the three discrete clusters in Fig. 3 *A*, JPD can also reveal continuous changes in persistence, as shown in Fig. 3 *B*. The distribution covers a large region, which starts with the state with random property around (1.855, 0.869) and ends up with the state with strong persistence around (24.592, 0.677). The results also conform to the prescribed persistence; i.e., the mean of the distributions ranges from 1.0 to 10.0 with an increment of 0.2, and the variance is 0.002.

Time-varying persistence of compliance with multiple trends

To test the performance of the approach used to compute time-varying persistence, we design four cases of persistence changes in this section. These cases are, respectively, 1) the constant persistence change, i.e., $P = 0.5$ min; 2) the linear persistence change, i.e., P increases from 0.5 to 1.0 min; 3) the abrupt persistence change, i.e., $P = 0.5$ in the first half of the simulation time and 1.0 min in the second half; and 4) the nonlinear persistence change, i.e., P first increases and then decreases according to a sine function.

Subsequently, the time-varying EAD is calculated for each case, as exhibited in Fig. 4. For the first case, the EAD is basically stable at around 0.890, indicating that the encoded persistence is generally unchanged and weak (Fig. 4 *A*). However, for the second case, the EAD gradually decreases from 0.887 to 0.797, implying that the persistence becomes stronger (Fig. 4 *B*). Different from the above linear changes in persistence, the sliding-window algorithm also exhibits good performance in processing nonlinear changes. For example, the EAD for the third case first starts at a high level and remains stable at around 0.890 and then drops to a low level and stabilizes at around 0.790, which to some extent reflects the transition of persistence (Fig. 4 *C*). This transition is also very similar to that of a single MDA-MB-231 cell migrating through a microstructured array (see Fig. 7 for more details). Moreover, the EAD for the last case first decreases from 0.880 to 0.779 and then increases in the opposite direction, indicating a nonmonotonic trend of persistence change (Fig. 4 *D*). Although the above four cases cannot represent all complex situations in natural conditions, they represent the main changes in persistence, including linearity, continuity, and monotonicity. Therefore, it is reasonable to assume that the approach is practical and robust. In the Supporting Material, we also explore the advantages of this EAD approach over some existing classical models, including the PRW, APRW, and TPRW models (see Fig. S1 and Table S1 for further analysis).

The EAD approach uncovers the critical role of the Arpin protein in governing the directional persistence of *D. discoideum* migration

With the biophysical insights encoded in EAD having been understood from the study of trajectory data simulated by the PRW/TPRW models in the previous sections, we now employ this EAD-based approach to investigate migration dynamics in terms of directional persistence regulated by intracellular proteins. In this section, all trajectory data are taken from the work of Dang et al. (64), which identifies a novel protein, Arpin. The Arpin inhibits the Arp2/3 complex in vitro and antagonizes an intrinsic positive feedback loop that sustains lamellipodial protrusion. In this work, we analyze the motility of *D. discoideum* amoebae, which contain three groups for experiment and control, i.e., wild-type (WT) amoebae, Arpin-knockout (KO) amoebae, and rescued by green fluorescent protein (GFP)-Arpin expression in knockout amoebae (rescue)

Fig. 5 shows comparisons of the results based on the EAD analysis for the three groups. On the one hand, the PDF of the WT amoeba (Fig. 5 *A*) has not only the intermediate-large values (~0.17) marked in red but also the intermediate-small values (~0.07) marked in blue when the time lag is less than ~5/6 min. In contrast, both the number of large and small values in the PDF of the WT amoeba are smaller than those

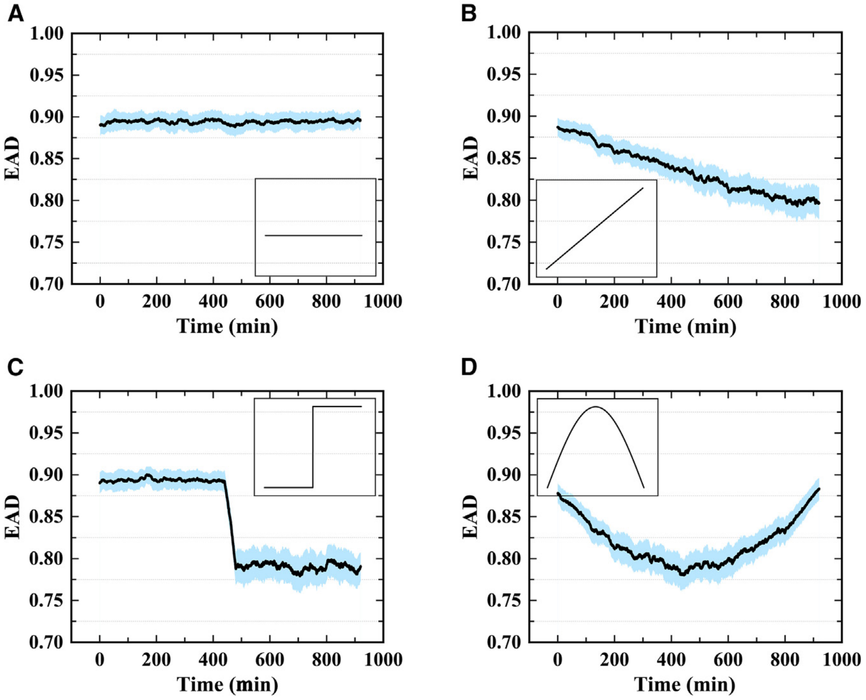

FIGURE 4 Time-varying EAD profiles reveal multiple changing trends of persistence prescribed in the TPRW model. The EAD profiles show the constant persistence changes (*A*), the linear persistence changes (*B*), the abrupt persistence changes (*C*), and the nonlinear persistence changes (*D*), respectively. The data are presented as mean ± 5×SE and indicated by the black lines and blue regions; the sample size is $n = 100$ for the ensemble-averaged EAD in each case. To see this figure in color, go online.

of the KO amoeba (Fig. 5 *D*), which has the most large ($\sim$0.23 marked in crimson) and small values ($\sim$0.05 marked in purple), and larger than those of the rescue amoeba (Fig. 5 *G*), which has the fewest large($\sim$0.13 marked in yellow) and small values ($\sim$0.09 marked in light blue). On the other hand, the number of values marked in green ($\sim$0.10) for the KO amoeba is significantly larger than those for the WT amoeba and the rescue amoeba, respectively, when the time lag is greater than $\sim$5/3 min. Overall, the PDF of the KO amoeba is more concentrated at around 0 rad, followed by the WT amoeba and the rescue amoeba. These characteristics qualitatively illustrate that the correlation encoded in the 3D PDF of the KO amoeba is stronger compared to the other two groups.

In terms of quantitative analysis, all EAD profiles of the three groups gradually increase and then remain stable. Nevertheless, there are still obvious differences that can be used to preliminarily estimate the degree of persistence. For example, when the time lag is 5/60 min (i.e., $\tau = 1$), the EAD is 0.942 ± 0.007 (mean ± SE) for the WT amoeba (Fig. 5 *B*), and it is significantly greater than 0.883 ± 0.009 for the KO amoeba ($p < 0.001$, *t*-test; Fig. 5 *E*) and less than 0.964 ± 0.004 for the rescue amoeba ($p < 0.01$, *t*-test; Fig. 5 *H*). Moreover, the time lag required for stabilization is about 30/60 min (i.e., $\tau = 6$) for the WT amoeba, whereas it is about 40/60 min (i.e., $\tau = 8$) for the KO amoeba and 25/60 min (i.e., $\tau = 5$) for the rescue amoeba. The results highlight the differences in directional persistence from an ensemble averaging perspective. In the Supporting Material, we also analyze the time-varying dynamics of individual trajectories and further illustrate the heterogeneity of cell population dynamics (see Fig. S2 for more details).

To further explore the details of directional persistence, we compare the normalized JPDs of these three groups. First, the JPDs mainly contain single clusters with the centers located at (0.354, 0.967) for the WT amoeba (Fig. 5 *C*) and (0.395, 0.977) for the rescue amoeba (Fig. 5 *I*), respectively. However, the JPD of the KO amoeba (Fig. 5 *F*) consists mainly of three clusters: 1) the first cluster is located at (0.838, 0.907) and has a large region (279,129, pixel^2); 2) the second cluster is located at (2.222, 0.882) and covers a relatively small area (150,268, pixel^2); and 3) the third cluster is shaped like a long strip (141,415, pixel^2) covering from $\sim$0.70 to $\sim$0.95 along the EAD$_1$ axis. In addition, the cluster for the WT amoeba has a larger region size than that for the rescue amoeba, i.e., 190,187 (pixel^2) for the former and 139,691 (pixel^2) for the latter, but both of them are smaller than 279,129 (pixel^2) of the first cluster for the KO amoeba. Here, the region areas are measured using the "Analyze>Measure" function in ImageJ. Note that the points at $TL_1 = 4.0$ min are the consequence of the issue that the actual TL_1 of some cells cannot be determined in the prescribed interval because all values in individual EAD profiles are smaller than the value used to determine the TL_1.

The EAD approach assesses the directional persistence of MCF-10A cell migration across varying conditions

In this section, we continue to investigate the characteristics of directional persistence of MCF-10A cells migrating on top of a 3D collagen gel (collagen) and a 2D Petri dish

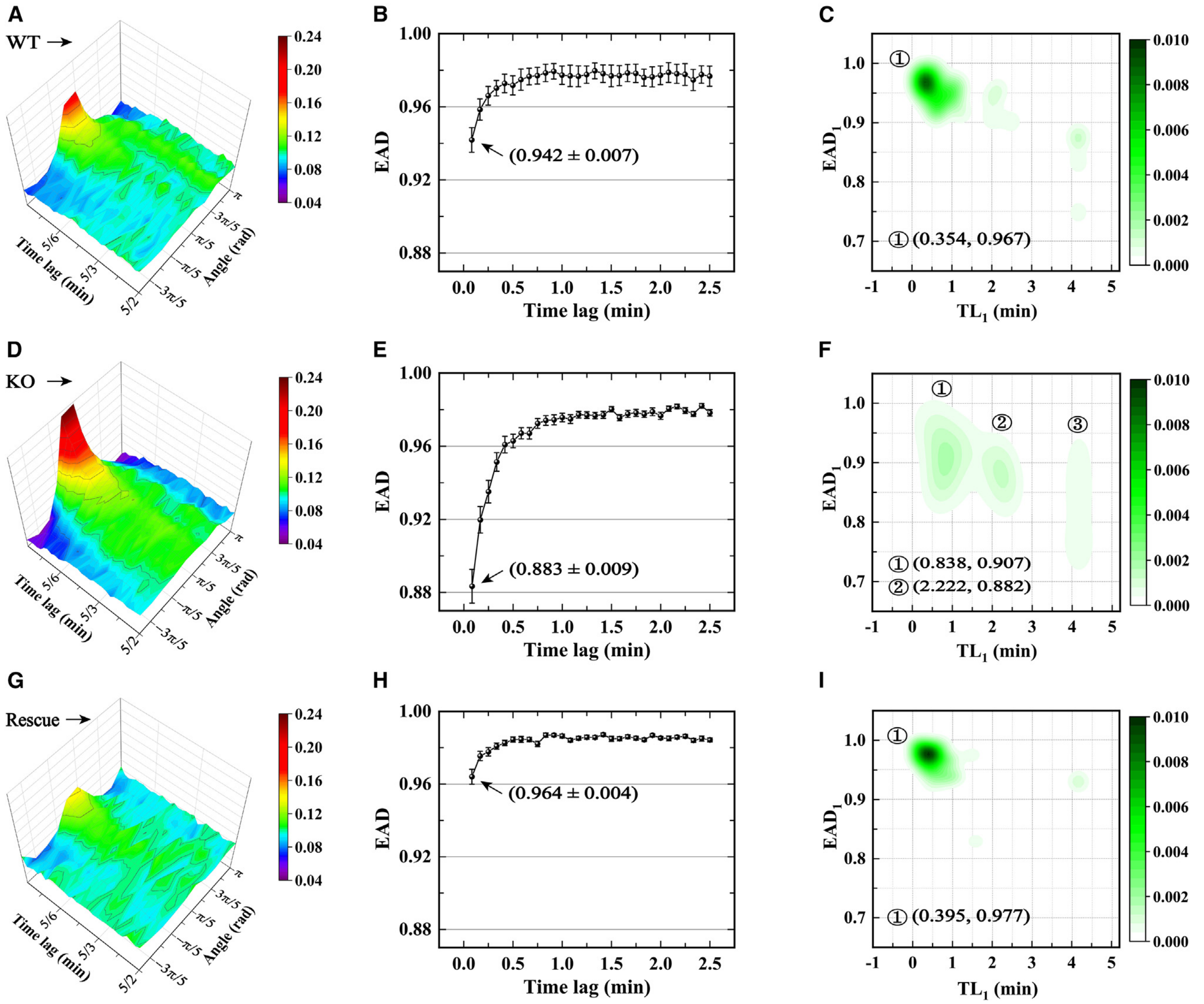

FIGURE 5 Arpin protein weakens the directional persistence of *D. discoideum* migration. (*A*) 3D PDF of angular displacements for the WT amoeba. (*B*) EAD as a function of time lag. (*C*) Normalized 2D JPD of TL_1 and EAD_1. Note that the results in (*B*) and (*C*) are derived from the data in (*A*). The other captions are the same as those in (*A*)–(*C*), but for the KO amoeba (*D*–*F*) and the rescue amoeba (*G*–*I*). The bin size is 10 for calculating the PDF; the data in (*B*), (*E*), and (*H*) are presented as mean ± SE; the sample sizes are $n = 43$, 38, and 45 for the WT, KO, and rescue amoeba, respectively; and the $EAD \geq 0.99$ for determining the TL_1. Trajectory data were taken with permission from the work (64). To see this figure in color, go online.

(substrate). See our previous work (30,34) for more details on the in vitro experiments.

Following the above procedures, we analyze the directional persistence of MCF-10A cell migration under different conditions, and the results are shown in Fig. 6. The 3D PDF of the collagen condition (Fig. 6 *A*) has more small values (~0.01) marked in purple when the time lag is less than ~40 min, in contrast to the less small values for the substrate condition (Fig. 6 *D*). Moreover, the large values (~0.03) marked in light blue for the collagen condition are more than those for the substrate condition when the time lag is greater than ~40 min. These features show that the PDF of the collagen condition is more concentrated at around 0 rad, meaning that the corresponding correlation is stronger when compared to the substrate condition.

Furthermore, the EAD profile of the collagen condition (Fig. 6 *B*) is integrally smaller than that of the substrate condition (Fig. 6 *E*), which is well quantified by the mean (0.892 ± 0.009) of the former and that (0.906 ± 0.008) of the latter, except for the values at time lag = 2 and 4 min (i.e., $\tau = 1$ and 2). More specifically for the exception, the EADs of the former are 0.793 ± 0.008 and 0.859 ± 0.008, both of which are greater than 0.781 ± 0.013 and 0.851 ± 0.012 of the latter. Although the term "greater" is not entirely infallible because of the error bars ($p > 0.5$, Wilcoxon rank sum test), it at least shows that some cells in the substrate condition possess the strongest persistence

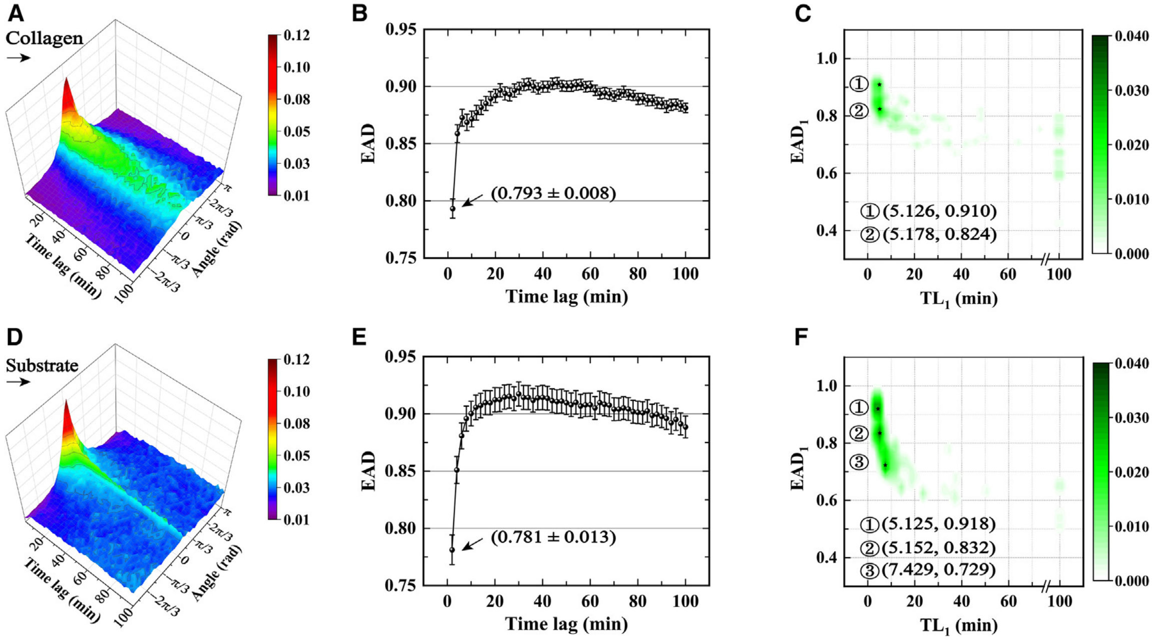

FIGURE 6 Collagen enhances the directional persistence of MCF-10A cell migration. The captions are the same as those in Fig. 5*A*–5*C*, but the results correspond to the migration on 3D collagen gel (*A*–*C*) and 2D Petri dish (*D*–*F*). The bin size is 36 for calculating the PDF; the data in (*B*) and (*E*) are presented as mean ± SE; the sample sizes are $n = 107$ and 105 for the collagen and the substrate conditions, respectively; and the EAD ≥ 0.9 for determining the TL_1. To see this figure in color, go online.

at the small scale of time lag. Interestingly, as the time lag gradually increases, the EAD of the collagen condition increases slowly and takes a total of 30 min (i.e., $\tau = 15$) to reach 0.90, whereas the EAD of the substrate condition increases rapidly and takes 10 min (i.e., $\tau = 5$). The above results illustrate 1) the directional persistence for the collagen condition is slightly weaker than that for the substrate condition at the small scale of time lag ($\leq$4 min), but 2) the persistence for the former decays more slowly (or lasts longer) when compared to that for the latter. This phenomenon may be explained by previous studies (29,36,68) that collagen fibers can limit protrusions and facilitate directed cell migration.

Finally, the JPDs of the two conditions also exhibit the statistical features of directional persistence and the significant differences between them. The JPD of the collagen condition is mainly composed of two clusters (marked by the small stars) in the interval of 0–70 min, and their centers are located at (5.126, 0.910) and (5.178, 0.824), respectively. However, it is noticeable that the JPD of the substrate condition contains three clusters with the centers located at (5.125, 0.918), (5.152, 0.832), and (7.429, 0.729). In contrast, the two clusters for the collagen condition correspond to the first and second clusters for the substrate condition, respectively, due to the similar joint parameters. Furthermore, these clusters have different cell proportions due to the different areas of the black regions, e.g., the second cluster at (5.178, 0.824) for the collagen condition and the first cluster at (5.125, 0.918) for the substrate condition represent higher proportions, whereas the clusters at (5.126, 0.910) and (5.152, 0.832) represent lower proportions. Thus, if we do not consider the third cluster at (7.429, 0.729) for the substrate condition and the more points at $TL_1 = 100$ min for the collagen condition, we could judge that the proportion of cells with stronger persistence is higher in the collagen condition than that in the substrate condition. Of course, the focus of JPD is not to estimate the proportions of cells with different persistence but to reveal how the persistence is distributed for a given cell population. Here, the analysis in terms of statistical features is also enhanced by the time-varying properties shown in Figs. S3 and S4.

The above results indicate that the directional persistence of cell migration is strictly regulated by intra/extracellular cues and validate the utility of the EAD-based approach in revealing the statistical properties of directional persistence.

The EAD metric captures transitions in the migration dynamics of individual MDA-MB-231 cells in a confined environment

In the previous sections, we have thoroughly analyzed the directional persistence of cell migration from the perspective of statistical features that are affected by different conditions. Now, let us shift our thinking to study the time-varying characteristics of directional persistence using the EAD-based approach. Here, the experimental trajectory data represent the positions of a single MDA-MB-231 cell

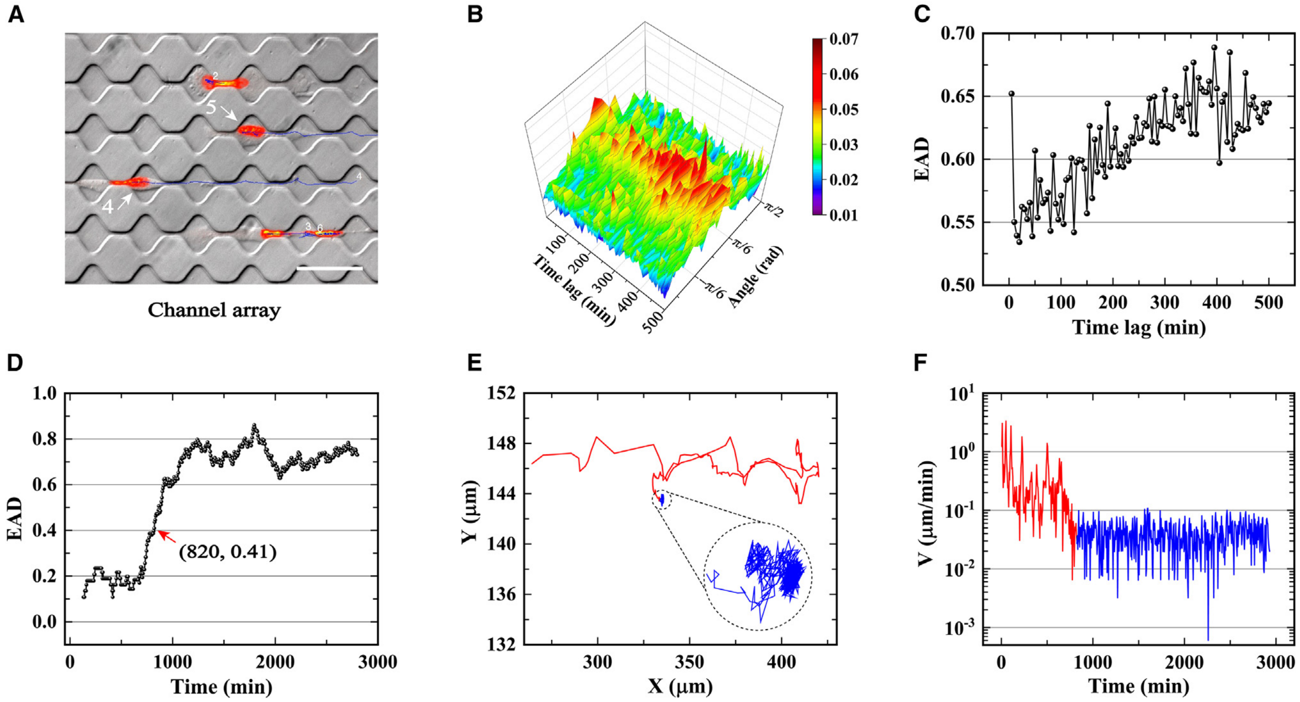

FIGURE 7 The EAD approach captures the transition of migration modes of a single MDA-MB-231 breast cancer cell in a channel array. (*A*) Invasive breast cancer cells migrate through a microstructured array of channels and chambers. Cell nuclei are stained with Hoechst and shown in red. Scale bar, 50 μm. (*B*) 3D PDF of angular displacements for the cell labeled 5. (*C*) EAD as a function of time lag. (*D*) Time-varying EAD. The red arrow indicates the transition point. (*E*) The migration trajectory is divided into two parts based on the transition point. The inset is an enlarged image of the second part. (*F*) The migration velocity is divided into two parts. The two parts are marked in red and blue, respectively; the bin size is 18 for calculating the PDF and the window length is 51. Data in (*A*) and (*E*) courtesy of Claus Metzner (57,63). To see this figure in color, go online.

nucleus stained in red with Hoechst as it squeezes through a one-dimensional (1D) array of sequential channels and chambers (see Fig. 7 *A*).

Fig. 7 *B* shows the 3D PDF of the cell labeled “5”; on the whole, it differs significantly from the other PDFs associated with the Arpin and collagen experiments because this PDF is the result of a single cell instead of an ensemble average. Although the PDF presents some irregular fluctuations or small peaks, it still has a noticeable trend; i.e., as the time lag increases, the sharp peak (with fewer red regions) centered at 0 rad gradually extends to two sides and forms a flatter peak (with more red regions). This change gradually transforms the uneven PDF at the small scale of time lag into the more uniform PDF at the large scale, implying that the correlation becomes progressively weaker. Subsequently, this result is further confirmed by the EAD profile (Fig. 7 *C*), which increases from ~0.55 to ~0.65 in a manner of fluctuation, meaning that the directional persistence decays with the increase of time lag.

More importantly, the time-varying EAD (Fig. 7 *D*) for this cell shows a noticeable transition from a low level (~0.2) to a high level (~0.7), indicating that this cell mainly performs two migration modes (i.e., modes with strong and weak persistence). Furthermore, this trend is highly similar to that shown in Fig. 4 *C*, which further verifies the reasonability of the abrupt persistence change. Next, we artificially determine the position of the transition point (820, 0.41) according to the trend of the EAD profile in Fig. 7 *D* and use it to divide the migration trajectory into two parts (Fig. 7 *E*). The first part (red line) describes the normal motility of this cell, including squeezing through narrow channels as well as migrating through wide chambers. However, the second part (blue line) shows that this cell persistently “wanders” in a tiny region without significant displacements. Similarly, the velocity profile (Fig. 7 *F*) also vividly manifests that the velocities are significantly larger (~3.0 μm/min) with gradually decreasing amplitudes in the first part, whereas they are smaller (<0.1 μm/min) but basically stable in the second part. See Fig. S5 for further analysis of the cell labeled “4.” Therefore, the EAD-based approach could be used to analyze the time-varying characteristics of individual migration trajectories, especially the transition of directional persistence.

DISCUSSION

In this paper, we have proposed the EAD-based approach, which mainly involves the analysis of cell angular displacement and the calculation of Shannon entropy. It can enable us to investigate cellular turning dynamics, especially the directional persistence of cell migration, including statistical and time-varying characteristics, and further reveal the effects of the local ECM on behavioral modes.

With the EAD-based approach, we first analyze the role of the intracellular crucial protein Arpin in regulating the directional persistence of *D. discoideum* migration (Fig. 5). On the whole, the results show that the EAD of the KO amoeba is smaller than that of the WT and the rescue amoeba, which directly illustrates that the directional persistence of the KO amoeba is stronger compared to the other two cases. In detail, we can qualitatively estimate the degree of EAD according to the shape of the 3D PDF of angular displacement. It provides a way to quickly understand the directional persistence when there is a lack of appropriate tools. Furthermore, the EAD profile derived from the 3D PDF can quantitatively show how the EAD changes with time lag. Significantly, one can easily compare the values of EAD_1 and TL_1 with other cases. On the one hand, it can be seen that the EAD_1 of the KO amoeba is significantly smaller than that of the WT and the rescue amoeba, indicating that the KO amoeba has a stronger short-term persistence; on the other hand, the TL_1 of the KO amoeba is more remarkable, implying that there is a long-term persistence.

Apart from the significant differences between the KO amoeba and the other two groups, there are also slight differences between the WT and the rescue amoeba that have not been reported in previous studies; i.e., the EAD_1 (0.942 ± 0.007) of the WT amoeba is significantly smaller than that (0.964 ± 0.004) of the rescue amoeba, whereas the TL_1 (30/60 min) of the former is larger than that (25/60 min) of the latter. The above relationships manifest that the WT amoeba possesses a stronger directional persistence compared to the rescue amoeba, which we believe may be mainly caused by two factors, namely more Arpin proteins expressed in the KO amoeba and/or the effects of the relevant operation.

Besides the characteristics encoded in the EAD profile, the JPD also vividly exhibits the statistical aspect of directional persistence. As shown in Fig. 5, the three clusters of the KO amoeba cover larger regions compared to the other two groups, illustrating that there are large heterogeneities in the directional persistence of the KO amoeba, in particular containing more cells with stronger persistence (see also Fig. S2). Here, the discrete clusters also represent a hierarchy in the dynamics for a given population, which may provide some insights into the formation of collective cell migration, e.g., the coexistence and role exchange of leading and following cells (69).

Except for the intracellular proteins, we also further study the effects of extracellular environments, i.e., 3D collagen and 2D substrate, on MCF-10A cell migration (Fig. 6). In general, the EAD of the collagen condition is smaller than that of the substrate condition, meaning that the collagen gel contributes to the enhancement of directional persistence. Interestingly, there are two aspects of the EAD profiles that are worth discussing: 1) although the EADs of the collagen condition are integrally smaller than those of the substrate condition, the EADs of some cells in the latter are minimal and correspond to the strongest persistence at time lag $= 2$ and 4 min. This feature shows that the persistence of the collagen condition is slightly weaker than that of the substrate condition at the small scale of time; however, it can last longer at the large scale of time. We believe that this difference is mainly due to collagen fibers, which can transmit mechanical signals (30,33) and further guide the directed cell migration (see also Fig. S4). 2) Both the EADs of the two conditions gradually decrease with the increasing time lag, indicating that the directional persistence becomes stronger. For this phenomenon, we speculate that there is a long-term factor that coexists with the persistence, which could be chemotaxis, systematic errors, etc. As the time lag increases, the effect of the persistence gradually becomes weaker, further causing this factor to become the main contributor.

To better illustrate the differences between the collagen and substrate conditions (see also Fig. S3), it is essential to examine the details of the JPDs. Overall, the first and second clusters of the collagen condition are basically consistent with those of the substrate condition due to the similar coordinates of the centers corresponding to the main migration modes. However, there is a third cluster of the substrate condition with the center at (7.429, 0.729), which means that some cells have stronger persistence. In contrast to the collagen condition, there is no prominent cluster, but more points at $TL_1 = 100$ min. These points result from the fact that all values in individual EAD profiles are smaller than the value used to determine TL_1, causing the problem that the value of TL_1 cannot be determined in the prescribed interval, indicating that the persistence of some cells can last longer and corresponds to the third cluster of the substrate condition.

Finally, to verify the performance of the EAD-based approach in analyzing time-dependent characteristics, we further study the invasive MDA-MB-231 cell migration in a microstructured array consisting of sequential channels and chambers by utilizing the sliding-window algorithm (Fig. 7). The results indicate that the cell undergoes two phases, i.e., phases of stronger and weaker persistence, and here the transition time is 820 min. According to the transition time, the migration trajectory and velocity profile are divided into two parts. The first part of the trajectory covers a larger distance, whereas the second part covers a tiny distance; likewise, the first part of the velocity profile differs significantly from the second part (see also Fig. 5). The above differences further confirm the effectiveness of the approach in analyzing the time-varying features of directional persistence. Nevertheless, we have to be aware of the problem that there are still some shortcomings of the approach, as follows: 1) it is not able to process the data in the first half of the first window and the second half of the last window, due to the presence of the window, and this further contributes to the loss of information. Therefore, it is better to record more trajectory data to compensate for the loss in obtaining the desired information. 2) It is challenging to determine the best window length because the small length will lead to inaccurate results and the ample length will have a poor temporal resolution.

CONCLUSIONS

In this paper, we propose an approach called EAD, which allows us to investigate cellular turning dynamics based on the angular displacement obtained from trajectory data, especially the directional persistence that strongly correlates with migration modes. Our analysis has verified the effectiveness and robustness of the EAD-based approach, which could first be utilized to study the effects of the intracellular protein, Arpin, on *D. discoideum* migration, indicating that the Arpin is negatively correlated with directional persistence. Furthermore, the approach also reveals the critical role of the extracellular local microenvironment, i.e., 3D collagen gel and 2D substrate, in regulating the directional persistence of MCF-10A cell migration. The results not only show how the directional persistence decays with time lag but also show its distribution to further distinguish the different environments. Additionally, we analyze the time-varying features of directional persistence during invasive MDA-MB-231 cell migration through a microstructured array, and the detected transition time identifies two prominent migration modes characterized by persistence. Overall, our approach contributes to precisely analyzing the directional persistence of cell migration, including statistical and time-varying characteristics, without any complex calculations, and further elucidates the effects of these multiscale cues on cellular migration modes in complex microenvironments.

DATA AND CODE AVAILABILITY

The data that support the results of this study are available from the corresponding author upon reasonable request.

The custom codes used in this study are available upon request from the corresponding author.

SUPPORTING MATERIAL

Supporting material can be found online at https://doi.org/10.1016/j.bpj.2024.02.010.

AUTHOR CONTRIBUTIONS

Y.L. conceived of the presented idea and designed the computational framework. Y.L., Y.J., and D.Q. performed the analytic calculations and the numerical simulations. X.L., G.L., and L.L. conceived and planned the experiments. G.L., Z.L., and L.Z. carried out the experiments. Y.L. wrote the manuscript with support from J.S., L.L., and W.W.. J.S. and Z.L. supervised and revised the manuscript. All authors discussed the results and contributed to the final manuscript.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (grant nos. 12347178, 62106032, 62171073, 12090052), the Natural Science Foundation of Chongqing, China (grant nos. CSTB2022NSCQ-MSX1260, CSTB2023NSCQ-MSX0861), the Science and Technology Research Program of Chongqing Municipal Education Commission (grant no. KJQN202300623), and the Key Project of Technology Innovation and Application Development of Chongqing, China (grant no. cstc2021jscx-gksbX0060). Y.J. does not have financial support for this work. We thank Ben Fabry, Claus Metzner, and Lena Lautscham for providing experimental data concerning channel arrays and Qihui Fan and Fangfu Ye for providing experimental data concerning correlated cell motility.

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPORTING CITATIONS

References (30,50–53,57,63–64) appear in the Supporting Material.

REFERENCES

1. Petrie, R. J., A. D. Doyle, and K. M. Yamada. 2009. Random versus directionally persistent cell migration. *Nat. Rev. Mol. Cell Biol.* 10:538–549. https://doi.org/10.1038/nrm2729.
2. Hanahan, D. 2022. Hallmarks of Cancer: New Dimensions. *Cancer Discov.* 12:31–46. https://doi.org/10.1158/2159-8290.Cd-21-1059.
3. Berg, H. C. 1987. Random Walks in Biology. https://doi.org/10.1063/1.2819954.
4. Franz, A., W. Wood, and P. Martin. 2018. Fat Body Cells Are Motile and Actively Migrate to Wounds to Drive Repair and Prevent Infection. *Dev. Cell.* 44:460–470.e3. https://doi.org/10.1016/j.devcel.2018.01.026.
5. Pawar, K. B., S. Desai, …, A. A. Deshmukh. 2021. Wound with Diabetes: Present Scenario and Future. *Curr. Diabetes Rev.* 17:136–142. https://doi.org/10.2174/1573399816666200703180137.
6. Friedl, P., and D. Gilmour. 2009. Collective cell migration in morphogenesis, regeneration and cancer. *Nat. Rev. Mol. Cell Biol.* 10:445–457. https://doi.org/10.1038/nrm2720.
7. Silva, C. G., E. Peyre, and L. Nguyen. 2019. Cell migration promotes dynamic cellular interactions to control cerebral cortex morphogenesis. *Nat. Rev. Neurosci.* 20:318–329. https://doi.org/10.1038/s41583-019-0148-y.
8. von Andrian, U. H., and C. R. Mackay. 2000. Advances in immunology: T-cell function and migration - Two sides of the same coin. *N. Engl. J. Med.* 343:1020–1033. https://doi.org/10.1056/nejm200010053431407.
9. Cyster, J. G. 1999. Chemokines - Chemokines and cell migration in secondary lymphoid organs. *Science.* 286:2098–2102. https://doi.org/10.1126/science.286.5447.2098.
10. Wu, P.-H., D. M. Gilkes, …, D. Wirtz. 2020. Single-cell morphology encodes metastatic potential. *Sci. Adv.* 6, eaaw6938. https://doi.org/10.1126/sciadv.aaw6938.
11. Sun, X., B. Wu, …, R. Li. 2021. Tumour DDR1 promotes collagen fibre alignment to instigate immune exclusion. *Nature.* 599:673–678. https://doi.org/10.1038/s41586-021-04057-2.
12. Weiger, M. C., S. Ahmed, …, J. M. Haugh. 2010. Directional Persistence of Cell Migration Coincides with Stability of Asymmetric Intracellular Signaling. *Biophys. J.* 98:67–75. https://doi.org/10.1016/j.bpj.2009.09.051.

13. Krause, M., and A. Gautreau. 2014. Steering cell migration: lamellipodium dynamics and the regulation of directional persistence. *Nat. Rev. Mol. Cell Biol.* 15:577–590. https://doi.org/10.1038/nrm3861.
14. Liu, J., Z. Liu, …, C. Wu. 2021. Kindlin-2 promotes rear focal adhesion disassembly and directional persistence during cell migration. *J. Cell Sci.* 134:jcs244616. https://doi.org/10.1242/jcs.244616.
15. Bahcecioglu, G., X. Yue, …, P. Zorlutuna. 2021. Aged Breast Extracellular Matrix Drives Mammary Epithelial Cells to an Invasive and Cancer-Like Phenotype. *Adv. Sci.* 8:e2100128. https://doi.org/10.1002/advs.202100128.
16. Cox, T. R. 2021. The matrix in cancer. *Nat. Rev. Cancer.* 21:217–238. https://doi.org/10.1038/s41568-020-00329-7.
17. De Martino, D., and J. J. Bravo-Cordero. 2023. Collagens in cancer: structural regulators and guardians of cancer progression. *Cancer Res.* 83:1386–1392. https://doi.org/10.1158/0008-5472.Can-22-2034.
18. Shaebani, M. R., M. Piel, and F. Lautenschläger. 2022. Distinct speed and direction memories of migrating dendritic cells diversify their search strategies. *Biophys. J.* 121:4099–4108. https://doi.org/10.1016/j.bpj.2022.09.033.
19. Zhang, Y., D. Wei, …, F. Ye. 2023. Run-and-Tumble Dynamics and Mechanotaxis Discovered in Microglial Migration. *Research.* 6, 0063. https://doi.org/10.34133/research.0063.
20. Wang, X., S. Chen, …, L. Liu. 2021. Abnormal Aggregation of Invasive Cancer Cells Induced by Collective Polarization and ECM-Mediated Mechanical Coupling in Coculture Systems. *Research.* 2021, 9893131. https://doi.org/10.34133/2021/9893131.
21. Nam, K.-H., P. Kim, …, D.-H. Kim. 2016. Multiscale Cues Drive Collective Cell Migration. *Sci. Rep.* 6, 29749. https://doi.org/10.1038/srep29749.
22. Roussos, E. T., J. S. Condeelis, and A. Patsialou. 2011. Chemotaxis in cancer. *Nat. Rev. Cancer.* 11:573–587. https://doi.org/10.1038/nrc3078.
23. Camley, B. A., J. Zimmermann, …, W. J. Rappel. 2016. Emergent Collective Chemotaxis without Single-Cell Gradient Sensing. *Phys. Rev. Lett.* 116. 098101. https://doi.org/10.1103/PhysRevLett.116.098101.
24. Shellard, A., and R. Mayor. 2021. Collective durotaxis along a self-generated stiffness gradient in vivo. *Nature.* 600:690–694. https://doi.org/10.1038/s41586-021-04210-x.
25. Isomursu, A., K.-Y. Park, …, D. J. Odde. 2022. Directed cell migration towards softer environments. *Nat. Mater.* 21:1081–1090. https://doi.org/10.1038/s41563-022-01294-2.
26. Mizoguchi, T., S. Ikeda, …, M. Itoh. 2017. Mib1 contributes to persistent directional cell migration by regulating the Ctnnd1-Rac1 pathway. *Proc. Natl. Acad. Sci. USA.* 114:E9280–E9289. https://doi.org/10.1073/pnas.1712560114.
27. Theisen, U., E. Straube, and A. Straube. 2012. Directional Persistence of Migrating Cells Requires Kif1C-Mediated Stabilization of Trailing Adhesions. *Dev. Cell.* 23:1153–1166. https://doi.org/10.1016/j.devcel.2012.11.005.
28. Tee, J. Y., and A. Mackay-Sim. 2021. Directional Persistence of Cell Migration in Schizophrenia Patient-Derived Olfactory Cells. *Int. J. Mol. Sci.* 22:9177. https://doi.org/10.3390/ijms22179177.
29. Ray, A., Z. M. Slama, …, P. P. Provenzano. 2017. Enhanced Directional Migration of Cancer Stem Cells in 3D Aligned Collagen Matrices. *Biophys. J.* 112:1023–1036. https://doi.org/10.1016/j.bpj.2017.01.007.
30. 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. Ed. Engl.* 60:11858–11867. https://doi.org/10.1002/anie.202016084.
31. Lin, J., Y. Shi, …, C. Zhang. 2020. Mechanical Roles in Formation of Oriented Collagen Fibers. *Tissue Eng., Part B.* 26 (2):116–128. https://doi.org/10.1089/ten.TEB.2019.0243.
32. Han, W., S. Chen, …, L. Liu. 2016. Oriented collagen fibers direct tumor cell intravasation. *Proc. Natl. Acad. Sci. USA.* 113:11208–11213. https://doi.org/10.1073/pnas.1610347113.
33. Zheng, Y., H. Nan, …, Y. Jiao. 2019. Modeling cell migration regulated by cell extracellular-matrix micromechanical coupling. *Phys. Rev. E.* 100. 043303. https://doi.org/10.1103/PhysRevE.100.043303.
34. Liu, Y., Y. Jiao, …, L. Liu. 2021. Shannon entropy for time-varying persistence of cell migration. *Biophys. J.* 120:2552–2565. https://doi.org/10.1016/j.bpj.2021.04.026.
35. Elkhatib, N., E. Bresteau, …, G. Montagnac. 2017. Tubular clathrin/AP-2 lattices pinch collagen fibers to support 3D cell migration. *Science.* 356, eaal4713. https://doi.org/10.1126/science.aal4713.
36. Ray, A., R. K. Morford, …, P. P. Provenzano. 2018. Dynamics of 3D carcinoma cell invasion into aligned collagen. *Integr. Biol.* 10:100–112. https://doi.org/10.1039/C7IB00152E.
37. Provenzano, P. P., K. W. Eliceiri, …, P. J. Keely. 2006. Collagen reorganization at the tumor-stromal interface facilitates local invasion. *BMC Med.* 4:38. https://doi.org/10.1186/1741-7015-4-38.
38. Pavithra, V., S. V. Sowmya, …, S. Nambiar. 2017. Tumor-associated Collagen Signatures: An Insight. *World J. Dent.* 8:224–230.
39. Brett, E. A., M. A. Sauter, …, D. Duscher. 2020. Tumor-associated collagen signatures: pushing tumor boundaries. *Cancer Metabol.* 8, 14.
40. 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. https://doi.org/10.1016/j.bpj.2019.11.3389.
41. Brückner, D. B., A. Fink, …, C. P. Broedersz. 2019. Stochastic nonlinear dynamics of confined cell migration in two-state systems. *Nat. Phys.* 15:595–601. https://doi.org/10.1038/s41567-019-0445-4.
42. Brückner, D. B., N. Arlt, …, C. P. Broedersz. 2021. Learning the dynamics of cell–cell interactions in confined cell migration. *Proc. Natl. Acad. Sci. USA.* 118, e2016602118. https://doi.org/10.1073/pnas.2016602118.
43. SenGupta, S., C. A. Parent, and J. E. Bear. 2021. The principles of directed cell migration. *Nat. Rev. Mol. Cell Biol.* 22:529–547. https://doi.org/10.1038/s41580-021-00366-6.
44. Li, L., S. F. Norrelykke, and E. C. Cox. 2008. Persistent Cell Motion in the Absence of External Signals: A Search Strategy for Eukaryotic Cells. *PLoS One.* 3:11.e2093. https://doi.org/10.1371/journal.pone.0002093.
45. Harris, T. H., E. J. Banigan, …, C. A. Hunter. 2012. Generalized Levy walks and the role of chemokines in migration of effector CD8(+) T cells. *Nature.* 486:545–U145. https://doi.org/10.1038/nature11098.
46. Reynolds, A. M. 2010. Can spontaneous cell movements be modelled as Levy walks? *Phys. Stat. Mech. Appl.* 389:273–277. https://doi.org/10.1016/j.physa.2009.09.027.
47. Yogurtcu, O. N., and S. X. Sun. 2012. A Two-State Eukaryotic Cell Migration Model. *Biophys. J.* 102:347A. https://doi.org/10.1016/j.bpj.2011.11.1899.
48. Gorelik, R., and A. Gautreau. 2014. Quantitative and unbiased analysis of directional persistence in cell migration. *Nat. Protoc.* 9:1931–1943. https://doi.org/10.1038/nprot.2014.131.
49. Tranquillo, R. T., D. A. Lauffenburger, and S. H. Zigmond. 1988. A stochastic model for leukocyte random motility and chemotaxis based on receptor binding fluctuations. *J. Cell Biol.* 106:303–309. https://doi.org/10.1083/jcb.106.2.303.
50. Parkhurst, M. R., and W. M. Saltzman. 1992. QUANTIFICATION OF HUMAN NEUTROPHIL MOTILITY IN 3-DIMENSIONAL COLLAGEN GELS - EFFECT OF COLLAGEN CONCENTRATION. *Biophys. J.* 61:306–315. https://doi.org/10.1016/s0006-3495(92)81838-6.
51. 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. https://doi.org/10.1073/pnas.1318967111.
52. 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. https://doi.org/10.1038/nprot.2015.030.
53. Liu, Y., Y. Jiao, …, L. Liu. 2021. Deriving time-varying cellular motility parameters via wavelet analysis. *Phys. Biol.* 18, 046007. https://doi.org/10.1088/1478-3975/abfcad.

54. Wu, P. H., D. M. Gilkes, and D. Wirtz. 2018. The Biophysics of 3D Cell Migration. *In* Annual Review of Biophysics, Annual Review of Biophysics. Annual Reviews, *Vol 47*. K. A. Dill, ed, pp. 549–567, Palo Alto.
55. Selmeczi, D., S. Mosler, …, H. Flyvbjerg. 2005. Cell motility as persistent random motion: Theories from experiments. *Biophys. J.* 89:912–931. https://doi.org/10.1529/biophysj.105.061150.
56. 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. https://doi.org/10.1103/PhysRevE.94.062401.
57. Metzner, C., C. Mark, …, B. Fabry. 2015. Superstatistical analysis and modelling of heterogeneous random walks. *Nat. Commun.* 6:7516. https://doi.org/10.1038/ncomms8516.
58. Codling, E. A., M. J. Plank, and S. Benhamou. 2008. Random walk models in biology. *J. R. Soc. Interface.* 5:813–834. https://doi.org/10.1098/rsif.2008.0014.
59. 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. Physiol.* 15, 13602. https://doi.org/10.1007/s11467-019-0929-9.
60. Li, L., E. C. Cox, and H. Flyvbjerg. 2011. 'Dicty dynamics': Dictyostelium motility as persistent random motion. *Phys. Biol.* 8, 046006. https://doi.org/10.1088/1478-3975/8/4/046006.
61. Fetics, S., A. Thureau, …, J. Cherfils. 2016. Hybrid Structural Analysis of the Arp2/3 Regulator Arpin Identifies Its Acidic Tail as a Primary Binding Epitope. *Structure.* 24:252–260. https://doi.org/10.1016/j.str.2015.12.001.
62. Buehler, M. J. 2006. Nature designs tough collagen: Explaining the nanostructure of collagen fibrils. *Proc. Natl. Acad. Sci. USA.* 103:12285–12290. https://doi.org/10.1073/pnas.0603216103.
63. 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. https://doi.org/10.1016/j.bpj.2015.07.025.
64. Dang, I., R. Gorelik, …, A. Gautreau. 2013. Inhibitory signalling to the Arp2/3 complex steers cell migration. *Nature.* 503:281–284. https://doi.org/10.1038/nature12611.
65. Liu, Y., D. He, …, L. Liu. 2021. Nonlinear dynamics of cell migration in anisotropic microenvironment. *Chin. Phys. B.* 30. 090505. https://doi.org/10.1088/1674-1056/ac11d5.
66. Leibovich, N., A. Dechant, …, E. Barkai. 2016. Aging Wiener-Khinchin theorem and critical exponents of 1/f (beta) noise. *Phys. Rev. E.* 94, 052130. https://doi.org/10.1103/PhysRevE.94.052130.
67. Maiuri, P., J. F. Rupprecht, …, R. Voituriez. 2015. Actin Flows Mediate a Universal Coupling between Cell Speed and Cell Persistence. *Cell.* 161:374–386. https://doi.org/10.1016/j.cell.2015.01.056.
68. 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. https://doi.org/10.1016/j.bpj.2014.10.035.
69. Zhang, J., K. F. Goliwas, …, C. A. Reinhart-King. 2019. Energetic regulation of coordinated leader-follower dynamics during collective invasion of breast cancer cells. *Proc. Natl. Acad. Sci. USA.* 116:7867–7872. https://doi.org/10.1073/pnas.1809964116.