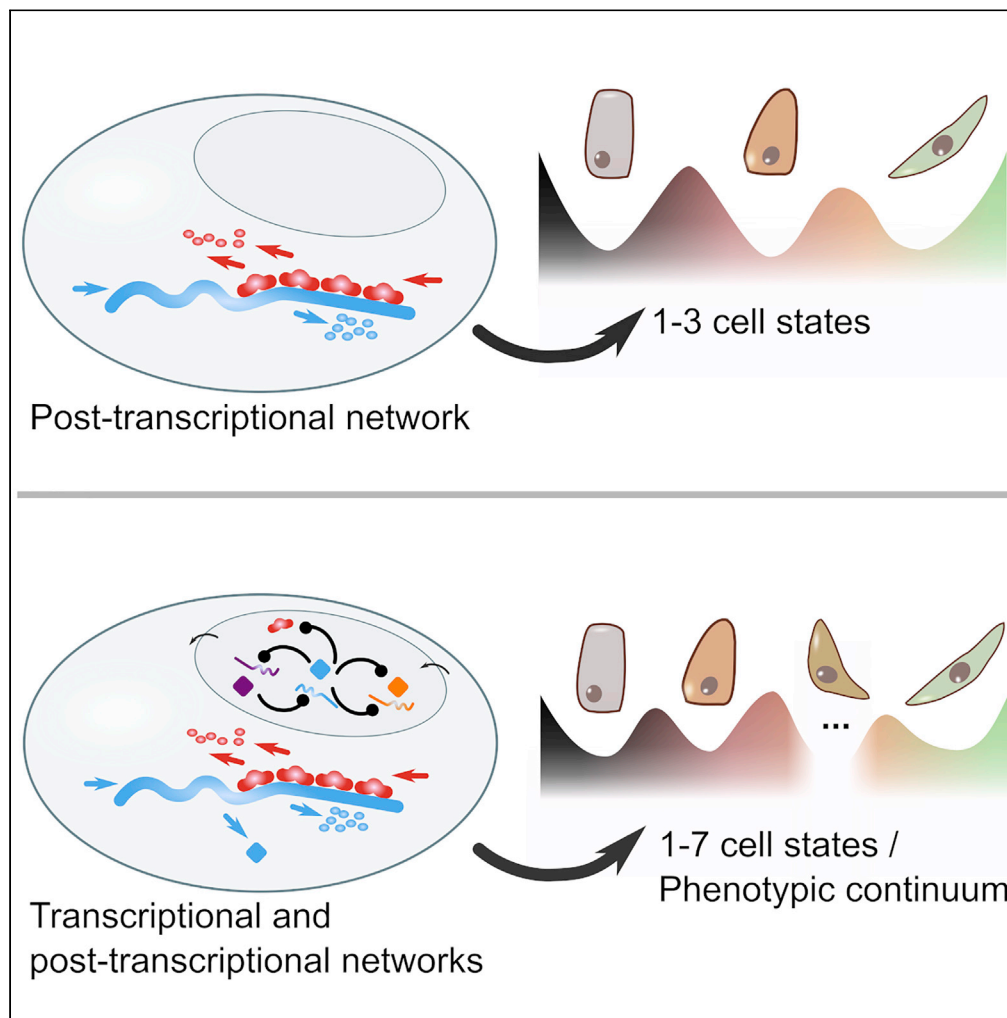


Article

# Cooperative RNA degradation stabilizes intermediate epithelial-mesenchymal states and supports a phenotypic continuum



Benjamin Nordick,  
Mary Chae-Yeon  
Park, Vito  
Quaranta, Tian  
Hong

hongtian@utk.edu

**Highlights**

Cooperative RNA degradation alone can generate three co-existing cell states

Transcription and RNA-decay circuits synergize to govern high-order multistability

RNA-decay circuits support continuum of epithelial-mesenchymal plasticity

Models and single-cell data show microRNA-101's role in controlling multistate EMT



## Article

## Cooperative RNA degradation stabilizes intermediate epithelial-mesenchymal states and supports a phenotypic continuum

Benjamin Nordick,<sup>1</sup> Mary Chae-Yeon Park,<sup>2</sup> Vito Quaranta,<sup>3</sup> and Tian Hong<sup>2,4,5,\*</sup>

## SUMMARY

Multiple intermediate epithelial-mesenchymal transition (EMT) states reflecting hybrid epithelial and mesenchymal phenotypes were observed in physiological and pathological conditions. Previous theoretical models explaining multiple EMT states rely on regulatory loops involving transcriptional feedback, which produce three or four attractors. This is incompatible with the observed continuum-like EMT spectrum. Here, we used mass-action-based models to describe post-transcriptional regulations, finding that cooperative RNA degradation via multiple microRNA binding sites can generate four-attractor systems without transcriptional feedback. Furthermore, the newly identified intermediates-enabling circuits are common in the EMT regulatory network, and they can synergize with transcriptional feedback to support phenotypic continuum. Finally, our model predicted a role of miR-101 in multistate EMT, and we identified evidence from single-cell RNA-sequencing data that support the prediction. Our work reveals a previously unknown role of cooperative RNA degradation and microRNAs in EMT, providing a framework that can bridge the gap between mechanistic models and single-cell experiments.

## INTRODUCTION

Epithelial-mesenchymal transition (EMT) is a cell state change required for embryogenesis, postnatal development, and some diseases' progression including metastasis (Thiery et al., 2009; Watanabe et al., 2014). During EMT, epithelial cells lose their apical-basal polarity and gain the ability to migrate. The transition is not a binary switch: intermediate cellular phenotypes between epithelial (E) and mesenchymal (M) states have been found in development, mammalian cell lines, fibrosis, and tumors (Aiello et al., 2018; Grande et al., 2015; Hong et al., 2015; Pastushenko et al., 2018; Zhang et al., 2014). These intermediate EMT states may possess characteristics of both E and M phenotypes, which can be important for cell- or tissue-level functions such as collective cell migration (Campbell et al., 2019). Recent single-cell transcriptomic studies using individual cell lines provide additional evidence for the existence of multiple intermediate EMT states (Cook and Vanderhyden, 2020, 2022; Deshmukh et al., 2021; Pastushenko et al., 2018). To investigate the mechanistic basis of the intermediate states, theoretical and experimental approaches were used to demonstrate that interconnected transcriptional feedback loops can support intermediate EMT states (Hong et al., 2015; Lu et al., 2013b; Steinway et al., 2015; Tian et al., 2013). These existing mechanistic models can produce only one or two intermediate EMT states with a set of biochemical rate constants representing one cell type. Contrary to these models, it was recently shown that accurate description of EMT dynamics in a single mammary epithelial cell line requires several more intermediate states (Goetz et al., 2020). This finding is consistent with transcriptomic data showing a continuum-like EMT spectrum which possibly contains many stable intermediate cell states (Cook and Vanderhyden, 2020, 2022; Deshmukh et al., 2021; Pastushenko et al., 2018). These observations indicate that an improved theoretical foundation of the intermediate EMT states is needed to bridge the gap between models and experimental data. The improved models will in turn provide new insights into the developmental and pathological processes governed by epithelial-mesenchymal cell plasticity (Aiello et al., 2018; Pastushenko et al., 2018).

It was shown that EMT involves extensive post-transcriptional regulation by microRNA: more than one hundred miRNAs were found to be significantly associated with EMT (Cursons et al., 2018; Hussen et al., 2021). Both epithelial-associated microRNAs and mesenchymal-associated microRNAs have been identified,

<sup>1</sup>School of Genome Science and Technology, The University of Tennessee, Knoxville, TN 37916, USA

<sup>2</sup>Department of Biochemistry & Cellular and Molecular Biology, The University of Tennessee, Knoxville, TN 37916, USA

<sup>3</sup>Department of Biochemistry, Vanderbilt University School of Medicine Basic Sciences, Nashville, TN 37232, USA

<sup>4</sup>National Institute for Mathematical and Biological Synthesis, Knoxville, TN 37916, USA

<sup>5</sup>Lead contact

\*Correspondence:  
hongtian@utk.edu

<https://doi.org/10.1016/j.isci.2022.105224>



inhibiting mesenchymal genes (M genes) and epithelial genes (E genes), respectively. Mechanistically, this microRNA-mediated inhibition often acts through mRNA degradation and translation repression upon microRNA-mRNA binding (Bartel, 2018). It was recently shown that multiple epithelial-associated microRNAs can inhibit in a strongly cooperative manner (Cursons et al., 2018), further supporting the roles of microRNAs in EMT. While existing models describing intermediate EMT states often include microRNA regulations, these models still require multiple transcriptional feedback loops to account for intermediate states (Hong et al., 2015; Lu et al., 2013b; Tian et al., 2013). It is therefore unclear whether intermediate EMT states can arise from simpler, more common gene regulatory networks. Recently, Li et al. used mass-action kinetics, a detailed description of biochemical reactions, to model microRNAs-mRNA reactions, including regulated degradation, with up to three microRNAs binding sites (Li et al., 2021). The models and subsequent experiments showed that these common microRNAs networks can generate two cell states (Li et al., 2021). This study and the extensive involvement of microRNAs in EMT suggest the possibility that reaction networks involving more than three microRNA-binding sites can be a hitherto unknown mechanism for intermediate EMT states.

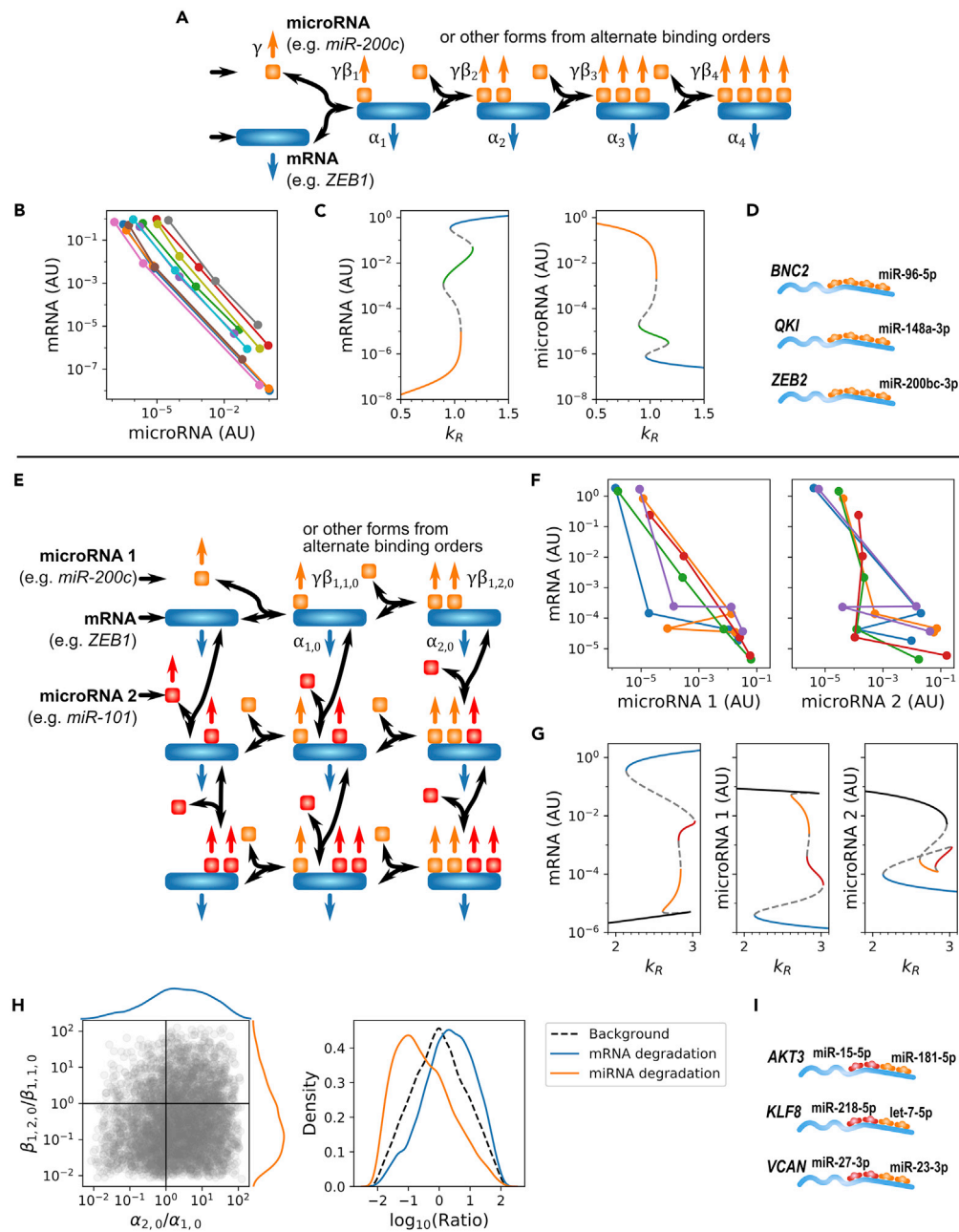
In this work, we use models describing elementary RNA reaction networks to show that cooperative RNA degradation can generate intermediate EMT states in the absence of transcriptional feedback. We use bioinformatic approaches to demonstrate that the structures of gene regulatory networks allowing multiple intermediate states are widespread in the EMT system. Furthermore, we demonstrate that the transcriptional and post-transcriptional mechanisms can be combined to support larger numbers of intermediate EMT states in both modular and emergent manners. Finally, we use a large EMT model to show that cooperative RNA degradation can facilitate the formation of a phenotypic continuum. This model predicts the role of miR-101 in multistate EMT, a prediction we validate with recent single-cell RNA-sequencing data. Our work reveals a previously unknown mechanism for stabilizing intermediate EMT states and provides a new theoretical framework for understanding the perplexing EMT spectra in development and disease progression.

## RESULTS

### Cooperative RNA degradation generates intermediate EMT states in the absence of transcriptional feedback

Based on recent theories and experiments showing post-transcriptional mechanisms for bistability (Li et al., 2021), we hypothesized that intermediate EMT states can arise without transcriptional feedback. To test the hypothesis, we first considered models with mass-action kinetics describing interactions between microRNAs and mRNAs as well as their synthesis and degradation. It was previously proved that with arbitrary positive rate constants, an mRNA with one microRNA-binding site (the MMI1 Model) can have only one stable steady state and that an mRNA with two microRNA-binding sites (the MMI2 Model) can have at most two stable steady states (Li et al., 2021). Systematic search with biologically relevant parameter sets showed that, like the MMI2 Model, an mRNA with three microRNA-binding sites (the MMI3 Model) can have at most two stable steady states (Li et al., 2021).

We therefore built a model containing an mRNA with four microRNA-binding sites: the MMI4 Model. In the first version of the MMI4 Model, the four microRNA-binding sites are bound by one type of microRNA. The two RNAs can form four types of complexes through complementarity-based binding, i.e. 1:  $n$  complex where  $n \in (1, 2, 3, 4)$  represents the number of microRNA molecules in each complex (Figure 1A, Table S1). The two RNAs are allowed to be degraded independently with distinct rate constants in the complexes. Details of all models can be found in the supplemental information. We randomly sampled parameter values within biologically plausible ranges (see STAR methods). Out of  $10^7$  sampled parameter sets, we found 1,732 that generated three stable steady states (i.e. three attractors, i.e. tristability, Figures 1B and 1C) and none that generated four attractors (tetrastability). The third attractor at which the concentration of an EMT marker gene is neither minimal (E state) nor maximal (M state) for the system represents an intermediate state. The tristable parameter sets, i.e. those permitting an intermediate state, are generally robust to changes of at least 10% in any single relative RNA degradation rate (Figure S1). Their degradation rate constants, while individually plausible, are however collectively unusual. For example, the most extreme degradation rates are often found in unsaturated complexes (Table S2), which would require the mRNA to be strongly stabilized by the binding of additional microRNA molecules. While the structure of the model is applicable to EMT (all 7 matching target genes shown in Figure 1D and Figure S6A), the



**Figure 1. Multistability in the MMI4 Models**

(A) Reactions in the one-microRNA MMI4 Model, which has four binding sites for the same microRNA on one mRNA. Wide rectangles, mRNA; squares, microRNA; horizontal arrows, transcription; colored arrows, RNA degradation; curved arrows, binding/unbinding.

(B) Example tristable systems from the one-microRNA MMI4 Model. Points represent attractors in the space of free mRNA vs. free microRNA concentration. Attractors of the same system/parameterization are joined by lines of the same color. AU, arbitrary units.

(C) Bifurcation diagrams showing the steady states of free mRNA (left) and free microRNA (right) as a function of the mRNA transcription rate  $k_R$  from the brown system in B. Each steady state is colored the same in both plots. Dashed lines, unstable steady states. Parameter values in Table S2.

(D) Other EMT-related examples of the one-microRNA MMI4 Model.

(E) Reactions in the two-microRNA MMI4 Model, which has two binding sites for each of two microRNAs on one mRNA.

**Figure 1. Continued**

(F) Example tetrastable systems from the two-microRNA MMI4 Model in the space of free mRNA vs. free microRNA 1 (left) or free microRNA 2 (right). Order of attractors connected by each line is the same in both subplots.

(G) Bifurcation diagrams showing the steady states of free mRNA (left), free microRNA 1 (middle), and free microRNA 2 (right) as a function of the mRNA transcription rate from the green system in F. Parameter values in [Table S4](#).

(H) Left: Scatterplot of functional cooperativities in mRNA degradation ( $\alpha_{2,0}/\alpha_{1,0}$ ) and microRNA 1 degradation ( $\beta_{1,2,0}/\beta_{1,1,0}$ ) rates due to second microRNA 1 binding in 5,000 3- or 4-attractor systems. Marginal distributions are Gaussian kernel density estimates. Right: Comparison of multistable cooperativity distributions for both microRNAs to distribution of 50,000 randomly sampled parameter sets.

(I) Other examples of the two-microRNA MMI4 Model in EMT genes.

See also [Tables S1–S4](#) and [Figures S1–S6](#).

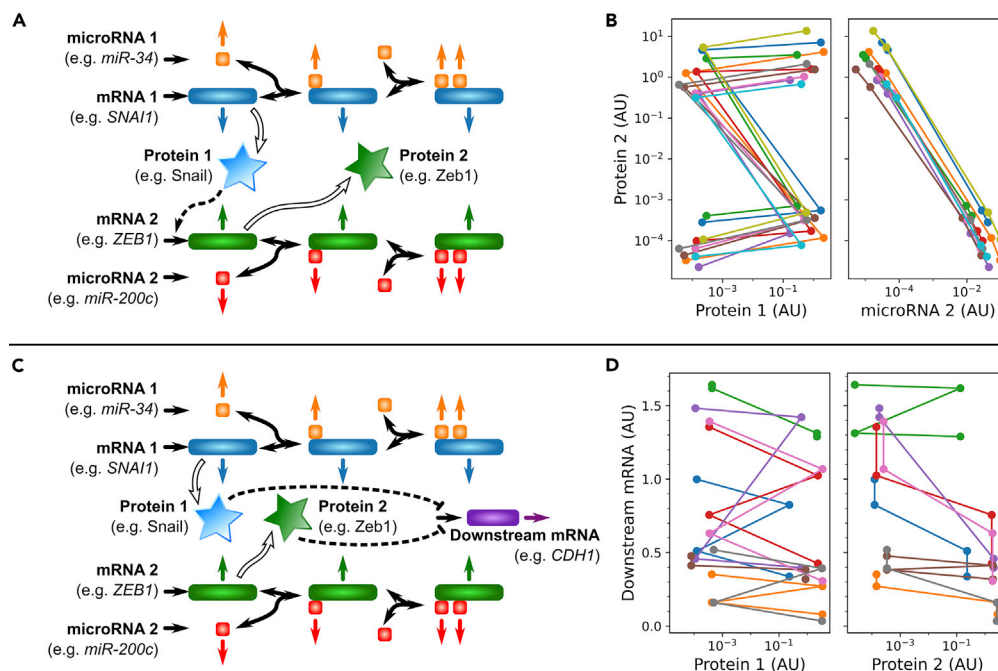
dependence on unusual parameter progression makes the single-microRNA MMI4 Model unlikely to drive multistability *in vivo*.

We considered that allowing the four binding sites to be bound by different microRNAs allows the model to represent more post-transcriptional EMT circuits with potentially more realistic parameter sets. We therefore considered a modified version of the MMI4 Model which describes two microRNAs each with two binding sites on the mRNA ([Figure 1D](#) and [Table S3](#)). This structure indeed allowed the MMI4 Model to produce three- and four-attractor systems with more biologically plausible parameter progressions ([Table S4](#)) and low dependence on any individual degradation rate ([Figure S3](#)). These systems contain intermediate EMT states that may be related to the widely observed phenotypes ([Figures 1F](#) and [1G](#)). To test the importance of functional cooperativity in stabilizing intermediate states, we randomly sampled parameter sets until 5,000 systems with at least three attractors had been obtained. Almost all, 4,973 of the 5,000, involved cooperative degradation of RNAs. The cooperativity is generated by the enhanced degradation rate constant of the mRNA in 1:2 complex with a microRNA compared to the 1:1 complex or by reduced degradation rate constant of the microRNA in 1:2 complex compared to the 1:1 complex ([Figures 1H](#) and [S5](#)). Cooperative mRNA degradation is observed experimentally ([Grimson et al., 2007](#)); functionally cooperative microRNA stabilization might arise through steric blocking of microRNA-degrading factors ([Kai and Pasquinelli, 2010](#)). Three instances of the two-microRNA MMI4 Model are shown in [Figure 1I](#); all 45 EMT genes matching the model are shown in [Figure S6B](#).

EMT is regulated by a large gene regulatory network, in which many genes are targeted by multiple microRNAs ([Hong et al., 2015](#); [Huang et al., 2017](#); [Lu et al., 2013b](#); [Tian et al., 2013](#)). Furthermore, the MMI2 Model can generate bistable systems ([Li et al., 2021](#)). We therefore asked whether connecting two MMI2 modules (the Chained-MMI2 Model, [Figure 2A](#) and [Table S5](#)) can enable intermediate EMT states. While this single connection can represent an EMT transcription factor regulating another, this network still does not contain any transcriptional feedback. We found that the Chained-MMI2 Model was able to generate intermediate EMT states in terms of the target gene expression ([Figure 2B](#)). Considering only direct regulations between EMT genes found in the TRRUST2 or OmniPath databases, we found 8 instances of the Chained-MMI2 Model in the EMT network (Methods, [Figure S9A](#)). If indirect regulations of up to 5 steps (i.e. 4 intermediate genes) are allowed, up to 171 combinations of EMT genes match the model.

We next considered another scenario where one EMT gene is regulated by two EMT transcription factors ([Batlle et al., 2000](#); [Vannier et al., 2013](#)), each involved in an MMI2 module (the Co-targeting-MMI2 Model, [Figure 2C](#) and [Table S7](#)). With this model, intermediate EMT states were observed in terms of the expression of the final target EMT gene ([Figure 2D](#)), arising in a combinatorial manner from the multiple upstream bistable systems. The 18 direct-regulation instances of this model in the EMT network are shown in [Figure S9B](#). If allowing indirect regulations of up to 5 steps, up to 1,312 sets of EMT genes match the Co-targeting-MMI2 Model.

In summary, we showed that a model with a total of four microRNA-binding sites on either one or two EMT genes can generate intermediate EMT states. In each of the four versions of the model, cooperative RNA degradation stabilizes these intermediate states. Intermediates-enabling topologies are common in the EMT system. In fact, 55 of 423 classically defined EMT genes are involved in direct-regulation versions of these topologies, so many seemingly unidirectional interactions between well-known EMT genes could potentially contribute to the multistate dynamics associated with feedback (see [STAR methods](#)).



**Figure 2. Tetrastability from transcriptionally connected MMI2 targets**

(A) Reactions in the Chained-MMI2 Model, in which the protein product of one MMI2 target gene transcriptionally regulates another MMI2 target gene. Either microRNA site on each mRNA is allowed to be bound first. Hollow arrows, translation; dashed pointed arrow, transcriptional activation; stars, proteins.

(B) Example tetrastable systems from the Chained-MMI2 Model in the space of Protein 2 vs. Protein 1 (left) or free microRNA 2 (right). Each system has four different expression levels of Protein 2, which are monotonically anticorrelated to free microRNA 2 levels.

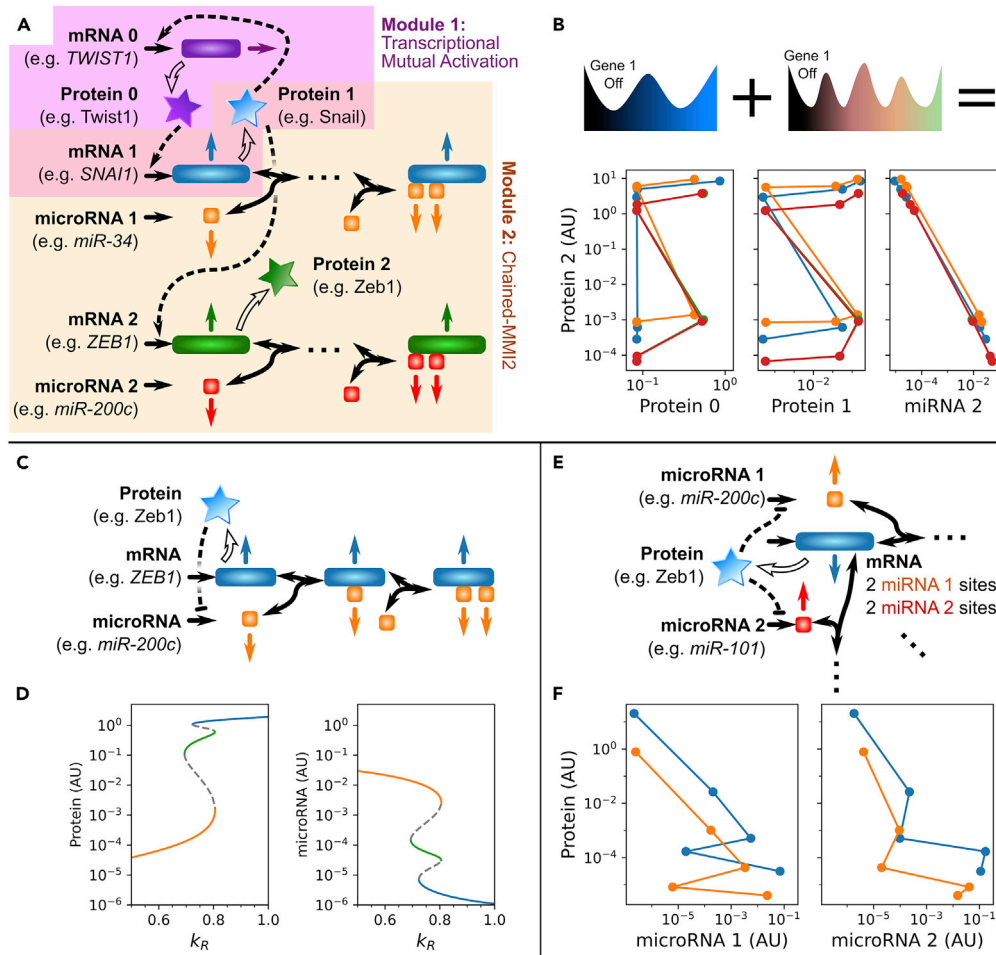
(C) Reactions in the Co-targeting-MMI2 Model, in which two MMI2 target genes encode proteins that both transcriptionally regulate a third downstream gene. Either microRNA site on each mRNA is allowed to be bound first. Dashed blunt arrow, transcriptional repression.

(D) Example tetrastable systems from the Co-targeting-MMI2 Model in the space of the downstream mRNA expression vs. Protein 1 (left) or Protein 2 (right). Each system has four different expression levels of the downstream gene.

See also [Tables S5–S8](#) and [Figures S7–S9](#).

### Modular and emergent synergies between post-transcriptional and transcriptional networks in generating EMT states

It is well known that transcriptional positive feedback loops can generate bistability (Gardner et al., 2000). This type of feedback loop is common in the core EMT regulatory network (Hong et al., 2015). Furthermore, hybrid transcriptional and post-transcriptional mechanisms can generate tristability according to previous models (Lu et al., 2013a, 2013b). We next asked whether combining a module consisting of a transcriptional feedback loop (transcriptional module) with a post-transcriptionally driven, intermediate-enabling module (post-transcriptional module) can generate even more intermediate EMT states (Jiménez et al., 2017). To test this, we first selected a bistability-enabling parameter set for a representative transcriptional module containing two mutually activating transcription factors (Figure 3A Module 1, Table S9). We next selected a parameter set for a tetrastability-enabling post-transcriptional module, the Chained-MMI2 Model (Figures 2A and 3A Module 2, Table S5). Without altering the values of parameters unique to each module, we considered new values for the parameters that conflict, namely the transcription and translation rates of the upstream gene in the Chained-MMI2 Model. We found that there existed values between each pair of original models' values that allowed the addition of at least one intermediate state to the existing states generated by the post-transcriptional module (Figure 3B). Even in 6-attractor systems, only the dependence of mRNA 0 transcription on Protein 1 is a very sensitive parameter for multistability (Figure S10). It is also remarkable that the addition of intermediate states was achieved without altering the biochemical rate constants within each module, suggesting feasibility of this phenotypic change through evolution.



**Figure 3. Synergies between transcriptional and post-transcriptional multistability**

(A) The model resulting from combining a transcriptional mutual activation module (genes 0 and 1, purple shading) with a Chained-MMI2 module (genes 1 and 2, orange shading).

(B) Example five- and six-attractor systems produced by combining bistable transcriptional mutual activation parameter sets with tetrastable Chained-MMI2 parameter sets. All concentrations are in arbitrary units.

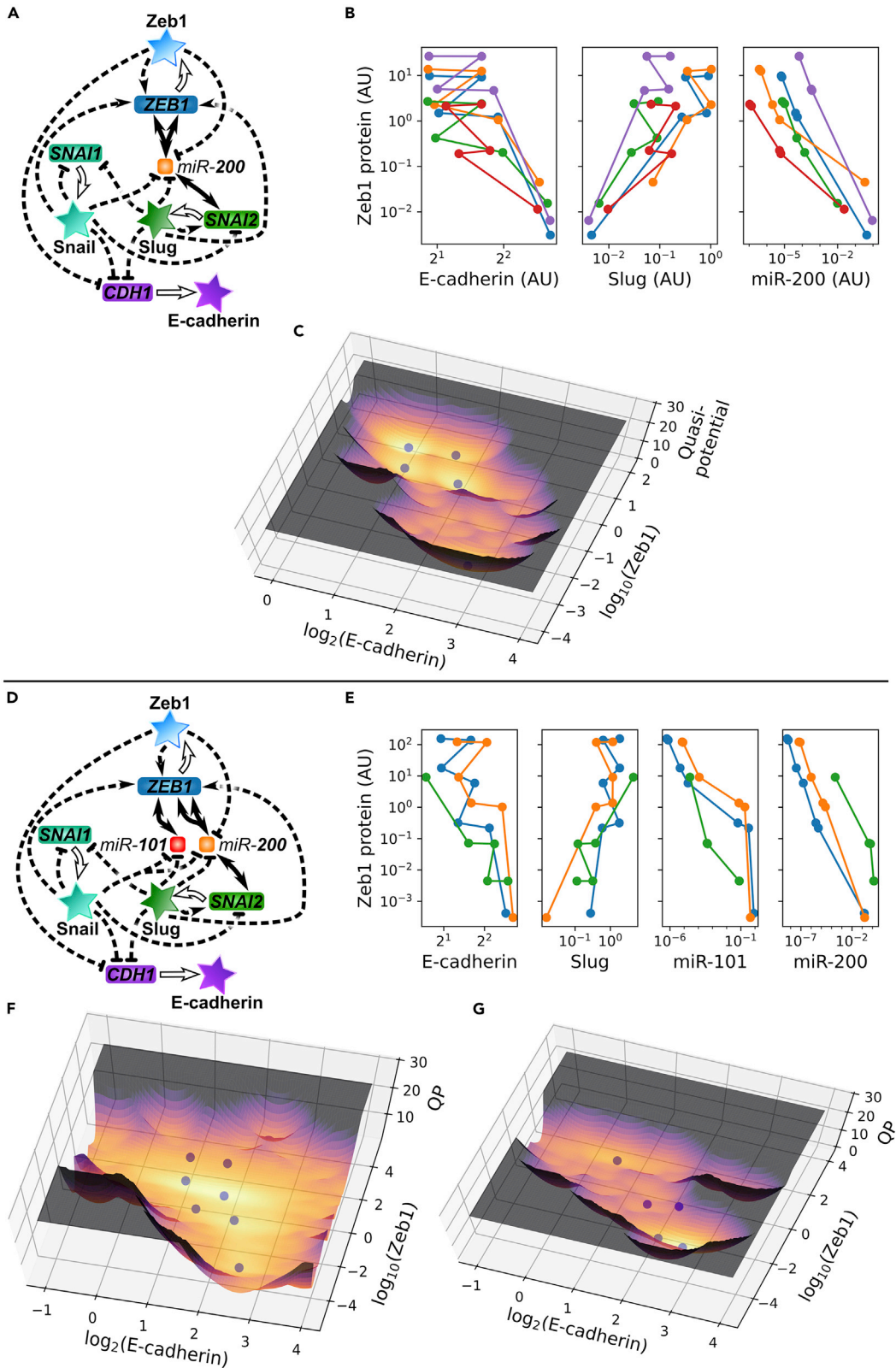
(C) The model resulting from adding transcriptional repression of the microRNA to the MMI2 Model.

(D) Bifurcation diagrams showing the steady states of protein (left) and free microRNA (right) levels with respect to the mRNA transcription rate. Tristability can emerge from the addition of the transcriptional repression to MMI2.

(E) The addition of transcriptional repressions to the two-microRNA MMI4 Model. The mRNA-microRNA complexes are hidden for compactness.

(F) Example five-attractor systems emerging from the addition of transcriptional repressions to the two-microRNA MMI4 Model. See also [Tables S9–S15](#) and [Figures S10–S13](#).

Modular combination of transcriptional and post-transcriptional mechanisms for generating multistable systems is applicable to EMT, but there might also exist “emergent” synergy between transcriptional and post-transcriptional regulations: a transcriptional module may not be multistable by itself, but when it is combined with a multistable, post-transcriptional module, additional attractor(s) can arise. Indeed, we obtained an intermediate state by extending the bistability-enabling MMI2 module with a single transcriptional repression of the microRNA by the gene product ([Figure 3D](#)). Mechanistically, the emergence of this intermediate cell state was due to the emergent feedback loop between the microRNA and mRNA, consisting of both transcriptional and post-transcriptional regulations ([Figure 3C](#) and [Table S12](#)). It was previously shown that this type of hybrid feedback system is common in biology ([Minchington et al., 2020](#)). Importantly, the well-known Zeb1-miR200 feedback loop contains this network topology ([Bracken et al., 2008](#)). Our results indicate that the previously known tristability-enabling structure of the





**Figure 4. Combining transcriptional and post-transcriptional regulation produces many attractors and a continuum in epithelial-mesenchymal space**

(A) Structure of the Subbalakshmi et al. model extended with two explicitly modeled *miR-200* binding sites on *ZEB1*. mRNA-microRNA complexes are hidden for compactness.  
 (B) Example 5-attractor systems from the model in A shown in the gene expression space of Zeb1 protein vs. E-cadherin protein, Slug protein, or free *miR-200*.  
 (C) Quasi-potential diagram showing the stochastic gene expression landscape of the blue system in B under multiplicative noise in the space of E-cadherin protein vs. Zeb1 protein. Deeper regions and brighter color correspond to more likely gene expression states. Spheres, deterministic attractors.  
 (D) Structure of the Subbalakshmi et al. model further extended with *miR-101* targeting one site on *ZEB1*, transcriptionally repressed by Slug and Snail.  
 (E) 7-attractor (blue), 6-attractor (orange), and 5-attractor (green) systems from the model in D. All concentrations are in arbitrary units.  
 (F) Quasi-potential (QP) diagram of the 7-attractor blue system in E.  
 (G) Quasi-potential diagram of the 5-attractor green system in E. Not every parameter set of the extended two-microRNA model exhibits the broad distribution found in F.

See also [Tables S16–S19](#), [Figures S14](#), and [S15](#).

Zeb1-miR200 feedback loop and Zeb1 self-regulation is not the minimal topology for generating an EMT intermediate state ([Lu et al., 2013b](#); [Tian et al., 2013](#)). Likewise, we found that adding transcriptional repression to the two-microRNA MMI4 Model permits the appearance of a fifth attractor ([Figures 3E, 3F](#), and [Tables S14](#), and [S15](#)).

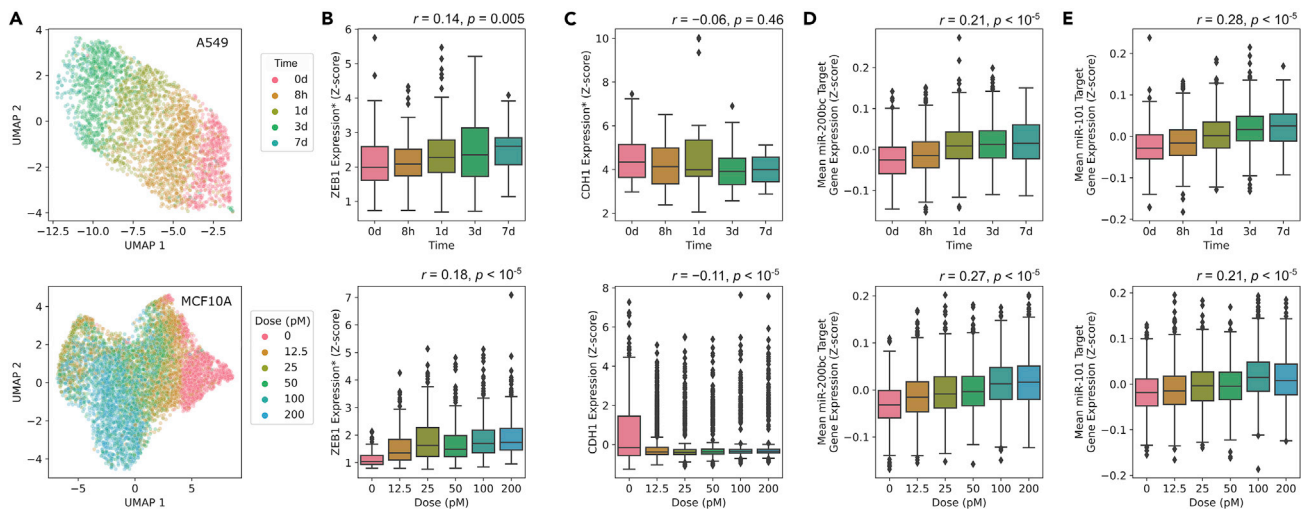
In summary, our results suggest that the post-transcriptional and transcriptional networks commonly observed in the EMT system can be combined in both modular and emergent fashions to generate additional intermediate phenotypes.

**A multiple-feedback-mechanism EMT network generates a 7-state EMT continuum**

How does the RNA degradation-based mechanism for multistability contribute to the formation of intermediate EMT states in networks larger than simple motifs? To address this question, we first selected an EMT model published recently: Subbalakshmi et al. showed that a network containing three transcription factors and thirteen transcriptional regulations can generate a three- or four-attractor EMT system with biologically meaningful parameters ([Subbalakshmi et al., 2022](#)). While there are numerous existing EMT models for tetrastable systems, we chose the Subbalakshmi et al. model because only one microRNA was considered in the model, so it serves as a good basal model for us to test the effect of cooperative RNA degradation by adding more biologically relevant microRNA-binding sites sequentially to the system.

As the base model contains inhibition of *ZEB1* and *SNAI2* by *miR-200* without explicit modeling of multiple complexes, we added an additional binding site on *ZEB1* to the model ([Figure 4A](#) and [Table S16](#)) and selected a parameter set representing cooperative RNA degradation. As expected, the model with two binding sites of *miR-200* produced a system with five attractors ([Figures 4B](#) and [4C](#)), i.e. an additional intermediate EMT state compared to the previously published results. We next considered another microRNA, *miR-101*, that regulates *ZEB1* ([Guo et al., 2014](#)) and is repressed by slug and snail ([Huang et al., 2017](#); [Zheng et al., 2015](#)) ([Figure 4D](#) and [Table S18](#)). Note that these additional regulations do not introduce any additional transcriptional feedback loops. The inclusion of this microRNA allowed yet two additional intermediate EMT states, resulting in a seven-attractor EMT system ([Figure 4E](#) blue).

To test the effect of cooperative RNA degradation with transcriptional noise, we performed stochastic simulations of the one-microRNA Subbalakshmi et al. network ([Figures 4A–4C](#)) and the extended model with two microRNAs ([Figures 4D–4G](#)), applying the same level of multiplicative noise to all RNAs, complexes, and proteins. Based on the steady-state distributions of molecules from at least 480 simulations each representing a cell from a population, we constructed the quasi-steady-state landscape to visualize the multi-attractor systems under the influence of noise. As expected, cooperative RNA degradation via multiple binding sites gave rise to gene expression states near the additional attractors ([Figure 4F](#) yellow). Notably, it also resulted in broader distributions of gene expression further from the attractors ([Figure 4F](#) orange, compare to [Figure 4C](#)). That is, a deterministic attractor—which may be highly sensitive to several aspects of Zeb1 regulation ([Figure S15](#))—is not required for a region of gene expression space to be populated in the presence of noise. While the wider distribution of epithelial and mesenchymal marker genes is not simply an artifact of simulating the more complex model ([Figure 4G](#)), the new complexes involving the additional microRNA do generally create more axes in concentration space. Fluctuations in gene expression space can then be amplified in impact by functional cooperativity in degradation rates, pushing the system toward a different state.



**Figure 5. EMT continuums and progression of microRNA activities validated by scRNA-seq data**

Two scRNA-seq datasets (GSE147405 and GSE213753) for A549 and MCF10A respectively were visualized with UMAP and expression profiles of individual genes.

(A) UMAP plots show distributions of 3568 A549 cells and 11,220 MCF10A cells. Colors indicate treatment conditions.

(B and C) Boxplots show medians and interquartile ranges of *ZEB1* and *CDH1* expression under each condition. \* indicates the exclusion of cells with no expression. *r* denotes the Pearson correlation coefficient for quantities on both axes.

(D and E) Boxplots show medians and interquartile ranges of miR-200c and miR-101 activities inferred from their target gene expressions (see STAR methods).

Our new model (Figure 4D) proposed a role of miR-101 and its stepwise changes of expression level in EMT progression. To test this prediction, we next analyzed two single-cell RNA-sequencing (scRNA-seq) datasets from TGF- $\beta$ -driven EMT. Transcriptome data for 3,568 A549 cells from a time course experiment (Cook and Vanderhyden, 2020) and for 11,220 MCF10A cells from a dose-dependent, nearly steady-state experiment (Panchy et al., 2022), were visualized using Uniform Manifold Approximation and Projection (UMAP). Consistent with earlier analyses (Cook and Vanderhyden, 2020; Panchy et al., 2022), cells were distributed in continuous space for each dataset (Figure 5A), suggesting an EMT continuum. Furthermore, median expression of *ZEB1* positively correlated with the progression of EMT, whereas that of *CDH1* had the opposite trend (Figures 5B and 5C) (the negative correlation of *CDH1* in A549 cells was not significant due to low detection rate). Interestingly, the mean target gene expressions for both miR-200b/c and miR-101 were positively correlated with the progression of EMT, suggesting stepwise changes of their expression levels (Figures 5D and 5E). These effects were not due to the time/dose-dependent changes of overall gene expression patterns (see STAR methods), and they are consistent with the predictions from our model (Figure 4E).

In summary, in a larger EMT network containing both transcriptional and post-transcriptional regulations, cooperative RNA degradation via multiple microRNA-binding sites gave rise to additional attractors and a broader distribution of gene expression, reflecting the EMT continuum observed in recent single-cell transcriptome data.

## DISCUSSION

Intermediate, or hybrid, EMT phenotypes have been widely observed in several biological contexts. In this work, we have used mathematical models to demonstrate a new mechanism for generating intermediate EMT states based on first principles of gene regulation. This finding can serve as a step toward the reconciliation of the observed EMT continuum with transcriptomic studies and the three or four discrete EMT states captured by previous models. While the observed EMT continuum may be alternatively explained by large subpopulations of cells *en route* to different attractors, recent work using cell-state transition models to explain experimental data showed that models with only a few states cannot describe the time course EMT data accurately (Goetz et al., 2020). The newly identified post-transcriptional mechanism for generating intermediate states provides a foundation for the additional EMT states necessary to explain expression data at the gene regulation level. A recent study showed that similar post-transcriptional

reactions can generate oscillations on slow timescales in addition to multistability (Nordick et al., 2022), which suggests another possibility that the EMT continuum may be supported by a combination of point attractors and cyclic attractors.

Cooperativity of microRNA-binding sites has been widely observed (Buck et al., 2014; Cursons et al., 2018; Lai et al., 2019). In particular, Cursons et al. showed high cooperativity between multiple microRNAs in controlling EMT (Cursons et al., 2018). In this work, we modeled the cooperativity of microRNA-binding sites in the form of synergistic mRNA degradation, which was also experimentally observed, and synergistic microRNA stabilization. There are other forms of biologically plausible synergy between binding sites, such as cooperative binding affinities, cooperative translational inhibition, and competition for one microRNA's binding to 3'UTRs of multiple mRNAs (e.g. miR-200 in the last model of this study) (Briskin et al., 2020; Grimson et al., 2007). We expect that some of these mechanisms can be used to support stable intermediate cell phenotypes. Future work is necessary to systematically compare the functions of these molecular mechanisms and to identify their existence in specific biological contexts. Nonetheless, the prevalence of the multi-site microRNA interactions with individual and groups of mRNAs in EMT and other systems suggests that these regulatory networks have nontrivial emergent functions.

In this study, we showed similar and modularizable performances of transcriptional and post-transcriptional mechanisms in generating intermediate EMT states. These two mechanisms are different in their cellular locations. While it may be beneficial for cells to combine both nuclear (transcriptional) and cytosolic (post-transcriptional) machineries to achieve the desired goal of stabilizing intermediate states, the post-transcriptional mechanism may be advantageous in terms of avoiding some sources of noise. This is because transcription is subject to significant noise levels due to the low numbers of DNA coding for the regulatory products, whereas post-transcriptional mechanisms involve larger numbers of molecules (Kataruka et al., 2022), which reduce intrinsic noise. Therefore, we expect that the proposed RNA-centric mechanism for stabilizing intermediate EMT states can be an efficient strategy for cells to adopt hybrid phenotypes with cytosolic reactions without the need for transcriptional regulatory systems.

The motivation to build our models derived from inconsistencies between existing EMT models that predict a paucity of EMT intermediate states, and experimental single-cell transcriptomic data that have been interpreted to support a wealth of states in a phenotypic continuum (Cook and Vanderhyden, 2020; Deshmukh et al., 2021; Panchy et al., 2022). It should be noted that some continuum-like expression distributions revealed by scRNA-seq may be due to technical noise, whose effects on EMT progression warrant future investigation. Nonetheless, it is plausible that some of the intermediate states should be favored, perhaps in relation with microenvironmental factors such as nutrient availability and cytokines. The models we present here can guide experimentation designed to validate the role of microRNAs to stabilize a constrained number of intermediate cell phenotypes both in physiologic and pathologic systems beyond EMT. For example, in small-cell lung cancer, a phenotypic continuum spanning neuroendocrine (NE) and non-neuroendocrine (non-NE) cell states was recently described based on archetype analysis (AA) of experimental data (Groves et al., 2022; Hausser and Alon, 2020). NE to non-NE transitions bear many similarities to EMT, particularly as it pertains increased metastatic properties of non-NE cells, and the similarity of transcriptional signatures (Ireland et al., 2020). Similar to the EMT phenotypic continuum, AA identified a phenotypic continuum for SCLC NE and non-NE subtypes, which gene expression enrichment links to cellular task they are optimized for, such as secretion, proliferation, or motility. In this continuum, cells may be specialists at one task or sub-optimal generalists at one or more tasks. Thus, the intermediate generalist phenotypes arise from task trade-offs, so that they can perform suboptimally to be in tune with the microenvironment that they experience. The continuum then becomes dominated by Pareto optimality. Interrogating the role of microRNAs in NE to non-NE transitions intermediates will provide mechanistic underpinnings for plasticity and high metastatic propensity of SCLC tumors. Our modeling approach is a valuable starting point for streamlining these experiments. In summary, future studies may reveal that post-transcriptional mechanisms are widely used by mammalian cells for generating intermediate states, both for EMT and other differentiation systems. Our mathematical models will aid in designing experiments to test this possibility.

### Limitations of the study

While it is expected that several gene regulatory mechanisms can generate multistable systems (Angeli et al., 2004; Li et al., 2017; Markevich et al., 2004), the predicted co-existing cell states from cooperative RNA degradation models were not validated experimentally in this study, and they warrant future work.

The observed EMT continuum at the transcriptome level (Cook and Vanderhyden, 2020; Deshmukh et al., 2021; Panchy et al., 2022) does not exclude the possibility that only a few cell states emerge post-translationally to govern cellular phenotypes in a more discrete manner. Future work is needed to quantitatively characterize the continuity of EMT spectrum in terms of functional phenotypes.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHOD DETAILS
  - Model construction
  - Testing model parameterizations for multiple attractors
  - Numerical bifurcation analysis
  - Stochastic simulation and quasi-potential landscape
  - Single-cell RNA-sequencing data origin
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Parameter sampling for multistable systems
  - Parameter sensitivity analysis
  - Enumeration of instances of network topologies
  - Single-cell RNA-sequencing data analysis

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2022.105224>.

## ACKNOWLEDGMENTS

This work was supported by National Institutes of Health grant R01GM140462 to T.H.

## AUTHOR CONTRIBUTIONS

Conceptualization: T.H. Data curation: M.C.Y.P. and T.H. Investigation: B.N. and T.H. Visualization: B.N., M.C.Y.P., and T.H. Supervision: V.Q. and T.H. Writing: B.N., V.Q., and T.H.

## DECLARATION OF INTERESTS

V.Q. is co-founder of Parthenon Therapeutics, Inc., and Duet BioSystems, Inc. The other authors declare no conflict of interest.

## INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: May 24, 2022

Revised: August 21, 2022

Accepted: September 23, 2022

Published: October 21, 2022

## REFERENCES

- Agarwal, V., Bell, G.W., Nam, J.-W., and Bartel, D.P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *Elife* 4, e05005.
- Angeli, D., Ferrell, J.E., and Sontag, E.D. (2004). Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems. *Proc. Natl. Acad. Sci. USA* 101, 1822–1827.
- Aiello, N.M., Maddipati, R., Norgard, R.J., Balli, D., Li, J., Yuan, S., Yamazoe, T., Black, T., Sahnoud, A., Furth, E.E., et al. (2018). EMT subtype influences epithelial plasticity and mode of cell migration. *Dev. Cell* 45, 681–695.e4.
- Bartel, D.P. (2018). Metazoan microRNAs. *Cell* 173, 20–51.
- Battle, E., Sancho, E., Francí, C., Domínguez, D., Monfar, M., Baulida, J., and García De Herreros, A. (2000). The transcription factor Snail is a repressor of E-cadherin gene expression in epithelial tumour cells. *Nat. Cell Biol.* 2, 84–89. <https://doi.org/10.1038/35000034>.
- Bracken, C.P., Gregory, P.a., Kolesnikoff, N., Bert, A.G., Wang, J., Shannon, M.F., and Goodall, G.J.

- (2008). A double-negative feedback loop between ZEB1-SIP1 and the microRNA-200 family regulates epithelial-mesenchymal transition. *Cancer Res.* 68, 7846–7854. <https://doi.org/10.1158/0008-5472.CAN-08-1942>.
- Bratley, P., and Fox, B.L. (1988). Algorithm 659: Implementing Sobol's quasirandom sequence generator. *ACM Trans. Math. Softw.* 14, 88–100. <https://doi.org/10.1145/42288.214372>.
- Briskin, D., Wang, P.Y., and Bartel, D.P. (2020). The biochemical basis for the cooperative action of microRNAs. *Proc. Natl. Acad. Sci. USA* 117, 17764–17774.
- Buck, A.H., Coakley, G., Simbari, F., McSorley, H.J., Quintana, J.F., Le Bihan, T., Kumar, S., Abreu-Goodger, C., Lear, M., Harcus, Y., et al. (2014). Exosomes secreted by nematode parasites transfer small RNAs to mammalian cells and modulate innate immunity. *Nat. Commun.* 5, 5488–5512.
- Campbell, K., Rossi, F., Adams, J., Pitsidianaki, I., Barriga, F.M., Garcia-Gerique, L., Batlle, E., Casanova, J., and Casali, A. (2019). Collective cell migration and metastases induced by an epithelial-to-mesenchymal transition in *Drosophila* intestinal tumors. *Nat. Commun.* 10, 2311–2410.
- Celià-Terrassa, T., Bastian, C., Liu, D.D., Ell, B., Aiello, N.M., Wei, Y., Zamalloa, J., Blanco, A.M., Hang, X., Kunisky, D., et al. (2018). Hysteresis control of epithelial-mesenchymal transition dynamics conveys a distinct program with enhanced metastatic ability. *Nat. Commun.* 9, 5005. <https://doi.org/10.1038/s41467-018-07538-7>.
- Chen, M., and Cao, Y. (2019). Analysis and remedy of negativity problem in hybrid stochastic simulation algorithm and its application. *BMC Bioinf.* 20, 315. <https://doi.org/10.1186/s12859-019-2836-z>.
- Choi, K., Medley, J.K., König, M., Stocking, K., Smith, L., Gu, S., and Sauro, H.M. (2018). Tellurium: an extensible python-based modeling environment for systems and synthetic biology. *Biosystems* 171, 74–79.
- Cook, D.P., and Vanderhyden, B.C. (2020). Context specificity of the EMT transcriptional response. *Nat. Commun.* 11, 2142–2149.
- Cook, D.P., and Vanderhyden, B.C. (2022). Transcriptional census of epithelial-mesenchymal plasticity in cancer. *Sci. Adv.* 8, eabi7640.
- Cursons, J., Pillman, K.A., Scheer, K.G., Gregory, P.A., Foroutan, M., Hediyyeh-Zadeh, S., Toubia, J., Crampin, E.J., Goodall, G.J., Bracken, C.P., and Davis, M.J. (2018). Combinatorial targeting by MicroRNAs Co-ordinates post-transcriptional control of EMT. *Cell Syst.* 7, 77–91.e7. <https://doi.org/10.1016/j.cels.2018.05.019>.
- Deshmukh, A.P., Vasaikar, S.V., Tomczak, K., Tripathi, S., Den Hollander, P., Arslan, E., Chakraborty, P., Soundararajan, R., Jolly, M.K., Rai, K., et al. (2021). Identification of EMT signaling cross-talk and gene regulatory networks by single-cell RNA sequencing. *Proc. Natl. Acad. Sci. USA* 118, e2102050118.
- Eichhorn, S.W., Guo, H., McGeary, S.E., Rodriguez-Mias, R.A., Shin, C., Baek, D., Hsu, S.-h., Ghoshal, K., Villén, J., and Bartel, D.P. (2014). mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues. *Mol. Cell* 56, 104–115.
- Gardner, T.S., Cantor, C.R., and Collins, J.J. (2000). Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 403, 339–342.
- Goetz, H., Melendez-Alvarez, J.R., Chen, L., and Tian, X.-J. (2020). A plausible accelerating function of intermediate states in cancer metastasis. *PLoS Comput. Biol.* 16, e1007682.
- 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., and Nieto, M.A. (2015). Snail1-induced partial epithelial-to-mesenchymal transition drives renal fibrosis in mice and can be targeted to reverse established disease. *Nat. Med.* 21, 989–997. <https://doi.org/10.1038/nm.3901>.
- Grimson, A., Farh, K.K.-H., Johnston, W.K., Garrett-Engle, P., Lim, L.P., and Bartel, D.P. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell* 27, 91–105.
- Groves, S.M., Ildefonso, G.V., McAtee, C.O., Ozawa, P.M.M., Ireland, A.S., Stauffer, P.E., Wasdin, P.T., Huang, X., Qiao, Y., Lim, J.S., et al. (2022). Archetype tasks link intratumoral heterogeneity to plasticity and cancer hallmarks in small cell lung cancer. *Cell Syst.* 13, 690–710.e17.
- Guo, F., Cogdell, D., Hu, L., Yang, D., Sood, A.K., Xue, F., and Zhang, W. (2014). MiR-101 suppresses the epithelial-to-mesenchymal transition by targeting ZEB1 and ZEB2 in ovarian carcinoma. *Oncol. Rep.* 31, 2021–2028.
- Hagberg, A.A., Schult, D.A., and Swart, P.J. (2008). Exploring network structure, dynamics, and function using NetworkX. In *Python in Science Conference, G. Varoquaux, T. Vaught, and J. Millman, eds.*
- Han, H., Cho, J.-W., Lee, S., Yun, A., Kim, H., Bae, D., Yang, S., Kim, C.Y., Lee, M., Kim, E., et al. (2018). TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res.* 46, D380–D386.
- Hausser, J., and Alon, U. (2020). Tumour heterogeneity and the evolutionary trade-offs of cancer. *Nat. Rev. Cancer* 20, 247–257.
- Hong, T., Watanabe, K., Ta, C.H., Villarreal-Ponce, A., Nie, Q., and Dai, X. (2015). An ovnl2-zeb1 mutual inhibitory circuit governs bidirectional and multi-step transition between epithelial and mesenchymal states. *PLoS Comput. Biol.* 11, e1004569. <https://doi.org/10.1371/journal.pcbi.1004569>.
- Huang, B., Lu, M., Jia, D., Ben-Jacob, E., Levine, H., and Onuchic, J.N. (2017). Interrogating the topological robustness of gene regulatory circuits by randomization. *PLoS Comput. Biol.* 13, e1005456.
- Huntzinger, E., and Izaurralde, E. (2011). Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nat. Rev. Genet.* 12, 99–110. <https://doi.org/10.1038/nrg2936>.
- Hussen, B.M., Shoorei, H., Mohaqiq, M., Dinger, M.E., Hidayat, H.J., Taheri, M., and Ghafouri-Fard, S. (2021). The impact of non-coding RNAs in the epithelial to mesenchymal transition. *Front. Mol. Biosci.* 8, 665199.
- Ireland, A.S., Micinski, A.M., Kastner, D.W., Guo, B., Wait, S.J., Spainhower, K.B., Conley, C.C., Chen, O.S., Guthrie, M.R., Soltero, D., et al. (2020). MYC drives temporal evolution of small cell lung cancer subtypes by reprogramming neuroendocrine fate. *Cancer Cell* 38, 60–78.e12.
- Jiménez, A., Cotterell, J., Munteanu, A., and Sharpe, J. (2017). A spectrum of modularity in multi-functional gene circuits. *Mol. Syst. Biol.* 13, 925.
- Kai, Z.S., and Pasquinelli, A.E. (2010). MicroRNA assassins: factors that regulate the disappearance of miRNAs. *Nat. Struct. Mol. Biol.* 17, 5–10. <https://doi.org/10.1038/nsmb.1762>.
- Kataruka, S., Kinterova, V., Horvat, F., Kulmann, M.I.R., Kanka, J., and Svoboda, P. (2022). Physiologically relevant miRNAs in mammalian oocytes are rare and highly abundant. *EMBO Rep.* 23, e53514.
- Lai, X., Eberhardt, M., Schmitz, U., and Vera, J. (2019). Systems biology-based investigation of cooperating microRNAs as monotherapy or adjuvant therapy in cancer. *Nucleic Acids Res.* 47, 7753–7766.
- Li, C.-J., Hong, T., Tung, Y.-T., Yen, Y.-P., Hsu, H.-C., Lu, Y.-L., Chang, M., Nie, Q., and Chen, J.-A. (2017). MicroRNA filters Hox temporal transcription noise to confer boundary formation in the spinal cord. *Nat. Commun.* 8, 14685. <https://doi.org/10.1038/ncomms14685>.
- Li, C., and Wang, J. (2014). Landscape and flux reveal a new global view and physical quantification of mammalian cell cycle. *Proc. Natl. Acad. Sci. USA* 111, 14130–14135.
- Li, C.J., Liao, E.S., Lee, Y.H., Huang, Y.Z., Liu, Z., Willems, A., Garside, V., McGlenn, E., Chen, J.A., and Hong, T. (2021). MicroRNA governs bistable cell differentiation and lineage segregation via a noncanonical feedback. *Mol. Syst. Biol.* 17, e9945. (Cover story).
- Lu, M., Jolly, M.K., Gomoto, R., Huang, B., Onuchic, J., and Ben-Jacob, E. (2013a). Tristability in cancer-associated microRNA-TF chimera toggle switch. *J. Phys. Chem. B* 117, 13164–13174. <https://doi.org/10.1021/jp403156m>.
- Lu, M., Jolly, M.K., Levine, H., Onuchic, J.N., and Ben-Jacob, E. (2013b). MicroRNA-based regulation of epithelial-hybrid-mesenchymal fate determination. *Proc. Natl. Acad. Sci. USA* 110, 18144–18149.
- Markevich, N.I., Hoek, J.B., and Kholodenko, B.N. (2004). Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades. *J. Cell Biol.* 164, 353–359. <https://doi.org/10.1083/jcb.200308060>.
- Minchington, T.G., Griffiths-Jones, S., and Papalopulu, N. (2020). Dynamical gene regulatory networks are tuned by transcriptional

autoregulation with microRNA feedback. *Sci. Rep.* 10, 12960–13013.

Nordick, B., and Hong, T. (2021). Identification, visualization, statistical analysis and mathematical modeling of high-feedback loops in gene regulatory networks. *BMC Bioinf.* 22, 481. <https://doi.org/10.1186/s12859-021-04405-z>.

Nordick, B., Yu, P.Y., Liao, G., and Hong, T. (2022). Nonmodular oscillator and switch based on RNA decay drive regeneration of multimodal gene expression. *Nucleic Acids Res.* 50, 3693–3708. <https://doi.org/10.1101/2022.01.12.475956>.

Panchy, N., Watanabe, K., Takahashi, M., Willems, A., and Hong, T. (2022). Comparative single-cell transcriptomes of dose and time dependent epithelial-mesenchymal spectrums. *NAR Genom. Bioinform.* 4, lqac072. <https://doi.org/10.1093/nargab/lqac072>.

Pastushenko, I., Brisebarre, A., Sifrim, A., Fioramonti, M., Revenco, T., Boumahdi, S., Van Keymeulen, A., Brown, D., Moers, V., Lemaire, S., et al. (2018). Identification of the tumour transition states occurring during EMT. *Nature* 556, 463–468.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.

Rackauckas, C., and Nie, Q. (2017). *Differentialequations.jl—a performant and feature-rich ecosystem for solving differential equations in julia.* *J. Open Res. Softw.* 5, 15.

Sánchez-Tilló, E., Lázaro, A., Torrent, R., Cuatrecasas, M., Vaquero, E.C., Castells, A., Engel, P., and Postigo, A. (2010). ZEB1 represses

E-cadherin and induces an EMT by recruiting the SWI/SNF chromatin-remodeling protein BRG1. *Oncogene* 29, 3490–3500.

Steinway, S.N., Zañudo, J.G.T., Michel, P.J., Feith, D.J., Loughran, T.P., and Albert, R. (2015). Combinatorial interventions inhibit TGFβ-driven epithelial-to-mesenchymal transition and support hybrid cellular phenotypes. *NPJ Syst. Biol. Appl.* 1, 15014.

Subbalakshmi, A.R., Sahoo, S., Biswas, K., and Jolly, M.K. (2021). A computational systems biology approach identifies SLUG as a mediator of partial Epithelial-Mesenchymal Transition (EMT). *Cells Tissues Organs* 211, 1–14.

Tan, T.Z., Miow, Q.H., Miki, Y., Noda, T., Mori, S., Huang, R.Y.J., and Thiery, J.P. (2014). Epithelial-mesenchymal transition spectrum quantification and its efficacy in deciphering survival and drug responses of cancer patients. *EMBO Mol. Med.* 6, 1279–1293.

Thiery, J.P., Acloque, H., Huang, R.Y.J., and Nieto, M.A. (2009). Epithelial-mesenchymal transitions in development and disease. *Cell* 139, 871–890. <https://doi.org/10.1016/j.cell.2009.11.007>.

Tian, X.-J., Zhang, H., and Xing, J. (2013). Coupled reversible and irreversible bistable switches underlying TGFβ-induced epithelial to mesenchymal transition. *Biophys. J.* 105, 1079–1089. <https://doi.org/10.1016/j.bpj.2013.07.011>.

Türei, D., Korcsmáros, T., and Saez-Rodriguez, J. (2016). OmniPath: guidelines and gateway for literature-curated signaling pathway resources. *Nat. Methods* 13, 966–967.

Vannier, C., Mock, K., Brabletz, T., and Driever, W. (2013). Zeb1 regulates E-cadherin and Epcam

(epithelial cell adhesion molecule) expression to control cell behavior in early zebrafish development. *J. Biol. Chem.* 288, 18643–18659. <https://doi.org/10.1074/jbc.M113.467787>.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* 17, 261–272. <https://doi.org/10.1038/s41592-019-0686-2>.

Watanabe, K., Villarreal-Ponce, A., Sun, P., Salmans, M.L., Fallahi, M., Andersen, B., and Dai, X. (2014). Mammary morphogenesis and regeneration require the inhibition of EMT at terminal end buds by *Ovol2* transcriptional repressor. *Dev. Cell* 29, 59–74. <https://doi.org/10.1016/j.devcel.2014.03.006>.

Wellner, U., Schubert, J., Burk, U.C., Schmalhofer, O., Zhu, F., Sonntag, A., Waldvogel, B., Vannier, C., Darling, D., Zur Hausen, A., et al. (2009). The EMT-activator ZEB1 promotes tumorigenicity by repressing stemness-inhibiting microRNAs. *Nat. Cell Biol.* 11, 1487–1495.

Zhang, J., Tian, X.J., Zhang, H., Teng, Y., Li, R., Bai, F., Elankumaran, S., and Xing, J. (2014). TGFβ-induced epithelial-to-mesenchymal transition proceeds through stepwise activation of multiple feedback loops. *Sci. Signal.* 7, ra91. <https://doi.org/10.1126/scisignal.2005304>.

Zheng, M., Jiang, Y.-p., Chen, W., Li, K.-d., Liu, X., Gao, S.-y., Feng, H., Wang, S.-s., Jiang, J., Ma, X.-r., et al. (2015). Snail and Slug collaborate on EMT and tumor metastasis through miR-101-mediated EZH2 axis in oral tongue squamous cell carcinoma. *Oncotarget* 6, 6797–6810. <https://doi.org/10.18632/oncotarget.3180>.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Deposited data</b>		
Single-cell RNA-seq data for TGF- $\beta$ induced EMT in A549 lung epithelial cell line	(Cook and Vanderhyden, 2020)	GEO: GSE147405
Single-cell RNA-seq data for TGF- $\beta$ induced EMT in MCF10A breast epithelial cell line	(Panchy et al., 2022)	GEO: GSE213753
<b>Software and algorithms</b>		
Multistable model parameterization search scripts	This paper	<a href="https://github.com/BenNordick/MMI4">https://github.com/BenNordick/MMI4</a>
Tellurium 2.2.0	(Choi et al., 2018)	<a href="https://tellurium.analogmachine.org/">https://tellurium.analogmachine.org/</a>
SciPy 1.1.0	(Virtanen et al., 2020)	<a href="https://scipy.org/">https://scipy.org/</a>
NetworkX 2.5.1	(Hagberg et al., 2008)	<a href="https://networkx.org/">https://networkx.org/</a>
Scikit-learn 0.20.0	(Pedregosa et al., 2011)	<a href="https://scikit-learn.org/">https://scikit-learn.org/</a>
Python 3.8.5	Python Software Foundation	<a href="https://python.org/">https://python.org/</a>

### RESOURCE AVAILABILITY

#### Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Tian Hong ([hongtian@utk.edu](mailto:hongtian@utk.edu)).

#### Materials availability

This study did not generate new unique reagents.

#### Data and code availability

- This paper analyzes existing, publicly available data. The accession numbers for the datasets are listed in the [key resources table](#).
- All original code has been deposited at GitHub and is publicly available as of the date of publication. The repository URL is listed in the [key resources table](#).
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

### METHOD DETAILS

#### Model construction

Each model is described by a system of ordinary differential equations (ODEs). Dynamics of molecular species are described by two types of function. Mass-action kinetics is used to model elementary reactions, i.e. the basal processes of RNA transcription/maturation, constitutive RNA decay, mRNA-microRNA binding and unbinding, regulated decay of each RNA member of each complex, translation, and protein decay. If applicable to the model, Hill functions are used to describe regulated transcription rates as in HiLoop models (Nordick and Hong, 2021):

$$f * \frac{1}{1 + (X/K)^n} \quad (\text{Equation 1})$$

for repression, or

$$f * \frac{(X/K)^n}{1 + (X/K)^n} \quad (\text{Equation 2})$$

for activation, where  $X$  is the concentration of the regulator protein,  $K$  is the threshold of regulation,  $n$  is the cooperativity of regulation, and  $f$  is the proportion of maximum transcription rate controlled by this regulator. Multiple regulatory influences are combined by adding their Hill functions.

The models based on the Subbalakshmi et al. network follow the form used there and by Lu et al. (Lu et al., 2013b; Subbalakshmi et al., 2022)

$$H^s(X, K, n, F) = H^-(X, K, n) + F(1 - H^-(X, K, n)) \quad (\text{Equation 3})$$

where

$$H^-(X, K, n) = \frac{1}{1 + (X/K)^n} \quad (\text{Equation 4})$$

and  $F$  is the fold-change in target expression resulting from overexpression of the regulator. Multiple regulations are combined by multiplying their  $H^s$  functions.

In all models, each type of RNA within each type of complex is assigned a regulated degradation factor,  $\alpha$  for mRNAs or  $\beta$  for microRNAs, representing the change in its degradation rate in the complex relative to its free state (Nordick et al., 2022).

Binding and unbinding in RNA complexes is always modeled explicitly to ensure that each microRNA molecule acts on only one mRNA at once, though Subbalakshmi et al.'s allowance of incomplete translation repression  $t$  by microRNA binding (Huntzinger and Izaurralde, 2011) is followed. For all mRNA-microRNA binding events in all models, the microRNA dissociation rate  $\kappa_{\text{off}} = 100$  and each association constant  $K = \kappa_{\text{on}}/\kappa_{\text{off}}$ .

For mRNAs with multiple binding sites for a type of microRNA, any binding/unbinding order is allowed, but complexes that differ only in the positions of sites bound by their microRNAs are grouped together into one variable, which represents the total concentration of all forms of the complex. This is accomplished by multiplying the single-microRNA association rate by the number of remaining free sites and multiplying the dissociation or microRNA degradation rate by the number of occupied sites.

The reactions, rate functions, and parameters in each model are listed in the [supplementary information](#). Antimony model code for use with the Tellurium modeling package (Choi et al., 2018) can be found in the associated GitHub repository. All variables and parameters are dimensionless quantities (Nordick et al., 2022).

For simulation, the set of reactions is converted to a system of ODEs, one for each RNA, protein, or complex. A species' rate of change is the sum of the rates of the reactions that produce or consume it, weighted by the net change in the species' amount caused by the reaction. As an example, the ODEs for the one-microRNA MMI4 Model (derived from [Table S1](#)) are

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

where  $R$  is the concentration of free mRNA,  $r$  is the concentration of free microRNA, and  $C_n$  is the total concentration of complexes with  $n$  microRNA molecules bound to an mRNA molecule except  $C_0 := R$  and  $C_5 := 0$ .

One reaction can contribute terms to multiple species' rate-of-change equations. The regulated mRNA decay reactions, for example, which occur at rate  $\alpha_n C_n$ , destroy one unit of complex  $C_n$  but produce  $n$  units of free microRNA  $r$ , so they contribute a  $-\alpha_n C_n$  term to  $dC_n/dt$  and a  $+n\alpha_n C_n$  term to  $dr/dt$ .



### Testing model parameterizations for multiple attractors

Each parameterized system was simulated starting from at least 100 initial conditions which span at least five powers of 5 in a log-uniform manner, drawn from a Sobol quasirandom sequence (Bratley and Fox, 1988). The initial concentration of each free RNA species was determined by one dimension of the Sobol hypercube, protein initial concentrations were set to their coding mRNA's initial concentration times the translation rate, and other species' initial concentrations were zero. Deterministic simulation with Tellurium (Choi et al., 2018) proceeded for at least 100 time units up to a maximum of 1,000 time units until the system reached a steady state as determined by the 2-norm of the derivatives vector falling below  $10^{-7}$ . Simulation endpoint concentration vectors  $a$  and  $b$  were considered equivalent attractors if their difference's 2-norm  $\|a - b\|_2$  was less than  $10^{-4}$  times the number of species  $n$ , or if  $\|a - b\|_2 < 0.3\min(\|a\|_2, \|b\|_2) < 0.01n$ .

### Numerical bifurcation analysis

Bifurcation diagrams were created with the AUTO2000 plugin for Tellurium 2.2.0 (Choi et al., 2018). For bifurcations with respect to transcription rate, continuation proceeded backwards from a high monostable signal value. For bifurcations with respect to degradation rate, continuation proceeded forward from a low degradation rate and was restarted from a different steady state when needed to follow all attractors.

### Stochastic simulation and quasi-potential landscape

We performed stochastic simulations for the modified 3-TF model with various microRNA binding sites by adding an independent multiplicative noise term to each ODE. Divergence arising from negative concentrations was reduced by applying the Zero-Reaction remedy (Chen and Cao, 2019). Starting with a population of an equal number of cells at each deterministic attractor, we solved the stochastic ODEs at a noise intensity of 0.2 for 200 time units using DifferentialEquations.jl (Rackauckas and Nie, 2017).

To visualize the epithelial-mesenchymal gene expression space in which the stochastic system fluctuates in the long term, the concentrations of Zeb1 and E-cadherin protein were extracted for each simulated cell at intervals of 5 time units starting at time 150. This pair of molecules were chosen because they were very widely studied and are considered core effectors/regulators of EMT (Celià-Terrassa et al., 2018; Sanchez-Tillo et al., 2010; Wellner et al., 2009). The concentrations of protein-coding genes are readily inferred from RNA-sequencing data. By contrast, microRNA concentrations are more difficult to obtain. To avoid distortions from temporarily negative concentrations, timepoints with either component less than  $10^{-7}$  were not used. Quasi-potential landscapes (Li and Wang, 2014) were rendered with potential  $U(x) = -\log P_S(x)$ , where  $P_S(x)$  is the probability density function computed by scikit-learn's 2-dimensional Gaussian kernel density estimate (Pedregosa et al., 2011) of bandwidth 0.14 on the base-2 logarithm of E-cadherin protein and the base-10 logarithm of Zeb1 protein, as Zeb1's expression was more variable under the parameter values tested.

### Single-cell RNA-sequencing data origin

We analyzed two datasets for validating some predictions from our final model that includes miR-101. Single-cell RNA-sequencing data for TGF- $\beta$  induced EMT in A549 lung epithelial cell line (Cook and Vanderhyden, 2020) (GSE147405), and in MCF10A breast epithelial cell line (Panchy et al., 2022) (GSE213753) were obtained from NCBI Gene Expression Omnibus (accessions in key resources table). The A549 cells were treated with a single dose of TGF- $\beta$  and their data have time labels of 0-7 days. The MCF10A cells were treated with multiple doses (0-200 p.m. chosen for this work) of TGF- $\beta$  for 14 days.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Parameter sampling for multistable systems

The likelihood of the single-microRNA MMI4 Model to generate multiple attractors was tested by sampling all regulated degradation factors independently from a log-uniform distribution on  $[2^{-3}, 2^4]$  (a range consistent with previous experiments and models (Eichhorn et al., 2014; Nordick et al., 2022)), the microRNA association constant from a log-uniform distribution on  $[10^3, 10^6]$ , and the microRNA transcription rate from a log-uniform distribution on  $[2^{-4}, 2^1]$ . The conditions for the two-microRNA MMI4 Model to produce multiple attractors were investigated by sampling from the same parameter regions independently for each microRNA, discarding systems with at most two attractors. Kernel density estimates were computed by SciPy (Virtanen et al., 2020) with standard settings.

When simply searching for example parameter sets in other models rather than characterizing the parameter space, subregions that provided computational efficiency were selected empirically from plausible regions similar to previous work (Nordick et al., 2022); all sampling distributions and parameter sets of interest for each model are listed in [supplementary tables](#).

### Parameter sensitivity analysis

To determine how much a parameter must change to abolish high multistability, the system was repeatedly re-simulated as described above. First, the original parameter value was doubled or halved up to twelve times until the perturbed parameter set had fewer attractors than the original. To find the precise limit, a binary search was then conducted between the last many-attractor and first fewer-attractor power of 2. The binary search was terminated when the base-2 logarithms of the narrowed parameter value endpoints differed by less than 0.0001.

To address derailment of the binary search from numerical imprecision in attractor detection, values 1% above and below each endpoint were re-simulated for verification. If a value inside the many-attractor range was found to lack an attractor, or if a value outside the range was found to still have many attractors, the limit search process was repeated with more stringent integrator settings. The initial time was increased to 500 units, the maximum time was increased to 4000 units, and the initial timestep was decreased to 0.25. Parameter limits that still failed verification, i.e. could not be determined to within 1%, were discarded.

### Enumeration of instances of network topologies

We examined a previously curated list of core EMT genes (Tan et al., 2014). Among them, 232 were classified as E genes and 191 as M genes. In addition, we used a list of 133 microRNAs, each with experimental evidence supporting its role in EMT (Cursons et al., 2018). With these lists and the TargetScan program for microRNA binding site prediction (Agarwal et al., 2015), we first identified 46 EMT genes regulated by a total of four binding sites of one or two EMT microRNAs, i.e. the MMI4 Models. Including those, we also identified 93 EMT genes matching the MMI2 Model.

For the topologies that combine transcriptional and posttranscription regulation, the TRRUST2 network (Han et al., 2018) and transcriptional subgraph of the OmniPath network (Türei et al., 2016) as preprocessed for HiLoop (Nordick and Hong, 2021) were added together except for regulations given opposite signs by the two networks. The existence of a regulatory path from each EMT MMI2 gene to every other, up to the specified path length of 1 or 5, was tested in the combined transcriptional network using NetworkX (Hagberg et al., 2008). Each ordered pair with such a directed path was considered an instance of the Chained-MMI2 topology.

The list of EMT MMI2 genes that, directly or indirectly via the combined network, transcriptionally regulates each EMT gene was similarly obtained. If an unordered pair of EMT MMI2 genes regulated the same EMT target gene without common dependencies on any intermediate genes, the partially ordered triple of regulators and target was considered an instance of the Co-targeting-MMI2 topology. The condition of no common intermediates avoids counting subnetworks in which one MMI2 gene regulates the target through the other MMI2 gene, an arrangement which may have different dynamics, and non-minimal subnetworks in which the target is downstream of an already co-targeted gene.

Note that these distinct instances are only based on different combinations of EMT genes. Considering combinations of EMT-associated microRNA binding sites would yield much larger numbers of instances.

### Single-cell RNA-sequencing data analysis

The Uniform Manifold Approximation and Projection (UMAP) was used to examine the overall distributions of cells for both datasets. Z-scores were computed for each gene across each dataset. Representative EMT genes were used to visualize the progression of EMT via their expression across time points and doses. For genes that were poorly detected (*ZEB1* and *CDH1* in A549 cells, and *ZEB1* in MCF10A cells), only cells with non-zero expression were used for this visualization. To infer microRNA expression levels, we computed the mean expression levels of the genes potentially targeted by miR-200b/c, and miR-101 respectively. We obtained lists of predicted target genes from TargetScan (Agarwal et al., 2015). To avoid the trivial case in which the mean target gene expression was merely driven by *ZEB1*

(regulated by the microRNAs in our model) and its downstream effector gene *VIM*, we excluded these two genes from this inference analysis. After the exclusion, miR-101 and miR-200b/c had 954 and 1194 predicted target genes respectively. We normalized the mean target gene expression by subtracting the overall mean expression for each time point or dose from the target gene expression.