



Article

Time-Series Clustering of Single-Cell Trajectories in Collective Cell Migration

Zhuohan Xin ^{1,†} , Masashi K. Kajita ^{2,3,†} , Keiko Deguchi ⁴, Shin-ichiro Suye ^{1,3,4} and Satoshi Fujita ^{1,3,4,*} 

¹ Department of Advanced Interdisciplinary Science and Technology, University of Fukui, Fukui 910-8507, Japan

² Department of Applied Chemistry and Biotechnology, University of Fukui, Fukui 910-8507, Japan

³ Organization for Life Science Advancement Programs, University of Fukui, Fukui 910-8507, Japan

⁴ Department of Frontier Fiber Technology and Science, University of Fukui, Fukui 910-8507, Japan

* Correspondence: fujitas@u-fukui.ac.jp; Tel.: +81-776-27-9969

† These authors contributed equally to this work.

Simple Summary: In this study, we normalized trajectories containing both mesenchymal and epithelial cells to remove the effect of cell location on clustering, and performed a dimensionality reduction on the time series data before clustering. When the clustering results were superimposed on the trajectories prior to normalization, the results still showed similarities in location, indicating that this method can find cells with similar migration patterns. These data highlight the reliability of this method in identifying consistent migration patterns in collective cell migration.

Abstract: Collective invasion drives multicellular cancer cells to spread to surrounding normal tissues. To fully comprehend metastasis, the methodology of analysis of individual cell migration in tissue should be well developed. Extracting and classifying cells with similar migratory characteristics in a colony would facilitate an understanding of complex cell migration patterns. Here, we used electrospun fibers as the extracellular matrix for the in vitro modeling of collective cell migration, clustering of mesenchymal and epithelial cells based on trajectories, and analysis of collective migration patterns based on trajectory similarity. We normalized the trajectories to eliminate the effect of cell location on clustering and used uniform manifold approximation and projection to perform dimensionality reduction on the time-series data before clustering. When the clustering results were superimposed on the trajectories before normalization, the results still exhibited positional similarity, thereby demonstrating that this method can identify cells with similar migration patterns. The same cluster contained both mesenchymal and epithelial cells, and this result was related to cell location and cell division. These data highlight the reliability of this method in identifying consistent migration patterns during collective cell migration. This provides new insights into the epithelial–mesenchymal interactions that affect migration patterns.

Keywords: collective cell migration; electrospinning; dimensionality reduction; clustering; migration pattern



Citation: Xin, Z.; Kajita, M.K.; Deguchi, K.; Suye, S.-i.; Fujita, S. Time-Series Clustering of Single-Cell Trajectories in Collective Cell Migration. *Cancers* **2022**, *14*, 4587. <https://doi.org/10.3390/cancers14194587>

Academic Editors: Mizuho Nishio and Alessandro Di Stefano

Received: 11 August 2022

Accepted: 16 September 2022

Published: 22 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Collective cell migration is a major mechanism of cell movement and is required for many physiological processes. This process is involved in biological processes, such as wound healing and tissue regeneration and several pathologies, including cancer metastasis. It relies on establishing polarized cell populations, which are asymmetrical in their front and rear. Polarization is caused by the presence of two types of cells, leader cells and follower cells, which differ in morphology and properties [1]. Leader cells with a large lamellipodial structure at the leading-edge guide follower cells through cell–cell interactions during collective migration [2]. This guidance has a prolonged effect on cell migration patterns. Although each cell migrates in different directions, once it becomes part of a

population, mutual coordination between cells allows it to migrate quickly and to maintain adhesion with neighboring cells [3]. The complex mechanism of collective migration is remains unknown, but migration patterns resulting from cellular interactions must be thoroughly investigated.

Epithelial–mesenchymal transition (EMT) involves a series of biological changes in which epithelial cells lose their epithelial properties and acquire mesenchymal properties [4]. EMT-induced cells change in cell functions and phenotypes such as losing apical-basal polarity and gaining front–back polarity, changes in cell morphology, and acquisition of invasiveness due to the downregulation of epithelial genes and upregulation of mesenchymal genes [5–7]. These experimental observations support the idea that EMT-induced cells, along with the fibrous structure of the extracellular matrix (ECM), promote colony migration as leader cells. To analyze the interaction between the leader cells and other cells, studies that tracked the position of the leading edge of a single cell have elucidated the mechanism by which the deformation of intracellular microtubules affects migration [8,9]. How the positional relationship of cells in population changes has not been clarified. This study focused on the positional relationship of individual cells in a cell population to elucidate their migration patterns.

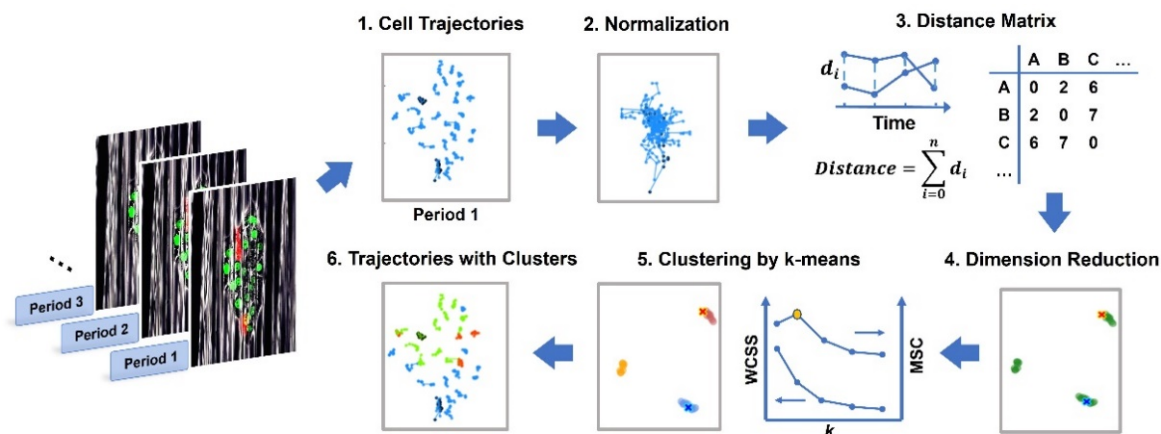
Cells migrate via actin polymerization at the leading edge of the colony to produce lamellipodia. For adhesion, they largely use a molecular mechanism involving integrin and related adhesion proteins [10–12]. In morphogenesis and cancer models, the loss of E-cadherin weakens cell connections, which causes cell detachment and migration during EMT [13,14]. E-cadherin is the main mediator of collective cell–cell interaction. The process of collective cell migration can be described to varying degrees using molecular mechanisms and mechanical models [15–17]. However, how these interactions integrate multiple factors that influence migration patterns requires further investigation [18]. We used a simpler approach based on the clustering of single-cell trajectories to explore the migration patterns between cells.

The study of cell migration patterns usually requires recording cell trajectories within observation windows comprising multiple sets of timeseries data. Timeseries data are characterized by high dimensionality and large data volumes [19]. Clustering such complex objects can reveal interesting patterns. As an unsupervised learning technique, k-means clustering is a commonly used clustering method. The main idea behind this clustering method is to minimize the total distance (usually the Euclidean distance) between all the objects in a cluster and their cluster centers. Cluster centers are defined as the average vectors of the objects in the cluster; however, time series clustering poses additional challenges. For cell trajectories, this increases the difficulty and computational effort of clustering based on the time dimension and poor clustering performance owing to the different locations of cell trajectories at each time record point, and similar problems have been observed in the phylogenetic analysis [20,21].

One solution is dimensionality reduction of time series data and subsequent clustering. Commonly used dimensionality reduction methods to maintain the two-by-two distance structure of the dataset include principal component analysis (PCA) [22], multi-dimensional scaling (MDS) techniques [23], and streamwise learning techniques such as t-distributed stochastic neighbor embedding (t-SNE), which can maintain the local structure of the data [24] and uniform manifold approximation and projection (UMAP) [25]. In particular, UMAP is efficient for processing large amounts of data in response to the global structure, providing higher-quality visualization. Clustering is often used in biological disciplines for the identification of functionally relevant genes and functional clustering of gene expression data [26,27]; however, to our knowledge, this is the first report to analyze the range of influence of leader cells on collective migration by classifying the movement pattern with clustering analysis of individual cell tracking.

In this study, we attempted to identify similar migration patterns among cells (Scheme 1). As an approach to quantitatively analyze cell–cell interactions, co-culture of EMT-induced mesenchymal cells by the addition of transforming growth factor β 1

(referred as TGF (+)) and EMT non-induced cells (TGF (−)) was performed to imitate colony migration on aligned electrospun fibers mimicking the ECM fibrous architecture. Tracking of cell migration by time-lapse observation followed by dimensionality reduction and clustering was performed to find similar migration patterns, where they are not only related to location but also influenced by cell division.



Scheme 1. Schematic of time-series clustering of cell trajectories in collective cell migration. After collecting cell migration movies by time-lapse microscopy, cell trajectories were generated manually by period using ImageJ. Second, the trajectories of each period were normalized to convert their starting points to the origin (0, 0). Third, the distance matrix is calculated from the Euclidean distance of each pair of trajectories. Fourth, the UMAP dimensionality reduction algorithm is used to visualize the data in two dimensions. Fifth, the data are clustered by k-means, where the number of clusters is determined by the mean silhouette coefficient. Finally, the clusters are combined with the original trajectories to analyze similar migration patterns in collective cell migration.

2. Materials and Methods

2.1. Electrospinning

To prepare the cell culture scaffold, 20 wt% polystyrene (PS) solutions dissolved in tetrahydrofuran (THF) were electrospun on a cover slip (18 × 24 mm) using a commercial electrospinning setup (NANON, MECC, Fukuoka, Japan) under the following conditions: flow rate, 0.1 mL/h; linear velocity of collector rotation, 15.7 m/s; electric field, 2.5 kV/cm; spinning time, 60 min. The electrospun fibers were treated with O₂ plasma (40 kHz/100 W, 25 Pa, 30 s) and coated with 10 µg/mL fibronectin (37 °C, 2 h).

2.2. Cell Culture and Time-Lapse Observation

We used the murine mammary gland cell line NMuMG (ATCC CRL-1636) and fluorescently labeled NMuMG cells (NMuMG-DsRed). The gene induction of pDsRed2-C1 was conducted according to the guidelines of the ethics committee of the institution. Cells were maintained in DMEM supplemented with 10% FBS at 37 °C, 5% CO₂, and 95% humidity. EMT was induced by adding 10 ng/mL TGF-β1 for 3 days [28]. To prepare cell aggregates, 500 µL of the cell suspension of non-labeled NMuMG cells (designated as TGF (−) cells) and EMT-induced NMuMG-DsRed cells (designated as TGF (+) cells) at a ratio of 1:4 at 1 × 10⁵ cells/mL were seeded onto a 24-well grided-plate (Elplasia #4445; Kuraray, Tokyo, Japan) and incubated for 3 h. Cell aggregates formed in the grids were harvested by gentle pipetting. They were suspended in 500 µL of medium containing 25 µL of CellLight™ Histone 2B-GFP and BacMam 2.0, (Thermo Fisher Scientific K.K., Tokyo, Japan), inoculated on fibronectin-coated fibers, cultured for 24 h, and then used for the migration assay. Time-lapse microscopy was performed for 24 h at intervals of 15 min using Biostation (Nikon, Tokyo, Japan) and analyzed using ImageJ (ver. 1.53 e). The captured fluorescence images were binarized and the cells were tracked using a plug-in (TrackMate). The location of the center of the cell nucleus was determined by manual tracking which started from the

last time slice to the first time slice backward. The cell was identified on the basis of the locations of the cell nucleus between two adjacent slices and assigned ID numbers to all cells. The slice when cell division occurred was determined by comparing the distance between neighboring cell marker points.

2.3. Cell Trajectory Data and Its Normalization

Let

$$V_i^{(p)} := \langle \mathbf{v}_{i,1}^{(p)}, \dots, \mathbf{v}_{i,n}^{(p)} \rangle \quad (1)$$

be the i th cell's trajectory of the p th time period, where $\mathbf{v}_{i,l}^{(p)} \in \mathbb{R}^2$ is the cell position at l th time slice of the p th time period and n is the number of time slices in a single time period.

The cells were divided several times during the observation. After manually marking the positions of the cell nuclei and finding mother cells with daughter cells, the trajectories of two daughter cells were unified with the mother cell prior to division. To remove the effect of the initial cell positions from the cell trajectory clustering, we first normalized the i th cell's trajectory of p th time period $V_i^{(p)}$ such that its initial position $\mathbf{v}_{i,1}^{(p)}$ was located at the origin $(0,0)$. The normalized trajectory is given by

$$Z_i^{(p)} := \langle \mathbf{z}_{i,1}^{(p)}, \dots, \mathbf{z}_{i,n}^{(p)} \rangle, \quad (2)$$

where $\mathbf{z}_{i,l}^{(p)} := \mathbf{v}_{i,l}^{(p)} - \mathbf{v}_{i,1}^{(p)} = (z_{i,l,1}^{(p)}, z_{i,l,2}^{(p)})^\top$.

2.4. Distance Matrix of Cell Trajectories

To evaluate the similarity of cell trajectories, we defined the pairwise distance between two trajectories as follows: the distance between two trajectories at the p th time period, $Z_i^{(p)}$ and $Z_j^{(p)}$, is defined by

$$D(Z_i^{(p)}, Z_j^{(p)}) := \sum_{l=1}^n \delta(\mathbf{z}_{i,l}^{(p)}, \mathbf{z}_{j,l}^{(p)}), \quad (3)$$

where $\delta(\mathbf{z}_{i,l}^{(p)}, \mathbf{z}_{j,l}^{(p)})$ is the distance between two points, $\mathbf{z}_{i,l}^{(p)}$ and $\mathbf{z}_{j,l}^{(p)}$, which are respectively i th and j th cell positions at l th time slice of p th time period.

In this study, we employed the Euclidean distance to define the distance function $\delta(\mathbf{z}_{i,l}^{(p)}, \mathbf{z}_{j,l}^{(p)})$:

$$\delta(\mathbf{z}_{i,l}^{(p)}, \mathbf{z}_{j,l}^{(p)}) := \sqrt{(z_{i,l,1}^{(p)} - z_{j,l,1}^{(p)})^2 + (z_{i,l,2}^{(p)} - z_{j,l,2}^{(p)})^2}. \quad (4)$$

Based on the pairwise distance between the two trajectories $D(Z_i, Z_j)$, the distance matrix is defined as follows:

$$D := \begin{bmatrix} D_{1,1} & \cdots & D_{1,N} \\ \vdots & \ddots & \vdots \\ D_{N,1} & \cdots & D_{N,N} \end{bmatrix}, \quad (5)$$

where $D_{i,j} := D(Z_i, Z_j)$ and N is the total number of cells and their trajectories.

2.5. Two-Dimensional Representation of Cell Trajectories Based on a Distance Matrix

To visualize the similarity of the cell trajectories, we embedded each cell trajectory $Z_i^{(p)}$ into a two-dimensional space based on the distance matrix D . We employed t-SNE [24,29–31] and UMAP [25] as the embedding methods using Python (scikit-learn 1.1.2;

umap-learn 0.5.3). By applying these two methods, we obtain a set of cell trajectories in p th time period represented in a two-dimensional space:

$$\mathcal{D}^{(p)} := \{x_1^{(p)}, \dots, x_N^{(p)}\}, \quad (6)$$

where $x_i^{(p)} \in \mathbb{R}^2$ is the i th cell trajectory of the p th time period in two-dimensional space.

2.6. Clustering of Cell Trajectories

Next, we classified cell trajectories based on their similarities. Here, we employed the k-means clustering method [32] for classification. Using k-means clustering, we partitioned the N samples of $\mathcal{D}^{(p)}$ into $K_p (\leq N)$ sets of the p th time period $M_p = \{M_1^{(p)}, M_2^{(p)}, \dots, M_{K_p}^{(p)}\}$ based on the similarities among the samples.

Let

$$\mathcal{M}_p = \{\mu_1^{(p)}, \dots, \mu_{K_p}^{(p)}\} \quad (7)$$

be a set of the mean of points in each cluster at the p th time period, where $\mu_k^{(p)}$ is the mean of points in the k th cluster at the p th time period. We also introduce a variable that denotes whether a set $M_k^{(p)}$ contains $x_i^{(p)}$ or not:

$$q_{i,k}^{(p)} = \begin{cases} 1 & (x_i^{(p)} \in M_k^{(p)}) \\ 0 & (x_i^{(p)} \notin M_k^{(p)}) \end{cases}. \quad (8)$$

Using the variables, the objective function of k-means clustering, the within-cluster sum of squares (WCSS), is defined as

$$J(q_{i,k}^{(p)}, \mu_k^{(p)}) = \sum_{i=1}^N \sum_{k=1}^{K_p} q_{i,k}^{(p)} \|x_i^{(p)} - \mu_k^{(p)}\|^2. \quad (9)$$

The optimal solution is a parameter that minimizes the objective function $J(q_{i,k}^{(p)}, \mu_k^{(p)})$. We numerically solved the optimization problem using the k-means algorithm. Notably, the solution of the k-means algorithm has an initial value dependency. Therefore, we tested several initial values and employed a clustering result that minimized the objective function.

The number of clusters at the p th time period K_p is a hyperparameter of k-means clustering. Therefore, we optimized K_p based on the silhouette analysis [33]. Silhouette analysis evaluates clustering results based on the degree of aggregation within clusters and the degree of separation between clusters.

The silhouette coefficients of the samples were calculated at different K_p values. This coefficient measures how similar an object is to its cluster compared with other clusters. After the samples were clustered into K_p clusters, the average distance of $x_i^{(p)}$ from the other samples in the cluster was calculated for the sample $x_i^{(p)}$ in cluster $M_I^{(p)}$. Within-cluster dissimilarity, which measures how well $x_i^{(p)}$ is assigned to its cluster, is defined as follows:

$$a(x_i^{(p)}) = \frac{1}{|M_I^{(p)}| - 1} \sum_{j \in M_I^{(p)}, j \neq i} \delta(x_i^{(p)}, x_j^{(p)}), \quad (10)$$

where $|M_I^{(p)}|$ is the number of samples belonging to cluster I . The definition of the distance function is the same as that of the distance function in Equation (4).

The between-cluster dissimilarity, the average distance between $\mathbf{x}_i^{(p)}$ and all samples in the other cluster $M_j^{(p)}$, are defined as follows:

$$b(\mathbf{x}_i^{(p)}) = \min_{J \neq I} \frac{1}{|M_J^{(p)}|} \sum_{j \in M_J^{(p)}} \delta(\mathbf{x}_i^{(p)}, \mathbf{x}_j^{(p)}). \quad (11)$$

The silhouette coefficient of sample $\mathbf{x}_i^{(p)}$ is defined as follows:

$$s(\mathbf{x}_i^{(p)}) = \begin{cases} 1 - \frac{a(\mathbf{x}_i^{(p)})}{b(\mathbf{x}_i^{(p)})}, & \text{if } a(\mathbf{x}_i^{(p)}) < b(\mathbf{x}_i^{(p)}) \\ 0, & \text{if } a(\mathbf{x}_i^{(p)}) = b(\mathbf{x}_i^{(p)}) \\ \frac{b(\mathbf{x}_i^{(p)})}{a(\mathbf{x}_i^{(p)})} - 1, & \text{if } a(\mathbf{x}_i^{(p)}) > b(\mathbf{x}_i^{(p)}) \end{cases} \quad (12)$$

The silhouette coefficient $s(\mathbf{x}_i^{(p)})$ is a measure of how well the data are clustered over the entire dataset. A larger value indicates a more reasonable clustering of $\mathbf{x}_i^{(p)}$.

Based on the silhouette coefficients, we obtained the optimal number of clusters K_p :

$$\hat{K}_p = \operatorname{argmax}_{K_p} \bar{s}(K_p), \quad (13)$$

where $\bar{s}(K_p)$ is the mean silhouette coefficient (MSC), defined as

$$\bar{s}(K_p) := \frac{1}{N} \sum_{i=1}^N s(\mathbf{x}_i^{(p)}). \quad (14)$$

The above equation represents the mean $s(\mathbf{x}_i^{(p)})$ over all cell trajectories for a specific number of clusters K_p . Cell–cell interactions can change during migration, and cell division also perturbs these interactions. Thus, the number of clusters K_p may vary over time. Therefore, this process is repeated for each period p and is used to determine the appropriate number of clusters \hat{K}_p .

2.7. Robustness of Our Method

Our method depends on the accuracy of the cell tracking data, particularly the accuracy of the estimated positions of the cell nucleus center. In this study, the positions of the cell nucleus centers were detected manually from bright field images, which might have been affected by the observation noise. Thus, we evaluated the robustness of our clustering method to the intensity of observation noise in the estimated cell nucleus positions.

As shown in Equation (1) $V_i^{(p)} = \langle \mathbf{v}_{i,1}^{(p)}, \dots, \mathbf{v}_{i,n}^{(p)} \rangle$ denotes the cell trajectories, where $\mathbf{v}_{i,l}^{(p)} \in \mathbb{R}^2$ is the cell position at the l th time slice of the p th time period and n is the number of time slices in a single time period.

We assume that the observation noise of the cell nucleus center position is modeled as two-dimensional Gaussian noise as follows:

$$\{\xi_l\} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}_2(\mathbf{0}, \Sigma). \quad (15)$$

Here, $\xi_l \in \mathbb{R}^2$ and $\mathcal{N}_2(0, \Sigma)$ denote a two-dimensional Gaussian distribution, where the mean $0, \Sigma$ denotes the variance–covariance matrix, which is defined as

$$\Sigma = \begin{bmatrix} \sigma & 0 \\ 0 & \sigma \end{bmatrix}. \quad (16)$$

The true nucleus center $v_{i,l}^{(p)}$ is unknown, but we assume that $v_{i,l}^{(p)}$ is the true center. We then generated the test dataset by adding Gaussian noise to the original single-cell trajectory dataset, as follows:

$$V_i^{(p)} = \langle v_{i,1}^{(p)} + \xi_1, \dots, v_{i,n}^{(p)} + \xi_n \rangle. \tag{17}$$

Next, we evaluated the consistency of the clustering results obtained from the datasets, original dataset, and generated test datasets. Let $n(M_k^{(p)})$ be the number of elements in the k th cluster at the p th time period, and $\mathbf{n}^{(p)}$ denote a vector whose element is $n(M_k^{(p)})$ for $1 \leq k \leq K_p$, which is defined as

$$\mathbf{n}^{(p)} := \left(n(M_1^{(p)}), n(M_2^{(p)}), \dots, n(M_{K_p}^{(p)}) \right)^\top, \tag{18}$$

where $n(M_k^{(p)})$ for $1 \leq k \leq K_p$ is sorted in descending order; thus, it satisfies $n(M_{k'}^{(p)}) \geq n(M_{k'+1}^{(p)})$.

Let $\mathbf{n}_u^{(p)}$ be $\mathbf{n}^{(p)}$ obtained from data $\mathcal{D}_u^{(p)}$, where $\mathcal{D}_u^{(p)}$ denotes a set of cell trajectories in the p th time period represented in two-dimensional space (see Equation (6)), $u \geq 1$. $u = 0$ denotes the index of the original dataset, and $u \geq 1$ denotes that of the test dataset generated by adding Gaussian noise to the original dataset. The consistency of the clustering results obtained from the original and test datasets is defined by the Euclidean distance between the clustering results obtained from the two datasets. Before calculating the consistency, we normalized $\mathbf{n}_u^{(p)}$ by the total number of cells in the p th time slice $N^{(p)}$. The normalized clustering result is defined as follows:

$$\mathbf{n}_u'^{(p)} := \left(\frac{n_u(M_1^{(p)})}{N^{(p)}}, \dots, \frac{n_u(M_{K_p}^{(p)})}{N^{(p)}} \right)^\top. \tag{19}$$

Then, the consistency between the original dataset and the test dataset is defined by the Euclidean distance between $\mathbf{n}_0'^{(p)}$ and $\mathbf{n}_u'^{(p)}$ ($u \geq 1$), which is

$$\delta(\mathbf{n}_0'^{(p)}, \mathbf{n}_{u \geq 1}'^{(p)}) = \sqrt{\sum_{k=1}^{K_p} \left(\frac{n_0(M_k^{(p)})}{N^{(p)}} - \frac{n_{u \geq 1}(M_k^{(p)})}{N^{(p)}} \right)^2}. \tag{20}$$

We generated several test datasets ($u \geq 1$) using Equation (17) and obtained a set of test dataset clustering results.

$$\mathcal{R}^{(p)} := \{ \mathbf{n}_1^{(p)}, \mathbf{n}_2^{(p)}, \dots, \mathbf{n}_h^{(p)} \}, \tag{21}$$

where h denotes the number of the test datasets. Finally, we defined the difference score as the mean value of $\delta(\mathbf{n}_0'^{(p)}, \mathbf{n}_{u \geq 1}'^{(p)})$ for $u \geq 1$ as follows. A lower difference score indicates a better robustness.

$$\bar{\delta}_{\mathcal{R}^{(p)}} := \frac{1}{h} \sum_{u=1}^h \delta(\mathbf{n}_0'^{(p)}, \mathbf{n}_u'^{(p)}). \tag{22}$$

3. Results and Discussion

3.1. Cell Tracking

The interaction between mesenchymal and epithelial cells is an important driving force for collective cell migration. Interactions promote the transmission of signals between cells, which guides the migration of follower cells. Thus, multiple cells follow the leader’s

migration. Here, the influence of leader cells was investigated by observing the cell population in a model in which TGF (+) cells (red labeled) and TGF (−) cells coexist (Figure 1A, Movie S1). Figure 1A,B indicate that the shape of the colony changed gradually from a circle to an ellipse, suggesting that the cell migration was not random but anisotropic due to the aligned electrospun fibers. The migration of each cell was one-directional toward the outside of the colony, which might be due to the local gradient of the soluble factors secreted by cells because the cell culture medium had no intentional mechanism to cause a chemotactic gradient; it is beyond the scope of this study but warrants further investigation.

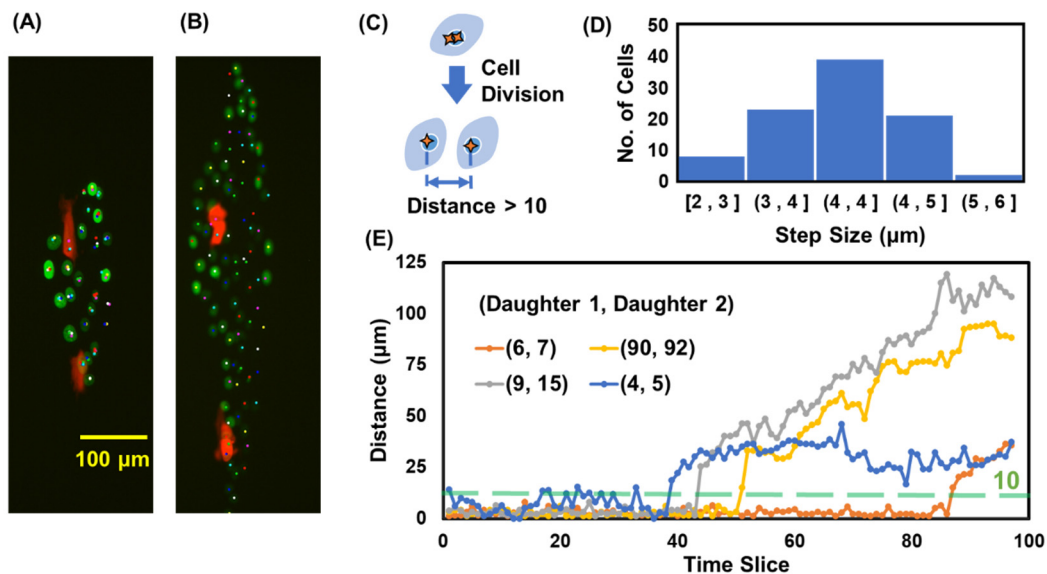


Figure 1. Daughter cell pairs were identified by calculating the “distance jump” between the markers. (A,B) Manual markers of cell position at time slice 1 and 96. (C) Schematic of calculation of distance between the marker of mother and daughter cell. (D) The distribution of step size (distance between two adjacent slices) of each cell before division. (E) Euclidean distance between two daughter cells.

At this time, TGF (+) cells were labeled with the red fluorescent protein DsRed to distinguish them from TGF (−). A population of 10–20 cells was observed because its size was the same as that of a colony in vivo. In the first time slice, there were 28 cells (Figure 1A). In the last time slice, we identified 92 cells (Figure 1B). Since the manual tracking started from the last time slice to the first time slice backward, we found 92 marker points of the cell nucleus in each time slice and assigned the ID number. In the cell division event, the cell before the division was defined as the mother and the two cells after division as daughter cells.

Cell division occurs during cell migration. For analysis, it is crucial to identify the correspondence between mother and daughter cells and the time at which cell division occurs. For all the marker points in each time slice, we calculated the Euclidean distance between each pair of marker points. A sudden change in the distance indicates that cell division had taken place (Figure 1C). The time-lapse observation recorded the image slices every 15 min. The distance between the two adjacent slices is defined as the step size. For a representative cell, the step size distribution for all pairs of slices is shown in Figure 1D. The maximum step size was 6 μm. The distances between the four groups of daughter cells are shown in Figure 1E. The graph shows a clear “distance jump” in the curve, where the distance before the “jump” is always very short. The “distance jump” indicates the occurrence of cell division. After the “jump” a completely different distance curve appears. The Euclidean distance between the marker points was greater than 10 μm after cell division. The maximum moving step size of 6 μm is below the “jump distance” of 10 μm, which means that the accuracy is acceptable for identifying when cell division occurs. In combination with fluorescence images, the division relationship between cells

can be determined. We identified divisions based on the calculation of the distance between the cell marker points and manually labeled cells. The marker points of the mother and two daughter cells were unified prior to cell division.

3.2. Dimension Reduction

After the cell tracking, the entire migration process was divided into multiple periods. The trajectories of TGF (+) (black) and TGF (−) (blue) cells during different periods are shown in Figure 2(A1–A5). Two groups of TGF (+) cells were observed from the initial position and marked separately. The cell trajectories within each period were normalized. The purpose of normalization is to eliminate the effect of location (Figure 2(B1–B5)), where “normalization” refers to translating all cell trajectories such that their starting points are located at the origin (0, 0).

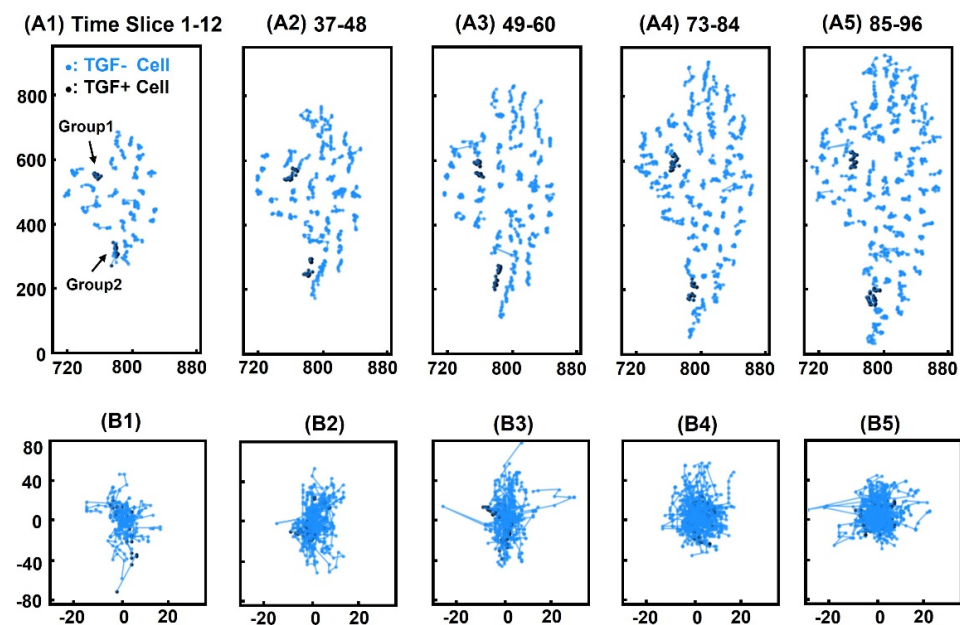


Figure 2. Before and after cell tracks’ positional normalization. (A1–A5) The positions of TGF (+) (black) and TGF (−) (blue) cells before normalization in different observation windows. (B1–B5) The positions after normalization, the initial position transferred to (0, 0).

The normalized cell trajectories are used to calculate the distance matrix, which is a series of high-dimensional datasets for which we need to reduce the dimensionality for visualization. There are two reasons for reducing the dimension of the data before performing k-means clustering, in addition to visualization. The first reason is data summarization. Our method aims to find a low-dimensional structure from high-dimensional cell-trajectory data. Reducing the dimensions before performing k-means clustering is beneficial for capturing the low dimensional structures. The second reason is robustness. Dimension reduction can help reduce the noise included in high-dimensional data and improve the robustness of our method. Dimension reduction before performing downstream analysis, such as clustering, is also used in single-cell RNA-sequencing data analysis [34,35]. We chose two as the number of the dimensions to ensure consistency with the visualization. The advantage of UMAP over t-SNE is its ability to reflect both local and global structures while maintaining relative global distances in a low-dimensional space. These capabilities allow us to obtain appropriate clustering results. Figure 3 shows the dimensionality reduction results for several periods with TGF (−) cells (green) and TGF (+) cells (red and blue). We also calculated using the t-SNE for comparison (Figure S1). As a result, UMAP showed a more apparent aggregation of data points in the same cluster. In addition, there were more apparent boundaries between the clusters.

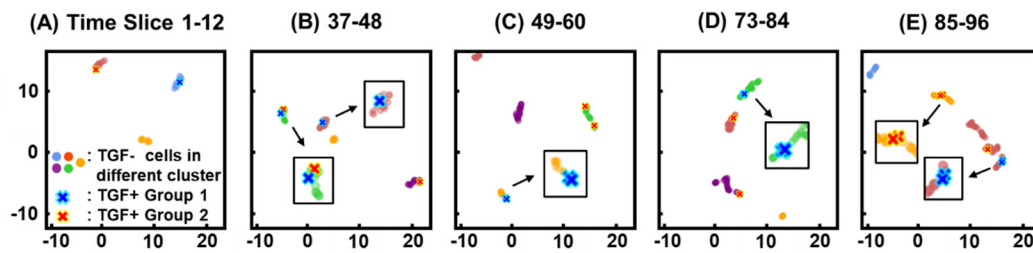


Figure 3. Visualization of cell tracks after dimensionality reduction. (A–E) 2D visualization of TGF (+) (“X” mark) and TGF (–) cell (colored points) trajectories after dimensionality reduction by UMAP and clustering by k-means. Each mark and point represent a trajectory. Colors represent different clusters. The insets show the magnified images of TGF (+) cells when they overlap (B–E).

The optimal observation window was selected such that the distance between the clusters was larger in the first and last time periods and the data points within the clusters were closer together (Figure S2). We selected an observation window of 12 time slices. As shown in Figure S3, the parameters for UMAP were chosen, such that the distance between clusters was larger and the data points within the clusters were closer.

3.3. Clustering

The clustering method used the k-means algorithm, and for each time period, we used the MSC method to determine the appropriate number of clusters (Figure S4). MSC $\bar{s}(k)$ is in the range of $(-1,1)$; therefore, the closer $\bar{s}(k)$ is to 1, the more reasonable the clustering is, at which point k is the appropriate number of clusters. The results of the clustering are shown in Figure 3A–E, where the different clusters are labeled with different colors. There were two TGF (+) cells in the first generation that were distant from each other. Multiple daughter cells were observed after cell division. Until the last time slice, five TGF (+) cells were present (refer to Figure S5). According to each division, TGF (+) was referred to as two groups, where Group 1 included TGF (+) Cell ID 4 and 5 and Group 2 included TGF (+) Cell ID 1 to 3.

3.4. Similarity of Migration Patterns

In each period in Figure 3A–E, it is clear that TGF (+) cells of the same group were not always in the same cluster at the same time. TGF (+) and TGF (–) cells assigned to the same cluster showed high similarity and maintained the same migration pattern in terms of collective cell migration. TGF (+) cells with more active migratory behavior exert some influence on TGF (–) cells. TGF (+) cells with a leading role showed interactions with TGF (–) cells through signaling. Pathways such as PI3K-Rac signaling are involved in actin remodeling and mediation of collective migration. Through PI3K-dependent integrin adhesion and modulation of Rho-GTPase signaling, cadherin-induced regulation of actomyosin contractility at more distant sites in the cell affects global cellular mechanics [2,36]. The present analysis found that cells around TGF(+) cells showed the same migration pattern, and they might have been under the influence of the leader cell. The behaviors observed here may correspond to reports that the leader cell leads the movement of the cell population such that the surrounding follower cells are in tow. Therefore, our methodology could be used to detect leader cells within a cell population. This maintained them in the same overall migration direction and distance.

In addition to clusters containing TGF (+) cells, some clusters contained only TGF (–) cells, illustrating that their migration pattern is different from that of TGF (+) cells. TGF (+) cells, as leaders of collective migration, affect nearby TGF (–) cells; however, the effect is relatively narrow. TGF (–) cells with fewer or no interactions migrated based on their migratory patterns.

3.5. Positional Similarity

It is expected that the influence of leader cells on follower cells decreases with distance and is stronger only within a certain range. Within the effective range, the interaction strengthened the connections between them. To verify these results, we superimposed the clustering results onto a cell location map. During the previous normalization process, the location information of the cell trajectories was reset and the similarity of the cell trajectories was determined from the clustering results. However, it is impossible to determine whether cell trajectories with similarities are close to each other in terms of their actual locations. Therefore, the clustering results and actual locations were combined, as shown in Figure 4. Notably, the clustering results in Figure 4 are the same as those in Figure 3; however, the color was adjusted for ease of comparison. The data points for TGF (+) cells in Figure 4 are marked in black, whereas the lines between the points remain colored.

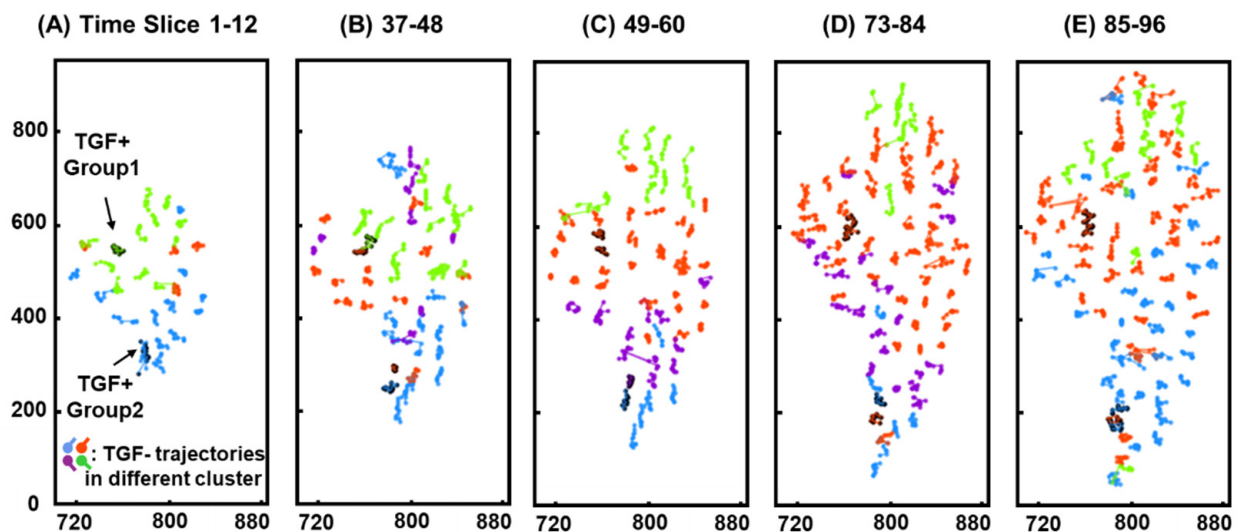


Figure 4. Cell trajectories' positions combined with clustering results. (A–E) Two groups of TGF (+) cells are marked black pointed by arrows. Colors represent different clusters.

Interestingly, after combining the clustering results with actual locations, most cells from the same cluster showed positional similarities. Our method reproduces positional similarity, even after normalization eliminates the location information. This worked for most cells, except for individual cells, such as the green cluster in Figure 4B, which contained cells belonging to the other clusters.

We used an aligned fiber to mimic the ECM in an in vitro model. This environment closely resembles the environment inside an organism. The migration direction of some cells may be influenced by fiber alignment; however, after normalization, the effect of position was eliminated. Normalization makes the method reliable and the trajectory independent of the ECM. In addition, the calculation of this method treats all cells equally and does not distinguish between leader and follower cells, but the results show cells with similar migration patterns. After superimposing the clustering with the actual locations, we found that in most cases, there were many TGF (–) cells around TGF (+) cells. From these results, we can see that clusters are not related to cells moving in a limited direction; they are related to the similarity of migration. TGF (+) cells were mostly not in the center of each cluster but at the edge of the cluster, which is consistent with the characteristics of the leader cells.

3.6. Cell Division

Cell division can occur during migration and may affect the surrounding micro-environment, resulting in altered migration characteristics. The direction and speed of the migration of individual cells in a colony can be altered by cell division. These changes may affect the migration pattern of the colony. Therefore, it is necessary to establish a cell lineage for analysis of cell migration patterns. The cell lineage was drawn to identify the specific time slices where cell division occurred and the correspondence between daughter and mother cells (Figure S5). Cell IDs 1–5 are TGF (+) cells, and the others are TGF (–) cells. The time of cell lineage (Figure S5) was divided into eight periods. The color in each time period (Figure S5) indicates the cluster to which the cell belongs, and the color remains the same as that in Figure 4.

In combination with Figure S5, TGF (+) cells ID 4 of Group 1 were divided into Cell ID 5 at time slice 38, and these cells were always surrounded by TGF (–) cells, although these cells were sometimes not fixed. Cell ID 3 of Group 2 was the first generation of cells, and the following two divisions occurred in time slices 32 and 88. Interestingly, the clusters of TGF (+) Group 2 and the surrounding TGF (–) cells in Figure 4B,D,E were different; the colors do not exactly match those of the surrounding TGF (–) cells. Our explanation for this phenomenon was that the state of the cells changed during cell division, which affected the migration pattern and trajectory of the cells. This led to differences in clustering. The state of cell division lasts for some time; therefore, the effect may occur over multiple periods. In contrast, Group 1 had only one cell division at time slice 38. Therefore, the clustering of Group 1 and the surrounding TGF (–) cells remained consistent for all periods, except for Figure 4B. The clusters of both daughter and parent cells essentially changed (e.g., Figure S5, Cell ID 34, 56, 59). Thus, cell division resets cell-cell interactions and creates a new surrounding environment, which is the key to changing the migration similarity.

3.7. Robustness

Whether this method is effective in identifying cells with similar migration patterns requires further validation. During manual marking, the accuracy of the marked cell nucleus position determines the degree of error. The robustness of the clustering results was verified by adding noise to the cell trajectories. The test data were generated by adding Gaussian noise with a fixed mean and standard deviation, σ (Figure 5A). Figure 5(B1) shows the original positions of the cell trajectories for time slices 49–60, and Figure 5(B2,B3) show the clustering results with $\sigma = 1$ and 2, respectively. A change in σ causes the similar trajectories to become dissimilar. The boundaries between clusters became indistinct and k increased, rendering the clustering meaningless. The maximum value of k for the original clustering results was 5. Therefore, we applied the maximum value of $k = 5$ to noise-containing clustering analysis. The trajectory difference score of the clustering results with noise compared to those without noise is shown in Figure 5D. Notably, the cluster color order (cluster index) changed after each adjustment of the σ value. Therefore, we recorded the ID of each trajectory in each cluster and then adjusted the color order as a whole to ensure reasonableness (Figure 5(B1–B3)).

As shown in Figure 5(B1–B3), the distribution of cluster positions did not differ significantly and only the clusters of individual cells changed. The observation error is mainly due to the accuracy of manual labeling. The radius of the nucleus was below 9 μm (Figure 5C). Considering the accuracy the manual labeling of the center of the cell nucleus, $\sigma = 2$ was sufficiently large for our test. The clustering results with noise and without noise were comparable in consistency. This difference may be related to noise altering the original migration trajectory. This is because the change in trajectory affects the migration pattern, including speed and direction. The internal connections between cells that originally belonged to the same cluster are broken. When σ is greater, noise causes severe distortion of the trajectory. Nevertheless, there were no significant difference in the position of each cluster. Therefore, the method is robust to observation errors that may be contained in the cell trajectory data.

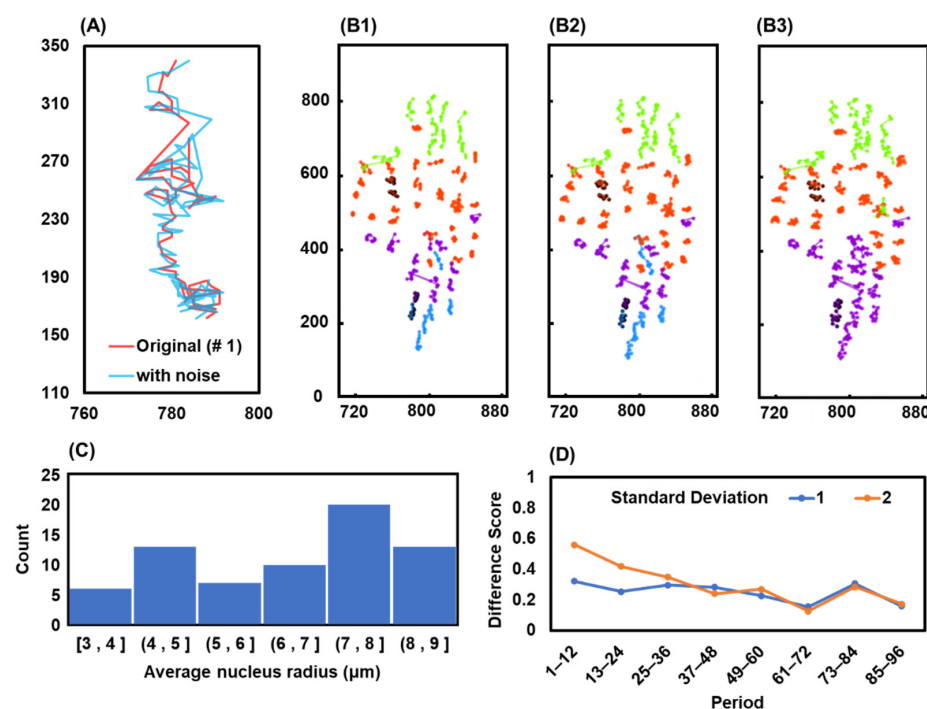


Figure 5. Robustness of the method. (A) The original trajectory of Cell ID 1 (red) and the trajectory with Gaussian noise of standard deviation $\sigma = 2$ (blue). (B1) Cell tracks' positions combined with clusters' result under time slice 49–60. (B2,B3) All trajectories are with noise of standard deviations = 1, 2. (C) Histogram of average radius of cell nucleus (μm). (D) The difference score of all clustering results under different standard deviations (from the mean of three calculations in each σ). The sample sizes for each time period are 92 cells.

4. Conclusions

Interactions between cells are essential for coordinated and directed collective movements. Although, collective cell migration has been extensively studied, this study focused on exploring the similarity of migration patterns based on migration trajectories. Using a combination of dimensionality reduction and clustering techniques, cells with similar migration patterns were observed to exhibit positional similarities. Cell division is an important factor that affects the migration patterns. TGF (+) and TGF (−) cells belonging to the same cluster showed similar migration patterns and, within a certain range, migration was consistent. Validation by adding noise illustrated the robustness of the proposed method. This provides a new perspective for a deeper understanding of collective cell migration. The method applies to collective cell migration on aligned fibers and models such as wound healing and migration on other substrates. It can also be extended to the cell migration model in a 3D hydrogel matrix if cells can be tracked, indicating its application in tissue engineering and organ development research.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/cancers14194587/s1>, Figure S1. The dimensionality reduction of cell trajectories by t-SNE under different parameters and observation windows. Figure S2. Cell tracks after dimensionality reduction when select different observation windows. The first and last periods showed that window 12 has clear boundaries of clusters and TGF (+) cells in the same group tend to the same cluster. Figure S3. The dimensionality reduction of cell trajectories under different UMAP parameters (observation window = 12, time slice 85–96). Figure S4. WCSS (within-cluster sum of squares of distance) (blue) and mean silhouette coefficient (MSC) (orange) of all samples with the number of k-means clusters of UMAP dimensionality reduction results under different observation periods. The K_p value shown by the red arrow is the optimal number of clusters. The initial number of TGF (+) and TGF (−) cells at each period is included in each figure. There were 5 TGF (+) cells

and 87 TGF (−) at the last slice (slice 96). Figure S5. Cell lineage tree. Daughter cells are divided from mother cells at different time slice (red font). Colors represent the same clusters as in Figure 5. Cell IDs 1–5 are TGF (+) cells (Group 1: ID 4–5, Group 2: ID 1–2), Cell ID 6–93 are TGF (−) cells. Movie S1. NMuMG cells migration on the PS fibers. TGF (+) cells (red labeled) and TGF (−) cells (green labeled).

Author Contributions: Conceptualization, Z.X., M.K.K. and S.F.; methodology, K.D., M.K.K. and S.F.; calculations, Z.X.; investigation, K.D. and S.F.; data curation, K.D.; writing—original draft preparation, Z.X., M.K.K. and S.F.; writing—review and editing, M.K.K., S.-i.S. and S.F.; visualization, Z.X.; validation, M.K.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The code to generate the data is available on the GitHub repository: <https://github.com/CollectiveCellMigration/CellTrajectoryClustering> (accessed on 15 September 2022).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jain, S.; Cachoux, V.M.L.; Narayana, G.H.N.S.; de Beco, S.; D'Alessandro, J.; Cellerin, V.; Chen, T.; Heuzé, M.L.; Marcq, P.; Mège, R.M.; et al. The Role of Single-Cell Mechanical Behaviour and Polarity in Driving Collective Cell Migration. *Nat. Phys.* **2020**, *16*, 802–809. [[CrossRef](#)]
2. Khalil, A.A.; de Rooij, J. Cadherin Mechanotransduction in Leader-Follower Cell Specification during Collective Migration. *Exp. Cell Res.* **2019**, *376*, 86–91. [[CrossRef](#)]
3. Collins, T.A.; Yeoman, B.M.; Katira, P. To Lead or to Herd: Optimal Strategies for 3D Collective Migration of Cell Clusters. *Biomech. Model. Mechanobiol.* **2020**, *19*, 1551–1564. [[CrossRef](#)]
4. Saénz-de-Santa-María, I.; Celada, L.; Chiara, M.-D. The Leader Position of Mesenchymal Cells Expressing N-Cadherin in the Collective Migration of Epithelial Cancer. *Cells* **2020**, *9*, 731. [[CrossRef](#)]
5. Gregory, P.A.; Bert, A.G.; Paterson, E.L.; Barry, S.C.; Tsykin, A.; Farshid, G.; Vadas, M.A.; Khew-Goodall, Y.; Goodall, G.J. The MiR-200 Family and MiR-205 Regulate Epithelial to Mesenchymal Transition by Targeting ZEB1 and SIP1. *Nat. Cell Biol.* **2008**, *10*, 593–601. [[CrossRef](#)]
6. Burk, U.; Schubert, J.; Wellner, U.; Schmalhofer, O.; Vincan, E.; Spaderna, S.; Brabletz, T. A Reciprocal Repression between ZEB1 and Members of the MiR-200 Family Promotes EMT and Invasion in Cancer Cells. *EMBO Rep.* **2008**, *9*, 582–589. [[CrossRef](#)]
7. Dongre, A.; Weinberg, R.A. New Insights into the Mechanisms of Epithelial–Mesenchymal Transition and Implications for Cancer. *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 69–84. [[CrossRef](#)]
8. Yang, H.; Ganguly, A.; Cabral, F. Inhibition of Cell Migration and Cell Division Correlates with Distinct Effects of Microtubule Inhibiting Drugs. *J. Biol. Chem.* **2010**, *285*, 32242–32250. [[CrossRef](#)]
9. Ganguly, A.; Yang, H.; Sharma, R.; Patel, K.D.; Cabral, F. The Role of Microtubules and Their Dynamics in Cell Migration. *J. Biol. Chem.* **2012**, *287*, 43359–43369. [[CrossRef](#)]
10. Doyle, A.D.; Wang, F.W.; Matsumoto, K.; Yamada, K.M. One-Dimensional Topography Underlies Three-Dimensional Fibrillar Cell Migration. *J. Cell Biol.* **2009**, *184*, 481–490. [[CrossRef](#)] [[PubMed](#)]
11. Caswell, P.T.; Zech, T. Actin-Based Cell Protrusion in a 3D Matrix. *Trends Cell Biol.* **2018**, *28*, 823–834. [[CrossRef](#)]
12. Borgne-Rochet, M.L.; Angevin, L.; Bazellieres, E.; Ordas, L.; Comunale, F.; Denisov, E.V.; Tashireva, L.A.; Perelmuter, V.M.; Bièche, I.; Vacher, S.; et al. P-Cadherin-Induced Decorin Secretion Is Required for Collagen Fiber Alignment and Directional Collective Cell Migration. *J. Cell Sci.* **2019**, *132*, jcs233189. [[CrossRef](#)]
13. Grünert, S.; Jechlinger, M.; Beug, H. Diverse Cellular and Molecular Mechanisms Contribute to Epithelial Plasticity and Metastasis. *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 657–665. [[CrossRef](#)]
14. Thompson, E.W.; Williams, E.D. EMT and MET in Carcinoma—Clinical Observations, Regulatory Pathways and New Models. *Clin. Exp. Metastasis* **2008**, *25*, 591–592. [[CrossRef](#)] [[PubMed](#)]
15. Kim, J.M.; Lee, M.; Kim, N.; Heo, W. Do Optogenetic Toolkit Reveals the Role of Ca²⁺ Sparklets in Coordinated Cell Migration. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 5952–5957. [[CrossRef](#)]
16. Becsky, D.; Szabo, K.; Gyulai-Nagy, S.; Gajdos, T.; Bartos, Z.; Balind, A.; Dux, L.; Horvath, P.; Erdelyi, M.; Homolya, L.; et al. Syndecan-4 Modulates Cell Polarity and Migration by Influencing Centrosome Positioning and Intracellular Calcium Distribution. *Front. Cell Dev. Biol.* **2020**, *8*, 575227. [[CrossRef](#)]
17. Morrison, J.A.; McLennan, R.; Wolfe, L.A.; Gogol, M.M.; Meier, S.; McKinney, M.C.; Teddy, J.M.; Holmes, L.; Semerad, C.L.; Box, A.C.; et al. Single-Cell Transcriptome Analysis of Avian Neural Crest Migration Reveals Signatures of Invasion and Molecular Transitions. *eLife* **2017**, *6*, e28415. [[CrossRef](#)] [[PubMed](#)]

18. Capuana, L.; Boström, A.; Etienne-Manneville, S. Multicellular Scale Front-to-Rear Polarity in Collective Migration. *Curr. Opin. Cell Biol.* **2020**, *62*, 114–122. [[CrossRef](#)]
19. Rani, S.; Sikka, G. Recent Techniques of Clustering of Time Series Data: A Survey. *Int. J. Comput. Appl.* **2012**, *52*, 1–9. [[CrossRef](#)]
20. Niennattrakul, V.; Ratanamahatana, C.A. Inaccuracies of Shape Averaging Method Using Dynamic Time Warping for Time Series Data. In *Proceedings of the Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Shi, Y., van Albada, G.D., Dongarra, J., Sloot, P.M.A., Eds.; Springer: Berlin/Heidelberg, Germany, 2007; Volume 4487, pp. 513–520.
21. Hozumi, Y.; Wang, R.; Yin, C.; Wei, G.W. UMAP-Assisted K-Means Clustering of Large-Scale SARS-CoV-2 Mutation Datasets. *Comput. Biol. Med.* **2021**, *131*, 104264. [[CrossRef](#)]
22. Wold, S.; Esbensen, K.; Geladi, P. Principal Component Analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37–52. [[CrossRef](#)]
23. Mead, A. Review of the Development of Multidimensional Scaling Methods. *Statistician* **1992**, *41*, 27. [[CrossRef](#)]
24. Laurens, V.D.M.; Hinton, G. Visualizing Data Using T-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
25. McInnes, L.; Healy, J.; Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv* **2018**. arXiv:1802.03426
26. Pyatnitskiy, M.; Mazo, I.; Shkrob, M.; Schwartz, E.; Kotelnikova, E. Clustering Gene Expression Regulators: New Approach to Disease Subtyping. *PLoS ONE* **2014**, *9*, e84955. [[CrossRef](#)]
27. Fujita, A.; Severino, P.; Kojima, K.; Sato, J.R.; Patriota, A.G.; Miyano, S. Functional Clustering of Time Series Gene Expression Data by Granger Causality. *BMC Syst. Biol.* **2012**, *6*, 137. [[CrossRef](#)]
28. Schneider, D.; Tarantola, M.; Janshoff, A. Dynamics of TGF- β Induced Epithelial-to-Mesenchymal Transition Monitored by Electric Cell-Substrate Impedance Sensing. *Biochim. Biophys. Acta—Mol. Cell Res.* **2011**, *1813*, 2099–2107. [[CrossRef](#)]
29. Lonseko, Z.M.; Adjei, P.E.; Du, W.; Luo, C.; Hu, D.; Zhu, L.; Gan, T.; Rao, N. Gastrointestinal Disease Classification in Endoscopic Images Using Attention-Guided Convolutional Neural Networks. *Appl. Sci.* **2021**, *11*, 11136. [[CrossRef](#)]
30. Hussain, S.M.; Buongiorno, D.; Altini, N.; Berloco, F.; Prencipe, B.; Moschetta, M.; Bevilacqua, V.; Brunetti, A. Shape-Based Breast Lesion Classification Using Digital Tomosynthesis Images: The Role of Explainable Artificial Intelligence. *Appl. Sci.* **2022**, *12*, 6230. [[CrossRef](#)]
31. Althuwaynee, O.F.; Aydda, A.; Hwang, I.T.; Lee, Y.K.; Kim, S.W.; Park, H.J.; Lee, M.S.; Park, Y. Uncertainty Reduction of Unlabeled Features in Landslide Inventory Using Machine Learning T-SNE Clustering and Data Mining Apriori Association Rule Algorithms. *Appl. Sci.* **2021**, *11*, 556. [[CrossRef](#)]
32. MacQueen, J. Some Methods for Classification and Analysis of Multivariate Observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Los Angeles, CA, USA, 1 January 1967; Volume 1, pp. 281–297.
33. Rousseeuw, P.J. Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53–65. [[CrossRef](#)]
34. Luecken, M.D.; Theis, F.J. Current Best Practices in Single-cell RNA-seq Analysis: A Tutorial. *Mol. Syst. Biol.* **2019**, *15*, e8746. [[CrossRef](#)]
35. Moon, K.R.; Stanley, J.S.; Burkhardt, D.; van Dijk, D.; Wolf, G.; Krishnaswamy, S. Manifold Learning-Based Methods for Analyzing Single-Cell RNA-Sequencing Data. *Curr. Opin. Syst. Biol.* **2018**, *7*, 36–46. [[CrossRef](#)]
36. Kölsch, V.; Charest, P.G.; Firtel, R.A. The Regulation of Cell Motility and Chemotaxis by Phospholipid Signaling. *J. Cell Sci.* **2008**, *121*, 551–559. [[CrossRef](#)] [[PubMed](#)]