**OPEN**

# Stochastic and Heterogeneous Cancer Cell Migration: Experiment and Theory

Taejin Kwon[1,4], Ok-Seon Kwon[2,4], Hyuk-Jin Cha [3*] & Bong June Sung[1*]

Cell migration, an essential process for normal cell development and cancer metastasis, differs from a simple random walk: the mean-square displacement ($\langle(\Delta r)^2(t)\rangle$) of cells sometimes shows non-Fickian behavior, and the spatiotemporal correlation function ($G(r, t)$) of cells is often non-Gaussian. We find that this intriguing cell migration should be attributed to heterogeneity in a cell population, even one with a homogeneous genetic background. There are two limiting types of heterogeneity in a cell population: *cellular heterogeneity* and *temporal heterogeneity*. Cellular heterogeneity accounts for the cell-to-cell variation in migration capacity, while temporal heterogeneity arises from the temporal noise in the migration capacity of single cells. We illustrate that both cellular and temporal heterogeneity need to be taken into account simultaneously to elucidate cell migration. We investigate the two-dimensional migration of A549 lung cancer cells using time-lapse microscopy and find that the migration of A549 cells is Fickian but has a non-Gaussian spatiotemporal correlation. We find that when a theoretical model considers both cellular and temporal heterogeneity, the model reproduces all of the anomalous behaviors of cancer cell migration.

Cell migration is essential to normal cell development[1–3], cancer metastasis[4–6] and wound healing[7–9]. Developing a mathematical model for cell migration[10–15] would help in understanding cell migration and designing new strategies to cure cancers and treat wounds. Because the trajectories of cells look quite similar to a random walk at certain timescales, most mathematical models are based on the diffusion equation and its equivalents such as a stochastic differential equation. Recent studies reported, however, that cell migration should be not only stochastic but also heterogeneous[16–18]: the cell-to-cell variation in migration capacity is significant, and/or a single cell undergoes temporal transitions in migration capacity. This heterogeneity makes the development of a mathematical model for cell migration tremendously challenging. In this study, we investigate the effects of heterogeneity on A549 lung cancer cell migration and compare various models to describe the heterogeneous trajectories of A549 cancer cells.

Heterogeneity is inherent in a cell population. The heterogeneity of a cancer cell population is significant even when the cancer cell line is derived from a single clone with the same genetic background[19–26]. Heterogeneity in a cell population may be categorized into two limiting cases: *cellular heterogeneity* and *temporal heterogeneity*. Cellular heterogeneity (population noise) accounts for time-independent cell-to-cell variation, while temporal heterogeneity (temporal noise) results from temporal fluctuations of single cells. Cellular and temporal heterogeneity are not mutually exclusive, as population and temporal noises may occur simultaneously in cell populations. Heterogeneity can result from a number of spatiotemporal factors such as different stages of the cell cycle[27], circadian rhythm[28] and the level of adenosine triphosphate (ATP)[29] or $Ca^{2+}$ [30] at an individual single cell level, which determine migration capacity. Therefore, an issue of fundamental importance is finding ways beyond population averaging to understand cell migration in terms of heterogeneity.

The diffusion equation, which describes the migration of particles, is based on two fundamental observations: (a) a particle is neither created nor destroyed, and (b) the flux of particles is proportional to a particle density gradient[31]. The diffusion equation leads to two important results: (1) the mean-square displacement ($\langle(\Delta r)^2(t)\rangle$), a measure of how far the particle diffuses in a given time $t$, should be linearly proportional to time $t$ at long times, i.e., $\langle(\Delta r)^2(t)\rangle \sim t^1$, and (2) the particle displacement should follow Gaussian statistics due to the central limit

[1]Department of Chemistry and Research Institute for Basic Science, Sogang University, Seoul, 04107, Republic of Korea. [2]Department of Life Sciences, Sogang University, Seoul, 04107, Republic of Korea. [3]College of Pharmacy, Seoul National University, Seoul, 08826, Republic of Korea. [4]These authors contributed equally: Taejin Kwon and Ok-Seon Kwon. *email: hjcha93@snu.ac.kr; bjsung@sogang.ac.kr
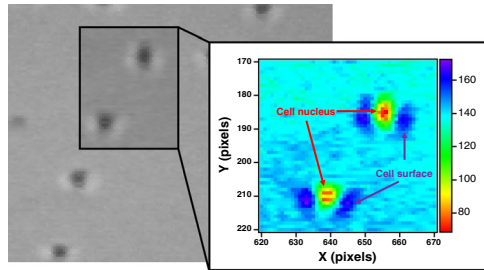
**Figure 1.** Image of A549 cells obtained by time-lapse microscope. (inset) Color code represents the brightness of pixels. The number of tracked A549 cells is 212.

theorem such that the spatiotemporal correlation function ($G(r, t)$) should be Gaussian. The concept of diffusion (and hence a random walk) has been employed to describe the cell migration. For example, a persistent random walk (PRW) model has been employed extensively to interpret and simulate cell migration[10,11,32–34]. Recent studies on cell migration showed, however, that cell migration is anomalous[12–15,35–38]: (1) cells sometimes showed non-Fickian behavior, i.e., $\langle(\Delta r)^2(t)\rangle \sim t^\alpha$ with $\alpha \neq 1$, and (2) $G(r, t)$ was non-Gaussian. We also find that A549 lung cancer cells migrate with a non-Gaussian $G(r, t)$. This finding implies that the random walk model (and the conventional PRW model) would not hold for cells and that a new mathematical model would be required.

Previous studies illustrated that if one were to incorporate the cellular heterogeneity into the PRW model with different parameters for different cells, the non-Gaussian $G(r, t)$ of HT1080 fibrosarcoma cells could be reproduced[15]. On the other hand, a single cell underwent temporal transitions between fast and slow states[36,39–42]. Metzner et al.[36] proposed a statistical framework to model and analyze heterogeneous cell migration of the breast carcinoma cell line MDA-MB-231. They found that in the presence of such temporal heterogeneity, the spatio-temporal correlation function could be non-Gaussian, which a simple random walk model can not elucidate. Scientific questions arise: how cellular and temporal heterogeneity relate to the anomalous cell migration, and whether one can develop a mathematical model that combines both cellular and temporal heterogeneity. In this work, to compare with A549 lung cancer cells and answer these questions, we employ the following four different theoretical models: (1) homogeneous (HO) model without any type of heterogeneity, (2) cellular heterogeneity (CH) model with only cellular heterogeneity considered, (3) temporal heterogeneity (TH) model with only temporal heterogeneity and (4) cellular and temporal heterogeneity (CTH) model with both cellular and temporal heterogeneity. We find that only the CTH model may reproduce the experimental results for not only $G(r, t)$ but also the spatiotemporal correlation function $g_i(r, t)$ of each single cell (not averaged over the population of cells) of A549 lung cancer cells, and that both cellular and temporal heterogeneity need to be taken into account to elucidate single-cell migration.

The rest of the paper is organized as follows. The experimental and cell tracking methods are described in the Materials and Methods section, the results are presented and discussed in the Results and Discussion section, and the conclusions are presented in the Conclusion section.

## Results and Discussion

**Migration of A549 cancer cells.** We obtain the trajectories of A549 cancer cells by using time-lapse microscopy (as shown in Figs 1 and 2(A)). We estimate the mean velocity ($\vec{v}_i(t)$) of each A549 cell during a $\Delta t = 34$ min interval, i.e.,

$$\vec{v}_i(t) = \frac{\vec{r}_i(t + \Delta t) - \vec{r}_i(t)}{\Delta t}. \tag{1}$$

Here, $\vec{r}_i(t)(= (\hat{x}_i, \hat{y}_i))$ denotes the position vector of the $i$th A549 cell at time $t$. Note that the time resolution is limited to 34 min in our experiment because the root of the mean-square displacement of A549 cells reaches the dimension (5 $\mu$m) of one pixel only after 34 min. See more details in Supporting Information. We estimate the mean-square displacement $\langle(\Delta r)^2(t)\rangle$ as follows:

$$\langle(\Delta r)^2(t)\rangle = \frac{1}{N}\sum_{i=1}^{N} \langle\vec{r}_i(t + t') - \vec{r}_i(t')\rangle^2_{t'}, \tag{2}$$

where $\langle\cdots\rangle_{t'}$ denotes a time average over different time origins $t'$ and $N$ is the number of A549 cells. The integral form of the time average is $\langle\cdots\rangle_{t'} = \frac{1}{\tau}\int_0^\tau \cdots dt'$, where $\tau$ is the total measurement time in our experiment. Therefore, $\langle(\Delta r)^2(t)\rangle$ is a quantity averaged over both times and all cells. We also obtain the spatiotemporal correlation function $g_i(r, t)$ of each A549 cell as follows:

$$g_i(r, t) = \langle\delta\{\vec{r} - [\vec{r}_i(t + t') - \vec{r}_i(t')]\}\rangle_{t'}, \tag{3}$$

where $\delta$ is the Dirac delta function. One can construct a histogram and obtain $g_i(r, t)$ numerically by checking whether the displacement ($|\vec{r}_i(t + t') - \vec{r}_i(t')|$) of the $i$th cell from $t'$ to $t' + t$ lies between a certain range. Note that $g_i(r, t)$ is averaged not over cells but over different time origins $t'$. The physical meaning of $2\pi r g_i(r, t)$ is the
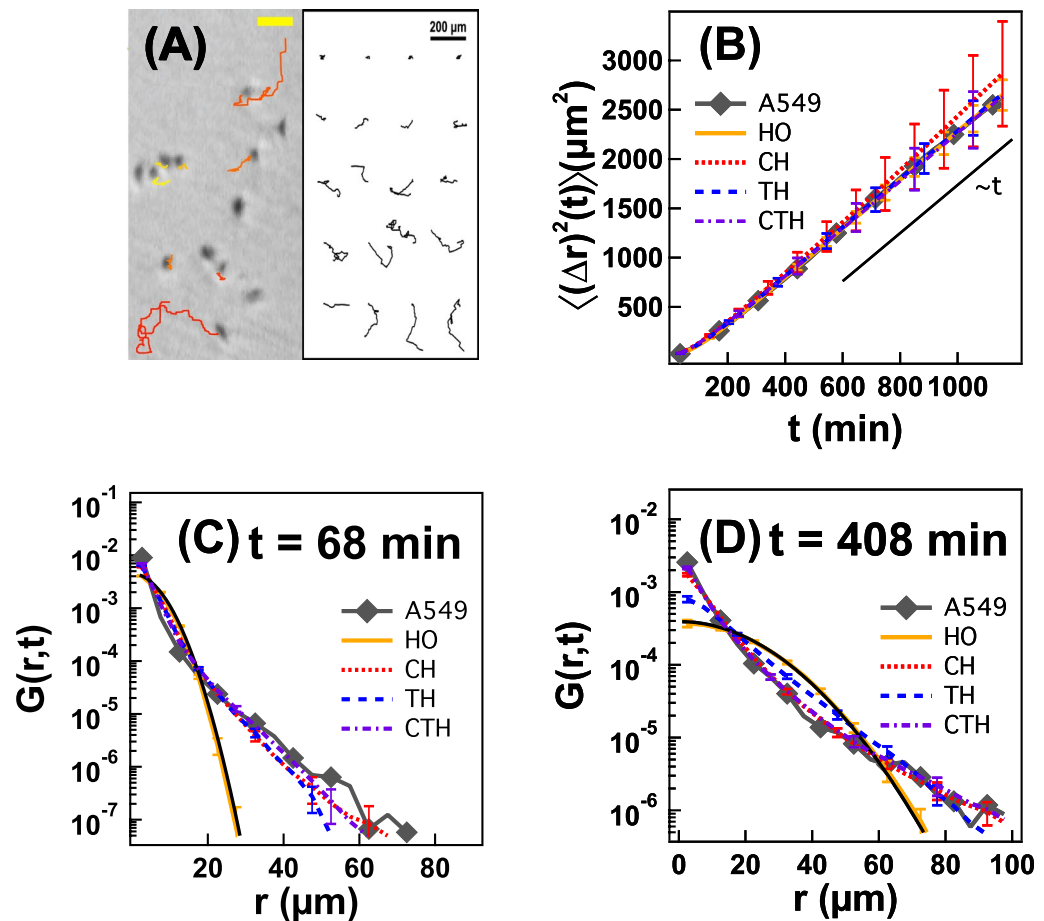
**Figure 2.** (**A**) Representative trajectories of A549 cells obtained from time-lapse microscopy. The yellow bar is 100 $\mu m$ long in the figure. (**B**) $\langle(\Delta r)^2(t)\rangle$ averaged over all cells as function of $t$. $G(r,t)$ averaged over all cells as function of $r$ at (**C**) $t = 68$ min and (**D**) $t = 408$ min. Symbols represent the results obtained from the trajectories of A549 cells. Lines represent the results obtained from stochastic simulations based on the HO, CH, TH and CTH models. Black solid lines are Gaussian guidelines. Error bars in figure are standard deviations obtained from 50 simulations for each model.

conditional probability that the $i$ th A549 cell migrates a distance $r$ during a time interval $t$. $G(r, t)$ is then obtained by ensemble-averaging $g_i(r, t)$ over all A549 cells in the population, i.e.,

$$G(r, t) = \frac{1}{N}\sum_{i=1}^{N} g_i(r, t).$$

(4)

Note that the physical meanings of $g_i(r, t)$ and $G(r, t)$ are identical. While $G(r, t)$ is a property averaged over the population of A549 cells, $g_i(r, t)$ is the property of each single A549 cell. If cells were to undergo a random walk and follow the diffusion equation, both $G(r, t)$ and $g_i(r, t)$ would be Gaussian.

Our experiment shows that A549 cells undergo Fickian yet non-Gaussian cell migration. $\langle(\Delta r)^2(t)\rangle$ becomes linearly proportional to time $t$ at long timescales, and A549 cells reach a Fickian regime (gray symbols in Fig. 2(B)). More interesting is that $G(r, t)$ for A549 cells is non-Gaussian. As depicted in Fig. 2(C) and (D), all A549 cells have a non-Gaussian $G(r, t)$ at $t = 68$ and 408 min, which results from the violation of the central limit theorem and is unexpected from the conventional PRW model. $t = 68$ min and $t = 408$ min represent time scales before and after persistent time ($P = 78$ min), respectively. Our migration results show that the cell migration of A549 cells is anomalous both before and after persistent time ($P = 78$ min).

The spatiotemporal correlation function $g_i(r, t)$ of each single A549 cell is also non-Gaussian at $t = 68$ min and $t = 408$ min. Figure 3 depicts rescaled $g_i(r, t)$ of A549 cells. Here, $r^*$ denotes the root-mean-square displacement of each single A549 cell at time t, i.e., $r^* = \sqrt{\langle(\Delta r)^2(t)\rangle}$. One can compare the $g_i(r, t)$ of A549 cells (black circles in Fig. 3(A)) with a solid Gaussian guideline and find that the rescaled $g_i(r, t)$ of A549 cells is far from being Gaussian at $t = 68$ min. Some A549 cells migrate a very long distance, up to $r/r^* = 4$ at $t = 68$ min, which is different from a simple random walk that follows Gaussian statistics.

Non-Gaussian migration has also been observed in complex systems such as nanoparticle diffusion in polymeric materials[43–46]. Previous studies took heterogeneity into account and suggested theoretical models to
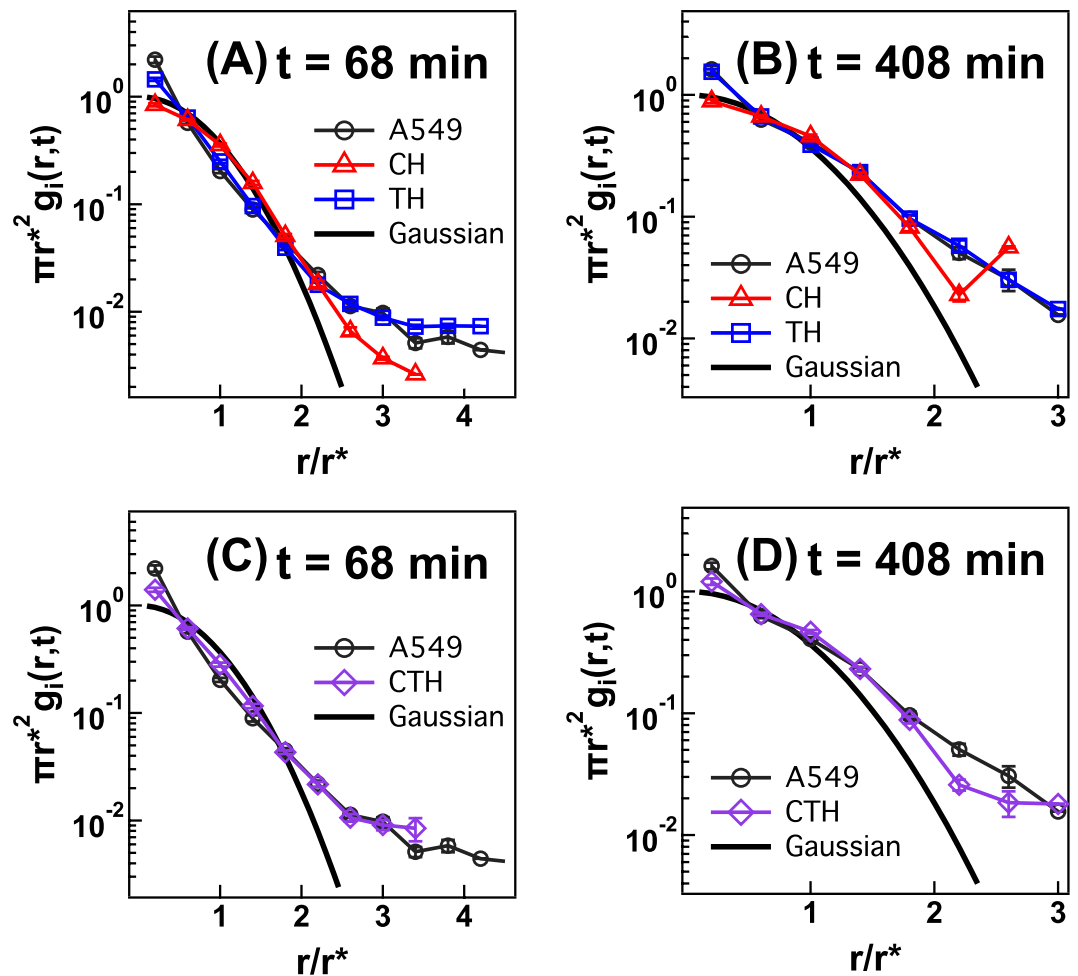
**Figure 3.** Rescaled spatiotemporal correlation functions ($\pi r^{*2} g_i(r, t)$) of individual A549 cells (black circles), which are not averaged over cell population. Rescaled $g_i(r, t)$'s are binned on the x-axis. Markers and error bars are the mean and standard error of the mean of binned data. Black solid lines are Gaussian guidelines. The simulation results for rescaled $g_i(r, t)$ obtained from the CH (red triangles) and TH (blue squares) models at (**A**) $t = 68$ min and (**B**) $t = 408$ min are presented. The simulation results for rescaled $g_i(r, t)$ obtained from the CTH (purple diamonds) model at (**C**) $t = 68$ min and (**D**) $t = 408$ min are also presented. $r^*(\equiv \sqrt{\langle (\Delta r)^2 (t) \rangle})$ is root-mean-square displacement of each cell trajectory at a given time $t$.

elucidate the non-Gaussian migration[47–49]. Chubynsky and his coworkers proposed a theory based on a stochastic differential equation where the diffusion coefficient of a particle also diffused, i.e., *diffusing diffusivity*[49]. In this approach, the temporal heterogeneity is incorporated, and the particle is supposed to undergo a temporal transition in the migration state. Other studies proposed that complex systems would consist of domains of different mobility: particles in fast domains migrate quickly and particles in slow domains migrate slowly, which is similar to the notion of cellular heterogeneity in the sense that different cells migrate with different migration capacity.

**Theoretical models for anomalous cell migration.**     To investigate how cellular and temporal heterogeneity affect cell migration, we consider four different theoretical models, as discussed below. We perform stochastic simulations based on those four theoretical models and compare the simulation results with the experimental results for A549 cells. The parameters required for stochastic simulations are obtained from the experiment on A549 cells. A summary of four models and parameters is described in the Table 1. In this study, the simulation time of all models is 1292 min, which is the same as the experimental time.

**HO model without heterogeneity.**     The PRW model has been employed extensively to interpret cell migration and is based on a stochastic differential equation as follows[11,15]:

$$\frac{dv_i(t)}{dt} = -\frac{1}{P}v_i + \frac{S}{\sqrt{P}}\hat{w},$$

(5)

where $v_i(t)$ is the velocity of the $i$ th cell. $\hat{w}$ represents the random noise of the Wiener process, and $P$ and $S$ denote the persistent time and the magnitude of the mean velocity of cells, respectively. The term $\frac{S}{\sqrt{P}}$ in front of the ran-

| | HO model | CH model | TH model | CTH model |
|---|---|---|---|---|
| Parameter | $S, P$ | $S_i, P_i$ | $A_{th}$ | $A_{cth}$ |
| How to get parameters | fitting to $\langle(\Delta r)^2(t)\rangle$ | $S_i$: mean speed of each cell $P_i$: decay time of $\langle\vec{v_i}(t)\vec{v_i}(0)\rangle/\langle\vec{v_i}^2(0)\rangle$ | fitting to $\langle(\Delta r)^2(t)\rangle$ | fitting to $\langle(\Delta r)^2(t)\rangle$ |
| $G(r,t)$ | Gaussian | non-Gaussian | non-Gaussian | non-Gaussian |
| $g_i(r,t)$ | Gaussian w/o cell-to-cell variation | Gaussian with cell-to-cell variation | non-Gaussian w/o cell-to-cell variation | non-gaussian with cell-to-cell variation |

**Table 1.** A summary of the HO, CH, TH, and CTH models.

dom noise corresponds to the magnitude of noise in HO model. One can find the relation between the magnitude ($S$) of the mean velocity and the magnitude of the noise by deriving the mean-squared velocity from the Eq. 5 [11]. We employ an integrator developed by Bussi et al.[50] to perform the numerical simulations of the Eq. 5. The integration time step is 0.01 min. If the cell-to-cell variation in cell migration was absent, all of the cells would have identical values of $P$ and $S$. For this homogeneous cell migration, the mean-square displacement $\langle(\Delta r)^2(t)\rangle$ in two dimensions is derived readily from the above stochastic differential equation as follows:

$$\langle(\Delta r)^2(t)\rangle = 2S^2P^2\left(e^{-\frac{t}{P}} + \frac{t}{P} - 1\right) + 4\sigma_{err}^2,$$
(6)

where $\sigma_{err}$ represents the magnitude of the localization error, which may arise due to spatial resolution in the experiment. Note that the first term of the Eq. 6 converges to zero as $t \to 0$, while the second term is a constant. The first term corresponds to the mean-square displacement of the *true position* of A549 cells. In the experiment, however, our position measurement of the A549 cell is unavoidably limited by the localization error due to the spatiotemporal resolution such that the localization error (the second term) should be incorporated into the Eq. 6 [13,15,51].

We obtain the values of $P$, $S$, and $\sigma_{err}$ by fitting $\langle(\Delta r)^2(t)\rangle$ of A549 cells with a sampling time of 34 min to Eq. 6 using the least squares method (See more details and Fig. S2 in Supporting Information): $S = 0.125$ $\mu$m/min, $P = 78$ min, and $\sigma_{err} = 1.66$ $\mu$m. Then, we perform stochastic simulations using the above HO model (the Eq. 5) and values of $P$, $S$, and $\sigma_{err}$. The localization error is added to the simulated trajectories to compare to the experiment as follows:

$$\hat{x}(t) = x(t) + \sigma_{err} \cdot W.$$
(7)

Here, $x(t)$ is the trajectory obtained from simulations, $W$ is white Gaussian noise of unit variance, and $\hat{x}(t)$ is the simulated trajectories with the localization error included. Even though $x(t)$ is the true position in simulations obtained by solving the Eq. 5, we need to compare the mean-square displacement of $\hat{x}(t)$ with $\langle(\Delta r)^2(t)\rangle$ of A549 cells because our position measurement for A549 cells is also limited by the localization error. We incorporate the localization error into the trajectories from all 4 models. As depicted schematically in Fig. 4(A), the HO model predicts that both $G(r,t)$ and $g_i(r,t)$ are identical to each other and are Gaussian.

**CH model with cellular heterogeneity.** Wu et al. proposed a theoretical approach based on the PRW model to incorporate cellular heterogeneity[15]. In this approach, they assumed that the cell-to-cell variation in migration capacity should be large enough that each cell would possess its own values of $P$ and $S$. The different values are assigned to the persistent time ($P_i$) and the magnitude of mean velocity ($S_i$) of the $i$ th cell. Then, the $i$ th cell obeys the following stochastic differential equation:

$$\frac{dv_i(t)}{dt} = -\frac{1}{P_i}v_i + \frac{S_i}{\sqrt{P_i}}\hat{w}.$$
(8)

In this CH model, the cell-to-cell variation in cell migration is incorporated into the distributions of $P_i$ and $S_i$.

To obtain the value of $P_i$ for the $i$ th A549 cell from time-lapse microscopy, we set the value of $P_i$ to a decay time when the normalized velocity autocorrelation function ($\langle\vec{v_i}(t)\vec{v_i}(0)\rangle/\langle\vec{v_i}^2(0)\rangle$) of the $i$ th cell equals $1/e$ (Fig. S3). Note that $\langle\vec{v_i}(t)\vec{v_i}(0)\rangle/\langle\vec{v_i}^2(0)\rangle$ of an individual A549 cell is not exponential but $P_i$ is still required to test the CH model. Therefore, we obtain a representative value for the persistent time $P_i$ by employing the equation $\langle\vec{v_i}(t)\vec{v_i}(0)\rangle/\langle\vec{v_i}^2(0)\rangle = 1/e$ using linear interpolation. Then, we round up $P_i$ to the multiples of the sampling time (34 min). Considering the spatial and temporal resolution in our experiment, the sampling time of 34 min is a sufficiently short timescale in this study. The values of the cell speed and continuous and discrete persistent times of individual cells are reported for comparison in the Supporting Information (Fig. S4). As discussed below, the CH model with the discrete persistent times obtained in this study successfully reproduces the $\langle(\Delta r)^2(t)\rangle$ and $G(r, t)$ of A549 cells. The value of $S_i$ is also obtained from the experiment by estimating the average magnitude of the $i$ th A549 cell's magnitude of mean velocity. The values of the cell speed and continuous and discrete persistent time of individual cells are reported in the Supporting Information (Fig. S4). We perform stochastic simulations by using Eq. 8 and the values of $P_i$ and $S_i$ obtained from A549 cells, which are conducted using the same integrator that was used in the HO model. The integration time step is 0.01 min.
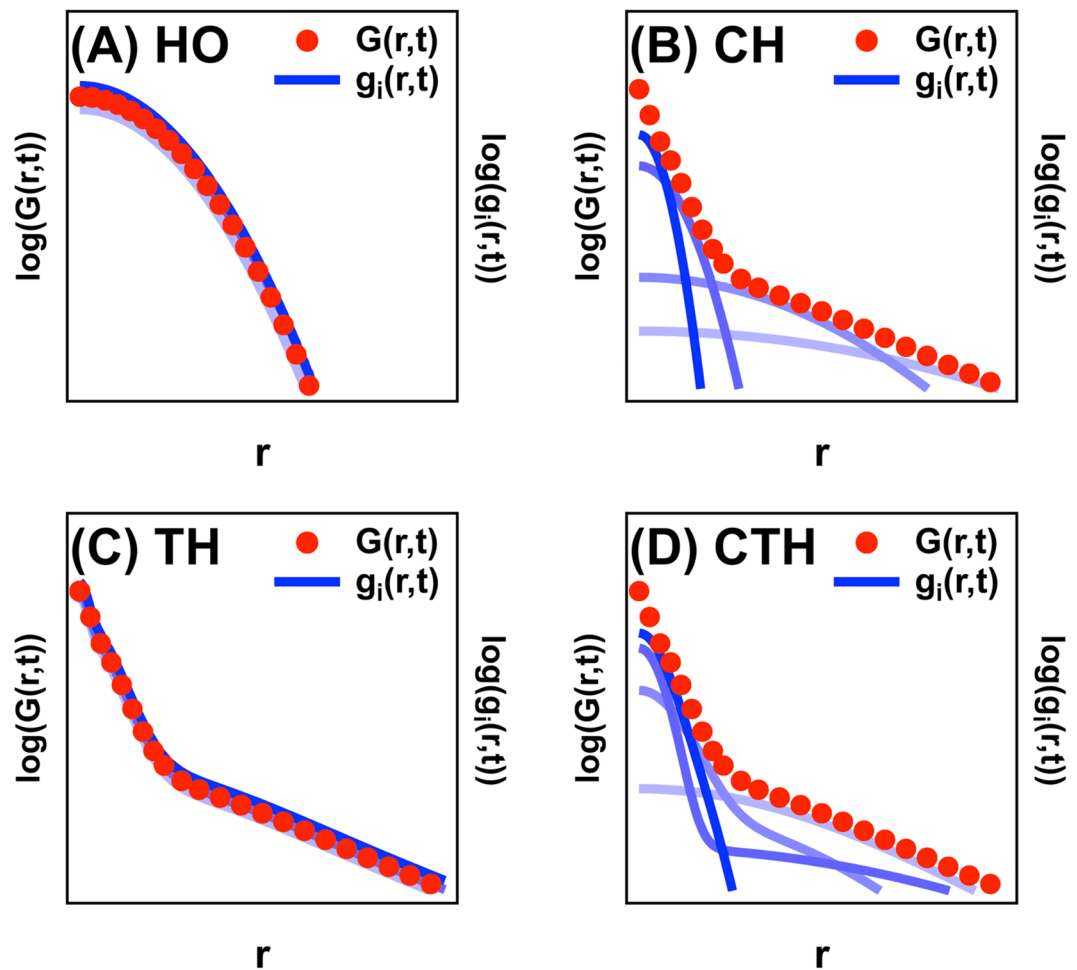
**Figure 4.** Schematic figures of the four different theoretical models employed in this study. Red symbols represent $G(r, t)$ while blue lines represent $g_i(r, t)$'s of individual single cells of each model. (**A**) HO model, where $G(r, t) = g_i(r, t)$ and both $G(r, t)$ and $g_i(r, t)$ are Gaussian. Since all cells are assumed to show identical dynamic behaviors, $g_i(r, t)$'s collapse onto one another. (**B**) In the CH model, $G(r, t) \neq g_i(r, t)$, and $G(r, t)$ is non-Gaussian, while $g_i(r, t)$ is Gaussian. Since cellular heterogeneity is introduced in the CH model, $g_i(r, t)$'s are different from each other. (**C**) In the TH model, $G(r, t) = g_i(r, t)$, and both $G(r, t)$ and $g_i(r, t)$ are non-Gaussian. Because cellular heterogeneity is not considered in the TH model, $g_i(r, t)$'s of individual cells collapse onto each other. (**D**) In the CTH model with both cellular and temporal heterogeneity, $G(r, t) \neq g_i(r, t)$, and both $G(r, t)$ and $g_i(r, t)$ are non-Gaussian. In the CTH model, cellular heterogeneity is incorporated such that $g_i(r, t)$'s of individual cells are different from each other.

In the CH model, each individual cell obeys the above stochastic differential equation based on the PRW model such that the $g_i(r, t)$ of each cell should be Gaussian. $g_i(r, t)$'s are all Gaussian but have different values for variance for different cells. However, $G(r, t)(\equiv \frac{1}{N}\sum_{i=1}^{N} g_i(r, t))$ becomes non-Gaussian because $G(r, t)$ is obtained by averaging over all cells (Fig. 4(B)).

**TH model with temporal heterogeneity.** If a cell were to undergo correlated transitions between different migration states, the conventional PRW model (with fixed values of $P$ and $S$ for each cell) would not describe the cell migration properly. The conventional PRW model assumes that the cells stay in a single migration state (characterized by $P$ and $S$). If the temporal heterogeneity were significant, the values of $P$ and $S$ of the cell would change with time such that the spatiotemporal correlation function $g_i(r, t)$ of the individual cell would be non-Gaussian, which is different from the CH model. Therefore, the conventional PRW stochastic differential equation (such as Eqs. 5 and 8) cannot be used.

To scrutinize whether the temporal heterogeneity alone could result in anomalous cell migration, we need to consider a case in which each cell undergoes the correlated transition in migration states but the population of cells is still homogeneous. This means that all of the cells in the population possess the same degree of the temporal heterogeneity such that $G(r, t) = g_i(r, t)$ for any $i$ (Fig. 4(C)). To realize such systems with temporal heterogeneity and *cellular homogeneity*, we propose the following stochastic differential equation (Eq. 9) and perform stochastic simulations in which the magnitude ($\beta(t)$) of the random noise changes with time $t$.

$$dr_i(t + dt) = \frac{dr_i(t)}{A_{th}} + \beta(t)\vec{\xi}, \tag{9}$$

Here, $dr_i(t)$ is the displacement vector of the $i$ th cell at time $t$. $\vec{\xi}$ is a unit vector with random orientation, which only determines the direction. $\beta(t)\vec{\xi}$ orients uniformly on the plane from 0 to $2\pi$ and is uncorrelated with that of previous steps at $t - dt$. $\beta(t)$ is the magnitude of the random noise and relates to the distance that the cell migrates during $dt$.

In order to simulate the TH model, one needs to obtain $\beta(t)$ and $A_{th}$ as discussed below. We sample $\beta(t)$ from the distribution $2\pi rG(r, t = P)$ of A549 cells using an inverse transform sampling method because the physical meaning of $2\pi rG(r, t = P)$ is the probability distribution function that the cell would migrate by $r$ during the time interval of $P$. Here, $P$ is 68 min due to a sampling time of 34 min (instead of $P = 78$ min obtained from fitting $\langle(\Delta r)^2(t)\rangle$). As discussed below, TH model using the discrete persistent time also reproduces successfully $\langle(\Delta r)^2(t)\rangle$ and $G(r, t)$ of A549 cells. In order to sample $\beta(t)$, we calculate the cumulative distribution function of $2\pi rG(r, t = P)$ and invert that function. The inverted cumulative distribution function transforms a random variable (uniformly sampled between 0 and 1) to a random variable $\beta(t)$. $\beta(t)$ changes with time $t$ such that each individual cell undergoes temporal transitions in migration states, which is not possible in the HO and CH models (See the Supporting Information for details). Note that $\beta(t)$ is not a fitting parameter but a stochastic variable sampled from $2\pi rG(r, t = P)$ of A549 cell trajectories. $A_{th}$ is a parameter that indicates how persistent the cell migration would be. In the OU process, for example, $A_{th}$ corresponds to $exp(dt/P)$. We obtain the value of $A_{th}$ ($= 2.5$) by fitting and reproducing the mean-square displacement of A549 cells (See more details and Fig. S6(A) in the Supporting Information.) We perform simulations of Eq. 9 using the Monte Carlo method. The integration time step, $dt$, is the same as the persistent time $P$.

As depicted schematically in Fig. 4(C), the TH model predicts that $G(r, t) = g_i(r, t)$ but both $G(r, t)$ and $g_i(r, t)$ are non-Gaussian. This is because all cells undergo the correlated transitions between different migration states in an identical fashion (Eq. 9) via $\beta(t)$.

### CTH model with both cellular and temporal Heterogeneity.

In the last theoretical model, we aim to combine both cellular and temporal heterogeneity by modifying the above TH model. In this model, each cell may undergo temporal changes in its migration state of different degrees for different cells. To fulfill both temporal and cellular heterogeneity in the numerical simulation, we propose the following stochastic differential equation and perform stochastic simulations:

$$\vec{dr_i}(t + dt) = \frac{\vec{dr_i}(t)}{A_{cth}} + \beta_i(t)\vec{\xi}, \tag{10}$$

where the value of $P_i$ is obtained from the $i$ th A549 cell and $\beta_i$ (the magnitude of the temporal noise of the $i$ th cell) is sampled randomly from $2\pi rg_i(r, t = P_i)$ of the $i$ th A549 cell instead of $2\pi rG(r, t = P)$ (As mentioned above, CTH model uses discrete $P_i$).

In order to simulate the CTH model, one needs to obtain $\beta_i(t)$ and $A_{cth}$. When sampling $\beta_i(t)$, we calculate the cumulative distribution function of $2\pi rg_i(r, t = P_i)$ and invert that function. The inverted cumulative distribution function transforms a random variable (uniformly sampled between 0 and 1) to a random variable $\beta_i(t)$. $\beta_i$ changes with time $t$ such that each cell may undergo a temporal transition between migration states (See in Fig. S5 in the Supporting Information). At the same time, $\beta_i$ is sampled from $g_i(r, t = P_i)$ of each cell, thus ensuring that the temporal transition varies with each cell. $\beta_i(t)$ is not a fitting parameter but a stochastic variable sampled from $2\pi rg_i(r, t = P_i)$. Therefore, one should obtain $P_i$ and $g_i(r, t = P_i)$ of each cell. $A_{cth} = 2.4$ is used in the CTH model to reproduce the averaged mean-square displacement of A549 cells (See Fig. S6(B) in Supporting Information). Note that $A_{cth}$ is identical for all cells and does not depend on each cell. But, $P_i$ and $g_i(r, t)$ of cells in the CTH model are heterogeneous, because $\beta_i$ is sampled from $2\pi rg_i(r, t = P_i)$ of each cell (See Figs S8(D) and S9 in Supporting Information). We also conduct simulations of the CTH model using the Monte Carlo method. In the CTH model, the integration time step is different for different cells because each cell has its own persistent time. The integration time step of the $i$ th cell is the persistent time of the $i$ th cell ($P_i$). The experimental results for $g_i(r, t = P_i)$ obtained from A549 cells are reproduced by the stochastic simulations based on the CTH model.

### Comparison of A549 cells and theoretical models.

*Average cell migration.* We perform stochastic simulations based on the four different theoretical models and compare the results to the experimental results for A549 cells. Even the HO model, into which we do not incorporate any heterogeneity, reproduces the experimental result for $\langle(\Delta r)^2(t)\rangle$ (Fig. 2(B)). This success of the HO model for $\langle(\Delta r)^2(t)\rangle$ has been well known. The other three models (with cellular and/or temporal heterogeneity considered) also reproduce $\langle(\Delta r)^2(t)\rangle$ successfully. To investigate how consistent the four theoretical models would be with the experiment for $\langle(\Delta r)^2(t)\rangle$, we estimate the p-values of the 4 models. We compare $\langle(\Delta r)^2(t)\rangle$ of A549 cells with $\langle(\Delta r)^2(t)\rangle$ obtained from 50 simulations for each model and calculate $\chi$-squared values[52]. The p-values (excluding the data point of $\langle(\Delta r)^2(t)\rangle$ at $t = 34$ min) are 0.4338 (HO model), 0.9742 (CH model), 1 (TH model), and 0.9792 (CTH model). The p-values of all of the models is more than 0.2 after 68 min, suggesting that $\langle(\Delta r)^2(t)\rangle$ from the simulations is consistent with the experiments for A549 cells. This indicates that $\langle(\Delta r)^2(t)\rangle$ is not a suitable physical quantity when one tries to investigate the effects of heterogeneity in the cell population on the cell migration.

The HO model fails, however, to reproduce the $G(r, t)$ of A549 cells at both short and long timescales. As shown in Fig. 2(C) and (D), the $G(r, t)$ obtained from the HO model is Gaussian at both $t = 68$ and 408 min. This is because the HO model (the conventional PRW model) should be, in principle, based on the random

| y | HO model | CH model | TH model | CTH model |
|---|---|---|---|---|
| $G(r, \ t = 68 \ min)$ | 0.0004976 | 0.0003021 | 0.0002649 | 0.0002018 |
| $G(r, t = 408 \ min)$ | 0.0002218 | 0.00008106 | 0.0001748 | 0.00004009 |
| Rescaled $g_i(r, t = 68 \ min)$ | | 0.0717 | 0.03452 | 0.03754 |
| Rescaled $g_i(r, t = 408 \ min)$ | | 0.05117 | 0.005978 | 0.02768 |

**Table 2.** Root mean-squared logarithmic error (RMSLE) of 4 models for $G(r, t)$ and rescaled $g_i(r, t)$.
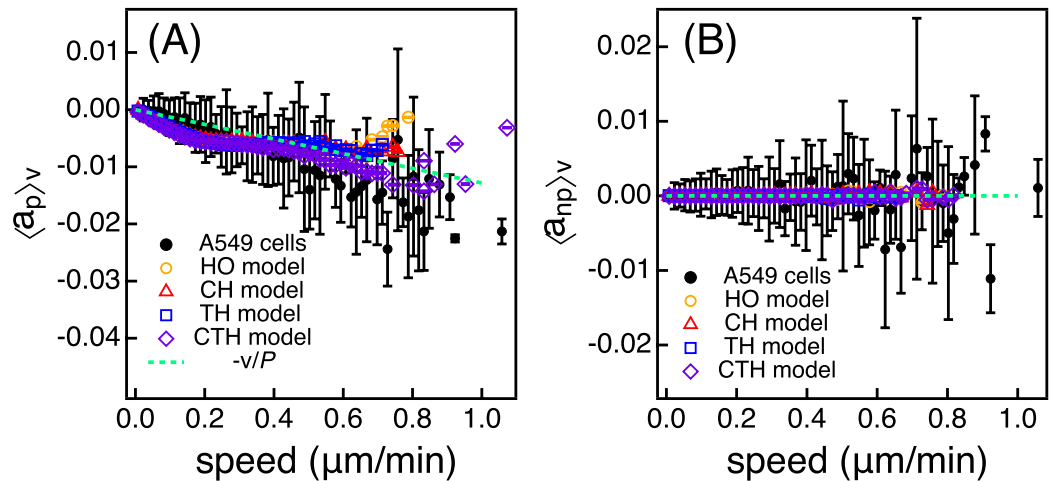


**Figure 5.** (**A**) $\langle a_p \rangle_v$ and (**B**) $\langle a_{np} \rangle_v$ as a function of cell speed for the HO (yellow empty circles), CH (red triangles), TH (blue squares), and CTH models (purple diamonds). Black filled circles represent A549 cells. Note that error bars in this figure indicate the standard errors of the mean. The dotted lines are guidelines with slopes of (**A**) $-v/P$ and (**B**) zero.

walk model such that $G(r, t)$ from the HO model is expected to be Gaussian. On the other hand, the $G(r, t)$ of A549 cells is non-Gaussian at $t = 68$ and $t = 408$ min. All of the other theoretical models (with heterogeneity considered to some extent) succeed in reproducing the non-Gaussian $G(r, t)$ at both short and long timescales. Wu et al. also showed that the CH model could explain the non-Gaussian $G(r, t)$ of HT1080 fibrosarcoma cells[15]. Metzner et al. illustrated that when they employed temporal heterogeneity, they could elucidate non-Gaussian $G(r, t)$ of MBA-MB-231 cells[36]. Our CTH model with both cellular and temporal heterogeneity also captures such non-Gaussian cell migration.

We estimate the root mean-squared logarithmic error (RMSLE) for $G(r, t)$ and rescaled $g_i(r, t)$ of four models as follows,

$$RMSLE = \sqrt{\frac{1}{N_t}\sum_{i=1}^{N_t}\{log(c_i + 1) - log(\hat{c}_i + 1)\}^2},$$

(11)

where $c$ denotes either $G(r, t)$ or rescaled $g_i(r, t)$ of four models, and $\hat{c}$ denotes the corresponding values from the experiment. $N_t$ is the number of data points in Fig. 2 and 3. As shown in Table 2, the CTH model shows the smallest value of RMSLE for $G(r, t = 68 \ min)$ and $G(r, t = 408 \ min)$. In case of the rescaled $g_i(r, t)$, the CH model has the largest RMSLE value while the RMSLE values of the TH and CTH model are comparable for the rescaled $g_i(r, t = 68 \ min)$. In case of rescaled $g_i(r, t = 408 \ min)$, the TH model produces a better result than the CTH and CH models. As shall be discussed in the following section, however, the TH model fails to capture the cellular heterogeneity of the A549 cell migration while the CTH model successfully reflects the cellular heterogeneity (See Fig. S8 in Supporting Information).

We estimate the magnitude of the deterministic term and the noise term by calculating the magnitude of the mean deviation of acceleration $(\vec{a}(t) = (\vec{v}(t + dt) - \vec{v}(t))/dt)$[13,51,53-55] to see whether the deterministic term and the noise term would be dependent on the cell speed. First, we calculate the component $(a_p \equiv \vec{a}(t) \cdot \vec{v}(t)/|\vec{v}(t)|)$ of the acceleration along the cell velocity $(\vec{v}(t))$ and then estimate the magnitude $(\langle a_p \rangle_v)$ of the conditional average of the component $a_p$ for a given $v = |\vec{v}(t)|$ and the magnitude $(|a_p - \langle a_p \rangle_v|)$ of the mean deviation of the component. From the component $(a_{np})$ of the acceleration orthogonal to the cell velocity, we estimate the conditional average $(\langle a_{np} \rangle_v)$ of the component $a_{np}$ for a given $v$ and the magnitude $(|a_{np} - \langle a_{np} \rangle_v|)$ of mean deviation of the component of acceleration orthogonal to cell velocity. Figure 5 shows that $\langle a_p \rangle_v$ of A549 cells and all models follows $-v/P$ and that $\langle a_{np} \rangle_v$ of A549 cells and all models is zero. Previous studies also showed that $\langle a_p \rangle_v$ followed $-v/P$ and $\langle a_{np} \rangle_v$ was zero regardless of speed[13,51], which is consistent with our results.
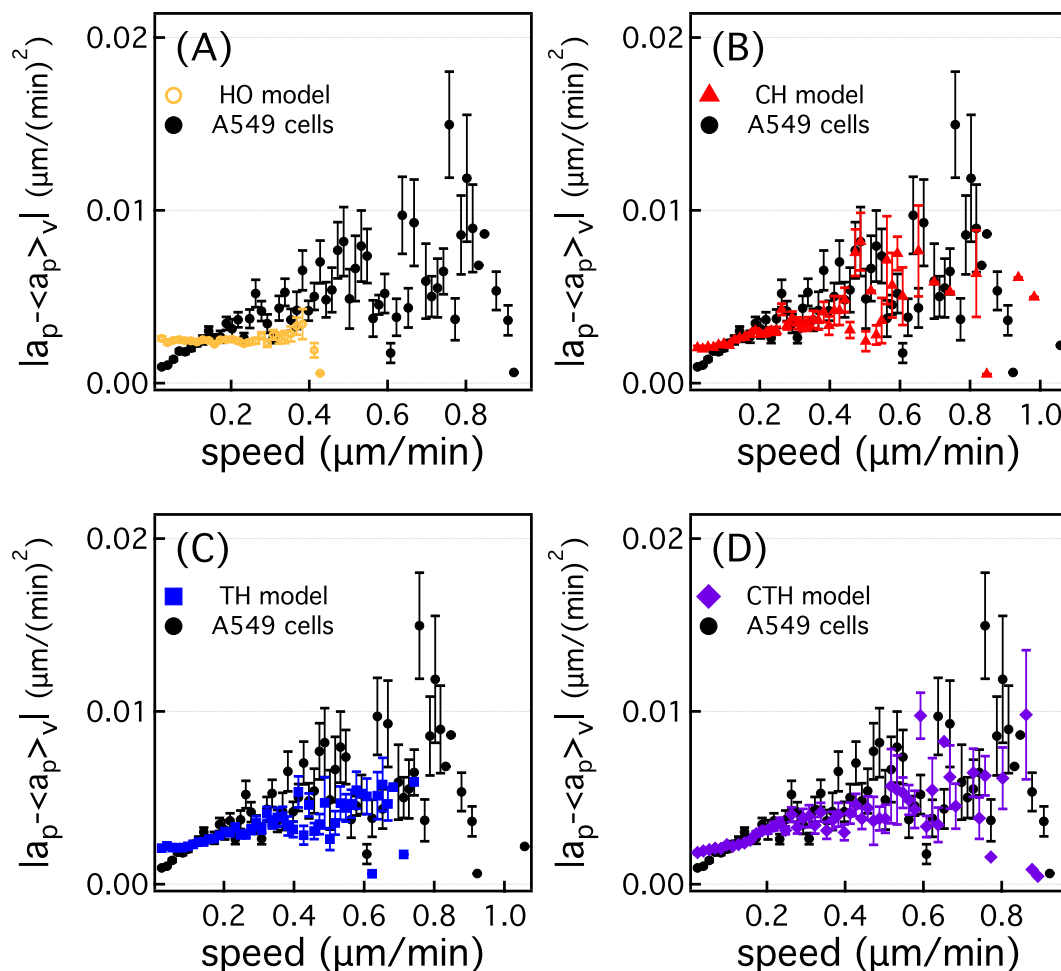
**Figure 6.** Magnitude of the mean deviation ($|a_p - \langle a_p \rangle_v|$) of the component of acceleration along the cell velocity as a function of cell speed for the (**A**) HO (yellow empty circles), (**B**) CH (red triangles), (**C**) TH (blue squares), and (**D**) CTH models (purple diamonds). Black filled circles represent A549 cells. Note that error bars in this figure indicate the standard error of the mean.

Figure 6(A) shows that the magnitude of noise of the HO model is uniform regardless of the speed, as expected. On the other hand, the magnitude of noise of A549 cells is dependent on the cell speed, unlike the HO model. As shown in Fig. 6(B), (C), and (D), the CH, TH and CTH models well reproduce $|a_p - \langle a_p \rangle_v|$ of A549 cells. In the case of the CH model, according to Eq. 8, the magnitude of the noise term is determined by $\frac{S_i}{\sqrt{P_i}}$. Therefore, each cell owns its own values for $S_i$ and $P_i$ such that cells with larger speed may have large $|a_p - \langle a_p \rangle_v|$, which makes the magnitude of noise of the CH model dependent on the speed. On the other hand, for the TH model, the magnitude of noise term of the TH model is determined by $\beta(t)$. If a cell were to have a large speed, the cell is supposed to migrate by a large distance during a given time and $\beta(t)$ may be large, too. This leads to the dependence of $|a_p - \langle a_p \rangle_v|$ on speed. Similarly, for the CTH model, the magnitude of noise term of the CTH model is determined by $\beta_i(t)$. CTH model takes into account the cellular heterogeneity such that a faster cell is likely to have a larger value for $\beta_i(t)$. This also leads to the dependence of $|a_p - \langle a_p \rangle_v|$ on speed. We also investigate the component of the acceleration orthogonal to the cell velocity (Fig. S10 in Supporting Information). Even for the orthogonal component, the CH, TH and CTH models reproduce the dependence of the noise on the cell speed.

Pedersen *et al.*[51] showed that $\langle a_p \rangle_v$ and $|a_p - \langle a_p \rangle_v|$ were dependent on the sampling time and the localization error. Our results could be also affected by the choice of sampling time and the magnitude of the localization error. Especially, $|a_p - \langle a_p \rangle_v|$ at a zero speed in all models is not zero due to the localization error. In addition, because we obtain the localization error from A549 cell data of the sampling time of 34 min, $|a_p - \langle a_p \rangle_v|$ of numerical simulations at time scales shorter than the sampling time is overestimated than that of A549 cells. Please note that, however, the CH, TH, and CTH models well describe the dependence of $|a_p - \langle a_p \rangle_v|$ on the cell speed of A549 cells after about 0.2 $\mu$m/min.

**Individual cell migration.** The stochastic simulations based on the CH model result in Gaussian $g_i(r, t)$ at short timescales of $t = 68$ min and long timescales of $t = 408$ min, which differs from that of A549 cells. Because we rescale both $g_i(r, t)$ and $r$ in Fig. 3(A), rescaled $g_i(r, t)$'s of different values of $P_i$ and $S_i$ in the CH model collapse

onto a single Gaussian curve (red symbols in Fig. 3(A)). The CH model can reproduce $G(r, t)$ but fails to reproduce $g_i(r, t)$. Because the CH model is based on the conventional PRW model but has different values of $P_i$ and $S_i$, each trajectory from the CH model should result in a Gaussian $g_i(r, t)$. This is different from A549 cells because the trajectory of each single A549 cell leads to non-Gaussian $g_i(r, t)$ at short timescales of $t = 68$ min and long timescales of $t = 408$ min (Fig. 3(B)).

On the other hand, the TH model captures the non-Gaussian behavior of the $g_i(r, t)$ of A549 cells. As shown in Fig. 3(A) and (B), rescaled $g_i(r, t)$'s of the TH model reproduce those of A549 cells. However, the TH model fails to capture the cell-to-cell variation in $g_i(r, t)$'s of A549 cells. As shown in Fig. S7, the 25th A549 cell migrates more than 40 $\mu$m at $t = 408$ min, while other A549 cells do not. The TH model does not reflect such cell-to-cell variation, and $g_i(r, t)$'s obtained from the TH model all collapse to a single curve (Fig. S8(C)).

The CTH model (which combines both cellular and temporal heterogeneity) reproduces not only rescaled $g_i(r, t)$'s at both short and long timescales (Fig. 3(C,D)) but also the cell-to-cell variation in $g_i(r, t)$. Rescaled $g_i(r, t)$'s from the CTH model are non-Gaussian and overlap with those of A549 cells at $t = 68$ and 408 min. As shown in Fig. S8(D), the $g_i(r, t)$ obtained using parameters extracted from the 25*th* cell is clearly distinguished from the $g_i(r, t)$ obtained using parameters extracted from the 212*th* cell, thus indicating that the CTH model captures well the cell-to-cell variation in $g_i(r, t)$.

The failure of the HO and TH models could be expected because those two models do not consider the cellular heterogeneity in a A549 cell population. Figure S9 in the Supporting Information depicts the distribution of the persistent time ($P_i$) obtained from the normalized velocity autocorrelation functions $\langle \vec{v_i}(t)\vec{v_i}(0)\rangle/\langle \vec{v_i}^2(0)\rangle$ of A549 cells and stochastic simulations. Because all of the cells in the TH model should undergo migration homogeneously, the $P_i$ values of these cells does not differ from one another much. On the other hand, $P_i$ of A549 cells has a broad distribution, which can be quantitatively reproduced only in the CH and CTH models. Our comparison of $G(r, t)$ and $g_i(r, t)$ between the experiment and simulations illustrates clearly that either cellular heterogeneity or temporal heterogeneity may explain $G(r, t)$ averaged over the A549 cells population. However, only when we incorporate both cellular and temporal heterogeneity together into the CTH model can we elucidate both individual ($g_i(r, t)$) and ensemble ($G(r, t)$) properties of the cell migration.

Previous studies for cell migration have also reported heterogeneous cell migration[15,36]. For example, Metzner et al.[36] proposed a statistical framework for modeling and analyzing heterogeneous cell migration, and showed that a cell migration model based on an autoregressive process of first order (AR-1 process) described the anomalous cell migration of MDA-MB-231 cells successfully. However, the cell migration model based on an AR-1 process still assumed that the migration of cells under the same condition would be determined by one single distribution, which corresponds to the TH model in this paper. On the other hands, the CTH model assumes that each cell would have different migration capacities by considering $P_i$ and $g_i(r, t)$ of an individual cell obtained in the experiment.

## Conclusion

We investigate the migration of A549 cells that possess intermediate characteristics between epithelial and mesenchymal states. A549 cells exhibit Fickian diffusion with $\langle (\Delta r)^2(t)\rangle \sim t^1$. However, the spatiotemporal correlation function ($G(r, t)$) averaged over the population of A549 cells is non-Gaussian at both short and long timescales. More interesting is that the spatiotemporal correlation function of individual cells ($g_i(r, t)$) is non-Gaussian at short and long timescales.

We find that such anomalous cell migration should be attributed to the heterogeneity in a A549 cell population. To elucidate the origin of the anomalous migration of A549 cells, we employ four different theoretical models and carry out stochastic simulations. The HO model does not take any type of heterogeneity into account: all of the cells are assumed to migrate with identical persistent times and magnitudes of the mean cell velocity. On the other hand, other theoretical models (CH, TH and CTH models) consider the cellular and/or temporal heterogeneity. The CH model considers the cellular heterogeneity with different values of persistent time ($P_i$) and magnitude of the mean cell velocity ($S_i$) for different cells. We obtain the values of $P_i$ and $S_i$ from the trajectory of each A549 cell, which are used again for the stochastic simulations based on Eq. 8.

In case of the TH model, the temporal heterogeneity, where cells may undergo transition between different migration states, is taken into account. In this case, the spatiotemporal correlation function ($g_i(r, t)$) of a single cell could be non-Gaussian. We assume in the TH model that all of the cells would undergo the temporal transition to the same extent such that the population of cells could be homogeneous. The quantity $\beta(t)$ (which is related to the magnitude of the mean velocity of cells) changes with time in this model to mimic the transition between migration states. For the CTH model, we incorporate both cellular and temporal heterogeneity into the model by allowing all of the cells to undergo temporal transitions in migration states but in their own ways.

All of the theories except the HO model successfully reproduce the ensemble migration properties ($\langle (\Delta r)^2(t)\rangle$ and $G(r, t)$) averaged over the cell population, which implies that one may employ either the cellular heterogeneity or the temporal heterogeneity to explain the average cell migration. However, when investigating the rescaled spatiotemporal correlation function $g_i(r, t)$ of individual cells, the CH model fails to reproduce rescaled $g_i(r, t)$ even qualitatively. Only the CTH model (with both cellular and temporal heterogeneity incorporated) reproduces the ensemble and individual spatiotemporal correlation functions of A549 cells successfully, which implies that both cellular and temporal heterogeneity together lead to anomalous cell migration.

Previous studies have reported that the temporal and/or cellular heterogeneity played critical roles in the cell migration of other cell lines[13,15,36]. The application of the CTH model to other cell lines should be a topic of interest. Previous studies focused mostly on the population averaged migration properties like $\langle (\Delta r)^2(t)\rangle$ and $G(r, t)$, for which models with only temporal or cellular heterogeneity (in case of this study, the CH and TH models) worked well. One has to obtain and investigate the migration properties of single cells (such as $g_i(r, t)$)

when trying to study the effects of both temporal and cellular heterogeneity on the cell migration in more details. Beyond studies on population-averaged cell migration, the CTH model may serve as a framework for the migration properties of single cells.

## Materials and Methods

**Cell culture.**    An experiment was performed with A549 lung adenocarcinoma cancer cell line. Cancer cells were maintained in Dulbecco's modified Eagle's medium (DMEM) (ThermoFisher, MA, USA) supplemented with 10% fetal bovine serum (FBS) and 0.1 % gentamycin at 37 °C in a $CO_2$ incubator. For real-time imaging, $5 \times 104$ cells, plated in the 60 *mm* cell culture plate, were monitored under an inverted time-lapse microscope (Lumascope 500, Etaluma, Carlsbad, USA) equipped with a 4× Olympus phase contrast objective. Time-lapse images ($1280 \times 800$ pixel) were obtained every 2 minutes for 24 or 48 hours with Lumaview 500 software.

In order to exclude the possibility that heterogeneity in A549 cell migration would be caused by cross-contamination of other cell types with a different genetic background, we performed short tandem repeat (STR) DNA fingerprint analysis using 16 STR loci on the chromosomes through the Korean Cell Line Bank (KCLB). The A549 cells employed in this experiment exhibited identical genetic markers to the reference A549 cells (Table S1 in Supporting Information).

The substrate was often coated with extracellular matrix (ECM) proteins such as collagen and fibronectin to mimic the ECM *in vitro*, which is normally broken by proteases secreted from the cancer cells in the cellular environment. However, in this study, we attempted to compare to theoretical models for cell migration and minimize the unpredictable factors. We therefore used commercially available tissue culture plates with no further treatment.

A549 lung cancer cells (drug resistant with K-Ras mutation) are one of the typical lung adenocarcinoma cell models and are widely used in cancer biology because their epithelial characteristics are converted readily to mesenchymal characteristics through epithelial mesenchymal transition (EMT) due to their plasticity[56]. Considering the dramatic change in cellular characteristics during EMT, which is closely associated with not only cancer malignancy but also metastasis[57], the development of a mathematical model for A549 cell movement would be important to further characterization of the movement of isogenic mesenchymal-type cancer cells[58].

**Cell tracking.**    When analyzing the trajectories of cells, we consider A549 cells that are not divided for a 24 h measurement period. We obtain the trajectories of 212 A549 cells in our experiment. The center of the cell nucleus is determined to be the position of the cell. The initial positions (or the centers of the cell nuclei) of A549 cells were identified manually, after which the cell positions were obtained by an automatic tracking method via a weight-average of pixel positions as follows:

$$\hat{x}_i = \frac{\int \int x S_i(x, y) dx dy}{\int \int S_i(x, y) dx dy},$$

where subscript $i$ is a cell index, $\hat{x}_i$ is the estimated x position of the $i$ th cell, and $(x, y)$ is the position vector of a pixel in the image from time-lapse microscopy. In our cell images, 1 pixel unit is 5 $\mu$m long. The score function ($S_i(x, y)$) represents the probability that the position of the pixel would correspond to the position of the $i$ th cell. The y position ($\hat{y}_i$) of the $i$ th cell can be estimated in the same way.

$S_i(x, y)$ consists of three functions of a pixel position $(x, y)$, i.e., $S_i(x, y) = S_i^1 S_i^2 S_i^3$. The first score function $S_i^1$ is used to identify the $i$ th cell at a current position at time $t$ with the $i$ th cell with a previous position at time $t-\Delta t$. Here, $\Delta t$ is the time resolution of the time-lapse microscopy in this study. Because $\Delta t = 2$ min is much smaller than the characteristic timescale of the cell migration, there should be little difference in the cell position between two consecutive images. Therefore, pixels around the previous position of cells should get a high score, for which we suggest the following score function for $S_i^1$:

$$S_i^1(x, y) = \frac{1}{N_1} e^{-\{(x-x_{i0})^2 + (y-y_{i0})^2\}/\sigma_1^2},$$

where $(x_{i0}, y_{i0})$ is the previous position of the $i$ th cell at time $t-\Delta t$, $N_1$ is a normalized constant and $\sigma_1 = 12$ pixels is a parameter used in this study. The second score function ($S_i^2$) is employed to locate the center of the cell nucleus. In the cell image obtained by the time-lapse microscope equipped with a phase contrast objective, a pixel around the center of cell nucleus has a lower degree of brightness ($f_i(x, y)$) than the pixels corresponding to the other parts of the cell (Fig. 1). Therefore, we define ($S_i^2$) as follows:

$$S_i^2(x, y) = \frac{1}{N_2} e^{-\{f_i(x,y)\}^2/\sigma_2^2},$$

where the degree of the brightness of a pixel $f_i(x, y)$ is rescaled from 0 to 100, $N_2$ denotes a normalized constant and $\sigma_2 = 10$ is a parameter for the brightness of pixels.

We introduce the third score function ($S_i^3$) to exclude pixels outside a cell, i.e.,

$$S_i^3(x, y) = \begin{cases} 1, & \text{inside cell area,} \\ 0, & \text{outside cell area.} \end{cases}$$

Here, we determine the pixel at position $(x, y)$ to be outside the cell if $f_i(x, y) - f_i(x_{i0}, y_{i0})$ would be larger than a threshold value ($=10$) because a pixel around the cell surface is brighter than pixels corresponding to the other parts of the cell (Fig. 1). All three score functions are estimated only when the pixel position is within a cutoff length ($=50\ \mu m$ (10 pixels)) from the previous cell position. Note that the size of A549 cells ranges from 70 to 100 $\mu$m. The representative trajectories of A549 cells are shown in Fig. 2(A).

In this study, we estimate and report the position of each cell by counting all of the certain digits in measurements plus the first uncertain digit. Because the estimated cell position is determined as the weighted average of pixel positions, the first uncertain digit corresponds to 1/10 pixel unit ($0.5\ \mu m$). (See the Supporting Information for details). We also report dynamic properties by setting the sampling time to 34 min even though time-lapse images were obtained every 2 minutes for 24 h. Two minutes is too short for a cell to migrate by more than a unit pixel (Fig. S2). Only after 34 min did the root of the mean-square displacement of A549 cells reach 5 $\mu$m (1 pixel unit). As shown in Fig. S2, the mean-square displacement of cells is independent of the sampling time.

## References

1. Detrich, H. *et al*. Intraembryonic hematopoietic cell migration during vertebrate development. *Proc. Natl. Acad. Sci. USA* **92**, 10713–10717 (1995).
2. Anderson, S. A., Eisenstat, D. D., Shi, L. & Rubenstein, J. L. R. Interneuron Migration from Basal Forebrain to Neocortex: Dependence on Dlx Genes. *Science* **278**, 474–476 (1997).
3. Young, H. M. *et al*. Dynamics of neural crest-derived cell migration in the embryonic mouse gut. *Dev. Biol.* **270**, 455–473 (2004).
4. Poste, G. & Fidler, I. J. The pathogenesis of cancer metasitasis. *Nature* **283**, 139–146 (1980).
5. Chicoine, M. R. & Silbergeld, D. L. Assessment of brain tumor cell motility *in vivo* and *in vitro*. *J. Neurosurg.* **82**, 615–622 (1995).
6. Wirtz, D., Konstantopoulos, K. & Searson, P. C. The physics of cancer: the role of physical interactions and mechanical forces in metastasis. *Nat. Rev. Cancer.* **11**, 512–522 (2011).
7. Krawczyk, W. S. A pattern of epidermal cell migration during wound healing. *J. Cell Biol.* **49**, 247–263 (1971).
8. Martin, P. Wound Healing–Aiming for Perfect Skin Regeneration. *Science* **276**, 75–81 (1997).
9. Farooqui, R. & Fenteany, G. Multiple rows of cells behind an epithelial wound edge extend cryptic lamellipodia to collectively drive cell-sheet movement. *J. Cell Sci.* **118**, 51–63 (2005).
10. Gail, M. H. & Boone, C. W. The locomotion of mouse fibroblasts in tissue culture. *Biophys. J.* **10**, 980–993 (1970).
11. Stokes, C. L., Lauffenburger, D. A. & Williams, S. K. Migration of individual microvessel endothelial cells: stochastic model and parameter measurement. *J. Cell Sci.* **99**, 419–430 (1991).
12. Thurner, S., Wick, N., Hanel, R., Sedivy, R. & Huber, L. Anomalous diffusion on dynamical networks: a model for interacting epithelial cell migration. *Physica A* **320**, 475–484 (2003).
13. Selmeczi, D., Mosler, S., Hagedorn, P. H., Larsen, N. B. & Flyvbjerg, H. Cell motility as persistent random motion: theories from experiments. *Biophys. J.* **89**, 921–931 (2005).
14. Dieterich, P., Klages, R., Preuss, R. & Schwab, A. Anomalous dynamics of cell migration. *Proc. Natl. Acad. Sci. USA* **105**, 459–463 (2008).
15. Wu, P. H., Giri, A., Sun, S. X. & Wirtz, D. Three-dimensional cell migration does not follow a random walk. *Proc. Natl. Acad. Sci. USA* **111**, 3949–3954 (2014).
16. Chen, Y.-C. *et al*. Single-cell Migration Chip for Chemotaxis-based Microfluidic Selection of Heterogeneous Cell Populations. *Sci. Rep.* **5**, 9980–9992 (2015).
17. Li, Y. *et al*. Cell migration microfluidics for electrotaxis-based heterogeneity study of lung cancer cells. *Biosensors and Bioelectronic* **89**, 837–845 (2017).
18. Poudineh, M. *et al*. Profiling Functional and Biochemical Phenotypes of Circulating Tumor Cells Using a Two-Dimensional Sorting Device. *Angew. Chem.* **56**, 163–168 (2017).
19. Banigan, E. J., Harris, T. H., Christian, D. A., Hunter, C. A. & Liu, A. J. Heterogeneous CD8+ T Cell Migration in the Lymph Node in the Absence of Inflammation Revealed by Quantitative Migration Analysis. *PLoS Comput Biol* **11**, e1004058–20 (2015).
20. Fricke, G. M., Letendre, K. A., Moses, M. E. & Cannon, J. L. Persistence and Adaptation in Immunity: T Cells Balance the Extent and Thoroughness of Search. *PLoS Comput. Biol.* **12**, e1004818–23 (2016).
21. Passucci, G., Brasch, M. E., Henderson, J. H., Zaburdaev, V. & Lisa, M. M. Identifying the mechanism for superdiffusivity in mouse fibroblast motility. *PLoS Comput. Biol.* **15**, e1006732 (2019).
22. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002).
23. Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L. & Leibler, S. Bacterial Persistence as a Phenotypic Switch. *Science* **305**, 1622–1625 (2004).
24. Kussell, E. & Leibler, S. Phenotypic Diversity, Population Growth, and Information in Fluctuation Environments. *Science* **309**, 2075–2078 (2005).
25. Huang, S. Non-genetic heterogeneity of cells in development: more than just noise. *Development* **136**, 3853–3862 (2009).
26. Altschuler, S. J. & Wu, L. F. Cellular Heterogeneity: Do Differences Make a Difference? *Cell* **141**, 559–563 (2010).
27. Sakaue-Sawano, A. *et al*. Visualizing Spatiotemporal Dynamics of Multicellular Cell-Cycle Progression. *Cell* **132**, 487–498 (2008).
28. Boucher, H., Vanneaux, V., Domet, T., Parouchev, A. & Larghero, J. Circadian Clock Genes Modulate Human Bone Marrow Mesenchymal Stem Cell Differentiation, Migration and Cell Cycle. *PLoS One* **11**, e0146674 (2016).
29. Wesley, U. V., Bove, P. F., Hristova, M., McCarthy, S. & van der Vliet, A. Airway Epithelial Cell Migration and Wound Repair by ATP-mediated Activation of Dual Oxidase 1. *J. Biol. Chem.* **282**, 3213–3220 (2007).
30. Minton, K. Cell migration: Coordinating calcium signalling. *Nat. Rev. Mol. Cell Biol.* **15**, 152 (2014).
31. Atkins, P. & Paula, D. J. *Atkins' physical chemistry* (Oxford University Press, New York, 2014).
32. Dunn, G. A. & Brown, A. F. A unified approach to analysing cell motility. *J. Cell Sci.* **8**, 81–102 (1987).
33. Tranquillo, R. T., Lauffenburger, D. A. & Zigmond, S. H. A Stochastic Model for Leukocyte Random Motility and Chemotaxis Based on Receptor Binding Fluctuations. *J. Cell Biol.* **106**, 303–309 (1988).
34. Hartman, R. S., Lau, K., Chou, W. & Coates, T. D. The fundamental motor of the human neutrophil is not random: evidence for local non-Markov movement in neutrophils. *Biophys. J.* **67**, 2535–2545 (1994).
35. Czirok, A., Schlett, K., Madarasz, E. & Viscek, T. Exponential Distribution of Locomotion Activity in Cell Cultures. *Phys. Rev. Lett.* **81**, 3038–3041 (1998).
36. Metzner, C. *et al*. Superstatistical analysis and modelling of heterogeneous random walks. *Nat. Commun.* **6**, 7516 (2015).
37. Thapa, S., Lukat, N., Selhuber-Unkel, C., Cherstvy, A. G. & Metzler, R. Transient superdiffusion of polydisperse vacuoles in highly motile amoeboid cells. *J. Chem. Phys.* **150**, 144901 (2019).

38. Cherstvy, A. G., Nagel, O., Beta, C. & Metzler, R. Non-Gaussianity, population heterogeneity, and transient superdiffusion in the spreading dynamics of amoeboid cells. *Phys. Chem. Chem. Phys.* **20**, 23034 (2018).
39. Banerjee, B., Bhattacharya, D. & Shivashankar, G. V. Chromatin Structure Exhibits Spatio-Temporal Heterogeneity within the Cell Nucleus. *Biophys. J.* **91**, 2297–2303 (2006).
40. Brock, A., Chang, H. & Huang, S. Non-genetic heterogeneity—a mutation-independent driving force for the somatic evolution of tumours. *Nat. Rev. Genet.* **10**, 336–342 (2009).
41. Yosef, N. & Regev, A. Impulse Control: Temporal Dynamics in Gene Transcription. *Cell* **144**, 886–896 (2011).
42. Nachman, I., Regev, A. & Ramanathan, S. Dissecting Timing Variability in Yeast Meiosis. *Cell* **131**, 544–556 (2007).
43. Yamamoto, R. & Onuki, A. Dynamics of highly supercooled liquids: Heterogeneity, rheology, and diffusion. *Phys. Rev. E* **58**, 3515–3529 (1998).
44. Wang, B., Anthony, S. M., Bae, S. C. & Granick, S. Anomalous yet Brownian. *Proc. Natl. Acad. Sci. USA* **106**, 15160–15164 (2009).
45. Bhattacharya, S. *et al.* Plasticization of Poly(vinylpyrrolidone) Thin Films under Ambient Humidity: Insight from Single-Molecule Tracer Diffusion Dynamics. *J. Phys. Chem. B* **117**, 7771–7782 (2013).
46. He, W. *et al.* Dynamic heterogeneity and non-Gaussian statistics for acetylcholine receptors on live cell membrane. *Nat. Commun.* **7**, 11701–11708 (2016).
47. Wang, B., Kuo, J., Bae, S. C. & Granick, S. When Brownian diffusion is not Gaussian. *Nat. Mater.* **11**, 481–485 (2012).
48. Paeng, K. & Kaufman, L. J. Single molecule rotational probing of supercooled liquids. *Chem. Soc. Rev.* **43**, 977–989 (2014).
49. Chubynsky, M. V. & Slater, G. W. Diffusing Diffusivity: A Model for Anomalous, yet Brownian, Diffusion. *Phys. Rev. Lett.* **113**, 098302 (2014).
50. Bussi, G. & Parrinello, M. Accurate sampling using Langevin dynamics. *Phys. Rev. E* **75**, 056707 (2007).
51. Pedersen, J. N. *et al.* How to connect time-lapse recorded trajectories of motile microorganisms with dynamical models in continuous time. *Phys. Rev. E* **94**, 062401 (2016).
52. Moninier, N. *et al.* Bayesian Approach to MSD-Based Analysis of Particle Motion in Live Cells. *Biophys. J.* **103**, 616–626 (2012).
53. Amselem, G., Theves, M., Bae, A., Bodenschatz, E. & Beta, C. A Stochastic Description of *Dictyostelium Chemotaxis*. *PLoS One* **7**, e37213 (2012).
54. Li, L., Cox, E. C. & Flyvbjerg, H. "Dicty Dynamics": Dictyostelium motility as persistent random motion. *Phys. Biol.* **8**, 046006 (2011).
55. Bödeker, H. U., Beta, C., Frank, T. D. & Bodenschatz, E. Quantitative analysis of random ameboid motion. *Europhys. Lett.* **90**, 28005 (2010).
56. Bae, G.-Y. *et al.* Chronic TGFβ stimulation promotes the metastatic potential of lung cancer cells by Snail protein stabilization through integrin β3-Akt-GSK3β signaling. *Oncotarget* **7**, 25366–25376 (2016).
57. Polyak, K. & Weinberg, R. A. Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. *Nat. Rev. Cancer* **9**, 265–273 (2009).
58. Wong, I. Y. *et al.* Collective and individual migration following the epithelial–mesenchymal transition. *Nat. Mater.* **13**, 1063–1071 (2014).

## Acknowledgements

## Author contributions

H.-J.C. and B.J.S. designed and directed the study. O.-S.K. and H.-J.C. performed the experiments on A549 lung cancer cells. T.K. and B.J.S. developed the theoretical models and performed stochastic simulations. All of the authors wrote, edited and approved the manuscript together.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at https://doi.org/10.1038/s41598-019-52480-3.

**Correspondence** and requests for materials should be addressed to H.-J.C. or B.J.S.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.