Multi-cell ECM compaction is predictable via superposition of nonlinear cell dynamics linearized in augmented state space

PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Multi-cell ECM compaction is predictable via superposition of nonlinear cell dynamics linearized in augmented state space

Michaëlle N. Mayalu[☯¤]*, **Min-Cheol Kim**[☯], **H. Harry Asada**[☯]*

Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America

☯ These authors contributed equally to this work.
¤ Current address: Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, California, United States of America
* mmayalu@caltech.edu (MNM); asada@mit.edu (HHA)

Citation: Mayalu MN, Kim M-C, Asada HH (2019) Multi-cell ECM compaction is predictable via superposition of nonlinear cell dynamics linearized in augmented state space. PLoS Comput Biol 15(9): e1006798. https://doi.org/10.1371/journal.pcbi.1006798

Editor: Philip K Maini, Oxford, UNITED KINGDOM

Received: January 17, 2019

Accepted: June 12, 2019

Published: September 20, 2019

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: This material is based on work supported by the National Science Foundation (NSF) under grant number CMMI-1762961, Singapore-MIT Alliance of Research and Technology (SMART), and NSF Science and Technology Center (STC) on Emergent Behaviors in Integrated Cellular Systems (EBICS) under Grant CBET-0939511. The funders had no role in study design, data collection and

Abstract

Cells interacting through an extracellular matrix (ECM) exhibit emergent behaviors resulting from collective intercellular interaction. In wound healing and tissue development, characteristic compaction of ECM gel is induced by multiple cells that generate tensions in the ECM fibers and coordinate their actions with other cells. Computational prediction of collective cell-ECM interaction based on first principles is highly complex especially as the number of cells increase. Here, we introduce a computationally-efficient method for predicting nonlinear behaviors of multiple cells interacting mechanically through a 3-D ECM fiber network. The key enabling technique is superposition of single cell computational models to predict multicellular behaviors. While cell-ECM interactions are highly nonlinear, they can be linearized accurately with a unique method, termed Dual-Faceted Linearization. This method recasts the original nonlinear dynamics in an augmented space where the system behaves more linearly. The independent state variables are augmented by combining auxiliary variables that inform nonlinear elements involved in the system. This computational method involves a) expressing the original nonlinear state equations with two sets of linear dynamic equations b) reducing the order of the augmented linear system via principal component analysis and c) superposing individual single cell-ECM dynamics to predict collective behaviors of multiple cells. The method is computationally efficient compared to original nonlinear dynamic simulation and accurate compared to traditional Taylor expansion linearization. Furthermore, we reproduce reported experimental results of multi-cell induced ECM compaction.

Author summary

Collective behaviors of multiple cells interacting through an ECM are prohibitively complex to predict with a mechanistic computational model due to its highly nonlinear

analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

dynamics and high dimensional space. We introduce a methodology where nonlinear dynamics of single cells are superposed to predict collective multi-cellular behaviors through a developed linearization method. We represent nonlinear single cell dynamics with linear state equations by augmenting the independent state variables with a set of auxiliary variables. We then transform the linear augmented state equations to a low-dimensional latent model and superpose the linear latent models of individual cells to predict collective behaviors that emerge from multi-cellular interactions. The method successfully reproduced experimental results of cell-induced ECM compaction.

Introduction

Cell-induced compaction of fibrous extracellular matrix (ECM) is an important mechanism for numerous processes such as wound healing and tissue development [1–3]. During wound healing, for example, traction forces exerted by fibroblasts and myofibroblasts result in ECM compaction at the site of injury [2, 3]. In vitro experiments using cell-populated collagen gel reveal global compaction of the matrix as a result of cooperative effect of multiple cells at the boundaries as well as propagation through the bulk [4–6]. Furthermore, matrix densification is observed in the regions around [7] and in-between cells. Here we examine the mechanical aspect of intercellular communication through the ECM and how contractile cells can induce emergent mechanical changes leading to matrix compaction. From a simplified mechanics point of view, compaction results when the traction forces exerted by the contractile cells embedded within the ECM overcome the resistive forces of the ECM structure, including viscoelastic forces and elastic energy forces. As a result the matrix is deformed from its original stress-free state and the elastic modulus increases [4–7].

In reality, the compaction process is far more complex. The ECM forms a network of cross-linked fibers that is highly nonlinear and intricate, but is critical for predicting large compaction and long-range transmission of forces [4]. As a large deformation is induced by contractile cells, the standard linear mechanics model yields substantial errors. The ECM fiber network is anisotropic and causes irreversible deformations as a large compaction takes place. This prominent nonlinearity prohibits use of simple methods for predicting the ECM compaction by a multitude of cells. In addition, cells can internally modulate their state in response to local mechanical stresses within the ECM, which influences cell polarity, contractility, stiffness and strength of focal adhesion's [8–11]. These cell properties are highly nonlinear and complex. Consideration of these nonlinear physical and physiological properties involved in the cell-ECM mechanics often result in differential equations that are intractably complex due to high-dimensional, nonlinear coupled dynamics.

Many in silico modeling approaches in the areas of wound healing and fibrotic disease have helped elucidate and explore the underlying phenomena involved in cell-induced ECM compaction, and have been used to supplement in vitro experiments for fast and inexpensive methods of evaluation. Approaches in previous works include: i) a hybrid continuum-discrete framework consisting of the macroscopic finite element domain and local microscopic fiber network [12], ii) rule-based models with deformable cells and ECM fibers to explain matrix remodeling and durotaxis [13, 14], iii) a discrete fiber model of cell populated fibrous matrix [15], and iv) continuum models of ECM gel compaction [7, 16, 17]. Even though these works provide many insights, they also simplify the ECM gel compaction mechanism by: a) 2-D

representation of a 3-D system, b) exclusion of intracellular mechanics such as mechanobiology of actin stress fibers, focal adhesions, and remodeling of cellular and nuclear membranes, and c) consideration of linear elastic spring model of ECM fibers without including the viscoelastic nature of the fibers. Consequently, these prior models abstract detailed cell-ECM interactions, resulting in limitations to understanding how these interactions enable characteristic gel compaction. In addition, models examining the complex dynamics surrounding cell morphology, contractility, and polarity based on finite element methods do exist for 2 dimensional cases [10]. And, focal adhesion-stress fiber dynamics have been modeled for 2-D PDMS substrates based on non-equilibrium thermodynamics [11].

In the current work, the ECM is modeled as a 3-D cross-linked network of discrete, viscoelastic fibers, and detailed mechanistic cell dynamics, including focal adhesion dynamics, cytoskeleton remodeling, actin motor activity and lamellipodia protrusion, are derived from basic principles. The resultant model is computationally complex, especially for a larger number of cells. The governing differential equations are highly nonlinear, coupled, and of high dimension. Here, we solved this difficulty by introducing a methodology having its disciplinary basis spanned in system dynamics, machine learning, and statistics.

It is known that a nonlinear system can behave more linearly when recast in a larger space [18]. In our approach, the original nonlinear dynamics derived from physical and physiological principles are recast in an enlarged state space by augmenting independent state variables with auxiliary variables that inform input-output characteristics of the nonlinear elements involved in the system. Once recast in the augmented space, the nonlinear system can be represented as an augmented set of linear dynamic equations. The linear representation facilitates model reduction using latent variable analysis, which can be shown is difficult to apply to highly nonlinear systems [19–22]. Furthermore, linearization in the augmented space allows for superposition of multiple subsystems. In the current work, collective behaviors of multiple cells are predicted via superposition of single cell subsystems through the linearization in the augmented state space. The proposed methodology is general, and is applicable to a broader class of problems where large-scale, collective behaviors must be predicted while retaining sufficient mechanistic details.

Results

Governing equations of collective cell behaviors in ECM fiber network

We construct a computational model for predicting cell-mediated gel compaction by multiple (n_{cell}) cells having a uniform phenotype and interacting through a surrounding 3-D ECM fiber network. The ECM is modeled as a network of many fibers connected at a large number of nodes ($N_e \approx 2000$), whereas each cell is represented with a mesh structure consisting of multiple nodes ($N_c \approx 200$) which forms the cell outer membrane (see Fig 1A). The cell outer membrane deforms and gains traction as the nodes on the membrane bond to the nodes of the surrounding ECM fiber network and form focal adhesions, which occur when bonding molecules (or integrins) on the cell membrane bind to ligands on ECM.

Consider the i-th outer membrane node of the k-th cell with three dimensional spatial coordinates $\boldsymbol{x}_i^{c,k} \in \Re^{3\times 1}$ (See Fig 1B). The forces acting on it include the cell's cortical tension force and elastic energy force (collectively denoted as $\mathbf{F}_{Cort-Elas,i}^{c,k} \in \Re^{3\times 1}$), focal adhesion force (denoted as $\mathbf{F}_{FA,i}^{c,k} \in \Re^{3\times 1}$), lamellipodium force ($\mathbf{F}_{L,i}^{c,k} \in \Re^{3\times 1}$), and frictional damping force ($\mathbf{F}_{Damp,i}^{c,k} \in \Re^{3\times 1}$) [23, 24]. Assuming that the mass of the node is negligibly small and the damping force is given by $\mathbf{F}_{Damp,i}^{c,k} = -D_c d\boldsymbol{x}_i^{c,k}/dt$, where D_c is damping constant, the equation of

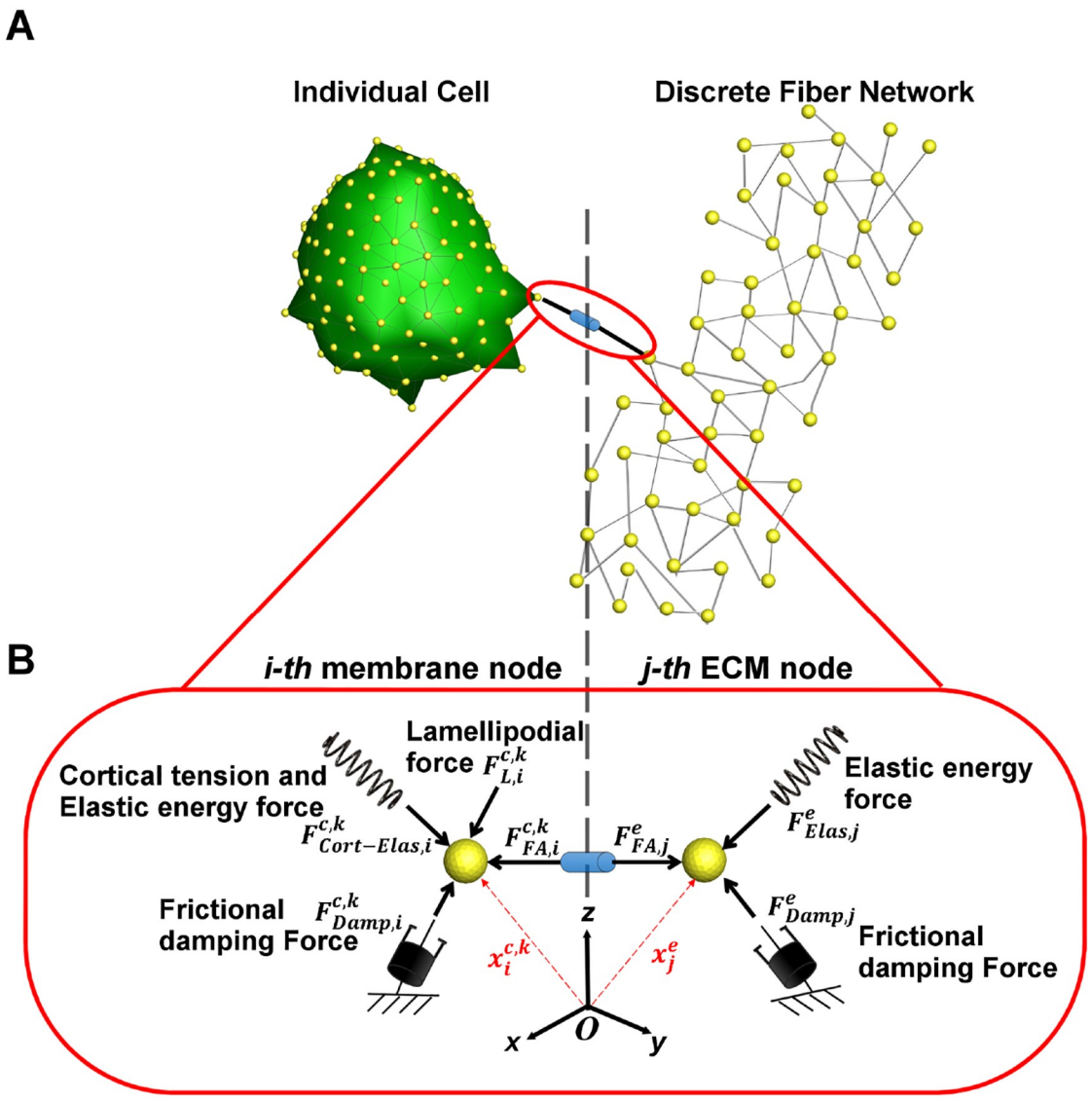

Fig 1. Schematic diagram of cell-ECM interaction. A: Each cell is represented by a mesh structure consisting of multiple nodes which are indicated by the yellow spheres. The ECM is modeled as a network of many fibers connected through a large number of nodes. The *i*-th membrane node is attached to the *j*-th ECM node through a focal adhesion connection. B: The forces acting on each membrane node include the cortical tension and membrane elastic energy force ($\mathbf{F}^{c,k}_{Cort-Elas,i}$), focal adhesion force ($\mathbf{F}^{c,k}_{FA,i}$), lamellipodium force ($\mathbf{F}^{c,k}_{L,i}$), and frictional damping force($\mathbf{F}^{c,k}_{Damp,i}$). The forces acting on each node within the ECM fiber network include the elastic energy force ($\mathbf{F}^{e}_{Elas,j}$), focal adhesion force ($\mathbf{F}^{e}_{FA,j}$), and damping force ($\mathbf{F}^{e}_{Damp,j}$).

https://doi.org/10.1371/journal.pcbi.1006798.g001

motion is given by:

$$\mathbf{F}^{c,k}_{Cort-Elas,i} + \mathbf{F}^{c,k}_{FA.i} + \mathbf{F}^{c,k}_{L,i} - D_c \frac{d\boldsymbol{x}^{c,k}_i}{dt} = 0$$
$$i = 1, \cdots, N_c,\ k = 1, \cdots, n_{cell} \tag{1}$$

The generation of lamellipodium force pertains to the polarity of the cell. Namely, lamellipodia extend in a particular direction of the cell determined by the cell's polarity [23–26]. The cell polarity and the lamellipodium forces can be treated as a cell's decision or, in the system

dynamics terminology, control inputs. Let $\boldsymbol{x}^{c,k} = \left(\boldsymbol{x}_1^{c,k\mathbf{T}} \cdots \boldsymbol{x}_{N_c}^{c,k\mathbf{T}}\right)^{\mathbf{T}} \in \Re^{3N_c \times 1}$ be a vector containing the 3-D coordinates of all the cell membrane nodes. Here the superscript in $\mathbf{X}^{\mathbf{T}}$ represents the transpose of matrix or vector $\mathbf{X}$. The above equation of motion can be written collectively as:

$$\frac{d\boldsymbol{x}^{c,k}}{dt} = \mathbf{W}_{CE}^{c}\mathbf{F}_{Cort-Elas}^{c,k} + \mathbf{W}_{FA}^{c}\mathbf{F}_{FA}^{c,k} + \mathbf{L}_c\mathbf{u}^k \qquad k = 1, \cdots, n_{cell} \tag{2}$$

where $\mathbf{F}_{Cort-Elas}^{c,k} \in \Re^{3N_C \times 1}$ is a vector comprising cortical tension and elastic energy forces for all the cell nodes ($i = 1, \cdots, N_C$), $\mathbf{F}_{FA}^{c,k} \in \Re^{3N_C \times 1}$ is a vector of focal adhesion forces at all the cell nodes, $\mathbf{u}^k \in \Re^{3N_C \times 1}$ is an input vector containing all the lamellipodium forces ($\mathbf{F}_{L,i}^{c,k}$), and $\mathbf{W}_{CE}^{c}$, $\mathbf{W}_{FA}^{c}$ and $\mathbf{L}_c$ are constant matrices of consistent dimensions.

The equation of motion of the surrounding ECM fiber network can be represented in a similar manner. The forces acting on the j-th node of the fiber network are the elastic energy forces, including both lateral restoring forces and the one associated with bending moments, ($\mathbf{F}_{Elas,j}^{e} \in \Re^{3\times 1}$), focal adhesion forces from the shared attachment with the cell ($\mathbf{F}_{FA,j}^{e} \in \Re^{3\times 1}$) and damping forces ($\mathbf{F}_{Damp,j}^{e} \in \Re^{3\times 1}$) [23–26]. The equation of motion can be written as:

$$\mathbf{F}_{Elas,j}^{e} + \mathbf{F}_{FA,j}^{e} - D_e\frac{d\boldsymbol{x}_j^{e}}{dt} = 0 \quad , j = 1, \cdots, N_e \tag{3}$$

Let $\boldsymbol{x}^{e} = (\boldsymbol{x}_1^{e\mathbf{T}} \cdots \boldsymbol{x}_{N_e}^{e\ \mathbf{T}})^{\mathbf{T}} \in \Re^{3N_e \times 1}$ be a vector containing the 3-D coordinates of all the ECM nodes. Then Eq 3 can be written as:

$$\frac{d\boldsymbol{x}^{e}}{dt} = \mathbf{W}_{Elas}^{e}\mathbf{F}_{Elas}^{e} + \mathbf{W}_{FA}^{e}\mathbf{F}_{FA}^{e} \tag{4}$$

The ECM elastic energy force is a nonlinear function of ECM coordinates $\boldsymbol{x}^{e}$. The cortical tension and elastic energy force of the k-th cell is a nonlinear function of its membrane coordinates $\boldsymbol{x}^{c,k}$. Here $\boldsymbol{x}^{e}$ and $\boldsymbol{x}^{c,k}$ are independent state variables of the multi-cell ECM system.

$$\mathbf{F}_{Elas}^{e} = \mathbf{F}_{Elas}^{e}(\boldsymbol{x}^{e}), \quad \mathbf{F}_{Cort-Elas}^{c,k} = \mathbf{F}_{Cort-Elas}^{c,k}(\boldsymbol{x}^{c,k}) \qquad k = 1, \cdots, n_{cell} \tag{5}$$

The focal adhesion force is modeled as a stochastic binding process between nodes on the cell membrane and those on the ECM. Using Monte Carlo simulations it has been found that focal adhesion forces can be approximated to a nonlinear algebraic function of cell membrane and ECM nodes as well as the biochemical parameters involved in integrin-ligand binding [23, 24].

$$\mathbf{F}_{FA}^{c,k} = \mathbf{F}_{FA}^{c,k}(\boldsymbol{x}^{c,k}, \boldsymbol{x}^{e}), k = 1, \cdots, n_{cell} \tag{6}$$

Assuming that no two cells bind to the same ECM node, we can find that the focal adhesion force of the i-th membrane node of the k-th cell attached to the j-th ECM node must satisfy:

$$\mathbf{F}_{FA,i}^{c,k} + \mathbf{F}_{FA,j}^{e} = 0 \tag{7}$$

Namely, $\mathbf{F}_{FA,i}^{c,k}$ and $\mathbf{F}_{FA,j}^{e}$ have the same magnitude with the opposite signs. Therefore, all the focal adhesion forces of the k-th cell can be mapped to the corresponding ECM nodes.

Collectively, the focal adhesion forces of all the nodes within the ECM may be written as:

$$\mathbf{F}_{FA}^{e} = \sum_{k=1}^{n_{cell}} \mathbf{P}_{map}^{k} \mathbf{F}_{FA}^{c,k} \tag{8}$$

where $\mathbf{P}_{map}^{k} \in \Re^{3N_e \times 3N_c}$ is a parameter matrix (consisting of either 0 or -1 elements) that maps the membrane focal adhesion forces of the individual cells ($\mathbf{F}_{FA}^{c,k}, k = 1, \ldots, n_{cell}$) to the corresponding ECM focal adhesion forces ($\mathbf{F}_{FA}^{e}$). The focal adhesion connections between the membrane nodes and ECM nodes change over time as the cell membrane deforms, gains traction and generates lamellipodial protrusions. Therefore, the mapping matrix $\mathbf{P}_{map}^{k}$ is updated at each time step. Details on the formation and structure of $\mathbf{P}_{map}^{k}$ are given in the Methods Section. The n_{cell} cells interact with each other through the surrounding ECM by generating focal adhesion forces, which propagate through the ECM fiber network and influence the other cells. The resultant collective behavior of the multiple cells is complex due to coupled, nonlinear dynamics.

Dual-faceted linearization

Although the governing equations derived above are rigorous and based on basic principles, they are complex and can become computationally expensive as the number of cells increases. Computational complexity is a key challenge in predicting collective behaviors of multiple cells. The number of state variables for the given system is $3N_e + 3N_c\, n_{cell}$, which is on the order of 7,000 for $n_{cell} = 2$ and 9000 for $n_{cell} = 5$. We aim to transform the governing equations into a linear latent variable representation in order to considerably reduce the number of state variables but also facilitate prediction of collective behaviors of the multiple cells through superposition of individual cell dynamics.

Model reduction is a challenging problem particularly for highly nonlinear, dynamical systems [21, 22, 27–29], as in the presented problem of collective behaviors of multiple cells within an ECM. If the system is linear or near linear, model reduction is more amenable and simple methods, such as Principal Component Analysis and Partial Least Squares, can reduce dimensionality. Here, we propose a unique linearization method, termed Dual-Faceted (DF) Linearization, and then apply a model reduction method to the linearized model. In DF Linearization, we represent the nonlinear dynamical system in an augmented space consisting of independent state variables ($\boldsymbol{x}^{e}$ and $\boldsymbol{x}^{c,k}$) and nonlinear forces ($\mathbf{F}_{Elas}^{e}, \mathbf{F}_{Cort-Elas}^{c,k}, \mathbf{F}_{FA}^{c,k}$) as the additional variables, termed auxiliary variables. Standard linearization, such as Taylor series expansion, is limited in accuracy, which may be valid only in the vicinity of a reference point. In DF Linearization, instead of taking "algebraic" linearization of these nonlinear terms, we consider "dynamic" linearization by representing their dynamic transitions using linear regressions.

Before formally introducing DF Linearization, let us consider a simple example that delineates the basic principle of the method. Suppose that the system consists of one spring and a damping element with negligibly small mass, $F - D\dot{x} = 0$. If the spring is a linear spring, $F = kx$, there is absolutely no difference between the equation in terms of state variable x, $\dot{x} = (k/D)x$, and the one in force F, $\dot{F} = (k/D)F$. However, it is not the case if the spring is nonlinear, for example a hard spring: $F = ax + bx^3$ where $a > 0$, $b > 0$. Representing the differential equation in two variables, one with the state variable x and the other with the auxiliary variable F, provide different equations.

$$\frac{dx}{dt} = \frac{a}{D}x + \frac{b}{D}x^3 \tag{9}$$

$$\frac{dF}{dt} = \frac{1}{D}\left(a + 3b(g(F))^2\right)F \tag{10}$$

where $x = g(F)$ s the inverse function of $F = ax + bx^3$. Both equations represent the same nonlinear system, yet the expressions are different, hence Dual-Faceted representations. Linearizing these differential equations lead to two linear differential equations viewed from the augmented space. Note that Eq 9 can be represented as a linear equation by using both state and auxiliary variables:

$$\frac{dx}{dt} = W_F F \tag{11}$$

where $W_F = 1/D$. The augmented state Eq 10 can be approximated to a linear regression:

$$\frac{dF}{dt} \simeq S_x x + S_F F \tag{12}$$

where S_x, S_F are regression parameters.

The expression given by Eq 12 differs from the one based on the first order Taylor expansion (or "algebraic" linearization) which yields:

$$F(x) \simeq F(\bar{x}) + \frac{dF}{dx}\Delta x \tag{13}$$

Furthermore, if we evaluate the derivative $J(x) \equiv dF/dx$ at a particular point, $\bar{J}(\bar{x})$, and then use Eq 13 to express the augmented state equation, it reduces to $\dot{F} = \bar{J}\,\dot{x}$. This implies that $\dot{F}$ and $\dot{x}$ are proportional. Using an "algebraic" linearization yields a differential equation representing the transition of F that is collinear to the one representing the transition of x, and thereby an auxiliary state equation would not provide any new information. Conversely, the regression model in Eq 12 provides us with a diverse view of the original nonlinear system, thus providing a richer representation of the nonlinearity than the standard first order Taylor expansion.

Applying the above principle of Dual Faceted Linearization to our problem, we note that the original state equations governing the transition of the independent state variables, 2 and 4, are linear in the augmented state space. All we need is to obtain the transition of the auxiliary variables. Let the regression of the dynamic transition of auxiliary variable $\mathbf{F}^e_{Elas}$, be expressed as:

$$\frac{d\mathbf{F}^e_{Elas}}{dt} \approx \mathbf{R}^e_x \boldsymbol{x}^e + \mathbf{R}^e_{F_{Elas}} \mathbf{F}^e_{Elas} \tag{14}$$

where $\mathbf{R}^e_x, \mathbf{R}^e_{F_{Elas}} \in \Re^{3N_e \times 3N_e}$ are parameter matrices. If an "algebraic" linearization using the Jacobian $\bar{J} = \partial \mathbf{F}^e_{Elas}/\partial \boldsymbol{x}^e|_{\bar{x}^e}$ was utilized, the above equation would be: $d\mathbf{F}^e_{Elas}/dt = \bar{J} \cdot d\boldsymbol{x}^e/dt$.

This state transition equation through "algebraic" linearization is equivalent to one of the original independent state Eq 4 because $d\mathbf{F}^e_{Elas}/dt$ and $d\boldsymbol{x}^e/dt$ are collinear within this formulation which renders it redundant. In contrast, the state transition equation presented in Eq 14 is not collinear, providing a diverse facet of the nonlinear system. Similarly, for the auxiliary

variables $\mathbf{F}^{c,k}_{Cort-Elas}$, $\mathbf{F}^{c,k}_{FA}$, let the regression equations be written as:

$$\frac{d\mathbf{F}^{c,k}_{Cort-Elas}}{dt} \simeq \mathbf{Q}^{c}_{x}\boldsymbol{x}^{c,k} + \mathbf{Q}^{c}_{F_{CE}}\mathbf{F}^{c,k}_{Cort-Elas} + \mathbf{Q}_{u}\mathbf{u}^{k} \tag{15}$$

$$\frac{d\mathbf{F}^{c,k}_{FA}}{dt} \simeq \mathbf{H}^{c}_{x}\boldsymbol{x}^{c,k} + \mathbf{H}^{e}_{x}\boldsymbol{x}^{e} + \mathbf{H}^{c}_{F_{FA}}\mathbf{F}^{c,k}_{FA} + \mathbf{H}_{u}\mathbf{u}^{k} \tag{16}$$

where $\mathbf{Q}^{\star}_{\star}$, $\mathbf{H}^{\star}_{\star}$ ($\star$ -corresponding subscript and superscript) are parameter matrices with consistent dimensions. The high-dimensional parameter matrices ($\mathbf{R}^{\star}_{\star}$, $\mathbf{Q}^{\star}_{\star}$, $\mathbf{H}^{\star}_{\star}$) do not need to be determined explicitly as discussed in the subsequent sections. DF Linearization represents a nonlinear dynamical system with two sets of differential equations. One set is the original state equations governing the transition of the independent state variables and the other set is the regression of the dynamics of auxiliary variables. The original state Eqs 2 and 4, are apparently linear in terms of the auxiliary variables and inputs. In these equations, all the forces acting on each node sum to zero. These are linear expressions when the nonlinear forces are treated as auxiliary variables. In addition, the auxiliary state transitions (Eqs 14–16) are given by linear regressions in the augmented space. Therefore, both differential equations are linear. The two linear differential equations represent different (or dual) facets of the original nonlinear system viewed from the augmented space, thus providing a richer representation of the nonlinearity.

Latent variable transformation and model order reduction

Now that the original nonlinear system has been represented as a linear dynamical system in the augmented space, we can apply a latent variable modeling method to reduce model order. Represented in the augmented space, the differential equations may contain similar modes, or some variables are close to collinear. These similar modes and collinear variables can be eliminated by model order reduction methods.

Let $\zeta^{c,k}$ be the augmented variable vector containing membrane node coordinates and forces of the k-th cell.

$$\zeta^{c,k} = \begin{pmatrix} \boldsymbol{x}^{c,k} \\ \mathbf{F}^{c,k}_{Cort-Elas} \\ \mathbf{F}^{c,k}_{FA} \end{pmatrix} \in \Re^{9N_c \times 1} \tag{17}$$

Here $\mathbf{u}^{k}$ (the cell's lamellipodial force) is treated as input variables that are excluded from the augmented state space. Similarly, let ζ^{e} be the augmented variable vector containing ECM node coordinates and forces:

$$\zeta^{e} = \begin{pmatrix} \boldsymbol{x}^{e} \\ \mathbf{F}^{e}_{Elas} \end{pmatrix} \in \Re^{6N_e \times 1} \tag{18}$$

Focal adhesion forces $\mathbf{F}^{e}_{FA}$ are determined by the individual cells by Eq 8 and, thereby, excluded from the augmented space of the ECM.

We apply latent space analysis to vectors $\zeta^{c,k}$ and ζ^{e}, respectively. First we generate data by using Eqs 2 and 4–7. Computation of the nonlinear state equations is amenable for a single cell interacting with ECM. A data set can be created by simulating those nonlinear equations by placing a single cell at diverse locations, i.e. repeating the simulation with different initial conditions. Let $\mathbf{C}^{c}_{\zeta\zeta}$, $\mathbf{C}^{e}_{\zeta\zeta}$ be the covariance matrices of simulation data sets of augmented states $\zeta^{c,k}$

and ζ^e, respectively. Each covariance matrix contains both independent state and auxiliary variables, where the latter is nonlinear functions of the former. If auxuliary variables were linear functions of the state variables, then the rank of the covariance matrix would be equal to the number of independent state variables. However due to the nonlinearity, the rank is higher. Details on the formation of $\mathbf{C}^c_{\zeta\zeta}$, $\mathbf{C}^e_{\zeta\zeta}$ are given in the Methods Section.

The covariance analysis also reveals that the system represented in the augmented space contains many components that may be negligibly small. Using Principal Component Analysis, the original data of $\zeta^{c,k}$ and ζ^e can be represented with latent variables of truncated dimension $m_c \ll 9N_c$ and $m_e \ll 6N_e$, respectively [29]:

$$\zeta^{c,k} = \underbrace{\begin{pmatrix} \mathbf{V}^c_x \\ \mathbf{V}^c_{F_{CE}} \\ \mathbf{V}^c_{F_{FA}} \end{pmatrix}}_{\mathbf{V}^c} \mathbf{z}^{c,k}, \quad \zeta^e = \underbrace{\begin{pmatrix} \mathbf{V}^e_x \\ \mathbf{V}^e_{F_{Elas}} \end{pmatrix}}_{\mathbf{V}^e} \mathbf{z}^e \tag{19}$$

where matrices $\mathbf{V}^c \in \Re^{9N_c \times m}$ and $\mathbf{V}^e \in \Re^{6N_e \times m}$ are orthogonal matrices comprising the eigenvectors of the covariance matrices, and

$$\begin{aligned} \mathbf{z}^{c,k} &= \begin{pmatrix} \mathbf{V}^c_x \\ \mathbf{V}^c_{F_{CE}} \\ \mathbf{V}^c_{F_{FA}} \end{pmatrix}^{\mathbf{T}} \begin{pmatrix} \boldsymbol{x}^{c,k} \\ \mathbf{F}^{c,k}_{Cort-Elas} \\ \mathbf{F}^{c,k}_{FA} \end{pmatrix} \in \Re^{m_c \times 1} \\ \mathbf{z}^e &= \begin{pmatrix} \mathbf{V}^e_x \\ \mathbf{V}^e_{F_{Elas}} \end{pmatrix}^{\mathbf{T}} \begin{pmatrix} \boldsymbol{x}^e \\ \mathbf{F}^e_{Elas} \end{pmatrix} \in \Re^{m_e \times 1} \end{aligned} \tag{20}$$

Differentiating the latent variable state vector $\mathbf{z}^{\mathbf{c,k}}$ and substituting Eqs 2, 15, 16 and 19 yields:

$$\begin{aligned} \frac{d\mathbf{z}^{c,k}}{dt} &= \mathbf{V}^{c\,\mathbf{T}}_x \frac{d\boldsymbol{x}^{c,k}}{dt} + \mathbf{V}^{c\ \mathbf{T}}_{F_{CE}} \frac{d\mathbf{F}^{c,k}_{Cort-Elas}}{dt} + \mathbf{V}^{c\ \mathbf{T}}_{F_{FA}} \frac{d\mathbf{F}^{c,k}_{FA}}{dt} \\ &= \mathbf{A}\,\mathbf{z}^{c,k} + \mathbf{B}\,\mathbf{u}^k + \mathbf{C}\mathbf{z}^e, \quad k = 1, \cdots, n_{cell} \end{aligned} \tag{21}$$

where:

$$\begin{aligned} \mathbf{A} &= \mathbf{V}^{c\,\mathbf{T}}_x (\mathbf{W}^c_{CE} \mathbf{V}^c_{F_{CE}} + \mathbf{W}^c_{FA} \mathbf{V}^c_{F_{FA}}) \\ &\quad + \mathbf{V}^{c\ \mathbf{T}}_{F_{CE}} (\mathbf{Q}^c_x \mathbf{V}^c_x + \mathbf{Q}^c_{F_{CE}} \mathbf{V}^c_{F_{CE}}) \\ &\quad + \mathbf{V}^{c\ \mathbf{T}}_{F_{FA}} (\mathbf{H}^c_x \mathbf{V}^c_x + \mathbf{H}^c_{F_{FA}} \mathbf{V}^c_{F_{FA}}) \\ \mathbf{B} &= \mathbf{V}^{c\,\mathbf{T}}_x \mathbf{L}_c + \mathbf{V}^{c\ \mathbf{T}}_{F_{CE}} \mathbf{Q}_u + \mathbf{V}^{c\ \mathbf{T}}_{F_{FA}} \mathbf{H}_u \\ \mathbf{C} &= \mathbf{V}^{c\ \mathbf{T}}_{F_{FA}} \mathbf{H}^e_x \mathbf{V}^e_x \end{aligned} \tag{22}$$

Eq 21 provides the latent variable state equation of the k-th cell interacting with the ECM. Given the latent variable state of ECM $\mathbf{z}^e$ and the input $\mathbf{u}^k$ reflecting the cell's decision, the transition of the cell's latent variable state is determined locally without directly including the states of the other cells. Cells interact indirectly through the strain field created by other cells over the ECM fiber network.

The ECM dynamics can be represented in the latent variable space spanned by $\mathbf{V}^e$. Differentiating the latent variable state vector and substituting Eqs 4, 14 and 19 yield:

$$\frac{d\mathbf{z}^e}{dt} = \mathbf{V}_x^{e\,\mathrm{T}} \frac{d\boldsymbol{x}^e}{dt} + \mathbf{V}_{F_{Elas}}^{e\,\mathrm{T}} \frac{d\mathbf{F}_{Elas}^e}{dt} = \mathbf{G}\,\mathbf{z}^e + \sum_{k=1}^{n_{cell}} \mathbf{D}^k \mathbf{z}^{c,k} \tag{23}$$

where:

$$\begin{aligned} \mathbf{G} &= \mathbf{V}_x^{e\,\mathrm{T}} \mathbf{W}_{Elas}^e \mathbf{V}_{F_{Elas}}^e + \mathbf{V}_{F_{Elas}}^{e\,\mathrm{T}} (\mathbf{R}_x^e \mathbf{V}_x^e + \mathbf{R}_{F_{Elas}}^e \mathbf{V}_{F_{Elas}}^e) \\ \mathbf{D}^k &= \mathbf{V}_x^{e\,\mathrm{T}} \mathbf{W}_{FA}^e \mathbf{P}_{map}^k \mathbf{V}_{F_{FA}}^c \end{aligned} \tag{24}$$

Fig 2 shows numerical examples of the DF linearization and subsequent latent variable transformation in reproducing accurate cell morphologies of the original nonlinear computational model over time. Remarkably, the DF linearized latent variable model can correctly reconstruct the complex cell membrane topology with $m = m_c + m_e = 50 + 50 = 100$ total latent variables. Fig 2C quantifies the root mean square error and computation time as a function of latent variable model dimension. As can be seen, the computation time for the latent variable cell-ECM system increases with increased latent variables while the root mean square error decreases. Conversely, the standard Taylor expansion linearization is not capable of representing cell morphologies without marked error which is quantified in Fig 2B. We also compare our DF linearization approach to a more sophisticated method for approximation of nonlinear systems termed trajectory piece-wise linear (TPWL) [30, 31]. This method uses collection of (algebraic) linearizations of the original nonlinear system about suitably selected states to approximate the nonlinear system. As can be seen from Fig 2B, although the TPWL method yields lower error than the first order taylor expansion, our DF linearized model still leads to significantly lower prediction error. This is because in DF linearization, instead of taking "algebraic" linearization of nonlinear terms, we consider "dynamic" linearization by representing their dynamic transitions using linear regressions. These results demonstrate the effectiveness

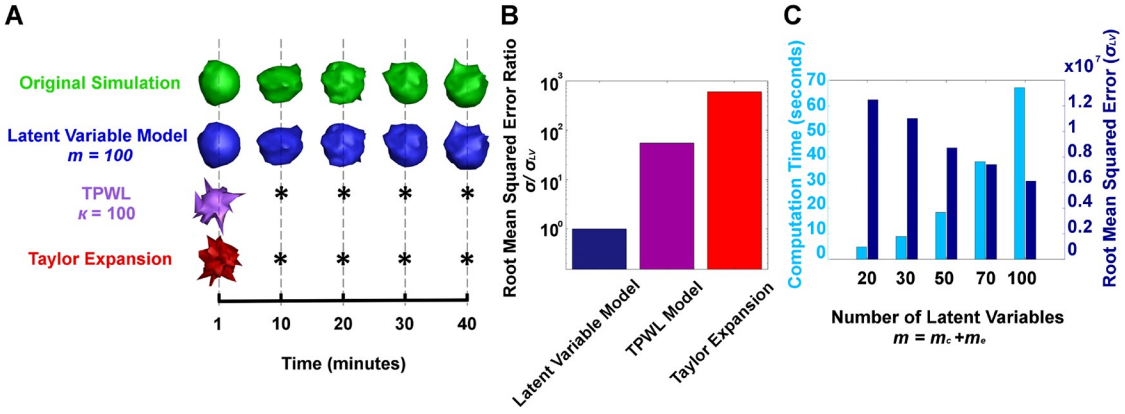

Fig 2. Comparison of nonlinear computational model, latent variable model, trajectory piece-wise linearization model, and taylor series expansion model for predicting a cell interacting with ECM. A: The cell morphologies over time for the original full nonlinear simulation (green), the latent variable simulation using 100 latent variables (blue), the trajectory piecewise-linear model (TPWL) with 100 linearization points (purple) and the Taylor series expansion model (red). The Taylor series and TPWL model are not capable of representing cell morphologies without significant error (represented by *). B: Comparison of root mean squared error of latent variable model using 100 latent variables, TPWL model using $\kappa = 100$ linearization points and first order Taylor expansion model. C: The computation time and root mean squared error (σ_{LV}) of the latent variable model as a function of the number of latent variables used to create the model.

https://doi.org/10.1371/journal.pcbi.1006798.g002

of DF linearization and model reduction in reconstructing simulations from a high dimensional complex nonlinear model.

This latent space model provides not only a low-dimensional structure for efficient computation, but also contains natural insights into the interactions among the multiple cells. Fig 3A shows the dynamic interactions in block diagram form based on Eqs 21 and 23. The ECM changes its latent variable state $\mathbf{z}^e$ with the autoregressive feedback through matrix **G** as well as with the forward path that collects the latent variable states of all the individual cells, as shown by the summing junction Σ. Each single cell changes its latent variable state $\mathbf{z}^{c,k}$ ($k = 1, \ldots, n_{cell}$) with autoregressive feedback through **A** as well as with two forward paths. The first path (fed through **C**) represents global feedback from the ECM ($\mathbf{z}^e$). The second path (fed through **B**) represents the updated lamellipodial forces $\mathbf{u}^k$ determined from ECM state $\mathbf{z}^e$. The lamellipodial forces can be thought of as the individual cell's decision based on its position and updated ECM properties as explained more in the following section. The actions taken by all the cells are integrated into the global ECM state transition, which is fed back to the individual cells. Therefore, each cell is connected to other cells through the global feedback of the ECM latent variable state $\mathbf{z}^e$. Fig 3A manifests the control-theoretic interpretation of multiple cells interacting through ECM.

Since the system is represented in a lower dimensional space, the high dimensional regression coefficient matrices ($\mathbf{R}^\star_\star$, $\mathbf{Q}^\star_\star$, $\mathbf{H}^\star_\star$) are not computed explicitly. Instead, the lower dimension coefficient matrices **A**, **B**, **C**, **G** are computed from numerical simulation data that can be transformed into latent variable space. Details are given in the Methods Section.

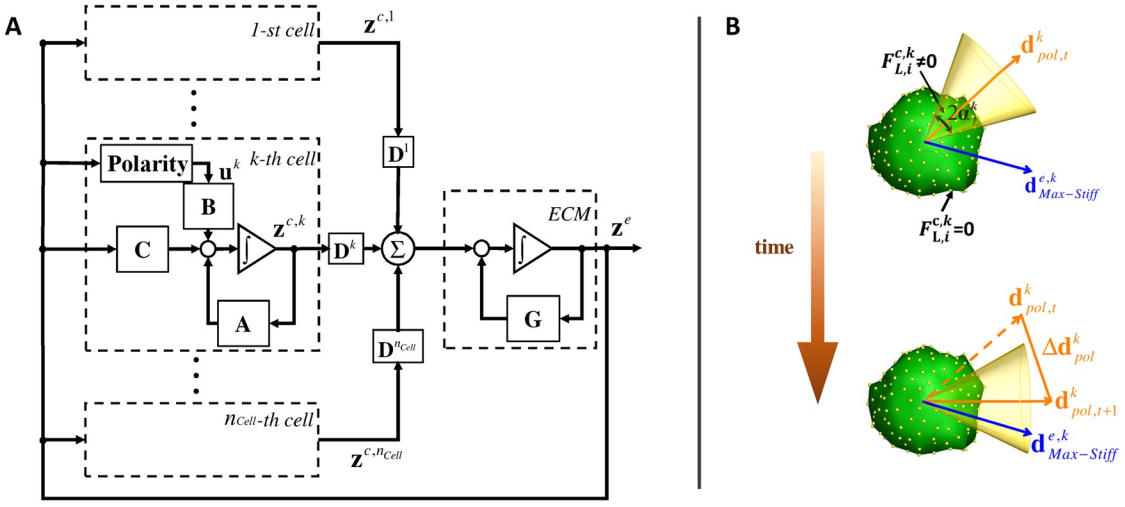

Fig 3. Block diagram of latent variable superposition model and schematic of the relationship between polarity direction, leading edge and direction of maximum stiffness. A: The ECM changes its latent state with the autoregressive feedback through matrix G as well as with the feedforward path which collects the latent variable states of all the individual cells $\mathbf{z}^{c,k}(k = 1, \ldots, n_{cell})$ through matrices $\mathbf{D}^k$. Each cell changes its latent variable state with autoregressive feedback through **A** and are exposed to the ECM forces represented by latent vector $\mathbf{z}^e$ in two separate paths. The path through the cell polarity block and matrix B can be viewed as an "active input". This feedback path includes a cell's internal decision as to which direction it extends lamellipodia. In contrast, the other feedback path through a gain matrix **C** does not have a high-level cell decision, but is reactive, playing a "passive role". B: The cell polarity direction rotates dynamically in such a way that the polarity vector $\mathbf{d}^k_{Pol}$ may align with the direction of the maximum stiffness $\mathbf{d}^{e,k}_{Max-Stiff}$. The leading edge of the cell is indicated by a right circular cone with apex angle $2\alpha^k_L$ having its centerline aligned with the polarity direction. The membrane nodes of the *k*-th cell within the cone have nonzero lamellipodial forces ($\mathbf{F}^{c,k}_{L,i} \neq 0$). Membrane nodes outside this cone have zero lamellipodial forces ($\mathbf{F}^{c,k}_{L,i} = 0$).

https://doi.org/10.1371/journal.pcbi.1006798.g003

Polarity and cell decisions

As discussed previously, input vector $\mathbf{u}^k$ pertains to the lamellipodial forces generated at each membrane node within the leading edge of the cell. The cell continuously updates its lamellipodial protrusions depending on the orientation of the leading edge as the cell's polarization (or polarity) changes. The polarity of a cell is important to determine the orientation of the leading edge and is influenced by the direction of local maximum stiffness in the ECM [23–26]. Here we aim to extend the dynamics model of cell polarity developed in [23–25] to predict the formation of lamellipodia.

Let $\mathbf{d}^k_{Pol} \in \Re^{3\times 1}$ be a 3-dimensional unit vector indicating the direction of polarity in the k-th cell and $\mathbf{d}^{e,k}_{Max-Stiff} \in \Re^{3\times 1}$ be a 3-dimensional unit vector pointing in the direction of the maximum stiffness of ECM in the vicinity of the k-th cell's current location. According to [25], the cell polarity rotates dynamically in response to ECM's local stiffness in such a way that the polarity vector may align with the direction of the maximum stiffness:

$$\frac{d\mathbf{d}^k_{Pol}}{dt} = \kappa \mathbf{d}^k_{Pol} \times \left(\mathbf{d}^{e,k}_{Max-Stiff} \times \mathbf{d}^k_{Pol} \right) \tag{25}$$

where $\times$ indicates vector product, and κ is a scalar parameter. Fig 3B illustrates this relationship. The polarity vector $\mathbf{d}^k_{Pol}$ tends to align with the maximum stiffness direction, $\mathbf{d}^{e,k}_{Max-Stiff}$.

The leading edge of the cell is indicated by a right circular cone with apex angle $2\alpha^k_L$ having its centerline aligned with the polarity direction. The membrane nodes of the k-th cell within the cone have nonzero lamellipodial forces ($\mathbf{F}^{c,k}_{L,i} \neq 0$). Membrane nodes outside this cone have zero lamellipodial forces ($\mathbf{F}^{c,k}_{L,i} = 0$). The direction of maximum ECM stiffness $\mathbf{d}^{e,k}_{Max-Stiff}$ depends on the stress field within the ECM, which pertains to the latent state vector of ECM $\mathbf{z}^e$. The details are given in Appendix C in S1 Text.

Computational analysis of ECM compaction by multiple cells

Compaction of the ECM by the collective efforts of multiple cells is numerically analyzed based on the model reduction and superposition of the nonlinear cell-ECM dynamics via DF Linearization. We first consider the case where two cells placed 30 μm apart are embedded in a 3-D cylindrical ECM that measures 40 μm in diameter and 100 μm in length as seen in Fig 4A. The boundary conditions of the ECM fiber network are set such that the two flat planes on sides are fixed to space (constrained), while the curved surface surrounding the ECM is kept free (unconstrained). The volume of the cylindrical ECM shrinks over time from its initial unstressed state as the cells interact with the surrounding ECM. To quantify the spatiotemporal compaction process, the original ECM cylinder is segmented into 10 slices of 10 μm thickness along its longitudinal axis as shown in Fig 4B. The volumetric changes to the individual slices are plotted in Fig 4C. The prediction of decreased cell volume by the latent variable superposition model (blue) agrees well with the ground-truth, full-scale nonlinear simulation results (green). This is further verified by the corresponding cross-sectional images of the 2-cell cylindrical ECM simulations in Fig 4C. The polarity directions of both cells (shown by red arrows initially pointing in arbitrary directions) shift to point inward, indicating that larger stresses are detected in the area between the cells. A video of the simulation comparison is shown in S1 Video.

The proposed model is able to reproduce collective behaviors of multiple cells causing the characteristic compaction of ECM gel, which is not observed for single isolated cell models. This is further verified in Fig 4D which compares the ECM compaction results between single cell and two cell models. As can be seen, the single cell model predicts more localized

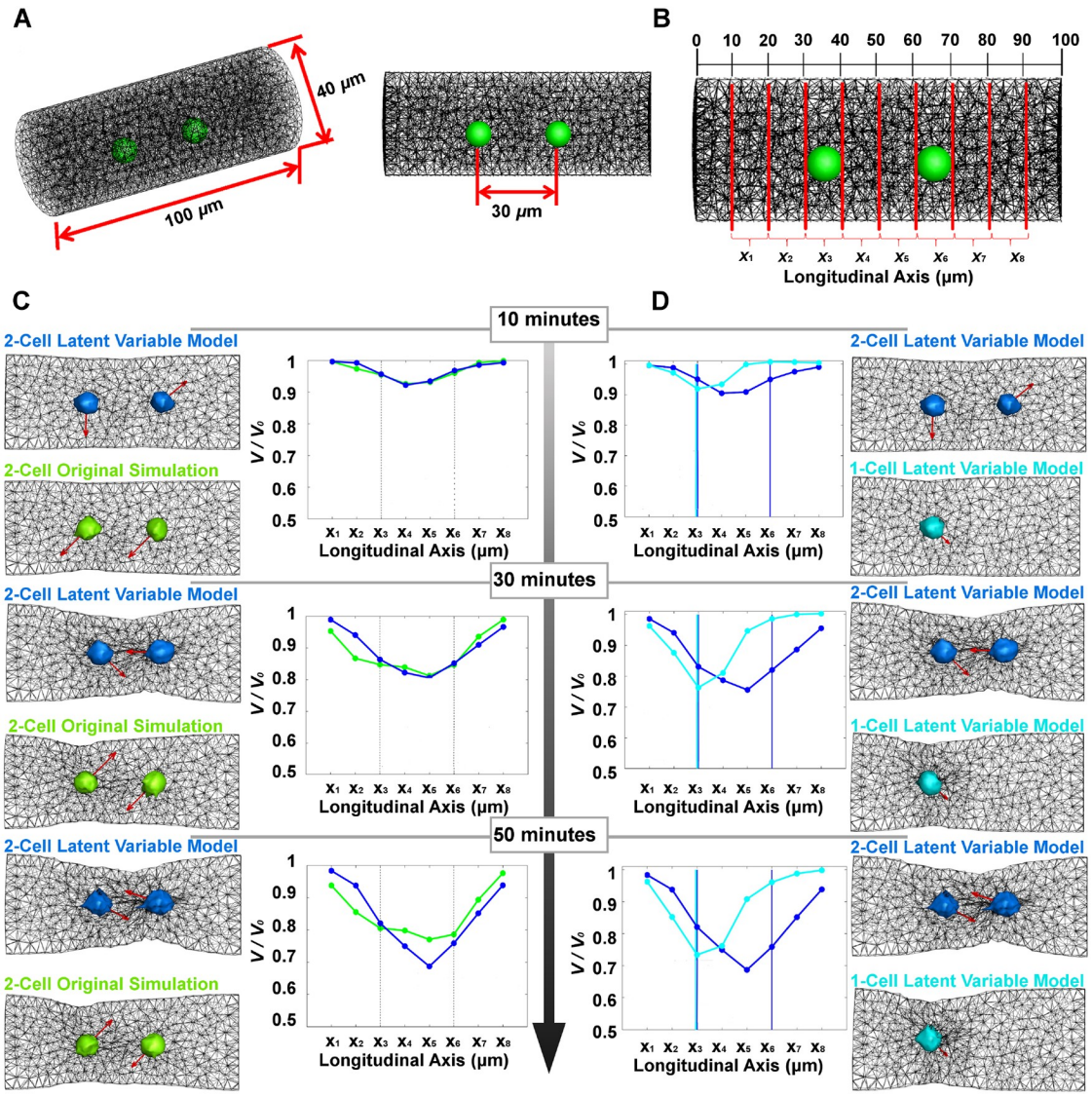

Fig 4. Comparison of ECM compaction between nonlinear computational model and linear latent variable models. A: Two cells placed 30 μm apart are embedded in a 3-D cylindrical ECM that measures 40 μm in diameter and 100 μm in length. B: The ECM is subdivided along the longitudinal axis into 10 μm length cylinders. The volume (V) of each subdivided cylinder may be estimated at multiple time points during the compaction simulation. C: ECM comparison at the subdivided segments at time t = 10 min, 30 min, and 50 min. Compaction predicted by the latent variable model simulations (blue) agrees well with the full nonlinear simulations (green). The compaction volume is normalized with the initial volume of each segment. The compaction is most significant in-between the cells (the region between the dashed lines in the plots). This is further verified by the corresponding cross-sectional images of the 2-cell cylindrical ECM simulations. Polarity directions of both cells (red arrows initially pointing in arbitrary directions) shift to point inward, indicating that larger stresses are detected in the area between the cells. D: Comparison of single cell (cyan) and two cell (blue) compaction predicted by the latent variable superposition model. As can be seen, the single cell model predicts more localized shrinkage of the ECM volume from its original unstressed state whereas the two cell model shows more global shrinkage extended to within the region between cells.

https://doi.org/10.1371/journal.pcbi.1006798.g004

shrinkage of the ECM volume whereas the two cell model shows more global shrinkage extended to within the region between the cells. A video of the simulation comparison is shown in S2 Video.

Fig 4D suggests the presence of more than one cell is necessary for the pronounced ECM compaction leading to emergent changes within the ECM. However, the emergence of

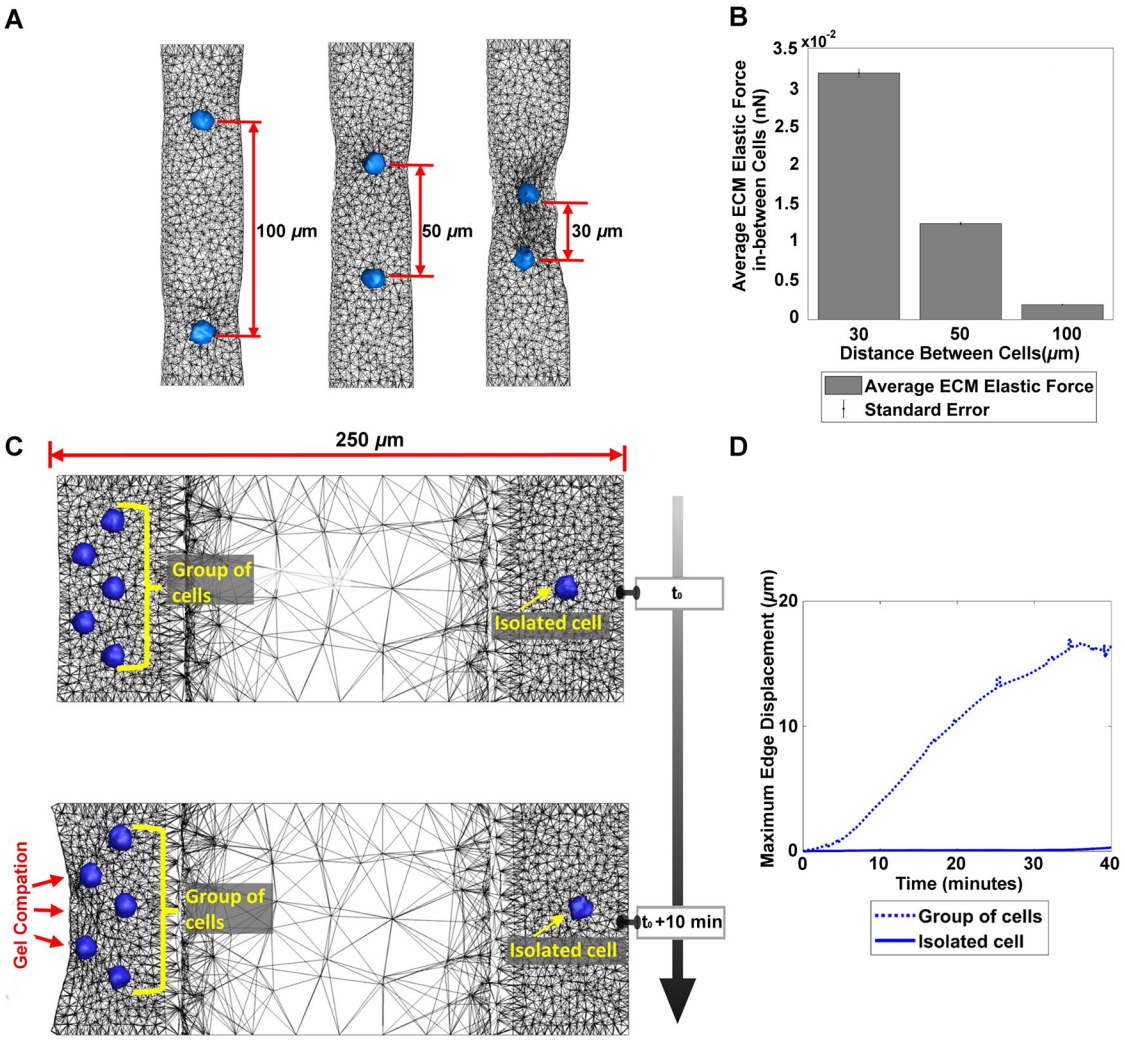

Fig 5. Effect of cell spacing and density on ECM compaction. (A) ECM compaction by two cells embedded within a cylindrical ECM spaced at 100μm, 50μm, and 30 μm. As the spacing between cells increases, compaction is less pronounced between them. (B) The average ECM elastic force in-between cells spaced at 100μm is an order of magnitude less than the elastic force in-between cells spaced at 30μm. (C) The latent variable superposition simulation (upper) can be used to reflect the behavior of heterogeneous planar distribution of MC3T3-E1 osteoblasts as shown in supplementary video in [5]. Whereas the group of cells at the left edge contract the gel, the isolated cell at the right edge did not contract the gel, indicating the importance of cell density for compaction. (D) For the latent variable superposition simulation the maximum displacement along the ECM edge noticeably increases over time for the group of cells and is minimal for the isolated cell.

https://doi.org/10.1371/journal.pcbi.1006798.g005

pronounced compaction entails not only plurality of cells but proper cell spacing. Fig 5A shows that, as the spacing between cells increases, compaction is less pronounced between them, indicating decreased interaction and integration of cell induced propagated forces. This is summarized in Fig 5B which quantifies the average ECM elastic force in-between cells against cell spacing. From Fig 5B, we see that the average ECM elastic force in-between cells spaced at 100 μm is an order of magnitude less than that of the cells spaced at 30 μm. A video of this simulation is shown in S3 Video.

The above computational results shown in Figs 4 and 5 verify that the proposed method can capture collective behaviors of multiple cells. The verification was made by comparing the

reduced-order superposition model using DF Linearization against the full-scale, nonlinear model. To further verify the capability of the reduced-order superposition model, a comparison is also made against in vitro experimental data of ECM compaction by a larger number of cells. As shown in Fig 5C and 5D, the computational model successfully reproduces the in vitro experiment conducted by Fernandez, et al [5], in which a heterogeneous planar distribution of MC3T3-E1 osteoblasts were plated in 3-D rectangular prism collagen gel of 50 μm height, 100μm width and 250 μm length. The boundary conditions of ECM in the computational model were set to be consistent with experimental conditions in [5].

The multi-cell latent variable simulation is able to predict the characteristics of the ECM over time. Whereas the group of 5 cells at the left edge exhibit anisotropic contraction of the ECM at the boundary, the single isolated cell at the right edge barely contracts the gel. Fig 5C compares the isolated cell to the group of cells in terms of maximum edge displacement. The isolated cell's ECM edge displacement is so small that five times its displacement is still substantially lower than the displacement of the group of cells. A video of this simulation is shown in the S4 Video. The presented method for predicting collective behaviors of cell-mediated ECM gel compaction is scalable. Since the individual cell-ECM interactions are local computations, as given by Eq 21, the computational complexity does not increase exponentially, although the number of cells increases.

Methods

Implementation of the presented method is enabled by four key constructs: 1) generation of simulated training data, 2) formation of the data covariance matrix necessary for latent variable transformations, 3) estimation of the parameter matrices involved in the latent variable space state equations (Eqs 21 and 23), and 4) association of cell and ECM focal adhesion force variables using mapping matrix. Each method is detailed below.

Generation of simulated training data based on the original nonlinear equations of single cells

The nonlinear state Eqs 2 and 4–6 were computed with custom C-code based on references [23, 24]. The computation of a single isolated cell embedded in the cylindrical ECM took approximately 24 hours for a single simulation of physical time $T \approx 3600$ seconds with sampling interval of 1 second. The simulation was repeated for $N \approx 10$ times at different initial cell locations each time. The simulations with a single isolated cell embedded in the large rectangular ECM for reproducing the experimental result were run for approximately 5 days (120 hours). The physical time of simulation was $T \approx 3600$. The simulation was repeated for $N \approx 10$ for various initial locations of a cell. For each simulation, the total number of sample points for all the variables was over 5,000,000. The number of sample points was 5×10^7. The computation was performed on Intel Xeon CPU E5-2687W @ 3.10 GHz (2 processors) with 32 logical cores. More details on the formulation of the full-scale nonlinear state equations are summarized in Appendix A in S1 Text.

Formation of the data covariance matrix necessary for latent variable transformations

We create the covariance matrix using simulated training data. The training data consists of 3,600 time points of both state and auxiliary variables of a cell embedded in an ECM environment. The simulation is repeated $N = 10$ times, each time with the cell embedded in distinct

locations within the ECM. Then the data covariance matrices may be formed:

$$C^c_{\zeta\zeta} = \frac{1}{K \cdot N \cdot T}\sum_{k=1}^{K}\sum_{n=1}^{N}\sum_{t=1}^{T}\tilde{\zeta}^{c,k,n}(t)\tilde{\zeta}^{c,k,n}(t)^{\mathbf{T}}$$
$$C^e_{\zeta\zeta} = \frac{1}{N \cdot T}\sum_{n=1}^{N}\sum_{t=1}^{T}\tilde{\zeta}^{e,n}(t)\tilde{\zeta}^{e,n}(t)^{\mathbf{T}} \tag{26}$$

where $\tilde{\zeta}^{c,n,k}(t)$ represents the mean centered t-th time sample (of augmented variable vector $\zeta^{c,k}$) for the k-th cell (here $K = 1$ or 2) embedded within for the ECM at the n-th simulation, and $\tilde{\zeta}^{e,n}(t)$ represents the mean centered t-th time sample of the augmented variable vector ζ_e in the n-th simulation. By performing eigen-decomposition on the covariance matrices we obtain the orthogonal matrix $\mathbf{V}^c$, $\mathbf{V}^e$ comprised the eigenvectors of the data covariance matrix:

$$C^c_{\zeta\zeta} \approx \mathbf{V}^c\mathbf{\Lambda}^c\mathbf{V}^{c\mathbf{T}} \in \Re^{m_c \times m_c}$$
$$C^e_{\zeta\zeta} \approx \mathbf{V}^e\mathbf{\Lambda}^e\mathbf{V}^{e\mathbf{T}} \in \Re^{m_e \times m_e} \tag{27}$$

where $\mathbf{\Lambda}^c$, $\mathbf{\Lambda}^e$ are diagonal matrices containing the largest m_c and m_e eigenvalues of the covariance matrices, respectively. It is important to check whether the covariance matrices contain sufficiently rich data, and their first m_c and m_e components are sufficient to capture the cell-ECM dynamics at any cell location within the ECM. Standard techniques can be applied to validate the data and truncation of components [29]. With these, the ECM dynamics of a cell embedded within the ECM at an arbitrary location will be well represented in linear latent variable space which is critical to the success of the method. Covariance matrix calculation were conducted by using Matlab.

Estimation of parameter matrices involved in latent variable space state equations

The following outlines the steps to compute coefficient matrices **A**, **B**, **C**, **G** by Eqs 21 and 23:

1. Create training data by simulating the original state Eqs, 2 and 4, using the full-scale, nonlinear model of single cells, as described above.
2. Compute covariance matrices $C^c_{\zeta\zeta}$, $C^e_{\zeta\zeta}$, and obtain eigenvalues and eigenvectors $\mathbf{V}^c$ and $\mathbf{V}^e$, as described in the Methods Section.
3. Transform the data of the augmented state variables to latent variable space ($\mathbf{z}^{c,k,n}(t)$ and $\mathbf{z}^{e,n}(t)$) using the orthogonal matrices $\mathbf{V}^c$ and $\mathbf{V}^e$.
4. Compute time derivatives $d\mathbf{z}^{c,k,n}/dt$ and $d\mathbf{z}^{e,n}/dt$, using latent variable space time samples and form a dataset:

$$Z_{Tr} = \{(\mathbf{z}^{e,n}(t), \mathbf{z}^{c,k,n}(t), \mathbf{u}^{k,n}(t), d\mathbf{z}^{e,n}/dt, d\mathbf{z}^{c,k,n}/dt) | k = 1, \cdots, K,\ n = 1, \cdots, N,\ t = 1, \cdots, T\}$$

identify the parameter matrices A, B, C, G involved in the latent variable space state equations using Least Squares Estimate. The details about the Least Squares Estimate computation are given in Appendix B in S1 Text.

Parameter matrices $\mathbf{A} \in \Re^{m_c \times m_c}$, $\mathbf{B} \in \Re^{m_c \times 3N_c}$, $\mathbf{C} \in \Re^{m_c \times m_e}$, $\mathbf{G} \in \Re^{m_e \times m_e}$ are substantially lower in dimension than the regression coefficient matrices $\mathbf{R}^*_*$, $\mathbf{Q}^*_*$, $\mathbf{H}^*_*$ given in Eqs 14, 15 and 16. Therefore, fewer data points allow us to determine these parameter matrices in the latent variable space. It should be noted that matrices $\mathbf{D}^k$'s are of high dimension, but are not

computed with regression since they consist of known matrices as shown in Eq 24. Matlab was used for estimation of parameter matrices and subsequent computations of the latent variable model. 3-D visualization of simulation data was conducted using Tecplot 360.

Association of cell and ECM focal adhesion force variables using mapping matrix

When a focal adhesion is formed between the i-th node of the k-th cell and the j-th node of ECM, the two focal adhesion forces sum to zero, as described previously ($\mathbf{F}^{c,k}_{FA,i} + \mathbf{F}^{e}_{FA,j} = 0$ where $\mathbf{F}^{c,k}_{FA,i}, \mathbf{F}^{e}_{FA,j} \in \Re^{3\times 1}$). Representing this relationship in terms of the collective focal adhesion force vectors, $\mathbf{F}^{c,k}_{FA} \in \Re^{3N_c\times 1}$ and $\mathbf{F}^{e}_{FA} \in \Re^{3N_e\times 1}$, requires a matrix $\mathbf{P}^{k}_{map} \in \Re^{3N_e\times 3N_c}$. Let $\mathbf{F}^{e,k}_{FA}$ be the forces acting on the ECM nodes caused by focal adhesion between the k-th cell and the ECM nodes. This can be written as

$$\mathbf{F}^{e,k}_{FA} = \begin{pmatrix} \vdots \\ \mathbf{F}^{e,k}_{FA,j} \\ \vdots \end{pmatrix} = \begin{matrix} \\ j \rightarrow \end{matrix} \underbrace{\begin{pmatrix} \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & -\mathbf{I}_{3\times 3} & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix}}_{\mathbf{P}^{k}_{map}} \begin{matrix} \\ i \\ \downarrow \end{matrix} \begin{pmatrix} \vdots \\ \\ \vdots \\ \mathbf{F}^{c,k}_{FA,i} \\ \vdots \end{pmatrix} \tag{28}$$

where $\mathbf{I}_{3\times 3}$ is the 3-dimensional identity matrix. Obtaining this mapping matrix $\mathbf{P}^{k}_{map}$ for all the cells, the complete focal adhesion forces in the ECM can be expressed in relation to the cell's focal adhesion forces.

$$\mathbf{F}^{e}_{FA} = \sum_{k=1}^{K} \mathbf{F}^{e,k}_{FA} = \sum_{k=1}^{K} \mathbf{P}^{k}_{map} \mathbf{F}^{c,k}_{FA} \tag{29}$$

As previously mentioned, the focal adhesion connections between the cell membrane nodes and ECM nodes can vary over time as the cell membrane deforms, gains traction, and generates lamellipodial protrusions. Therefore, $\mathbf{P}^{k}_{map}$ is updated to reflect the new focal adhesion attachments and detachments at each time step. The original nonlinear computational model has developed a functional relationship between the focal adhesion force, number of integrins, and distance between the membrane and ECM node (see details supplementary materials Appendix A in S1 Text). In the presented framework, the change in focal adhesion attachments can be derived from simulated training of the nonlinear computational model.

Discussion

The collective ECM compaction by multiple cells is predicted through superposition of individual cells' contributions in latent variable space. This is made possible by DF Linearization, latent variable transformation and subsequent superposition of single-cell models to predict the collective behavior among multiple cells.

As shown in Fig 3A, the DF linearization has two-order-of-magnitude higher accuracy than the first-order Taylor expansion, and can approximate the original full scale model with a reasonable root-mean-square error. This representation of nonlinear dynamics is markedly different from standard linearization methods. The DF linearization was also compared to the TPWL method. The figure shows an order-of-magnitude better result for the DF linearization compared to the sophisticated technique. Note that the TPWL does not yield a linear model since the state equation includes a product of two state functions. Therefore, superposition as applied with our DF linearization approach cannot be applied.

As applied to the analysis of multi-cell ECM compaction, linear augmented equations describing single cell-ECM interactions were derived from DF linearization, and then converted to a reduced-order linear representation by transformation onto a basis of eigenvectors derived from simulated data set. Unlike model reduction of nonlinear dynamical systems, which still remains a challenging problem in the field [19–22], the model reduction of a linear system through DF Linearization is straightforward. It allows for the evolution of independent and auxiliary states to be described within a lower dimensional linear manifold. The resulting reduced order latent variable model is capable of reproducing nonlinear dynamics, and the linearized structure of individual models facilitated their integration to describe multi-cell behaviors. The prediction of collective behaviors of a group of cells was achieved by superposing contributions of individual cells represented by latent variables $\mathbf{z}^{c,k}$, which evolves based on their own dynamics in response to the global ECM state represented by latent variable $\mathbf{z}^{e}$.

The linear representation of the collective multi-cell-ECM interactions manifests the two types of feedback actions by the individual cells. As shown in the block diagram in Fig 3A, the individual cells are exposed to the ECM forces represented by latent variable vector $\mathbf{z}^{e}$ in two separate paths. The path through the cell polarity block and matrix **B**, leading to lamellipodia formation, can be viewed as an "active input" as addressed in [5]. This feedback path includes a cell's internal decision as to which direction it extends lamellipodia. In contrast, the other feedback path through a gain matrix **C** does not have a high-level cell decision, but is reactive, playing a "passive role" [5]. These feedback interactions support the prior experimental work [5]. It is interesting to note that ECM compaction begins almost instantaneously, but the magnitude of compaction is rather limited. Once the "active" feedback loop is initiated in, the ECM compacts further, resulting in a large deformation. As the polarity dynamics are rather slow, the second stage ECM compaction does not start immediately. The time scale is determined by the constant κ involved in the polarity dynamics Eq 25. Using the proposed methodologies, we are able to reproduce intercellular mechanical interactions consistent with published experimental observations. In particular, the global compaction of gel volume via collective cell-contractile activities is characteristically different from local deformations of single isolated cells embedded within the same gel. Through study of emergent behaviors of groups of cells embedded in a 3-D ECM fiber network, we can advance our understanding of intercellular mechanical signaling during tissue formation [1–7]. There are a few limitations to our method, however. While the presented method can predict complex nonlinear behaviors, the method is still a type of approximation. Care must be taken with the validity period. In Fig 4C at the sample time of $t = 50$ minutes, the latent variable superposition simulation over predicts the volume shrinkage by 12%.

With the current mathematical formulation, we have not yet incorporated the degradation of ECM fibers through matrix metalloproteinases. ECM degradation would be necessary to reproduce sustained movement and migration of the cells particularly in 3-D embedded matrices [32]. Since ECM degradation continuously changes the fiber connectivity through ECM remodeling, a methodology to update the node grid structure describing the ECM field would need to be developed. However, ECM degradation may not be necessary for predicting gel

compaction since a cluster of cells remains stationary when contracting the surrounding gel [5]. Finally, in the current work, it was assumed that the cell's polarity mechanism is a dominating internal response to mechanical cues. Cells change their internal state through a complex process of mechanotransduction and intracellular signaling. Incorporating these more complex mechanisms is an exciting avenue for future research. While the method has been developed and demonstrated for multi-cellular interactions with 3D ECM, the basic methodology is applicable to a broad range of systems where nonlinear dynamics of many interacting subsystems are prohibitively complex to compute.

Supporting information

S1 Fig. Focal adhesion dynamics on an elastic substrate. Schematic showing integrin molecules on the cellular membrane interacting with an extracellular matrix fiber, and illustrating a stochastic ligand-receptor bonding process at the focal adhesion site.
(PDF)

S2 Fig. Composition of ECM fiber network model. A: Segmented ECM fibers were generated between crosslink nodes. Yellow spheres indicate segmented ECM fiber nodes. B: A magnified view in blue circle mark in A showing examples of three fibers' connectivity with a crosslink node. Blue lines indicated crosslinks between an ECM fiber node and a crosslink node.
(PDF)

S1 Video. Comparison between original nonlinear simulation and latent variable superposition simulation of two-cell interaction embedded within cylindrical ECM. This video depicts the cross-sectional view of the 3-D visualization of simulation of a cylindrical ECM with 2 cells embedded within it. The prediction of decreased cell volume by the latent variable superposition model (blue) agrees well with the ground-truth, full-scale nonlinear simulation results (green). The polarity directions are shown by red arrows. The polarity directions of both cells (initially pointing in arbitrary directions) shift to point inward, indicating that larger stresses are detected in the area between the cells.
(MP4)

S2 Video. Comparison between two-cell latent variable superposition simulation and single cell latent variable simulation. As can be seen from the cross-sectional view of the 3-D visualization of the simulations, the single cell model predicts more localized shrinkage of the ECM volume whereas the two cell model shows more global shrinkage extended to within the region between the cells. This suggests the presence of more than one cell is necessary for the pronounced ECM compaction leading to emergent changes within the ECM.
(MP4)

S3 Video. Two-cell latent variable superposition simulation at varied spacing between 2 cells embedded within cylindrical ECM. This video depicts the cross-sectional view of the 3-D visualization of simulation of a cylindrical ECM with 2 cells embedded within it. As the spacing between cells increases, compaction is less pronounced between them, indicating decreased interaction and integration of cell induced propagated forces.
(MP4)

S4 Video. Multi-cell latent variable superposition simulation depicting comparison of ECM compaction between heterogeneous distributions of cells. This video depicts the cross-sectional view of the 3-D visualization of simulation of a ECM with multiple cells embedded within it. The computational model successfully reproduces the in vitro experiment conducted by Fernandez, et at [5] in which a heterogeneous planar distribution of MC3T3-E1

osteoblasts where plated in 3-D rectangular prism collagen gel. Whereas the group of 5 cells at the left edge exhibit anisotropic contraction of the ECM at the boundary, the isolated cell at the right edge does not contract the gel.
(MP4)

S1 Text. Contains Appendix A: Nonlinear dynamics of cell-ECM interaction for computational model, Appendix B: Least squares estimation for identification of the parameter matrices A, B, C, G involved in the latent space state equations, Appendix C: Implementing polarity model and lamellipodial force generation.
(PDF)

S1 Table. List of simulation parameters.
(PDF)

Acknowledgments

The authors would like to thank Prof. Roger D. Kamm (MIT) and Taher Saif (UIUC) for their biological insights and advice of the studied system.

Author Contributions

Conceptualization: Michaëlle N. Mayalu, H. Harry Asada.

Data curation: Michaëlle N. Mayalu.

Formal analysis: Michaëlle N. Mayalu, H. Harry Asada.

Funding acquisition: H. Harry Asada.

Investigation: Michaëlle N. Mayalu, H. Harry Asada.

Methodology: Michaëlle N. Mayalu, H. Harry Asada.

Project administration: H. Harry Asada.

Resources: Min-Cheol Kim.

Software: Min-Cheol Kim.

Supervision: Min-Cheol Kim, H. Harry Asada.

Validation: Michaëlle N. Mayalu, Min-Cheol Kim.

Visualization: Michaëlle N. Mayalu, Min-Cheol Kim.

Writing – original draft: Michaëlle N. Mayalu, Min-Cheol Kim, H. Harry Asada.

Writing – review & editing: Michaëlle N. Mayalu, Min-Cheol Kim, H. Harry Asada.

References

1. Gjorevski N, Nelson CM. Bidirectional extracellular matrix signaling during tissue morphogenesis. Cytokine & growth factor reviews. 2009; 20(5-6):459–465.
2. Li B, Wang JHC. Fibroblasts and Myofibroblasts in Wound Healing: Force Generation and Measurement. Journal of tissue viability. 2011; 20(4):108–120. https://doi.org/10.1016/j.jtv.2009.11.004 PMID: 19995679
3. Enever PAJ, Shreiber DI, Tranquillo RT. A novel implantable collagen gel assay for fibroblast traction and proliferation during wound healing. The Journal of Surgical Research. 2002; 105(2):160–172. https://doi.org/10.1006/jsre.2002.6372 PMID: 12121703

4. Ma X, Schickel ME, Stevenson M, Sarang-Sieminski A, Gooch K, Ghadiali S, et al. Fibers in the Extracellular Matrix Enable Long-Range Stress Transmission between Cells. Biophysical Journal. 2013; 104(7):1410–1418. https://doi.org/10.1016/j.bpj.2013.02.017 PMID: 23561517
5. Fernandez P, Bausch AR. The compaction of gels by cells: a case of collective mechanical activity. Integrative Biology. 2009; 1(3):252–259. https://doi.org/10.1039/b822897c PMID: 20023736
6. Bell E, Ivarsson B, Merrill C. Production of a tissue-like structure by contraction of collagen lattices by human fibroblasts of different proliferative potential in vitro. Proceedings of the National Academy of Sciences of the United States of America. 1979; 76(3):1274–1278. https://doi.org/10.1073/pnas.76.3.1274 PMID: 286310
7. Stevenson MD, Sieminski AL, McLeod CM, Byfield FJ, Barocas VH, Gooch KJ. Pericellular conditions regulate extent of cell-mediated compaction of collagen gels. Biophysical Journal. 2010; 99(1):19–28. PMID: 20655829
8. Vader D, Kabla A, Weitz D, Mahadevan L. Strain-Induced Alignment in Collagen Gels. PLOS ONE. 2009; 4(6):e5902. https://doi.org/10.1371/journal.pone.0005902 PMID: 19529768
9. Zemel A, Safran SA. Active self-polarization of contractile cells in asymmetrically shaped domains. Physical Review E. 2007; 76(2):021905. https://doi.org/10.1103/PhysRevE.76.021905
10. Rosenthal DT, Iyer H, Escudero S, Bao L, Wu Z, Ventura AC, et al. p38 Promotes Breast Cancer Cell Motility and Metastasis through Regulation of RhoC GTPase, Cytoskeletal Architecture, and a Novel Leading Edge Behavior. Cancer Research. 2011; 71(20):6338–6349. https://doi.org/10.1158/0008-5472.CAN-11-1291 PMID: 21862636
11. Maraldi M, Valero C, Garikipati K. A Computational Study of Stress Fiber-Focal Adhesion Dynamics Governing Cell Contractility. Biophysical Journal. 2014; 106(9):1890–1901. https://doi.org/10.1016/j.bpj.2014.03.027 PMID: 24806921
12. Aghvami M, Barocas VH, Sander EA. Multiscale Mechanical Simulations of Cell Compacted Collagen Gels. Journal of Biomechanical Engineering. 2013; 135(7):0710041–0710049. https://doi.org/10.1115/1.4024460
13. Reinhardt JW, Krakauer DA, Gooch KJ. Complex matrix remodeling and durotaxis can emerge from simple rules for cell-matrix interaction in agent-based models. Journal of biomechanical engineering. 2013; 135(7):071003. https://doi.org/10.1115/1.4024463
14. Reinhardt JW, Gooch KJ. Agent-Based Modeling Traction Force Mediated Compaction of Cell-Populated Collagen Gels Using Physically Realistic Fibril Mechanics. Journal of Biomechanical Engineering. 2014; 136(2):021024. https://doi.org/10.1115/1.4026179 PMID: 24317298
15. Abhilash AS, Baker BM, Trappmann B, Chen CS, Shenoy VB. Remodeling of fibrous extracellular matrices by contractile cells: predictions from discrete fiber network simulations. Biophysical journal. 2014; 107(8):1829–1840. https://doi.org/10.1016/j.bpj.2014.08.029 PMID: 25418164
16. Moon AG, Tranquillo RT. Fibroblast-populated collagen microsphere assay of cell traction force: Part 1. Continuum model. AIChE Journal. 1993; 39(1):163–177. https://doi.org/10.1002/aic.690390116
17. Barocas VH, Moon AG, Tranquillo RT. The fibroblast-populated collagen microsphere assay of cell traction force–Part 2: Measurement of the cell traction parameter. Journal of Biomechanical Engineering. 1995; 117(2):161–170. https://doi.org/10.1115/1.2795998 PMID: 7666653
18. Cover TM. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. Electronic Computers, IEEE Transactions on. 1965;(3):326–334. https://doi.org/10.1109/PGEC.1965.264137
19. Schölkopf B, Smola A, Müller KR. Nonlinear Component Analysis as a Kernel Eigenvalue Problem. Neural Computation. 1998; 10(5):1299–1319. https://doi.org/10.1162/089976698300017467
20. Bartenhagen C, Klein HU, Ruckert C, Jiang X, Dugas M. Comparative study of unsupervised dimension reduction techniques for the visualization of microarray gene expression data. BMC bioinformatics. 2010; 11(1):567. https://doi.org/10.1186/1471-2105-11-567 PMID: 21087509
21. Kerschen G, Golinval Jc, Vakakis AF, Bergman LA. The method of proper orthogonal decomposition for dynamical characterization and order reduction of mechanical systems: an overview. Nonlinear dynamics. 2005; 41(1-3):147–169. https://doi.org/10.1007/s11071-005-2803-2
22. Ilbeigi S, Chelidze D. Persistent model order reduction for complex dynamical systems using smooth orthogonal decomposition. Mechanical Systems and Signal Processing. 2017; 96:125–138. https://doi.org/10.1016/j.ymssp.2017.04.005
23. Kim MC, Whisler J, Silberberg YR, Kamm RD, Asada HH. Cell Invasion Dynamics into a Three Dimensional Extracellular Matrix Fibre Network. PLOS Computational Biology. 2015; 11(10):e1004535. https://doi.org/10.1371/journal.pcbi.1004535 PMID: 26436883

24. Kim MC, Silberberg YR, Abeyaratne R, Kamm RD, Asada HH. Computational modeling of three-dimensional ECM-rigidity sensing to guide directed cell migration. Proceedings of the National Academy of Sciences. 2018; p. 201717230.
25. Borau C, Kamm RD, García-Aznar JM. Mechano-sensing and cell migration: a 3D model approach. Physical Biology. 2011; 8(6):066008. https://doi.org/10.1088/1478-3975/8/6/066008 PMID: 22120116
26. De R, Zemel A, Safran SA. Dynamics of cell orientation. Nature Physics. 2007; 3(9):655. https://doi.org/10.1038/nphys680
27. Proctor JL, Brunton SL, Kutz JN. Dynamic Mode Decomposition with Control. SIAM Journal on Applied Dynamical Systems. 2016; 15(1):142–161. https://doi.org/10.1137/15M1013857
28. Rathinam M, Petzold L. A New Look at Proper Orthogonal Decomposition. SIAM Journal on Numerical Analysis. 2003; 41(5):1893–1925. https://doi.org/10.1137/S0036142901389049
29. Kruger U, Xie L. Partial Least Squares. In: Statistical Monitoring of Complex Multivariate Processes. John Wiley & Sons, Ltd; 2012. p. 375–409.
30. Bond BN, Daniel L. A Piecewise-Linear Moment-Matching Approach to Parameterized Model-Order Reduction for Highly Nonlinear Systems. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. 2007; 26(12):2116–2129. https://doi.org/10.1109/TCAD.2007.907258
31. Cardoso MA, Durlofsky LJ. Linearized reduced-order models for subsurface flow simulation. Journal of Computational Physics. 2010; 229(3):681–700. https://doi.org/10.1016/j.jcp.2009.10.004
32. Lu P, Takai K, Weaver VM, Werb Z. Extracellular Matrix Degradation and Remodeling in Development and Disease. Cold Spring Harbor Perspectives in Biology. 2011; 3(12). https://doi.org/10.1101/cshperspect.a005058 PMID: 21917992