

## REVIEW

# Data- and theory-driven approaches for understanding paths of epithelial–mesenchymal transition

Tian Hong<sup>1</sup> | Jianhua Xing<sup>2,3,4</sup> 

<sup>1</sup>Department of Biochemistry & Cellular and Molecular Biology, The University of Tennessee, Knoxville, Tennessee, USA

<sup>2</sup>Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

<sup>3</sup>UPMC-Hillman Cancer Center, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

<sup>4</sup>Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

## Correspondence

Tian Hong, Department of Biochemistry & Cellular and Molecular Biology, The University of Tennessee, Knoxville TN, USA.

Email: [hongtian@utk.edu](mailto:hongtian@utk.edu)

Jianhua Xing, Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, PA, USA.

Email: [xing1@pitt.edu](mailto:xing1@pitt.edu)

## Funding information

National Science Foundation, Grant/Award Numbers: 2243562, DMS2325149, MCB2205148; National Institute of General Medical Sciences, Grant/Award Numbers: R35GM149531, R01GM148525; National Institute of Diabetes and Digestive and Kidney Diseases, Grant/Award Number: R56DK119232

## Summary

Reversible transitions between epithelial and mesenchymal cell states are a crucial form of epithelial plasticity for development and disease progression. Recent experimental data and mechanistic models showed multiple intermediate epithelial–mesenchymal transition (EMT) states as well as trajectories of EMT underpinned by complex gene regulatory networks. In this review, we summarize recent progress in quantifying EMT and characterizing EMT paths with computational methods and quantitative experiments including omics-level measurements. We provide perspectives on how these studies can help relating fundamental cell biology to physiological and pathological outcomes of EMT.

## KEYWORDS

fate specification process, genomics process, transcription process

## 1 | INTRODUCTION

Epithelial–mesenchymal transition (EMT) is a process in which epithelial cells lose tight cell junctions and gain the ability to migrate. It has been long known that EMT is involved in development, regeneration, and diseases such as fibrosis and cancer (Nieto et al., 2016; Thiery et al., 2009). For example, tumor cells can undergo EMT and become more invasive, which contributes to metastasis (Celià-Terrassa et al., 2018; Krebs et al., 2017). The reversal of EMT, that is, MET, was also proposed to play critical roles in the settlement of circulating tumor cells at metastatic sites (Ocaña et al., 2012). EMT is controlled

by an intricate network of genes whose interactions at transcriptional, post-transcriptional, and post-translational levels enable both the transition and its dynamics. Examples of widely recognized EMT-related genes include those coding for transcription factors such as ZEB and SNAIL families, and effector proteins such as vimentin (Nieto, 2013). In addition, many microRNAs, for example, miR-200 family, were shown to play a role in EMT/MET (Cursors et al., 2018). Nonetheless, the entire regulatory network of EMT may involve several hundred protein-coding genes and regulatory RNAs (Tan et al., 2014).

EMT was first viewed as a binary switch from E to M. In the past decade, however, researchers have been using interdisciplinary

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Authors. *genesis* published by Wiley Periodicals LLC.

approaches to challenge the binary switch model and transforming the concept of EMT into a system with a wide range of cell phenotypes spanning the phenotypic space between E and M states. In particular, several groups used dynamical systems theory and models based on core regulatory network involving a few players (ZEB1, miR-200, etc.) to show that intermediate, partial, or hybrid EMT states can be stable (Hong et al., 2015; Lu et al., 2013; Steinway et al., 2015; Tian et al., 2013). The notion of intermediate EMT states is supported by experimental data from both cell lines and in vivo samples (Grande et al., 2015; Hong et al., 2015; Pastushenko et al., 2018; Zhang et al., 2014). More recently, quantitative, and high-throughput data as well as theoretical development allowed the characterization of the trajectory of EMT involving complex state transitions. In this review, we will highlight recent innovations in data analytical approaches and theories for studying multistate EMT and provide perspectives for future developments in this field.

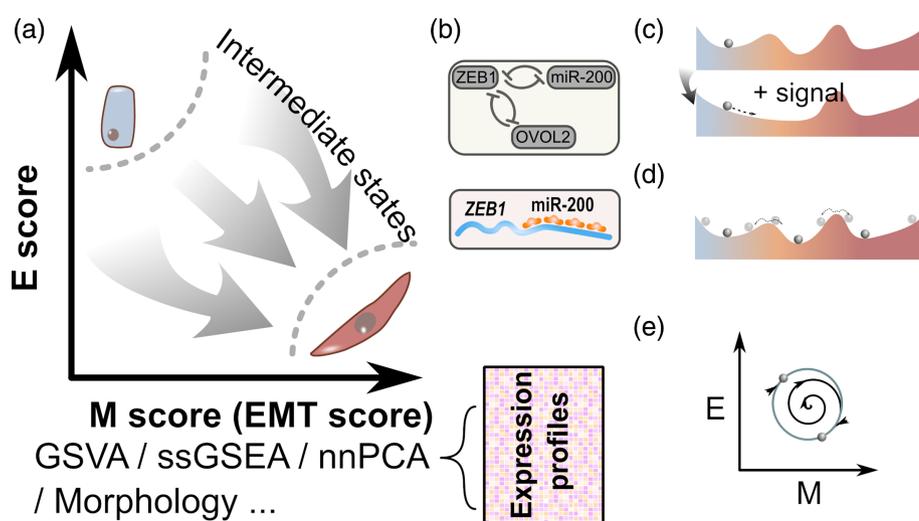
## 2 | SCORING METHODS FOR THE DEGREE OF EMT

Because EMT is now recognized as a continuous spectrum, there is a need for scoring individual samples or cells for assessing its extent of EMT. Expression levels of phenotypic proteins, such as E-cadherin and Vimentin, as well as key transcription factors, have been used to assess EMT. However, the scoring process is challenged by the diversity and complexity of EMT program. As a form of dramatic phenotypic change, the EMT program involves coordinated regulations of many genes, which may generate multiple stable or meta-stable cell states. Therefore, using individual “marker” expression may produce misleading results because the expression of a specific marker may not be significantly altered in a partial EMT process. Tan et al. proposed an EMT scoring approach based on a curated list of EMT signature genes and Kolmogorov–Smirnov (KS) tests (Tan et al., 2014) (Figure 1a). This approach and the associated gene lists have been widely used for assessing EMT progression. For example, Foroutan

et al. found wide score distributions for multiple cancer cell lines with TGF- $\beta$  induced gene signature (Foroutan et al., 2017). As expected, in most cases when epithelial and mesenchymal signatures were separately considered, cancer cells with varying degrees of EMT have inversely correlated E and M scores. The separation of E and M signatures also allows the identification of EMT trajectories in a two-dimensional (2D), E and M space (EM space). For example, using time-course bulk RNA-seq data, Panchy et al. found that TGF- $\beta$  induced EMT involves transient states at the high-E–high-M region in the EM space (Panchy et al., 2020). The high-E–high-M states represent most, but not all, breast cancer tumors' expression profiles that show intermediate EMT phenotypes (i.e., not extreme E- or M-like).

While the KS test-based approaches are useful for assessing degrees and trajectories of EMT, there has been a concern on the suitability of these methods for analyzing single-cell transcriptomes due to the efficiency and the bias toward upregulated genes (Noureen et al., 2022). A method based on principal component analysis (PCA) (non-negative PCA) was proposed to rapidly score EMT progression (Panchy et al., 2021). While this method produced similar results in terms of the EMT progression in the EM 2D space when the leading principal component (PC) was used for gene set (E or M) scoring (Figure 1a), it allows the investigation of gene set scores from more than one PC ordered by the explained variance. Because it has been shown the transcriptional program of EMT is context-dependent and the activation of M genes can be diverse (Cook & Vanderhyden, 2020; Watanabe et al., 2019), the advantage of the multi-PC feature is that it can detect divergent EMT trajectories in a single dataset. For example, in the analysis of a time course, single-cell RNA-seq dataset for small cell lung cancer (SCLC), two PCs from a M gene set produced comparable variances explained, and they showed distinct sub-modules of mesenchymal-like transcriptional programs used by different SCLC subtypes that correspond to intermediate EMT states (Groves et al., 2022).

To simplify the assessment of partial or hybrid EMT phenotypes in datasets, a single score for partial EMT was constructed by taking the minimum of E and M scores (Sacchetti et al., 2021). This was



**FIGURE 1** Illustration of EMT paths and dynamical systems underpinning EMT/MET. (a) An EMT scoring system allows the identification of intermediate EMT states and paths of EMT. (b) Two regulatory networks each capable of producing a tristable EMT system. (c–e) Three plausible modes of EMT/MET: signal-driven disappearance of stable steady state (c), noise-driven transitions among stable steady states (d), and limit-cycle-like orbits (e).

helpful for identifying partial EMT states during colon cancer progression (Sacchetti et al., 2021). A categorial scoring scheme based on E, M, and E/M phenotypes was also proposed, and it was shown to be useful for predicting survival outcomes of cancer patients (George et al., 2017).

In addition to the scoring methods based on transcriptomic profiles, morphological features can be important to determine cell phenotypes in the EMT spectrum. For example, Wang et al. used a scoring method based on live cell images to track trajectories of EMT, revealing crucial temporal dynamics not captured by snapshot data (Wang et al., 2020). In the future, connecting the morphological space to omics' layers will be helpful to gain insights into the characteristics of the intermediate EMT states.

### 3 | DYNAMICAL SYSTEMS THEORY-BASED MECHANISTIC MODEL STUDIES OF PARTIAL EMT

The early notion of binary switch between E and M states resulted in confusions and controversies regarding the role of EMT in metastasis and other pathological processes. This necessitates the rigorous description of EMT as a dynamical system that allows the assignment of cellular states to (stable) steady states representing the long-term phenotypes of cells. This application of theory to biology yielded successful insights that shed light on the complex process of EMT.

Two pioneering models provided the first theoretical frameworks for a stable, hybrid/partial EMT state (Lu et al., 2013; Tian et al., 2013). Both models considered networks containing transcriptional regulations by key EMT factors ZEB1 and SNAIL, as well as post-transcriptional regulations by miR-200 and miR-34. With ordinary differential equation (ODE) models and bifurcation analysis, the authors showed a gradual increase of EMT signals such as TGF- $\beta$  can trigger a two-step, sequential transition from E to M states. The partial EMT state was experimentally observed by Zhang et al. (2014). The regulatory network structures supporting the 3-state EMT systems are centered at the regulation of ZEB1, which is involved in several feedback loops with other factors via mutual-repression or mutual-activation circuits (Figure 1b) (Hong et al., 2015). Describing these regulatory units with nonlinear functions (e.g., Hill functions) in ODE models is sufficient to give rise to three stable states. For example, we consider the ODE system

$$\begin{aligned} dx/dt &= \eta f(y, n_1, K_1) f(z, n_2, K_2) - x \\ dy/dt &= f(x, n_3, K_3) - y \\ dz/dt &= f(x, n_4, K_4) - z, \end{aligned} \quad (1)$$

where  $f(v, n, K) = 1/(1 + (v/K)^n)$ ,  $v$  represents the concentration of a regulator,  $n$  is the Hill exponent, and  $K$  is the repression threshold. The system describes the network shown in the yellow box of Figure 1b, that is,  $x$ ,  $y$ , and  $z$  represent the concentrations of ZEB1, OVOL2, and miR-200, respectively. The model can produce a tristable system with a specific choice of parameter set (Figure 1c).

However, most EMT models used more detailed descriptions for microRNA-mediated regulations. This approach requires the separation of mRNA and protein species in the model, as well as mass-action kinetics as opposed to nonlinear functions with explicitly sigmoidal response curve. For example, the interactions between miR-200 and ZEB1 mRNA harboring four miR-200 binding sites (Figure 1b, pink box) can be described by

$$\begin{aligned} dR/dt &= k_R - R - 4\kappa_{on}Rr + \kappa_{off}C_1 + \beta_1\gamma C_1 \\ dr/dt &= k_r - \gamma r + \sum_{n=1}^4 (-5-n)\kappa_{on}C_{n-1}r + n\kappa_{off}C_n + n\alpha_n C_n \\ dC_n/dt &= (5-n)\kappa_{on}C_{n-1}r - (4-n)\kappa_{on}C_n r - n\kappa_{off}C_n + (n+1)\kappa_{off}C_{n+1} \\ &\quad - \alpha_n C_n - n\beta_n\gamma C_n + (n+1)\beta_{n+1}\gamma C_{n+1}, \quad n \in (1, 2, 3, 4) \end{aligned} \quad (2)$$

where  $R$  is the concentration of free ZEB1 mRNA,  $r$  is the concentration of free microRNA, and  $C_n$  is the total concentration of complexes with microRNA molecules bound to an mRNA molecule except  $C_0 := R$  and  $C_5 := 0$ ,  $k_R$  is the transcription rate constant of ZEB1,  $\kappa_{on}$  and  $\kappa_{off}$  are the binding and unbinding rate constants of mRNA and microRNA,  $\gamma$  is the degradation rate constant of free microRNA, and  $\alpha_n$  and  $\beta_n$  are the degradation rate constants of mRNA and microRNA bound to a complex with respect to their free-form degradation rate constants. The advantage of using this type of model is that one can relate the mechanisms at the level of molecular interactions to the EMT cellular phenotypes. Even more interestingly, this type of model for post-transcriptional regulations can enable 3-state EMT systems even without transcriptional regulations or explicit feedback loops (Nordick, Park, et al., 2022). Equations (1) and (2) are typically combined to describe both transcriptional and post-transcriptional regulations in a single model (Hong et al., 2015; Lu et al., 2013; Tian et al., 2013), and a recent study shows that this modeling approach can yield up to seven stable states in the EMT spectrum (Nordick, Park, et al., 2022), which supports the notion of an EMT continuum in the gene expression space. While the ODE-based models provide useful mechanistic insights into the basis of hybrid EMT states, it is challenging to parameterize and simulate such systems when the number of variables is large. To consider a more comprehensive EMT system, Steinway et al. used Boolean models containing up to 69 nodes, including both transcription factors and signaling proteins, to predict multiple EMT states, including hybrid ones, in liver cancer cell lines (Steinway et al., 2015). The team validated their predictions regarding combinatorically activated nodes with experiments (Steinway et al., 2015). In addition to these modeling approaches for single-cell signaling networks, cellular level, and multiscale (gene and cell levels) models have also been used to characterize dynamics of EMT in terms of phenotypic compositions in populations containing proliferating cells at various EMT states (Sha et al., 2020; Ta et al., 2016).

The connection between hybrid cellular states and high-order multistability (more than two co-existing attractors) provides a theoretical basis for studying the complex EMT system. In this framework, transitions among the EMT states can be driven by signals that trigger the disappearance of certain attractors (Wang et al., 2022) (Figure 1c),

or stochastic gene expression that permits “jumps” among co-existing attractors (Bhatia et al., 2023; Li et al., 2016) (Figure 1d). A recent work showed an additional type of transitions driven by limit cycle oscillations with diverging periods (Nordick, Yu, et al., 2022) (Figure 1e). Interestingly, this dynamical system can be produced by microRNA-mediated regulations (Equation (2)), and it has higher robustness of regenerating diverse gene expression patterns compared with the stochasticity-driven transitions (Nordick, Yu, et al., 2022). Future experiments will be needed to distinguish these different modes of EMT/MET transitions in various biological scenarios.

## 4 | UNRAVEL EMT TRANSITION PATHS THROUGH SNAPSHOT CAPTURE OF CELL STATES

To investigate how EMT proceeds, the first question is how to define and distinguish between an epithelial and a mesenchymal phenotype experimentally. Traditionally, one often monitors a small panel of EMT markers, for example, a decrease of E-cadherin and an increase of Vimentin as the epithelial and mesenchymal markers, respectively (Zhang et al., 2014). Accumulating studies reveal that the EMT program is context-dependent and it generally requires a large panel of markers for the state definition (Yang et al., 2020). For example, Taube et al. identified an EMT signature composed of downregulated genes and 87 upregulated genes from analyzing the expression profiles of a panel of claudin-low and metaplastic breast cancer lines (Taube et al., 2010). Using a small number of EMT markers may mislead interpretation of the experimental results and erroneous conclusions, as exemplified in two studies (Fischer et al., 2015; Zheng et al., 2015). The necessity of using a large panel imposes a technical challenge of using fluorescence-labeling-based live-cell imaging approaches for tracking individual cells over time to monitor the transition process. Advances of fixed single-cell techniques, such as various omics approaches, provide transcriptomic, proteomic, and/or epigenomic profiling of cell states. For these methods cells need to be fixed thus the state of individual cells cannot be tracked over time. Consequently, numerous computational approaches have been developed to infer the transition dynamics and transition paths from the snapshot data, and several studies on EMT have been reported. Below we will briefly discuss a few such studies.

Pseudotime analysis refers to a class of computational methods of trajectory inference. To perform such analyses, one first specifies a starting cell state and an ending cell state, then uses computational algorithms to assign a scalar value as a metric measuring the similarity between a given cell state and the initial state relative to that between it and the final state. Pseudotime is therefore an unfortunate misnomer with no direct relation to time. It corresponds to the reaction coordinate in chemistry, which is a one-dimensional abstract coordinate quantifying the progression of a chemical reaction from a reactant to a product state along a specific reaction route. Another misnomer in the single-cell field is to term such an averaged transition

path as a trajectory, while in physics and chemistry objects follow individual trajectories to complete a transition process. McFaline-Figueroa et al. performed pseudotime analysis on MCF10A scRNA-seq data and concluded that EMT proceeds through a one-dimensional continuum (McFaline-Figueroa et al., 2019).

While pseudotime trajectories are inferred single or multiple average transition paths, several other methods aim to infer individual single-cell transition trajectories to unravel the heterogeneity of transition processes. A Markov state transition model first clusters single cell data into groups, and assumes that unidirectional or bidirectional transitions can take place between neighboring groups in the cell state space. Based on single-cell genomics data, a general method MuTrans uses multiscale reduction of Markov model to identify attractors and reveal multiple transition paths of EMT, as well as transitional cells in-between (Zhou et al., 2021). One uses computational approaches to infer the transition rates from the measured stationary or time series cluster populations. Karacosta et al. analyzed time course mass cytometry data of lung cancer cells using a Markov transition model, and concluded similarly that both EMT and the reverse MET proceed through distinct one-dimensional paths (Karacosta et al., 2019). Optimal transport-based approaches are another class of computational approaches that uses the cell state distributions from time series single cell data to infer transition dynamics between different cell states. Using optimal transport analysis, Cheng et al. examined the TGF- $\beta$  treated MCF10A cells and identified distinct paths leading to three final cell states with different mesenchymal features (Cheng et al., 2023). Using the method of matrix decomposition, Sha et al. studied transition cells along the EMT trajectories based on single-cell data as well as benchmarked the methods using agent-based EMT models (Sha et al., 2020).

All the above approaches use the expression information only and unavoidably impose some approximations to infer the cellular dynamics. Different class of approaches uses additional information for the inference. Specifically, La Manno et al. showed that one can estimate the instant velocity of RNA copy number change termed RNA velocities from the snapshot single-cell sequence data (La Manno et al., 2018). The work inspired an explosive number of studies to improve the estimation accuracy. With all these efforts of improving the estimation of the instant RNA velocities, Qiu et al. developed a theoretical framework, Dynamo, which addresses a separate question of how to extract quantitative mechanistic information from the single cell expression states and instant velocities. The method takes (a) the discrete, sparse, and noisy single-cell gene expression state  $x$  measured in scRNA-seq and (b) the estimated genome-wide RNA turnover dynamics (i.e., RNA velocity  $dx/dt$ ) as input and robustly reconstructs a set of dynamical equations  $dx/dt = F(x)$  through machine learning (Qiu et al., 2022). Contrary to some widespread misconception, the input of Dynamo can be RNA velocities estimated from various approaches, for example, the original splicing-based RNA velocities or those from metabolic labeling data. Dynamo provides a continuous and analytical vector field function  $F$ , which quantitatively describes (generally nonlinear) gene regulation relations in the single-cell transcriptome. That is, it is a generative

model to provide the values of  $F$  at cell states not included in the original data points. Therefore, Dynamo is a framework to systematically construct a data-driven mechanistic model such as the one described in Equation (1), but at a larger (genomewide) scale. With the vector field  $F$ , one can make *in silico* predictions on the effects of gene perturbations. For example, the Jacobian analysis can be viewed as an *in silico* experiment: for a gene  $i$ , the Jacobian reveals how changing the expression of gene  $j$ , while keeping other genes unperturbed, affects the transcription rate of gene  $i$ , with a positive value indicating activation and a negative value for repression. The vector field is analogous to the force field in molecular dynamics simulations, and one can simulate how individual cells evolve over time during processes such as differentiation and reprogramming. Using the approaches, Hu et al. analyzed MCF10A cells treated with low-dose of TGF- $\beta$ , and revealed that EMT proceeds through either G1/S or G2/M cell cycle arrest (Hu et al., 2023).

The work of Qiu et al. set a general framework of analyzing single-cell data under dynamical systems theories, and inspired further development of the detailed algorithms. Bocci et al. provide a SpliceJAC method that approximates the gene regulation as being linear (Bocci et al., 2022). This approximation is valid near a fixed point such as a stable phenotype excluding cell cycle. With SpliceJAC the researchers demonstrated that they could recover gene regulation relations from single-cell data, and identified driver genes that regulate partial and full EMT. Wang et al. also developed a linearized version of the vector field that applies for cell states away from the fixed points (Wang et al., 2021). They first identified one-dimensional transition paths connecting the epithelial and mesenchymal regions in the cell state space, called reaction coordinates in the context of studying chemical reactions, using an algorithm originally developed for studying chemical reactions. They constructed an approximate linear model describing gene regulation relations and binarized the gene expressions into on and off states. The directed regulations are turned on or off by the switch of genes' on-and-off states. That is, the binarized gene states give rise to dynamical gene regulation networks along the transition paths. Additionally, they grouped the genes into different communities for EMT and several other developmental processes and found that at the stable cell type regions there are more regulations between genes within the same community that corresponds to the functionality of the cell type. This modular structure has been proposed to be a general feature of biological networks for minimizing interferences between different cellular processes (Norman et al., 2013; Ravasz et al., 2002). Interestingly, along a transition path the intercommunity regulations first increase upon moving away from the initial cell state region and then decrease upon approaching the final stable cell state region. This result suggests cell phenotypic transitions require transient cooperation between different cellular programs.

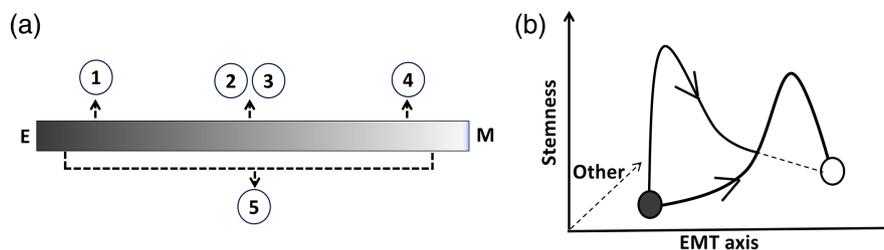
Another group of methods that go beyond the pseudotime analyses use additional information one can extract from the time course data of measured single-cell distributions in the cell state space (Cheng et al., 2023; Schiebinger et al., 2019; Sha et al., 2024; Yeo et al., 2021). Consider a cellular process starts with a swarm of cells

forming a cloud in the cell state space. The cell density distribution evolves over time. Computational approaches have been developed to learn the cell-state transition dynamics of underlying cellular processes including gene-gene regulations and stochastic processes, as well as death and proliferation of cells at specific cell states, from the measured temporal evolution of cell density distributions. Specifically, Sha et al. developed a Trajectory Inference with Growth via Optimal transport and Neural network (TIGON) that infer cell transition dynamics, gene regulation relations, and population growth simultaneously (Sha et al., 2024). They applied to EMT datasets and demonstrated that the method recovers gene regulation relations such as activation of SNAIL1 on mesenchymal genes. Separately, Zhang et al. also developed a graph Dynamo formalism that describes the full dynamics including convection from the vector field, diffusion from stochastic processes, and cell birth-death in a multidimensional cell state space (Zhang et al., 2023). Graph Dynamo also uses both the temporal cell density distributions and RNA velocities to constrain model parameters. The framework also addresses an important technical problem on how to transform the velocity vectors and the equations between different representations, for example, between principal component space and UMAP.

## 5 | UNRAVEL EMT TRANSITION PATHS THROUGH TRACING STATES OF INDIVIDUAL LIVE CELLS OVER TIME

A fundamental limitation of the above approaches is that the dynamical information is inferred from snapshot data under certain model assumptions. Direct experimental measurements of the dynamics of cellular state change require monitoring cells over time. Live-cell imaging with fluorescently labeled proteins or mRNAs allows tracking cell states with molecular specificity but with a limited number of molecular species to define a cell state.

Notice that cellular functions are executed through various cellular machineries, and one expects that a cell state change accompanies with noticeable changes of morphological change of cellular structures such as membranes and organelles. Indeed, morphological profiling with fluorescently labeled and label-free optical imaging has been used to characterize cell states complementing to expression profiling methods (Way et al., 2022). For EMT studies, Burute et al. showed that centrosome position is an indicator of EMT progression (Burute et al., 2017). Montell described a framework of using morphological state space to describe EMT (Montell, 2021). Amack reviewed efforts of *in vivo* live cell imaging on EMT, specifically cytoskeletal reorganization (Amack, 2021). Zhao et al. reviewed efforts of using morphological features as readouts for screening drugs that can reverse EMT (Zhao et al., 2021). Devaraj and Bose (2019) and Mandal et al. (2016) used morphological features to define different EMT states and used Markov transition models to infer transition dynamics between different states from measured cell state distributions. Mandal et al. also characterized distinct types of cell-cell connectivity during EMT (Mandal et al., 2020). These reports demonstrate that cell



**FIGURE 2** Possible mechanistic origin of the conflicting reports on EMT-stemness coupling. (a) Summary of reports on cell states showing stemness couples with different EMT states along the EMT axis. 1: ref. (Celia-Terrassa et al., 2012), 2: ref. (Grosse-Wilde et al., 2015), 3: ref. (Kröger et al., 2019), 4: ref. (Ocaña et al., 2012), 5: ref. (Liu et al., 2014). (b) Schematic illustration that the different reports on the EMT-stemness coupling may come from existence of distinct transition paths due to coupling between EMT and other cellular programs, and the weight of each path may be context-dependent.

morphological features can be used to specify cell states, particularly in EMT studies. Notice that morphological features can be monitored through transmission light imaging without the need of fluorescent labeling, Wang et al. used combined fluorescent- and label-free imaging to monitor individual single cells over time in a multidimensional composite cell feature space (Wang et al., 2020; Wang et al., 2022). The studies reveal two distinct classes of single-cell EMT trajectories, and further studies are needed to examine whether they correspond to the two EMT paths corresponding to G1/S and G2/M arrest from scRNA-seq data analysis (Hu et al., 2023).

## 6 | PERSPECTIVE

It is widely accepted that EMT is a stepwise process with a spectrum of intermediate phenotypes (Nieto et al., 2016; Yang et al., 2020; Zhang et al., 2014). That is, EMT transitions may originate from different epithelial states and end at different mesenchymal/partial mesenchymal states through multiple convergent and divergent paths. EMT also couples to various other cellular processes such as metabolism and stemness, further increasing the number of possible transition paths. Traditionally, one pre-select a panel of markers representing each process for subsequent analyses, and biases may be introduced with different choices of the markers. One example is the two controversial reports about the role of EMT on cancer metastasis (Fischer et al., 2015; Zheng et al., 2015). Later studies suggest that the conclusion was misled by using mesenchymal markers incapable of reflecting partial EMT states, which turn out to contribute to metastasis. With advances of single-cell techniques, now we are able to study how the EMT regulatory program couples to and coordinates with other processes systematically.

Consider the question on EMT and stemness that has attracted extensive attention. An influential study of Weinberg and coworkers (Mani et al., 2008) led to an explosive number of studies on the relationship between EMT and cancer stem cell generation (Scheel & Weinberg, 2012), which is hypothesized to contribute to the complexity of EMT landscape. However, mechanistic understanding on the coupling among the processes remains unclear. While it is well established in the EMT field that EMT proceeds through a continuum

spectrum of intermediate states, conflicting reports exist on where cell stemness couples to the EMT axis, from coupling to the epithelial end (Celia-Terrassa et al., 2012), the hybrid (Grosse-Wilde et al., 2015; Kröger et al., 2019), the mesenchymal end (Mani et al., 2008), or both epithelial and mesenchymal states (Liu et al., 2014) (Figure 2a). One possible explanation is that coupling between EMT and other cellular programs including stemness results in multiple transition paths, and these paths and associated intermediate states may be populated differently at different cellular contexts (Figure 2b). That is, one contribution to the different observations may come from the oversimplification of using a one-dimensional EMT axis to describe the transition progression. Several modeling studies have analyzed small-scale EMT-stemness regulatory network models and concluded that the partial EMT states are the most stem-like (Jolly et al., 2014). This conclusion resonates with the observation of Wang et al. (2021) that the crosstalk between different cellular programs transiently increases during EMT. With single-cell omics data one can perform systematic studies on how various cellular programs work cooperatively to execute large-scale cell expressions and function reprogramming, such as the context-dependent coupling between EMT programs and stemness regulation programs, as exemplified by a few existing such studies (Acosta et al., 2013; Bocci et al., 2021; Hari et al., 2022; Hari et al., 2023; Hu et al., 2023; Watanabe et al., 2019).

Therefore, it is a high priority for EMT studies to have a complete map of the EMT states and their transitions (Yang et al., 2020), or the EMT landscape as it is referred in some literature. Advances of single-cell techniques, including multiomics techniques, spatial genomics, and live cell imaging studies, will provide unprecedented cellular details at both in vivo and in vitro contexts. Analyzing the data within the framework of dynamical systems theories, that is, data-driven systems biology modeling (Xing, 2022), will reveal mechanisms on how the EMT transition is regulated in a context- and microenvironment-dependent manner, and guide effective strategies for modulating the transitions.

## ACKNOWLEDGMENTS

We thank Weikang Wang and Qing Nie for helpful discussions. This work was partially supported by National Institute of General Medical

Sciences (R35GM149531) and National Science Foundation (2243562) to TH, and National Institute of Diabetes and Digestive and Kidney Diseases (R56DK119232), National Institute of General Medical Sciences (R01GM148525), and National Science Foundation (DMS2325149, MCB2205148) to JX.

## DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

## ORCID

Jianhua Xing  <https://orcid.org/0000-0002-3700-8765>

## REFERENCES

- Acosta, J. C., Banito, A., Wuestefeld, T., Georgilis, A., Janich, P., Morton, J. P., Athineos, D., Kang, T. W., Lasitschka, F., Andriulis, M., Pascual, G., Morris, K. J., Khan, S., Jin, H., Dharmalingam, G., Snijders, A. P., Carroll, T., Capper, D., Pritchard, C., ... Gil, J. (2013). A complex secretory program orchestrated by the inflammasome controls paracrine senescence. *Nature Cell Biology*, *15*, 978–990.
- Amack, J. D. (2021). Cellular dynamics of EMT: Lessons from live in vivo imaging of embryonic development. *Cell Communication and Signaling*, *19*, 79.
- Bhatia, S., Gunter, J. H., Burgess, J. T., Adams, M. N., O'Byrne, K., Thompson, E. W., & Duijff, P. H. G. (2023). Stochastic epithelial-mesenchymal transitions diversify non-cancerous lung cell behaviours. *Translational Oncology*, *37*, 101760.
- Bocci, F., Zhou, P., & Nie, Q. (2021). Single-cell RNA-Seq analysis reveals the acquisition of cancer stem cell traits and increase of cell-cell signaling during EMT progression. *Cancers (Basel)*, *13*, 5726.
- Bocci, F., Zhou, P., & Nie, Q. (2022). spliceJAC: Transition genes and state-specific gene regulation from single-cell transcriptome data. *Molecular Systems Biology*, *18*, e11176.
- Burute, M., Prioux, M., Blin, G., Truchet, S., Letort, G., Tseng, Q., Bessy, T., Lowell, S., Young, J., Filhol, O., & Théry, M. (2017). Polarity reversal by centrosome repositioning primes cell scattering during epithelial-to-mesenchymal transition. *Developmental Cell*, *40*, 168–184.
- Celià-Terrassa, T., Bastian, C., Liu, D. D., Ell, B., Aiello, N. M., Wei, Y., Zamalloa, J., Blanco, A. M., Hang, X., Kunisky, D., Li, W., Williams, E. D., Rabitz, H., & Kang, Y. (2018). Hysteresis control of epithelial-mesenchymal transition dynamics conveys a distinct program with enhanced metastatic ability. *Nature Communications*, *9*, 5005.
- Celià-Terrassa, T., Meca-Cortés, O., Mateo, F., Martínez de Paz, A., Rubio, N., Arnal-Estapé, A., Ell, B. J., Bermudo, R., Díaz, A., Guerra-Rebollo, M., Lozano, J. J., Estarás, C., Ulloa, C., Álvarez-Simón, D., Milà, J., Vilella, R., Paciucci, R., Martínez-Balbás, M., de Herreros, A. G., ... Thomson, T. M. (2012). Epithelial-mesenchymal transition can suppress major attributes of human epithelial tumor-initiating cells. *The Journal of Clinical Investigation*, *122*, 1849–1868.
- Cheng, Y.-C., Zhang, Y., Tripathi, S., Harshavardhan, B. V., Jolly, M. K., Schiebinger, G., Levine, H., McDonald, T. O., & Michor, F. (2023). Reconstruction of single cell lineage trajectories and identification of diversity in fates during the epithelial-to-mesenchymal transition. *bioRxiv Preprint*, bioRxiv 2023.09.19.558325.
- Cook, D. P., & Vanderhyden, B. C. (2020). Context specificity of the EMT transcriptional response. *Nature Communications*, *11*, 2142.
- Cursons, J., Pillman, K. A., Scheer, K. G., Gregory, P. A., Foroutan, M., Hedyeh-Zadeh, S., Toubia, J., Crampin, E. J., Goodall, G. J., Bracken, C. P., & Davis, M. J. (2018). Combinatorial targeting by microRNAs co-ordinates post-transcriptional control of EMT. *Cell Systems*, *7*, 77–91.
- Devaraj, V., & Bose, B. (2019). Morphological state transition dynamics in EGF-induced epithelial to mesenchymal transition. *Journal of Clinical Medicine*, *8*, 911.
- Fischer, K. R., Durrans, A., Lee, S., Sheng, J., Li, F., Wong, S. T. C., Choi, H., el Rayes, T., Ryu, S., Troeger, J., Schwabe, R. F., Vahdat, L. T., Altorki, N. K., Mittal, V., & Gao, D. (2015). Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature*, *527*, 472–476.
- Foroutan, M., Cursons, J., Hedyeh-Zadeh, S., Thompson, E. W., & Davis, M. J. (2017). A transcriptional program for detecting TGFβ-induced EMT in cancer. *Molecular Cancer Research*, *15*, 619–631.
- George, J. T., Jolly, M. K., Xu, S., Somarelli, J. A., & Levine, H. (2017). Survival outcomes in cancer patients predicted by a partial EMT gene expression scoring metric. *Cancer Research*, *77*, 6415–6428.
- Grande, M. T., Sánchez-Laorden, B., López-Blau, C., de Frutos, C. A., Boutet, A., Arévalo, M., Rowe, R. G., Weiss, S. J., López-Novoa, J. M., & Nieto, M. A. (2015). Snail1-induced partial epithelial-to-mesenchymal transition drives renal fibrosis in mice and can be targeted to reverse established disease. *Nature Medicine*, *21*, 989–997.
- Grosse-Wilde, A., Fouquier d'Hérouël, A., McIntosh, E., Ertaylan, G., Skupin, A., Kuestner, R. E., del Sol, A., Walters, K. A., & Huang, S. (2015). Stemness of the hybrid epithelial/mesenchymal state in breast cancer and its association with poor survival. *PLoS One*, *10*, e0126522.
- Groves, S. M., Panchy, N., Tyson, D. R., Harris, L. A., Quaranta, V., & Hong, T. (2022). Involvement of epithelial-mesenchymal transition genes in small cell lung cancer phenotypic plasticity. *Cancers*, *15*, 1477.
- Hari, K., Harlapur, P., Saxena, A., Haldar, K., Girish, A., Malpani, T., Levine, H., & Jolly, M. K. (2023). Low dimensionality of phenotypic space as an emergent property of coordinated teams in biological regulatory networks. *bioRxiv Preprint*, bioRxiv 2023.02.03.526930.
- Hari, K., Ullanat, V., Balasubramanian, A., Gopalan, A., & Jolly, M. K. (2022). Landscape of epithelial-mesenchymal plasticity as an emergent property of coordinated teams in regulatory networks. *eLife*, *11*, e76535.
- Hong, T., Watanabe, K., Ta, C. H., Villarreal-Ponce, A., Nie, Q., & Dai, X. (2015). An *Ovol2-Zeb1* mutual inhibitory circuit governs bidirectional and multi-step transition between epithelial and mesenchymal states. *PLoS Computational Biology*, *11*, e1004569.
- Hu, S., et al. (2023). Inferring cell state transition dynamics from single cell data requires proper representation and dynamical information. [to be submitted].
- Jolly, M. K., Huang, B., Lu, M., Mani, S. A., Levine, H., & Ben-Jacob, E. (2014). Towards elucidating the connection between epithelial-mesenchymal transitions and stemness. *Journal of the Royal Society Interface*, *11*, 20140962.
- Karacosta, L. G., Anchang, B., Ignatiadis, N., Kimmey, S. C., Benson, J. A., Shrager, J. B., Tibshirani, R., Bendall, S. C., & Plevritis, S. K. (2019). Mapping lung cancer epithelial-mesenchymal transition states and trajectories with single-cell resolution. *Nature Communications*, *10*, 5587.
- Krebs, A. M., Mitschke, J., Lasierra Losada, M., Schmalhofer, O., Boerries, M., Busch, H., Boettcher, M., Mouggiakakos, D., Reichardt, W., Bronsert, P., Brunton, V. G., Pilarsky, C., Winkler, T. H., Brabletz, S., Stemmler, M. P., & Brabletz, T. (2017). The EMT-activator *Zeb1* is a key factor for cell plasticity and promotes metastasis in pancreatic cancer. *Nature Cell Biology*, *19*, 518–529.
- Kröger, C., Afeyan, A., Mraz, J., Eaton, E. N., Reinhardt, F., Khodor, Y. L., Thiru, P., Bierie, B., Ye, X., Burge, C. B., & Weinberg, R. A. (2019). Acquisition of a hybrid E/M state is essential for tumorigenicity of basal breast cancer cells. *Proceedings of the National Academy of Sciences of the United States of America*, *116*, 7353–7362.
- La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M. E., Lönnerberg, P., Furlan, A., Fan, J., Borm, L. E., Liu, Z., van Bruggen, D., Guo, J., He, X., Barker, R., Sundström, E., Castelo-Branco, G., ... Kharchenko, P. V. (2018). RNA velocity of single cells. *Nature*, *560*, 494–498.

- Li, C., Hong, T., & Nie, Q. (2016). Quantifying the landscape and kinetic paths for epithelial–mesenchymal transition from a core circuit. *Physical Chemistry Chemical Physics*, 18, 17949–17956.
- Liu, S., Cong, Y., Wang, D., Sun, Y., Deng, L., Liu, Y., Martin-Trevino, R., Shang, L., McDermott, S. P., Landis, M. D., Hong, S., Adams, A., D'Angelo, R., Ginestier, C., Charafe-Jauffret, E., Clouthier, S. G., Birnbaum, D., Wong, S. T., Zhan, M., ... Wicha, M. S. (2014). Breast cancer stem cells transition between epithelial and mesenchymal states reflective of their normal counterparts. *Stem Cell Reports*, 2, 78–91.
- Lu, M., Jolly, M. K., Levine, H., Onuchic, J. N., & Ben-Jacob, E. (2013). MicroRNA-based regulation of epithelial-hybrid-mesenchymal fate determination. *Proceedings of the National Academy of Sciences of the United States of America*, 110, 18144–18149.
- Mandal, M., Ghosh, B., Anura, A., Mitra, P., Pathak, T., & Chatterjee, J. (2016). Modeling continuum of epithelial mesenchymal transition plasticity. *Integrative Biology*, 8, 167–176.
- Mandal, M., Ghosh, B., Rajput, M., & Chatterjee, J. (2020). Impact of intercellular connectivity on epithelial mesenchymal transition plasticity. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1867, 118784.
- Mani, S. A., Guo, W., Liao, M. J., Eaton, E. N., Ayyanan, A., Zhou, A. Y., Brooks, M., Reinhard, F., Zhang, C. C., Shipitsin, M., Campbell, L. L., Polyak, K., Brisken, C., Yang, J., & Weinberg, R. A. (2008). The epithelial–mesenchymal transition generates cells with properties of stem cells. *Cell*, 133, 704–715.
- McFaline-Figueroa, J. L., Hill, A. J., Qiu, X., Jackson, D., Shendure, J., & Trapnell, C. (2019). A pooled single-cell genetic screen identifies regulatory checkpoints in the continuum of the epithelial-to-mesenchymal transition. *Nature Genetics*, 51, 1389–1398.
- Montell, D. J. (2021). EMT, one of many morphological transitions in cellular phase space. In K. Campbell & E. Theveneau (Eds.), *The epithelial-to-mesenchymal transition: Methods and protocols* (pp. 13–18). Springer US.
- Nieto, M. A. (2013). Epithelial plasticity: A common theme in embryonic and cancer cells. *Science*, 342, 1234850.
- Nieto, M. A., Huang, R. Y., Jackson, R. A., & Thiery, J. P. (2016). EMT: 2016. *Cell*, 166, 21–45.
- Nordick, B., Park, M. C.-Y., Quaranta, V., & Hong, T. (2022). Cooperative RNA degradation stabilizes intermediate epithelial–mesenchymal states and supports phenotypic continuum. *iScience*, 25, 105224.
- Nordick, B., Yu, P. Y., Liao, G., & Hong, T. (2022). Nonmodular oscillator and switch based on RNA decay drive regeneration of multimodal gene expression. *Nucleic Acids Research*, 50, 3693–3708.
- Norman, T. M., Lord, N. D., Paulsson, J., & Losick, R. (2013). Memory and modularity in cell-fate decision making. *Nature*, 503, 481–486.
- Noureen, N., Ye, Z., Chen, Y., Wang, X., & Zheng, S. (2022). Signature-scoring methods developed for bulk samples are not adequate for cancer single-cell RNA sequencing data. *eLife*, 11, e71994.
- Ocaña, O. H., Córcoles, R., Fabra, Á., Moreno-Bueno, G., Aclouque, H., Vega, S., Barrallo-Gimeno, A., Cano, A., & Nieto, M. A. (2012). Metastatic colonization requires the repression of the epithelial–mesenchymal transition inducer Prrx1. *Cancer Cell*, 22, 709–724.
- Panchy, N., Azeredo-Tseng, C., Luo, M., Randall, N., & Hong, T. (2020). Integrative transcriptomic analysis reveals a multiphasic epithelial–mesenchymal spectrum in cancer and non-tumorigenic cells. *Frontiers in Oncology*, 9, 1479.
- Panchy, N., Watanabe, K., & Hong, T. (2021). Interpretable, scalable, and transferrable functional projection of large-scale transcriptome data using compressed matrix decomposition. *Frontiers in Genetics*, 12, 1555.
- Pastushenko, I., Brisebarre, A., Sifrim, A., Fioramonti, M., Revenco, T., Boumahdi, S., van Keymeulen, A., Brown, D., Moers, V., Lemaire, S., de Clercq, S., Minguijón, E., Balsat, C., Sokolow, Y., Dubois, C., de Cock, F., Scozzaro, S., Sopena, F., Lanas, A., ... Blanpain, C. (2018). Identification of the tumour transition states occurring during EMT. *Nature*, 556, 463–468.
- Qiu, X., Zhang, Y., Martin-Rufino, J. D., Weng, C., Hosseinzadeh, S., Yang, D., Pogson, A. N., Hein, M. Y., Hoi (Joseph) Min, K., Wang, L., Grody, E. I., Shurtleff, M. J., Yuan, R., Xu, S., Ma, Y., Replogle, J. M., Lander, E. S., Darmanis, S., Bahar, I., ... Weissman, J. S. (2022). Mapping transcriptomic vector fields of single cells. *Cell*, 185, 690–711.
- Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., & Barabási, A. L. (2002). Hierarchical organization of modularity in metabolic networks. *Science*, 297, 1551–1555.
- Sacchetti, A., Teeuwssen, M., Verhagen, M., Joosten, R., Xu, T., Stabile, R., van der Steen, B., Watson, M. M., Gusinac, A., Kim, W. K., Ubink, I., van de Werken, H. J. G., Fumagalli, A., Paauwe, M., van Rheenen, J., Sansom, O. J., Kranenburg, O., & Fodde, R. (2021). Phenotypic plasticity underlies local invasion and distant metastasis in colon cancer. *eLife*, 10, e61461.
- Scheel, C., & Weinberg, R. A. (2012). Cancer stem cells and epithelial–mesenchymal transition: Concepts and molecular links. *Seminars in Cancer Biology*, 22, 396–403.
- Schiebinger, G., Shu, J., Tabaka, M., Cleary, B., Subramanian, V., Solomon, A., Gould, J., Liu, S., Lin, S., Berube, P., Lee, L., Chen, J., Brumbaugh, J., Rigollet, P., Hochedlinger, K., Jaenisch, R., Regev, A., & Lander, E. S. (2019). Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176, 928–943.
- Sha, Y., Qiu, Y., Zhou, P., & Nie, Q. (2024). Reconstructing growth and dynamic trajectories from single-cell transcriptomics data. *Nature Machine Intelligence*, 6, 25–39.
- Sha, Y., Wang, S., Zhou, P., & Nie, Q. (2020). Inference and multiscale model of epithelial-to-mesenchymal transition via single-cell transcriptomic data. *Nucleic Acids Research*, 48, 9505–9520.
- Steinway, S. N., Zañudo, J. G. T., Michel, P. J., Feith, D. J., Loughran, T. P., & Albert, R. (2015). Combinatorial interventions inhibit TGFβ-driven epithelial-to-mesenchymal transition and support hybrid cellular phenotypes. *npj Systems Biology and Applications*, 1, 15014.
- Ta, C. H., Nie, Q., & Hong, T. (2016). Controlling stochasticity in epithelial–mesenchymal transition through multiple intermediate cellular states. *Discrete & Continuous Dynamical Systems - Series B*, 21, 2275–2291.
- Tan, T. Z., Miow, Q. H., Miki, Y., Noda, T., Mori, S., Huang, R. Y. J., & Thiery, J. P. (2014). Epithelial–mesenchymal transition spectrum quantification and its efficacy in deciphering survival and drug responses of cancer patients. *EMBO Molecular Medicine*, 6, 1279–1293.
- Taube, J. H., Herschkowitz, J. I., Komurov, K., Zhou, A. Y., Gupta, S., Yang, J., Hartwell, K., Onder, T. T., Gupta, P. B., Evans, K. W., Hollier, B. G., Ram, P. T., Lander, E. S., Rosen, J. M., Weinberg, R. A., & Mani, S. A. (2010). Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. *Proceedings of the National Academy of Sciences*, 107, 15449–15454.
- Thiery, J. P., Aclouque, H., Huang, R. Y. J., & Nieto, M. A. (2009). Epithelial–mesenchymal transitions in development and disease. *Cell*, 139, 871–890.
- Tian, X.-J., Zhang, H., & Xing, J. (2013). Coupled reversible and irreversible bistable switches underlying TGFβ-induced epithelial to mesenchymal transition. *Biophysical Journal*, 105, 1079–1089.
- Wang, W., Douglas, D., Zhang, J., Kumari, S., Enuameh, M. S., Dai, Y., Wallace, C. T., Watkins, S. C., Shu, W., & Xing, J. (2020). Live-cell imaging and analysis reveal cell phenotypic transition dynamics inherently missing in snapshot data. *Science Advances*, 6, eaba9319.
- Wang, W., Poe, D., Ni, K., & Xing, J. (2021). Transiently increased intercommunity regulation characterizes concerted cell phenotypic transition. *bioRxiv Preprint*, bioRxiv 2021.09.21.461257.

- Wang, W., Poe, D., Yang, Y., Hyatt, T., & Xing, J. (2022). Epithelial-to-mesenchymal transition proceeds through directional destabilization of multidimensional attractor. *eLife*, 11, e74866.
- Watanabe, K., Panchy, N., Noguchi, S., Suzuki, H., & Hong, T. (2019). Combinatorial perturbation analysis reveals divergent regulations of mesenchymal genes during epithelial-to-mesenchymal transition. *npj Systems Biology and Applications*, 5, 21.
- Way, G. P., Natoli, T., Adeboye, A., Litichevskiy, L., Yang, A., Lu, X., Caicedo, J. C., Cimini, B. A., Karhohs, K., Logan, D. J., Rohban, M. H., Kost-Alimova, M., Hartland, K., Bornholdt, M., Chandrasekaran, S. N., Haghighi, M., Weisbart, E., Singh, S., Subramanian, A., & Carpenter, A. E. (2022). Morphology and gene expression profiling provide complementary information for mapping cell state. *Cell Systems*, 13, 911–923.
- Xing, J. (2022). Reconstructing data-driven governing equations for cell phenotypic transitions: Integration of data science and systems biology. *Physical Biology*, 19, 061001.
- Yang, J., Antin, P., Bex, G., Blanpain, C., Brabletz, T., Bronner, M., Campbell, K., Cano, A., Casanova, J., Christofori, G., Dedhar, S., Derynck, R., Ford, H. L., Fuxe, J., García de Herreros, A., Goodall, G. J., Hadjantonakis, A. K., Huang, R. Y. J., Kalcheim, C., ... *EMT International Association (EMTIA)*. (2020). Guidelines and definitions for research on epithelial–mesenchymal transition. *Nature Reviews. Molecular Cell Biology*, 21, 341–352.
- Yeo, G. H. T., Saksena, S. D., & Gifford, D. K. (2021). Generative modeling of single-cell time series with PRESCIENT enables prediction of cell trajectories with interventions. *Nature Communications*, 12, 3222.
- Zhang, J., Tian, X. J., Zhang, H., Teng, Y., Li, R., Bai, F., Elankumaran, S., & Xing, J. (2014). TGF- $\beta$ -induced epithelial-to-mesenchymal transition proceeds through stepwise activation of multiple feedback loops. *Science Signaling*, 7, ra91.
- Zhang, Y., Qiu, X., Weissman, J. S., Bahar, I., & Xing, J. (2023). Graphdynamo: Learning stochastic cellular state transition dynamics from single cell data. *bioRxiv Preprint*, bioRxiv 2023.09.24.559170.
- Zhao, N., Powell, R. T., Yuan, X., Bae, G., Roarty, K. P., Stossi, F., Strempl, M., Toneff, M. J., Johnson, H. L., Mani, S. A., Jones, P., Stephan, C. C., & Rosen, J. M. (2021). Morphological screening of mesenchymal mammary tumor organoids to identify drugs that reverse epithelial–mesenchymal transition. *Nature Communications*, 12, 4262.
- Zheng, X., Carstens, J. L., Kim, J., Scheible, M., Kaye, J., Sugimoto, H., Wu, C. C., LeBleu, V. S., & Kalluri, R. (2015). Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. *Nature*, 527, 525–530.
- Zhou, P., Wang, S., Li, T., & Nie, Q. (2021). Dissecting transition cells from single-cell transcriptome data through multiscale stochastic dynamics. *Nature Communications*, 12, 5609.

**How to cite this article:** Hong, T., & Xing, J. (2024). Data- and theory-driven approaches for understanding paths of epithelial–mesenchymal transition. *genesis*, 62(2), e23591. <https://doi.org/10.1002/dvg.23591>