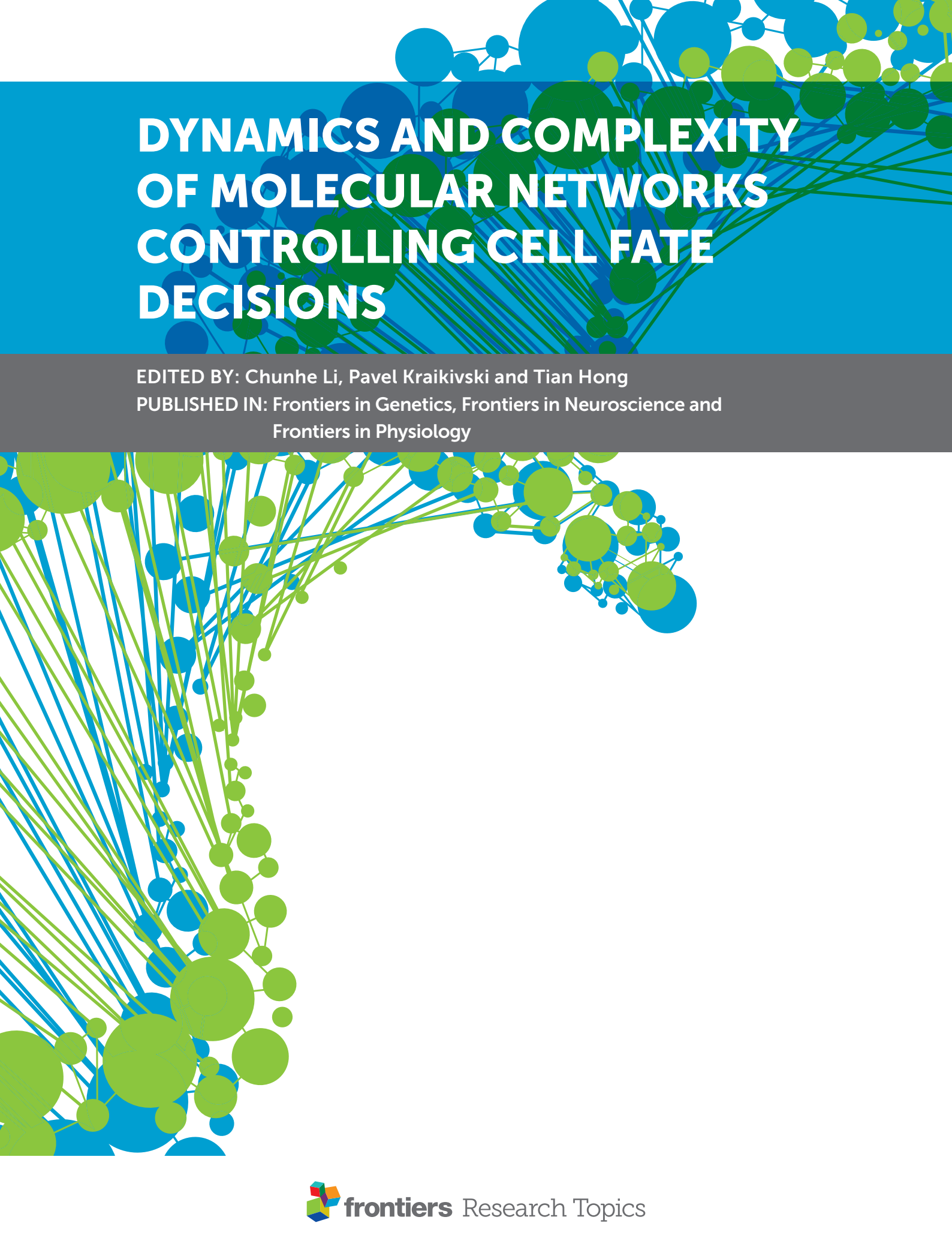




DYNAMICS AND COMPLEXITY OF MOLECULAR NETWORKS CONTROLLING CELL FATE DECISIONS

EDITED BY: Chunhe Li, Pavel Kraikivski and Tian Hong

PUBLISHED IN: *Frontiers in Genetics*, *Frontiers in Neuroscience* and
Frontiers in Physiology



frontiers

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence.

The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714

ISBN 978-2-83250-655-4

DOI 10.3389/978-2-83250-655-4

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

DYNAMICS AND COMPLEXITY OF MOLECULAR NETWORKS CONTROLLING CELL FATE DECISIONS

Topic Editors:

Chunhe Li, Fudan University, China

Pavel Kraikivski, Virginia Tech, United States

Tian Hong, The University of Tennessee, Knoxville, United States

Citation: Li, C., Kraikivski, P., Hong, T., eds. (2022). Dynamics and Complexity of Molecular Networks Controlling Cell Fate Decisions. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83250-655-4

Table of Contents

05	<i>Compressive Sensing Inference of Neuronal Network Connectivity in Balanced Neuronal Dynamics</i>
	Victor J. Barranca and Douglas Zhou
15	<i>Revealing Dynamic Mechanisms of Cell Fate Decisions From Single-Cell Transcriptomic Data</i>
	Jiajun Zhang, Qing Nie and Tianshou Zhou
28	<i>Analysis of Single-Cell Gene Pair Coexpression Landscapes by Stochastic Kinetic Modeling Reveals Gene-Pair Interactions in Development</i>
	Cameron P. Gallivan, Honglei Ren and Elizabeth L. Read
44	<i>Quantification of the Underlying Mechanisms and Relationships Among Cancer, Metastasis, and Differentiation and Development</i>
	Chong Yu, Qiong Liu, Cong Chen and Jin Wang
57	<i>A Computational Model of the Endothelial to Mesenchymal Transition</i>
	Nathan Weinstein, Luis Mendoza and Elena R. Álvarez-Buylla
83	<i>Searching for the Mechanisms of Mammalian Cellular Aging Through Underlying Gene Regulatory Networks</i>
	Wenbo Li, Lei Zhao and Jin Wang
97	<i>Stochastic Dynamics of Gene Switching and Energy Dissipation for Gene Expression</i>
	Quan Liu, FengZhen Yu, Liang Yi, Yijun Gao, Rong Gui, Ming Yi and Jianqiang Sun
106	<i>A Probabilistic Approach to Explore Signal Execution Mechanisms With Limited Experimental Data</i>
	Michael A. Kochen and Carlos F. Lopez
123	<i>Understanding the Principles of Pattern Formation Driven by Notch Signaling by Integrating Experiments and Theoretical Models</i>
	Federico Bocci, José Nelson Onuchic and Mohit Kumar Jolly
147	<i>A Guard Cell Abscissic Acid (ABA) Network Model That Captures the Stomatal Resting State</i>
	Parul Maheshwari, Sarah M. Assmann and Reka Albert
167	<i>Orchestration of lincRNA-p21 and miR-155 in Modulating the Adaptive Dynamics of HIF-1α</i>
	Cheng-Yuan Sun, Xiao-Peng Zhang, Feng Liu and Wei Wang
179	<i>A Systems Biology Approach to Identifying a Master Regulator That Can Transform the Fast Growing Cellular State to a Slowly Growing One in Early Colorectal Cancer Development Model</i>
	Jihye Choi, Jeong-Ryeol Gong, Chae Young Hwang, Chang Young Joung, Soobeom Lee and Kwang-Hyun Cho
192	<i>Quantifying the Stability of Coupled Genetic and Epigenetic Switches With Variational Methods</i>
	Amogh Sood and Bin Zhang
201	<i>Bistable Mathematical Model of Neutrophil Migratory Patterns After LPS-Induced Epigenetic Reprogramming</i>
	Stanca M. Ciupe, Brittany P. Boribong, Sarah Kadelka and Caroline N. Jones

- 212** *DEEPsc: A Deep Learning-Based Map Connecting Single-Cell Transcriptomics and Spatial Imaging Data*
Floyd Maseda, Zixuan Cang and Qing Nie
- 226** *Exact Probability Landscapes of Stochastic Phenotype Switching in Feed-Forward Loops: Phase Diagrams of Multimodality*
Anna Terebus, Farid Manuchehrfar, Youfang Cao and Jie Liang
- 242** *A Modeling and Machine Learning Pipeline to Rationally Design Treatments to Restore Neuroendocrine Disorders in Heterogeneous Individuals*
Tongli Zhang



Compressive Sensing Inference of Neuronal Network Connectivity in Balanced Neuronal Dynamics

Victor J. Barranca^{1*} and Douglas Zhou^{2,3,4*}

¹ Department of Mathematics and Statistics, Swarthmore College, Swarthmore, PA, United States, ² School of Mathematical Sciences, Shanghai Jiao Tong University, Shanghai, China, ³ Ministry of Education Key Laboratory of Scientific and Engineering Computing, Shanghai Jiao Tong University, Shanghai, China, ⁴ Institute of Natural Sciences, Shanghai Jiao Tong University, Shanghai, China

OPEN ACCESS

Edited by:

Chunhe Li,
Fudan University, China

Reviewed by:

Changsong Zhou,
Hong Kong Baptist University,
Hong Kong
Mainak Jignesh Patel,
College of William & Mary,
United States

*Correspondence:

Victor J. Barranca
vbarran1@swarthmore.edu
Douglas Zhou
zdz@sjtu.edu.cn

Specialty section:

This article was submitted to
Systems Biology,
a section of the journal
Frontiers in Neuroscience

Received: 18 August 2019

Accepted: 30 September 2019

Published: 17 October 2019

Citation:

Barranca VJ and Zhou D (2019)
Compressive Sensing Inference of
Neuronal Network Connectivity in
Balanced Neuronal Dynamics.
Front. Neurosci. 13:1101.
doi: 10.3389/fnins.2019.01101

Determining the structure of a network is of central importance to understanding its function in both neuroscience and applied mathematics. However, recovering the structural connectivity of neuronal networks remains a fundamental challenge both theoretically and experimentally. While neuronal networks operate in certain dynamical regimes, which may influence their connectivity reconstruction, there is widespread experimental evidence of a balanced neuronal operating state in which strong excitatory and inhibitory inputs are dynamically adjusted such that neuronal voltages primarily remain near resting potential. Utilizing the dynamics of model neurons in such a balanced regime in conjunction with the ubiquitous sparse connectivity structure of neuronal networks, we develop a compressive sensing theoretical framework for efficiently reconstructing network connections by measuring individual neuronal activity in response to a relatively small ensemble of random stimuli injected over a short time scale. By tuning the network dynamical regime, we determine that the highest fidelity reconstructions are achievable in the balanced state. We hypothesize the balanced dynamics observed *in vivo* may therefore be a result of evolutionary selection for optimal information encoding and expect the methodology developed to be generalizable for alternative model networks as well as experimental paradigms.

Keywords: neuronal networks, balanced networks, signal processing, network dynamics, connectivity reconstruction

1. INTRODUCTION

The connectivity of neuronal networks is fundamental for establishing the link between brain structure and function (Boccaletti et al., 2006; Stevenson et al., 2008; Gomez-Rodriguez et al., 2012); however, recovering the structural connectivity in neuronal networks is still a challenging problem both theoretically and experimentally (Salinas and Sejnowski, 2001; Song et al., 2005; Friston, 2011; Kleinfeld et al., 2011; Bargmann and Marder, 2013). Recent experimental advances, such as diffusion tensor imaging (DTI), dense electron microscopy (EM), and highly resolved tracer injections, have facilitated improved measurement of network connectivity, but constructing complete neuronal wiring diagrams for networks of thousands or more neurons is currently infeasible due largely to the small spatial scale and the dense packing of nervous tissue (Lichtman and Denk, 2011; Sporns, 2011; Briggman and Bock, 2012; Markov et al., 2013). Likewise, modern

mathematical approaches for recovering network connectivity based on measured neuronal activity, such as Granger causality, information theory, and Bayesian analysis, typically demand linear dynamics or long observation times (Aertsen et al., 1989; Sporns et al., 2004; Timme, 2007; Eldawlatly et al., 2010; Friston, 2011; Hutchison et al., 2013; Zhou et al., 2013b, 2014; Goñi et al., 2014). Is it possible to achieve the successful reconstruction of network connectivity from the measurement of individual non-linear neuronal dynamics within a short time scale?

To address this central question, we develop a novel theoretical framework for the recovery of neuronal connectivity based on both network sparsity and balanced dynamics. Sparse connectivity among neurons is widely observed on large (inter-cortical) and small (local circuit) spatial scales (Mason et al., 1991; Markram et al., 1997; Achard and Bullmore, 2007; He et al., 2007; Ganmor et al., 2011), and, therefore, the amount of observed activity required to reconstruct network connectivity may be significantly smaller than suggested by estimates using only the total network size. *Compressive sensing* (CS) theory has emerged as a useful methodology for sampling and reconstructing sparse signals (Candes et al., 2006; Donoho, 2006; Gross et al., 2010; Wang et al., 2011b) and has primarily been utilized in estimating the connectivity of linear or time-invariant network models (Hu et al., 2009; Mishchenko and Paninski, 2012). In the case of realistic neuronal networks, their non-linear dynamics in time pose a major conceptual difficulty, particularly in isolating the impact of direct network connections on recorded activity from the effects of indirect neuronal interactions and the external drive.

We demonstrate that the reconstruction of neuronal connectivity based on compressive sensing of non-linear network dynamics is indeed possible in an appropriately *balanced dynamical regime* in which fluctuations in neuronal input largely drive firing events. Numerous experimental studies demonstrate that neuronal firing events are generally irregular, with large excitatory and inhibitory inputs dynamically balanced such that the voltage of a neuron typically resides near the resting potential for a broad class of external stimulation (Britten et al., 1993; Shadlen and Newsome, 1998; Compte et al., 2003; Haider et al., 2006; Tan and Wehr, 2009; London et al., 2010; Runyan et al., 2010; Isaacson and Scanziani, 2011; Xue et al., 2014). Theoretical work corroborates the existence of this operating regime for balanced network models in which neurons are sparsely connected while strongly coupled, such that neuronal activity is highly variable and heterogeneous across the network (van Vreeswijk and Sompolinsky, 1996, 1998; Troyer and Miller, 1997; Vogels and Abbott, 2005; Miura et al., 2007; Mongillo et al., 2012). Here, we utilize the same binary-state network model as such previous studies and demonstrate that using a small ensemble of random inputs and corresponding time-averaged measurements of neuronal dynamics collected over a short time scale, it is possible to achieve high fidelity reconstructions of recurrent connectivity for sparsely connected networks of excitatory and inhibitory neurons. We show that the quality of this reconstruction improves as the network dynamics are further balanced, expecting that for physiological networks, once in the balanced state, CS-based estimates of network connectivity

are feasible. We hypothesize that the balanced operating regime may have arisen in sensory systems from evolution as a means of optimally encoding both connectivity and stimulus information through network dynamics.

2. RESULTS

2.1. Compressive Sensing of Balanced Dynamics

To investigate the reconstruction of neuronal network connectivity in the balanced state, we consider a mechanistic binary-state model with non-linear dynamics (van Vreeswijk and Sompolinsky, 1996, 1998). The model network is composed of N neurons, such that N_E neurons are excitatory (E) and N_I neurons are inhibitory (I). The state of the i th neuron in the k th population ($k = E, I$) at time t is prescribed by

$$\sigma_k^i(t) = H(\mu_k^i(t) - \theta_k), \quad (1)$$

where $H(\cdot)$ denotes the Heaviside function and θ_k is the firing threshold for the neurons in population k . The total synaptic drive $\mu_k^i(t)$ into the i th neuron in the k th population at time t is

$$\mu_k^i(t) = \sum_{j=1}^{N_E} R_{kE}^{ij} \sigma_E^j(t) + \sum_{j=1}^{N_I} R_{kI}^{ij} \sigma_I^j(t) + (\mathbf{Fp})_k^i, \quad (2)$$

where R_{kl}^{ij} denotes the connection strength between the i th post-synaptic neuron in the k th population and the j th pre-synaptic neuron in the l th population ($l = E, I$), and $(\mathbf{Fp})_k^i$ is the total external input into the i th neuron in the k th population. The connection strength R_{kl}^{ij} is chosen to be $R_{kl}/\sqrt{K}$ with probability K/N_l and 0 otherwise. In this case, the excitatory connection strength $R_{kE} > 0$ and the inhibitory connection strength $R_{kI} < 0$. Since each neuron is expected to receive projections from K pre-synaptic excitatory neurons and K pre-synaptic inhibitory neurons, sparse connectivity is reflected by the assumption that $K \ll N_E, N_I$. In advancing the model dynamics for each neuron, the mean time between subsequent updates is $\tau_E = 10$ ms for excitatory neurons and $\tau_I = 9$ ms for inhibitory neurons, reflecting experimental estimates of cortical membrane potential time constants (McCormick et al., 1985; van Vreeswijk and Sompolinsky, 1996; Shelley et al., 2002). Based on the total synaptic drive at each time the system is updated, a given neuron is either in a quiescent ($\sigma_k^i(t) = 0$) or firing ($\sigma_k^i(t) = 1$) state.

To partition the model across the two subpopulations, the neurons and their corresponding activity variables may also be indexed from $l = 1, \dots, N$, with the first N_E indices corresponding to neurons in the excitatory population and the second N_I indices corresponding to neurons in the inhibitory population. Using this choice of indexing, $\mathbf{R}$ is the $N \times N$ recurrent connectivity matrix and $\mathbf{p}$ is the N -vector of static external inputs for the network. The external input $\mathbf{p}$ is selected such that $(\mathbf{Fp})_k^i$ is $\mathcal{O}(\sqrt{K})$ for each neuron, thereby comparable to the total synaptic drive from each population. Analogously, the feed-forward connectivity matrix $\mathbf{F}$ is $N \times N$ and diagonal, such that diagonal entries $F_{ii} = f_E$ for $i = 1, \dots, N_E$ and $F_{ii} = f_I$ for

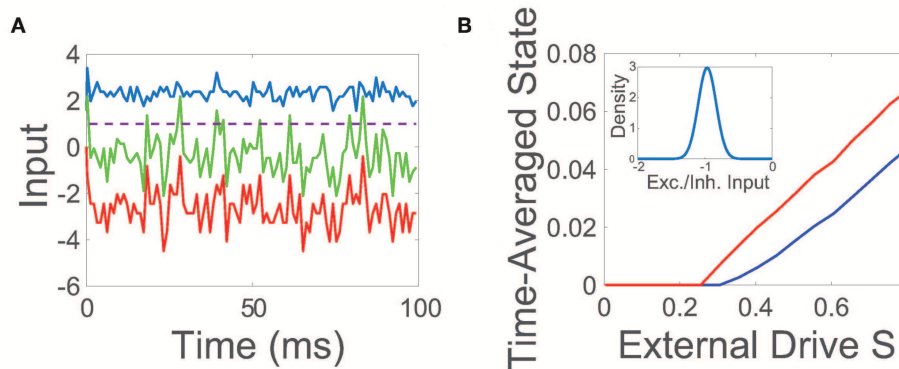


FIGURE 1 | Balanced network dynamics. **(A)** Excitatory (blue), inhibitory (red), and net (green) inputs into a sample excitatory neuron in a balanced network. The dashed line indicates the firing threshold. **(B)** The time-averaged state of the excitatory (blue) and inhibitory (red) population as a function of external input scaling strength S . Inset: Probability density of the time-averaged ratio of excitatory and inhibitory inputs across the network. Unless otherwise specified, parameters utilized are $R_{EE} = R_{IE} = 1$, $R_{II} = -1.8$, $R_{EI} = -2$, $N_E = 800$, $N_I = 200$, $f_E = 1.2$, $f_I = 1$, $K = 0.03N_E$, $\theta_E = 1$, and $\theta_I = 0.7$. The external drive into a neuron in the k^{th} population is an excitatory constant current that is independently generated and identically uniformly-distributed with $\mathcal{O}(\sqrt{K})$ magnitude scaled by f_k .

$i = N_E + 1, \dots, N_E + N_I$, thereby scaling the relative external input strength for each respective population. Since the absolute scale of the neuronal input is inconsequential in this non-dimensional model, we assume connectivity parameters $R_{EE} = R_{IE} = 1$, so the primary parameters that determine the inhibition relative to excitation are the post-synaptic connection strengths for the inhibitory neurons and the external input strengths.

Since R_{kl} as well as θ_k are $\mathcal{O}(1)$ and the external drive is $\mathcal{O}(\sqrt{K})$, if the excitatory and inhibitory inputs are not balanced, the total synaptic drive is $\mathcal{O}(\sqrt{K})$ and thus each neuron either fires with an excessively high rate or remains nearly quiescent. In the balanced operating regime, however, the excitatory and inhibitory inputs instead dynamically cancel and produce physiological firing dynamics, leaving the mean synaptic input nearly vanishing with relatively large $\mathcal{O}(1)$ input fluctuations responsible for the exact timing of firing events and their irregular distribution. This leads to theoretical conditions on the connection strength parameters (van Vreeswijk and Sompolinsky, 1996, 1998):

$$\frac{f_E}{f_I} > \frac{R_{EI}}{R_{II}} > 1. \quad (3)$$

The net input into a representative neuron in the balanced state is plotted in **Figure 1A**, demonstrating a dynamic tracking of excitatory and inhibitory inputs such that the mean total input is far below threshold. On the larger scale of the entire network, an equilibrium between excitation and inhibition is also achieved in the balanced regime, with the time-averaged mean of the ratio between the excitatory and inhibitory input (E/I input ratio) across the network narrowly distributed near -1 .

While the sparsity of $\mathbf{R}$ in principle reduces the necessary data for a successful reconstruction of the network connectivity, compressive sensing theory generally only applies to the recovery of sparse inputs into linear and time-invariant systems (Candes et al., 2006; Donoho, 2006), rather than from measurements of the non-linear and time-evolving dynamics of a neuronal

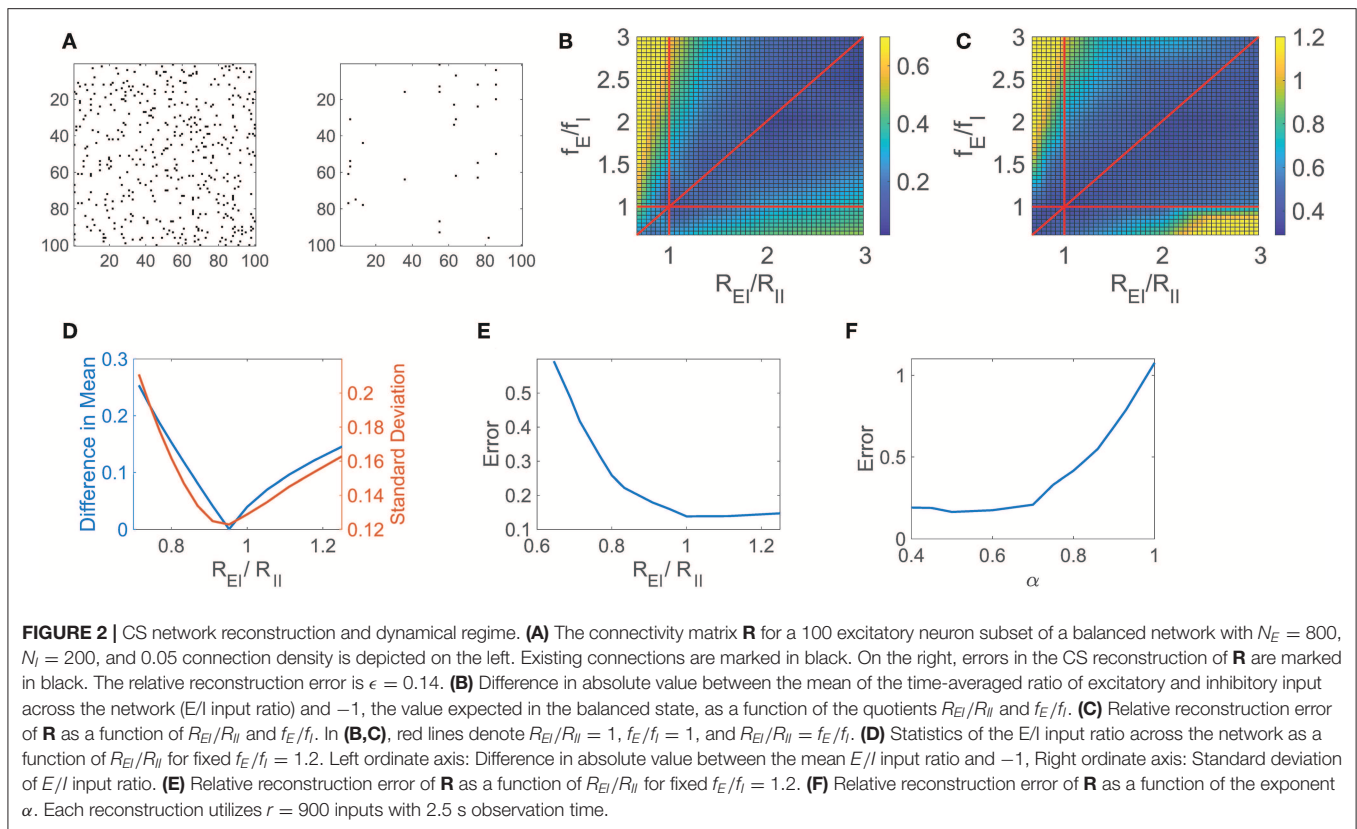
network. To overcome this theoretical challenge, it is important to note that for a broad class of physiological neurons as well as realistic neuron models, the neuronal firing activity exhibits linear dependence on relatively strong external inputs in the proper dynamical regime (Brunel and Latham, 2003; Rauch et al., 2003; Fourcaud-Trocmé and Brunel, 2005; La Camera et al., 2006; Barranca et al., 2014a). Considering the dynamic balance between excitatory and inhibitory inputs facilitates a rapid and robust linear response to external inputs (van Vreeswijk and Sompolinsky, 1996, 1998), we hypothesize that balanced neuronal network dynamics are critical to the efficient CS reconstruction of sparse network connectivity.

For the binary-state balanced network model, the temporal expectation of Equation (2) yields a natural linear input-output mapping in response to a single input vector $\mathbf{p}$

$$\boldsymbol{\mu} = \mathbf{R}\mathbf{x} + \mathbf{F}\mathbf{p}, \quad (4)$$

where $\boldsymbol{\mu}$ is an N -dimensional vector denoting the time-averaged total input into each neuron and $\mathbf{x}$ is an N -dimensional vector denoting the time-averaged state of each neuron.

To demonstrate the generality of our network reconstruction framework with respect to external inputs and to avoid specializing their design, we drive the network with an ensemble of r random input vectors with independent identically uniformly distributed elements, denoted by $\{\mathbf{p}^{(i)}\}_{i=1}^r$, and measure the evoked time-averaged net input and state of the neurons, denoted by $\{\boldsymbol{\mu}^{(i)}\}_{i=1}^r$ and $\{\mathbf{x}^{(i)}\}_{i=1}^r$, respectively, over a short time duration. From a physiological standpoint, on a given trial, we inject into each neuron a distinct constant current of magnitude determined by a uniformly distributed random variable and measure the evoked dynamics across the network, subsequently reconstructing the network connectivity from a linear mapping relating these quantities. To facilitate efficient recovery, the number of trials utilized $r \ll N^2$. Here the N^2 entries of $\mathbf{R}$ are to be recovered using only Nr state measurements,



leading to a highly underdetermined inverse problem. However, since $\mathbf{R}$ is sparse, CS may still potentially yield a successful reconstruction (see the Methods section for details).

While conventional balanced network theory assumes a constant and homogeneous excitatory external input is injected into each population (van Vreeswijk and Sompolinsky, 1996, 1998), note that we choose the excitatory external input vector $\mathbf{p}$ to be composed of independent and identically distributed random variables. Even for these heterogeneous external inputs, balanced dynamics are still well-maintained under population scalings with $f_E > f_I$. The maintenance of balance across the majority of the network can be seen in the inset of **Figure 1B**, plotting the mean E/I input ratio across the network, which closely resembles the distribution for the $\mathcal{O}(\sqrt{K})$ constant homogeneous input case in its narrow peak near -1 (van Vreeswijk and Sompolinsky, 1996, 1998; Barranca et al., 2019). To further probe the evoked network dynamics, we empirically examine the response of the network to increasingly large random external inputs in **Figure 1B**, adjusting scaled external input $S\mathbf{p}$ by increasing the scaling strength S . We observe that as the external drive strength is increased, the time-averaged state of both the excitatory and inhibitory populations intensifies linearly with S for sufficiently large inputs, thereby demonstrating linear gain in agreement with Equation (4) and as expected theoretically in the large network limit in the case of homogeneous external inputs (van Vreeswijk and Sompolinsky, 1996, 1998).

With linear input-output mapping (4), we obtain a system of equations relating the network input, evoked dynamics, and the

connectivity structure of $\mathbf{R}$. To recover the i^{th} row of $\mathbf{R}$ in this case, denoted $\mathbf{R}_{i*}$, it is necessary to utilize the full set of inputs, $\mathbf{P} = [\mathbf{p}^{(1)} \dots \mathbf{p}^{(r)}]$, the respective time-averaged inputs into the i^{th} neuron, $\mathbf{U}_i = [\mu_i^{(1)} \dots \mu_i^{(r)}]$, and the respective evoked time-averaged states of the i^{th} neuron, $\mathbf{X}_i = [x_i^{(1)} \dots x_i^{(r)}]$.

The resultant underdetermined linear system in recovering the i^{th} row, $\mathbf{R}_{i*}$ of the recurrent connectivity matrix is

$$\mathbf{R}_{i*}\mathbf{X} = \mathbf{U}_i - (\mathbf{F}\mathbf{P})_{i*}. \quad (5)$$

Since $\mathbf{R}$ is sparse and the respective average states in $\mathbf{X}$ are approximately uncorrelated in the balanced regime (van Vreeswijk and Sompolinsky, 1996, 1998), the optimal row reconstruction is the solution to Equation (5) with minimal L_1 norm (Candes et al., 2006; Donoho, 2006) in accordance with CS theory. Considering the resultant L_1 minimization problem is solvable in polynomial time (Donoho and Tsaig, 2008) and since Equation (5) represents a sequence of independent linear systems with respect to the row index i , parallelization furnishes a computationally efficient reconstruction of $\mathbf{R}$.

In **Figure 2A**, we consider a sparsely connected network with balanced dynamics and 0.05 connection density, and reconstruct its connectivity matrix composed of $N^2 = 10^6$ entries using Equation (5) for $i = 1, \dots, N$, recording the network response to $r = 900$ random inputs for 2.5 s each. The connectivity matrix for a subset of 100 excitatory neurons is depicted alongside the corresponding reconstruction

error, demonstrating that the majority of connections, or lack thereof, are indeed captured. Improving significantly upon preexisting approaches for reconstructing network connectivity, which commonly require long observation times and focus primarily on excitatory networks (Timme, 2007; Eldawlatly et al., 2010; Hutchison et al., 2013; Zhou et al., 2013b; Goñi et al., 2014), this reconstruction framework successfully distinguishes between excitatory and inhibitory connection types over short observation times. Since the neuron types are not assumed to be known *a priori*, we note that while there is no constraint that excitatory and inhibitory connections are of the appropriate sign directly enforced in solving optimization problem (5) via L_1 minimization, with sufficiently rich measurements of the network dynamics, the connectivity reconstructions nevertheless are generally able to successfully identify both connection signs and magnitudes, as indicated by the small relative error obtained in recovering the connectivity matrix.

To quantify the accuracy of the entire connectivity matrix reconstruction, $\mathbf{R}^{\text{recon}}$, we measure the relative reconstruction error, $\epsilon = \|\mathbf{R} - \mathbf{R}^{\text{recon}}\| / \|\mathbf{R}\|$, using the Frobenius norm, $\|\mathbf{R}\| = \sqrt{\sum_i \sum_j R_{ij}^2}$. In this particular case, utilizing significantly less trials than entries in $\mathbf{R}$, the network relative reconstruction error is only $\epsilon = 0.14$, yielding close agreement with the original connection matrix. We remark that in this network the ratio of excitatory to inhibitory neurons is chosen to be 4:1 in agreement with estimates in the primary visual cortex (Gilbert, 1992; Liu, 2004; Cai et al., 2005; Zhou et al., 2013a), though this framework is adaptable to other distributions of neuron types corresponding to alternative cortical regions. While in this work we specifically consider the role of balanced dynamics in the context of an analytically tractable binary-state model setting, the compressive sensing reconstruction framework naturally generalizes to alternative model networks. In the case of the integrate-and-fire model (Lapicque, 1907; Burkitt, 2006; Mather et al., 2009; Barranca et al., 2014a), for example, rather than requiring detailed knowledge of the networks' inputs as in the binary-state model, the network input-output mapping may instead involve the time-averaged neuronal membrane potentials and firing rates (Barranca et al., 2014b), yielding a framework that is more amenable to experimental settings.

2.2. Balanced Network Characteristics for Optimal Reconstruction

We posit that the network functioning in the balanced operating regime is fundamental to the success of the CS reconstruction and demonstrate that the relative reconstruction error indeed increases as the network departs from the balanced state. We confirm the central role of the balanced state in network reconstruction by varying several network connectivity parameters, which crucially determine the network operating state, and examining the resultant impact on the CS reconstruction of $\mathbf{R}$.

For the network dynamics to be appropriately balanced in the large K limit, Equation (3) gives restrictions on the external and cortical input strengths for the network. These

parameter restrictions hold approximately for the sparsely-connected networks of large yet finite size that we examine, and we analyze the impact of these parameters on the network reconstruction accuracy. Since we are considering the connectivity reconstruction for networks composed of a finite number of neurons and therefore Equation (3) only holds approximately, in many cases the dynamics may be well-balanced even though the corresponding theoretical condition in the large network limit is violated (van Vreeswijk and Sompolinsky, 1996; Gu et al., 2018). For this reason, to gauge the degree to which a finite-sized network exhibits balanced dynamics, we analyze the absolute difference between the mean E/I input ratio for all neurons and -1 , the expected value for balanced dynamics, as depicted in **Figure 2B** across network parameters. Here we vary the quotients, R_{EI}/R_{II} and f_E/f_I , which are each crucial to Equation (3), observing a clear region of well-balanced dynamics. Investigating the impact of the network dynamical regime on the CS reconstruction of $\mathbf{R}$, we plot in **Figure 2C** the corresponding relative reconstruction error over the same parameter space. The highest quality reconstructions are generally achieved when the mean E/I input ratio is near -1 , and the network is consequently in the balanced operating regime, with degradation in accuracy incurred as the mean E/I input ratio departs from -1 .

Similarly, we examine a detailed one-dimensional slice of these plots in **Figures 2D,E**, respectively, as we fix $f_E/f_I = 1.2$ and vary the quotient R_{EI}/R_{II} . In particular, we plot the absolute difference between the mean E/I input ratio and -1 as well as the standard deviation of the E/I input ratio across the network in **Figure 2D** to further classify the network operating state. We observe that when Equation (3) is approximately satisfied, the difference between the mean E/I input ratio and -1 is small. In this same regime, the standard deviation of the E/I input ratio is also near zero, indicating a dynamic balance between the excitatory and inhibitory inputs over the entire network. For nearly identical parameter choices as those producing balanced dynamics, we observe that the corresponding relative reconstruction error, depicted in **Figure 2E**, is minimal. As the reconstruction accuracy diminishes, increasingly large proportions of neurons remain either active with unrealistically high firing rate or are completely quiescent. Since the relatively rare and irregular threshold crossings due to input fluctuations in the balanced regime largely reflect the impact of the network connectivity on dynamics, nearly frozen or excessively high neuronal activity results in significantly diminished reconstruction quality.

Another crucial assumption in formulating the balanced network model is strong synapses. Similar models could be formulated with connection strengths of form R_{kl}/K^α . However, the dynamics are only well-balanced in the large K limit for $\alpha = 1/2$. For $1/2 < \alpha \leq 1$, the weaker synapse case, the temporal input fluctuations decrease with K , scaling as $K^{1/2-\alpha}$, leading to mean-driven dynamics in the large K limit. In contrast, for $0 < \alpha < 1/2$, the stronger synapse case, input fluctuations instead grow with K , and thus the net input wildly fluctuates well above and below threshold.

Using our CS framework to reconstruct the network connectivity $\mathbf{R}$, we examine the reconstruction error achieved for

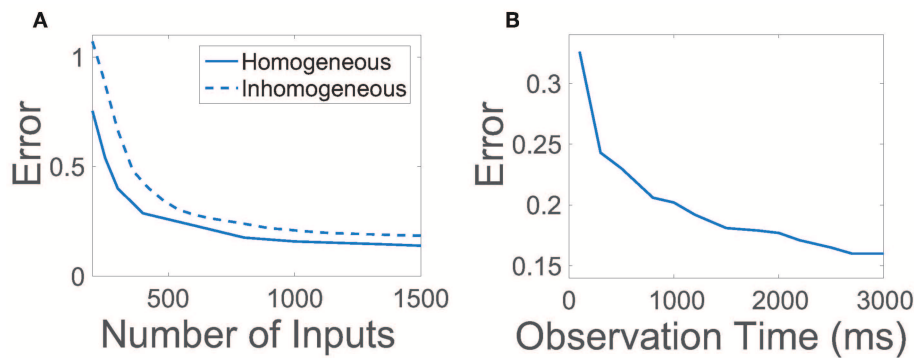


FIGURE 3 | Efficiency and robustness of CS network reconstruction. **(A)** Relative reconstruction error as a function of the number of random input vectors r utilized. Solid line depicts error using homogeneous thresholds $\theta_E = 1, \theta_I = 0.7$. Dashed line depicts error using inhomogeneous thresholds such that $\theta_k^i = \theta_k + \delta_k^i$, where inhomogeneities $\delta_k^i \in (0, d)$ are uniformly distributed random variables and $d = 0.3$. In each case, the observation time is 2.5 s. **(B)** Relative reconstruction error as a function of observation time. In each case, 900 random input vectors are utilized.

the network model initialized across choices of α in **Figure 2F** while fixing K and the remaining model parameters. The optimal reconstruction is achieved near $\alpha = 0.5$, when the network is in the balanced operating regime, with error generally increasing as α moves away from 0.5 and the mean E/I input ratio deviates from balance. Note that while here we study the impact of α for network realizations with a fixed and finite choice of K , the theoretical considerations in the large network limit suggest that these effects become more pronounced for larger networks with correspondingly larger K . Considering that the reconstruction error increases especially rapidly as $\alpha \rightarrow 1$, we hypothesize that weaker synapses in non-balanced network models are not conducive to the reconstruction of network connectivity, particularly in the mean-field limit. While mean-driven dynamics generally well encode information regarding network inputs and feed-forward connectivity (Barranca et al., 2016b), in this case, we instead observe that the balanced dynamical regime is better suited for encoding recurrent interactions in the network dynamics.

2.3. Robustness of Connectivity Reconstruction

For efficiency, it is desirable to achieve an accurate reconstruction of the network connections using a relatively small number of random inputs and also by collecting the evoked network activity over a small observation time. In **Figure 3A**, we plot the relative reconstruction error for $\mathbf{R}$ as the number of input vectors is increased given a fixed observation time. Initially, as the number of inputs is increased, the error rapidly decreases. Once the number of inputs utilized is sufficiently large, near ~ 800 , more marginal improvements are garnered, at which point additional experiments are of less utility. Hence, only a relatively small number of trials are necessary to yield near-maximum reconstruction quality.

Given a sufficient number of inputs such that the reconstruction error saturates, we next examine the duration of time over which data must be recorded for successful connectivity reconstruction in **Figure 3B**. The relative

reconstruction error precipitously drops for small observation times, leveling off for sufficiently large time durations over 2 s. Thus, for each set of inputs utilized, it is only necessary to record neuronal activity over a short time duration.

Similar dependence on input ensemble size and observation time holds for networks of alternative sizes with analogous connection density and architecture, yielding comparably accurate reconstructions by using a relatively small number of input vectors. As the irregular dynamics of neurons in the balanced state is crucial to the success of the CS recovery framework, we note that regardless of the observation time and number of inputs utilized, reconstruction of $\mathbf{R}$ remains intractable if the network dynamics are not sufficiently well-balanced.

In our original binary-state model, we had assumed that all excitatory neurons and inhibitory neurons are statistically homogeneous. We now examine the effect of inhomogeneity in the network on the reconstruction of $\mathbf{R}$ by varying the firing threshold for each neuron. In this case, thresholds are chosen such that the firing threshold for the i th neuron in the k th population is $\theta_k^i = \theta_k + \delta_k^i$, with inhomogeneities prescribed by identically uniformly distributed random variables $\delta_k^i \in (0, d)$. In **Figure 3A**, we plot the reconstruction error dependence on the number of random inputs for inhomogeneity strength $d = 0.3 \approx 0.43\theta_I$, observing only a minor degradation in reconstruction quality relative to the homogeneous threshold case. Thus, we expect that even if a network is composed of neurons of many types, as long as the neuronal dynamics are robustly balanced, it is possible to still utilize our CS framework to reconstruct the network connectivity.

3. DISCUSSION

Addressing the current theoretical and experimental difficulties in measuring the structural connectivity in large neuronal networks, we show that the high degree of sparsity in network connections makes it feasible to accurately reconstruct network

connectivity from a relatively small number of measurements of evoked neuronal activity via CS theory. The success of this reconstruction depends on the dynamical regime of the network, with the balanced operating state facilitating optimal recovery. Just as the connectivity matrix $\mathbf{R}$ may be recovered from dynamical activity based on an underlying linear mapping, such as Equation (4), unknown network feed-forward connectivity as well as natural stimuli may analogously be reconstructed (Barranca et al., 2016b). We have empirically verified such reconstructions are also improved when the network is in the balanced operating regime. In light of this, we hypothesize that evolution may have fine-tuned much of the cortical network connectivity to optimize both the encoding of sensory inputs as well as local connectivity based on balanced network dynamics.

It is important to note that while the compressive sensing theory leveraged in this work is well-suited for the reconstruction of sparse signals, the reconstruction of densely-connected neuronal networks in the brain with potentially strongly correlated dynamics remains a challenging area for future investigation (Wang et al., 2011a; Markov et al., 2013; Yang et al., 2017). Though we considered a balanced network model with statistically homogeneous random connectivity among neuron types, physiological neuronal circuits observed in experiment typically exhibit a complex network structure (Massimini et al., 2005; Bonifazi et al., 2009; Markov et al., 2013), which may induce stronger correlations and oscillations in neuronal dynamics (Honey et al., 2007; Wang et al., 2011a; Yang et al., 2017). While prolonged synchronous dynamics may make it infeasible to reconstruct network connections using our methodology, intermixed periods of irregular dynamics may provide sufficient neuronal interaction data or, otherwise, connections between functional modules may be potentially identifiable. Recent theoretical analysis demonstrates that even for networks with small-world or scale-free structure, balanced dynamics can persist in these neuronal networks with various types of single-neuron dynamics, particularly in an embedded active core of neurons hypothesized to play a key role in sparse coding (Gu et al., 2018). For such a balanced core in a network with heterogeneous connectivity, the primary dynamical assumptions of our reconstruction framework hold as does compressive sensing theory in the presence of mildly structured sampling matrices (Elad, 2007; Barranca et al., 2016a; Adcock et al., 2017), and thus it may be possible to extend our framework in recovering connections within the balanced core.

While this work utilized specific modeling choices for which the balanced state is well-characterized, in alternative settings, a similar framework can potentially be utilized to reconstruct sparse network connectivity as long as the dynamics are in the balanced operating regime. Linear mappings in the balanced state similar to Equation (4) have been well-established for various classes of neuronal network models, including those with more physiological dynamics (Brunel and Latham, 2003; Fourcaud-Trocme and Brunel, 2005; Barranca et al., 2014a, 2019; Gu et al., 2018), and experimental measurements of neuronal firing-activity also generally exhibit a similar linear

dependence on input strength (Rauch et al., 2003; La Camera et al., 2006). Advances in multiple neuron recording, such as multiple-electrode technology, optical recording with fast voltage-sensitive dyes, and light-field microscopy, have facilitated the recording of increasingly large numbers of neurons simultaneously (Stevenson and Kording, 2011; Prevedel et al., 2014; Frost et al., 2015), and combined with new optogenetic as well as optochemical techniques for precisely stimulating specific neurons (Banghart et al., 2004; Rickgauer et al., 2014; Packer et al., 2015), we expect the theoretical framework developed to be generalizable by combining these techniques in experiment. To circumvent potential experimental difficulties in simultaneously stimulating specific neurons and recording their evoked dynamics, we expect it to be also possible to extend our theoretical framework by driving a subset of neurons and recording the response of a random group of neurons in each trial. Since a particular subnetwork of neurons in the brain generally receives inhomogeneous and unknown input from external neurons, the development and utilization of an accurate input-output mapping involving only the recorded network dynamics and applied drive in experiments marks a key area for future exploration. While there are known mappings that make no assumption of the detailed input into each neuron, they do assume that external inputs are fully characterized. Since such mappings are quite robust in the presence of noise (Barranca et al., 2014b,c, 2016b), it may still be possible to well discern recurrent connections even in the presence of unknown external neuronal inputs for sufficiently strong forcing applied in experiments.

4. METHODS

4.1. Compressive Sensing Theory

Compressive sensing theory states that for sparse data, the number of measurements required for a successful reconstruction in a static and linear system is determined by the number of dominant non-zero components in the data (Candes et al., 2006; Donoho, 2006). Using this reasoning, optimally reconstructing sparse data from a small number of samples requires selecting the sparsest reconstruction consistent with the measured data, since such a signal is most compressible. CS theory thus provides a significant improvement in sampling efficiency from the conventional Shannon-Nyquist theorem, which asserts that the sampling rate should instead be determined by the full bandwidth of the data (Shannon, 1949).

The reconstruction of time-invariant data from a small number of samples in a linear system can be considered an underdetermined inverse problem. For an n -component signal, $\mathbf{y}$, m discrete samples of $\mathbf{y}$ can be represented by $\mathbf{A}\mathbf{y}$, where $\mathbf{A}$ is an $m \times n$ measurement matrix composed of rows which are each a set of measurement weights. This yields an m -component measured signal, $\mathbf{b}$. Reconstructing the true data $\mathbf{y}$ from the measured data $\mathbf{b}$ is therefore equivalent to solving

$$\mathbf{A}\mathbf{y} = \mathbf{b}. \quad (6)$$

When the number of samples taken is significantly smaller than the number of components in $\mathbf{y}$, i.e., $m \ll n$, the above system is highly underdetermined with an infinite number of possible solutions. While one approach to selecting the most compressible solution is to choose the sparsest $\mathbf{y}$ satisfying Equation (6), this is generally too computationally expensive for real-world signals.

For sufficiently sparse $\mathbf{y}$ and a broad class of measurement matrices, CS theory shows that a viable surrogate is in fact

minimizing $\|\mathbf{y}\|_{L_1} = \sum_{i=1}^n |y_i|$ (Candes and Wakin, 2008),

which is efficiently solvable in polynomial time using numerous algorithms (Tropp and Gilbert, 2007; Donoho and Tsaig, 2008). From an experimental standpoint, it is relatively straightforward to devise sampling schemes such that the corresponding measurement matrices are amenable to CS. Measurement matrices exhibiting randomness in their structure are particularly viable candidates (Baraniuk, 2007; Candes and Wakin, 2008; Barranca et al., 2016a), and, consequently, the response matrix $\mathbf{X}$ in the left-hand side of Equation (5) is suited for CS reconstructions in the balanced regime since $\mathbf{X}$ demonstrates little correlation among its entries.

REFERENCES

- Achard, S., and Bullmore, E. (2007). Efficiency and cost of economical brain functional networks. *PLoS Comput. Biol.* 3:e17. doi: 10.1371/journal.pcbi.0030017
- Adcock, B., Hansen, A. C., Poon, C., and Roman, B. (2017). "Breaking the coherence barrier: a new theory for compressed sensing," in *Forum of Mathematics, Sigma*, Vol. 5. (Cambridge, UK: Cambridge University Press).
- Aertsen, A. M., Gerstein, G. L., Habib, M. K., and Palm, G. (1989). Dynamics of neuronal firing correlation: modulation of "effective connectivity". *J. Neurophysiol.* 61, 900–917.
- Banghart, M., Borges, K., Isacoff, E., Trauner, D., and Kramer, R. H. (2004). Light-activated ion channels for remote control of neuronal firing. *Nat. Neurosci.* 7, 1381–1386. doi: 10.1038/nn1356
- Baraniuk, R. (2007). Compressive sensing. *IEEE Signal Process. Mag.* 24, 118–120. doi: 10.1109/MSP.2007.4286571
- Bargmann, C. I., and Marder, E. (2013). From the connectome to brain function. *Nat. Methods* 10, 483–490. doi: 10.1038/nmeth.2451
- Barranca, V. J., Huang, H., and Li, S. (2019). The impact of spike-frequency adaptation on balanced network dynamics. *Cogn. Neurodyn.* 13, 105–120. doi: 10.1007/s11571-018-9504-2
- Barranca, V. J., Johnson, D. C., Moyher, J. L., Sauppe, J. P., Shkarayev, M. S., Kovačič, G., et al. (2014a). Dynamics of the exponential integrate-and-fire model with slow currents and adaptation. *J. Comput. Neurosci.* 37, 161–180. doi: 10.1007/s10827-013-0494-0
- Barranca, V. J., Kovačič, G., Zhou, D., and Cai, D. (2014b). Network dynamics for optimal compressive-sensing input-signal recovery. *Phys. Rev. E* 90:042908. doi: 10.1103/PhysRevE.90.042908
- Barranca, V. J., Kovačič, G., Zhou, D., and Cai, D. (2014c). Sparsity and compressed coding in sensory systems. *PLoS Comput. Biol.* 10:e1003793. doi: 10.1371/journal.pcbi.1003793
- Barranca, V. J., Kovačič, G., Zhou, D., and Cai, D. (2016a). Improved compressive sensing of natural scenes using localized random sampling. *Sci. Rep.* 6:31976. doi: 10.1038/srep31976
- Barranca, V. J., Zhou, D., and Cai, D. (2016b). Compressive sensing reconstruction of feed-forward connectivity in pulse-coupled nonlinear networks. *Phys. Rev. E* 93:060201. doi: 10.1103/PhysRevE.93.060201

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the manuscript/supplementary files.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This work was supported by NSF DMS-1812478 (VB), by a Swarthmore Faculty Research Support Grant (VB), by NSFC-11671259, NSFC-11722107, NSFC-91630208, SJTU-UM Collaborative Research Program, and by the Student Innovation Center at Shanghai Jiao Tong University (DZ).

ACKNOWLEDGMENTS

We thank the late David Cai for initial thoughtful discussions of this work.

- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D.-U. (2006). Complex networks: structure and dynamics. *Phys. Rep.* 424, 175–308. doi: 10.1016/j.physrep.2005.10.009
- Bonifazi, P., Goldin, M., Picardo, M. A., Jorquera, I., Cattani, A., Bianconi, G., et al. (2009). GABAergic hub neurons orchestrate synchrony in developing hippocampal networks. *Science* 326, 1419–1424. doi: 10.1126/science.1175509
- Briggman, K. L., and Bock, D. D. (2012). Volume electron microscopy for neuronal circuit reconstruction. *Curr. Opin. Neurobiol.* 22, 154–161. doi: 10.1016/j.conb.2011.10.022
- Britten, K. H., Shadlen, M. N., Newsome, W. T., and Movshon, J. A. (1993). Responses of neurons in macaque MT to stochastic motion signals. *Vis. Neurosci.* 10, 1157–1169. doi: 10.1017/S0952523800010269
- Brunel, N., and Latham, P. (2003). Firing rate of the noisy quadratic integrate-and-fire neuron. *Neural Comp.* 15, 2281–2306. doi: 10.1162/089976603322362365
- Burkitt, A. N. (2006). A review of the integrate-and-fire neuron model: I. Homogeneous synaptic input. *Biol. Cybern.* 95, 1–19. doi: 10.1007/s00422-006-0068-6
- Cai, D., Rangan, A. V., and McLaughlin, D. W. (2005). Architectural and synaptic mechanisms underlying coherent spontaneous activity in V1. *Proc. Natl. Acad. Sci. U.S.A.* 102, 5868–5873. doi: 10.1073/pnas.0501913102
- Candes, E. J., Romberg, J. K., and Tao, T. (2006). Stable signal recovery from incomplete and inaccurate measurements. *Commun. Pur. Appl. Math.* 59, 1207–1223. doi: 10.1002/cpa.20124
- Candes, E. J., and Wakin, M. B. (2008). An introduction to compressive sampling. *Signal Process. Mag.* 25, 21–30. doi: 10.1109/MSP.2007.914731
- Compte, A., Constantinidis, C., Tegner, J., Raghavachari, S., Chafee, M. V., Goldman-Rakic, P. S., et al. (2003). Temporally irregular mnemonic persistent activity in prefrontal neurons of monkeys during a delayed response task. *J. Neurophysiol.* 90, 3441–3454. doi: 10.1152/jn.00949.2002
- Donoho, D. L. (2006). Compressed sensing. *IEEE Trans. Inform. Theory* 52, 1289–1306. doi: 10.1109/TIT.2006.871582
- Donoho, D. L., and Tsaig, Y. (2008). Fast solution of l_1 -norm minimization problems when the solution may be sparse. *IEEE Trans. Inform. Theory* 54, 4789–4812. doi: 10.1109/TIT.2008.929958
- Elad, M. (2007). Optimized projections for compressed sensing. *IEEE Trans. Signal Process.* 55, 5695–5702. doi: 10.1109/TSP.2007.900760

- Eldawlatly, S., Zhou, Y., Jin, R., and Oweiss, K. G. (2010). On the use of dynamic Bayesian networks in reconstructing functional neuronal networks from spike train ensembles. *Neural Comput.* 22, 158–189. doi: 10.1162/neco.2009.11-08-900
- Fourcaud-Trocmé, N., and Brunel, N. (2005). Dynamics of the instantaneous firing rate in response to changes in input statistics. *J. Comput. Neurosci.* 18, 311–321. doi: 10.1007/s10827-005-0337-8
- Friston, K. J. (2011). Functional and effective connectivity: a review. *Brain Connect.* 1, 13–36. doi: 10.1089/brain.2011.0008
- Frost, W. N., Brandon, C. J., Bruno, A. M., Humphries, M. D., Moore-Kochlacs, C., Sejnowski, T. J., et al. (2015). Monitoring spiking activity of many individual neurons in invertebrate ganglia. *Adv. Exp. Med. Biol.* 859, 127–145. doi: 10.1007/978-3-319-17641-3_5
- Ganmor, E., Segev, R., and Schneidman, E. (2011). The architecture of functional interaction networks in the retina. *J. Neurosci.* 31, 3044–3054. doi: 10.1523/JNEUROSCI.3682-10.2011
- Gilbert, C. D. (1992). Horizontal integration and cortical dynamics. *Neuron* 9, 1–13.
- Gomez-Rodriguez, M., Leskovec, J., and Krause, A. (2012). Inferring networks of diffusion and influence. *T. Knowl. Discov. D.* 5:21. doi: 10.1145/2086737.2086741
- Goni, J., van den Heuvel, M. P., Avena-Koenigsberger, A., Velez de Mendizabal, N., Betzel, R. F., Griffa, A., et al. (2014). Resting-brain functional connectivity predicted by analytic measures of network communication. *Proc. Natl. Acad. Sci. U.S.A.* 111, 833–838. doi: 10.1073/pnas.1315529111
- Gross, D., Liu, Y. K., Flammia, S. T., Becker, S., and Eisert, J. (2010). Quantum state tomography via compressed sensing. *Phys. Rev. Lett.* 105:150401. doi: 10.1103/PhysRevLett.105.150401
- Gu, Q. L., Li, S., Dai, W. P., Zhou, D., and Cai, D. (2018). Balanced active core in heterogeneous neuronal networks. *Front. Comput. Neurosci.* 12:109. doi: 10.3389/fncom.2018.00109
- Haider, B., Duque, A., Hasenstaub, A. R., and McCormick, D. A. (2006). Neocortical network activity *in vivo* is generated through a dynamic balance of excitation and inhibition. *J. Neurosci.* 26, 4535–4545. doi: 10.1523/JNEUROSCI.5297-05.2006
- He, Y., Chen, Z. J., and Evans, A. C. (2007). Small-world anatomical networks in the human brain revealed by cortical thickness from MRI. *Cereb. Cortex* 17, 2407–2419. doi: 10.1093/cercor/bhl149
- Honey, C. J., Kötter, R., Breakspear, M., and Sporns, O. (2007). Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proc. Natl. Acad. Sci. U.S.A.* 104, 10240–10245. doi: 10.1073/pnas.0701519104
- Hu, T., Leonardo, A., and Chklovskii, D. (2009). “Reconstruction of sparse circuits using multi-neuronal excitation (RESCUME),” in *Advances in Neural Information Processing Systems* 22 (Vancouver, BC), 790–798.
- Hutchison, R. M., Womelsdorf, T., Allen, E. A., Bandettini, P. A., Calhoun, V. D., Corbetta, M., et al. (2013). Dynamic functional connectivity: promise, issues, and interpretations. *Neuroimage* 80, 360–378. doi: 10.1016/j.neuroimage.2013.05.079
- Isaacson, J. S., and Scanziani, M. (2011). How inhibition shapes cortical activity. *Neuron* 72, 231–243. doi: 10.1016/j.neuron.2011.09.027
- Kleinfeld, D., Bharioke, A., Blinder, P., Bock, D. D., Briggman, K. L., Chklovskii, D. B., et al. (2011). Large-scale automated histology in the pursuit of connectomes. *J. Neurosci.* 31, 16125–16138. doi: 10.1523/JNEUROSCI.4077-11.2011
- La Camera, G., Rauch, A., Thurbon, D., Lüscher, H. R., Senn, W., and Fusi, S. (2006). Multiple time scales of temporal response in pyramidal and fast spiking cortical neurons. *J. Neurophysiol.* 96, 3448–3464. doi: 10.1152/jn.00453.2006
- Lapicque, L. (1907). Recherches quantitatives sur l’excitation électrique des nerfs traitée comme une polarisation. *J. Physiol. Pathol. Gen.* 9, 620–635.
- Lichtman, J. W., and Denk, W. (2011). The big and the small: challenges of imaging the brain’s circuits. *Science* 334, 618–623. doi: 10.1126/science.1209168
- Liu, G. (2004). Local structural balance and functional interaction of excitatory and inhibitory synapses in hippocampal dendrites. *Nat. Neurosci.* 7, 373–379. doi: 10.1038/nn1206
- London, M., Roth, A., Beeren, L., Häusser, M., and Latham, P. E. (2010). Sensitivity to perturbations *in vivo* implies high noise and suggests rate coding in cortex. *Nature* 466, 123–127. doi: 10.1038/nature09086
- Markov, N. T., Ercsey-Ravasz, M., Van Essen, D. C., Knoblauch, K., Toroczkai, Z., and Kennedy, H. (2013). Cortical high-density counterstream architectures. *Science* 342:1238406. doi: 10.1126/science.1238406
- Markram, H., Lübke, J., Frotscher, M., Roth, A., and Sakmann, B. (1997). Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex. *J. Physiol.* 500, 409–440.
- Mason, A., Nicoll, A., and Stratford, K. (1991). Synaptic transmission between individual pyramidal neurons of the rat visual cortex *in vitro*. *J. Neurosci.* 11, 72–84.
- Massimini, M., Ferrarelli, F., Huber, R., Esser, S. K., Singh, H., and Tononi, G. (2005). Breakdown of cortical effective connectivity during sleep. *Science* 309, 2228–2232. doi: 10.1126/science.1117256
- Mather, W., Bennett, M. R., Hasty, J., and Tsimring, L. S. (2009). Delay-induced degrade-and-fire oscillations in small genetic circuits. *Phys. Rev. Lett.* 102:068105. doi: 10.1103/PhysRevLett.102.068105
- McCormick, D. A., Connors, B. W., Lighthall, J. W., and Prince, D. A. (1985). Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex. *J. Neurophysiol.* 54, 782–806. doi: 10.1152/jn.1985.54.4.782
- Mishchenko, Y., and Paninski, L. (2012). A Bayesian compressed-sensing approach for reconstructing neural connectivity from subsampled anatomical data. *J. Comput. Neurosci.* 33, 371–388. doi: 10.1007/s10827-012-0390-z
- Miura, K., Tsubo, Y., Okada, M., and Fukai, T. (2007). Balanced excitatory and inhibitory inputs to cortical neurons decouple firing irregularity from rate modulations. *J. Neurosci.* 27, 13802–13812. doi: 10.1523/JNEUROSCI.2452-07.2007
- Mongillo, G., Hansel, D., and van Vreeswijk, C. (2012). Bistability and spatiotemporal irregularity in neuronal networks with nonlinear synaptic transmission. *Phys. Rev. Lett.* 108:158101. doi: 10.1103/PhysRevLett.108.158101
- Packer, A. M., Russell, L. E., Dalgleish, H. W., and Häusser, M. (2015). Simultaneous all-optical manipulation and recording of neural circuit activity with cellular resolution *in vivo*. *Nat. Methods* 12, 140–146. doi: 10.1038/nmeth.3217
- Prevedel, R., Yoon, Y. G., Hoffmann, M., Pak, N., Wetzstein, G., Kato, S., et al. (2014). Simultaneous whole-animal 3D imaging of neuronal activity using light-field microscopy. *Nat. Methods* 11, 727–730. doi: 10.1038/nmeth.2964
- Rauch, A., La Camera, G., Lüscher, H.-R., Senn, W., and Fusi, S. (2003). Neocortical pyramidal cells respond as integrate-and-fire neurons to *in vivo*-like input currents. *J. Neurophysiol.* 90, 1598–1612. doi: 10.1152/jn.00293.2003
- Rickgauer, J. P., Deisseroth, K., and Tank, D. W. (2014). Simultaneous cellular-resolution optical perturbation and imaging of place cell firing fields. *Nat. Neurosci.* 17, 1816–1824. doi: 10.1038/nn.3866
- Runyan, C. A., Schummers, J., Van Wart, A., Kuhlman, S. J., Wilson, N. R., Huang, Z. J., et al. (2010). Response features of parvalbumin-expressing interneurons suggest precise roles for subtypes of inhibition in visual cortex. *Neuron* 67, 847–857. doi: 10.1016/j.neuron.2010.08.006
- Salinas, E., and Sejnowski, T. J. (2001). Correlated neuronal activity and the flow of neural information. *Nat. Rev. Neurosci.* 2, 539–550. doi: 10.1038/35086012
- Shadlen, M. N., and Newsome, W. T. (1998). The variable discharge of cortical neurons: implications for connectivity, computation and information coding. *J. Neurosci.* 18, 3870–3896.
- Shannon, C. E. (1949). Communication in the Presence of Noise. *Proc. IRE* 37, 10–21.
- Shelley, M., McLaughlin, D., Shapley, R., and Wielaard, J. (2002). States of high conductance in a large-scale model of the visual cortex. *J. Comp. Neurosci.* 13, 93–109. doi: 10.1023/A:1020158106603
- Song, S., Sjöström, P. J., Reigl, M., Nelson, S., and Chklovskii, D. B. (2005). Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biol.* 3:e68. doi: 10.1371/journal.pbio.0030068
- Sporns, O. (2011). The human connectome: a complex network. *Ann. N. Y. Acad. Sci.* 1224, 109–125. doi: 10.1016/j.tics.2004.07.008
- Sporns, O., Chialvo, D. R., Kaiser, M., and Hilgetag, C. C. (2004). Organization, development and function of complex brain networks. *Trends Cognit. Sci.* 8, 418–425. doi: 10.1038/nn.2731
- Stevenson, I. H., and Kording, K. P. (2011). How advances in neural recording affect data analysis. *Nat. Neurosci.* 14, 139–142. doi: 10.1038/nn.2731

- Stevenson, I. H., Rebesco, J. M., Miller, L. E., and Kording, K. P. (2008). Inferring functional connections between neurons. *Curr. Opin. Neurobiol.* 18, 582–588. doi: 10.1016/j.conb.2008.11.005
- Tan, A. Y., and Wehr, M. (2009). Balanced tone-evoked synaptic excitation and inhibition in mouse auditory cortex. *Neuroscience* 163, 1302–1315. doi: 10.1016/j.neuroscience.2009.07.032
- Timme, M. (2007). Revealing network connectivity from response dynamics. *Phys. Rev. Lett.* 98:224101. doi: 10.1103/PhysRevLett.98.224101
- Tropp, J. A., and Gilbert, A. C. (2007). Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit. *IEEE Trans. Inform. Theory* 53, 4655–4666. doi: 10.1109/TIT.2007.909108
- Troyer, T. W., and Miller, K. D. (1997). Physiological gain leads to high ISI variability in a simple model of a cortical regular spiking cell. *Neural Comput.* 9, 971–983.
- van Vreeswijk, C., and Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* 274, 1724–1726.
- van Vreeswijk, C., and Sompolinsky, H. (1998). Chaotic balanced state in a model of cortical circuits. *Neural Comput.* 15, 1321–1371.
- Vogels, T. P., and Abbott, L. F. (2005). Signal propagation and logic gating in networks of integrate-and-fire neurons. *J. Neurosci.* 25, 10786–10795. doi: 10.1523/JNEUROSCI.3508-05.2005
- Wang, S. J., Hilgetag, C. C., and Zhou, C. (2011a). Sustained activity in hierarchical modular neural networks: self-organized criticality and oscillations. *Front. Comput. Neurosci.* 5:30. doi: 10.3389/fncom.2011.00030
- Wang, W.-X., Lai, Y.-C., Grebogi, C., and Ye, J. (2011b). Network reconstruction based on evolutionary-game data via compressive sensing. *Phys. Rev. X* 1:021021. doi: 10.1103/PhysRevX.1.021021
- Xue, M., Atallah, B. V., and Scanziani, M. (2014). Equalizing excitation-inhibition ratios across visual cortical neurons. *Nature* 511, 596–600. doi: 10.1038/nature13321
- Yang, D. P., Zhou, H. J., and Zhou, C. (2017). Co-emergence of multi-scale cortical activities of irregular firing, oscillations and avalanches achieves cost-efficient information capacity. *PLoS Comput. Biol.* 13:e1005384. doi: 10.1371/journal.pcbi.1005384
- Zhou, D., Rangan, A., McLaughlin, D., and Cai, D. (2013a). Spatiotemporal dynamics of neuronal population response in the primary visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* 110, 9517–9522. doi: 10.1073/pnas.1308167110
- Zhou, D., Xiao, Y., Zhang, Y., Xu, Z., and Cai, D. (2013b). Causal and structural connectivity of pulse-coupled nonlinear networks. *Phys. Rev. Lett.* 111:054102. doi: 10.1103/PhysRevLett.111.054102
- Zhou, D., Xiao, Y., Zhang, Y., Xu, Z., and Cai, D. (2014). Granger causality network reconstruction of conductance-based integrate-and-fire neuronal systems. *PLoS ONE* 9:e87636. doi: 10.1371/journal.pone.0087636

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Barranca and Zhou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Revealing Dynamic Mechanisms of Cell Fate Decisions From Single-Cell Transcriptomic Data

Jiajun Zhang^{1,2}, Qing Nie^{3,4} and Tianshou Zhou^{1,2*}

¹ School of Mathematics, Sun Yat-Sen University, Guangzhou, China, ² Guangdong Province Key Laboratory of Computational Science and School of Mathematics and Computational Science, Sun Yat-Sen University, Guangzhou, China, ³ Department of Developmental and Cell Biology, University of California, Irvine, Irvine, CA, United States, ⁴ Department of Mathematics, University of California, Irvine, Irvine, CA, United States

OPEN ACCESS

Edited by:

Chunhe Li,
Fudan University, China

Reviewed by:

Lin Wan,
Academy of Mathematics and
Systems Science (CAS), China
Xiuwei Zhang,
Georgia Institute of Technology,
United States

*Correspondence:

Tianshou Zhou
mcszhtsh@mail.sysu.edu.cn

Specialty section:

This article was submitted to
Systems Biology,
a section of the journal
Frontiers in Genetics

Received: 18 August 2019

Accepted: 21 November 2019

Published: 23 December 2019

Citation:

Zhang J, Nie Q and Zhou T (2019)
Revealing Dynamic Mechanisms of
Cell Fate Decisions From Single-Cell
Transcriptomic Data.
Front. Genet. 10:1280.
doi: 10.3389/fgene.2019.01280

Cell fate decisions play a pivotal role in development, but technologies for dissecting them are limited. We developed a multifunction new method, Topographer, to construct a “quantitative” Waddington’s landscape of single-cell transcriptomic data. This method is able to identify complex cell-state transition trajectories and to estimate complex cell-type dynamics characterized by fate and transition probabilities. It also infers both marker gene networks and their dynamic changes as well as dynamic characteristics of transcriptional bursting along the cell-state transition trajectories. Applying this method to single-cell RNA-seq data on the differentiation of primary human myoblasts, we not only identified three known cell types, but also estimated both their fate probabilities and transition probabilities among them. We found that the percent of genes expressed in a bursty manner is significantly higher at (or near) the branch point (~97%) than before or after branch (below 80%), and that both gene-gene and cell-cell correlation degrees are apparently lower near the branch point than away from the branching. Topographer allows revealing of cell fate mechanisms in a coherent way at three scales: cell lineage (macroscopic), gene network (mesoscopic), and gene expression (microscopic).

Keywords: cell fate decision, single-cell data, developmental landscape, cell-type dynamics, cellular process

INTRODUCTION

Multi-cell organisms start as a single cell that matures through complex dynamic processes involving multiple cell fate decision points, leading to functionally different cell types, many of which have yet to be defined (Trapnell, 2015). While cellular processes such as proliferation, differentiation, and reprogramming are governed by complex gene regulatory programs, each cell makes its own fate decisions by integrating a wide array of signals and executing a complex choreography of gene regulatory changes (Moris et al., 2016; Tanay and Regev, 2017). Since the structure of a multi-cell tissue is tightly linked with its function (Perié et al., 2015), elucidating the integrative (from gene to cell) mechanism of cell fate decisions is crucial yet challenging.

Single-cell measurement technologies (Svensson et al., 2017; Ziegenhain et al., 2017) which can simultaneously measure the expressions of many genes in a large number of single cells, provide an unprecedented opportunity to elucidate developmental pathways and dissect cell fate decisions.

Several algorithms [see a recent review (Saelens et al., 2019)] have been developed to organize single cells in pseudo-temporal order based on transcriptomic divergence and cell-state classification. It has been a major challenge to illuminate the dynamic mechanisms of cellular programs governing fate transitions from single-cell data that lacks temporal information (Trapnell, 2015). The current methods have mainly focused on identifying trajectories between the most phenotypically distant cell states, and they are usually less robust in reconstructing trajectories from early states towards intermediate or transitory cell states [e.g., Wishbone (Setty et al., 2016), Diffusion Pseudotime (Haghverdi et al., 2016), Cyclus (Gut et al., 2015), and CellRouter (Lummertz da Rocha et al., 2018)]. Some of the methods have focused on gaining insights into the regulatory mechanisms driving cell differentiation [e.g., Monocle (Trapnell et al., 2014), ERA (Kafri et al., 2013), Waterfall (Shin et al., 2015), and PIDC (Chan et al., 2017)], and they seem not to consider how discontinuous, stochastic fate transition events are driven by the dynamic nature of the developmental landscape (which can change in response to activity of gene regulatory networks and extracellular signals) and reflected in the observed increased transcriptional heterogeneity at transition points. In all the existing methods, cell-type dynamics are mainly characterized qualitatively, providing little quantitative information on in-depth characterization of complex cellular ecosystems involving cell fate decisions. For a system of multiple cell fate decision points, it has been difficult for the current methods to estimate cell types and their transitions. How fate transitions in the single cell data are related to cell-state gene regulatory networks and the characteristics of transcriptional bursting remains largely unknown.

To overcome the above challenges and to address the important issues on cell fate decisions, we developed *Topographer*, an integrated pipeline. It first constructs a data-driven “quantitative” (i.e., each cell is endowed with spatiotemporal information) developmental landscape, which provides a global view for differentiation processes together with the cartoon landscapes (Waddington, 1957) and the model-driven landscapes (Wang et al., 2011; Li and Wang, 2013; Li and Wang, 2013; Li and Wang, 2014; Li and Wang, 2015), and then reveals stochastic dynamics of cell types by estimating both their fate probabilities and transition probabilities among them, and infers dynamic characteristics of transcriptional bursting kinetics along the identified developmental trajectory. In addition, it can also both identify various branched (e.g., bi- and tri-) cell-state transition trajectories with multiple branching points from single-cell data and infer networks of marker genes and their pseudo-temporal changes. Together, *Topographer* enables construction of complex cell lineages, resolving intermediate developmental stages, and revealing multilayer mechanisms of cell fate decisions in a coherent way at three different levels: cell lineage, gene network, and transcriptional burst (referring to **Supplementary Figure 1**).

We demonstrated effectiveness of *Topographer* by analyzing single-cell RNA-seq data on the differentiation of primary human myoblasts (Trapnell et al., 2014) while showing

applications to other examples in **Supplementary Information**. We first identified three known cell types: proliferating cells, differentiating myoblasts, and interstitial mesenchymal cells, and then constructed a quantitative developmental landscape where each cell is endowed with spatiotemporal information. Furthermore, by estimating the fate probabilities of the identified cell types and transition probabilities among them, we found that the probability of transition from the proliferating cell type to the interstitial mesenchymal cell type was approximately twice that of transition from the former to the differentiating myoblast type, and that the fate probability of the differentiating myoblast type was approximately equal to that of the interstitial mesenchymal cell type. We also found that the relative number of the genes expressed in a bursty manner was apparently higher at (or near) the branch point (~97%) than before or after branch (below 80%). In addition, the mean burst size (MBS)/mean burst frequency (MBF) monotonically decreased/decreased before branch but monotonically increased/decreased after branch, with the identified trajectories.

RESULTS

The Outline of Topographer

In order to infer the stochastic dynamics of cell fate decisions from single-cell transcriptomic data, *Topographer* makes the following assumption about the data: the information on the entire development process is adequate, or a snapshot of primary tissue represents a complete developmental process. The data need pre-processing (**Supplementary Information** for detail) so that *Topographer* achieve a good performance. The overall *Topographer*, a multifaceted single-cell analysis platform, comprises five functional modules: (Trapnell, 2015) the backbone module (**Figure 1B**); (Tanay and Regev, 2017) the landscape module (**Figure 1C**); (Moris et al., 2016) the dynamics module (**Figure 1D**); (Perié et al., 2015) the network module (**Figure 1E**); and (Svensson et al., 2017) the burst module (**Figure 1F**). The backbone module is independent of the remaining 4 modules that depends on the former since they make use of information on cell-state transition trajectories identified in the first module. All the five modules are logically related but each module achieves an independent function.

Two important notes on this method are (Trapnell, 2015) *Topographer* is unsupervised and needs no prior knowledge of specific genes that distinguish cell fates, and is thus suitable for studying a wide array of dynamic processes involving fate transitions. (Tanay and Regev, 2017) Except for the backbone module, the other four functional modules only use the pseudotime information derived in that module (*Materials and Methods*), so they can also use the result on pseudo-temporal ordering of single cells obtained by other existing algorithms (Saelens et al., 2019) to achieve their respective purposes. However, the backbone module is established based on a different approach (see the following content for details), and has its own advantages, e.g., it can identify not only cell-state

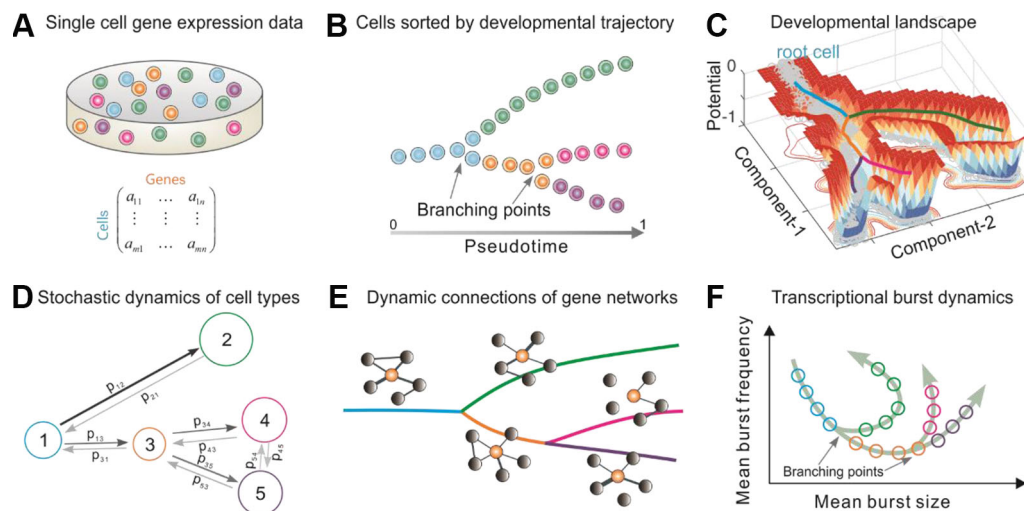


FIGURE 1 | Overview of *Topographer*. *Topographer* comprises five functional modules with each (**B**, **C**, **D**, **E**, or **F**) achieving an independent function. (**A**) Single-cell data are represented by a matrix. (**B**) The backbone module identifies the main cell trajectories from the data. (**C**) The landscape module constructs a quantitative Waddington's landscape where each cell is endowed with spatiotemporal information (*Materials and Methods*), and the thick colored lines represent the backbone of cell trajectories identified in the backbone module. This panel is not schematic, but is plotted using an artificial set of data generated by a toy model (*Supplementary Eq. (24)*). (**D**) The dynamics module reveals stochastic dynamics of cell types by estimating the fate probabilities of cell types and the transition probabilities (indicated by symbols) among them (*Materials and Methods*), where numbers 1–5 represent cell types, the size of circle represents that of fate probability, and the thickness of line with arrow represents the size of transition probability. (**E**) The network module infers marker gene networks and their changes along the identified cell trajectories (or along the pseudotime), where the orange ball represents a marker gene, and the thickness of connection line represents the strength of correlation. (**F**) The burst module infers dynamic characteristics of transcriptional bursting kinetics (characterized by both burst size and burst frequency) along the pseudotime, where arrows represent the pseudotime direction.

transition trajectories with multiple branching points, but also intermediate or transitory cell states.

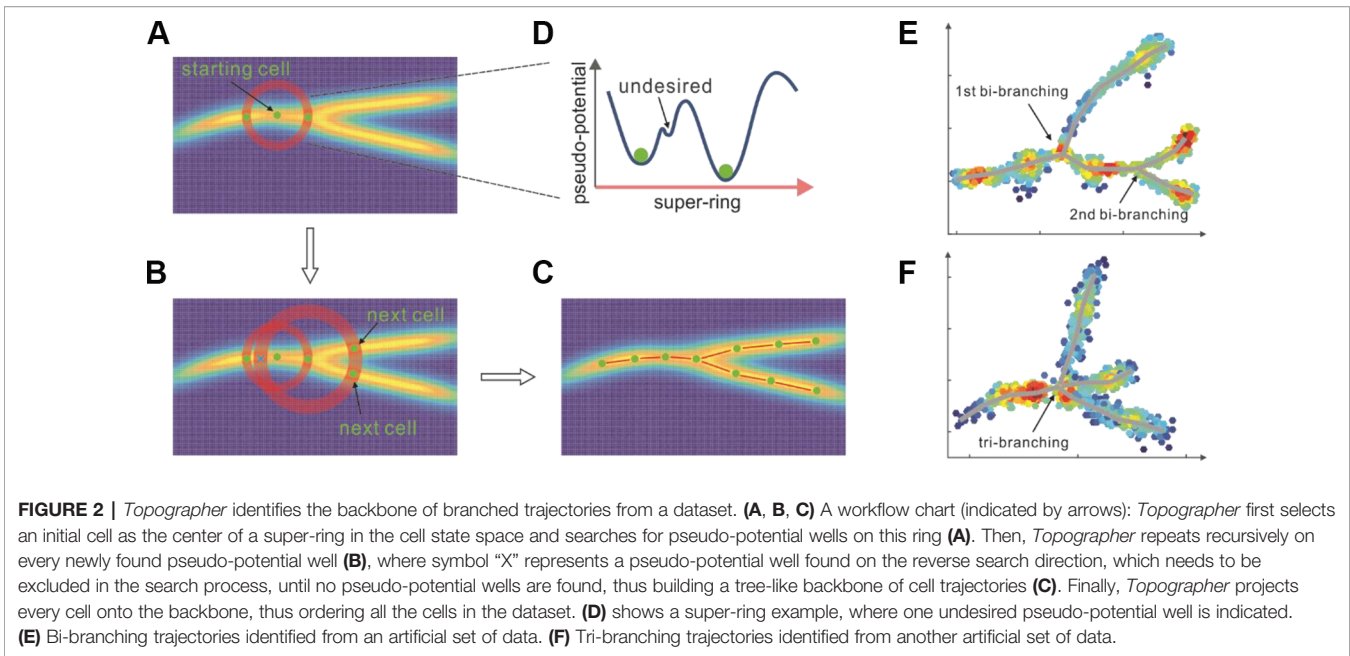
Below we introduce each of the five functional modules separately (*Materials and Methods* give more details and *Supplementary Information* provides a complete description).

Identifying the Backbone of Cell Trajectories From Single-Cell Data

The backbone module is a fast and local pseudo-potential-based algorithm. Here the pseudo-potential is defined as the negative of the logarithm of a local density function (Eq. (1), *Materials and Methods*), which aims to identify the “backbone” (i.e., “planimetric” contour) of cell-state transition trajectories cross development and find valley floors in a developmental landscape from single-cell data.

Starting from an initial cell (**Figure 2A**) selected either based on the global minimal pseudo-potential or the prior knowledge, *Topographer* calculates an adaptive step (*Supplementary Eq. (5)*) and searches for pseudo-potential wells (i.e., “pits” where pseudo-potentials are relatively lower) on a super-ring (i.e., a high-dimensional circular tube, referring to **Figure 2A**, which shows a flatten super-ring) centered at this initial cell and with the radius equal to the step length (also **Figure 2A**). In this search method, which clusters cells on super-rings, cluster centers are characterized by a lower pseudo-potential than their neighbors and by a relatively larger distance from points with lower pseudo-potentials (e.g., the only two pseudo-potential wells with “green ball” in **Figure 2D** are desired),

providing the basis of a procedure to find pseudo-potential wells on a super-ring. In this procedure, the number of pseudo-potential wells arises naturally, outliers are automatically spotted, and pseudo-potential wells are recognized regardless of their shape and the dimensionality of the space in which they are embedded. We stress that although there is an analogy between our method and a density-based approach developed originally by Rodriguez and colleagues (Rodriguez and Laio, 2014), the difference is that the former is carried out on a super-ring rather than in the full cell state space. Clearly, if the number of the found pseudo-potential wells (but not including the one found on the “reverse” search direction) is more than one, this implies the occurrence of branch. The segments linking the center and the newly found pseudo-potential well/or wells on the super-ring can be viewed as part/or parts of the entire developmental trajectory. Similar processes are repeated recursively on sequential super-rings along search directions until no new pseudo-potential wells are found (**Figure 2B**). By linking all the centers and all the pseudo-potential wells found on super-rings, *Topographer* thus builds a tree-like developmental backbone (**Figure 2C**). Note that the identified backbone is actually a projection of the pseudo-potential landscape. By projecting every cell onto this backbone (see subsection *Cell Projection and Pseudotime Assignment*, *Materials and Methods*) and by selecting a root node in the tree (e.g., based on the prior knowledge), *Topographer* thus orders all the single cells in the dataset, and equips each cell with a pseudotime if this root node is set as an



initial moment (without loss of generality, the full pseudotime may be set as the interval between 0 and 1).

Figures 2E, F showed respectively a doubly bi-branched trajectory identified from one simulated dataset and a tri-branched trajectory identified from another artificial set of data. **Figure 3A** below demonstrated a two-dimensional projection of the *de novo* cell trajectories identified from single-cell RNA-seq data on the differentiation of primary human myoblasts and **Figure 3B** demonstrated the evolutions of five marker genes (MYOG, MYF5, MYH2, CDK1, and MEF2C) with branches along the pseudotime. **Supplementary Figure 12** and **Figure 14** demonstrated results of other two examples, which further showed the power of *Topographer* in pseudo-temporally ordering single cells in single-cell data.

Because of its ability to find pseudo-potential wells on super-rings, *Topographer* can identify *de novo* developmental trajectories with non-, bi-, and multi-branches (referring to **Figures 1E, F**) (note: a low resolution of experimentally sampling data may lead to tri-branches).

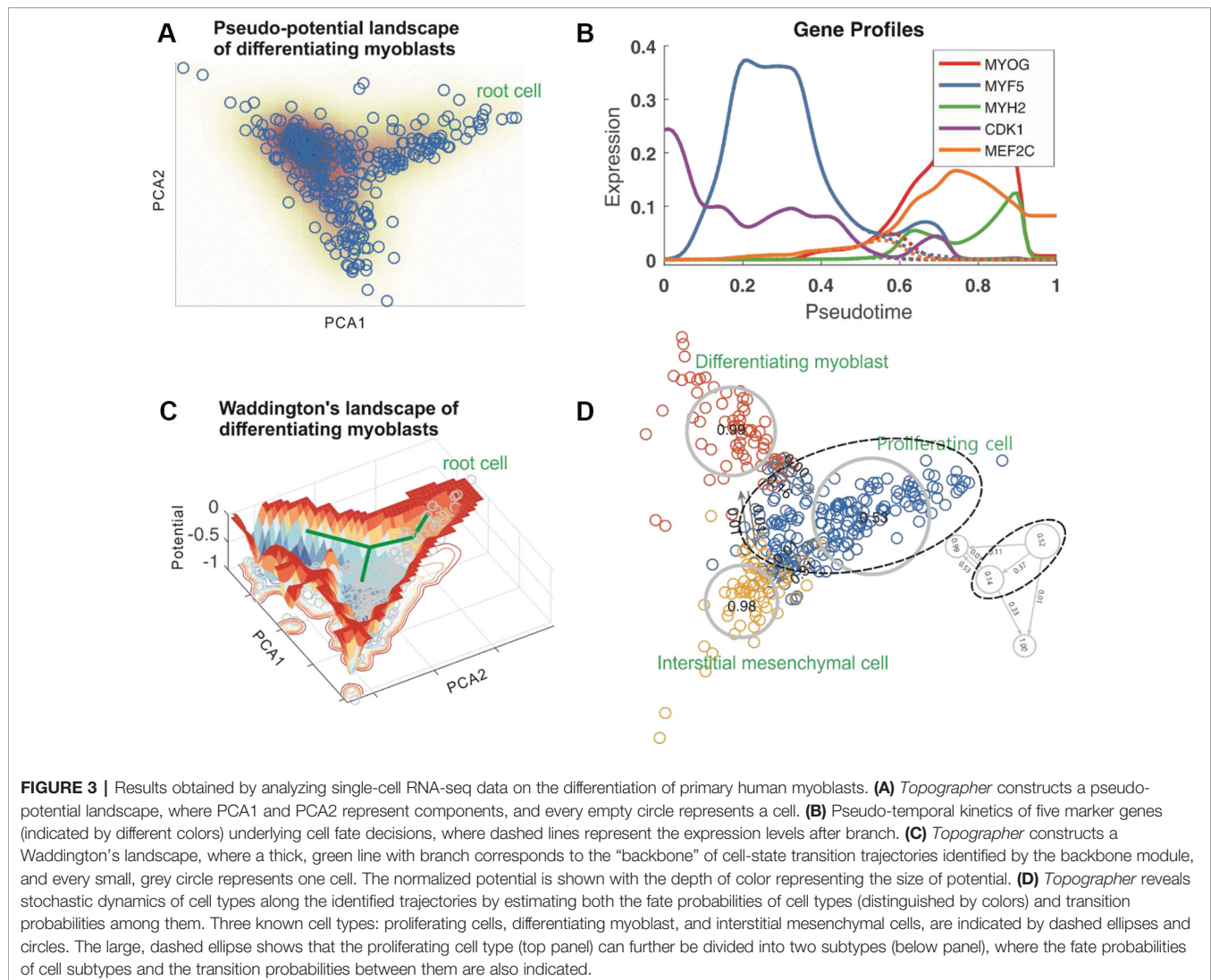
Constructing a Quantitative Developmental Landscape of Single-Cell Data

The backbone module used pseudo-potentials to construct the contour of cell-state transition trajectories, which extracted the information on both branch and cellular ordering from single-cell data. Note that this kind of potential would not correctly reflect transitions between cells since the probability fluxes would exist between them due to cell division, cell death and/or other factors, and have been quantified from gene network models (Li and Wang, 2014). For example, precursor cells should in principle have higher pseudo-potentials (Eq. (8), *Materials and Methods*) in a developmental landscape in contrast to their

generations, but if the precursor cells have higher densities, they have lower pseudo-potentials. Apparently, both are inconsistent. In addition, pseudo-potential lacks the temporal information on differentiation or development.

Because of both the above shortcoming of pseudo-potential and the intuition of the Waddington's developmental landscape (in fact, it has been extensively viewed as a powerful metaphor for how differentiated cell types emerge from a single, totipotent cell 1), the landscape module (an algorithm) is designed to construct a “stereometric” developmental landscape (by “stereometric” we mean that each cell is loaded with spatiotemporal information) in contrast to the “planimetric” contour identified by the backbone module. This constructed landscape can provide a more intuitive understanding for the whole developmental process. The principle of the landscape module is simply stated below.

Since single-cell data are noisy due to both cellular heterogeneity and gene expression noise, transitions among the cells scattered randomly in the cell state space can be considered as a random walker (this consideration is inspired by Rosvall and Bergstrom's work 23). *Topographer* first constructs a weighted directed graph based on the pseudotime information obtained in the backbone module, and then defines a conditional probability (Eq. (Svensson et al., 2017), *Materials and Methods*) that the random walker moves from one cell to another with relative weight strengths. Then, *Topographer* estimates the visit probability for each cell by solving a master equation (Eq. (6), *Materials and Methods*), and determines the potential of every cell in the dataset, where the potential is defined as the negative of the logarithm of the visit probability, seeing Eq. (8) with Eq. (7) in *Materials and Methods*. All these potentials are then used to construct a Waddington's developmental landscape. For this, a dimension reduction (van der Maaten and Hinton, 2008) is used for visualization, the nearest neighbor interpolation is used to fit a landscape function of



two variables in a 2-dimension space, and a Gaussian kernel is applied to smooth interpolation (see subsection *Scatter Plot of Developmental Landscape, Materials and Methods* or subsection *Scatter Plot of Developmental Landscape, Supplementary Information*). In this constructed landscape, each cell is equipped with both potential and pseudotime: two important attributes of a cell. Therefore, the identified backbone of cell-state transition trajectories, which considers pseudo-potentials rather than potentials, can be viewed as an aerial photograph of the constructed Waddington's developmental landscape (comparing **Figure 3A** with **Figure 3C**).

To demonstrate effectiveness of the landscape module, we analyzed two examples: the one for the same set of artificial data used in **Figure 2E**, with the result demonstrated in **Figure 1C**, and the other for a set of single-cell data on the differentiation of primary human myoblasts, with the results demonstrated in **Figure 3C**. Consequently, we constructed a Waddington's developmental landscape shown in **Figure 3C** from a realistic set of data. Note

that it is different from a cartoon landscape, such as **Figure 5** in references (Olsson et al., 2016). **Supplementary Figure 13** demonstrated another Waddington's developmental landscape constructed using single-cell data on the development of somatic stem cells.

It is worth noting that: (1) In contrast to the backbone module that is mainly used to identify a main "road" but ignores "bumpiness" of the road, the landscape module considers both the road (actually a valley floor of the constructed Waddington's landscape) and its bumpiness (reflected by the height of potentials). (2) Both modules can identify cell-state transition trajectories from a dataset, but the former uses pseudo-potentials that rely on neither pseudotime nor cell type whereas the latter uses potentials that depend on both pseudotime and cell type (Eq. (8) with Eq. (4), *Materials and Methods*). (3) Pseudo-potential cannot correctly reflect the motion of a "ball" (i.e., progenitor cell progression) in the constructed Waddington's landscape in which the ball has lower potential at the beginning than at the end, since a lower cell density

means a higher pseudo-potential. **Supplementary Figure 5** shows a difference between potential and pseudo-potential.

Estimating Fate and Transition Probabilities From Single-Cell Data

Gene regulatory programs underlying cell fate decisions drive one cell type toward another. Quantifying such a transition using single-cell data is challenging due to both cellular heterogeneity and the noise in gene expression in the data.

In order to estimate cell-type dynamics characterized by fate and transition probabilities from single-cell RNA-seq data, it is first needed to determine types of the cells in the dataset. *Topographer* determines cell type according to the following rules: (1) each branch of the identified developmental trajectory is viewed as a cell type with a different branch representing a different cell type; (2) At each branch, the found potential well is taken as a cell subtype with a different potential well representing a different cell subtype. Thus, the number of cell types is equal to the number of branches whereas the total number of cell subtypes is equal to that of potential wells. We will not distinguish cell type and cell subtype unless confusion arises. The cell types determined using this method depend on the shapes of rugged potential wells (prior knowledge can provide additional information in some cases). Therefore, the classification of cell types in this approach is relative rather than “absolute”. For example, in **Figure 3D**, the proliferating cell type indicated by a dashed ellipse can be further divided into two subtypes. In some situations, a potential well in the constructed Waddington’s developmental landscape might not be apparent, but still represents a small cell subtype or an intermediate cell state, which may have important biological implications.

In the dynamics module, *Topographer* considers that transitions among the cells scattered randomly in the cell state space is a random walker who randomly moves from a cell state to another, and then estimates two kinds of probabilities: the fate probability for each cell type and the transition probabilities between every two cell types (*Materials and Methods*). In these estimations, *Topographer* makes use of the cell-state transition trajectories identified in the backbone module.

Specifically, *Topographer* first defines a weight of the directed edge from one cell to another based on the pseudotime (Eq. (4), *Materials and Methods*), and then uses all the possible weights to estimate the visit probability that the random walker visits a cell in the state space, and further the conditional probability defined as a relative link weight (Eq. (5), *Materials and Methods*). With these two kinds of probabilities, *Topographer* further estimates the probability that the random walker visits each cell type, and the transition probabilities between every two cell types (Eq. (9), *Materials and Methods*). These estimations indicate that transitions between cell types are in general not deterministic, but stochastic (referring to **Figure 3D**). In addition, *Topographer* estimates the fate probability of each cell type (Eq. (12), *Materials and Methods*).

In order to demonstrate stochastic cell-type dynamics estimated by the dynamics module, we again analyzed a simulated data with results shown in **Supplementary Figure 6**, and a realistic set of

single-cell RNA-seq data on the differentiation of primary human myoblasts with results demonstrated in **Figure 3D** (as well as another realistic set of single-cell RNA-seq data on the development of somatic stem cells, with results demonstrated in **Supplementary Figure 13**). From **Figure 3D**, we observed that the fate probability (~ 0.53) for the proliferating cell type is about the half of that for the differentiating or interstitial mesenchymal cell type (this is not strange since the proliferating cells are root ones), but the fate probabilities for the latter two (~ 0.99 and ~ 0.98 , respectively) are approximately equal. In addition, the proliferating cells differentiate into the differentiating cells at the ~ 0.16 probability, but the inverse differentiation probability is very small (~ 0.001). On the other hand, the proliferating cells differentiate into the interstitial mesenchymal cells at the ~ 0.31 probability but the inverse differentiation probability is also very small (~ 0.01), implying that the proliferating cells tend to differentiate into the interstitial mesenchymal cells. **Figure 3D** also showed the fate probabilities of cell subtypes and the transition probabilities between them (low panel).

Apart from the above three main functional modules, *Topographer* can also infer both marker gene networks and their pseudo-temporal changes as well as pseudo-temporal characteristics of transcriptional bursting kinetics. We point out that these inferences can in turn be used to infer whether and when (along pseudotime) the branches of a developmental trajectory occur.

Inferring Marker Gene Networks and Their Pseudo-Temporal Changes

The network module aims to infer the trend of how marker gene networks dynamically change along the identified cell-state transition trajectories. For this, *Topographer* uses the network neighborhood analysis method (Li and Horvath, 2007) (or section *The Network Module Infers Marker Gene Networks and Their Pseudo-Temporal Changes, Materials and Methods*) to explore dynamic changes in gene regulatory networks (GRNs) across development.

First, *Topographer* uses GENIE3 (Huynh-Thu et al., 2010) to generate a series of GRNs along the pseudotime. Then, based on these GRNs, *Topographer* further analyzes the covariation partners of some particular gene (or genes) using a topological network analysis scheme (Klein et al., 2015) that can identify those genes most closely correlated with a given gene (or genes) of interest and most closely correlate to each other (See *Materials and Methods* for details). We stress that before these two steps, transcriptomic data of interest need pre-processing (**Supplementary Information**) since they are noisy and would contain many zeros that must be removed in our method.

Here, we used the network module to analyze single-cell data on the differentiation of primary human myoblasts, and obtained dynamic changes in the connections of marker gene networks along the identified cell-state transition trajectories (**Figure 4A**, where the PEBP1 gene is a core node of the networks). From the dependences of mean gene-gene correlation degrees (**Figure 4B**) and mean cell-cell correlation degrees (**Figure 4C**) on the pseudotime, we observed that before branch, both degrees were

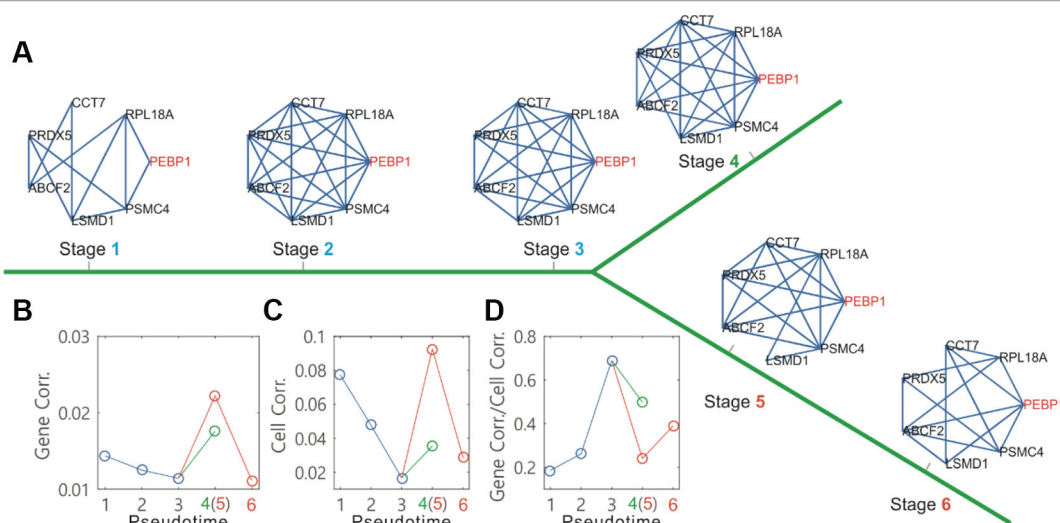


FIGURE 4 | *Topographer* infers dynamic changes in the local connection network of a marker gene along the pseudotime from single-cell transcriptomic data on the differentiation of primary human myoblasts. **(A)** Dynamic changes in a connection network of seven genes along the pseudotime, where the PEBP1 gene (orange) is taken as a core node of neighborhood networks. **(B)** Dynamic changes in the gene-gene correlation degree along the pseudotime before and after branch (different colors), where 6 empty circles correspond to the networks at 6 stages indicated in **(A)**, respectively. **(C)** Dynamic changes in the cell-to-cell correlation degree along the pseudotime before and after branch. **(D)** Dynamic changes in the ratio of the gene-to-gene correlation degree over the cell-to-cell correlation degree along the pseudotime before and after branch.

a monotonically decreasing function in pseudotime (the blue line, **Figure 4B** or **C**), but after branch, each became first monotonically increasing and then monotonically decreasing on one branch (the orange line, **Figure 4B** or **C**), and monotonically increasing on the other branch (the green line, **Figure 4B** or **C**). However, the change tendency for the ratio of the gene-gene correlation degree over the cell-cell correlation degree was just opposite to that described above (**Figure 4D**). Note that a decrease in the overall gene-gene correlation indicates that there are less regulations in the cells. And a decrease in the cell-cell correlation reflects an increase in the amplitude of random fluctuation in gene expression due to the weakening attracting force in the flattening basin of attraction prior to the bifurcation. The ratio of GeneCorr./CellCorr. is a quantitative index for predicting critical transitions. This index increases toward a maximum when cells go through the critical state transition that is similar to the index proposed by Chen et al., (2012) and Mojtahedi et al., (2016).

Inferring Pseudo-Temporal Characteristics of Transcriptional Bursting Kinetics

Transcription occurs often in a bursty manner, and single-cell measurements have provided evidence for transcriptional bursting both in bacteria and in eukaryotic cells (Larson, 2011). By analyzing a simplified stochastic model of gene expression, Xie, et al. previously showed (Friedman et al., 2006) that the number of mRNAs produced in the bursty fashion following a Gamma distribution determined by two parameters: MBF (i.e., the mean number of mRNA production bursts per cell cycle) and MBS (i.e., the average size of the mRNA bursts). We point out that if a Beta-

Poisson distribution (Kim and Marioni, 2013) is used or other distributions are used, the result is similar (data are not shown).

There is great interest in analyzing single cell data to understand the transcriptional changes that occur as cells differentiate and the genes and regulatory mechanisms controlling differentiation processes and cell-fate transition points (Moignard and Göttgens, 2016; Tanay and Regev, 2017). The burst module is designed to infer the trend of how transcriptional bursting kinetics dynamically changes across development. For this, *Topographer* uses the maximum likelihood method (Cam, 1991) to infer the two parameters of MBF and MBS from single-cell RNA-seq data (see section *The Burst Module Infers Pseudo-Temporal Characteristics of Transcriptional Bursting Kinetics, Materials and Methods*), thus revealing dynamic characteristics of transcriptional bursting kinetics before branch, near the branching point and after branch of the developmental trajectory.

We used the burst module to analyze single-cell data on the differentiation of primary human myoblasts. **Figures 5A–E** showed how the cells at four pseudotime points (two before branch, one at branch point, and one after branch) were distributed in the logarithmic plane of BF and BS. A reference system (two orthogonal blue lines indicated by blue: the horizontal line for BF and the vertical line for BS) was used to guide visual comparison between the rates (i.e., the percents indicated) of gene numbers over the total gene number at a particular pseudotime point. The four quadrants of the reference system clearly showed how the genes in the dataset were expressed, e.g., the fourth quadrant showed that the genes were expressed in a manner of high frequency (i.e., the BF is more than 0.33) and small burst (i.e., the BS number is less than 200). We observed that the genes expressed in a bursty manner (i.e.,

the other three cases except for the case in the fourth quadrant) were more at the branching point (97%) than before or after branch (approximate or below 80%). In other words, the percent of the genes expressed with high frequency and small burst was apparently lower at the branching point. From these figures, we can conclude that during the differentiation of primary human myoblasts, there are more genes expressed in a bursty manner at the branching point than before or after branch.

From the dependences of MBF and MBS on the pseudotime (Figures 5F, G), we observed that there were apparently different change trends before and after branch. Figure 5H showed the dependence of the mean mRNA expression level on the pseudotime, demonstrating a change tendency opposite to that of MBF. Although fundamentally similar to the change trend of MBS on the whole, the mean mRNA level (which is approximately equal to the product of the MBS and the MBF) for the branched pseudo trajectory of points 1, 2, 3, and 4 has an increasing tendency with the increase of the pseudotime (Figure 5H). These three subfigures implied that MBF or MBS can be taken as a better indicator of the branch occurrence than the mean mRNA expression level. They also imply that cell fate decisions would not be inferred from the changes in the mean gene expression levels but can be inferred from the changes in the transcriptional bursting kinetics characterized by BS and BF. Recently, Larsson, et al., showed that a separation of expression into bursting kinetics was required to identify the effects of core promoter elements on transcriptional dynamics that were masked at mean expression levels (Larsson, 2019).

DISCUSSION

We have developed a computational pipeline—*Topographer* for construction of developmental landscapes, identification of *de novo* continuous developmental trajectories, and quantification of fate transitions. One unique feature of *Topographer* is its capability of characterizing both transcriptional bursting kinetics and changes in connections of marker gene networks along developmental trajectory. When identifying the backbone of cell-state transition trajectories from single-cell data, *Topographer* was robust to the noise in the dataset (Supplementary Figures 8–10). When applied to the differentiation of primary human myoblasts, *Topographer* first constructed an intuitive developmental landscape for an order and timing of events that closely recapitulated previous studies of this system. In addition, it estimated the fate probabilities for cell types and the transition probabilities between them. Together, the results suggested that the fate transition during the differentiation of primary human myoblasts occurred in a probabilistic rather than deterministic manner, and the transitions between cell types might be unidirectional and bidirectional. These two new insights challenge the traditional view that the development of primary human myoblasts was tree-like or that the process from a predecessor to its generations was both deterministic and unidirectional (Svensson et al., 2017).

When ordering single cells, *Topographer* (like existing methods in the literature) needs to assume sufficient number of cells in the dataset because the backbone module is established essentially

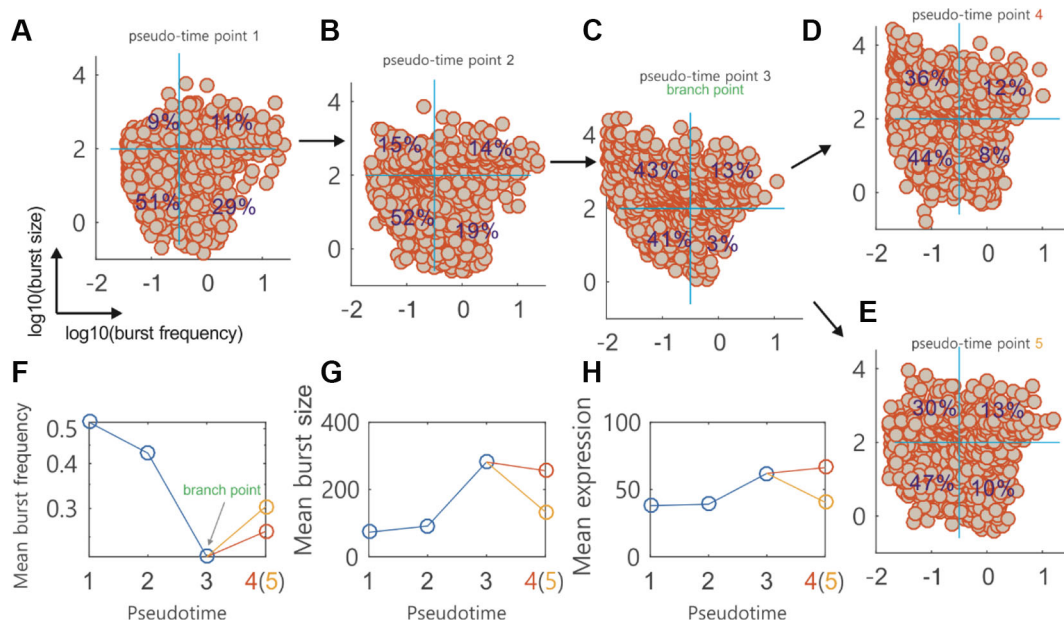


FIGURE 5 | *Topographer* infers dynamic characteristics of transcriptional bursting kinetics along the pseudotime from single-cell RNA-seq data on the differentiation of primary human myoblasts. (A–E) Scatter plot of the genes in the logarithmic plane of burst size (BS) and burst frequency (BF) at four pseudotime points, where every circle represents a gene in the dataset. Four percents are indicated in a reference system (two orthogonal blue lines at every subfigure, which correspond to mean BS and BF, respectively). Numbers 4 and 5 actually represent the same pseudotime point. (F) Evolution of the mean BF along the pseudotime, where the branching point is indicated and two empty circles after the branching point correspond to (D and E), respectively. (G) Evolution of the mean BS along the pseudotime (H) Evolution of the mean mRNA expression level along the pseudotime.

based on the estimation of cell density. A small number of cells (e.g., less than 100), would lead to inaccuracy of finding pseudo-potential well/or wells on a super-ring in the backbone module. As more cells can simultaneously be measured (Klein et al., 2015), the accuracy of *Topographer* will improve. In principle, *Topographer* can also be used to analyze other single-cell data such as mass cytometry data and single-cell PCR data (Bendall et al., 2011).

Cell fate decisions may involve hierarchy of cell types including intermediate cell states or cell subtypes. Identifying such (e.g. rare) sub-cell types is important yet challenging. *Topographer* has shown its ability to identify cell subtypes, which may correspond to shallow or small potential wells in the constructed developmental landscape (right below, **Figure 3D**). Moreover, *Topographer* can estimate the fate probability of each identified cell subtype and the transition probabilities between every two identified cell subtypes (right below, **Figure 3D**), which is one main advantage of *Topographer* compared to many existing methods (Saelens et al., 2019). In particular, *Topographer* enables identification of non-, bi-, and multi-branches (**Figures 2C, D**).

It is worth noting that *Topographer* only provides a general framework connecting three interplayed major components based on single-cell data: cell lineage committing dynamics (macroscopic), gene network dynamics (mesoscopic), and transcriptional bursting kinetics (microscopic). First, *Topographer* provides useful information on their relationships that are implied by the pseudotime, but this kind of time only reflects the impact of the former on each of the latter two. The issues of how and in what degree the inferred gene connection networks or/and transcriptional bursting kinetics influence or/and determine cell fates in the underlying developmental process, remain unexplored. In order to study the relationship between the mesoscope/microscope and the macroscope, a possible way is to establish the so-called balance equation (Wu and Tzanakakis, 2012). Second, in order to estimate the fate probabilities of cell types and the transition probabilities between them (Eq. (12) and Eq. (9), *Materials and Methods*, respectively), *Topographer* makes an assumption that the transition from one cell to another along a cell-state transition trajectory is linear (Eq. (4) in *Materials and Methods* or **Supplementary Eq. (6)**). In many cases, such transition may be nonlinear due to, e.g., cell-cell communication through signal molecules. Third, the traditional Waddington's landscape (e.g., tumor pathobiology 39) is often used as an intuitive tool to describe a differentiation process through the trajectory of a ball into branching valleys with each representing a developmental state (Furusawa and Kaneko, 2012). *Topographer* uses the potential of each cell to quantify developmental landscape, which allows estimation of the transition probabilities between cell types and their fate probabilities to characterize cell lineage committing dynamics. These probabilities have physical meanings as they actually represent the Krammer escape rates (van Kampen, 1992) between potential wells. However, how cell fate decisions including cell-state dynamics are related to Krammer escape rates remain unclear. Fourth, based on the transition probabilities between cell types, one can establish a model of cell population dynamics [referring to **Supplementary Eq. (21)**], and further study

stochastic state transitions from a dynamical-system perspective. Fifth, *Topographer* uses a simple Gamma distribution to infer transcriptional bursting kinetics. For this reference, a more reasonable distribution used would be a Beta-Poisson distribution, but the result is similar (data are not shown). In fact, our reference method is suitable to any distribution.

Finally, using “relatively smaller pseudo-potential and relatively larger distance” (*Materials and Methods*) as a rule in the backbone module is a robust approach in finding cell trajectories (referring to **Supplementary Figures 8–10**); In our method, the transitions among cells are considered as a random walker who moves randomly between the data points scattered in the cell state space. These two ground rules used in *Topographer* can be viewed as new principles of mining single-cell data to uncover mechanisms of cell fate decisions.

CONCLUSION

As the single-cell field progresses towards analyzing the transcriptomic data of large-scale individual cells in parallel, it will become increasingly important to develop statistical methods to reveal cell fate mechanisms in a coherent way at three levels: cell lineage (macroscopic), gene network (mesoscopic), and gene expression (microscopic). In this context, we anticipate that the *Topographer* presented here, and other related approaches, will be vital in maximizing the amount of biological insight that can be obtained from these data.

MATERIALS AND METHODS

The overall *Topographer*, a multifunctional algorithm, comprises five functional modules: the backbone module, the landscape module, the dynamics module, the network module, and the burst module. Main details of these modules are separately stated below and the complete description including data pre-processing is given in **Supplementary Information**.

The Backbone Module Identifies the Backbone of Cell-State Transition Trajectories From Single-Cell Data

Assume that there are m cells and n genes in single-cell RNA-seq data of interest, which can in principle be represented as m points in the n -dimensional space (X) of gene expression (called the cell state space for convenience).

The backbone module aims to identify the backbone of cell-state transition trajectories across development from the dataset. The essence is to find valley floors in a developmental landscape. Specifically, *Topographer* finds valleys with local minimal pseudo-potentials, where pseudo-potential is defined as

$$\tilde{E}(\mathbf{x}) = -\log \rho(\mathbf{x}) \quad (1)$$

with

$$\rho(\mathbf{x}) = \sum_{\mathbf{y} \in X} \exp\left(-\frac{d(\mathbf{x}, \mathbf{y})^2}{2\sigma^2}\right). \quad (2)$$

In Eq. (2), d is the Euclidean distance between two state points x and y in the cell state space X (note: other kinds of distances are also suitable for *Topographer*). Note that ρ represents the local cell density, mostly accounting for the number of cells in a neighborhood defined by σ . The value of parameter σ is set as the corresponding quantile for all pair-wise distances of cell states in the dataset.

Roughly speaking, *Topographer* starts by cell state x_0 (i.e., an initial cell) and then searches for pseudo-potential wells on super-rings (which are actually circular tubes in the cell state space) by recursively applying an extended version of the cluster algorithm (Rodriguez and Laio, 2014). Finally, all the centers of the super-rings are represented in a tree, T . Main details are stated below and more details are given in **Supplementary Information**.

Constructing a Developmental Tree

Starting by an initial cell that has the global minimal pseudo-potential or by the cell that the user chooses according to the prior knowledge, *Topographer* calculates an adaptive radius or an adaptive step length (see subsection *Setting Step Length*, **Supplementary Information**) and searches for pseudo-potential wells on a super-ring centered at this cell and with the radius (referring to **Figure 2A**). The search method (called the pseudo-potential well search algorithm) is based on the idea that cluster centers on the super-ring are characterized by a lower pseudo-potential than their neighbors and by a relatively larger distance from points with locally lower pseudo-potentials. Specifically, *Topographer* first defines

$$\delta(x) = \begin{cases} \max_{y \neq x} d(x, y) & \text{if } \tilde{E}(x) = \min_{y \in X} \tilde{E}(y) \\ \min_{y: \tilde{E}(y) < \tilde{E}(x)} d(x, y) & \text{otherwise,} \end{cases} \quad (3)$$

and then finds local pseudo-potential well/or wells on the super-ring, based on the combination of relatively smaller $\tilde{E}$ and relatively larger δ . Therefore, there is an analogy between the pseudo-potential well search algorithm and a density-based approach developed originally by Rodriguez and colleagues (Rodriguez and Laio, 2014). The segments linking the center and the pseudo-potential wells found on the super-ring can be taken as approximate part/or parts of the entire developmental trajectory.

Then, taking every found pseudo-potential well as the center of a new super-ring with a new adaptive radius, *Topographer* performs similar calculations as at the previous step, thus finding pseudo-potential well/or wells on this new super-ring. Again, the segments linking the new center and the newly found pseudo-potential wells on the new super-ring can be taken as other approximate part/or parts of the entire developmental trajectory. This process is repeated until no new pseudo-potential wells are found. By linking the cluster centers, *Topographer* thus builds a tree-like developmental backbone, which is actually composed of valleys.

Note that for a super-ring center, rather than the starting point, the newly found valleys would include valleys on the “reverse direction” in the processes of searching for local pseudo-potential wells on super-rings, which are not expected in our algorithm. To handle such an exception, *Topographer* excludes those valleys that are too close to the found valleys. In addition, any two newly found valleys with the distance of smaller than the step length are merged by discarding the valleys with larger pseudo-potentials. Such a

treatment may greatly improve the algorithm’s robustness against the noise in the dataset (referring to **Supplementary Figures 8–10**).

Also note that a complete valley floor is constructed by terminating the recursive process for some super-ring on which no desired pseudo-potential wells can be found. Since no loops are assumed to exist in the developmental trajectory, there is definitely a boundary, implying that the search process necessarily stops within finite steps.

After the above search process is completed, all the found pseudo-potential valley floors are represented in an undirected acyclic graph (a tree with branches). To achieve better accuracy and coverage, *Topographer* refines a pseudo-potential valley tree by searching for pseudo-potential well/or wells on the line linking two centers on every edge of the tree (referring to **Supplementary Figures 9 and 10**). To that end, *Topographer* finishes construction of the backbone of a developmental tree from a given set of single-cell data.

Cell Projection and Pseudotime Assignment

After constructing a developmental tree, *Topographer* then projects every cell point in the cell state space onto some edge of the tree according to the shortest distance principle (i.e., the perpendicular distance from the cell point to the edge is shortest). Thus, every cell has its unique relative position in the identified backbone (or in the constructed tree).

Next, *Topographer* assigns a pseudotime for every cell in the dataset. Before that, however, it is needed to determine a root node in the constructed tree. *Topographer* chooses a root cell in such a manner that the distances between this cell and those cells that are initially set according to, e.g., the prior knowledge, are as short as possible. An initial pseudotime is first assigned to this root node. Every other cell in the dataset is then assigned in order with a pseudotime according to its relative position in the constructed tree. Without loss of generality, the full pseudotime may be set as the interval between 0 and 1 (i.e., $0 \leq \tau \leq 1$).

The Landscape Module Constructs a Quantitative Waddington’s Developmental Landscape of Single-Cell Data

Calculation of Cell Potential

After the backbone of a developmental trajectory has been identified and every cell has been endowed with a pseudotime value, the landscape module first estimates the potential of every cell in the dataset and then uses these potentials to construct a quantitative developmental landscape. It is expected that the potential to be introduced can be avoid shortcomings of the pseudo-potential as pointed out in the main text. For this estimation, *Topographer* analogizes transitions between cells at distinct stages of the differentiation process to a random walker who moves randomly between the data points that are randomly scattered in the cell state space. This analogy, which is inspired by Rosvall and Bergstrom’s work (Rosvall and Bergstrom, 2008), is reasonable due to both cellular heterogeneity and gene expression noise in the dataset. In addition, it is important that *Topographer* uses the pseudotime information to construct a weighted directed graph W .

Specifically, *Topographer* defines the weight of the directed edge from cell α to cell β as

$$W_{\alpha \rightarrow \beta} = W_0 e^{-\chi(\tau_\alpha - \tau_\beta)}. \quad (4)$$

(**Supplementary Information** gives a reason for this definition), where τ_α and τ_β represents the pseudotime points for cells α and β respectively, and positive constant χ represents a linearly changing rate that cell α transitions to cell β (this setting implies an assumption, i.e., the evolutionary process from one cell to another along the pseudotime is assumed to be linear). The setting of the χ value in general depends on the dataset under consideration (see subsection 3.2.1 in **Supplementary Information** gives a simple discussion) but it may be set as 30 in our cases (i.e., $\chi=30$). It is worth pointing out that the weight defined in such a manner has used the information on the pseudo-temporally ordered cell trajectories, which is a key for the entire calculation.

Then, in order to estimate cell visit probability on a random walk, *Topographer* defines a conditional probability that the random walker moves from cell β to cell α as the relative link weight, given by

$$p_{\beta \rightarrow \alpha} = \frac{W_{\beta \rightarrow \alpha}}{\sum_{\beta} W_{\beta \rightarrow \alpha}}, \quad (5)$$

which is apparently independent of initial W_0 . If the stationary visit probability of cell α is denoted by p_ω then p_α can in principle be derived from a recursive system of the form

$$p_\alpha = \sum_{\beta} p_{\beta} p_{\beta \rightarrow \alpha}, \quad (6)$$

which represents the probability that the random walker visits the α cell from all the other possible cells. Note that Eq. (6) is actually a master equation (van Kampen, 1992) and can efficiently be solved with the power-iteration method (Booth, 2006). However, to ensure that the unique solution of this equation is independent of the starting node in the directed network, the random walker instead teleports to a random node at a small rate ϵ with $0 < \epsilon < 1$ (in simulation, we set $\epsilon = 0.01$). In addition, to obtain more robust results that depend less on the teleportation parameter ϵ , it is most often to use teleportation to a node proportional to the total weight of the links to the node (Rosvall and Bergstrom, 2008). Because of these considerations, the resulting stationary visit probability for cell α is modified as

$$p_\alpha = (1 - \epsilon) \sum_{\beta} p_{\beta} p_{\beta \rightarrow \alpha} + \epsilon \frac{\sum_{\beta} W_{\alpha \rightarrow \beta}}{\sum_{\alpha, \beta} W_{\beta \rightarrow \alpha}}. \quad (7)$$

Finally, *Topographer* quantifies the potential of every cell in the dataset, according to

$$E_\alpha = -\log p_\alpha, \quad (8)$$

where p_α is given by Eq. (7). Apparently, the potential defined in such a manner has again made use of the information on the identified cell-state transition trajectories due to Eq. (4). We point out that the potential of a cell depends on pseudotime but the pseudo-potential lacks the information on pseudotime.

Scatter Plot of Developmental Landscape

After all the cells in the dataset have been equipped with potentials, all these potentials are then used to construct a Waddington's

landscape for the developmental process. The method is stated as follows. First, dimension reduction is needed for visualization (the tSNE method (van der Maaten and Hinton, 2008) or the PCA method (Hastie et al., 2001) may be used to achieve this purpose). In general, dimension reduction cannot explicitly reflect the information on coordinates in a visualized landscape, e.g., PCA1 and PCA2 in **Figure 3C** do not actually represent components in the dataset. Second, *Topographer* uses the nearest neighbor interpolation method to perform interpolation on a 3-dimensional scattered data set. Specifically, *Topographer* uses ScatteredInterpolant (a function of the MATLAB software) to establish the corresponding relationships between a set of points, (x, y) , and a set of cell potentials, E . These relationships, denoted by $E = F(x, y)$, in principle define a curved surface in the 3-dimensional space for the developmental landscape, which in return passes through all the sampling points in the space under consideration. *Topographer* then uses the nearest neighbor interpolation to evaluate this surface at any query point (x_q, y_q) , thus obtaining an interpolating value of every known potential given by Eq. (8), i.e., $E_q = F(x_q, y_q)$. Third, a Gaussian kernel is used to smooth interpolation. Finally, the identified developmental trajectory is drawn on the constructed developmental landscape (referring to the thick colored line in **Figure 1A** or the thick green line in **Figure 3C**).

We point out that pseudo-potential cannot correctly reflect the motion of a “ball” in the constructed Waddington's developmental landscape in which the moving ball has lower potential at the beginning than at the end, since a lower cell density implies a higher pseudo-potential according to definitions. **Supplementary Figure 5** shows a difference between potential and pseudo-potential.

The Dynamics Module Estimates Fate Probabilities of Cell Types and Transition Probabilities Between Them From Single-Cell Data

Determining Cell Types

Cell-type dynamics can be characterized by fate and transition probabilities. In order to estimate these probabilities, it is first needed to determine the types of cells in the dataset. For this, *Topographer* adopts the following rules: First, each branch in the identified developmental trajectory is defined as one cell type, and a different branch is defined as a different cell type. Then, each potential well on each branch is defined as one cell subtype, and a different potential well is defined as a different cell subtype. These definitions imply that the number of cell types is equal to that of branches whereas the number of cell subtypes is equal to that of potential wells. It should be pointed out that the cell type determined in such a manner is not unique but depends on the choice of $\tilde{E}$ and δ (their respective definitions above). In the following, we will not distinguish cell type and cell subtype unless confusion arises.

Estimating Transition Probabilities Between Cell Types

Equation (5) has given the conditional probability ($p_{\beta \rightarrow \alpha}$) that the random walker moves from cell β to cell α , whereas Eq. (7) has given the stationary visit probability of cell α , i.e., p_α . On the basis of these, *Topographer* estimates the transition probability at which a random walker visits the j th cell type from the i th cell

type (denoted by $q_{i \rightarrow j}$), according to

$$q_{i \rightarrow j} = \sum_{\alpha \in i, \beta \neq j} q_{\alpha \rightarrow \beta}, \quad (9)$$

and the transition probability at which the random walker exits the i th cell type (denoted by $q_{i \rightarrow \cdot}$), according to

$$q_{i \rightarrow \cdot} = \sum_{\alpha \in i, \beta \neq i} q_{\alpha \rightarrow \beta}, \quad (10)$$

where the unrecorded visit rate on a link, $q_{\beta \rightarrow \alpha}$ is given by

$$q_{\beta \rightarrow \alpha} = p_{\beta} p_{\beta \rightarrow \alpha}. \quad (11)$$

Estimating Fate Probabilities of Cell Types

The fate probability for cell type i , denoted by $fate_i$, is defined as

$$fate_i = 1 - q_{i \rightarrow \cdot}, \quad (12)$$

which implies that a larger transition probability at which the random walker exits cell type i corresponds to a smaller fate probability for this cell type. This definition is in accordance with our intuition, so it is reasonable. Again, we emphasize that the above formulae for transition probability ($q_{i \rightarrow j}$) and fate probability ($fate_i$) have all made use of the pseudotime information.

The Network Module Infers Marker Gene Networks and Their Pseudo-Temporal Changes

In a complex mixture of cells, correlations of gene expression patterns would arise from differences between different cell lineages. To explore the correlation between the patterns of gene expression across development, *Topographer* constructs a series of GRNs along the pseudotime, which are directed networks for gene-gene interactions. Unsupervised GRNs are then created by GENIE3 (Huynh-Thu et al., 2010) that takes advantage of the random forest machine learning algorithm.

Based on the constructed GRNs, *Topographer* further explores the covariation partners of some particular gene (or genes) using a topological network analysis scheme (Li and Horvath, 2007). The method is to identify the set of those genes that are most closely correlated with a given gene (or genes) of interest and that most closely correlate to each other, at a given pseudotime point (in practical calculations, we use data at a pseudotime window to improve accuracy) (in **Figure 4** of the main text, however, we showed how neighborhood networks of a marker gene change at several representative pseudotime points). **Supplementary Information** provides more details of the method.

The Burst Module Infers Pseudo-Temporal Characteristics of Transcriptional Bursting Kinetics

Transcriptional bursting kinetics can be characterized by BS and BF. As is well known, Gamma distributions can well capture this bursty expression in some cases. *Topographer* uses a Gamma distribution to infer dynamic characteristics of transcriptional bursting kinetics along the cell-state transition trajectories

identified from single-cell RNA-seq data. Assume that this distribution takes the form (Friedman et al., 2006)

$$p(x) = \frac{x^{\alpha-1}}{b^{\alpha} \Gamma(\alpha)} e^{-\frac{x}{b}}, \quad (13)$$

where x represents the number of transcripts, α represents the mean BF (i.e., the mean number of mRNA production bursts per cell cycle) whereas b does the mean BS (i.e., the average size of the mRNA bursts), and $\Gamma(\cdot)$ is the common Gamma function.

Thus, in order to infer pseudo-temporal characteristics of transcriptional bursting dynamics, the key is to estimate two parameters α and b from the dataset at every pseudotime point. For this, *Topographer* makes use of the maximum likelihood method (Hastie et al., 2001). Since the number of cells at a single pseudotime point would be very few, *Topographer* uses the cell data in a window of this point to obtain more reliable estimations of α and b .

DATA AVAILABILITY STATEMENT

Single-cell data on development of primary human myoblasts can be downloaded from doi:10.1038/nbt.2859 (Trapnell et al., 2014). A MATLAB package for the *Topographer* algorithm is available through github (<https://github.com/cellfate/topographer>).

AUTHOR CONTRIBUTIONS

JZ and TZ conceived and designed the work. JZ carried out computer implementation and data analysis. JZ, QN, and TZ interpreted the simulation results. TZ supervised the project, JZ and TZ wrote the original manuscript, and QN contributed to the writing of final manuscript.

FUNDING

This work was partially supported by the National Natural Science Foundation of China (11931019, 11775314, 91530320, 11631005), and the Science and Technology Program of Guangzhou (201707010117). NIH grant U01AR073159, NSF grant DMS1763272, The Simons Foundation grant (594598,QN), the National Key Basic Research Program of China (2014CB964703).

ACKNOWLEDGMENTS

We thank Mr. Xiaoming Lai's partial computer program codes for the manuscript. This manuscript has been released as a pre-print at bioRxiv (Zhang et al., 2018).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.01280/full#supplementary-material>

It provides a complete description of *Topographer* and supplements analysis of artificial and real examples.

REFERENCES

- Bendall, S. C., Simonds, E. F., Qiu, P., el-AD, Amir, P. O., Krutzik, Finck, R., et al. (2011). Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* 332, 687–696. doi: 10.1126/science.1198704
- Booth, T. E. (2006). Power iteration method for the several largest eigenvalues and eigenfunctions. *Nucl. Sci. Eng. J. Am. Nucl. Soc.* 154, 48–62. doi: 10.13182/NSE05-05
- Cam, L. L. (1991). Maximum likelihood—an introduction. *Int. Statist. Rev.* 58, 153–171. doi: 10.2307/1403464
- Chan, T. E., Stumpf, M. P. H., and Babbie, A. C. (2017). Gene regulatory network inference from single-cell data using multivariate information measures. *Cell Syst.* 5, 251–267. doi: 10.1016/j.cels.2017.08.014
- Chen, L., Liu, R., Liu, Z. P., Li, M., and Aihara, K. (2012). Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers. *Sci. Rep.* 2, 18–20. doi: 10.1038/srep00342
- Friedman, N., Cai, L., and Xie, X. S. (2006). Linking stochastic dynamics to population distribution: An analytical framework of gene expression. *Phys. Rev. Lett.* 97, 168302. doi: 10.1103/PhysRevLett.97.168302
- Furusawa, C., and Kaneko, K. (2012). A dynamical-systems view of stem cell biology. *Science* 338, 215–217. doi: 10.1126/science.1224311
- Gupta, P. B., Fillmore, C. M., Jiang, G., Shapira, S. D., Tao, K., Kuperwasser, C., et al. (2011). Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell* 146, 633–644. doi: 10.1016/j.cell.2011.07.026
- Gut, G., Tadmor, M. D., Pe'er, D., Pelkmans, L., and Liberali, P. (2015). Trajectories of cell-cycle progression from fixed cell populations. *Nat. Methods* 12, 951–954. doi: 10.1038/nmeth.3545
- Haghverdi, L., Büttner, M., Wolf, F. A., Büttner, F., and Theis, F. J. (2016). Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods* 13, 845–848. doi: 10.1038/nmeth.3971
- Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning* (New York: Springer). doi: 10.1007/978-0-387-21606-5
- Huynh-Thu, V. A., Irrthum, A., Wehenkel, L., and Geurts, P. (2010). Inferring regulatory networks from expression data using tree-based methods. *PLoS One* 5, e12776. doi: 10.1371/journal.pone.0012776
- Kafri, R., Levy, J., Ginzberg, M. B., Oh, S., Lahav, G., and Kirschner, M. W. (2013). Dynamics extracted from fixed cells reveal feedback linking cell growth to cell cycle. *Nature* 494, 480–483. doi: 10.1038/nature11897
- Kim, J. K., and Marioni, J. C. (2013). Inferring the kinetics of stochastic gene expression from single-cell RNA-sequencing data. *Genome Biol.* 14, R7. doi: 10.1186/gb-2013-14-1-r7
- Klein, A. M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., et al. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cell. *Cell* 161, 1187–1101. doi: 10.1016/j.cell.2015.04.044
- Larson, D. R. (2011). What do expression dynamics tell us about the mechanism of transcription? *Curr. Opin. Genet. Dev.* 21, 591–599. doi: 10.1016/j.gde.2011.07.010
- Larsson, A. J. M., Johnsson, P., Hagemann-Jensen, M., Hartmanis, L., Faridani, O. R., et al. (2019). Genomic encoding of transcriptional burst kinetics. *Nature* 565 (7738), 251–254. doi: 10.1038/s41586-018-0836-1
- Li, A., and Horvath, S. (2007). Network neighborhood analysis with the multimode topological overlap measure. *Bioinformatics* 23, 222–231. doi: 10.1093/bioinformatics/btl581
- Li, C., and Wang, J. (2013). Quantifying cell fate decisions for differentiation and reprogramming of a human stem cell network: landscape and biological paths. *PLoS Comput. Biol.* 9, e1003165. doi: 10.1371/journal.pcbi.1003165
- Li, C., and Wang, J. (2013). Quantifying Waddington landscapes and paths of non-adiabatic cell fate decisions for differentiation, reprogramming and transdifferentiation. *J. R. Soc. Interface.* 10, 20130787. doi: 10.1098/rsif.2013.0787
- 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. U.S.A.* 111 (39), 14130–14135. doi: 10.1073/pnas.1408628111
- Li, C., and Wang, J. (2015). Quantifying the landscape for development and cancer from a core cancer stem cell circuit. *Cancer Res.* 75, 2607–2618. doi: 10.1158/0008-5472.CAN-15-0079
- Lummertz da Rocha, E., Rowe, R. G., Lundin, V., Malleshaiah, M., Jha, D. K., Rambo, C. R., et al. (2018). Reconstruction of complex single-cell trajectories using CellRouter. *Nat. Commun.* 9, 892. doi: 10.1038/s41467-018-03214-y
- Moignard, V., and Göttgens, B. (2016). Dissecting stem cell differentiation using single cell expression profiling. *Curr. Opin. Cell Biol.* 43, 78–86. doi: 10.1016/j.celb.2016.08.005
- Mojtahedi, M., Skupin, A., Zhou, J., Castaño, I. G., Leong-Quong, R. Y. Y., Chang, H., et al. (2016). Cell fate decision as high-dimensional critical state transition. *PLoS Biol.* 14 (12), e2000640. doi: 10.1371/journal.pbio.2000640
- Moris, N., Pina, C., and Arias, A. M. (2016). Transition states and cell fate decisions in epigenetic landscapes. *Nat. Rev. Genet.* 17, 693–603. doi: 10.1038/nrg.2016.98
- Olsson, A., Venkatasubramanian, M., Chaudhri, V. K., Aronow, B. J., Salomonis, N., Singh, H., et al. (2016). Single-cell analysis of mixed-lineage states leading to a binary cell fate choice. *Nature* 537, 698–702. doi: 10.1038/nature19348
- Perié, L., Duffy, K. R., Kok, L., and de Boer, R. J. (2015). Schumacher TN. the branching point in erythro-myeloid differentiation. *Cell* 163, 1655–1662. doi: 10.1016/j.cell.2015.11.059
- Rodriguez, A., and Laio, A. (2014). Clustering by fast search and find of density peaks. *Science* 344, 1492–1496. doi: 10.1126/science.1242072
- Rosvall, M., and Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci. U.S.A.* 105, 1118–1123. doi: 10.1073/pnas.0706851105
- Saelens, W., Cannoodt, R., Todorov, H., and Saeys, Y. (2019). A comparison of single-cell trajectory inference methods: towards more accurate and robust tools. *Nat. Biotechnol.* 37 (5), 547–554. doi: 10.1038/s41587-019-0071-9
- Setty, M., Tadmor, M. D., Reich-Zeliger, S., Angel, O., Salame, T. M., Kathail, P., et al. (2016). Wishbone identifies bifurcating developmental trajectories from single-cell data. *Nat. Biotechnol.* 34, 637–645. doi: 10.1038/nbt.3569
- Shin, J., Berg, D. A., Zhu, Y., Shin, J. Y., Song, J., Bonaguidi, M. A., et al. (2015). Single-cell RNA-Seq with waterfall reveals molecular cascades underlying adult neurogenesis. *Cell Stem Cell* 17, 360–372. doi: 10.1016/j.stem.2015.07.013
- Svensson, V., Natarajan, K. N., Ly, L. H., Miragaia, R. J., Labalette, C., Macaulay, I. C., et al. (2017). Power analysis of single-cell RNA-sequencing experiments. *Nat. Methods* 14, 381–387. doi: 10.1038/nmeth.4220
- Tanay, A., and Regev, A. (2017). Scaling single-cell genomics from phenomenology to mechanism. *Nature* 541, 331–338. doi: 10.1038/nature21350
- Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., et al. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* 32, 381–386. doi: 10.1038/nbt.2859
- Trapnell, C. (2015). Defining cell types and states with single-cell genomics. *Genome Res.* 25, 1491–1498. doi: 10.1101/gr.190595.115
- van der Maaten, L. J. P., and Hinton, G. E. (2008). Visualizing high-dimensional data using t-SNE. *J. Machin. Learn. Res.* 9, 2579–2505.
- van Kampen, N. G. (1992). *Stochastic Process in Physics and Chemistry* (Amsterdam: North-Holland).
- Waddington, C. H. (1957). *The strategy of the genes: a discussion of some aspects of theoretical biology* (London: Allen and Unwin).
- Wang, J., Zhang, K., Xu, L., and Wang, E. (2011). Quantifying the Waddington landscape and biological paths for development and differentiation. *Proc. Natl. Acad. Sci. U.S.A.* 108, 8257–8262. doi: 10.1073/pnas.1017017108
- Wu, J. C., and Tzanakakis, E. S. (2012). Contribution of stochastic partitioning at human embryonic stem cell division to NANOG heterogeneity. *PLoS One* 7, e50715. doi: 10.1371/journal.pone.0050715
- Zhang, J. J., Nie, Q., and Zhou, T. S. (2018). Topographer reveals dynamic mechanisms of cell fate decisions from single-cell transcriptomic data. *bioRxiv* 215205. doi: 10.1101/251207
- Ziegenhain, C., Vieth, B., Parekh, S., Reinis, B., Guillaumet-Adkins, A., Smets, M., et al. (2017). Comparative analysis of single-cell RNA sequencing methods. *Mol. Cell* 65, 631–643. doi: 10.1016/j.molcel.2017.01.023

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor declared a past co-authorship with one of the authors QN.

Copyright © 2019 Zhang, Nie and Zhou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Analysis of Single-Cell Gene Pair Coexpression Landscapes by Stochastic Kinetic Modeling Reveals Gene-Pair Interactions in Development

Cameron P. Gallivan¹, Honglei Ren^{2,3} and Elizabeth L. Read^{1,2*}

¹ Department of Chemical & Biomolecular Engineering, University of California, Irvine, CA, United States, ² NSF-Simons Center for Multiscale Cell Fate, University of California, Irvine, CA, United States, ³ Mathematical and Computational Systems Biology Graduate Program, University of California, Irvine, CA, United States

OPEN ACCESS

Edited by:

Tian Hong,
The University of Tennessee,
Knoxville, United States

Reviewed by:

Jiajun Zhang,
Sun Yat-sen University, China
Davit Potoyan,
Iowa State University, United States

*Correspondence:

Elizabeth L. Read
elread@uci.edu

Specialty section:

This article was submitted to
Systems Biology,
a section of the journal
Frontiers in Genetics

Received: 16 October 2019

Accepted: 18 December 2019

Published: 31 January 2020

Citation:

Gallivan CP, Ren H and Read EL
(2020) Analysis of Single-Cell
Gene Pair Coexpression
Landscapes by Stochastic Kinetic
Modeling Reveals Gene-Pair
Interactions in Development.
Front. Genet. 10:1387.
doi: 10.3389/fgene.2019.01387

Single-cell transcriptomics is advancing discovery of the molecular determinants of cell identity, while spurring development of novel data analysis methods. Stochastic mathematical models of gene regulatory networks help unravel the dynamic, molecular mechanisms underlying cell-to-cell heterogeneity, and can thus aid interpretation of heterogeneous cell-states revealed by single-cell measurements. However, integrating stochastic gene network models with single cell data is challenging. Here, we present a method for analyzing single-cell gene-pair coexpression patterns, based on biophysical models of stochastic gene expression and interaction dynamics. We first developed a high-computational-throughput approach to stochastic modeling of gene-pair coexpression landscapes, based on numerical solution of gene network Master Equations. We then comprehensively catalogued coexpression patterns arising from tens of thousands of gene-gene interaction models with different biochemical kinetic parameters and regulatory interactions. From the computed landscapes, we obtain a low-dimensional “shape-space” describing distinct types of coexpression patterns. We applied the theoretical results to analysis of published single cell RNA sequencing data and uncovered complex dynamics of coexpression among gene pairs during embryonic development. Our approach provides a generalizable framework for inferring evolution of gene-gene interactions during critical cell-state transitions.

Keywords: stochastic modelling, gene expression noise, gene regulatory networks, single-cell data, scRNA-seq

INTRODUCTION

In recent years, single-cell-resolution measurements have revealed unprecedented levels of cell-to-cell heterogeneity within tissues. The discovery of this ever-present heterogeneity is driving a more nuanced view of cell phenotype, wherein cells exist along a continuum of cell-states, rather than conforming to discrete classifications. The comprehensive view of diverse cell states revealed by

single cell measurements is also affording new opportunities to discover molecular regulators of cell phenotype and dynamics of lineage commitment (Trapnell et al., 2014; Olsson et al., 2016; Briggs et al., 2018). For example, single cell transcriptomics have revealed the widespread nature of *multilineage priming* (MLP), a phenomenon wherein individual, multipotent cells exhibit “promiscuous” coexpression of genes associated with distinct lineages prior to commitment (Nimmo et al., 2015). In principle, mathematical modeling of gene regulatory network dynamics can provide a theoretical foundation for understanding cell heterogeneity and gene expression dynamics, by quantitatively linking molecular-level regulatory mechanisms with observed cell states. However, due to the molecular complexity of gene regulatory mechanisms, it remains challenging to integrate such models with single-cell data.

Mathematical models of gene regulatory network dynamics can account for (and at least partially reproduce) observed cellular heterogeneity in two primary ways. First, gene network models are multistable dynamical systems, meaning a given network has the potential to reach multiple stable states of gene expression. These states arise from the dynamic interplay of activation, inhibition, feedback, and nonlinearity (Kauffman, 1969; MacArthur et al., 2009; Huang, 2012). Second, some mathematical models inherently treat cellular noise. This noise, or stochasticity, is modeled in various ways depending on assumptions about the source (Peccoud and Ycart, 1995; Arkin et al., 1998; Kepler and Elston, 2001; Swain et al., 2002). Discrete, stochastic models of gene regulation, which track discrete molecular entities, regulatory-protein binding kinetics, and binding states of promoters controlling gene activity, have formed the basis of biophysical theories of gene expression noise due to so-called *intrinsic* molecular noise (Peccoud and Ycart, 1995; Thattai and van Oudenaarden, 2001; Kepler and Elston, 2001; Pedraza and Paulsson, 2008). Such stochastic gene-regulation mechanisms have also been incorporated into larger regulatory network models using the formalism of stochastic biochemical reaction networks, and have been utilized to explore how molecular fluctuations can cause heterogeneity within phenotype-states and promote stochastic transitions between phenotypes (Feng and Wang, 2012; Sasai et al., 2013; Zhang and Wolynes, 2014; Tse et al., 2015).

The quantitative *landscape* of cellular states is another concept that is increasingly utilized to describe cellular heterogeneity. Broadly, the cellular potential landscape (first conceptualized by Waddington (Wang et al., 2011; Huang, 2012; Waddington, 2014) is a function in high-dimensional space (over many molecular observables, typically expression levels of different genes), that quantifies the stability of a given cell-state. In analogy to potential energy (gravitational, chemical, electric, etc.), cell states of higher potential are less stable than those of lower potential. The landscape concept inherently accounts for cellular heterogeneity, since it holds that a continuum of states is theoretically accessible to the cell, with low-potential states (in “valleys”) more likely to be observed than high-potential states. The landscape is a rigorously defined function derived from the dynamics of the underlying gene network model, according to some choice of mathematical formalism (Wang et al., 2011; Bhattacharya et al., 2011;

Huang, 2012; Zhou et al., 2016). For stochastic gene network models that inherently treat noise, the landscape is directly obtained from the computed probability distribution over cell-states (Cao and Liang, 2008; Micheelsen et al., 2010; Feng and Wang, 2012; Tse et al., 2015).

Stochastic modeling of gene network dynamics has been employed in various forms for analysis of single cell measurements. For example, application of noisy dynamical systems theory has shed light on cell-state transitions (Mojtahedi et al., 2016; Jin et al., 2018; Lin et al., 2018). Stochastic simulations of gene network dynamics have been used to develop and/or benchmark tools for network reconstruction (Schaffter et al., 2011; Dibaeinia and Sinha, 2019; Bonnaïffoux et al., 2019). Stochastic model-aided analysis of single-cell measurements has been demonstrated to yield insights on gene regulatory mechanisms (Munsky et al., 2018). However, few existing analysis methods utilize discrete-molecule, stochastic models, which fully account for intrinsic gene expression noise and its impact on cell-state, to aid in the interpretation of noisy distributions recovered from single cell RNA sequencing data. There exists an opportunity to link such biophysical, stochastic models, which reproduce intrinsic noise and cell heterogeneity *in silico*, to single cell datasets that characterize cell heterogeneity *in vivo*. In particular, the landscape of heterogeneous cell-states computed from discrete stochastic models can be directly compared to single-cell measurements.

In this work, we present a method for analyzing single-cell gene pair coexpression patterns that is founded on biophysical theory of stochastic gene networks. In our approach, the key object linking the models to the data is the gene-pair coexpression landscape, which is derived directly from the bivariate distribution of expression states, and which is computed from a stochastic model or extracted from single cell measurements. The rationale underlying the method is two-fold: (1) information on gene-gene interactions can be inferred from the distinctive characteristics of noise in single-cell data (i.e., from the “shape” of the landscape); (2) existing analysis techniques are relatively insensitive to landscape shape. We first comprehensively compute and classify the landscapes produced by a family of ~40,000 stochastic two-gene regulatory network models. We then use the model-derived classification to analyze published data from vertebrate development. In so doing, we uncover both expected and novel patterns of coexpression in development. While our analysis here is proof-of-principle, and limited to two-gene interactions, the conceptual framework could be expanded to include multibody gene interactions in the future.

METHODS

Discrete, Stochastic Models of Two-Gene Regulatory Networks

We first developed a family of stochastic models of gene-gene interactions (see **Figure 1** for model schematic), which is based

on previously published models (Feng and Wang, 2012; Zhang and Wolynes, 2014). We label two genes X and Y . Each gene encodes a protein, which acts as a transcription factor (TF) that potentially regulates its own expression as well as that of the other gene. Each gene has a promoter (or more generally, regulatory regions of DNA) that can be bound by any combination of its own expressed protein and/or the other gene's expressed protein. The promoter states are thus labeled as: X_{00} (neither transcription factor is bound to X 's promoter), X_{0x} (X 's own protein is bound, resulting in autoregulation of gene expression), X_{y0} (Y 's protein is bound to X 's promoter, resulting in cross-regulation), X_{yx} (both proteins are bound to X 's promoter, resulting in combinatorial regulation). (The promoter states for gene Y are defined in a symmetric manner.) The regulatory effect of each promoter state (i.e., the effect of having none, one, or both proteins bound on the gene's expression) is accounted by the transcription rate g_{ij} corresponding to each possible promoter state: e.g., when gene X 's promoter is unbound, it transcribes at rate g_{00}^X . Binding of Y 's protein changes the transcription rate to g_{y0}^X , which may be lower, higher, or the same, depending on whether the effect of Y on X is assumed to be repressing, activating, or not impacting. (All other transcription rates for each promoter state and for gene Y are defined similarly.) The model involves three classes of reactions: mRNA synthesis, mRNA degradation, and promoter-state-change reactions. mRNA synthesis reactions are given by:



where x and y denote mRNA transcripts which will be translated into the transcription factors encoded by genes X and Y , respectively. mRNA degradation reactions are given by:

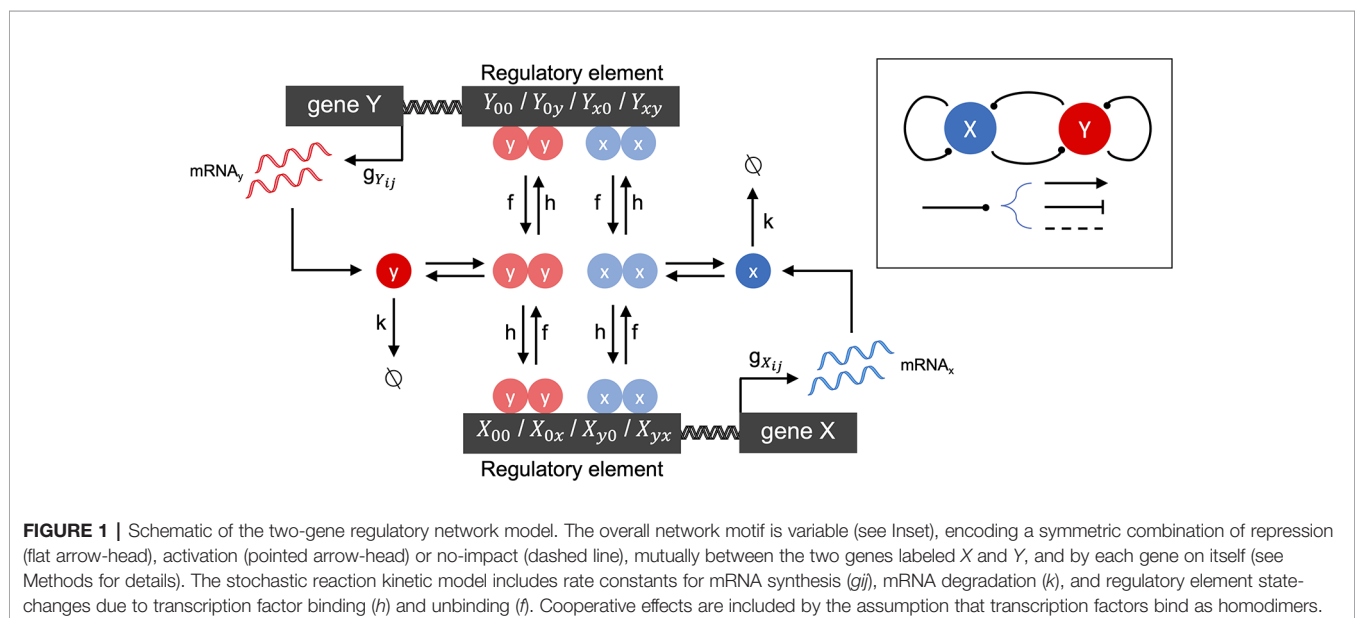


Promoter-state-change reactions are given by, e.g.:



which represents the change of promoter-state (and corresponding regulatory impact) on gene X when Y 's transcription factor binds (forward reaction) or unbinds (reverse reaction). All other promoter-state-change reactions for X and Y are defined similarly. The changes of promoter state occur with forward rates $hy^2/2$ or $hx^2/2$ (when the change of state occurs due to binding of transcription factor from gene Y or X , respectively) and f (when the change of state occurs due to an unbinding event). The model tracks copy numbers of individual mRNA molecules in the cell, to enable direct comparison with single cell transcriptomic data, but translation of mRNA into protein is not explicitly accounted for. Instead, transcription factor (protein) levels are assumed to be linearly proportional to mRNA, and this proportionality constant is subsumed into the binding rate h . The quadratic dependence of the forward binding rates on x or y arises from the assumption that homodimeric transcription factors regulate gene expression, which is a general and convenient way to include cooperativity in the model.

We assign rate constants to intracellular processes that are in line with experimental estimates from vertebrates, where possible (see **Table 1**). (For full details of model reactions and parameter derivations, see **Supplement**). Rates of mRNA synthesis and degradation are relatively well characterized, though they vary considerably for different transcripts (Schwanhäusser et al., 2011). Rates of promoter-state-change are less well-defined,



since promoter-state-changes that ultimately impact gene expression may be attributed to a variety of molecular processes, including: (a) relatively fast processes of TF binding or unbinding from DNA (b) relatively slow chromatin remodeling processes that may be initiated or facilitated by TF binding, require multiple steps and cooperative interactions, and are generally poorly understood. In our models, to account for this range of possible mechanisms, we consider a wide range of parameter values h , f for promoter-state-changes. (The significance of these fast and slow regimes, termed the *adiabatic* and *nonadiabatic* regimes, respectively, to cell-state stability has been studied previously by stochastic modeling (Sasai and Wolynes, 2003; Feng and Wang, 2012; Sasai et al., 2013; Zhang and Wolynes, 2014). We here define the “fast” regime as determined by measured parameter values of protein binding/unbinding DNA (e.g., from Geertz et al., 2012), occurring with timescales of minutes, seconds, or faster. We define the “slow” regime more broadly as any epigenetic/chromatin changes occurring on timescales of hours, days, or longer. For example, in mammalian cells, changes of chromatin state during cell-fate specification were estimated to be on the order of several days (Mariani et al., 2010; Hathaway et al., 2012), while theoretical studies predicted timescales on the order of the cell cycle time (i.e., hours to days, Sasai et al., 2013).

We define two types of model systems. The Mutual Inhibition/Self-Activation (MISA) model encodes a common network motif that is understood to control a variety of cell fate decisions (Graf and Enver, 2009; Huang, 2013) and has been extensively studied by mathematical modeling (Huang et al., 2007; Feng and Wang, 2012; Chu et al., 2017). In contrast, the Two-Gene Flex model flexibly encodes a variety of regulatory interactions, as described below.

Mutual Inhibition/Self-Activation Model

In all models, promoter activity is assumed to be either high (transcription rate g_{hi}) or low (g_{lo}) (giving a relatively fast or slow rate of mRNA synthesis, respectively). To encode MISA regulatory logic, mRNA synthesis rates for each promoter state are $\{g_{00}^X, g_{0x}^X, g_{y0}^X, g_{yx}^X\} = \{g_{lo}, g_{hi}, g_{lo}, g_{lo}\}$. Transcription rates for gene Y are defined symmetrically, $\{g_{00}^Y, g_{0y}^Y, g_{x0}^Y, g_{yx}^Y\} = \{g_{lo}, g_{hi}, g_{lo}, g_{lo}\}$. The high rate corresponds to maximal activity, whereas the low rate is

effectively off (but is nonzero to allow for some leakiness in the promoter). Thus, binding of the self-TF turns the gene on, but subsequent binding of the other TF turns the gene off. The relative strengths and kinetics of the activating (self-regulatory) and repressing (cross-regulatory) interactions are encoded in the rates of binding/unbinding of regulators. Autoregulatory binding and unbinding rates (symmetric on both genes) are denoted by h_a and f_a , respectively. Cross-regulatory rates are denoted by h_r and f_r . The model is thus fully specified by 7 parameters: $\{g_{lo}, g_{hi}, k, h_a, f_a, h_r, f_r\}$. We computed landscapes for ~22,000 unique parameter combinations for the MISA regulatory logic (see **Table 1** for parameter value ranges). We studied only symmetric network motifs, but asymmetry between the genes is accounted for by allowing the “on” transcription rate g_{hi} to be asymmetric between the two genes (in case of asymmetry in g_{hi} the model is specified by eight parameters).

Two-Gene Flex Model

The Two-Gene Flex model is identical to MISA in all ways except the regulatory logic. Instead of the transcription rates being $\{g_{lo}, g_{hi}, g_{lo}, g_{lo}\}$, all 16 logical combinations of four promoter states and two activity-levels are included. Within these combinations, various behavior is encoded including self-activation, self-repression, mutual activation, mutual repression, no interaction (self- or cross-), and dual-effects (where a TF has a distinct effect whether bound alone or in combination with the other). Note that the MISA logic is contained within these 16 combinations. Note also that the promoter states for X and Y are always defined symmetrically, i.e., only symmetric motifs are included. We computed landscapes for ~34,000 unique parameter combinations for the Two-Gene Flex Model (including all network motif variants). Our aim with the Two-gene Flex model was to comprehensively encode all possible logical combinations within the constraints of the symmetric two-gene model. Note that these combinations encompass several cis-regulatory motifs that have been described previously. For example, $\{g_{00}, g_{0y}, g_{x0}, g_{yx}\} = \{g_{hi}, g_{lo}, g_{hi}, g_{lo}\}$, corresponds to a “simple repressor” motif where Y is the repressor, and $\{g_{00}, g_{0y}, g_{x0}, g_{yx}\} = \{g_{hi}, g_{lo}, g_{lo}, g_{lo}\}$, corresponds to a “dual repressor” motif (Bintu et al., 2005). Our Two-Gene Flex model also encompasses various biologically inspired logic gates for combinatorial cis-regulation studied previously (Zhang et al., 2009).

TABLE 1 | Rate Parameters used in gene regulatory network models.

Rate constant	Symbol	Units	Value	Comments/Source
mRNA synthesis (not repressed)	g_{hi}	mRNA/hr	0.8 – 1.4*	Schwanhäusser et al. (2011)
mRNA synthesis (repressed)	g_{lo}	mRNA/hr	0.001	see text
mRNA degradation	k	/hr	0.2 [†]	Schwanhäusser et al. (2011)
Promoter state change (unbinding)	$f^{\dagger}$	/hr	(fast) 10 – 10 ⁵ (slow) 10 ^{–6} – 10	Geertz et al. (2012) see text
Promoter state change (binding)	$h^{\dagger}$	hr ^{–1} mRNA ^{–2}	(fast) 10 – 500 (slow) 10 ^{–6} – 10	Geertz et al. (2012) see text

Parameter values are derived from experimental measurements in vertebrates, where possible. See Methods text for details. *Measured rates of mRNA synthesis varied, with a median of 2/h Schwanhäusser et al. (2011). We use lower values (within experimental range) to roughly match observed counts in scRNA-seq data, which may be lower than expected because of dropouts or other technical issues. [†]Corresponds to mRNA half-life of 3.5 h, which is well within experimentally measured values but shorter than the median value of 9 h, assuming that transcriptional regulators have shorter-than-average half-lives in the cell. [‡]Promoter state change rates f and k are reported in fast and slow regimes. Fast promoter state changes are assumed to occur due to TF-DNA unbinding or binding events, with rate parameters chosen based on values reported in Geertz et al. (2012) (see **Supplement** for details on parameter derivation and unit conversion). Slow promoter state changes are thought to involve collective changes in epigenetic marks and rearrangement of chromatin.

Mathematical Framework: Chemical Master Equation

Chemical Master Equation

Stochastic dynamics for the above-described network motifs are modeled by a Chemical Master Equation (CME) (alternatively known as a discrete space, continuous time Markov Chain). The instantaneous state of the system is given by the vector $\mathbf{n}$, which enumerates the mRNA copy numbers and promoter-states of both genes, i.e., $\mathbf{n} = [n_x, n_y, X_{ij}, Y_{ij}]$, where n_x is the mRNA copy number for gene X , X_{ij} is the promoter state for gene X , and so on. The CME gives the probability for the system to exist in a given state at a given time, $\mathbf{p}(\mathbf{n}, t)$. The CME can be written in vector-matrix form as a linear system

$$\frac{dp(\mathbf{n}, t)}{dt} = \mathbf{K}p(\mathbf{n}, t), \quad (4)$$

where $\mathbf{K}$ is the reaction rate-matrix. Each off-diagonal element K_{lm} gives the rate of transitioning from state m to l (nonzero values correspond to allowed state transitions with rates according to reactions 1–3 above), while the diagonal elements are the summed rates for exiting each state, $K_{ll} = -\sum_{m \neq l} K_{ml}$. Transition rates are computed according to standard stochastic chemical kinetic rate laws (Gillespie (1977)). If both types of mRNA are assumed to exist in the cell in copy numbers that never exceed $M - 1$, then the total size of the enumerated space including all possible states is $N = M \times M \times 4 \times 4$ (note that the total number of mRNA copy number states includes the state of 0 copies, thus $n_x, n_y \in \{0, 1, \dots, M - 1\}$). The assumption that mRNAs never exceed $M - 1$ is equivalent to assuming reflective boundary conditions on the enumerated state-space. That is, it assumes the propensity of reactions that lead to mRNA numbers exceeding $M - 1$ is 0. This assumption is justified when M is chosen to be sufficiently large compared to g/k (Chu et al., 2017). We confirmed that the probability of mRNAs exceeding $M - 1$ for our parameter values is negligible (Supplement, Section 2.2) and we further confirmed that increasing M (from 21, the value used in calculations throughout the manuscript, to 36) had negligible impact on quasipotential landscape shape and all subsequent analysis of single cell RNA sequencing (scRNA-seq) data (Figure S2). Note that an algorithm has been published recently that provides rigorous error bounds on steady-state solutions to the CME (Gupta et al., 2017) though we do not make use of the algorithm here.

Computing Gene Pair Coexpression Landscapes

The complete steady state probability to find a cell in state $\mathbf{n}$ is given by the vector $\boldsymbol{\pi}(\mathbf{n}) = \mathbf{p}(\mathbf{n}, t \rightarrow \infty)$, which is obtained from Eq. 4 using eigenvalue routines in numpy and scipy (McKinney, 2010; van der Walt et al., 2011). Each individual model requires solution of an N -state system, where N is (e.g., assuming the probability to have mRNA exceed 25 is negligible, then $N = 10,816$). Efficient computation of the landscapes over tens of thousands of model variants/parameter combinations was achieved using routines compiled with the numba library (Lam et al., 2015) and parallelization using Python's multiprocessing library to distribute the workload across the available cores.

To mimic experimental scRNA-seq data, the probability is projected onto the mRNA subspace by summation over all promoter state combinations. We hereon define the gene pair coexpression landscape as the steady-state probability to find a cell with mRNA count numbers (n_x, n_y) . More precisely, the *probability landscape* is the vector $\boldsymbol{\pi}$ with each element π_i giving the steady-state probability for the cell to be found in state i with the combination of mRNA counts (n_x, n_y) from genes X and Y , and $i \in 1, \dots, M^2$. Alternatively, the *quasipotential landscape* is log-transformed, given by the vector ϕ where $\phi_i = -\ln(\pi_i)$.

scRNA-Seq Data Acquisition, and Landscape Estimation

Experimental data is obtained from the published scRNA-seq measurements of Briggs et al. (2018). The dataset “Corrected_combined.annotated_counts.tsv” was used which provides the normalized transcriptome profiles for *Xenopus tropicalis* at single cell resolution for ten different stages of embryonic development, with labeled cell types and parent cell types. We analyzed 1,380 gene pairs, which were identified as putative MLP pairs in Briggs et al. (2018), based on their estimated changes in coexpression over the course of development. Gene pairs were identified by their developmental stage and lineage branch point in which coexpression was maximal. Cell types from other stages were then included in the lineage if they were a parent (preceding in development) cell type or daughter (descendant later in development) cell type. After selecting the desired gene pair and cell/tissue/cluster type of interest, gene pair counts were combined and summed resulting in ten gene pair landscapes, one for each stage of development, in cells of the relevant lineage.

To directly compare computed coexpression landscapes with experimental data, we extracted cell count matrices for each gene pair, and where necessary, truncated to mRNA count numbers $\leq M - 1$ (truncation eliminated less than 0.5% of cells in the data, across all gene pairs and cell stages). This produces an $M \times M$ (including zeros) count matrix that serves as a sampled estimator of the steady-state distribution, $\tilde{\boldsymbol{\pi}}(n)$, of the same size as computed landscapes. In order to compute the sampled quasipotential landscape, we use $\tilde{\phi}(n) = -\ln \tilde{\boldsymbol{\pi}}(n)$, after replacing the not-observed count-combinations with a low but nonzero estimate of these probabilities (since log of zero is undefined). We use a general estimate of $1\text{E-}6$ for nonobserved counts, both because it is in line with the predictions of the theoretical models for the low probability edges of the distributions, and because it is less than the lowest estimable probability (i.e., observation of one cell in a given matrix position, given total cell counts on the order of 10^5 , would correspond to an estimated probability of $1\text{E-}5$).

Dimensionality Reduction for Landscape Shape-Space

We apply Principal Component Analysis (PCA) to the theoretically computed landscapes over the model sets to achieve a reduced-dimension description of landscape shape. All PCA training and dimensionality reduction was performed using the decomposition module of the python package scikit

learn. Each unique model is treated as a replicate and the steady-state probability π_i (or alternatively, quasipotential ϕ_i) of each of the $M \times M$ possible mRNA copy-number states (n_x, n_y) is treated as a feature.

The principal components obtained from the model set were then used to fit the experimental data, where each landscape from each gene-pair/stage is a replicate. Note that in our application, we have opted to use a “theory-driven” analysis of landscape shape-space, where the PCA training set consists of theoretically computed probability (or quasipotential) landscapes. The experiment-derived landscapes are then projected into this theory-driven shape-space, which enables linking of experimentally measured gene-pair landscapes with possible model logic/parameter combinations that could produce observed landscape shapes. Alternatively, a “data-driven” analysis is possible, wherein the PCA training set consists of experiment-derived landscapes. Such an analysis makes no connection between theoretical models and experimental data, but can still be useful in revealing shape-features present in experimental data. We show results from data-driven analysis in **Supplement** section 2.4 and **Figure S5**.

Clustering Of Developmental Landscape-Shape Trajectories

By viewing the time-ordered coexpression landscapes of a given gene pair in PCA space, termed “landscape-shape trajectories”, one can gain insight into the genes' roles in development. The trajectories were hierarchically clustered based on their geometric distance in PCA space. More specifically, the *fcluster* method in scikit-learn package was used in hierarchical clustering (McKinney, 2010), and the geometric distance between trajectories A and B were defined as the sum of the pair-wised Euclidean distance between two corresponding stages, i.e.,

$$\|A - B\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n (A_{ij} - B_{ij})^2} \quad (5)$$

where $\|\cdot\|_F$ is the Frobenius norm, A and B are two trajectories represented by m by n matrices, m is the number of developmental stages in single cell data, n is the number of PCA components used in clustering.

RESULTS

Stochastic Two-Gene Network Models Show A Variety Of Coexpression Landscape Shapes, Distinguishable by PCA

Our modeling framework enabled efficient computation of coexpression landscapes resulting from discrete, stochastic gene network models. This in turn enabled us to compute landscapes for tens of thousands of parameter sets, encompassing both various relative strengths and kinetics of regulatory interactions, as well as different schemes of regulatory logic among the two genes (see Methods). This approach

afforded a comprehensive view of theoretically predicted landscape shapes resulting from gene-gene interactions (within the assumptions of the current model system).

We applied PCA to the computed probability landscapes for Two-Gene Flex, in order to find a low-dimensional description of their shapes (**Figure 2**). The first two PCA components encompass 98% of total covariance, and all models fall within a triangular region of this 2D subspace. The vertices of the triangle correspond generally to landscapes with: (1) very low expression of both genes (i.e., transcript levels of X/Y are *lo/lo*, **Figure 2E**), (2) high simultaneous expression of both genes (*hi/hi*, **Figure 2C**), and (3) expression of only one gene at a time (*hi/lo* and *lo/hi*, **Figure 2A**). Landscapes located away from the vertices are thus well-described by some linear combination of these three shapes, consistent with PCA, and supported by visual inspection. In all, the results reveal that two-gene interaction motifs can encode a wide variety of patterns of coexpression, including mixtures of all combinations of *lo/lo*, *hi/hi* and *lo/hi*, *hi/lo* phenotypes (e.g., **Figure 2B**). At the same time, this variety of shapes is well-described by a small number of principal components (which form a basis for what we term the “shape-space”), and we hereon use the magnitudes along these components as measures of landscape shape.

Shape Measures of Coexpression Landscapes Distinguish Different Types of Mutual Gene–Gene Interactions

We sought to understand how different regulatory motifs contributed to landscape shape. Projecting the landscapes arising from each network motif separately revealed distinctive patterns (i.e., occupying distinct, but overlapping, regions of the PCA triangle) (approximately 2,000 landscapes were computed for each network motif, i.e., ~2,000 models that share the regulatory logic but have different kinetic parameters). We grouped all motifs according to their region of occupancy within the PCA triangle, and discovered logical consistency among the groups (see **Figure 3**). For example, all motifs with some type of mutual activation were found to co-occupy a region of PCA shape-space in the lower part of the triangle (3A). This result is consistent with the intuition that motifs with mutual activation cannot produce the apparent bistability seen in landscapes at the *hi/lo-lo/hi* vertex of the triangle. The other three motif groupings include motifs with some type of mutual repression, motifs with no inter-gene interactions, and incoherent motifs with dual-interactions (when the regulator bound by itself has the opposite effect of the regulator bound in combination with the other TF). Note that 2 of the 16 logical combinations of promoter binding-states in the Two-Gene Flex models are not included here, since they effectively encode no gene-gene interactions (the “always on” or “always off” logic, $\{g_{hi}, g_{hi}, g_{hi}, g_{hi}\}$ or $\{g_{lo}, g_{lo}, g_{lo}, g_{lo}\}$). Note that here we assess all kinetic parameter combinations associated to one regulatory motif; these parameters tune the strength of different interactions. As such, the analysis of **Figure 3** assumes fixed network topologies but variable weights on network edges, accounting for the overlap between different motifs. These results indicate that

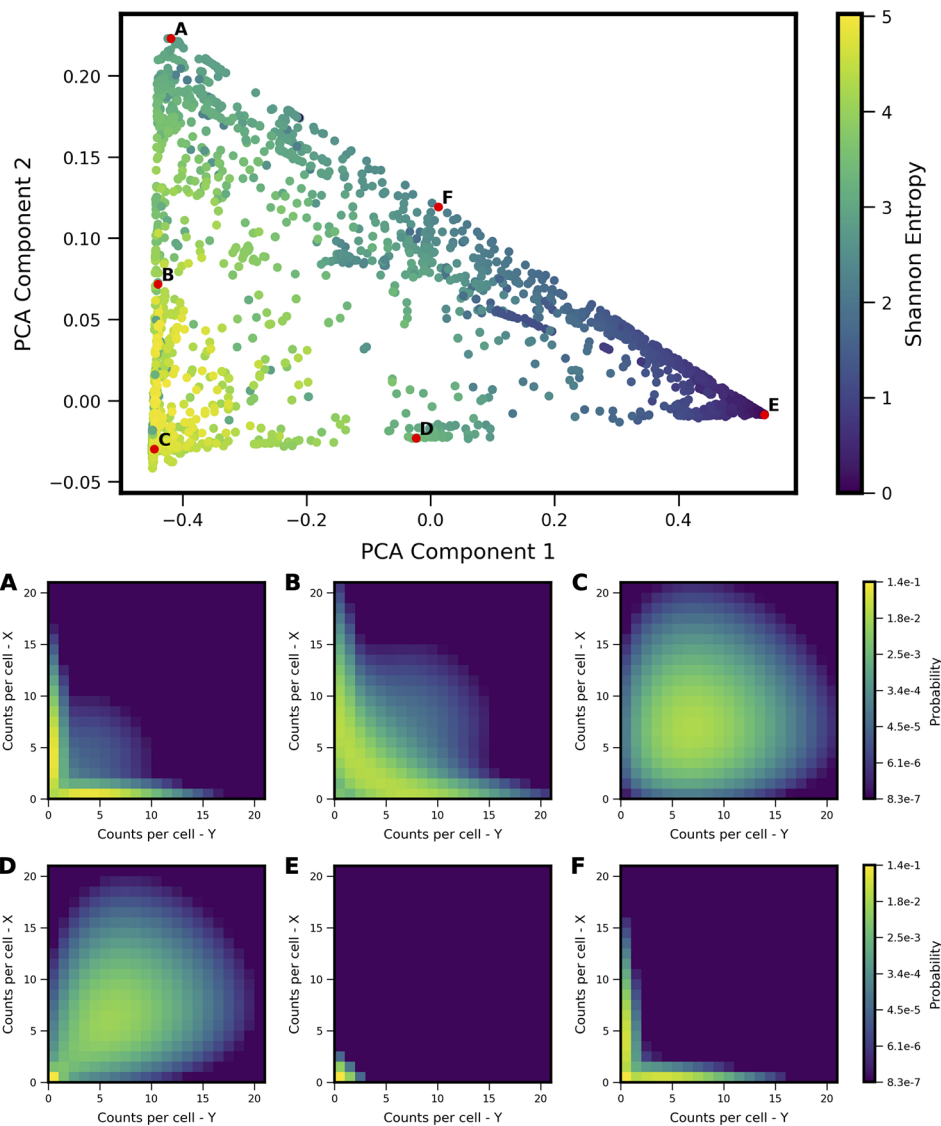


FIGURE 2 | Shape-space of simulated Two-Gene Flex coexpression landscapes analyzed by Principal Component Analysis (PCA). Coexpression landscapes were computed for 34,097 unique two-gene stochastic network models with varying regulatory interactions and kinetic rate parameters (see Model schematic in **Figure 1**). (Top) All model landscapes projected onto the first two principal components. Each dot corresponds to one model, colored by the model's Shannon Entropy. (Bottom) Representative quasipotential landscapes $\phi(n)$ (see Methods) of individual models from different regions of PCA component-space. color of each discrete grid space in $\{x,y\}$ corresponds to computed probability (in log-scale) to find a single cell with the corresponding numbers of $\{x,y\}$ transcripts. Each landscape is labeled (**A-F**) to indicate the corresponding point where it appears in PCA space above.

landscape shape can to some extent be used to distinguish regulatory interactions between pairs of genes, despite variable and/or unknown kinetics governing the interactions.

Commonly Used Pairwise Metrics Are Relatively Insensitive to Coexpression Landscape Shape

In order to analyze how previously applied measures of gene-gene interactions align with landscape shape, we computed a set of metrics for each model landscape and visualized the resultant

values projected onto the PCA subspace. We chose four metrics: Shannon Entropy, Pearson Correlation Coefficient, Mutual Information, and a Coexpression Index (see **Figure 4**, note Shannon Entropy is visualized also in **Figures 2** and **3**). The first three of these are obtained directly from the computed bivariate probability distributions according to standard definitions; the Coexpression Index has been used previously (Briggs et al., 2018) and is given by the conditional probability to find cells with nonzero counts of both mRNA x and y (conditioned on the cells having nonzero counts of at least one

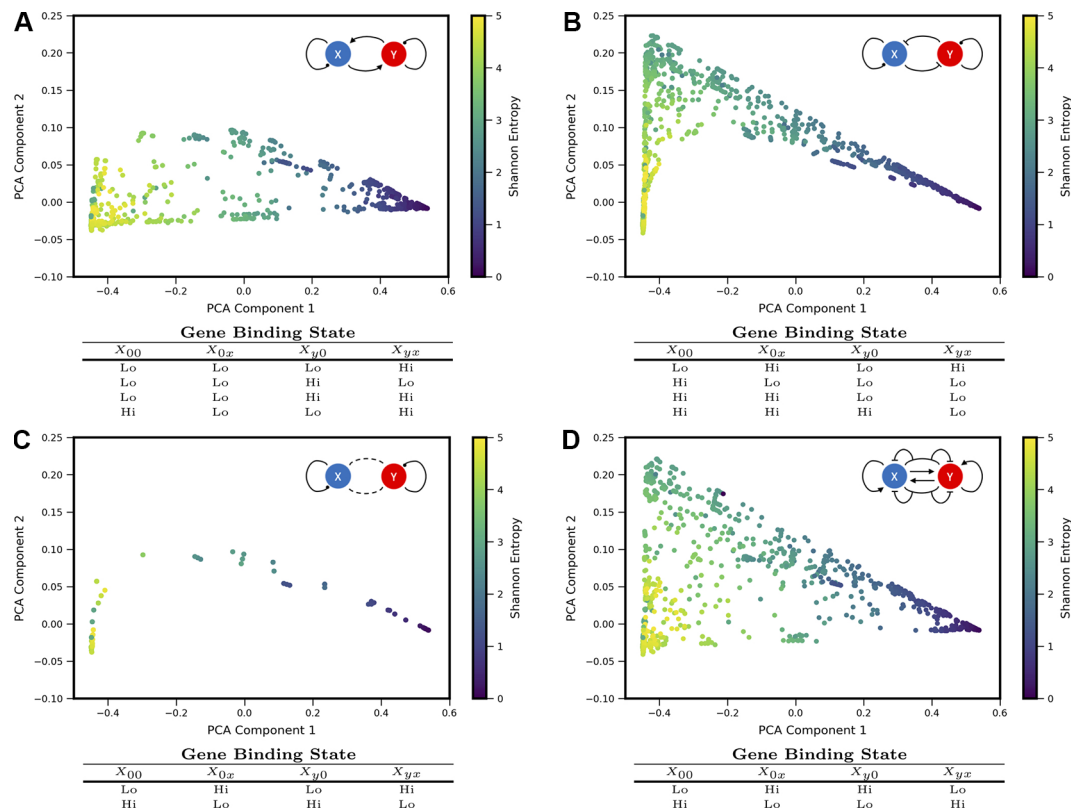


FIGURE 3 | Coexpression landscapes computed from the Two-Gene Flex models show distinctive shapes that depend on the regulatory logic of gene-gene interactions. The Two-Gene Flex model encodes 16 logical combinations (2^4) of gene-gene interactions, corresponding to four possible promoter-binding states and two possible levels of transcription activity (low and high). These 16 model variants can be grouped into motif classes: **(A)** Models with mutual activation. **(B)** Models with mutual repression. **(C)** No mutual gene-gene interactions. **(D)** "Incoherent" models, where the combinatorial-binding state has the opposite behavior of both of the singly bound states (see text). Within each motif class, different kinetic parameters serve to modify the relative strength of interactions (i.e., different weights on the edges). Each motif class occupies a distinct, but overlapping, region of the shape-space (with the exception of the Incoherent motif, which can reach all areas of the shape-space).

of genes X or Y). Here, for a given model j , we derive this metric from the probability landscape π over count-states i by:

$$m_{j, \text{Coex. Index}} = \frac{\sum_{i \in n_x > 0 \cap n_y > 0} \pi_i}{\sum_{i \in n_x > 0 \cup n_y > 0} \pi_i}. \quad (6)$$

We estimate the value of each metric as a function of landscape shape (that is, we estimate the function $m(c_1, c_2)$, where m is a given metric and (c_1, c_2) are the coordinate values in PCA components 1 and 2). For each of the four metrics, we estimate and visualize this function by first computing each of the four metrics from the probability landscapes $\pi(n_x, n_y)$ corresponding to each of the 34,097 models. We then project the models onto the first two principal components, with a given metric serving as the color scale (e.g., as shown with Shannon Entropy, **Figure 2** top). The continuous surface $m(c_1, c_2)$ is then estimated by local averaging and interpolation over the computed results for each individual model landscape with the *tricontourf* routine from the matplotlib package. We found that each metric aligns in distinctive, and generally intuitive, ways with the PCA landscape shape-space. High or low values of each

metric were to some extent localized to particular sub-regions of the triangle, and thus could be understood to be arising from landscapes of similar shape. However, numerous examples can also be found of models colocated (or nearly colocated) in the triangle but having different values of a given metric, so the functional dependence $m(c_1, c_2)$ is noisy.

For Shannon entropy, the highest values are generally seen near the *hi/hi* vertex of the triangle, while the lowest values are seen near the *lo/lo* vertex. This reflects the amount of disorder in the *hi/hi* state of expression, in which a broad range of count-values are possible for each gene, whereas in the *lo/lo* vertex, count values are always zero or near-zero. The noise in expression levels can be quantified more precisely for the subset of models in the "slow-binding" regime ($h, f \ll g, k$). In this parameter regime, cells show distinctive high ("hi") and/or low ("lo") expression states with mean counts g_{hi}/k and g_{lo}/k , respectively, and the disorder in each expression state can be quantified as Poisson birth/death noise (Al-Radhawi et al., 2019), such that variance scales linearly with the expression rate g . Sources of disorder contributing to higher values of Shannon Entropy include both noisy expression within a given phenotype

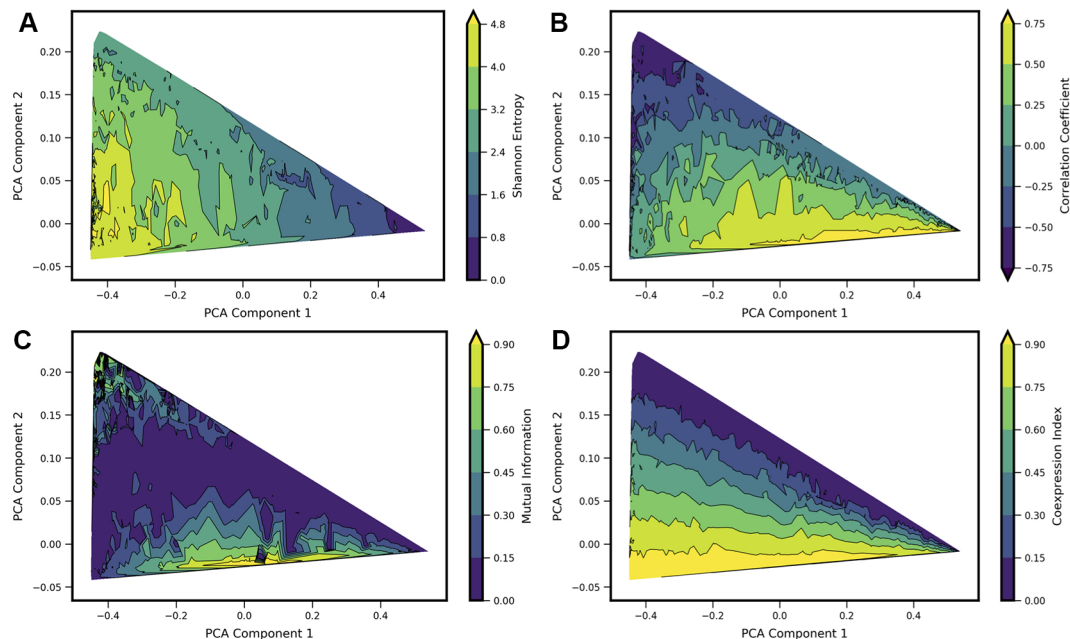


FIGURE 4 | Comparison of four standard metrics of gene-gene coexpression with landscape shape. Metrics include: **(A)** Shannon Entropy. **(B)** Correlation Coefficient. **(C)** Mutual Information. **(D)** Coexpression Index (see text for details). Each metric was computed for each computed model landscape, using the same set of 34,097 Two-Gene Flex models as in **Figures 2** and **3**. Contour plots show each metric as a function of principal components 1 and 2, obtained by local averaging and interpolation over the results from individual model landscapes. Taken together with **Figure 2**, the results show how these metrics correspond with landscape shape.

state and the ability for cells to exist in multiple different phenotype states (i.e., the breadth of a valley in the potential landscape, and the number of different valleys). Notably, in the parameter regimes studied here, the highest Shannon Entropy models are single-phenotype (*hi/hi*), indicating that the noise in this one state contributes more disorder than does noise from multiple phenotype-states. As such, models with two or more accessible states have intermediate values of Shannon Entropy.

A strongly negative correlation coefficient between the two genes is found near the *lo/hi-hi/lo* vertex of the triangle, which is occupied by models showing bistability (cells can express one gene or the other, but not both simultaneously) resulting from mutual repression in the network motif. Landscapes with high positive correlation tend to be those that combine expression in the *hi/hi* and *lo/lo* quadrants of the two dimensional subspace (see, e.g., 4B and 2D), resulting from mutual activation in the network motif. Mutual Information aligns somewhat with large absolute values of Correlation Coefficients, but cannot distinguish high positive from high negative correlation. This arc-shaped region bisecting the triangle also overlaps with the models lacking interactions between the two genes (see **Figure 3C**).

The Coexpression Index shows the smoothest functional dependence on PCA components (c_1 , c_2). Of note, the model-subspace of high coexpression is not fully overlapping with the subspace of high correlation coefficients. This reflects the fact that high simultaneous expression occurs in both genes in an

uncorrelated manner, since the noise arises from aforementioned birth-death noise of mRNA transcription/degradation.

None of the four metrics are by themselves able to fully differentiate between landscape shapes. For example, model landscapes with similarly high values of Mutual Information include both *hi/lo-lo/hi* landscapes from mutual repression motifs and *hi/hi-lo/lo* landscapes from mutual activation motifs. (see, e.g., **Figures 4A, B**). Model landscapes with similar intermediate values of Coexpression Index also encompass a variety of landscape shapes, including some that arise from different network motifs (see, e.g., **Figures 4C, D**). Taken together, these results show that these four single metrics are not reliable determinants of landscape shape. They furthermore show that a given value for commonly used measures, as obtained from experimental data, can potentially arise from a variety of regulatory scenarios.

Stochastic Theory-Based Analysis of Coexpression Landscapes From Single-Cell Experiments Reveals Distinct Developmental “Landscape Shape” Trajectories

We applied the landscape shape analysis framework, developed above on the basis of theoretical models, to publicly available single cell RNA sequencing data in vertebrate development. We applied the analysis to putative MLP gene pairs in *Xenopus tropicalis* development collected at ten stages of embryonic development (Stages 8,10,11,12,13,14,16,18,20,22) (Briggs et al.,

2018). To carry out the analysis, we first analyzed the landscape shape-space for a restricted set of theoretical models, which encode only the MISA interaction motif. The MISA motif has been previously discovered to operate at critical cell-fate branch points (Graf and Enver, 2009) and has potential to enable both antagonistic expression and coexpression of genes in individual cells (depending on kinetic parameters), as is characteristic of MLP gene-pairs. We first generated a MISA-model training set (**Figure 5**) and the Two-Gene Flex-model training set (**Figure 2**). For MISA, we utilized quasipotential landscapes, rather than probability landscapes, in order to increase sensitivity to rarer cell-states (i.e., weaker landscape features). We furthermore restricted the kinetic parameters h , f to the fast (adiabatic) regime (see **Table 1**), in order to use the models to analyze time-resolved data. That is, the experiments measure embryos at different developmental stages, which are roughly 1–3 h apart in time. We compare the steady-state landscapes from stochastic models to the experiment-derived landscapes at different time points by applying a quasi-steady-state assumption: we assume that the promoter-binding states (which govern gene activity) reach equilibrium faster than the progression of developmental stage, which is valid only in the adiabatic regime. Despite these modifications to the model training set, the projection of models onto the PCA subspace for MISA (**Figure 5**) shows qualitative similarity to that of Two-Gene Flex [(**Figure 2**), including delineation of a subregion of a triangle (note that the triangle is inverted between the two figures, which is an arbitrary result of eigenvector sign invariance)]. However, antagonistic expression of the two genes is a stronger feature across models in the MISA training set, such that the *hi/hi* vertex of the triangle for MISA still shows considerable probability for cells to antagonistically express one gene or the other (**Figure 5F**).

We extracted two-gene coexpression quasipotential landscapes corresponding to distinct developmental stages from the dataset of Briggs et al. We then projected the landscapes onto the PCA subspace, and thereby derived developmental trajectories through landscape shape-space. By way of illustration, we first present developmental trajectories for three representative gene pairs (**Figure 6**). *Gata5* and *pax8* were identified (in Briggs et al.) as being antagonistically expressed within the intermediate mesoderm lineage, in cardiac mesoderm and pronephric mesenchyme cell subtypes, respectively. In contrast, *lhx1* and *pax8* were shown to coexpress in cells of the pronephric mesenchyme. Finally, the gene pair *sox2* and *brachyury* (*t*) has been identified as influencing the cell fate decision between the neural plate and the dorsal marginal zone (Wardle and Smith, 2004), and was identified as presenting MLP behavior, characterized by high coexpression at some stage of development, followed by antagonistic expression at a later stage (Briggs et al., 2018). We found that these three gene pairs showed distinctive trajectories through PCA subspace. All of the genes showed low expression early in development (stage 8) and their landscapes were colocated near the *lo/lo* vertex in the model subspace. Their trajectories then diverged: *gata5-pax8* travels along the bistable edge of the triangle, increasing expression of both genes over the course of development, but in largely

nonoverlapping subpopulations of cells. In contrast, *lhx1-pax8* shows strong coexpression starting at stage 14, and continues thereafter to move toward increasing values of PCA component 2, which coincides with increasing coexpression. (*lhx1-pax8* landscapes for some of the measured developmental stages fall slightly outside the area reached by MISA models in the training set, suggesting that the interaction is likely not well described by a MISA motif). Finally, *sox2-t* shows a cyclic pattern in the shape subspace, where landscapes move towards *hi/hi*, and then back towards the antagonistic *lo/hi-hi/lo* region, landing in a similar area to *gata5-pax8*. Relating these landscape-shape dynamics to the stochastic MISA model parameters suggests that the gene-pairs undergo changes in the relative balance of mutual inhibition versus self-activation as development progresses (see **Figure S1**).

The experiment-derived developmental trajectories can be further understood by considering the features extracted by individual (by definition orthogonal) PCA components. Visualization of the first three PCA eigenvectors (**Figure 7**) reveals that the first component (69.3% of covariance across the training set) can be summarized as separating landscapes with more or less expression overall, regardless of whether expression occurs in individual genes or both simultaneously. By contrast, the second component (15.6% of covariance) separates landscapes with coexpression versus antagonistic expression. The third component (6.8% of covariance) distinguishes landscapes with asymmetry between the two genes (subsequent components that describe less of the covariance displayed more complex shapes, and are not shown here). Comparison of the PCA scores versus developmental stage (**Figure 7**, right) to the experiment-derived landscapes of **Figure 6** confirms visually that the PCA components extract the above-described features. For example, all three gene pairs show varying degrees of asymmetry (imbalance in expression levels of the two genes). *Gata5-pax8* shows generally increasing positive amplitude of asymmetry, corresponding to stronger *pax8* expression. At later stages, the other two gene-pairs show asymmetry in the other direction, corresponding to negative amplitude in component 3. *Sox2-t* exhibits a switch in asymmetry between stage 10 ($t > \text{sox2}$) and later stages ($\text{sox2} > t$).

Developmental trajectories through the coexpression shape-space were compiled for 1,380 gene pairs (putative MLP pairs in *Xenopus tropicalis* identified by Briggs et al., 2018). By applying the developmental trajectory clustering procedure described in Methods, we found that the trajectories of multiple gene pairs across different lineages display conserved patterns of coexpression dynamics. Twenty-four clusters were identified (see **Supplemental Figures S3 and S4**), four of which are shown in **Figure 8**; these clusters are chosen as representative of the different types of dynamic patterns obtained. The clusters display a variety of behaviors. For example, the cluster of **Figure 8B** shows behavior that is consistent with MLP, i.e., genes are first increasingly coexpressed in single cells, followed by a switch towards antagonistic expression, similar to the cycle in PCA space delineated by *sox2-t* in **Figure 6**. Surprisingly, we also observed clusters that show “inverted MLP” behavior (**Figure 8A**) where the genes initially turn on in nonoverlapping subsets

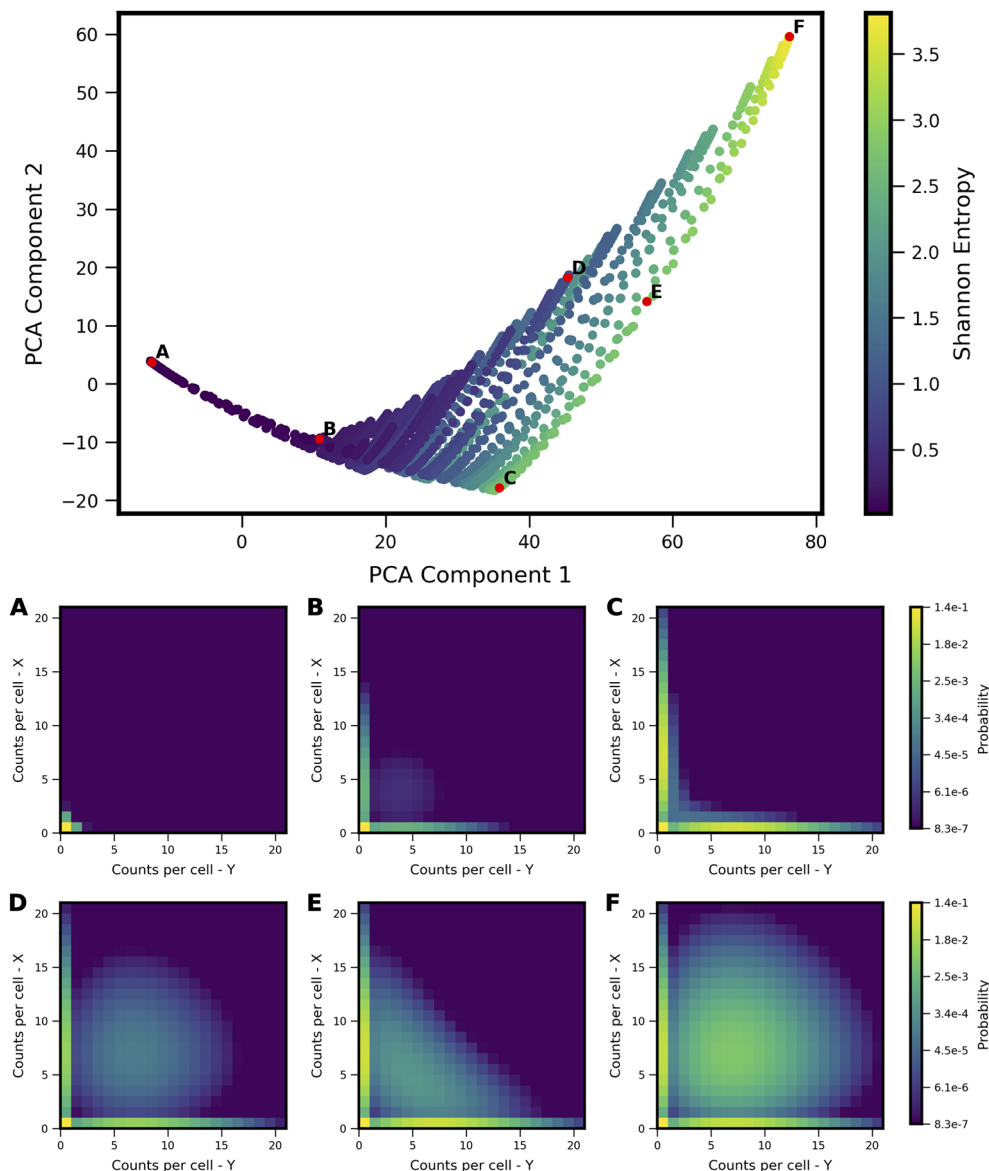


FIGURE 5 | Shape-space of simulated MISA coexpression landscapes analyzed by Principal Component Analysis (PCA). Coexpression landscapes were computed for 22,718 unique two-gene stochastic network models with Mutual Inhibition/Self-Activation (MISA) logic and varying kinetic rate parameters. Promoter-state change rates were restricted to the fast regime (see **Table 1**). (Top) All model landscapes projected onto the first two principal components. Each dot corresponds to one model, colored by the model's Shannon Entropy. (Bottom) Representative quasipotential landscapes $\phi(n)$ (see Text) of individual models from different regions of PCA component-space. Color of each discrete grid space in $\{x, y\}$ corresponds to computed probability (in log-scale) to find a single cell with the corresponding numbers of $\{x, y\}$ transcripts. (Analogous to **Figure 2**). Each landscape is labeled (A–F) to indicate the corresponding point where it appears in PCA space above.

of cells (i.e., increasing antagonism), but later show increasing coexpression in single cells. A number of the analyzed gene pairs showed generally antagonistic expression (**Figure 8C**), reminiscent of *gata5-pax8*. Others showed behavior consistent with the dynamics of MLP (i.e., first coexpression, later antagonistic expression), but with coexpression being only weakly detectable (**Figure 8D**). The gene pairs represented in these clusters include (but are not limited to) regulators of embryonic development including *zic3*, *hoxc10*, and *neurog1*.

The full list of clusters and their associated gene pairs are listed in the **Supplementary File 1**.

DISCUSSION

In this work, we comprehensively studied theoretically predicted single-cell gene-gene coexpression landscapes based on a class of stochastic gene regulation models, and applied the theory to

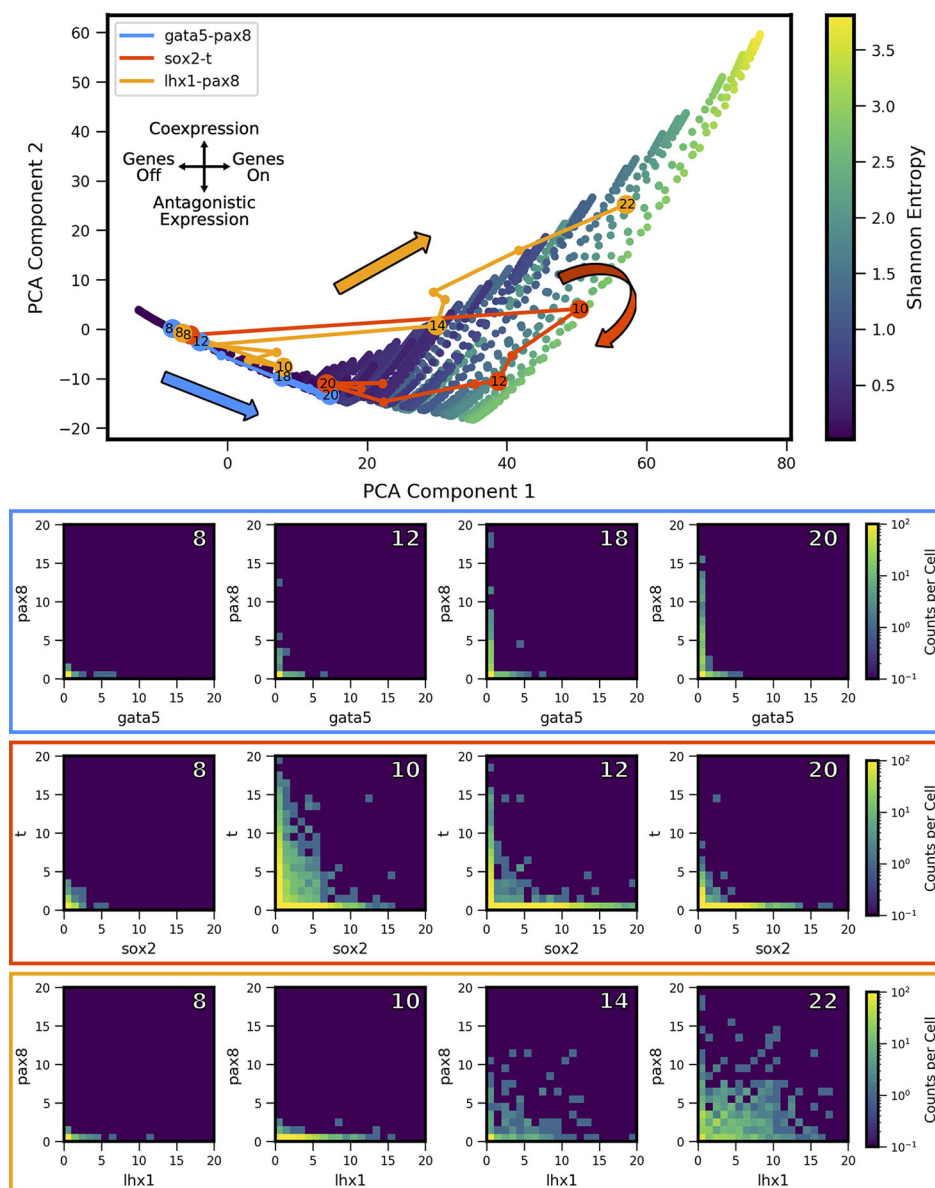


FIGURE 6 | Landscape-shape trajectories of three representative gene pairs from scRNA-seq measurements in *Xenopus tropicalis* embryonic development. (Top) Developmental trajectories of three different gene pairs, plotted in principal component-space. Stages of interest shown below are labeled with the corresponding stage. Note the three stage 8 points are overlapping near the origin as a result of low expression. (Bottom) Coexpression quasipotential landscapes extracted from experimental measurements for the three gene pairs at different labeled stages of embryonic development (white numbers indicate developmental stage). The experiment-derived landscapes were trained on the principal components generated from the simulated MISA dataset of **Figure 5**. Principal component 1 corresponds to overall level of expression, while component 2 separates antagonistic vs. coexpression (see **Figure 7**). The landscape of *gata5*-*pax8* (blue) shows increasing antagonistic expression, consistent with movement along the lower left edge of the triangle in Principal Component Analysis (PCA) shape-space. *Sox2*-*t* (red) shows high coexpression at stage 10, followed by later antagonistic expression, corresponding to a partial loop through PCA space, consistent with Multilineage Priming behavior. *Lhx1*-*pax8* (orange) shows consistently increasing coexpression, corresponding to a mostly steady increase in principal components 1 and 2. (Data from Briggs et al., 2018).

analyze two-gene coexpression landscapes from single cell measurements. From a training set of tens of thousands of computed, theoretical landscapes, we identify Principal Components of landscape covariance that serve as simple “fingerprints” of landscape shape and reflect underlying gene-

gene interaction dynamics. We then apply the theoretically derived framework to scRNA-seq data from vertebrate development. In so doing, we uncover distinctive and novel developmental trajectories of gene-gene coexpression. Specifically, our framework reveals a nuanced picture of

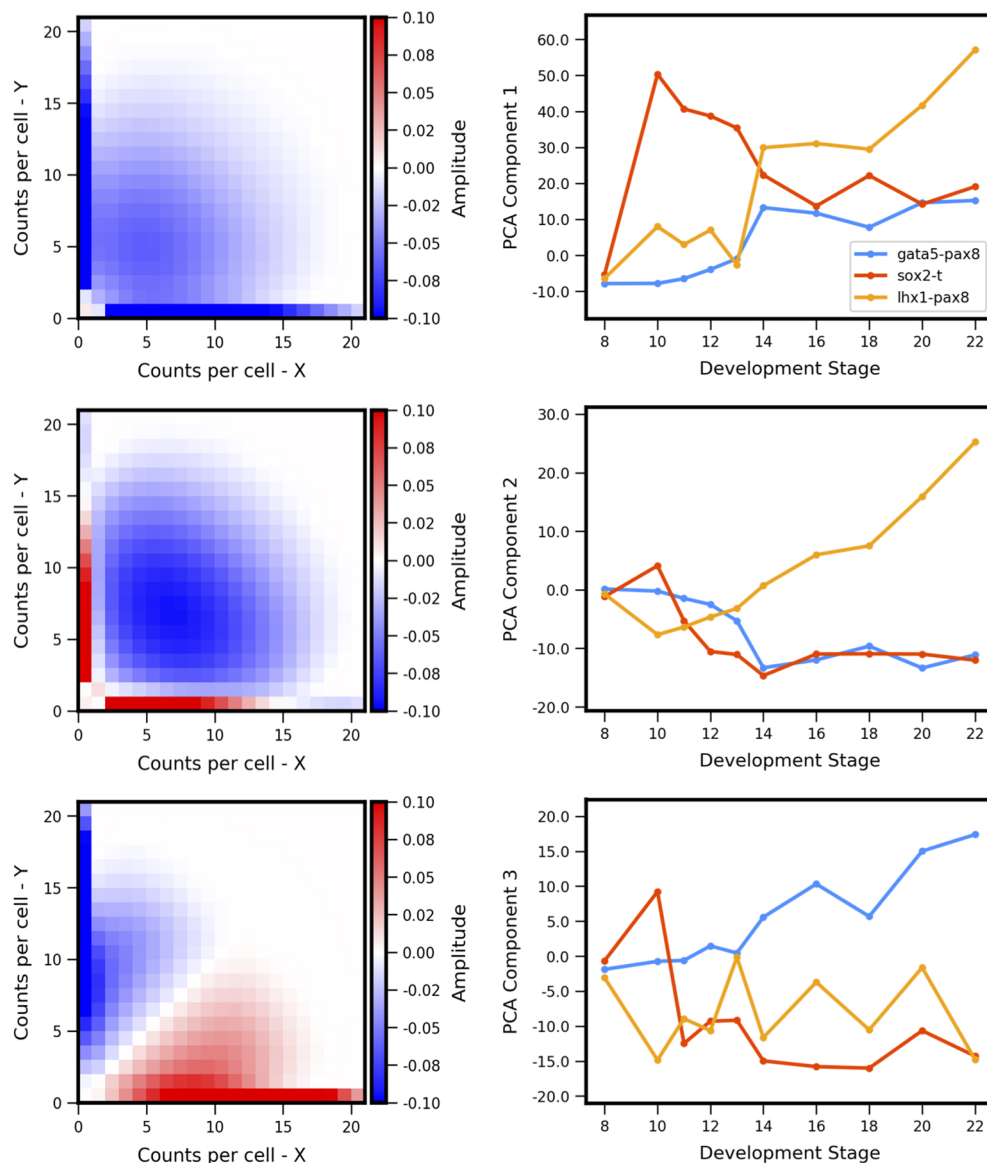


FIGURE 7 | Principal components of landscape shape features. (Left Column) The reshaped Principal Component Analysis (PCA) principal axes in feature space which represent the maximum variance in the data, specifically which features of the coexpression landscape that each component is accounting for. (Right Column) Magnitude or positive/negative value shift in observed variance for the respective component for each gene pair, versus developmental stage. Each component summarizes a landscape shape features: (Top Row) The overall amount of gene expression, (Middle Row) Antagonistic Expression vs. Coexpression of the two genes, and (Bottom Row) degree of asymmetric expression between the two genes.

multilineage priming, where the relative balance between expression of gene pairs simultaneously (in the same cells) versus antagonistically (in different cells) within a lineage shows complex dynamics during development, for example, revealing that simultaneous coexpression occurs either earlier or later than antagonism. Based on the results, we propose that the framework developed here can be generalized to other single cell datasets and stochastic network models to analyze the

evolution of gene-gene regulatory interactions over the course of development.

The theoretical framework applied here—discrete, stochastic reaction kinetic modelling—is well-suited to aid interpretation of single cell measurements: first, because it inherently captures cell population heterogeneity and second, because of the direct correspondence between the computed quantities (e.g., probability to find a given number of mRNAs in a cell) and

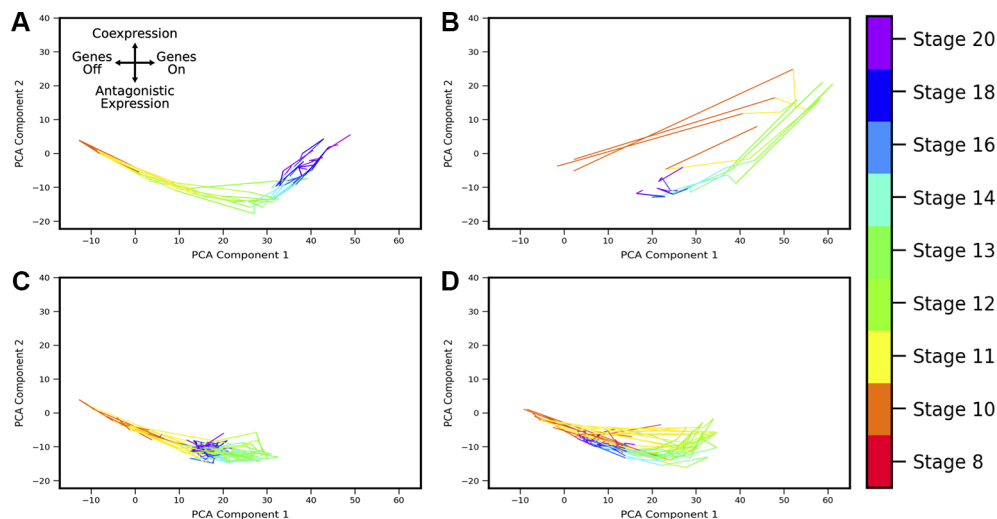


FIGURE 8 | Landscape shape trajectory clustering reveals conserved patterns of gene-pair coexpression dynamics during development. Four representative trajectory clusters showing distinct dynamics are presented (full list of 24 clusters and associated gene pairs in **Supplement**). Gene pairs in cluster **(A)** display behavior of an “inverted MLP”: first undergoing increasing antagonistic expression which then switches to increasing coexpression around stage 13. Gene pairs in cluster **(B)** follow the typical MLP behavior, with highest coexpression taking place around stage 10 followed by antagonistic expression at later stages. Cluster **(C)** shows consistent antagonistic expression (negative component 2), with nonmonotonic overall expression (a switch-back in component 1 around stage 12). **(D)** shows cyclic behavior similar to **(B)**, with highest coexpression at stage 12, but overall expression and relative amount of coexpression is lower.

experimentally measured transcript counts in scRNA-seq. The theoretical models can partially reproduce true cell population heterogeneity, but also neglect many sources of noise, both biological and technical. We employ models that treat intrinsic noise but neglect sources of persistent cell-to-cell variability (i.e., extrinsic noise) (Swain et al., 2002), which is known to contribute to noise in gene expression. For example, one source of extrinsic noise would be asynchronicity between cells, where individual cells might be at different stages of progression in development. Here, we opted to use a relatively simplistic model framework (i.e., no additional noise assumptions beyond intrinsic noise of biomolecular interactions, relatively few reactions describing molecular mechanisms of gene regulation, etc.) to minimize the number of model parameters while still enabling study of a variety of “rules” for gene regulatory logic. The framework presented here could be expanded in the future by integration of additional types of mechanistic assumptions and noise sources in the stochastic models.

The models also neglect technical noise/measurement errors arising from experiments (Grün et al., 2014). For example, scRNA-seq measurements face a well-known technical issue of drop-outs (Kharchenko et al., 2014), which we have not included in our modeling. Future efforts may improve the presented modeling framework by inclusion of these additional sources of noise, or by additional data-processing steps for imputation of missing data points (Gong et al., 2018). However, such an approach would also present challenges by necessarily introducing additional assumptions about cell population heterogeneity, which is still not fully understood. Given the danger of false signals (Andrews and Hemberg, 2019), we opted here to utilize minimal data processing in comparing our theoretical results to a public dataset. We also note

that the discrete stochastic modeling framework advanced in this work has potential to shed new light on the drop-outs issue: a relatively large proportion of “zeros” arises naturally from discrete stochastic models, depending on the regulatory interactions among genes, suggesting that perhaps biological variability plays a larger role in producing dropouts than has previously been supposed. Overall, despite the lack of additional biological/technical noise sources in our models, we note that our computed landscapes qualitatively reproduce the noise characteristics of the scRNA-seq measurements, in that they showed similarly broad distributions of coexpression. Thus we conclude that the simplistic models employed here are sufficient for the current application, which focused on characterization of coexpression landscape shape and its evolution in development, but we also foresee that incorporation of additional noise sources in the model might improve the practical utility of our proposed coexpression-shape-based analysis.

We focused here on two-gene models and pairwise interactions, because (1) certain gene-pairs are known to play a critical role in development (Graf and Enver, 2009) (2) the edges (pairwise interactions) are the elemental units or building blocks of larger regulatory networks. However, the focus on pairwise interactions has potential drawbacks: it does not elucidate how gene-pair interactions are modified when embedded in a larger network. In the same vein, it does not differentiate between direct or indirect interactions between genes (e.g., by direct transcriptional regulation versus molecular intermediaries). In principle, the framework presented here could be expanded to treat “3-body” (or higher order) interactions among genes, though this presents several computational challenges. For example, solution of the CME becomes intractable already for 3-gene networks, such that advanced approximation methods (Zhang and Wolynes, 2014) or more costly simulations

(Tse et al., 2018) become necessary. Nevertheless, expansion of the approach to higher-order interactions is feasible, and recent work has revealed how such an approach might proceed, for example, by incorporating developments in multivariate information measures (Chan et al., 2017).

In this work, linear PCA was used to identify shape features of gene-pair coexpression landscapes, and this approach was useful for separating landscapes with, e.g., more simultaneous coexpression versus more antagonistic expression for a given gene pair. Another possible extension of the method in the future could be to test alternative, nonlinear dimensionality reduction strategies for potential improvements in classifying coexpression landscapes based on desired features.

DATA AVAILABILITY STATEMENT

The simulation codes produced for this study can be found at https://github.com/Read-Lab-UCI/Stochastic_Coexpression_Landscape_Analysis.

REFERENCES

- Al-Radhawi, M. A., Vecchio, D. D., and Sontag, E. D. (2019). Multi-modality in gene regulatory networks with slow promoter kinetics. *PLoS Comput. Biol.* 15, e1006784. doi: 10.1371/journal.pcbi.1006784
- Andrews, T. S., and Hemberg, M. (2019). False signals induced by single-cell imputation. *F1000 Res.* 7, 1740. doi: 10.12688/f1000research.16613.2
- Arkin, A., Ross, J., and McAdams, H. H. (1998). Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. *Genetics* 149, 1633–1648.
- Bhattacharya, S., Zhang, Q., and Andersen, M. E. (2011). A deterministic map of Waddington's epigenetic landscape for cell fate specification. *BMC Syst. Biol.* 5, 85. doi: 10.1186/1752-0509-5-85
- Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J., et al. (2005). Transcriptional regulation by the numbers: models. *Curr. Opin. In Genet. Dev.* 15, 116–124. doi: 10.1016/j.gde.2005.02.007
- Bonnaïfoux, A., Herbach, U., Richard, A., Guillemin, A., Gonin-Giraud, S., Gros, P.-A., et al. (2019). WASABI: a dynamic iterative framework for gene regulatory network inference. *BMC Bioinf.* 20, 220. doi: 10.1186/s12859-019-2798-1
- Briggs, J. A., Weinreb, C., Wagner, D. E., Megason, S., Peshkin, L., Kirschner, M. W., et al. (2018). The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science* 360, eaar5780. doi: 10.1126/science.aar5780
- Cao, Y., and Liang, J. (2008). Optimal enumeration of state space of finitely buffered stochastic molecular networks and exact computation of steady state landscape probability. *BMC Syst. Biol.* 2, 30. doi: 10.1186/1752-0509-2-30
- Chan, T. E., Stumpf, M. P., and Babbie, A. C. (2017). Gene Regulatory Network Inference from Single-Cell Data Using Multivariate Information Measures. *Cell Syst.* 5, 251–267.e3. doi: 10.1016/j.cels.2017.08.014
- Chu, B. K., Tse, M. J., Sato, R. R., and Read, E. L. (2017). Markov State Models of gene regulatory networks. *BMC Syst. Biol.* 11, 14. doi: 10.1186/s12918-017-0394-4
- Dibaeinia, P., and Sinha, S. (2019). A single-cell expression simulator guided by gene regulatory networks. *Bioinformatics*. doi: 10.1101/716811. preprint.
- Feng, H., and Wang, J. (2012). A new mechanism of stem cell differentiation through slow binding/unbinding of regulators to genes. *Sci. Rep.* 2, 550. doi: 10.1038/srep00550
- Geertz, M., Shore, D., and Maerkl, S. J. (2012). Massively parallel measurements of molecular interaction kinetics on a microfluidic platform. *Proc. Natl. Acad. Sci.* 109, 16540–16545. doi: 10.1073/pnas.1206011109

AUTHOR CONTRIBUTIONS

Conceptualization and study design, CG and ER. Coding, CG. Analysis, CG, HR, and ER. Visualization, CG and HR. Writing, review, and editing, CG, HR, and ER. All authors read and approved the final version of the manuscript.

FUNDING

This work was (partially) supported by a NSF grant DMS 1763272 and a grant from the Simons Foundation (594598, QN). CG was provided support by a GAANN fellowship funded by the U.S. Department of Education.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.01387/full#supplementary-material>

- Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* 81, 2340–2361. doi: 10.1021/j100540a008
- Gong, W., Kwak, I.-Y., Pota, P., Koyano-Nakagawa, N., and Garry, D. J. (2018). DrImpute: imputing dropout events in single cell RNA sequencing data. *BMC Bioinf.* 19, 220. doi: 10.1186/s12859-018-2226-y
- Grün, D., Kester, L., and van Oudenaarden, A. (2014). Validation of noise models for single-cell transcriptomics. *Nat. Methods* 11, 637–640. doi: 10.1038/nmeth.2930
- Graf, T., and Enver, T. (2009). Forcing cells to change lineages. *Nature* 462, 587–594. doi: 10.1038/nature08533
- Gupta, A., Mikelson, J., and Khammash, M. (2017). A finite state projection algorithm for the stationary solution of the chemical master equation. *J. Chem. Phys.* 147, 154101. doi: 10.1063/1.5006484
- Hathaway, N., Bell, O., Hodges, C., Miller, E., Neel, D., and Crabtree, G. (2012). Dynamics and memory of heterochromatin in living cells. *Cell* 149, 1447–1460. doi: 10.1016/j.cell.2012.03.052
- Huang, S., Guo, Y.-P., May, G., and Enver, T. (2007). Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Dev. Biol.* 305, 695–713. doi: 10.1016/j.ydbio.2007.02.036
- Huang, S. (2012). The molecular and mathematical basis of Waddington's epigenetic landscape: a framework for post-Darwinian biology? *BioEssays* 34, 149–157. doi: 10.1002/bies.201100031
- Huang, S. (2013). Hybrid T-helper cells: stabilizing the moderate center in a polarized system. *PLoS Biol.* 11, e1001632. doi: 10.1371/journal.pbio.1001632
- Jin, S., MacLean, A. L., Peng, T., and Nie, Q. (2018). scEpath: energy landscape-based inference of transition probabilities and cellular trajectories from single-cell transcriptomic data. *Bioinformatics* 34, 2077–2086. doi: 10.1093/bioinformatics/bty058
- Kauffman, S. (1969). Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theor. Biol.* 22, 437–467. doi: 10.1016/0022-5193(69)90015-0
- Kepler, T. B., and Elston, T. C. (2001). Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophys. J.* 81, 3116–3136. doi: 10.1016/S0006-3495(01)75949-8
- Kharchenko, P. V., Silberstein, L., and Scadden, D. T. (2014). Bayesian approach to single-cell differential expression analysis. *Nat. Methods* 11, 740–742. doi: 10.1038/nmeth.2967
- Lam, S. K., Pitrou, A., and Seibert, S. (2015). “Numba: A llvm-based python jit compiler,” in *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC* (New York, NY, USA: ACM), 1–7. doi: 10.1145/2833157.2833162

- Lin, Y. T., Hufton, P. G., Lee, E. J., and Potoyan, D. A. (2018). A stochastic and dynamical view of pluripotency in mouse embryonic stem cells. *PLoS Comput. Biol.* 14, e1006000. doi: 10.1371/journal.pcbi.1006000
- MacArthur, B. D., Ma'ayan, A., and Lemischka, I. R. (2009). Systems biology of stem cell fate and cellular reprogramming. *Nat. Rev. Mol. Cell Biol.* 10, 672–681. doi: 10.1038/nrm2766
- Mariani, L., Schulz, E. G., Lexberg, M. H., Helmstetter, C., Radbruch, A., Löhning, M., et al. (2010). Short-term memory in gene induction reveals the regulatory principle behind stochastic IL-4 expression. *Mol. Syst. Biol.* 6, 359. doi: 10.1038/msb.2010.13
- McKinney, W. (2010). "Data structures for statistical computing in python," in *Proceedings of the 9th Python in Science Conference*, vol. 51–56. S. van der Walt and J. Millman
- Micheelsen, M. A., Mitarai, N., Sneppen, K., and Dodd, I. B. (2010). Theory for the stability and regulation of epigenetic landscapes. *Phys. Biol.* 7, 026010. doi: 10.1088/1478-3975/7/2/026010
- Mojtahedi, M., Skupin, A., Zhou, J., Castaño, I. G., Leong-Quong, R. Y. Y., Chang, H., et al. (2016). Cell fate decision as high-dimensional critical state transition. *PLoS Biol.* 14, e2000640. doi: 10.1371/journal.pbio.2000640
- Munsky, B., Li, G., Fox, Z. R., Shepherd, D. P., and Neuert, G. (2018). Distribution shapes govern the discovery of predictive models for gene regulation. *Proc. Natl. Acad. Sci.* 115, 7533–7538. doi: 10.1073/pnas.1804060115
- Nimmo, R. A., May, G. E., and Enver, T. (2015). Primed and ready: understanding lineage commitment through single cell analysis. *Trends In Cell Biol.* 25, 459–467. doi: 10.1016/j.tcb.2015.04.004
- Olsson, A., Venkatasubramanian, M., Chaudhri, V. K., Aronow, B. J., Salomonis, N., Singh, H., et al. (2016). Single-cell analysis of mixed-lineage states leading to a binary cell fate choice. *Nature* 537, 698–702. doi: 10.1038/nature19348
- Peccoud, J., and Ycart, B. (1995). Markovian Modeling of Gene-Product Synthesis. *Theor. Popul. Biol.* 48, 222–234. doi: 10.1006/tpbi.1995.1027
- Pedraza, J. M., and Paulsson, J. (2008). Effects of molecular memory and bursting on fluctuations in gene expression. *Science* 319, 339–343. doi: 10.1126/science.1144331
- Sasai, M., and Wolynes, P. G. (2003). Stochastic gene expression as a many-body problem. *Proc. Natl. Acad. Sci.* 100, 2374–2379. doi: 10.1073/pnas.2627987100
- Sasai, M., Kawabata, Y., Makishi, K., Itoh, K., and Terada, T. P. (2013). Time scales in epigenetic dynamics and phenotypic heterogeneity of embryonic stem cells. *PLoS Comput. Biol.* 9, e1003380. doi: 10.1371/journal.pcbi.1003380
- Schaffter, T., Marbach, D., and Floreano, D. (2011). GeneNetWeaver: in silico benchmark generation and performance profiling of network inference methods. *Bioinformatics* 27, 2263–2270. doi: 10.1093/bioinformatics/btr373
- Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., et al. (2011). Global quantification of mammalian gene expression control. *Nature* 473, 337–342. doi: 10.1038/nature10098
- Swain, P. S., Elowitz, M. B., and Siggia, E. D. (2002). Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci.* 99, 12795–12800. doi: 10.1073/pnas.162041399
- Thattai, M., and van Oudenaarden, A. (2001). Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci.* 98, 8614–8619. doi: 10.1073/pnas.151588598
- Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., et al. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* 32, 381–386. doi: 10.1038/nbt.2859
- Tse, M., Chu, B., Roy, M., and Read, E. (2015). DNA-binding kinetics determines the mechanism of noise-induced switching in gene networks. *Biophys. J.* 109, 1746–1757. doi: 10.1016/j.bpj.2015.08.035
- Tse, M. J., Chu, B. K., Gallivan, C. P., and Read, E. L. (2018). Rare-event sampling of epigenetic landscapes and phenotype transitions. *PLoS Comput. Biol.* 14, e1006336. doi: 10.1371/journal.pcbi.1006336
- van der Walt, S., Colbert, S. C., and Varoquaux, G. (2011). The numpy array: a structure for efficient numerical computation. *Comput. In Sci. Eng.* 13, 22–30. doi: 10.1109/MCSE.2011.37
- Waddington, C. H. (2014). *The strategy of the genes*. OCLC: 861212016.
- Wang, J., Zhang, K., Xu, L., and Wang, E. (2011). Quantifying the Waddington landscape and biological paths for development and differentiation. *Proc. Natl. Acad. Sci.* 108, 8257–8262. doi: 10.1073/pnas.1017017108
- Wardle, F. C., and Smith, J. C. (2004). Refinement of gene expression patterns in the early *Xenopus* embryo. *Development* 131, 4687–4696. doi: 10.1242/dev.01340
- Zhang, B., and Wolynes, P. G. (2014). Stem cell differentiation as a many-body problem. *Proc. Natl. Acad. Sci.* 111, 10185–10190. doi: 10.1073/pnas.1408561111
- Zhang, J., Yuan, Z., and Zhou, T. (2009). Synchronization and clustering of synthetic genetic networks: a role for cis-regulatory modules. *Phys. Rev. E* 79, 041903. doi: 10.1103/PhysRevE.79.041903
- Zhou, J. X., Samal, A., d'Hérousel, A. F., Price, N. D., and Huang, S. (2016). Relative stability of network states in Boolean network models of gene regulation in development. *Biosystems*, 142–143, 15–24. doi: 10.1016/j.biosystems.2016.03.002

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Gallivan, Ren and Read. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Quantification of the Underlying Mechanisms and Relationships Among Cancer, Metastasis, and Differentiation and Development

Chong Yu^{1,2}, Qiong Liu¹, Cong Chen³ and Jin Wang^{3*}

¹ State Key Laboratory of Electroanalytical Chemistry, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun, China, ² University of Science and Technology of China, Hefei, China, ³ Department of Chemistry and of Physics and Astronomy, State University of New York at Stony Brook, Stony Brook, NY, United States

OPEN ACCESS

Edited by:

Pavel Kraikivski,
Virginia Tech, United States

Reviewed by:

Areejit Samal,
Institute of Mathematical
Sciences, India
Dongya Jia,
Rice University, United States

*Correspondence:

Jin Wang
jin.wang.1@stonybrook.edu

Specialty section:

This article was submitted to
Systems Biology,
a section of the journal
Frontiers in Genetics

Received: 11 October 2019

Accepted: 19 December 2019

Published: 02 March 2020

Citation:

Yu C, Liu Q, Chen C and Wang J
(2020) Quantification of the Underlying
Mechanisms and Relationships
Among Cancer, Metastasis, and
Differentiation and Development.
Front. Genet. 10:1388.
doi: 10.3389/fgene.2019.01388

Recurrence and metastasis have been regarded as two of the greatest obstacles to cancer therapy. Cancer stem cells (CSCs) contribute to cancer development, with the distinctive features of recurrence and resistance to popular treatments such as drugs and chemotherapy. In addition, recent discoveries suggest that the epithelial mesenchymal transition (EMT) is an essential process in normal embryogenesis and tissue repair, as well as being a required step in cancer metastasis. Although there are many indications of the connections between metastasis and stem cells, these have often been studied separately or at most bi-laterally, not in an integrated way. In this study, we aimed to explore the global mechanisms and interrelationships among cancer, development, and metastasis, which are currently poorly understood. First, we constructed a core gene regulatory network containing specific genes and microRNAs of CSCs, EMT, and cancer. We uncovered seven distinct states emerging from the underlying landscape, denoted normal, premalignant, cancer, stem cell, CSC, lesion, and hyperplasia. Given the biological definition of each state, we also discuss the metastasis ability of each state. We show how and which types of cells can be transformed to a cancer state, and the connections among cancer, CSCs, and EMT. The barrier height and flux of the kinetic paths are explored to quantify how and which cells switch stochastically between the states. Our landscape model provides a quantitative approach to reveal the global mechanisms of cancer, development, and metastasis.

Keywords: landscape, kinetic path, cancer stem cell, epithelial mesenchymal transition, differentiation, metastasis

INTRODUCTION

Cell phenotypes change during the development of cellular differentiation (Wang et al., 2011; Xu et al., 2014). Differentiation starts from an oosperm, which develops into a complex biont system and continues into adulthood as stem cells (SCs) divide and generate differentiated daughter cells during tissue repair and cell regeneration (Sell, 2004). Induced pluripotent SCs (iPS) provide an opportunity for therapeutic use (Takahashi et al., 2007).

Adult cells were reprogrammed into pluripotent SCs in 2006 (Takahashi and Yamanaka, 2006). This was a significant step in SC and regenerative biology, as the cell type switching could skip many intermediate steps. This lineage reprogramming technology may also have profound implications for cancer biology.

Cancer is one of the most deadly diseases in humans. Studies show that there are multiple factors associated with recurrence and metastasis. Moreover, cancer is fatal mainly owing to metastasis (Cowin et al., 2005). Many studies have focused on the genetic origins of cancer (Muller and Vousden, 2013; Martincorena and Campbell, 2015). The accumulation of mutations leads to malignant transformation, which has been described as a disease of clonal evolution. Through such mutation and selection, cells acquire the hallmarks of cancer (Lynch et al., 1998; Hanahan and Weinberg, 2000). Some cells may acquire hypoxic and fast-growing characteristics, or may develop new blood vessels and so on. This is a widely accepted aspect of the generation of cancer. On the other hand, many observations have demonstrated that cancer could be thought of as an intrinsic state which emerges from underlying gene regulation networks (Kauffman, 1971; Spano et al., 2012), which control a series of cellular activities and biological processes. The network can provide regulatory instructions which may affect early events of cancer (Blancafort et al., 2013). These network environmental and epigenetic effects can result in not only silencing of tumor suppressors but also reactivation of the silenced regions, which could prime subsequent events in the development of cancer (Liu et al., 2008; Rodriguez-Paredes and Esteller, 2011).

Cancer SCs (CSCs) can be defined as cells with the characteristics of cancerousness and stemness (Tang, 2012). Although cancer cells might be killed during chemotherapy or immune surveillance, CSCs can survive as “seeds” of the cancer (Hanahan and Coussens, 2012; Kreso et al., 2013), explaining the recurrence of cancer after treatment. Although the CSC theory was reported as early as 1952 (Hewitt, 1952), its importance has only recently been understood. CSCs have been shown to serve as the basis of cancer development, maintenance, metastasis, and recurrence (Dragu et al., 2015). In general, the differentiation and development process is due to primary SCs, and reprogramming is vice versa; this is important in tissue reengineering. Cellular reprogramming involves iPS, indicating the possibility of cell fate switching and transformation (Kondo and Raff, 2000). However, reprogramming often results in a cancer state, resulting in the transformed progenitors acquiring self-renewal and cancerous characteristics (Pavlova and Thompson, 2016). Furthermore, CSCs facilitate the primary tumor cells to migrate from one location to another, which is a key step in the metastatic cascade.

Epithelial mesenchymal transition (EMT) is an essential process through which most adult tissues maintain their migratory capacity in normal embryogenesis, wound healing, and tissue repair (Morel et al., 2012). CSCs can also implant into another organ through the EMT process (Wang et al., 2015).

In EMT, a set of transcription factors (TFs) induce the early steps of metastasis (Scheel and Weinberg, 2012). Through EMT-TFs, differentiated epithelial cells can obtain mesenchymal traits to colonize foreign tissues and create new tumor sites in distant organs. Moreover, the EMT process is also the means by which non-SCs are transformed into SC states. Experiments have shown that inducing an EMT process during normal mammary epithelial cell differentiation can cause generation of mammary epithelial stem-like cells (Mani et al., 2008). This kind of experimental phenomena can be observed in both normal and cancerous tissues (Morel et al., 2008). Thereby, EMT is an important process which not only contributes to creating metastatic CSCs but also has a close relationship with CSCs (Chaffer et al., 2011).

Despite many results indicating the connections between metastasis and CSCs, or cancer and differentiation and development (Li and Wang, 2015), cancer, metastasis, and differentiation and development are rarely studied in an integrated way. Here, we aim to explore the connections among cancer, differentiation and development, and metastasis in a systematic and quantitative way. We start by constructing a core gene regulation network. In order to characterize the key points of the dynamic process, some specific genes and microRNAs of cancer, CSC, and EMT are included. In this work, we quantify the underlying landscape of cancer, metastasis, and differentiation and development. Furthermore, we include regulatory binding and unbinding information to make the model more precise. Seven states emerge from the landscape, which are quantified by the basins of attractions representing the normal, premalignant, cancer, SC, CSC, lesion, and hyperplasia states. In certain previous studies (Yu and Wang, 2016), normal, premalignant and cancer states were explored. In another model (Li and Wang, 2015), normal, cancer, CSC, and stem cell states were found. The lesion and hyperplasia states were not found in the previous theoretical studies but were observed in the experiments. They are predicted in our studies. We define these states by gene expression levels and biological significance. We also discuss the metastatic ability of these states. There are three pathways from the normal to the cancer state. Two kinetic paths which connect CSC state show the formation of cancer SCs from two sources. The optimal paths and barrier heights between the states illustrate how and which cells will be able to transform into the cancer state, and why cancer is so difficult to cure. This leads to a quantitative understanding of the degree of difficulty in curing the cancer. Moreover, the quantified landscape provides a portrait of the dynamic interrelationships among the biological processes of CSCs, EMT, and cancer. Finally, we use global sensitivity analysis to explore which regulatory process is more relevant to cancer therapy, which may provide guidance for future clinical experiments. This work helps to elucidate the origins of cancer, as well as the processes of differentiation and development in cancer and metastasis. This has clear clinical significance in understanding the role of CSCs in treatment response, therapeutic resistance, and cancer relapse.

RESULTS AND DISCUSSION

Model Construction

To emphasize the characteristics of CSCs, EMT, and cancer, a core gene regulatory network was constructed to cover specific genes and microRNAs of the three aspects, as shown in **Figure 1**. MDM2 is an oncogene of cancer, P53 is a well-known tumor suppressor gene (Yu and Wang, 2016), ZEB is an EMT activator gene which suppresses the stemness-inhibition of a microRNA (mir-200) (Wellner et al., 2009), OCT4 is an essential gene which mediates phenotype self-renewal and stemness (Kumar et al., 2012), and mir-145 and mir-200 are two important microRNAs with vital roles in both CSC and EMT regulation (Liu et al., 2015). The arrows represent activation and the short bars represent repression. The details of the regulatory network and gene function can be seen in **Tables S1** and **S2** in the **Supporting Information**.

The network includes ZEB–|mir-145–|OCT4 and OCT4–|mir-145–|ZEB regulation, indicating that ZEB effectively activates ($\rightarrow$) OCT4, while OCT4 effectively activates ($\rightarrow$)

ZEB. Therefore, stemness and metastasis promote each other: stemness can induce metastasis, and metastasis can also induce stemness. The regulation of OCT4–|mir-145–|MDM2 and MDM2–|P53–|OCT4 and ZEB–|mir-145–|MDM2 indicates that OCT4 effectively activates ($\rightarrow$) MDM2, while MDM2 effectively activates (MDM2 $\rightarrow$) OCT4, and ZEB effectively activates (ZEB $\rightarrow$) MDM2. MDM2 is known to be an oncogene. Therefore, stemness and metastasis can induce cancer, and cancer can also induce stemness. The regulation of ZEB–|mir-145–|MDM2 and MDM2–|P53 $\rightarrow$ mir145–|ZEB also shows that ZEB effectively activates ($\rightarrow$) MDM2, and MDM2 effectively activates ($\rightarrow$) ZEB. Thus, metastasis and cancer can promote each other. From the network wiring, we can gain certain information about the reinforcing relationships among stemness, cancer, and metastasis.

Methods

In our previous study (Yu and Wang, 2016), as the strong interaction, proteins and genes are treated as the same identity. We used differential equations to describe the gene regulatory network. The parameters are activation, repression and degradation items which describe the activation regulation rate, repression regulation rate and self-degradation rate, respectively. In this study, we use the chemical reactions to describe not only the protein concentration dynamics but also explicitly the underlying gene regulations (protein binding to the genes) and defined a set of rate parameters for each reaction to describe the gene regulatory process, which is stochastic. The underlying chemical reactions of gene regulation can be described as follows:

$$\mathcal{G}_1^{0\alpha\beta} + (n+1)P_1 \xrightleftharpoons[f_1]{h_1} \mathcal{G}_1^{1\alpha\beta} + (n)P_1 \quad (1)$$

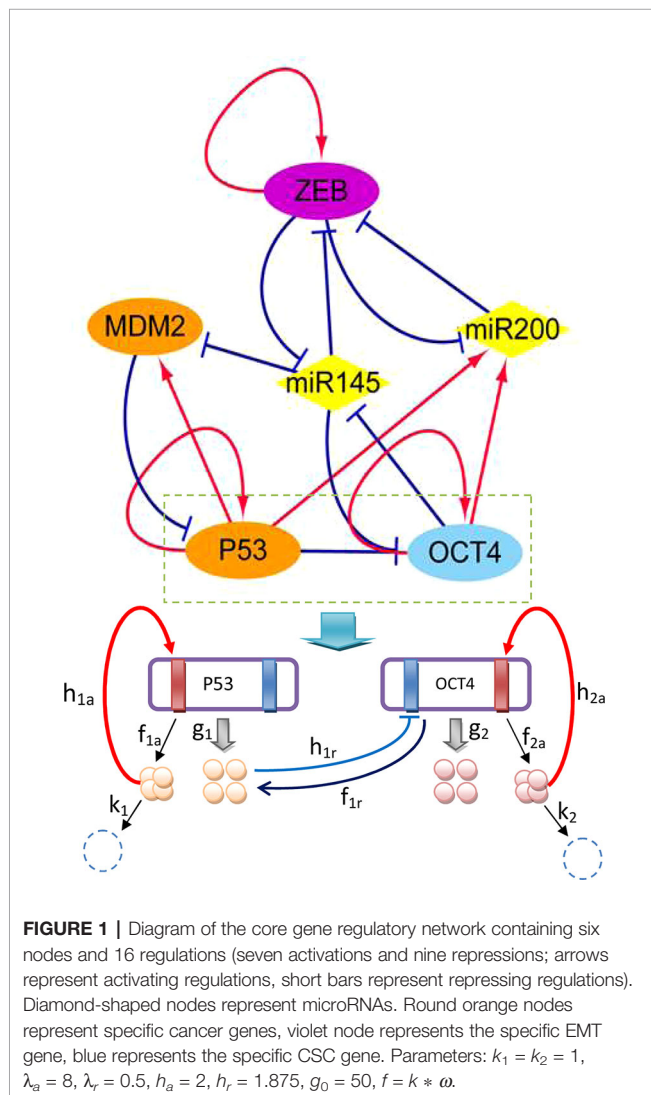
$$\mathcal{G}_1^{\alpha0\beta} + (n+2)P_2 \xrightleftharpoons[f_2]{h_2} \mathcal{G}_1^{\alpha1\beta} + (n)P_2 \quad (2)$$

$$\mathcal{G}_1^{\alpha\beta0} + (n+4)P_3 \xrightleftharpoons[f_3]{h_3} \mathcal{G}_1^{\alpha\beta1} + (n)P_3 \quad (3)$$

$$(n)\mathcal{G}_1 \xrightleftharpoons[k]{g} (n+1)\mathcal{G}_1, \quad (4)$$

where $\mathcal{G}_1$ represents a gene with three binding sites, 0 indicates the binding site which is unoccupied, and 1 indicates the binding site which is occupied. In the chemical reactions, the first binding site of $\mathcal{G}_1$ can be occupied by a monomer, the second binding site of $\mathcal{G}_1$ can be occupied by a dimer, and the third binding site can be occupied by a tetramer. P_i ($i = 1, 2, 3$) represents the type of the protein regulator. The parameter g represents the protein synthesis rate and k represents the protein degradation rate; h represents the binding rate and f is the unbinding rate of regulatory proteins to the target genes.

In **Figure 1**, we take gene regulation of P53 and OCT4 as an example to illustrate the regulatory process. The red rectangles



indicate activated binding sites for the genes, while the blue rectangles are repressed binding sites. P53 and OCT4 have protein synthesis rates of g_1 and g_2 , and protein degradation rates of k_1 and k_2 , respectively. The P53 and OCT4 proteins have binding rates of h_{1a} (a represents activation) and h_{2a} to their own activated binding sites, and unbinding rates of f_{1a} and f_{2a} from the binding sites. The P53 protein has a rate of h_{1r} (r represents repression) for binding to the repressing binding site of gene OCT4, and a rate of f_{1r} for unbinding from the binding site of OCT4.

For the first reaction (monomer binding site), the binding rate is given as $h_1 = h_{1n_1}$. For the second reaction (dimer binding site), the binding rate is given as $h_1 = h_{1n_2}(n_2 - 1)/2$. For the third reaction (tetramer binding site), the binding rate is given as $h_1 = h_{1n_3}(n_3 - 1)(n_3 - 2)(n_3 - 3)/6$. The protein synthesis rate is influenced by the regulated molecular number and regulated type. There are two regulated types: binding state and unbinding state. If the gene has n binding sites, it can give rise to 2^n synthesis rates. The synthesis rate can be increased by a factor of λ_a (a represents activation) or decreased by a factor of λ_r (r represents repression). If there are two binding sites, one for activation and the other for repression, the four synthesis rates are set as: $g_{00}, g_{01} = g_{00}\lambda_a, g_{10} = g_{00}\lambda_r$, and $g_{11} = g_{00}\lambda_a\lambda_r$. We define the equilibrium constant $X_{eq} = f/h$ and the adiabatic parameter $\omega = f/k$. The latter is used to quantify the ratio of the unbinding rate of a protein to the gene and its degradation rate. If the value of ω is large, the regulatory processes are relatively fast compared with synthesis and degradation; this is sometimes termed adiabatic. If the value of ω is small, it means the regulatory processes are relatively slow. In this model, we set the parameters $k = 1, g_0 = 50, \lambda_a = 8, \lambda_r = 0.5, h_a = 2$, and $h_r = 1.875$. If the protein switches on and off to the target gene relatively slowly, then the regulation process is non-adiabatic. In this work, we mainly focus on fast binding and unbinding, that is, the adiabatic case ($\omega = 1000$).

In the adiabatic case, the stochastic reactions can be described by a master equation (Gillespie, 2000), with a probability $P(x, t)$ of reaching the state x of the system at time t . The transition rates $M(x|x')$ are given as a matrix, for a system changing from state x to state x' , where

$$M(x|x') = \begin{cases} \geq 0, & \text{if } x \neq x' \\ -\sum_{x \neq x'} M(x|x'), & \text{if } x = x' \end{cases}$$

The master equation can be expressed as the rate of change of $P(x, t)$ for the combinations of possible transitions of x :

$$\partial P / \partial t = \sum_{x \in X} M(x|x') P(x, t) \quad (5)$$

The master equation can be further written in a more explicit form as:

$$\partial P / \partial t = (M_0 + M_b)P, \quad (6)$$

where the probability P is a state vector. Each component of P represents the probability of the system with a protein number at a gene state. M_0 is the diagonal part of the matrix M that represents the protein synthesis and degradation processes. M_b is the non-diagonal part of the matrix M which represents the

binding and unbinding regulation reactions. The binding and unbinding parts represent the reactions between gene states. The potential landscape of the gene regulation system can be defined as $U = -\ln P$ (Wang et al., 2008). In practice, we use Gillespie algorithm (Gillespie et al., 1977) to simulate the gene regulatory network and effectively solve the master equation (see details in the **Supporting Information**).

Definition and Metastatic Ability of Each Steady State and the Kinetic Paths of the Landscape

There are six nodes in our network. As it is difficult to visualize a six-dimensional space, we chose to discuss three specific genes, P53, ZEB, and OCT4, reflecting the cancer, EMT, and differentiation and development (with CSCs) aspects. P53 is a tumor suppressor gene. Normally functioning cells often have high gene expression levels of P53. Low gene expression of P53 is a general characteristic of cancer (Yang et al., 2013; Yu and Wang, 2016). OCT4 is a signature gene of SCs. Many studies have shown that OCT4 is critically involved in self-renewal and is a critical gene for cell differentiation and reprogramming (Lin et al., 2012). High gene expression of OCT4 indicates that cells have self-renewing ability, multi-differentiating potential, and strong proliferative ability. ZEB is a critical gene of the EMT process. The expression of ZEB can activate EMT, which is a required step in metastasis (Lamouille et al., 2014). The gene expression level of ZEB is a metastatic signature.

As shown in **Figure 2**, seven states emerge, which are denoted normal, premalignant, cancer, SC, CSC, lesion, and hyperplasia. In the normal state, the gene expression level of P53 is high, and those of OCT4 and ZEB are low. Thus, if cells remain in the normal state, they maintain normal function and do not have the characteristics of SCs such as self-renewal or reprogramming, nor do they have metastasis ability. This is consistent with the regulation in the network, in which OCT4 and ZEB effectively

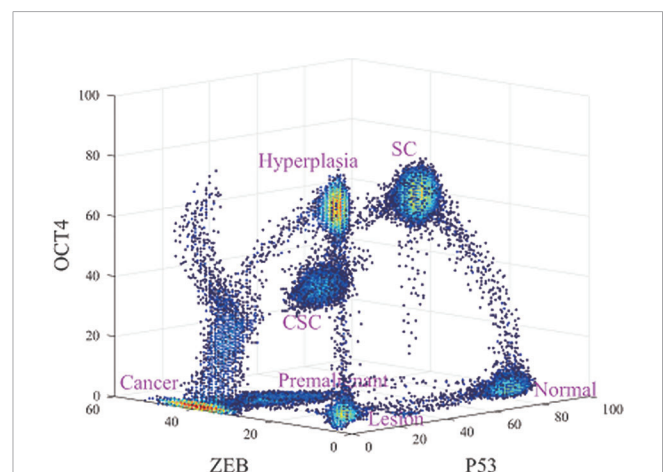


FIGURE 2 | Three-dimensional landscape showing the normal, premalignant, cancer, SC, Cancer stem cell (CSC), lesion, and hyperplasia states, and optimal paths among these states.

activate each other and both repress P53 (ZEB–|mir-145–|OCT4, OCT4–|mir-145–|ZEB; this is because OCT4 and ZEB activate each other. ZEB–|mir-145–|MDM2–|P53 can be seen as ZEB–|P53, and OCT4–|mir-145–|MDM2–|P53 can be seen as OCT4–|P53). Therefore, OCT4 and ZEB can both have low expression. In this case, there is no further repression of P53, resulting in high P53 expression. Overall, the gene expression levels indicate that the cells in the normal state are in a healthy condition without metastasis or self-renewal capability.

The lesion state involves low expression levels of P53, OCT4, and ZEB. Low gene expression levels of OCT4 and ZEB indicate that the cells do not have the characteristics of SCs or metastatic ability. Low gene expression levels of P53 indicate that the cells do not have normal function, which may be caused by inflammation, pH, hypoxia, and so on (Jeremy et al., 2003). This is again consistent with the gene regulatory wiring. Mutual effective activation can give rise to a low expression state for both OCT4 and ZEB, as discussed earlier. However, MDM2 represses P53 and P53 activates MDM2. It is possible for MDM2 expression to be high while P53 expression is low, or for MDM2 expression to be low while P53 expression is high. The former corresponds to the lesion state, while the latter corresponds to a normal cell state. In this case, the gene expression levels indicate that the cells are not in a healthy condition but do not have metastatic or self-renewal capability.

In the hyperplasia state, compared with the lesion state, the expression level of OCT4 is high and the expression levels of P53 and ZEB are low. A high gene expression level of OCT4 implies that the cells have the characteristics of SCs, such as self-renewal or reprogramming. The hyperplasia state can be seen as the accumulation of cell damage, while the tissues which are inflamed start the self-repair process which helps to produce new cells to replace the pathological cells. In this process, OCT4 is also a significant player in self-repair and DNA replication (Rizzino, 2013). Low gene expression levels of ZEB indicate that the metastasis is not significant. A low expression level of P53 indicates that the cells are still in an abnormal condition. This is consistent with the gene regulatory wiring. High OCT4 expression can repress P53 expression, maintaining the low expression of P53. OCT4 can also repress ZEB by another route to keep ZEB expression levels low (OCT4→mir200–|ZEB, the same as the SC state). OCT4 can self-activate to keep its expression high. The gene expression levels indicate that the cells in the hyperplasia state are not in a healthy condition and have strong self-renewal but not metastatic capability. We considered the hyperplasia state to be a tumor state without metastasis. Cells in the lesion and hyperplasia states both have a degree of damage, as the gene expression levels of P53 are low. In general, they can be reversed to a normal state by the self-healing system, as the expression level of ZEB is low, indicating that metastasis has not yet started. Tumors exist mainly in the hyperplasia state according to our definition, in which they have certain hallmarks of cancer such as overgrowth in some organs. However, tumors are not fatal until they are metastatic. When the tumor is already in a metastatic condition, it is considered to be cancer.

In the cancer state, the gene expression level of ZEB is high, and the gene expression levels of OCT4 and P53 are low. For cells in the cancer state, the tumor suppressor gene P53 shows low expression levels, but the metastatic ability is obvious (high gene expression level of ZEB). Moreover, cancer cells in the terminally differentiated stage have lost the ability to proliferate or to alter their destiny; their stemness ability is relatively low as well. Thus, the gene expression level of OCT4 is low. That is consistent with the gene regulatory wiring. High ZEB expression will repress P53 so as to keep P53 at a low expression level. OCT4 expression may be low owing to self-degradation, despite effective activation by ZEB. The low expression levels of P53 and OCT4 mean they cannot effectively repress ZEB by another regulation route, so ZEB expression remains at a high level. The gene expression levels indicate that the cells in the cancer state have very significant cancerous characteristics and metastatic capability. Thus, the cancer state represents a tumor with metastasis.

The premalignant state is a transition state between the normal and cancer states. In the premalignant state, the expression level of P53 decreases and that of ZEB increases when the cells transform from the normal to the cancer state, that is, the cancerization and metastasis become increasingly significant. The metastatic ability of the premalignant state is intermediate, bridging those of the normal state and the complete cancer/metastasis state. Moreover, the intermediate expression level of ZEB indicated that the EMT is also in an intermediate state, which known as the partial (hybrid) epithelial/mesenchymal (E/M) state (Kumar Jolly, 2015; Pastushenko et al., 2018). The partial EMT can be considered as primary bad actors of metastases. This is consistent with the gene regulatory wiring. Relatively higher ZEB expression will repress P53 so as to keep P53 expression at a relatively low level. OCT4 expression may be low owing to self-degradation despite effective activation by ZEB. The relatively lower expression of P53 and low expression of OCT4 mean that they cannot effectively repress ZEB by another regulatory route, so ZEB expression remains at a relatively high level. The gene expression levels indicate that cells in the premalignant state have certain cancerous characteristics, partial EMT phenotype and metastatic capability. Therefore, we considered the premalignant state to represent tumors with a certain level of metastasis.

In the SC state, the gene expression levels of P53 and OCT4 are high, and that of ZEB is low. Cells in this state have stemness activity, so the expression levels of OCT4 are high. The expression level of P53 is also high and that of ZEB is low, indicating that the cells are functioning normally and their metastatic ability is not active. This is consistent with the regulatory wiring. When both OCT4 and ZEB have high expression as a result of their effective mutual activation, their repression leads to low P53 expression. When the expression of ZEB is low, its repression of P53 is weak, leading to high expression of P53, which represses OCT4. However, OCT4 can sustain its high level of expression through self-activation. OCT4 is involved in an alternative route (OCT4→mir200–|ZEB); the

protein concentration determines which path is dominant. If the concentration of mir-200 is dominant, the route is repression. This route can repress ZEB and keep its expression levels low. The gene expression levels indicate that the cells in the SC state are in a healthy condition with strong self-renewal capability, but without the metastatic capability.

In the CSC state, the expression levels of P53, OCT4, and ZEB are all intermediate. These cells are in a transition between the SC and the cancer state. CSCs show some characteristics of cancerization and self-renewal (stemness), as their gene expression level of P53 is lower and that of OCT4 is higher than in the normal state. Moreover, the elevated gene expression of ZEB indicates that the cells have a certain metastatic ability and an intermediate EMT phenotype, in between those of the SC and cancer state. Many studies suggest that partial EMT associates with Stemness. Cells in partial EMT state are most likely to gain stemness (Strauss et al., 2011; Kumar Jolly et al., 2014; Grosse-Wilde et al., 2015). This is consistent with the regulatory wiring. When both OCT4 and ZEB have high expression levels as a result of their effective mutual activation, their repression leads to low P53 expression. The gene expression levels indicate that cells in the CSC state have certain cancerous characteristics, self-renewal capability, a partial EMT phenotype and metastatic ability. Therefore, the CSC state represents a tumor with a certain degree of stemness and metastasis.

We also compared our landscape with the experimental data. To quantify the landscape from the experimental results, we

project the data to 3 dimensions in expression levels of P53, ZEB, and OCT4. This projection for the landscape can be used to reflect the characteristic features of the cancer, EMT, and differentiation/development (with CSCs). The RNA-seq data can only reflect the gene expressions at the transcriptional level. The protein concentrations reflect the gene expressions at the post translational level. Due to the post transcriptional and post translational influences, the RNA-seq data may not be able to completely determine the activities of these genes.

Instead of directly using the individual gene RNA-seq data, we consider some other genes which are regulated by or indirectly regulated by the individual gene (P53 representing the cancer group for example). The downstream genes transcription levels (18 of them related to P53 in this example) are determined by the upstream genes protein activities (post translation level). Therefore, these 18 genes are also cancer related genes and their genes transcriptional data can be used to reflect p53s post translation level gene expressions in some respect. This can lead to more complete information on P53 gene expressions at the post translation level beyond the transcription level which is crucial for describing the function. It serves as the rationale for choosing more genes (total of 40) instead of individual genes (six genes) we focused on at the beginning of the analysis. We then analyzed these three groups of experimental data using principal component analysis. By selecting the first principal component for each group, respectively, the RNA-seq data could be reduced to three

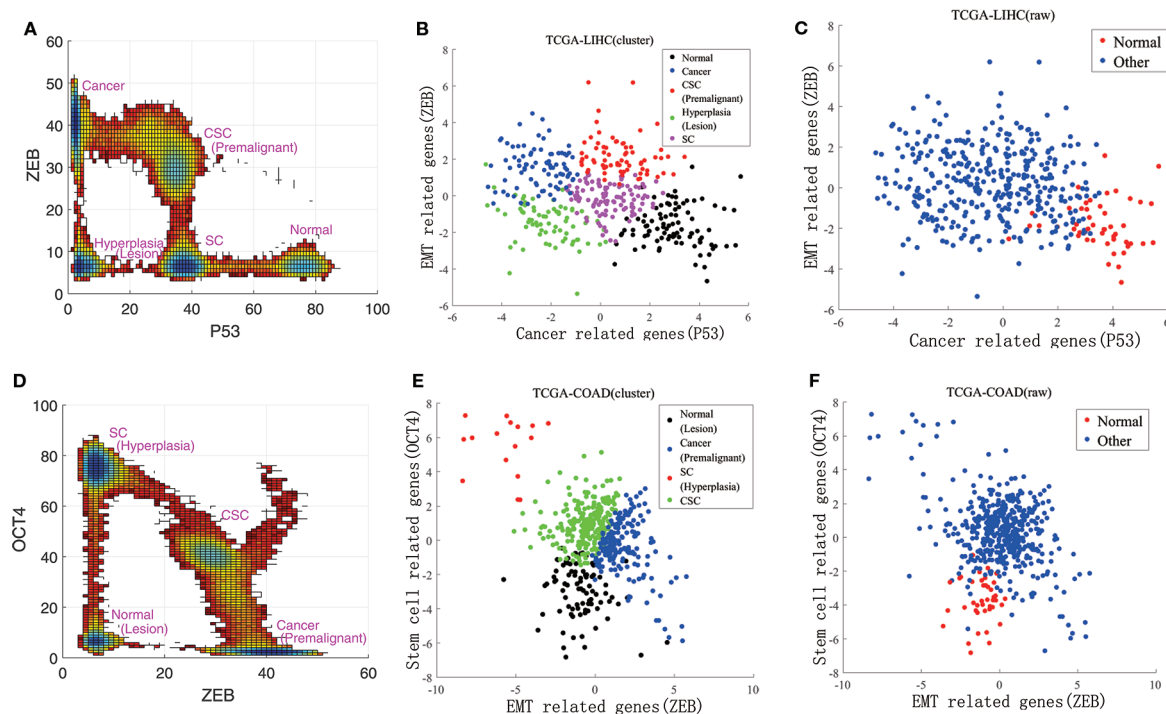


FIGURE 3 | Comparisons of experimental data and the steady states of our landscape. (A) is our landscape projection to X and Y axes. (B, C) show the data clustering and raw data for LIHC. (D) is our landscape projection to Y and Z axes. (E, F) show the data clustering and raw data for COAD.

dimensions. In **Figures 3B, C, E, and F**, RNA-seq data are represented by their first principal component. **Figure 3A** shows our landscape projection to the X and Y axes. As the CSC state coincided with the premalignant state, and the hyperplasia state coincided with the lesion state, there were five steady states: normal, CSC (pre-malignant), cancer, hyperplasia (lesion), and SC. In **Figure 3B**, there are five clusters which correspond to the five states. **Figure 3D** shows our landscape projection to the Y and Z axes. As the normal state coincided with the lesion state, the hyperplasia state coincided with the SC state, and the cancer state was connected with the premalignant state, there were four steady states: normal (lesion), cancer (pre-malignant), CSC, and SC (hyperplasia). In **Figure 3E**, there are four clusters which correspond to the four states. **Figures 3C, F** show the raw data for LIHC and COAD, which were used to validate the clustering results in **Figures 3B, E**. The normal state positions of the cluster results coincide with those of the raw data.

From a landscape perspective, there were several major kinetic paths that could be quantitatively explored. When the expression level of ZEB increases, the paths from the SC to CSC and the CSC to cancer states become prominent. These two paths show that the formation of CSCs has two main sources. One route of CSC generation involves somatic SCs with self-renewal capabilities; these have the potential to divide into both SCs and specialized somatic cells, which are destined to stop proliferating or die (Lobo et al., 2007). If these SCs are out of control with respect to stopping division, but still keep their self-renewal and differentiation abilities, they become CSCs (Ponti et al., 2005; Ye and Weinberg, 2015). Another route for generating CSCs exists owing to a minor proportion of cancer cells with the capacity for self-renewal and differentiation in their progeny (Liu et al., 2015). Many experiments have demonstrated that terminally differentiated cancer cells can gain SC properties under specific epigenetic conditions (Tang, 2012). These SC-like cancer cells drive cell growth and metastasis, and are considered to be CSCs. Many reports have shown that cancer cells undergoing EMT can obtain SC-like characteristics (Mani et al., 2008), demonstrating the connection between EMT and CSC. These have been found in hematopoietic (Bonnet and Dick, 1997) and solid tumors such as brain (Singh et al., 2004) and breast cancers (Alhajj et al., 2003). These two paths driving CSC generation lead independently to the capacity for self-renewal, differentiation, and migration. The kinetic paths in the landscape view illustrate the dynamic transitions of SCs, CSCs, and cancer. Owing to these diversifications, CSCs present a major challenge with respect to drug resistance and cancer recurrence.

The landscape view also shows that there is more than one pathway from the normal to the cancer state. There are at least three major paths. The first is from the SC to CSC to cancer state. Stem cells can gain cancer characteristics and become CSCs. Recently, some studies tracing CD133+ cells have provided direct evidence that SCs are susceptible to cancerous transformation (Medema, 2013; Zhu et al., 2016). CSCs inherit many characteristics of SCs, including self-renewal and differentiation. Moreover, CSCs have cancerization characteristics such as uncontrollable growth and

metastasis. CSCs can be asymmetrically divided into cancer cells and CSCs (Sell, 2004). Thus, CSCs can be seen as the seeds of cancer cells. This path involves both stemness through CSCs and the metastasis (or EMT) process (half-metastasis state for CSCs). When cells are in the SC state, the gene expression levels of P53 and OCT4 are high, but that of ZEB low. This indicates that the cells are in a healthy condition and have stemness but not the metastatic feature. When the cells transform to the CSC state, the gene expression levels of P53 and OCT4 both decrease, and that of ZEB increases. This indicates that the cells in the CSC state become cancerous and have a certain metastatic ability.

The second pathway is from the normal to premalignant to cancer state. This can be seen as a cancerous process. On this path, the gene expression level of P53 decreases and that of ZEB increases. This indicates that the cells not only show a trend of pathological changes but also have metastatic characteristics. This path involves metastasis or EMT, since the premalignant state is a half-cancer and half-metastasis. The cells in the normal state are in a healthy condition and do not have the metastatic or stemness features, as the gene expression level of P53 is relatively high but those of OCT4 and ZEB are low. When the cells transform to a premalignant state, the gene expression level of P53 decreases and that of ZEB increases. This indicates that the cells exhibit half-cancerous and half-metastatic features. When the cells are in the cancer state, the gene expression level of P53 is low and that of ZEB is high. OCT4 gene expression is also low. This indicates that the cells in the cancer state gain a strong metastatic ability and are differentiated without stemness.

The third pathway is from the normal to lesion to hyperplasia to cancer state. This can be seen as a process by which normal cells develop into the lesion state, gaining proliferation ability (hyperplasia), then turning malignant and eventually achieving the cancer state. Some experiments have shown that a lesion often occurred before the hyperproliferative changes (Jeremy et al., 2003). Hyperplasia is accumulated to a certain degree; cells possess metastatic ability and ultimately transform to the cancer state. This pathway involves stemness and the EMT process during its last stage. The cells in the lesion state are not in a healthy condition and do not have the stemness or metastasis features, as the gene expression level of P53, ZEB, and OCT4 are low. When the cells are in the hyperplasia state, the gene expression level of OCT4 becomes high, although the expression levels of the other two genes do not change significantly. This indicates that the cells in the hyperplasia state have significant stemness features, as cell damage induces their self-renewal ability to enable self-repair (Rizzino, 2013). However, the gene expression level of ZEB remains low. This indicates that cells in the hyperplasia state do not have the metastatic feature. When the cells reach the cancer state, the gene expression level of ZEB becomes very high and metastasis is obvious. These three paths address a central question in cancer biology: how and which cells can be transformed to cancer. These results also indicate that cancer is difficult to cure because the formation of these paths.

Barrier Heights and Flux of Kinetic Paths

Figure 4 shows the barrier heights between the normal, premalignant, cancer, SC, CSC, lesion, and hyperplasia states. In path 1, the barrier height from the SC to the CSC state was 6.0189, and that from the CSC to cancer state was 3.5048. We can describe the carcinogenesis of SCs with a high barrier as less likely to occur, as strong regulatory and environmental conditions are required to make the SCs cancerous (Tang, 2012). With a lower barrier, differentiation of CSCs to cancer cells is an easy process, as the CSCs can generate cancer cell progeny when they divide. A CSC can be asymmetric divided into a cancer cell and another CSC (Sell, 2004). The barrier height from the SC to the normal state was 6.0509, which is comparable to the barrier from the SC to CSC state. The SC state has two choices: to become a normal differentiated cell or a CSC, both with certain degrees of difficulty. In adults, somatic SCs are always dormant; specific conditions are required to induce them to divide. On the other hand, reprogramming requires specific gene regulation. Therefore, the barriers for both differentiation and reprogramming are relatively high. When SCs are activated, they are asymmetric divided into SCs and normal somatic cells. It appears that in the SC state, the cell can switch to either a differentiated cell or a CSC. The paths connecting the CSC state to the SC state and the cancer state had barriers of 3.0438 and 3.5048, respectively. This illustrates that when cells stay in the CSC state, they are both very unstable owing to the low barrier height and more likely to transform to the cancer state or back to the SC state. The barrier height from the cancer to the CSC state was also high, at 9.11, which means it is difficult for cancer cells to transform back to CSCs. Experiments have revealed that only a minor proportion of cancer cells have the capacity for self-renewal and differentiation in their progeny (Liu et al., 2015).

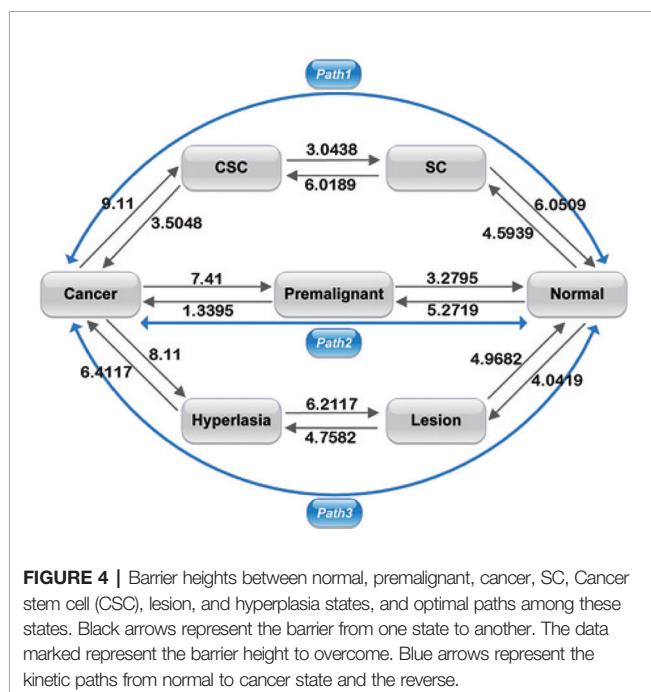


FIGURE 4 | Barrier heights between normal, premalignant, cancer, SC, Cancer stem cell (CSC), lesion, and hyperplasia states, and optimal paths among these states. Black arrows represent the barrier from one state to another. The data marked represent the barrier height to overcome. Blue arrows represent the kinetic paths from normal to cancer state and the reverse.

Therefore, the switching from cancer cells to CSCs is not easily realized. We can state that path 1 has the characteristics of both SCs and metastasis. Cells going through path 1 from the normal state can acquire stemness and metastasis, and eventually reach the cancer state.

In path 2, the barrier heights from the premalignant to the normal state and from the normal to the premalignant state were 3.2795 and 5.2719, respectively. The lower barrier height from the premalignant to the normal state compared with that from the normal to the premalignant state shows that it is relatively difficult for a cell in the normal state to be transformed to the premalignant state, whereas a cell in the premalignant state can relatively easily revert back to the normal state. Moreover, the barrier heights from the premalignant to the cancer state and from the cancer state to the premalignant state were 1.3395 and 7.41, respectively. This illustrates that it is much easier for a cell in the premalignant state with an intermediate level of metastasis to transform to the cancer state than the reverse from the cancer state back to the premalignant state. The barriers between the premalignant state and the normal and cancer states were lower; thus, a cell state can transform to the normal or cancer state relatively easily. The fatality of cancer is due to uncontrolled diffusion and metastasis; if cells are in the cancer state, metastasis is obvious. So, the premalignant state with an intermediate level of tumor characteristics and metastasis has vital clinical significance with respect to early diagnosis and prevention of cancer, as cells in the premalignant state can transform to cancer or revert back to a normal state easily. Therefore, path 2 has the characteristics of metastasis. Cells going through path 2 reflect the metastatic process, as the premalignant state is an intermediate state of metastasis. The importance of the premalignant state was discussed in our previous study (Yu and Wang, 2016).

In path 3, the barrier heights between the normal, lesion, and hyperplasia states were 4.0419–4.9682 and 4.7582–6.2117, that is, they were not very high. This means that it is not very difficult for cells to transform from one state to another. Experiments have shown that lesions commonly occur before hyperproliferative changes (Jeremy et al., 2003). However, the barrier heights between the hyperplasia and cancer states were 6.4117–8.11. This means that transformation from hyperplasia to cancer and the reversion from cancer to hyperplasia are both difficult, and the cancer to hyperplasia transition is unlikely to occur. That is, it is not difficult for cells to transform from one state to another before metastasis (transformation to the cancer state) occurs. If the cells have not reached metastasis, it is relatively easy for the cancer to be cured (reversion of cells to a normal state). When cells become cancerous owing to hyperplasia, this is a difficult process, but it is even more difficult to escape from the cancer state to hyperplasia as a very high barrier needs to be overcome. Therefore, path 3 reflects the process of accumulated cell damage resulting in metastasis. Cells in path 3 go through increasing pathological changes and eventually reach the cancer state.

As shown in **Figure 4**, the paths connecting the cancer state to the CSC, premalignant, and hyperplasia state had relatively high barriers of 9.11, 7.41, and 8.11, respectively. That is, the barriers

of the cancer state are all very high. Thus, cells cannot easily escape the cancer state, which explains why cancer is so difficult to cure.

We also calculated the correlation of the transition time with the barrier height, obtaining a correlation coefficient of 0.80. As **Figure 5** shows, the transition time and the barrier height show almost the same trend.

We also compared the flux of the three paths (from the normal to the cancer state, and the reverse). The flux of each path indicates which path is more important in cancer formation. According to the transition time and the probability of each pathway, we could quantify the flux of each path. The transition time in our work depended on the landscape topography, which is reflected by the barrier heights of each state. The transition rate k of a pathway is the reciprocal of the transition time. The details and data can be seen in the **Supporting Information**. This method was used in our previous work (Wang et al., 2013). The flux of the path normal $\rightarrow$ SC $\rightarrow$ CSC $\rightarrow$ cancer was 2.2157×10^{-10} . The probability of this path was 0.0719. The flux of the path normal $\rightarrow$ premalignant $\rightarrow$ cancer was 2.6227×10^{-9} . The probability of this path was 0.8509. The flux of the path normal $\rightarrow$ lesion $\rightarrow$ hyperplasia $\rightarrow$ cancer was 2.3813×10^{-10} . The probability of this path was 0.0773. The flux and the probability of the path normal $\rightarrow$ premalignant $\rightarrow$ cancer account for the vast majority of the three. Thus, this path is dominant for the transition from the normal to the cancer state and should therefore be the focus to prevent cancer formation. We demonstrated the importance of the premalignant state for cancer prevention in our previous work (Yu and Wang, 2016).

In the same way, we could also quantify the flux from the cancer to the normal state. The flux of the path cancer $\rightarrow$ CSC $\rightarrow$ SC $\rightarrow$ normal was 2.1830×10^{-9} . The probability of this path was 0.4243. The flux of the path cancer $\rightarrow$ premalignant $\rightarrow$ normal was 9.3693×10^{-10} . The probability of this path was 0.1821. The

flux of the path cancer $\rightarrow$ hyperplasia $\rightarrow$ lesion $\rightarrow$ normal was 2.0245×10^{-9} . The probability of this path was 0.3935. The flux of path 1 (cancer $\rightarrow$ CSC $\rightarrow$ SC $\rightarrow$ normal) and that of path 3 (cancer $\rightarrow$ hyperplasia $\rightarrow$ lesion $\rightarrow$ normal) were very similar, and both were higher than that of path 2 (cancer $\rightarrow$ premalignant $\rightarrow$ normal). These paths are important in the reversal of the cancer state back to the normal state. The flux and probability of path 1 were higher than those of path 3, so this path is dominant for the cancer to normal state transition. These results indicate the importance of CSCs in cancer therapy.

These three paths can be used to address a central question in cancer biology, how and which cells can be transformed to cancer, in a quantitative way. The barrier heights describe the basin depths of the landscape and help in understanding the tendency of the cells to transform from one state to another. Furthermore, the barrier heights of the cancer state are all very high, meaning that the cells in the cancer state transform less readily to others. The presence of multiple cancer formation paths explains the various mechanisms of cancer formation, which are among the reasons that cancer is difficult to prevent. The flux of the paths indicate which path is dominant in cancer formation and help to describe in a quantitative way the difficulty of curing a particular cancer.

Finding Key Regulations by Global Sensitivity Analysis of Landscape Topography

To gain further insight into cancer formation, we explored the network to find the key regulations by global sensitivity analysis of the landscape topography. In the network, each gene and regulation contributes to the network dynamics. Variation of the regulatory strengths will influence the barrier heights between attractor basins. In this way, we could determine which regulations were more sensitive for cancer formation in the network. The results may provide a reference for drug design for cancer therapy.

Figures 6A, B display the variation of the regulation miR200 $\downarrow$ ZEB; regulation 1 is miR200 $\downarrow$ ZEB in **Figures 6A, C**, and **D** display the variation of regulation OCT4 $\rightarrow$ OCT4; regulation 1 is OCT4 $\rightarrow$ OCT4 in **Figures 6C, E**, and **F** display the variation of regulation P53 $\rightarrow$ P53; regulation 1 is P53 $\rightarrow$ P53 in **Figure 6E**. In **Figures 6A, C**, and **E**, the control regulations 2–13 are P53 $\rightarrow$ miR200, P53 $\rightarrow$ miR145, P53 $\rightarrow$ MDM2, miR145 $\downarrow$ ZEB, miR145 $\downarrow$ OCT4, miR145 $\downarrow$ MDM2, ZEB $\downarrow$ miR200, ZEB $\downarrow$ miR145, ZEB $\rightarrow$ ZEB, OCT4 $\rightarrow$ miR200, OCT4 $\rightarrow$ miR145, and MDM2 $\downarrow$ P53, respectively.

As shown in **Figure 6A**, we increased the regulation strength to 1.5 times. In regulation 1 (miR200 $\downarrow$ ZEB), the barrier height from the premalignant state to the cancer state increased significantly, and the barrier height from the cancer state to the premalignant state decreased slightly. Although regulation 6 also changed very significantly, this was discarded as it changed in the same direction. ZEB is an EMT activator gene; when its gene expression level is high, metastasis becomes obvious. Thus, when we increased the suppression strength of ZEB, the expression level decreased, leading to weaker metastasis. In that case, it is much more difficult for the cell state to move from premalignant to cancer,

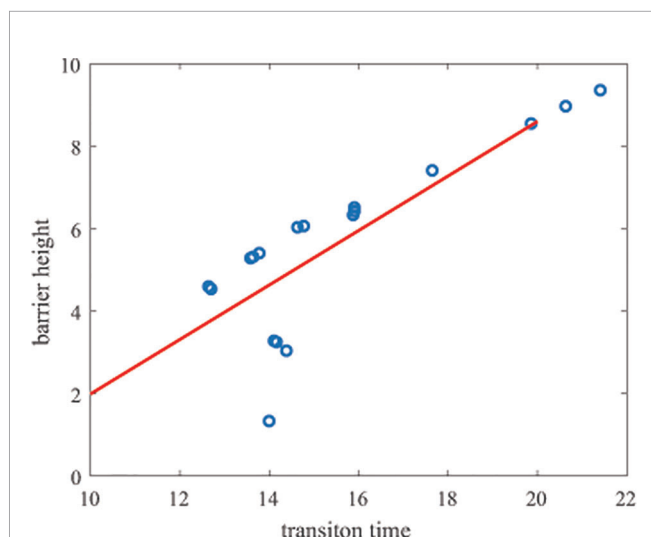


FIGURE 5 | Correlation of the transition time and barrier heights. The y-axis represents the barrier height and the x-axis represents the natural logarithm of the transition time.

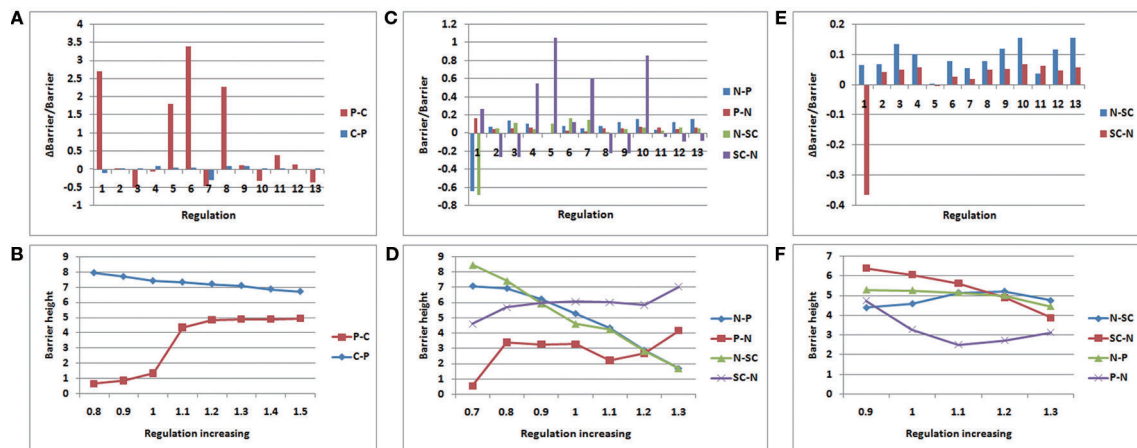


FIGURE 6 | Variation rate of barrier height with regulation strength. $P - C$ ($C - P$) denotes the barrier height from premalignant to cancer (cancer to premalignant) state. $N - P$ ($P - N$) denotes the barrier height from normal to premalignant (premalignant to normal) state. $N - SC$ ($SC - N$) denotes the barrier height from normal to SC (SC to normal) state. (A, B):miR200-ZEB; (C, D):OCT4→OCT4; (E, F):P53→P53.

and easier for it to move from cancer to premalignant, which is beneficial to cancer recovery. As shown in **Figure 6B**, we also varied the regulation strength from 0.8 to 1.5 times. When the regulation strength decreased, the barrier height from the premalignant to the cancer state also decreased, and the barrier height from the cancer to the premalignant state increased. This illustrates that the regulation is associated with the variation of the barrier height between the premalignant and cancer states. This variation of regulation miR200-ZEB indicates how metastasis could be controlled.

As shown in **Figure 6C**, we increased the regulation strength to 1.3 times. OCT4 is a signature gene of SCs. If the expression level of OCT4 is high, the stemness of the cell is obvious. When the regulation strength increased, the expression level of OCT4 increased, the barrier height from the normal to the SC state decreased, and the barrier height from the SC to the normal state increased significantly. This means that a cell in the normal state could more easily move to the SC state, but it was more difficult for cells in the SC state to move to the normal state. At the same time, the barrier height from the normal to the premalignant state decreased, and that from the premalignant to the normal state increased. This indicates that a cell in the normal state is more likely to become cancerous, and a cell in the premalignant state is less likely to move back. These results are consistent with those of experiments involving iPS. Many studies have shown that the cellular reprogramming of iPS often leads to cells with cancerous characteristics, which eventually reach the cancer state (Kondo and Raff., 2000; Pavlova and Thompson, 2016). As shown in **Figure 6D**, when the regulation strength decreased to 0.7 times, the barrier height from the normal to the premalignant state increased, and that from the premalignant to the normal state decreased. This indicates that the cells in the normal state are more stable, and those in the premalignant state are more likely to transform to the normal state. The barrier height from the normal to the SC state increased, and that from the SC to the normal state decreased. Thus, cells in the

normal state are less likely to switch to the SC state, and those in the SC state are more apt to transform to the normal state. The regulation OCT4→OCT4 reflects the connection between SCs and metastasis. This may guide the search for cancer treatments involving SCs.

As shown in **Figure 6E**, we also increased the regulation strength to 1.3 times. The variation of regulation 1 (P53→P53) increased the barrier height from the normal to the SC state, and decreased that from the SC to the normal state. This indicates that when the expression level of P53 increases, the cells in the normal state are less likely to transform to the SC state, whereas the cells in the SC state were more likely to transform to the normal state. Experiments have shown that P53 is a major driving force for the differentiation of embryonic SCs (ESCs). Spontaneous differentiation of hESCs reduced significantly when P53 expression decreased (Qin et al., 2007). P53 also can provide an effective barrier for the generation of stemness cells from terminally differentiated cells (Solozobova and Blattner, 2011). The variation of regulation P53→P53 illustrates the importance of P53 not only for cancer but also for SC processes.

P53 is a tumor suppressor, as shown in **Figure 6F**. When the regulation strength decreased to 0.9 times and P53 abundance was reduced, the barrier height from the normal to the premalignant state barely changed, but the barrier height from the premalignant to the normal state increased significantly. In this situation, the cells in the premalignant state are less likely to move back to the normal state. When the regulation strength was increased to 1.1 times, the barrier height from the normal to the premalignant state showed no significant change, whereas that from the premalignant to the normal state was reduced. That means the cells in the premalignant state would more easily transition to the normal state. When the regulation strength was increased to 1.2 and 1.3 times, the barrier height between the normal and premalignant

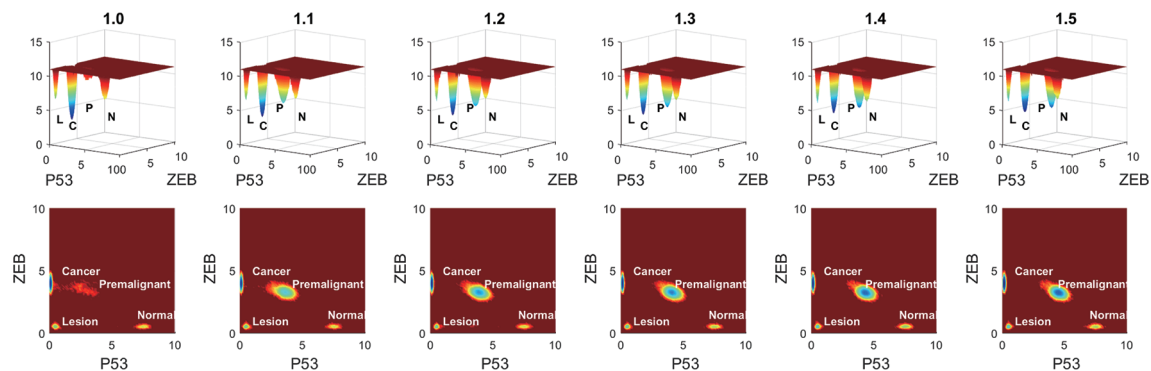


FIGURE 7 | Variation of barrier height when the regulation strength changes from 1.0 to 1.5 times. *L* represents the lesion state, *P* represents the premalignant state, *C* represents the cancer state, and *N* represents the normal state.

states varied only slightly. When the concentration of P53 reaches a very high level, its tumor suppressor characteristics become less obvious and other characteristics are present, such as inducing apoptosis (Haupt et al., 2003).

To see the variation more clearly, we depicted the landscape topography of $\text{miR200} \rightarrow \text{ZEB}$ from regulation strength 1.0 to 1.5 times. As shown in **Figure 7**, the depth of the basin of the premalignant state increased significantly when the regulation strength increased, and the depth of the basin of the cancer state decreased.

CONCLUSIONS

Cancer is a complex and fatal disease. Its features of metastasis, drug resistance, and recurrence, which are related to CSCs, cause cancer to be a major health threat. Recent studies have shown that EMT has a vital role in inducing the early stage of metastasis, as well as being a way for non-SCs to transform into SCs. In this study, we developed a dynamic model which includes specific genes and microRNAs for CSC, EMT, and cancer, with the aim of uncovering the connections among cancer, metastasis, and differentiation and development in CSC, EMT, and cancer. We quantified the underlying landscape to explore differentiation and development and metastasis, thereby elucidating the origin of cancer. The kinetic paths and barrier heights between each state were quantified. The barrier heights determine the stability of the state and relate to the switching frequency of the cells from one state to another. Multiple cancer formation pathways were observed. The flux of each path (from normal to cancer, and the reverse) was calculated using the statistics of the path transitions. This was used to determine which path is more important in cancer formation and treatment, and could also help to quantify the degree of difficulty of curing a particular cancer. Furthermore, we used global sensitivity analysis to find key regulations which are vital for cancer formation. Three regulations, $\text{miR200} \rightarrow \text{ZEB}$, $\text{OCT4} \rightarrow \text{OCT4}$, and $\text{P53} \rightarrow \text{P53}$ were more sensitive than other regulations. These regulations may provide a reference for the treatment of cancer. This work

studied the functional dynamics and physical mechanisms of differentiation and development in cancer and metastasis in a quantitative way, and may serve as a guide for clinical therapy of cancer.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found in TCGA-LIHC and TCGA-COAD projects.

ETHICS STATEMENT

The authors declare that the study doesn't include animal and human experiments that violate ethics.

AUTHOR CONTRIBUTIONS

Article written by CY and JW. Model constructed by CC, CY, QL, and JW. Data analysis by CY, QL, and JW. Tools analysis provided by JW.

FUNDING

CY and QL were supported by NSFC grant no. 21721003, MOST China grant no. 2016YFA0203200. JW thanks NSF-PHY-76066 and NSF-CHE1808474 for supports.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.01388/full#supplementary-material>

REFERENCES

- Alhajj, M., Wicha, M. S., Benito Hernandez, A., Morrison, S. J., and Clarke, M. F. (2003). Prospective identification of tumorigenic breast cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* 100 (7), 3983–3988. doi: 10.1073/pnas.0530291100
- Blancafort, P., Jin, J., and Frye, S. (2013). Writing and rewriting the epigenetic code of cancer cells: from engineered proteins to small molecules. *Mol. Pharmacol.* 83 (3), 563. doi: 10.1124/mol.112.080697
- Bonnet, D., and Dick, J. E. (1997). Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat. Med.* 3 (7), 730–737. doi: 10.1038/nm0797-730
- Chaffer, C. L., Brueckmann, I., Scheel, C., Kaestli, A. J., Wiggins, P. A., Rodrigues, L. O., et al. (2011). Normal and neoplastic nonstem cells can spontaneously convert to a stem-like state. *Proc. Natl. Acad. Sci.* 108 (19), 7950–7955. doi: 10.1073/pnas.1102451108
- Cowin, P., Rowlands, T. M., and Hatsell, S. J. (2005). Cadherins and catenins in breast cancer. *Curr. Opin. Cell Biol.* 17 (5), 499–508. doi: 10.1016/j.ccb.2005.08.014
- Dragu, D. L., Necula, L. G., Bleotu, C., Diaconu, C. C., and Chivu-Economescu, M. (2015). Therapies targeting cancer stem cells: Current trends and future challenges. *World J. Stem Cells* 7 (9), 1185. doi: 10.4252/wjsc.v7.i9.1185
- Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* 81 (25), 2340–2361. doi: 10.1021/j100540a008
- Gillespie, D. T. (2000). The chemical langevin equation. *J. Chem. Phys.* 113 (1), 297–306. doi: 10.1063/1.481811
- Grosse-Wilde, A., d'Hérou, A. F., McIntosh, E., Ertaylan, G., Skupin, A., Kuestner, R. E., et al. (2015). Stemness of the hybrid epithelial/mesenchymal state in breast cancer and its association with poor survival. *PLoS One* 10 (5), e0126522. doi: 10.1371/journal.pone.0126522
- Hanahan, D., and Coussens, L. M. (2012). Accessories to the crime: functions of cells recruited to the tumor microenvironment. *Cancer Cell* 21 (3), 309–322. doi: 10.1016/j.ccr.2012.02.022
- Hanahan, D., and Weinberg, R. A. (2000). The hallmarks of cancer. *Cell* 100 (1), 57–70. doi: 10.1016/S0092-8674(00)81683-9
- Haupt, S., Berger, M., Goldberg, Z., and Haupt, Y. (2003). Apoptosis - the p53 network. *J. Cell Sci.* 116 (20), 4077–4085. doi: 10.1242/jcs.00739
- Hewitt, H. B. (1952). Transplantation of mouse sarcoma with small numbers of single cells. *Nature* 170 (4328), 622–623. doi: 10.1038/170622a0
- Jeremy, A. H., Holland, D. B., Roberts, S. G., Thomson, K. F., and Cunliffe, W. J. (2003). Inflammatory events are involved in acne lesion initiation. *J. Invest. Dermatol.* 121 (1), 20. doi: 10.1046/j.1523-1747.2003.12321.x
- Kauffman, S. (1971). Differentiation of malignant to benign cells. *J. Theor. Biol.* 31 (3), 429–451. doi: 10.1016/0022-5193(71)90020-8
- Kondo, T., and Raff, M. (2000). Oligodendrocyte precursor cells reprogrammed to become multipotential CNS stem cells. *Science* 289 (5485), 1754–1757. doi: 10.1126/science.289.5485.1754
- Kreso, A., O'Brien, C. A., Van Galen, P., Gan, O. I., Notta, F., Brown, A. M. K., et al. (2013). Variable clonal repopulation dynamics influence chemotherapy response in colorectal cancer. *Science* 339 (6619), 543–548. doi: 10.1126/science.1227670
- Kumar, S. M., Liu, S., Lu, H., Zhang, H., Zhang, P. J., Gimotty, P. A., et al. (2012). Acquired cancer stem cell phenotypes through oct4-mediated dedifferentiation. *Oncogene* 31 (47), 4898–4911. doi: 10.1038/onc.2011.656
- Kumar Jolly, M., Huang, B., Lu, M., Mani, S. A., Levine, H., and Ben-Jacob, E. (2014). Towards elucidating the connection between epithelial–mesenchymal transitions and stemness. *J. R. Soc. Interface* 11 (101), 20140962. doi: 10.1098/rsif.2014.0962
- Kumar Jolly, M. (2015). Implications of the hybrid epithelial/mesenchymal phenotype in metastasis. *Front. In Oncol.* 5, 5–155. doi: 10.3389/fonc.2015.00155
- Lamouille, S., Xu, J., and Derynck, R. (2014). Molecular mechanisms of epithelial–mesenchymal transition. *Nat. Rev. Mol. Cell Biol.* 15 (3), 178–196. doi: 10.1038/nrm3758
- Li, C., and Wang, J. (2015). Quantifying the landscape for development and cancer from a core cancer stem cell circuit. *Cancer Res.* 75 (13), 2607–2618. doi: 10.1158/0008-5472.CAN-15-0079
- Lin, Y., Yang, Y., Li, W., Chen, Q., Li, J., Pan, X., et al. (2012). Reciprocal regulation of akt and oct4 promotes the self-renewal and survival of embryonal carcinoma cells. *Mol. Cell* 48 (4), 627–640. doi: 10.1016/j.molcel.2012.08.030
- Liu, S., Ginestier, C., Charafe-Jauffret, E., Foco, H., Kleer, C. G., Merajver, S. D., et al. (2008). Brca1 regulates human mammary stem/progenitor cell fate. *Proc. Natl. Acad. Sci. U. States America* 105 (5), 1680–1685. doi: 10.1073/pnas.0711613105
- Liu, H.-X., Li, X.-L., and Dong, C.-F. (2015). Epigenetic and metabolic regulation of breast cancer stem cells. *J. Zhejiang Univ. Sci. B.* 16 (1), 10–17. doi: 10.1631/jzus.B1400172
- Lobo, N. A., Shimono, Y., Qian, D., and Clarke, M. F. (2007). The biology of cancer stem cells. *Annu. Rev. Cell Dev. Biol.* 23, 675–699. doi: 10.1146/annurev.cellbio.22.010305.104154
- Lynch, M., Latta, L., Hicks, J., and Giorgianni, M. (1998). Mutation, selection, and the maintenance of life-history variation in a natural population. *Evolution* 52 (3), 727–733. doi: 10.2307/2411267
- Mani, S. A., Guo, W., Liao, M.-J., Ng Eaton, E., Ayyanan, A., Zhou, A. Y., et al. (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* 133 (4), 704–715. doi: 10.1016/j.cell.2008.03.027
- Mani, S. A., Guo, W., Liao, M. J., Eaton, E. N., Ayyanan, A., Zhou, A. Y., et al. (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* 133 (4), 704. doi: 10.1016/j.cell.2008.03.027
- Martincoren, I., and Campbell, P. J. (2015). Somatic mutation in cancer and normal cells. *Science* 349 (6255), 1483. doi: 10.1126/science.aab4082
- Medema, J. P. (2013). Cancer stem cells: the challenges ahead. *Nat. Cell Biol.* 15 (4), 338–344. doi: 10.1038/ncb2717
- Morel, A.-P., Lièvre, M., Thomas, C., Hinkal, G., Ansieau, S., and Puisieux, A. (2008). Generation of breast cancer stem cells through epithelial-mesenchymal transition. *PLoS One* 3 (8), e2888. doi: 10.1371/journal.pone.0002888
- Morel, A.-P., Hinkal, G. W., Thomas, C., Fauvet, F., Courtois-Cox, S., Wierinckx, A., et al. (2012). EMT inducers catalyze malignant transformation of mammary epithelial cells and drive tumorigenesis towards claudin-low tumors in transgenic mice. *PLoS Genet.* 8 (5), e1002723. doi: 10.1371/journal.pgen.1002723
- Muller, P. A. J., and Vousden, K. H. (2013). p53 mutations in cancer. *Nat. Cell Biol.* 15 (1), 2. doi: 10.1038/ncb2641
- Pastushenko, I., Brisebarre, A., Sifrim, A., Fioramonti, M., Revenco, T., Boumahdi, S., et al. (2018). Identification of the tumour transition states occurring during EMT. *Nature* 556 (7702), 463–468. doi: 10.1038/s41586-018-0040-3
- Pavlova, N. N., and Thompson, C. B. (2016). The emerging hallmarks of cancer metabolism. *Cell Metab.* 23 (2), 24–47. doi: 10.1016/j.cmet.2015.12.006
- Ponti, D., Costa, A., Zaffaroni, N., Pratesi, G., Petrangolini, G., Coradini, D., et al. (2005). Isolation and *in vitro* propagation of tumorigenic breast cancer cells with stem/progenitor cell properties. *Cancer Res.* 65 (13), 5506–5511. doi: 10.1158/0008-5472.CAN-05-0626
- Qin, H., Yu, T., Qing, T., Liu, Y., Zhao, Y., Cai, J., et al. (2007). Regulation of apoptosis and differentiation by p53 in human embryonic stem cells. *J. Biol. Chem.* 282 (8), 5842. doi: 10.1074/jbc.M610464200
- Rizzino, A. (2013). Concise review: the sox2-oct4 connection: critical players in a much larger interdependent network integrated at multiple levels. *Stem Cells* 31 (6), 1033–1039. doi: 10.1002/stem.1352
- Rodriguez-Paredes, M., and Esteller, M. (2011). Cancer epigenetics reaches mainstream oncology. *Nat. Med.* 17 (3), 330. doi: 10.1038/nm.2305
- Scheel, C., and Weinberg, R. A. (2012). “Cancer stem cells and epithelial–mesenchymal transition: concepts and molecular links,” in *Seminars in cancer biology*, vol. 22. (Elsevier), 396–403. doi: 10.1016/j.semcancer.2012.04.001
- Sell, S. (2004). Stem cell origin of cancer and differentiation therapy. *Crit. Rev. In Oncol./hematol.* 51 (1), 1–28. doi: 10.1016/j.critrevonc.2004.04.007
- Singh, S. K., Kawkins, C., Clarke, I. D., Squire, J. A., Bayani, J., Hide, T., et al. (2004). Identification of human brain tumour initiating cells. *Nature* 432 (7015), 396–401. doi: 10.1038/nature03128
- Solozobova, V., and Blattner, C. (2011). p53 in stem cells. *World J. Biol. Chem.* 2 (9), 202–214. doi: 10.4331/wjbc.v2.i9.202
- Spano, D., Heck, C., De Antonellis, P., Christofori, G., and Zollo, M. (2012). Molecular networks that regulate cancer metastasis. *Semin. In Cancer Biol.* 22 (3), 234. doi: 10.1016/j.semcancer.2012.03.006
- Strauss, R., Li, Z.-Y., Liu, Y., Beyer, I., Persson, J., Sova, P., et al. (2011). Analysis of epithelial and mesenchymal markers in ovarian cancer reveals phenotypic heterogeneity and plasticity. *PLoS One* 6 (1), e16186. doi: 10.1371/journal.pone.0016186

- Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126 (4), 663–676. doi: 10.1016/j.cell.2006.07.024
- Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., et al. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 131 (5), 861–872. doi: 10.1016/j.cell.2007.11.019
- Tang, D. G. (2012). Understanding cancer stem cell heterogeneity and plasticity. *Cell Res.* 22 (3), 457–472. doi: 10.1038/cr.2012.13
- Wang, J., Xu, L., and Wang, E. K. (2008). Potential landscape and flux framework of nonequilibrium networks: robustness, dissipation, and coherence of biochemical oscillations. *Proc. Natl. Acad. Sci. U. States America* 105 (34), 12271–12276. doi: 10.1073/pnas.0800579105
- Wang, J., Zhang, K., Xu, L., and Wang, E. (2011). Quantifying the waddington landscape and biological paths for development and differentiation. *Proc. Natl. Acad. Sci. U. States America* 108 (20), 8257–8262. doi: 10.1073/pnas.1017017108
- Wang, Y., Gan, L. F., Wang, E. K., and Wang, J. (2013). Exploring the dynamic functional landscape of adenylate kinase modulated by substrates. *J. Chem. Theory Comput.* 9 (1), 84–95. doi: 10.1021/ct300720s
- Wang, S.-S., Jiang, J., Liang, X.-H., and Tang, Y.-L. (2015). Links between cancer stem cells and epithelial–mesenchymal transition. *OncoTargets Ther.* 8, 2973. doi: 10.2147/OTT.S91863
- Wellner, U., Schubert, J., Burk, U. C., Schmalhofer, O., Zhu, F., Sonntag, A., et al. (2009). The emt-activator zeb1 promotes tumorigenicity by repressing stemness-inhibiting microRNAs. *Nat. Cell Biol.* 11 (12), 1487–1495. doi: 10.1038/ncb1998
- Xu, L., Zhang, K., and Wang, J. (2014). Exploring the mechanisms of differentiation, dedifferentiation, reprogramming and transdifferentiation. *PloS One* 9 (8), e105216. doi: 10.1371/journal.pone.0105216
- Yang, P., Du, C. W., Kwan, M., Liang, S. X., and Zhang, G. J. (2013). The impact of p53 in predicting clinical outcome of breast cancer patients with visceral metastasis. *Sci. Rep.* 3(1). doi: 10.1038/srep02246
- Ye, X., and Weinberg, R. A. (2015). Epithelial–mesenchymal plasticity: a central regulator of cancer progression. *Trends In Cell Biol.* 25 (11), 675–686. doi: 10.1016/j.tcb.2015.07.012
- Yu, C., and Wang, J. (2016). A physical mechanism and global quantification of breast cancer. *PloS One* 11 (7), e0157422. doi: 10.1371/journal.pone.0157422
- Zhu, L., Finkelstein, D., Gao, C., Shi, L., Wang, Y., Lopez-Terrada, D., et al. (2016). Multi-organ mapping of cancer risk. *Cell* 166 (5), 1132. doi: 10.1016/j.cell.2016.07.045

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Yu, Liu, Chen and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



A Computational Model of the Endothelial to Mesenchymal Transition

Nathan Weinstein^{1,3*}, Luis Mendoza² and Elena R. Álvarez-Buylla^{1,3}

¹ Instituto de Ecología, Universidad Nacional Autónoma de México, Mexico City, Mexico, ² Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, Mexico City, Mexico, ³ Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, Mexico City, Mexico

OPEN ACCESS

Edited by:

Chunhe Li,
Fudan University, China

Reviewed by:

Xiaopeng Zhang,
Nanjing University, China
Debashis Barik,
University of Hyderabad,
India

*Correspondence:

Nathan Weinstein
nathan.weinstein@c3.unam.mx

Specialty section:

This article was submitted to
Systems Biology,
a section of the journal
Frontiers in Genetics

Received: 28 November 2019

Accepted: 14 January 2020

Published: 12 March 2020

Citation:

Weinstein N, Mendoza L
and Álvarez-Buylla ER (2020)
A Computational Model
of the Endothelial to
Mesenchymal Transition.
Front. Genet. 11:40.
doi: 10.3389/fgene.2020.00040

Endothelial cells (ECs) form the lining of lymph and blood vessels. Changes in tissue requirements or wounds may cause ECs to behave as tip or stalk cells. Alternatively, they may differentiate into mesenchymal cells (MCs). These processes are known as EC activation and endothelial-to-mesenchymal transition (EndMT), respectively. EndMT, Tip, and Stalk EC behaviors all require SNAI1, SNAI2, and Matrix metalloproteinase (MMP) function. However, only EndMT inhibits the expression of VE-cadherin, PECAM1, and VEGFR2, and also leads to EC detachment. Physiologically, EndMT is involved in heart valve development, while a defective EndMT regulation is involved in the physiopathology of cardiovascular malformations, congenital heart disease, systemic and organ fibrosis, pulmonary arterial hypertension, and atherosclerosis. Therefore, the control of EndMT has many promising potential applications in regenerative medicine. Despite the fact that many molecular components involved in EC activation and EndMT have been characterized, the system-level molecular mechanisms involved in this process have not been elucidated. Toward this end, hereby we present Boolean network model of the molecular involved in the regulation of EC activation and EndMT. The simulated dynamic behavior of our model reaches fixed and cyclic patterns of activation that correspond to the expected EC and MC cell types and behaviors, recovering most of the specific effects of simple gain and loss-of-function mutations as well as the conditions associated with the progression of several diseases. Therefore, our model constitutes a theoretical framework that can be used to generate hypotheses and guide experimental inquiry to comprehend the regulatory mechanisms behind EndMT. Our main findings include that both the extracellular microenvironment and the pattern of molecular activity within the cell regulate EndMT. EndMT requires a lack of VEGFA and sufficient oxygen in the extracellular microenvironment as well as no FLI1 and GATA2 activity within the cell. Additionally Tip cells cannot undergo EndMT directly. Furthermore, the specific conditions that are sufficient to trigger EndMT depend on the specific pattern of molecular activation within the cell.

Keywords: endothelial-mesenchymal transition, systems biology, angiogenesis, Boolean network, endothelial cell plasticity, heart development, fibrosis

INTRODUCTION

The circulatory system allows the body to efficiently transport oxygen and nutrients to all the constituent cells of animals through an intricate network of blood vessels. Capillaries are the smallest blood vessels, communicating arterioles and venules; they are composed of a single layer of endothelial cells (ECs), and are partially covered by mural cells called pericytes (PCs). ECs and PCs are in close proximity to most cells in multicellular animals and are some of the most important cells involved in wound healing and tissue regeneration. Thus, alterations that affect these cells result in several pathological processes (Eming et al., 2014; Birbrair et al., 2015).

While ECs and PCs are fully differentiated cell types, they have the notable capacity to trans-differentiate into each other (Nakagomi et al., 2015; Chen et al., 2016; Jackson et al., 2017), and are also capable of differentiating into hematopoietic stem cells, mesenchymal stem cells, and several other cell types (van Meeteren and Ten Dijke, 2012; Birbrair et al., 2017; Dejana et al., 2017). Notably, ECs differentiate into PCs in a process called endothelial to mesenchymal transition (EndMT), which is very similar to the epithelial-to-mesenchymal transition (EMT) (Lamouille et al., 2014; Méndez-López et al., 2017). Like EMT, EndMT is a reversible process, and the opposite mechanism is denominated mesenchymal-to-endothelial transition (MEnT) (Sánchez-Duffhues et al., 2018). EndMT is triggered either by changes in the concentration of WNT, NOTCH, FGF, or TGF ligands in the extracellular microenvironment, reduced oxygen availability or shear stress. These changes lead to the activation of

the transcription factors SNAIL1, SNAIL2, TWIST1, ZEB1, and SPI1(ZEB2), resulting in the repression of the expression of endothelial markers, specifically VEGFR2, PECAM1, VE-Cadherin, TIE1, TIE2, and vWF accompanied by the augmented expression of mesenchymal markers including α -SMA, N-cadherin, and Collagen I/II. During EndMT, ECs lose cell-to-cell adhesion and luminobasal polarity, gaining migratory and invasive potential (**Figure 1B**) (Gong et al., 2017; Jackson et al., 2017).

EndMT is a key process; physiologically, it is present during the development of the heart. The formation and maturation of the endocardial cushion leads to the formation of the septa and valves. First, the endocardial cells located at the atrioventricular canal (AVC)—including endocardial ECs—separate from the myocardial cells that cover them. Then, the endocardial and myocardial cells secrete extracellular matrix (ECM) components that accumulate to form and expand the cardiac matrix jelly that separates them. After that, AVC myocardial cells secrete bone morphogenetic proteins (BMPs), causing AVC ECs to undergo EndMT. Lastly, the mesenchymal cells resulting from the EndMT differentiate into the cells that compose the cardiac septa and heart valves (Kaneko et al., 2008). From the pathological perspective, EndMT alterations are involved in many cardiovascular disorders including atherosclerosis, congenital heart disease, myocardial fibrosis, myocardial infarctions, and pulmonary arterial hypertension.

Stable vascular networks are lined by a layer of quiescent ECs called *Phalanx cells* that are tightly bound to each other and to the basement membrane, as well as being at least partially

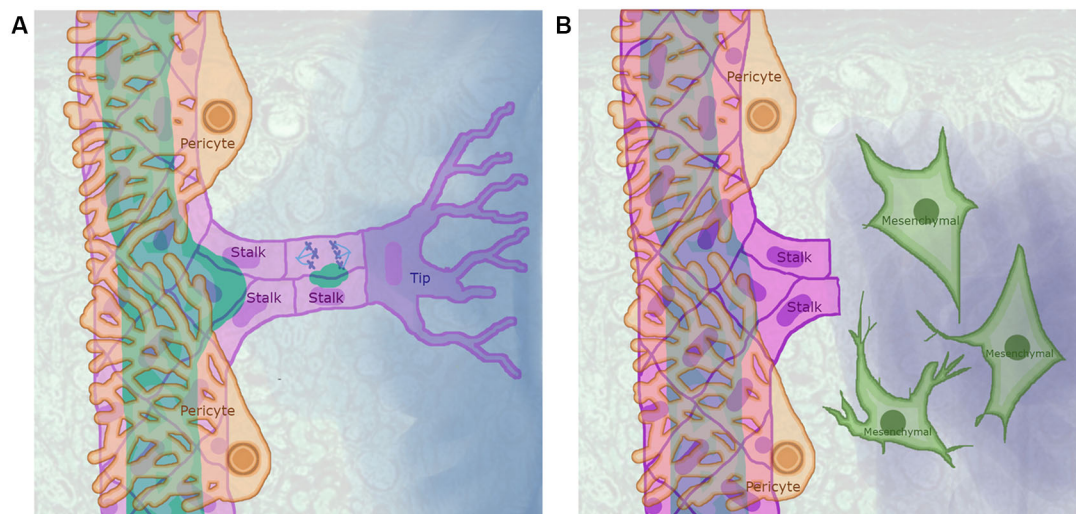


FIGURE 1 | Sprouting angiogenesis as partial endothelial-to-mesenchymal transition (EndMT): **(A)** In a precapillary arteriole with an angiogenic sprout, the pericytes (light orange cells that surround the arteriole) detach from a region of the arteriole exposed to a concentration of angiogenic signal that exceeded a certain threshold leading to the activation of an endothelial cell (EC) that became a Tip cell (purple) that extends filopodia to sense the angiogenic signal gradient. The ECs that surrounded the Tip cell were induced to become Stalk cells (pink) that proliferate, elongate, secrete vacuoles, and trail the tip cell as it migrates following the angiogenic signal gradient. **(B)** The EndMT process is similar to sprouting angiogenesis, as ECs have to be activated and secrete Matrix metalloproteinases that degrade the basement membrane to increase their motility and proliferate. However, in contrast to Tip and Stalk cells, ECs that undergo EndMT completely detach from other ECs and stop expressing EC markers.

covered by PCs. These Phalanx ECs do not proliferate, however, they do exhibit lumen to basal membrane polarity, and express EC markers (Korn and Augustin, 2015; Betz et al., 2016). Either hypoxia or the lack of sufficient nutrients may cause cells that surround a microvascular network to secrete angiogenic factors, triggering sprouting angiogenesis. In this process, certain ECs are induced to become migratory, invasive *Tip cells* (TCs), while adjacent PCs detach from the capillary segment. Each TC induces abutting ECs to become *Stalk cells* (SCs). Then, both the TC and SCs detach from the basement membrane and the TC migrates toward the source of the angiogenic signal trailing SCs that elongate and proliferate (**Figure 1A**). The new sprout continues to grow until the TC reaches either another blood vessel or the TC leading another sprout. Then, the lumen of the new segment is formed from the fusion of vacuoles (Jianxin et al., 2015; Kim et al., 2017) and flow-mediated apical membrane invagination (Gebala et al., 2016). Lastly, the new capillary segment is stabilized and surrounded by PCs.

During sprouting angiogenesis TCs and SCs detach from the basement membrane, migrate, and lose their luminobasal polarity. Furthermore, TCs are invasive and secrete MMPs that degrade the ECM while SCs proliferate. However, during angiogenesis, ECs continue to express their characteristic molecular markers, and the adherens and tight junctions that bind ECs remain intact, thus suggesting that TC and SC behavior involves partial EndMT (Welch-Reardon et al., 2015). Both TCs and SCs express *SNAI1* and *SNAI2*, and silencing either of these genes inhibits angiogenic sprout formation, TC migration, and affects lumen formation. *SNAI2* directly regulates the expression of *MT1-MMP*, the protein encoded by this gene cleaves and activates *MMP2* and *MMP9*. These are two proteases involved in ECM degradation during sprouting angiogenesis (Welch-Reardon et al., 2014).

As summarized above, a large set of molecules has been described to be involved in angiogenesis and EndMT. Nonetheless, the integrated dynamical mechanisms that underlie full or partial EndMT are still not well understood (Welch-Reardon et al., 2015). We propose that theoretical and system-biology approaches, such as those proposed by (Álvarez-Buylla Rocas et al., 2018; Yang and Albert, 2019), can help us elucidate the molecular mechanisms involved in EndMT regulation. Cell types and behaviors are defined by a combination of morphological, behavioral, genetic, and epigenetic traits (Pavillon and Smith, 2019). In molecular regulatory network models, cell types and behaviors are represented by fixed and cyclic patterns of molecular activation called attractors. Both ECs and MCs are very diverse groups of cells with different developmental origins and exhibit many patterns of gene expression and molecular activation (Chi et al., 2003; Ho et al., 2018). Therefore, we expect the underlying molecular mechanism involved in EC and MC identity and behavior regulation to be multistable.

Due to the enormous biological and medical importance of angiogenesis and EMT, both processes have been widely explored through the simulation of models at the molecular and cellular levels (Peirce, 2008; Qutub et al., 2009; Lu et al., 2013;

Steinway et al., 2014; Heck et al., 2015; Li et al., 2016; Méndez-López et al., 2017; Weinstein et al., 2017; Suzuki et al., 2018). In contrast, to the best of our knowledge, simulation or formal analyses of the molecular mechanism that control EndMT are lacking. To this aim, we inferred the regulatory network of EndMT by undertaking an exhaustive search of published data, and formalizing it as a dynamical network system to study its behavior under wild type and mutant backgrounds. The model is able to recover the expression patterns that characterize the main cell types during normal and pathological angiogenesis. Importantly, the model can be used as a tool to generate hypotheses regarding molecular and cellular effects of a large group of perturbations, such as mutations and pharmacological manipulations. Our main findings are that the specific conditions sufficient to trigger EndMT and MEnT depend on the pattern of molecular activation within the cell. EndMT requires a lack of *FLI1* and *GATA2* activity within the cell and also requires the absence of *VEGFA* and the presence of sufficient oxygen in the extracellular microenvironment. Additionally *Tip cells* cannot undergo EndMT directly.

METHODS

We assembled the molecular regulatory network of EndMT using information available in the literature, focusing on the incorporation of key molecules and their regulatory interactions. Then, the inferred network was transformed into a discrete dynamical system in the form of a Boolean network (BN). We analyzed the dynamical behavior of the model to find and classify the stationary and cyclic patterns of molecular activation. Thereafter, we studied the conditions that led to changes in the behavior or identity of the cells. Also, we evaluated the robustness of the model to single gain- and loss-of-function mutations, as well as its robustness to changes in the components of the logical update rule. Besides the study of these properties of the system, the model was compared with the expected effect of the extracellular microenvironments, gain- and loss-of-function mutations, and mechanical forces associated to several diseases in humans.

Regarding the validation of our model, the standard way of doing it is by comparing the specific effects of gain and loss-of-function mutations as reported in the references with their simulated effect. Furthermore, we also simulated the conditions that have been associated with several diseases related to EndMT and compared the simulated dynamic behavior of our model with the clinical observations of the pathologies.

Formalization of the Molecular Regulatory Network as a Discrete Dynamical System

By assuming that every molecule in a regulatory network has a concentration threshold that must be exceeded in order to have an effect, it makes sense to use the formalism of a BN, where each molecule is represented by a node that can be either active or inactive, represented by 1 or 0, respectively. Let $\mathbb{B} = \{0, 1\}$ and

$\mathbb{Z}_{\leq n}^+ = \{1, 2, \dots, n\}$ a set of labels. A *state* of a BN is an n -tuple $x = (x_1, x_2, \dots, x_n)$ such that $x \in \mathbb{B}^n$, and each component x_i of a state x , represents the *activation state of variable i* . To relate a synchronous BN with a molecular network, we interpret that variable i denotes a molecule included in the network. A BN is then a set of functions that contains for each variable i in the network an update rule $f_i: \mathbb{B}^k \rightarrow \mathbb{B}$ where k is the number of nodes that regulate variable i , and the n -tuple is an ordered list of the states of the nodes that regulate node i . The dependency of the state of activation of each node on the discrete time parameter t is denoted as $x_i(t)$, and obeys the update rule given by f_i , such that for all $t \in \mathbb{Z}$:

$$x_i(t+1) = f(x(t))_i = f_i(r_i(t)).$$

When no race conditions or important cyclic behaviors are expected from the simulated dynamic behavior of the model, it is convenient to update all nodes simultaneously obtaining a deterministic discrete dynamic system. A *synchronous* BN with n components is a function $f: \mathbb{B}^n \rightarrow \mathbb{B}^n$, where the i -th component *off* is a function f_i such that $f_i(x) = f(x)_i$. That is for all $t \in \mathbb{Z}$

$$x(t+1) = f(x(t)) = (f_1(r_1(t)), f_2(r_2(t)), \dots, f_n(r_n(t))).$$

BNs encode regulatory interactions among the molecules that compose them. Node j *activates* node i if there exists a pair of network states x, y that differ only in the state of activation of variable j , where $x_j = 0$ and $y_j = 1$, such that $f_i(y) - f_i(x) > 0$. Conversely, node j *inhibits* node i if there exists a pair of network states x, y that differ only in the state of activation of variable j . Specifically, $x_j = 0$ and $y_j = 1$, such that $f_i(y) - f_i(x) < 0$. Node i both activates and inhibits node j if there exists a pair of network states x, y that differ only in the state of activation of variable j . Specifically, $x_j = 0$ and $y_j = 1$, such that $f_i(y) - f_i(x) > 0$, and there exists another pair of network states p, q that differ only in the state of activation of variable j . Specifically, $p_j = 0$ and $q_j = 1$, such that $f_i(q) - f_i(p) < 0$. An *interaction* denoted as the pair (i, j) , $i, j \in \mathbb{N}_{\leq n}$ is *functional* if variable j activates or inhibits variable i , or both.

BN models as defined above are deterministic and finite systems, thus simulating the dynamic behavior from any given initial state of the network leads to an attractor. A *fixed point attractor* is a state $s \in \mathbb{B}^n$ such that $f(s) = s$. If we define f^{ol} as the l -th iterate of the function f such that $f^{ol} = f(f^{ol-1})$. Then, an *attractor* is a set of states $A \subseteq \mathbb{B}^n$, such that $f^{ol}(x) = x$ for any state $x \in A$. Furthermore, l is the size of the attractor and for any $j \in \mathbb{N}_{\leq n}^+$, $f^{oj}(x) \in A$.

It is a standard practice to interpret fixed point attractors as the stationary patterns of molecular activation observed in a given regulatory network, and attractors of larger order as cyclic patterns of molecular activation (Álvarez-Buylla Rocas et al., 2018; Yang and Albert, 2019). In the present study, we were able to assign to all attractors a biological interpretation in term of either a cell type or a cellular behavior.

We defined each component f_i of the update rule f as follows: In the simplest case, the node $N1$ is only regulated by $R1$, then $f_{N1} = x_{N1}(t+1) = x_{R1}(t)$. However, when the number of regulators is greater than one, we find groups of active and

inactive regulators that are sufficient to activate a given node. We then represent such group as a logical expression where if all the regulators of the group are active or inactive, as needed, then the node is active. For instance, if $N2$ is regulated by the activators $A1, A2$, and $A3$ that form a complex, and the formation of such complex is inhibited by $I1$, then $f_{N2} = x_{N2}(t+1) = x_{A1}(t) \wedge x_{A2}(t) \wedge x_{A3}(t) \wedge \neg x_{I1}(t)$. If there are several groups of molecules that are sufficient to activate the node, then those groups form an OR expression. For example, if $N3$ represents a gene that can be activated either by $A4$ if $I2$ is absent, or independently by $A5$, then $f_{N3} = x_{N3}(t+1) = (x_{A4}(t) \wedge \neg x_{I2}(t)) \vee x_{A5}(t)$. Additionally, some nodes are regulated at transcriptional, posttranslational and protein levels and can be formalized using an AND expression. For example, if the transcription of node $N4$ is regulated by $TF1$ or $TF2$, its splicing is regulated by $SF1$, and also $MPK1$ activates the protein by phosphorylation and $PF1$ causes its proteolysis. Then $f_{N4} = x_{N4}(t+1) = (x_{TF1}(t) \vee x_{TF2}(t)) \wedge x_{SF1}(t) \wedge (x_{MPK1}(t) \wedge \neg x_{PF1}(t))$.

The molecular basis of our regulatory network is sufficient to specify the direction and sign for most of the interactions, as well as to specify most of the components of the logical update rule of the model. Nevertheless, in some cases the published information was not sufficient to unequivocally determine the sign of an interaction or an update rule. In these cases, we adjusted the system by assuming that the dynamic behavior of our model must reach fixed or cyclic patterns of molecular activation that correspond to the expected cell marker expression for Phalanx, Stalk, and Tip EC behaviors, as well as mesenchymal cells.

For the interested reader, the BoolNet, and GINsim versions of the discrete model are available for download at <https://github.com/NathanWeinstein/EndMT>.

Molecular Pattern Identification

We labeled the attractors according to the molecular activation patterns associated to specific cell types or cell states. Notably, these labels are not mutually excluding; a given network state may fit more than one label. In the following paragraphs, we describe the possible labels that might be assigned to network states. Furthermore, some of the attractors are cyclic in nature, therefore, we applied a label to a cyclic attractor only if it was possible to apply the label to each one of the states that composes the cyclic attractor.

It is known that all ECs express VE-cadherin, PECAM1, TIE2, and VEGFR2. These molecules, in turn, are activated by the combined presence of the transcription factors GATA2, and FLI1. Hence, whenever a network state has these two nodes in an active state, we say that such network represents an EC. Some mesenchymal cells express GATA2 and FLI1, but they also express fibroblast specific protein-1 (FSP-1), α smooth muscle actin (α SMA), Smooth muscle-22 α (SM22 α), encoded by transgelin (TAGLN), and fibronectin (Kamata et al., 2014). The precise mechanism by which mesenchymal markers are expressed during EndMT has not been fully elucidated. However, SNAIL1, SNAIL2, TWIST1, ZEB1, and ZEB2, which

are also expressed by certain ECs, have been used experimentally as mesenchymal markers (Magenta et al., 2011; Welch-Reardon et al., 2014; Mahmoud et al., 2016). Because of these, we identify as a mesenchymal cell all those network states where ZEB1, ZEB2, TWIST1, and either SNAI1 or SNAI2 are active. Phalanx ECs are the quiescent and tightly-bound ECs that form a layer that functions as a barrier. We identify as Phalanx ECs those states where there is an absence of NPR1, CTNNB, SNAI1, SNAI2, while GATA2 and FLI1 are present. The absence of the first set of markers is important because SNAI1 and SNAI2 inhibit the transcription of VE-cadherin (Lopez et al., 2009; Cheng et al., 2013), and other important components of endothelial adherence and tight junctions (Laakkonen et al., 2017). Also, CTNNB activates the transcription of SNAI2 and TWIST1, while CTNNB and LEF1 induce EC proliferation by activating the transcription of *Cyclin D1*. Finally, NRP1 is a marker for Tip EC behavior (Aspalter et al., 2015), (Phng et al., 2009). Stalk cells are activated ECs that trail ECs. These cells express FLI1, GATA2, and JAG1, yet they do not express NPR1 (del Toro et al., 2010; Blancas et al., 2012). Finally, Tip cells are activated ECs that grow filipodia. Here, we use the presence of FLI1, GATA2, NRP1, and ETS1 to identify Tip cells. NRP1 is a recognized Tip cell marker (del Toro et al., 2010; Blancas et al., 2012) and Tip cells must express DLL4, which requires ETS1 activity (Wythe et al., 2013).

The *basin of attraction* of an attractor is the group of states that converges to that attractor. These states include the attractor itself. In models where an attractor corresponds to just one cell type (see for example Weinstein et al. (2015)), it is customary to characterize the basins of attraction. In the present model, however, a given attractor may correspond to more than one label, and *vice versa*, one label can be assigned to more than one attractor. Henceforth, it is necessary to define a *trap space* of any given cell type or behavior *c*. This trap space is the union of the basins of attraction of the fixed and cyclic behaviors that can be classified as *c*. We estimated the size of each trap space by first generating 10^7 random network states. For each state, we simulated the behavior of our model until reaching an attractor. We then classified the attractor and calculated the fraction of the sampling space covered by each trap space.

Robustness of the Model

Evolution has made biological organisms resilient to several perturbations such as mutations and fluctuations in the concentration or level of molecular activation, while at the same time remaining sensitive to changes in the concentration of key molecular signals used to regulate its development. We refer to this property as selective robustness. Specifically, the systems affected by EndMT resist most changes in the extracellular microenvironment, single gain and loss-of-function mutations, as well as parameter variation. Substantial alterations occur only when a critical molecule or interaction is affected, or when several molecules are affected simultaneously. Therefore, the molecular mechanisms involved in EndMT regulation exhibit selective robustness. For clarity, we need to

specify the trait we test for robustness, as well as the nature of the perturbations we use to assess such robustness. Moreover, it is also necessary to define a method to quantify robustness (Félix and Barkoulas, 2015). Hence, we measured the robustness of the network in the following ways:

1. The robustness of the cell types, as measured by the percentage of gain- or loss-of-function mutations the system is able to resist without the loss of a specific stationary or cyclic pattern of molecular activation.
2. The robustness of the cell types to random changes in the update rule. This was done by generating a population of 100,000 instances of the models, but each instance affected by a single bit-flip in a random component of the update rule. The mean number of attractors for all the networks in the population were calculated. We say in this case that a cell type is robust if the mean of the population is closer to the nonperturbed model, and also if the variance is small.
3. The sensitivity of each component of the update rule to molecular activation noise. For each update rule component, namely each $f_i \in f$, we generated 500,000 random initial states, and for each one of those initial states *s*, a variant *s'* is generated with a one bit flip. Then, we applied the update rule to both *s* and *s'* and calculated the sensitivity of f_i as the fraction of initial states where $f(s)_i \neq f(s')_i$. Additionally, we calculated the sensitivity of each update rule component when flipping from 2 to 15 bits of *s* to obtain *s'* in order to observe how the sensitivity of each update rule is affected by different levels of molecular activation noise. For each component and each number of flipped bits we used 20,000 random initial states.
4. The robustness of each cell type in response to perturbations in the molecules that represent the extracellular microenvironment and the main transcription factors involved in maintaining EC identity. Such nodes are *DLL4*, *FGF2*, *FLI1*, *GATA2*, *HIF1 α* , *PDGF_AB*, *TGF β* , *VEGFA*, *WNT5b*, and *WNT7a*. For each of the patterns classified as a cell type or cellular state, we tested all possible combinations of perturbations in the aforementioned nodes and let the system converge. Here, the robustness is the fraction of the perturbations that were absorbed by the system, such that the network reached the original cell type or behavior before the perturbation.

Libraries for the Dynamical Analysis

We used *GINsim* (Naldi et al., 2009) to find and analyze the feedback circuits of our model. Then, we used the R package *BoolNet* (Müssel et al., 2010) to find the attractors using a heuristic method that formulates the attractor search as a boolean satisfiability (SAT) problem that is solved using the PicoSAT solver (Biere, 2008; Dubrova and Teslenko, 2011). We also used *BoolNet* to simulate mutations and perturbations. The analysis of the perturbations that cause changes in cell type and behavior required preparing a function for parallel processing, and for this we used the R package *doParallel* (Weston and Calaway, 2019). We also used the R package *ggplot2* (Wickham, 2011) to create

graphics. Lastly, we used the R package *xtable* (Swinton, 2014) to export matrices and data frames from R into LaTeX. The scripts and the data generated by the scripts are freely available at: <https://github.com/NathanWeinstein/EndMT>.

RESULTS

Molecular Basis of the Regulatory Network

EndMT is defined by the loss of EC adhesion, the conversion of endothelial apical-basal polarity to front end-back end polarity, and a marked decrease in EC markers accompanied by increased MC marker expression. During EndMT, the signaling pathways of TGF, WNT, NOTCH, VEGF, FGF, TNF, and PDGF modulate the activity of the transcription factors FLI1 and GATA2 that are essential for EC identity, as well as the activity of SNAI1, SNAI2, TWIST1, ZEB1, ZEB2, and LEF1 that are necessary for mesenchymal cell differentiation. Importantly, these transcription factors form a complex regulatory network that we have uncovered here. The following sections include the mechanism by which these and other relevant molecules regulate each other.

EC Adhesion

In stable and mature blood vessels, ECs are interconnected, forming a barrier that separates blood or lymph from the surrounding tissue. Additionally, ECs are covered by a basement membrane, and at least partially covered by mural cells. Many of the proteins that compose the transmembrane complexes that bind ECs together are expressed only in ECs, and are thus used as EC markers. Both EndMT and EC activation reduce EC adhesion and increase the EC barrier permeability; however, only EndMT causes ECs to completely detach from the endothelial monolayer.

EndMT represses the expression of many of the proteins that compose intraendothelial junctions resulting in loss of EC adhesion and identity. ECs are connected by junctional proteins, which assemble to form adherens junctions (AJs) that link the cytoskeletons of adjacent ECs; by tight junctions (TJs) that function as a selectively permeable barrier between ECs; and by gap junctions (GJs) that function as selective ion channels (Radeva and Waschke, 2018). Furthermore, focal adhesions (FAs) anchor ECs to the basement membrane, but they can also be located between ECs where they function as important regulators of the microvascular function (Wu, 2005).

Vascular endothelial cadherin (VE-cadherin) is one of the main components of endothelial AJs (Giannotta et al., 2013). α , β and γ -catenins, α -actinin and vinculin anchor VE-cadherin to actin. VE-cadherin can also recruit the desmosomal proteins desmoplakin and vimentin. Intermediate filaments composed of vimentin may be linked to endothelial AJs via plakoglobin/desmoplakin or p0071 forming junctional structures called *complexus adherens* (Wallez and Huber, 2008). Moreover, VE-PTP inhibits VEGFA-mediated phosphorylation of VE-cadherin, thus stabilizing endothelial AJs (Bazzoni and Dejana, 2004). Furthermore, VE-cadherin, PECAM1, and VEGFR2 form

a junctional mechanosensory complex (Conway et al., 2013; Kutys and Chen, 2016). Nectins are one of the main components of AJs, are linked to actin microfilaments through Afadin, and also form interendothelial bonds.

Tight junctions also include proteins that form bonds at the interendothelial cleft, forming a physical barrier that prevents solutes and water from freely crossing the EC sheet. The number of TJs at an interendothelial cleft is proportional to the shear stress applied to the endothelial sheet by blood flow. The proteins that compose TJs include Claudins, Occludin, JAMS, ESAM, and Nectins. Those proteins are linked to numerous intracellular partners, including AF-6/afadin, cingulin, the antigen 7H6, PAR-3, ZO-1, ZO-2, and ZO-3, forming a molecular complex (Wallez and Huber, 2008). The barrier forming Claudins CLDN3, CLDN5, and CLDN11 are expressed by ECs. Occludin (OCLN) increases TJ barrier function and is one of the main molecules involved in the regulation of endothelial layer permeability. The expression of occludin is upregulated by Angiopoietin 1 (ANGPT1), and further stabilized by angiotensin-2 (AT2) binding to type 1 angiotensin receptor (ATR). VEGFA downregulates OCLN by inducing OCLN proteolysis through activation of the urokinase (uPA)/uPAR system and also by PKC-mediated phosphorylation. OCLN is also regulated by monocyte chemoattractant protein-1 (MCP-1/CCL-2), histamine, oxidized phospholipids, lysophosphatidic acid, and shear-stress (González-Mariscal et al., 2008; Wallez and Huber, 2008; Radeva and Waschke, 2018). The junctional adhesion proteins F11R (JAM-A), JAM2 (VE-JAM or JAM-B), JAM3 (JAM-C), and the related protein ESAM (EC adhesion protein) from the immunoglobulin superfamily are important components of endothelial TJs that regulate paraendothelial permeability, leukocyte trafficking and TJ dynamics (Wallez and Huber, 2008; Rahimi, 2017).

FAs are composed of α and β integrin heterodimers that bind several ECM components, as well as TJ components and several intracellular proteins. Those adhesive integrin interactions contribute to the maintenance of endothelial barrier function, and the loss of integrin-matrix adhesion results in leaky microvessels (Wu, 2005; Izawa et al., 2018). ECs express multiple integrins that assemble into several different heterodimers. The extracellular domains of many integrins have a high binding affinity for the Arg-Gly-Asp (RGD) sequence and are able to interact with several matrix proteins. However, certain heterodimers exhibit a higher affinity for a specific protein including $\alpha 6 \beta 1$ and $\alpha 6 \beta 4$ that favor laminin, $\alpha 1 \beta 1$ and $\alpha 1 \beta 2$ that tend to bind collagen, $\alpha v \beta 3$ and $\alpha v \beta 5$ that exhibit affinity to vitronectin, as well as $\alpha 3 \beta 1$ and $\alpha 5 \beta 1$ that favor fibronectin (Wu, 2005). Focal adhesion kinase (FAK) is another important FA component. The N-terminal domain of FAK contains a region called FERM homology that exhibits a high binding affinity for growth factor receptors and integrins. The C-terminal domain contains a noncatalytic region, also referred to as FRNK (FAK-related nonkinase), that carries a FAT sequence that directs FAK to adhesion complexes and provides docking sites for other cytoplasmic proteins. FAK activation, triggered by phosphorylation regulates endothelial barrier function either

increasing or decreasing permeability depending on the site of phosphorylation and the context. When VEGFA binds VEGFR2, it causes a conformation change that exposes an integrin $\alpha v\beta 3$ binding site. Integrin $\alpha v\beta 3$ then binds VEGFR2, recruits FAK and promotes the activation of several signaling pathways that lead to increased microvascular permeability. VEGFA also causes phosphorylation-coupled FAK activation and relocation from the cytoplasm to focal contacts.

EC Polarization

Certain cellular processes including asymmetric cell division, cell migration, and barrier formation require the asymmetric organization of components within a cell. In stable blood vessels, ECs have an apical (luminal) membrane domain, an interendothelial (lateral) membrane domain, and a basal membrane domain. This organization results in a luminobasal or apicobasal EC polarity. During angiogenesis, the cytoskeleton of tip cells and stalk cells undergoes several changes that result in transient front-to-rear EC polarity which is necessary for collective directed migration (Ebnet et al., 2018). Many of the molecules involved in EC polarization are implicated in lumen formation and also regulate endothelial permeability linking these processes (Lizama and Zovein, 2013).

Both angiogenesis and vasculogenesis involve cord hollowing, a process that results in lumen formation. Prior to lumen formation, the ECs that compose the segment must acquire an apicobasal polarity (Lizama and Zovein, 2013; Ebnet et al., 2018). The molecular signaling pathways involved in EC polarization and lumen formation are largely unknown and are subject to current research (Norden et al., 2016; Szymborska and Gerhardt, 2018). During early embryonic vasculogenesis, $\beta 1$ integrin (ITGB1), RAS interacting protein 1 (RASIP1), and partitioning defective 3 (PAR3) interact to establish EC apicobasal polarity before epithelial lumen formation (Herbert and Stainier, 2011). VE-cadherin acts as a positional cue to attract and organize the proteins involved in EC polarization. Accordingly, loss of VE-cadherin function prevents apicobasal EC polarization and EC agglomerations from developing a vascular lumen. VE-cadherin directly interacts with many proteins involved in EC polarization such as PAR3, PARD6A (PAR6), MPP5 (PALS1), and KRIT1 (CCM1) (Giannotta et al., 2013; Lizama and Zovein, 2013; Brinkmann et al., 2016). VE-cadherin recruits the sialomucins CD34 and PODXL (Podocalyxin) to EC-cell contact sites. Sialomucins contain negative charges that cause repulsive forces and initiate adjacent EC membrane separation. Later, VE-cadherin is involved in Moesin, F-actin and nonmuscle myosin II recruitment to induce lumen expansion and stabilization. Other proteins involved in lumen expansion and stabilization include Protein kinase C (PKC) that links CD34 to the actin cytoskeleton through Moesin phosphorylation, and ROCK, that is also necessary for nonmuscle myosin II recruitment (Lizama and Zovein, 2013).

During the initial stages of angiogenesis, tip cells form filopodia and lamellipodia and orient them following the gradient of a vascular growth factor, typically VEGFA. The Ras homologue gene (Rho) and Ras-related protein (Rap) families of small G proteins are important mediators of VEGFA signaling in

ECs (Shimizu et al., 2018). The Rho GTPases RhoA, Rac1, and Cdc42 interact with integrins at FAs where actin accumulates to initiate the formation of filopodia and lamellipodia (Lizama and Zovein, 2013). Tip cells then migrate toward the source of the morphogen trailing stalk cells. During sprout elongation, the elastic properties of the cytoskeleton of the ECs that conform the sprout have to be tightly regulated (Szymborska and Gerhardt, 2018). *In vitro*, EC sprout elongation requires a reduction of EC contractility mediated by the downregulation of Rho kinase (ROCK) and myosin light chain 2 (MLC2). Another important molecular mechanism that increases EC contractility involves RAP1, which induces the formation of a RAF1-VE-cadherin complex that recruits ROCK (Szymborska and Gerhardt, 2018). KRIT1 is an effector of RAP1, which upon activation interacts with β -catenin and afadin. Additionally, KRIT1 stabilizes endothelial junctions by recruiting RAP1 that stabilizes and concentrates VE-cadherin. KRIT1 also recruits CCM2 to the junction where it inhibits RHOA to further stabilize the junction. Another important function of KRIT1 is to prevent the activation of the canonical WNT signaling pathway by sequestering β -catenin (Wilson and Ye, 2014).

Key Transcription Factors for Endothelial and Mesenchymal Identities

The specification and maintenance of EC identity requires the function of ETV2, FLI1, ERG, ETS1, and other members of the E26 transforming specific (ETS) family of transcription factors; all of them share a core GGAA/T DNA-binding motif (Craig and Sumanas, 2016). ETV2 function is required for endothelial specification during early embryonic development in both mice and zebrafish (Abedin et al., 2014), and it is also necessary for vascular regeneration after an injury (Park et al., 2016). ETV2 directly binds to the promoters of *Cdh5* (VE-cadherin), *Tie2*, *Kdr* (VEGFR2), *Scf*, *Gata2*, *Meis1*, *Dll4*, *Notch1*, *Nrp1/2*, *Flt4*, *RhoJ*, *Mapk*, and *Fli1* (Oh et al., 2015). Later, during embryonic development, ETV2 is no longer expressed and FLI1 maintains endothelial identity by binding to the promoters of *Cdh5*, *Tie2*, *Cd31* (PECAM1), *Erg* and *Fli1*, activating their expression as well as its own (Abedin et al., 2014). Notably, diminishing the expression of FLI1 and ERG triggers the EndMT (Nagai et al., 2018).

ETS1 exhibits functional redundancy with ETS2, is expressed during angiogenesis, and is involved in the regulation of EC survival, migration, and proliferation. ETS1 induces the expression of several matrix metalloproteinases (MMPs), integrins, and *NRP1* (Teruyama et al., 2001; Craig and Sumanas, 2016). Then, GATA2 belongs to the C2H2 zinc-finger class of transcription factors and is also involved in the regulation of EC identity. Importantly, the loss of GATA2 in ECs triggers EndMT. In ECs, GATA2 activates the transcription of *Emcn* (Endomucin, interferes with FJ assembly), *Cdh5*, *Pecam1*, *Vegfr2*, *Nrp1*, *vWF*, and *Gata2* itself (Kanki et al., 2011; Coma et al., 2013). It is also important to mention that GATA2 and FLI1 activate the transcription of each other (Pimanda et al., 2007b).

Five transcription factors have been associated with EndMT. Four of them, SNAI1 (SNAIL), SNAI2 (SLUG), ZEB1, and ZEB2 (SIP1), contain four to six E2â€ box DNA binding zinc fingers,

and a SNAG domain involved in transcriptional repression. The other transcription factor is the basic helix-loop-helix (bHLH) TWIST1 (Gong et al., 2017; Jackson et al., 2017; Sánchez-Duffhues et al., 2018). SNAI1, SNAI2, and TWIST directly repress the transcription of VE-cadherin (Lopez et al., 2009; Cheng et al., 2013). Other components of endothelial AJs and TJs are also downregulated during EndMT. However, in most cases, the molecular mechanism has not been fully elucidated. For instance, CLDN5 is downregulated by SNAI1 (Kokudo et al., 2008) and SNAI2 (Laakkonen et al., 2017), yet it is well recognized that VE-cadherin is a key component of endothelial junctions that integrates molecular and mechanical signals. VE cadherin is involved in EC identity, quiescence, migration and polarization. Therefore, loss of VE-cadherin function explains several of the cellular processes involved during EndMT.

Both SNAI1 and SNAI2 proteins bind to E2 boxes in promoters that regulate *Snai1* and *Snai2* expression (Chen and Gridley, 2013b). SNAI1 and SNAI2 directly suppress each other's transcription during chondrogenesis (Chen and Gridley, 2013b; Chen and Gridley, 2013a). SNAI1 (Peiro et al., 2006) and TWIST1 (Yu et al., 2013; Forghanifard et al., 2017) directly repress the transcription of *Snai1*. However, E47 binds TWIST1 forming a dimer that binds to the *Snai1* promoter and activates its expression (Yu et al., 2013). In certain tumor cells, SNAI1 upregulates ZEB1 and ZEB2 expression (Guaita et al., 2002; Takkunen et al., 2006). In contrast, in melanoma cell lines, SNAI1 does not activate the transcription of ZEB1 (Wels et al., 2011), thus, we have not included this interaction in our model. SNAI2 (Kumar et al., 2015) and TWIST1 (Casas et al., 2011) directly activate the transcription of *Snai2*. SNAI2 also directly induces the transcription of ZEB1 (Wels et al., 2011). The molecular mechanism that causes loss of FLI1, ERG, and GATA2 expression to induce EndMT remains obscure. Nonetheless, it is known that GATA2 siRNA leads to increased SNAI1 and SNAI2 expression, and GATA2 binds to the proximal promoter of SNAI2 (Kanki et al., 2011). Additional interactions have been reported for other cell types. In hematopoietic stem cells, for example, TWIST1 binds to the promoter of *Gata2* and induces its transcription (Kulkeaw et al., 2017), while in nasopharyngeal carcinoma cells, GATA2 induces EMT by binding to the promoter of *Twist1* and activating its expression (Wang et al., 2017b). Furthermore, ETS1 and ZEB2 activate each other's transcription (Katoh and Katoh, 2009; Yalim-Camci et al., 2019).

The Molecular Signaling Pathways Involved in EndMT Regulation

In a previous model of endothelial behavior during angiogenesis (Weinstein et al., 2017), the TGF, NOTCH, WNT, VEGF, FGF, and HIF signaling pathways were described in detail. Thus, we will focus now on their roles during the EndMT.

The TGF signaling pathway is of central importance for the regulation of EC plasticity and EndMT (Dejana et al., 2017). When a TGF or a BMP ligand binds to a TGF receptor complex, it causes the activation of several signaling pathways that mediate TGF-induced EndMT, among them SMAD, MEK, p38 MAPK,

and PI3K signaling (Medici et al., 2011). Some of the key components of TGF signaling involved in the regulation of EndMT include the ligand TGF β 2 (Chen et al., 2012), type I receptors ALK1 and ALK5 (TGFBR1), the type II receptor TGF β R2, as well as SMAD2, SMAD3, and SMAD4 (Medici et al., 2011; Chen et al., 2012). SMAD2 and SMAD3 activate the transcription of SNAI2, while SMAD4—which is a co-SMAD that allows other SMADs to activate the transcription of target genes—is required for TGF-induced SNAI1 expression (Cooley et al., 2014). The expression of ZEB2 is induced by TGF signaling and its promoter contains SMAD binding sites (Katoh and Katoh, 2009). Furthermore, ZEB1 and ZEB2 bind SMADs forming transcriptional regulation complexes (Grabitz and Duncan, 2012). Also, TGF β 2 also induces inhibitory VEGFA splicing (Weinstein et al., 2017).

FGF signaling modulates EC and PC function and behavior. When an FGF ligand, like FGF2, binds to an FGF receptor such as FGFR2, it causes FRS2-mediated ERK and PI3K signaling pathway activation (Yang et al., 2015). FGF signaling inhibits EndMT by downregulating TGF signaling; FGF2 activates the transcription of miRNAs from the *let7* family, especially *let7b* and *let7c*, which prevents the expression of TGFBR1 (Chen et al., 2012). FGF2 also increases the expression of *mir-20a*, another miRNA that prevents the expression of TGFBR1, TGFBR2 and SARA (Smad anchor for receptor activation) (Correia et al., 2016). In addition to RNA silencing, FGF2 activates the Ras-MAPK signaling pathway that regulates TGFBR1-induced SMAD2 phosphorylation in lymphatic ECs (Ichise et al., 2014). Another important function of FGF signaling in ECs is to activate the transcription of VEGFR2. FGF activates ERK signaling, which then activates several transcription factors from the ETS family including ETS1 and ETV2 that activate *Vegfr2* transcription (Murakami et al., 2011; Yang et al., 2015).

Insufficient oxygen availability (hypoxia) in the cells that compose the tissue surrounding a network of capillaries triggers angiogenesis. HIF1, composed of subunits HIF1 α and HIF1 β , is a key mediator of the cellular response to hypoxia. Hypoxia prevents the PHD-mediated proteasomal degradation of HIF1 α , a molecule that directly activates the transcription of *Vegfa* (Forsythe et al., 1996; Kumar et al., 2014). When ECs themselves are exposed to hypoxia, it may cause senescence, increased apoptosis and necrosis rates due to augmented oxidative stress and irreparable DNA damage, or angiogenesis and proliferation, depending on the duration and severity of the reduction in oxygen availability (Baldea et al., 2018). Under certain circumstances, hypoxia causes EndMT. In this case, HIF1 α directly binds to the promoter region of *Snai1* and induces its transcription (Xu et al., 2015). Hypoxia also induces the expression of SNAI2 and TWIST1 in ECs (Xu et al., 2015). Additionally, during EMT (Yang and Wu, 2008) and also during EndMT associated with pulmonary arterial hypertension, HIF α directly induces the expression of TWIST1 (Zhang et al., 2018). Furthermore, the proximal promoter region of ZEB2 contains a HIF1 α -binding site (Katoh and Katoh, 2009). Finally, HIF1 is an important inducer of EC differentiation since HIF1 α binds to the *Etv2*

promoter and activates its transcription. ETV2, in turn, activates the transcription of *Fli1* (Oh et al., 2015).

VEGF signaling is involved in EC activation during vascular remodeling. Typically, during angiogenesis VEGFA binds to a VEGFR2 homodimer and activates PLC γ , and TSAd-AKT signaling (Simons et al., 2016). VEGFA signaling strengthens the EC identity by activating the expression of GATA2 (Coma et al., 2013). Further, VEGFA-VEGFR2 signaling phosphorylates and activates STAT3 (Chen et al., 2008), which then activates the transcription of SNAI1 in HeLa cells (Saitoh et al., 2016). Additionally, the VEGF co-receptor NRP1 is a key molecule that promotes tip cell behavior and inhibits stalk cell behavior by limiting SMAD2/3 phosphorylation (Aspalter et al., 2015). However, NRP1 also acts as a co-receptor for TGF β 1 and is necessary for TGF β -mediated EndMT (Matkar et al., 2016).

Notch signaling is required for EndMT by the cardiac cushions during early cardiac valve development. The signaling of this pathway is initiated when a ligand (DLL4) binds to a Notch receptor (NOTCH1). Then, the receptor is cleaved into an intracellular domain, a transmembrane domain, and an extracellular domain. NOTCH1 activation leads to increased SNAI2 expression (Niessen et al., 2008), as well as increased SNAI1 stability and nuclear retention. The intracellular domain of NOTCH1 forms a complex with β -catenin and TCF4 that activates the transcription of AKT2. This molecule then inhibits glycogen synthase kinase 3 (GSK3 β)-mediated proteolysis and translocation of SNAI1 from the nucleus to the cytoplasm (Frias et al., 2016). Furthermore, Notch signaling induces the transcription of both subunits of the nitric oxide (NO) receptor soluble guanylyl cyclase (sGC), namely GUCY1A3 and GUCY1B3. Also, this signaling induces Activin A, consequently promoting both NO production and the transcription of its receptor, which are necessary for EndMT to occur in the developing AVC (Chang et al., 2011). In response to an increase in shear stress, NOTCH1 activation leads to the formation of GTPase signaling complexes at AJs composed of the NOTCH1 transmembrane domain, VE-cadherin, the guanine nucleotide exchange factor Trio, and the tyrosine phosphatase LAR that activates RAC1 to stabilize adherens junctions (Fischer and Braga, 2018). NOTCH also induces the transcription of *Vegfr1*. VEGFR1 inhibits VEGFA-VEGFR2 signaling by reducing the amount of VEGFA available to bind VEGFR2 (Funahashi et al., 2010). The Notch-regulated ankyrin repeat protein (NRARP) links NOTCH and WNT signaling. Dll4-NOTCH1 signaling induces *Nrarp* expression in ECs. NRARP negatively regulates Notch signaling by destabilizing the Notch intracellular domain and positively regulates Wnt signaling by increasing the stability of the LEF1 protein (Ishitani et al., 2005; Phng et al., 2009). Finally, another important function of NOTCH signaling in stalk cells is to negatively regulate the expression of NRP1 (Aspalter et al., 2015).

Canonical Wnt signaling is initiated by a WNT ligand, which is usually WNT7A or WNT3, and leads to the stabilization of CTNNB (β -catenin). Like Notch signaling, canonical Wnt signaling also causes GSK3 β phosphorylation, allowing the

accumulation and nuclear localization of SNAI1 and SNAI2 (Wu et al., 2012; Menezes, 2014). Further, the complex formed by CTNNB and TCF activates the transcription of many of the genes regulated by canonical Wnt signaling (Menezes, 2014), including SNAI2 (Conacci-Sorrell et al., 2003), TWIST1 (Howe et al., 2003), and ZEB1 (Sánchez-Tilló et al., 2011; Sanchez-Tillo et al., 2014). CTNNB and TCF also induce the transcription and activation of LEF1. During EMT, LEF1 activates the transcription of *Snai2* and *Zeb1* even in the absence of both β and γ -catenins (Kobayashi and Ozawa, 2018). Other WNT ligands including WNT5a, WNT5b, and WNT11 activate the noncanonical planar cell polarity (PCP) and CA²⁺ WNT signaling pathways that also activate the Activator protein 1 (AP-1) transcription factor (Nishita et al., 2010). AP-1 binding sites exist in the promoter regions of *Snai1* and *Snai2*, and the inhibition of AP-1 results in reduced SNAI1 expression in mesenchymal cells (Nguyen et al., 2013). Moreover, WNT5b induces EndMT and SNAI1 expression in lymphatic ECs through the activation of WNT/ β -catenin and PCP pathways. WNT5b also induces inhibitory VEGFA splicing through noncanonical WNT signaling (Weinstein et al., 2017).

The PDGF signaling pathway is involved in the regulation of pericyte recruitment during microvascular maturation, and EndMT-mediated pericyte differentiation from ECs (Gaengel et al., 2009; Chen et al., 2016). The signaling is initiated by a PDGF ligand that can be the PDGF-AB heterodimer, or one of four homodimers, namely PDGF-AA, -BB, -CC, and -DD. The tyrosine kinase receptors PDGFR α and PDGFR β dimerize after ligand binding. PDGF-AA forms PDGFR $\alpha\alpha$. PDGF-BB can form either PDGFR $\alpha\alpha$, PDGFR $\beta\beta$ or PDGFR $\alpha\beta$ dimers. PDGF-CC forms PDGFR $\alpha\alpha$, or PDGFR $\alpha\beta$ receptors. PDGF-DD signals specifically via the PDGFR $\beta\beta$ receptor, but is able to form the PDGFR $\alpha\beta$ heterodimer. PDGF-AB forms PDGFR $\alpha\alpha$, or PDGFR $\alpha\beta$ receptors. After activation and dimerization, PDGFRs can interact with signaling proteins that contain an SH2 domain, including FER, PI3K, PLC, SHP2, and SRC, leading to the activation of several signaling pathways, such as MAPK, PI3K-AKT-NF- κ B and PLC γ (Romashkova and Makarov, 1999; Papadopoulos and Lennartsson, 2018). ECs weakly express PDGFR α and PDGFR β . However, when brain ECs are exposed to PDGF-AB, it causes the activation of the transcription factor NF- κ B, which binds to the promoter of *Snai1* and activates its transcription, leading to EndMT (Liu et al., 2018). In spite of the fact that in human breast cancer cells, NF- κ B binds to the promoter regions of *Snai2*, *Twist1*, and *ZEB2* and activates their transcription (Pires et al., 2017), PDGF-AB does not increase the expression of *Snai2*, *Twist1*, and *ZEB2* in brain ECs (Liu et al., 2018). Additionally, NF- κ B directly activates the transcription of *Lef1* in chondrocytes (Yun et al., 2007).

During acute inflammation, TNF α and IL-1 β cause NF- κ B-mediated EndMT by inducing the degradation of the inhibitory κ B (I κ B α) protein, which sequesters NF- κ B in the cytosol (Sánchez-Duffhues et al., 2018). Furthermore, inflammation may suffice to determine if an EC is activated or if it undergoes full EndMT. TNF α induces VE-cadherin internalization and degradation. Additionally, TNF α inhibits VE-cadherin

expression by activating the transcription of hsa-miR-6086 (Cai et al., 2018). Shear stress and cyclic strain also modulate EndMT. Laminar shear stress activates the mechanosensitive transcription factors KLF2 and KLF4 that inhibit EndMT by downregulating AP1 and NFκB. Also, KLF2 induces the expression of *Smad6*, *Smad7* and *VegfA*, which inhibit SMAD2 activity. Further, KLF4 activates the transcription of VE-cadherin, prevents the expression of genes regulated by SMADs by binding to the TGFβ control element, and also impedes the transcription of mesenchymal genes by binding SMAD3. Cycle strain induces EndMT by Rho mediated VE-cadherin translocation from the membrane into the cytoplasm, causing the concentration of β-catenin in the nucleus to increase (Krenning et al., 2016). For simplicity, we only take into account one activating signal for AP-1, β-catenin, and NF-κB.

BN Model Assembly

As summarized in the previous section, a very large number of molecular components and pathways have been described to be involved in the regulation of EndMT and angiogenesis. In order to integrate their roles and understand their concerted action, we propose here a BN approach. For simplicity, we selected a subset of molecules. Specifically, we incorporated into our model only those molecules that are essential either due to their biological function, or due to their effect in the simulated dynamic behavior of our model. As a result, the regulatory network of EndMT includes 29 molecules connected by 77 regulatory interactions, as shown in **Figure 2**. The model encompasses molecules necessary for EC identity, the ligands that activate the VEGF, HIF, NOTCH, FGF, TGF, WNT, and PDGF signaling pathways, as

well as the main transcription factors involved in EndMT. We did not include many EC and MC markers because they act as network sinks, and their activity can be inferred from that of the included transcription factors. Most of the 77 interactions represent direct transcriptional or posttranscriptional regulations. However, the interactions that connect ligands directly to transcription factors represent entire linear signaling pathways.

After the reconstruction of the regulatory network, we translated the information to construct a Boolean model, as described in Section 2.1. We used the molecular information outlined in Section 3.1 to obtain the logical rules. Additionally, the references we used to define each component of the update rule are specified in **Table 1**. However, in order for our model to reach fixed or cyclic patterns of molecular activation that correspond to the expected cell marker expression for Phalanx, Stalk and Tip EC behaviors as well as mesenchymal cells, we had to fix the rules in three instances (**Table 2**). As a result, the components of the update rule of the network are shown as follows in equations 1–29.

Our Model Formalized as a Set of Boolean Equations

$$AP1(t+1) = WNT5b(t) \vee SMAD2(t) \quad (1)$$

$$CTNNB(t+1) = WNT5b(t) \vee WNT7a(t) \quad (2)$$

$$DLL4(t+1) = DLL4(t) \quad (3)$$

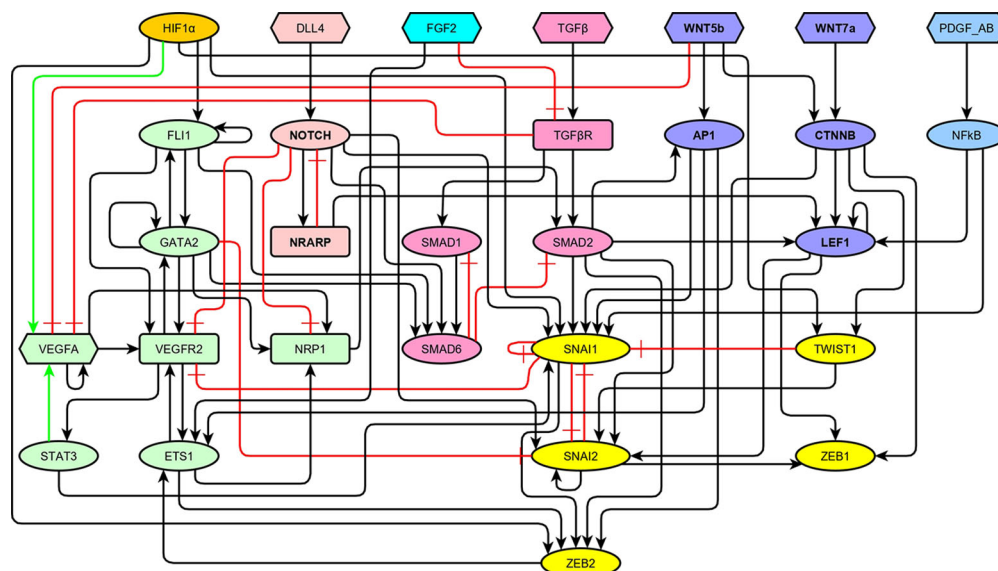


FIGURE 2 | The topology of our model of the network of molecules involved in the regulation of endothelial-to-mesenchymal transition (EndMT) represented as a signed directed graph: Black arrows represent positive regulations, green arrows represent positive autocrine regulations, and red blunt arrows represent inhibitions. The VEGF signaling pathway and the main transcription factors involved in endothelial cell (EC) identity are shown in green, HIF1α is shown in orange, the NOTCH signaling pathway is shown in light red, FGF2 is shown in turquoise, the TGFβ signaling pathway is shown in pale magenta-pink, the WNT signaling pathway is shown in lavender, the PDGF signaling pathway is shown in light cyan-blue, and the main transcription factors involved in EndMT are shown in yellow.

TABLE 1 | References that serve as a base for each component of the update rule.

Molecule	References
AP1	Nishita et al. (2010); Sundqvist et al. (2013); Zhao et al. (2014)
CTNNB	Wu et al. (2012); Menezes (2014)
DLL4	Niessen et al. (2008); Chang et al. (2011)
ETS1	Murakami et al. (2011); Hollenhorst (2012); Chen et al. (2017); Yalim-Camci et al. (2019)
FGF2	Murakami et al. (2011); Chen et al. (2012); Ichise et al. (2014); Yang et al. (2015); Correia et al. (2016)
FLI1	Lelièvre et al. (2002); Pimanda et al. (2007b); Abedin et al. (2014); Oh et al. (2015); Tsang et al. (2017)
GATA2	Pimanda et al. (2007b); Kanki et al. (2011); Coma et al. (2013)
HIF1a	Kumar et al. (2014); Xu et al. (2015); Baldea et al. (2018)
LEF1	Ishitani et al. (2005); Medici et al. (2006); Yun et al. (2007); Phng et al. (2009)
NFκB	Liu et al. (2018)
NOTCH	Phng et al. (2009)
NRARP	Phng et al. (2009)
NRP1	Teruyama et al. (2001); Coma et al. (2013); Aspalter et al. (2015)
PDGF_AB	Liu et al. (2018)
SMAD1	van Meeteren and Ten Dijke (2012); Pardali et al. (2017)
SMAD2	van Meeteren and Ten Dijke (2012); Aspalter et al. (2015); Pardali et al. (2017)
SMAD6	Ishida et al. (2000); Pimanda et al. (2007a); Mouillesseaux et al. (2016)
SNAI1	Peiro et al. (2006); Julien et al. (2007); Kokudo et al. (2008); Medici et al. (2011); Yu et al. (2013); Chen and Gridley (2013b); Menezes (2014); Frias et al. (2015); Xu et al. (2015); Saitoh et al. (2016); Wang et al. (2017a); Liu et al. (2018)
SNAI2	Niessen et al. (2008); Lambertini et al. (2010); Casas et al. (2011); Chen and Gridley (2013b); Cooley et al. (2014); Kumar et al. (2015); Welch-Reardon et al. (2015); Kobayashi and Ozawa (2018)
STAT3	Chen et al. (2008)
TGFB	van Meeteren and Ten Dijke (2012); Pardali et al. (2017)
TGFBR	Chen et al. (2012); van Meeteren and Ten Dijke (2012); Pardali et al. (2017)
TWIST1	Howe et al. (2003); Zhang et al. (2018)
VEGFA	Forsythe et al. (1996); Chang et al. (2004); Harper and Bates (2008); Kumar et al. (2014); Simons et al. (2016); Weinstein et al. (2017)
VEGFR2	Funahashi et al. (2010); Kanki et al. (2011); Murakami et al. (2011); Abedin et al. (2014); Simons et al. (2016); Weinstein et al. (2017); Liu et al. (2018)
WNT5b	Wu et al. (2012); Menezes (2014); Wang et al. (2017a)
WNT7a	Menezes (2014)
ZEB1	Sánchez-Tilló et al. (2011); Wels et al. (2011); Sanchez-Tillo et al. (2014); Kobayashi and Ozawa (2018)
ZEB2	Takkunen et al. (2006); Katoh and Katoh (2009); Grabitz and Duncan (2012); Zhao et al. (2014)

TABLE 2 | The changes to the update rule components necessary in order to reach a fixed pattern of molecular activation for each expected cell type or behavior.

Modification	Reason or desired effect
In ECs, E47 should be absent so that TWIST1 inhibits the transcription of SNAI1	Otherwise TWIST1 would activate SNAI1 in Stalk ECs.
GATA2 must not activate the transcription of TWIST1 in ECs	Prevents TWIST1, SNAI1 and SNAI2 activation in Phalanx cells.
Both SNAI1 and GATA2 should be required to inhibit SNAI2 expression	to preserve SNAI2 expression in Stalk cells.

$$ETS1(t+1) = VEGF2(t) \vee FGF2(t) \vee ZEB2(t) \vee AP1(t) \quad (4)$$

$$FGF2(t+1) = FGF2(t) \quad (5)$$

$$FLI1(t+1) = FLI1(t) \vee GATA2(t) \vee HIF1\alpha(t) \quad (6)$$

$$GATA2(t+1) = FLI1(t) \vee GATA2(t) \vee VEGFR2(t) \quad (7)$$

$$HIF1\alpha(t+1) = HIF1\alpha(t) \quad (8)$$

$$LEF1(t+1) = LEF1(t) \vee NRARP(t) \vee SMAD2(t) \vee NF\kappa B(t) \vee CTNNB(t) \quad (9)$$

$$NF\kappa B(t+1) = PDGF_AB(t) \quad (10)$$

$$NOTCH(t+1) = DLL4(t) \wedge \neg NRARP(t) \quad (11)$$

$$NRARP(t+1) = NOTCH(t) \quad (12)$$

$$NRP1(t+1) = VEGFA(t) \wedge (ETS1(t) \vee GATA2(t)) \wedge \neg NOTCH(t) \quad (13)$$

$$PDGF_AB(t+1) = PDGF_AB(t) \quad (14)$$

$$SMAD1(t+1) = TGF\beta R(t) \wedge \neg SMAD6(t) \quad (15)$$

$$SMAD2(t+1) = TGF\beta R(t) \wedge \neg SMAD6(t) \wedge NRP1(t) \quad (16)$$

$$SMAD6(t+1) = FLI1(t) \wedge GATA2(t) \wedge (NOTCH(t) \vee SMAD1(t)) \quad (17)$$

$$SNAI1(t+1) = (HIF1\alpha(t) \vee STAT3(t) \vee CTNNB(t) \vee AP1(t) \vee NF\kappa B(t) \vee SMAD2(t)) \wedge (NOTCH(t) \vee CTNNB(t)) \wedge \neg TWIST1(t) \wedge \neg SNAI1(t) \wedge \neg SNAI2(t) \quad (18)$$

$$SNAI2(t+1) = \neg(SNAI1(t) \wedge GATA2(t)) \wedge (SMAD2(t) \vee SNAI2(t) \vee TWIST1(t) \vee LEF1(t) \vee NOTCH(t)) \quad (19)$$

$$STAT3(t+1) = VEGFR2(t) \quad (20)$$

$$TGF\beta(t+1) = TGF\beta(t) \quad (21)$$

$$TGF\beta R(t+1) = TGF\beta(t) \wedge \neg FGF2(t) \quad (22)$$

$$TWIST1(t+1) = CTNNB(t) \vee HIF1\alpha(t) \quad (23)$$

$$VEGFA(t+1) = (\neg WNT5b(t) \wedge \neg TGF\beta R(t) \wedge (STAT3(t) \vee HIF1\alpha(t))) \vee VEGFA(t) \quad (24)$$

$$VEGFR2(t+1) = VEGFA(t) \wedge \neg SNAI1(t) \wedge \neg NOTCH(t) \wedge (FLI1(t) \vee GATA2(t) \vee ETS1(t)) \quad (25)$$

$$WNT5b(t+1) = WNT5b(t) \quad (26)$$

$$WNT7a(t+1) = WNT7a(t) \quad (27)$$

$$ZEB1(t+1) = SNAI2(t) \vee CTNNB(t) \vee LEF1(t) \quad (28)$$

$$ZEB2(t+1) = HIF1\alpha(t) \vee ETS1(t) \vee SMAD2(t) \vee AP1(t) \vee SNAI1(t) \quad (29)$$

Feedback Circuits

The regulatory network, as shown in **Figure 2**, contains a total of 74 feedback circuits. However, only 11 circuits are functional, eight of them positive and three negative (**Supplementary Table S1**). The three functional negative circuits are of particular importance because they originate the cyclic behavior in the dynamical model. Specifically, *a*) SNAI1 inhibits itself; *b*) NOTCH activates NRARP, which in turn inhibits NOTCH; and *c*) SMAD1 activates SMAD6, which inhibits SMAD1. Additionally, the microenvironment is defined by the pattern of activation of seven source molecules, and since there are possible microenvironments, the minimum number of attractors is 128. However, the simulated dynamic behavior results in 444 attractors due to the effect of the functional positive feedback circuits (Azpeitia et al., 2017; Rozum and Albert, 2018). This is in qualitative accordance with the large diversity of EC and MC patterns of molecular activation that has been reported in the literature (Chi et al., 2003; Ho et al., 2018).

Fixed and Cyclic Patterns of Molecular Activation

The analysis of the dynamical behavior of the model shows that the system has 444 attractors, 169 of which are fixed points, 18 are cyclic attractors of size 2, and 257 are cyclic attractors of size 4. These attractors correspond to stationary or cyclic patterns of molecular activation, which in turn can be identified with specific cell types and cellular states. Using the procedure described in Section 2.3, these attractors can be identified as belonging to Endothelial, Mesenchymal, Phalanx, Stalk, and Tip sets, which intersect each other but that can be dissected into nine disjoint sets, as shown in **Figure 3**. The specific active and inactive molecules for all these sets are shown in **Table 3**.

The presence or absence of a seven ligands in the extracellular microenvironment together with the pattern of molecular activation within the cell define the attractor reached after a simulation of the dynamic behavior of our model. In order to illustrate how this process functions, we simulated the behavior

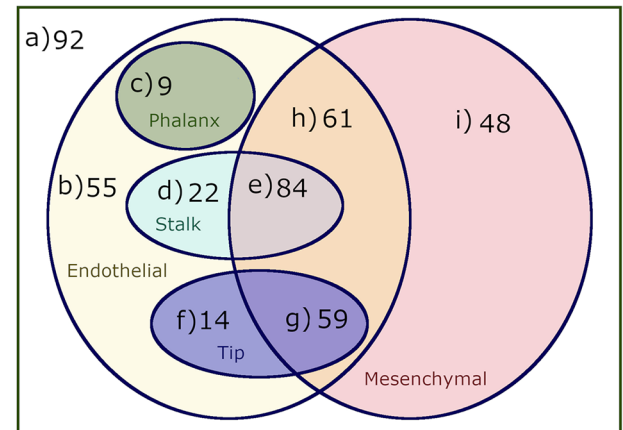


FIGURE 3 | A venn diagram of the attractors reached by simulating the dynamic behavior of our Boolean model. We classified the attractors as mesenchymal, endothelial, phalanx, stalk, and tip cells, forming nine disjoint sets that represent the following cell types and behaviors: *a*) Cell types that are neither endothelial nor mesenchymal, *b*) Endothelial cell types that are not mesenchymal and do not behave as phalanx, stalk or tip cells, *c*) Endothelial and nonmesenchymal phalanx cell types, *d*) Endothelial and nonmesenchymal stalk cell types, *e*) Endothelial and mesenchymal stalk cell types, *f*) Endothelial and nonmesenchymal tip cell types, *g*) Mesenchymal and endothelial tip cell types, *h*) Endothelial and mesenchymal cell types that do not exhibit tip stalk or phalanx cell behavior, and *i*) Mesenchymal only cell types.

of our model cell in an EndMT-inducing extracellular microenvironment where HIF1 and FGF2 are absent while DLL4, TGFβ, WNT5b, WNT7a, and PDGF_AB are present. The attractors reached by our model under such conditions are shown in **Table 4**. Attractor 1 corresponds to the expected pattern of expression of a mesenchymal stalk cell. Note that here, FLI1 and GATA2 are active, and their activity is sustained by three positive feedback circuits. Attractor 2 represents the pattern of expression of an EC that competes with its neighbors to become a tip cell, and cannot fully become a tip cell due to the paracrine effect of the DLL4 ligand expressed by its neighbors (Jakobsson et al., 2010). Note that in Attractor 2, in addition to GATA2 and FLI1, VEGFA is active, and its activity is sustained by a positive feedback circuit. Attractor 3 represents a nonendothelial mesenchymal cell where FLI1, GATA2, and VEGFA are inactive.

Robustness Analysis

The first type of robustness analysis was the evaluation of the effects on cell types and behaviors caused by the simulation of all possible single loss and gain-of-function mutations in the model. These are summarized in **Table 5**. Observe that only 24 of 58 possible single mutations do not alter the qualitative behavior of the model, as measured by the type of resulting attractors. The relative low robustness to gene mutations is likely to be due to the fact that we only included in our model molecules with an important biological role. Furthermore, the simulation of the other single mutations all results in the disappearance of certain cell types. However, according to our model, each cell type or

TABLE 3 | Simulated cell type characteristics: Each cell type is represented by a group of attractors sorted as explained in *Molecular Pattern Identification*

Cell type	Active molecules	Inactive molecules	Fraction of the state space covered by the trap space
Non-EC and Non-MC	N.A.	AP1, FLI1, GATA2, HIF1 α , NRP1, SMAD2, SMAD6, SNAI1, STAT3, VEGFR2, WNT5b	0.64206%
ECs	FLI1, GATA2	SNAI1	97.24696%
ECs only	FLI1, GATA2, SNAI2, ZEB1	CTNNB, HIF1 α , TWIST1, WNT5b, WNT7a	7.12915%
MCs	ETS1, SNAI2, TWIST1, ZEB1, ZEB2	SNAI1	87.32046%
MCs only	CTNNB, ETS1, LEF1, SNAI2, TWIST1, ZEB1, ZEB2	FLI1, GATA2, HIF1 α , NRP1, SMAD2, SMAD6, SNAI1, STAT3, VEGFA, VEGFR2	2.11098%
ECs and MCs	ETS1, FLI1, GATA2, SNAI2, TWIST1, ZEB1, ZEB2	SNAI1	85.20948%
ECs and MCs only	ETS1, FLI1, GATA2, SNAI2, TWIST1, ZEB1, ZEB2	SNAI1	29.79242%
Phalanx	FLI1, GATA2	AP1, CTNNB, DLL4, HIF1 α , LEF1, NFkB, NOTCH, NRARP, NRP1, PDGF_AB, SMAD2, SNAI1, SNAI2, STAT3, TWIST1, VEGFA, VEGFR2, WNT5b, WNT7a, ZEB1	0.00322%
Stalk	CTNNB, FLI1, GATA2, LEF1, SNAI2, TWIST1, ZEB1	NRP1, SMAD2, SNAI1, STAT3, VEGFA, VEGFR2	26.56887%
Stalk MCs	CTNNB, ETS1, FLI1, GATA2, LEF1, SNAI2, TWIST1, ZEB1, ZEB2	NRP1, SMAD2, SNAI1, STAT3, VEGFA, VEGFR2	26.43057%
Stalk Non-MCs	CTNNB, FLI1, GATA2, LEF1, SNAI2, TWIST1, WNT7a, ZEB1	AP1, FGF2, HIF1 α , NRP1, SMAD2, SNAI1, STAT3, VEGFA, VEGFR2, WNT5b	0.13830%
Tip	ETS1, FLI1, GATA2, NRP1, STAT3, VEGFA, VEGFR2, ZEB2	DLL4, NOTCH, NRARP, SNAI1	33.7533%
Tip MCs	ETS1, FLI1, GATA2, NRP1, SNAI2, STAT3, TWIST1, VEGFA, VEGFR2, ZEB1, ZEB2	DLL4, NOTCH, NRARP, SNAI1	28.98649%
Tip Non-MCs	ETS1, FLI1, GATA2, NRP1, STAT3, VEGFA, VEGFR2, ZEB2	CTNNB, DLL4, HIF1 α , NOTCH, NRARP, SNAI1, TWIST1, WNT5b, WNT7a	4.76681%

TABLE 4 | The attractors reached by our model in an endothelial-to-mesenchymal transition (EndMT)-inducing extracellular microenvironment where HIF1 and FGF2 are absent while DLL4, TGFB, WNT5b, WNT7a, and PDGF_AB are present.

Attractor	1	1	1	1	2	2	2	2	3	3	3	3
AP1	1	1	1	1	1	1	1	1	1	1	1	1
CTNNB	1	1	1	1	1	1	1	1	1	1	1	1
DLL4	1	1	1	1	1	1	1	1	1	1	1	1
ETS1	1	1	1	1	1	1	1	1	1	1	1	1
FGF2	0	0	0	0	0	0	0	0	0	0	0	0
FLI1	1	1	1	1	1	1	1	1	0	0	0	0
GATA2	1	1	1	1	1	1	1	1	0	0	0	0
HIF1a	0	0	0	0	0	0	0	0	0	0	0	0
LEF1	1	1	1	1	1	1	1	1	1	1	1	1
NFkB	1	1	1	1	1	1	1	1	1	1	1	1
NOTCH	0	1	1	0	0	0	1	1	0	1	1	0
NRARP	0	0	1	1	1	0	0	1	0	0	1	1
NRP1	0	0	0	0	0	1	1	0	0	0	0	0
PDGF_AB	1	1	1	1	1	1	1	1	1	1	1	1
SMAD1	0	1	1	0	0	0	1	1	1	1	1	1
SMAD2	0	0	0	0	0	0	1	1	0	0	0	0
SMAD6	0	0	1	1	1	0	0	1	0	0	0	0
SNAI1	0	0	0	0	0	0	0	0	0	0	0	0
SNAI2	1	1	1	1	1	1	1	1	1	1	1	1
STAT3	0	0	0	0	0	0	1	1	0	0	0	0
TGFB	1	1	1	1	1	1	1	1	1	1	1	1
TGFBF	1	1	1	1	1	1	1	1	1	1	1	1
TWIST1	1	1	1	1	1	1	1	1	1	1	1	1
VEGFA	0	0	0	0	1	1	1	1	0	0	0	0
VEGFR2	0	0	0	0	0	1	1	0	0	0	0	0
WNT5b	1	1	1	1	1	1	1	1	1	1	1	1
WNT7a	1	1	1	1	1	1	1	1	1	1	1	1
ZEB1	1	1	1	1	1	1	1	1	1	1	1	1
ZEB2	1	1	1	1	1	1	1	1	1	1	1	1

Active molecules are shown in white, and inactive molecules are shown in gray.

behavior is very robust to single gain- and loss-of-function gene mutations. Notably, the larger the number of attractors classified as a cell type or behavior, the larger the robustness of the cell type is to gene mutations.

As for the robustness of each cell type against noise in the update rule, in all cases, the original model reached slightly more attractors than the mean of the 100 000 networks with perturbed update rules, as shown in **Figure 4**. Observe that the standard deviation in the number of attractors for all cell types and behaviors is relatively large, and therefore the robustness of the number of attractors that represent each cell type or behavior is low. The maximum numbers of attractors for each of the cell types were the following: nECs/MCs 238, EOnly 243, Phalanxes 27, nMCStalks 106, MCStalks 216, nMCTips 70, MCTips 149, MCEOnly 129, MCs/ECs 237, while the minimum values reached 0 for all cell types and behaviors.

Regarding the robustness of the components of the update rule to noise in the activation value, all such components are sensitive to less than 4.6% of all single bit perturbations that are the most likely to occur, as shown in **Figure 5**. Notice that the components corresponding to ZEB2, LEF1, and ETS1 are under

TABLE 5 | The simulated single gain and loss-of-function mutations that affect each cell type.

Effect	Mutations	Robustness
Wild type	AP1-, DLL4-, FGF2-, HIF1a-, LEF1-, NFkB-, NOTCH-, NRARP-, PDGF_AB-, SMAD1-, SMAD1+, SMAD2-, SMAD6-, SMAD6+, SNAI1-, STAT3-, STAT3+, TGFB-, TGFB+, TGFB-, TGFB+, VEGFR2-, WNT5b-, ZEB1+	41.38%
Loss of nonendothelial and nonmesenchymal cells	FLI1+, GATA2+, HIF1a+, VEGFR2+, WNT5b+	91.38%
Loss of nonmesenchymal, nonphalanx, nontip, and nonstalk ECs	CTNNB+, FLI1-, GATA2-, HIF1a+, WNT5b+, WNT7a+	89.66%
Loss of phalanx cells	CTNNB+, DLL4+, FLI1-, GATA2-, HIF1a+, LEF1+, NFkB+, NOTCH+, NRARP+, NRP1+, PDGF_AB+, SMAD2+, SNAI1+, SNAI2+, TWIST1+, VEGFA+, WNT5b+, WNT7a+	68.97%
Loss of nonmesenchymal stalk cells	AP1+, CTNNB-, ETS1+, FGF2+, FLI1-, GATA2-, HIF1a+, NRP1+, SMAD2+, SNAI1+, SNAI2-, VEGFA+, VEGFR2+, WNT5b+, WNT7a-, ZEB2+	72.41%
Loss of mesenchymal stalk cells	CTNNB-, FLI1-, GATA2-, NRP1+, SNAI1+, SNAI2-, TWIST1-, VEGFA+, ZEB1-, ZEB2-	82.76%
Loss of nonmesenchymal tip cells	CTNNB+, DLL4+, ETS1-, FLI1-, GATA2-, HIF1a+, NOTCH+, NRP1-, TWIST1+, VEGFA-, WNT5b+, WNT7a+	79.31%
Loss of mesenchymal tip cells	DLL4+, ETS1-, FLI1-, GATA2-, NOTCH-, NRP1-, SNAI2-, TWIST1-, VEGFA-, ZEB1-, ZEB2-	81.03%
Loss of mesenchymal ECs that are neither phalanx, tip nor stalk cells	FLI1-, GATA2-, NRP1+, SNAI2-, TWIST1-, ZEB1-, ZEB2-	87.93%
Loss of nonendothelial mesenchymal cells	CTNNB-, FLI1+, GATA2+, HIF1a+, SNAI2-, TWIST1-, VEGFA+, VEGFR2+, ZEB1-, ZEB2-	82.76%

2%; SNAI1, FLI1, VEGFR2, GATA2, SMAD2, and ZEB1 have a sensitivity of about 2.5%, while most of the other components have a sensitivity between 3.4% and 3.5% except for SNAI2 and VEGFA that have a sensitivity of over 4%. Nonetheless, the sensitivity of all the components increases as the number of flipped bits increases (Figure 6). When the activity of 15 nodes is affected, the components can be grouped by their sensitivity into 5 categories: 1) VEGFA, STAT3, NRARP, FGF2, PDGF_AB, HIF1a, DLL4, WNT7a, WNT5b, and NFkB, TGFβ with a sensitivity between 49.4% and 52%. 2) TGFBR, TWIST1, CTNNB, AP1, NOTCH, SMAD1, and SNAI2 with a sensitivity between 38.1% and 40.55%. 3) SMAD6, and NRP1 with a sensitivity of 31.1%, and 31.2% respectively. 4) SNAI1, VEGFR2, SMAD2, ZEB1, GATA2, and FLI1 with a sensitivity between 17.7% and 22.8%. And 5) ETS1, LEF1, and ZEB2 with a sensitivity between 5.9% and 11.8%. Note that there exists a trend that is independent of the number of flipped bits, where the sensitivity for the components that represent ligands that define the extracellular microenvironment is high, and the sensitivity of the components that represent molecules used as cell type

markers is low. The very low sensitivity of the components that represent ETS1, LEF1, and ZEB2 is in accordance with the importance that of the three transcription factors not only during EndMT, but also during other cell differentiation processes. Specifically, ZEB2 is involved in T cell differentiation (Goossens et al., 2019) and neurological development (Epifanova et al., 2018). LEF1 is important during osteogenesis (Li et al., 2018), immune cell regulation (Chae and Bothwell, 2018), and hair follicle development (Abaci et al., 2018). ETS1 is an important regulator of lymphatic cell differentiation and physiology (Garrett-Sinha et al., 2016).

Finally, one of the goals of this modeling effort was to understand the conditions that cause an EC cell to change, either by behaving differently or by differentiating partially or fully into a mesenchymal cell. Further, EndMT is a gradual, and reversible process and therefore we also aimed to fathom the conditions that cause MEnT. Moreover, the intermediate states reached through partial-EndMT are important due to their physiological role during sprouting angiogenesis (Welch-Reardon et al., 2014), and due to the similarity between EndMT and EMT; it seems likely that the intermediate states are also important from a dynamic perspective (Lu et al., 2013; Li et al., 2016). In order to grasp the conditions that lead to EndMT and MEnT, for each cyclic or fixed pattern of molecular activation of our model, we simulated all possible perturbations in the molecules that are either microenvironmental signals or the main transcription factors involved in the regulation of EC identity. Specifically DLL4, FGF2, FLI1, GATA2, HIF1α, PDGF_AB, TGFβ, VEGFA, WNT5b, and WNT7a. The possible effects of the 1024 perturbations are available for the interested reader as the 81 **Supplementary Files** in the folder https://github.com/NathanWeinstein/EndMT/T_Results.zip in the format used to export R objects, namely, RData and are summarized in **Table 6** which can be interpreted as a cell type or behavior transition graph (Figure 7).

Model Validation

An exhaustive comparison between the global effect of all possible single gene mutations in the model and the reported experimental results are presented in **Supplementary Table S2**, and summarized in **Table 7**. Overall, the behavior of the model is very good at recovering the effect of a large proportion of the reported mutants. Notice that several of the discrepancies are because the model does not incorporate multicellular or morphological effects, or because the reported effects involve some molecules not included in the model. This is encouraging given the qualitative nature of the model. Of the 58 possible mutations, we successfully simulated the specific effects reported for 37 (63.8%) of them. Furthermore, the effects of 4 (6.9%) mutations constitute predictions of our model. 13 (22.4%) mutations cause multicellular effects that we could not reproduce using our model. Two mutations (3.45%) cause morphological changes in the shapes of cells that are also beyond the scope of our model. 4 (6.9%) mutations affect molecules that we did not include in the model. 4 (6.9%) mutations have an effect that was only observed in lymphatic

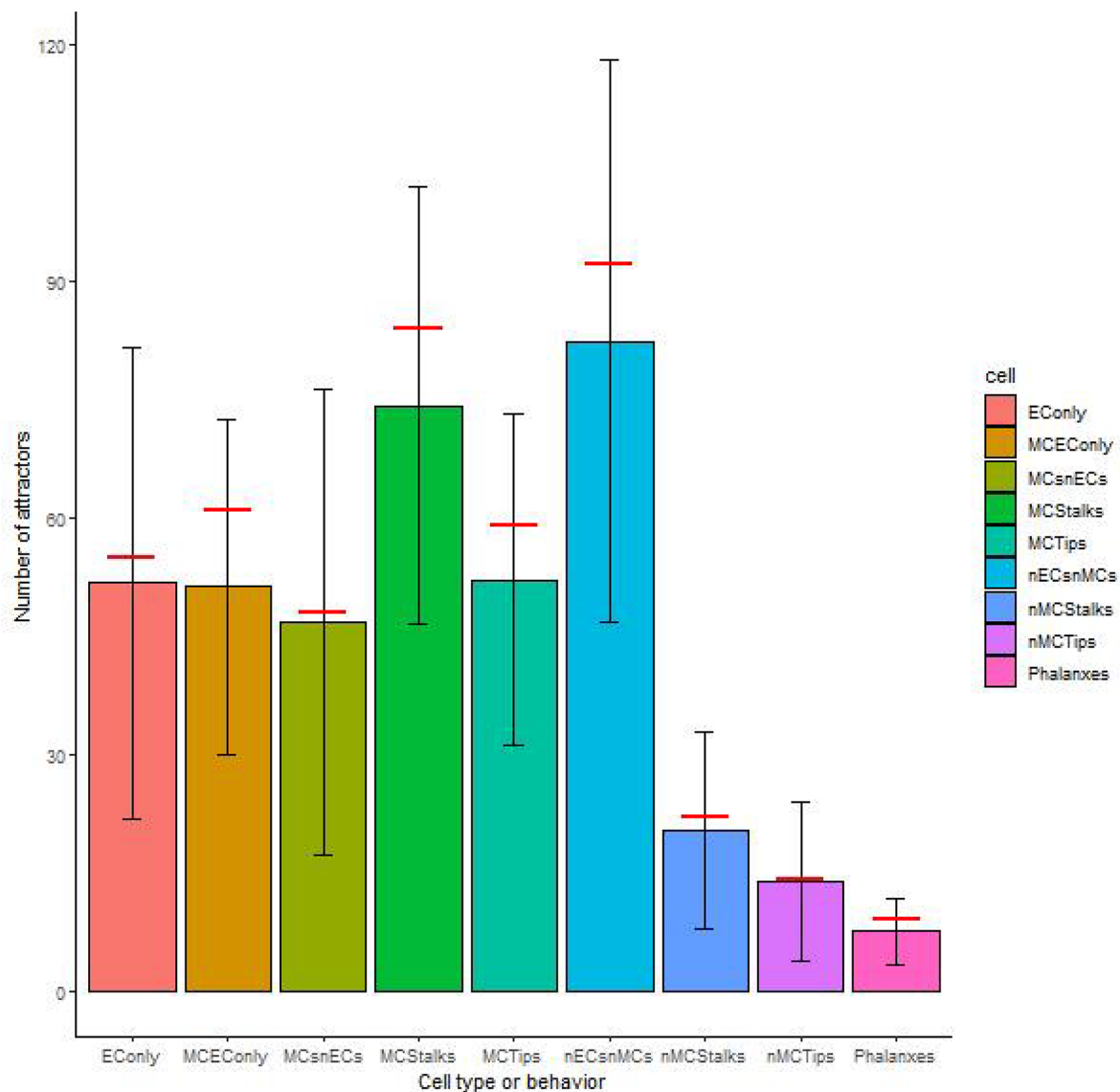


FIGURE 4 | The robustness of the cell types and behaviors to changes in the update rule: The height of the bars represents the median number of attractors of each cell type or behavior, the error bars represent one standard deviation over and under the mean respectively, and the red horizontal line segments represent the number of attractors of each cell type or behavior on our original model.

ECs. Some of the reported effect of 7 (12.1%) mutations was not recovered by the simulated behavior of our model. For two mutations (3.45%), there are conflicting effects reported in the literature.

Simulating EC Behavior and Differentiation During Developmental Processes and the Progression of Diseases Related to EndMT

During early heart valve formation, *embryonic heart cushion EndMT* is triggered by TGF, WNT, and NOTCH signaling and is inhibited by VEGF signaling (von Gise and Pu, 2012). This behavior is recovered by the simulated dynamic behavior of our

model, (TGF β +, WNT5b+, and NOTCH+ increase the fraction of mesenchymal attractors, and VEGFA+ prevents full-EndMT). TGF β 2-, ALK1-, ALK5-, SMAD4-, SMAD6+, NOTCH1-, VEGFA+, CTNNB-, and PDGF_AB- inhibit EndMT, and cause endocardial cushion hypoplasia. In contrast SMAD6- causes heart valve hyperplasia by increasing the number of MCs (von Gise and Pu, 2012). According to the simulated dynamic behavior of our model, the simulated loss of TGF β , which represents TGF- β 2, and the loss of TGF β R, which represents all TGF receptors including ALK1 and ALK2, reduces the fraction of mesenchymal attractors. Further, the loss of the cofactor SMAD4 can be simulated as the loss of

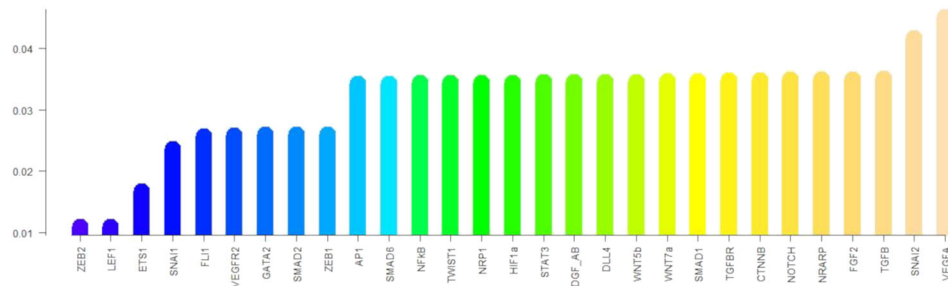


FIGURE 5 | The sensitivity of each component of the update rule: The height of the bars represents the sensitivity of the components of the update rule to perturbations that affect one node.

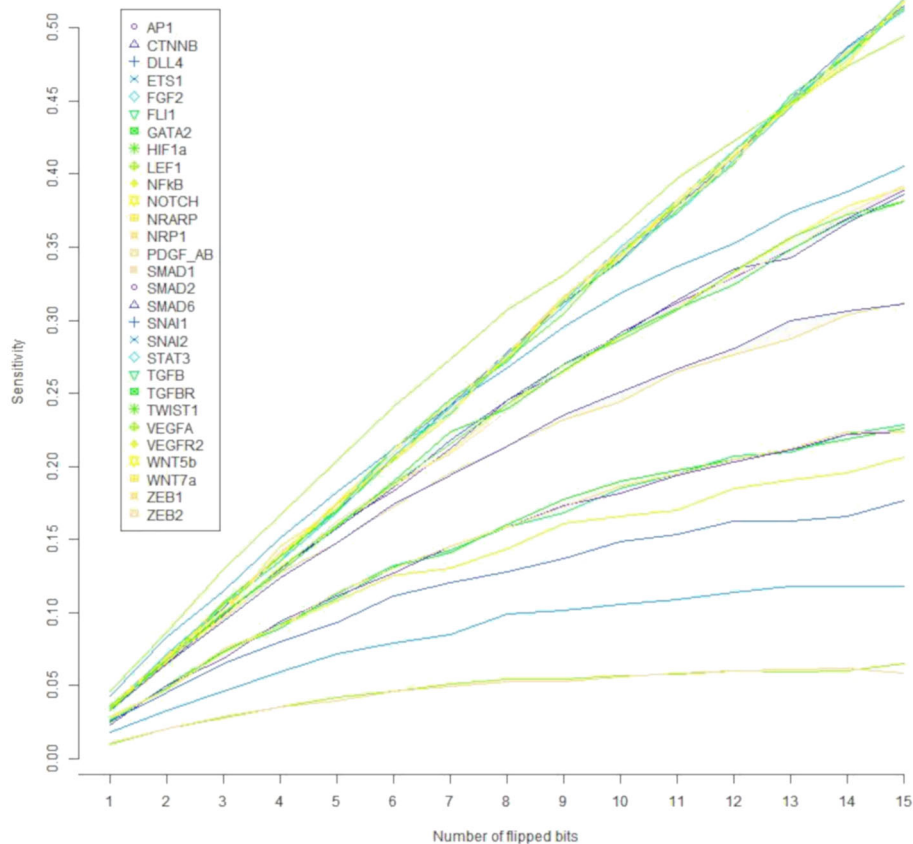


FIGURE 6 | The effect of the number of flipped bits on the sensitivity of the update rule components. Note that the components segregate according to their sensitivity to molecular activation noise into five categories.

both SMAD1 and SMAD2 function and does not affect the fraction of mesenchymal attractors. The simulated SMAD6 gain of function mutation also does not affect the fraction of mesenchymal attractors. Simulated NOTCH loss of function reduces the fraction of mesenchymal attractors. Moreover, simulated VEGFA gain of function and CTNNB loss of function prevent the existence of nonendothelial mesenchymal

attractors. Additionally, the simulated loss of PDGF_AB reduces the fraction of mesenchymal attractors. Lastly, the simulated loss of SMAD6 function exhibits a slight increase in the fraction of mesenchymal attractors.

The initial stages of *atherosclerosis and vascular calcification* are characterized by neointimal hyperplasia. Local disparity in shear stress is associated to neointimal lesions. While most

TABLE 6 | The number and the characteristics of the perturbations in the activation state of the molecules DLL4, FGF2, FLI1, GATA2, HIF1 α , PDGF_AB, TGF β , VEGFA, WNT5b, and WNT7a that cause cell type or cell behavior changes: Each cell contains first the number of perturbations that trigger the transitions, if the number is bigger than 0, the cell contains the molecules that are active in all perturbations +(), as well as the molecules that are inactive in all perturbations -().

	nECsnMCs	EConly	Phalanxes	nMCStalks	MCStalks	nMCTips	MCTips	MCEConly	MCsnECs
nECsnMCs	40, -(FLI1, GATA2, HIF1a, WNT5b)	80, -(HIF1a, WNT5b, WNT7a)	12, -(DLL4, HIF1a, PDGF_AB, VEGFA, WNT5b, WNT7a)	24, +(WNT7a) -(FGF2, HIF1a, VEGFA, WNT5b)	288, -(VEGFA)	32, +(VEGFA) -(DLL4, HIF1a, WNT5b, WNT7a)	288, -(DLL4)	296	48, -(FLI1, GATA2, HIF1a, VEGFA)
EConly	40, -(FLI1, GATA2, HIF1a, WNT5b)	90, -(HIF1a, WNT5b, WNT7a)	0	24, +(WNT7a) -(FGF2, HIF1a, VEGFA, WNT5b)	324, -(VEGFA)	64, -(DLL4, HIF1a, WNT5b, WNT7a)	320, -(DLL4)	328	48, -(FLI1, GATA2, HIF1a, VEGFA)
Phalanxes	30, -(FLI1, GATA2, HIF1a, WNT5b)	68, -(HIF1a, WNT5b, WNT7a)	12, -(DLL4, HIF1a, PDGF_AB, VEGFA, WNT5b, WNT7a)	24, +(WNT7a) -(FGF2, HIF1a, VEGFA, WNT5b)	288, -(VEGFA)	32, +(VEGFA) -(DLL4, HIF1a, WNT5b, WNT7a)	288, -(DLL4)	296	48, -(FLI1, GATA2, HIF1a, VEGFA)
nMCStalks	40, -(FLI1, GATA2, HIF1a, WNT5b)	80, -(HIF1a, WNT5b, WNT7a)	0	24, +(WNT7a) -(FGF2, HIF1a, VEGFA, WNT5b)	264, -(VEGFA)	32, +(VEGFA) -(DLL4, HIF1a, WNT5b, WNT7a)	288, -(DLL4)	296	40, -(FLI1, GATA2, HIF1a, VEGFA)
MCStalks	16, -(FLI1, GATA2, HIF1a, VEGFA, WNT5b, WNT7a)	80, -(HIF1a, WNT5b, WNT7a)	0	0	288, -(VEGFA)	32, +(VEGFA) -(DLL4, HIF1a, WNT5b, WNT7a)	288, -(DLL4)	296	48, -(FLI1, GATA2, HIF1a, VEGFA)
nMCTips	0	72, -(HIF1a, WNT5b, WNT7a)	0	0	288, -(VEGFA)	64, -(DLL4, HIF1a, WNT5b, WNT7a)	320, -(DLL4)	328	0
MCTips	0	72, -(HIF1a, WNT5b, WNT7a)	0	0	288, -(VEGFA)	64, -(DLL4, HIF1a, WNT5b, WNT7a)	320, -(DLL4)	328	0
MCEConly	16, -(FLI1, GATA2, HIF1a, VEGFA, WNT5b, WNT7a)	90, -(HIF1a, WNT5b, WNT7a)	0	0	324, -(VEGFA)	64, -(DLL4, HIF1a, WNT5b, WNT7a)	320, -(DLL4)	328	48, -(FLI1, GATA2, HIF1a, VEGFA)
MCsnECs	16, -(FLI1, GATA2, HIF1a, VEGFA, WNT5b, WNT7a)	80, -(HIF1a, WNT5b, WNT7a)	0	0	288, -(VEGFA)	32, +(VEGFA) -(DLL4, HIF1a, WNT5b, WNT7a)	288, -(DLL4)	296	48, -(FLI1, GATA2, HIF1a, VEGFA)

The perturbations that do not cause a change in cell type or behavior are shown in white, those that cause a full endothelial-to-mesenchymal transition (EndMT) transition are shaded in dark cyan, those that represent a partial EndMT are shaded in medium cyan, those that represent an endothelial activation are shown in light cyan, those that represent an endothelial deactivation are shown in light amber, those that represent a partial mesenchymal-to-endothelial transition are shown in amber, full mesenchymal-to-endothelial transitions are shown in dark amber, and other perturbations are shown in gray.

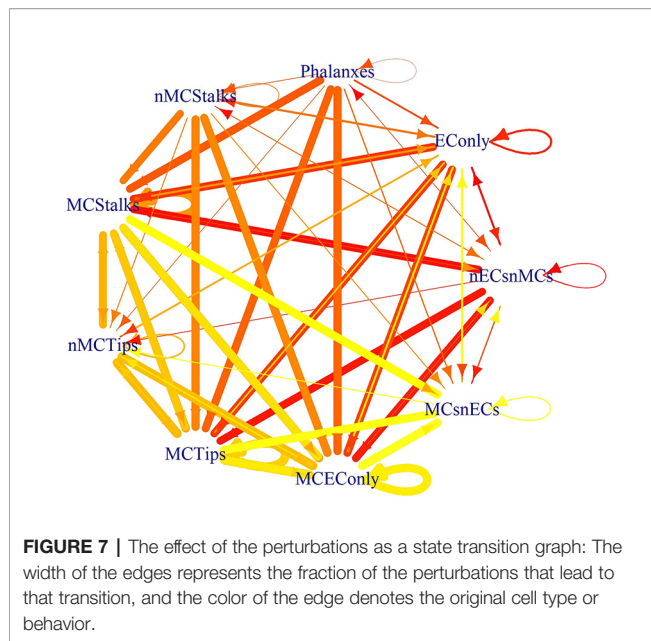


TABLE 7 | The capacity of our model to simulate the effects of mutations as reported in the literature.

Successfully simulated	CTNNB+, DLL4+, ETS1-, FGF2-, FLI1-, FLI1+, GATA2-, HIF1 α -, HIF1 α +, LEF1+, NF κ B+, NOTCH1-, NOTCH+, NRARP+, PDGF_AB-, PDGF_AB+, SMAD6-, SMAD6+, SNAI1+, SNAI2-, SNAI2+, STAT3-, TGF β -, TGF β +, TGF β R-, TGF β R+, TWIST1-TWIST1+, VEGFA-WNT5b-WNT5b+, WNT7a+, ZEB1-, ZEB2-, ZEB2+.
Affect the likelihood of transient patterns of expression	AP1-, NF κ B-, STAT3+.
Predictions of our model	AP1+, GATA2+, SMAD1+, SMAD2+.
Unable to simulate the multicellular effect	CTNNB-, DLL4-, DLL4-, LEF1-, NRARP-, SMAD6-, SNAI2-, SNAI2+, VEGFA+, VEGFR2-, VEGFR2+, WNT7a-, ZEB2-.
Unable to simulate the morphological cell change	SNAI1-, SNAI1+
Affects molecules not included in our model	ETS1+, SNAI2-, SNAI2+, ZEB2+
Some effects were only observed in lymphatic ECs	FGF2-, FGF2+, WNT5b-, WNT5b+
Some effects not simulated	SMAD1-, SMAD2-, STAT3+, TGF β R+, TWIST1+, VEGFA+, ZEB1+
Conflicting effects reported in the literature	NRP1-, NRP1+

neointimal cells originate from smooth muscle, some neointimal cells might arise from ECs that undergo EndMT. ECs that are exposed to disturbed flow undergo EndMT; conversely, uniform laminar shear stress hinders EndMT through KLF2, KLF4, MEK5, and ERK5 (Moonen et al., 2015). Molecularly, ERK5 is the main mitogen-activated protein kinase (MAPK) involved in the regulation of cardiovascular development (Nishimoto and

Nishida, 2006), and VEGF/MAPK signaling activates the transcription of several transcription factors from the ETS family including *ets1* and *fli1* (Wythe et al., 2013). Furthermore, the low shear stress caused by disturbed nonlaminar flow at the sites where neointimal hyperplasia occurs leads to a decrease in *ets1*, and *fli1* expression. Therefore, we can simulate uniform laminar shear stress as the double gain of function mutation *ets1+/fli1+*, and a disturbed nonlaminar flow as the double mutant *ets1-/fli1-*. The simulated effect of *ets1-/fli1-* is the loss of all endothelial attractors, the number of nonendothelial mesenchymal attractors increases from 48 to 347, and the fraction of nonendothelial mesenchymal attractors increases from 0.108 to 0.69. This behavior can be interpreted as an increase in full EndMT resulting from nonlaminar flow. Moreover, the simulated effect of *ets1+/fli1+* is the loss of all nonendothelial mesenchymal attractors as well as an increase in the fraction of mesenchymal attractors from 0.567 to 0.816. This behavior can be interpreted as an increase in partial-EndMT and a complete inhibition of full EndMT. Therefore, according to our model, nonlaminar flow triggers full EndMT, and uniform laminar shear stress prevents full EndMT and upregulates angiogenesis-related partial EndMT. These results are in direct correspondence with the observed effect of uniform laminar and disturbed nonlaminar flow (Wragg et al., 2014).

Another important EndMT-related disease is *pulmonary arterial hypertension*, which is defined as a sustained pulmonary arterial pressure of more than 25 mm Hg at rest or more than 30 mm Hg during exercise, with a left ventricular pressure at the end of the diastole and a mean pulmonary-capillary wedge pressure lower than 15 mm Hg. The lung tissue of patients affected with pulmonary arterial hypertension is characterized by increased medial thickness, intimal fibrosis, plexiform lesions, and pulmonary arteriolar occlusion (Farber and Loscalzo, 2004). EndMT is involved in many of the pathological mechanisms associated with pulmonary arterial hypertension (Kovacic et al., 2019). At the molecular level, most cases of heritable pulmonary arterial hypertension involve mutations that affect the bone morphogenic protein (BMP) branch of the TGF signalling pathway including *ACVRL1(ALK1)*, *BMPRII*, *ENG*, *SMAD1*, *SMAD4*, and *SMAD9*. Furthermore, *BMPRII* siRNA increases the expression of SNAI2 (Hopper et al., 2016). According to our model, the simulated gain of function mutation for SNAI2 increases the fraction of mesenchymal attractors, which is consistent with the experimental evidence.

Finally, hypoxia-induced EndMT is another important mechanism involved in the pathophysiology of pulmonary arterial hypertension. HIF-1 α directly binds to the promoter of TWIST1 and activates its expression (Zhang et al., 2018). Pulmonary arterial hypertension patients exhibit high levels of the cytokines IL-1 β and TNF α that in the presence of TGF β induce EndMT in pulmonary arterial ECs *in vitro* (Good et al., 2015). IL-1 β and TNF α induce EndMT by stabilizing NF- κ B (Sánchez-Duffhues et al., 2018). According to the simulated dynamic behavior of our model, a gain-of-function mutation of NF- κ B induces EndMT and elevates the expression of ZEB2.

DISCUSSION

The Model as Theoretical Framework

Our model of the molecular regulatory network involved in the control of EndMT and EC activation integrates a vast amount of published experimental results. Therefore, our model constitutes a theoretical framework that summarizes the current knowledge and allows for the simulation of experiments that explore the molecular mechanisms involved in the regulation of EndMT *in silico*.

Many important questions about EndMT remain unanswered (Welch-Reardon et al., 2015). While such questions require a experimental approach to obtain a conclusive answer, models like the one presented here can be used to generate hypotheses to direct, or at least restrict, all the possible venues of experimental inquiry. In this sense, our model provides an important theoretical framework to understand the regulatory mechanisms behind EndMT. The following paragraphs provide testable hypotheses on some key aspects, according to our model.

Are SNAI1, SNAI2, TWIST1, ZEB1, and ZEB2 all required for EndMT? According to the dynamic behavior of the model, the loss of any of the transcription factors SNAI2, TWIST1, ZEB1, and ZEB2 prevents mesenchymal cell differentiation. Experimentally, the loss of SNAI2 (Niessen et al., 2008), TWIST1 (Mammoto et al., 2018), or ZEB1 (Sanchez-Tillo et al., 2010) prevents EndMT. ZEB2 has many functions in addition to its role during EndMT, its loss causes severe neurodevelopmental defects and cardiovascular malformations (Epifanova et al., 2018), while its specific effect during EndMT still needs to be elucidated. Conversely, the gain of ZEB2 function is sufficient to trigger EndMT (DaSilva-Arnold et al., 2018). SNAI1 over-expression can rescue the heart valve defects caused by the loss of SNAI2 (Niessen et al., 2008). According to our model, SNAI1 gain-of-function increases the fraction of mesenchymal attractors. This implies that it can trigger EndMT under certain circumstances. However, it cannot replace SNAI2 in fixed or cyclic patterns of expression because it inhibits its own expression.

Do SNAI1, SNAI2, TWIST1, ZEB1, and ZEB2 work sequentially, in parallel or in feedback circuits? TWIST1 regulates the transcription of both SNAI1 and SNAI2 (Yu et al., 2013; Forghanifard et al., 2017), while these two inhibit each other (Peiro et al., 2006; Chen and Gridley, 2013b). Then, SNAI1 activates the transcription of ZEB2 (Guaita et al., 2002; Takkunen et al., 2006), and SNAI2 activates the transcription of ZEB1 (Wels et al., 2011). The regulatory network presented here shows that SNAI1 and SNAI2 form part of several other circuits, including two functional feedback circuits where SNAI1 inhibits its own expression and SNAI2 activates its own expression. Furthermore, TWIST1, ZEB1, and ZEB2 appear to work sequentially with SNAI1 and SNAI2. In this context, our model contributes to the unraveling of several molecular circuits relevant for EndMT.

What regulates the expression of EndMT-promoting transcription factors? According to experimental observations (Piera-Velazquez and Jimenez, 2019) captured by our model,

nonlaminar blood flow, inflammation, as well as TGF, WNT and NOTCH signaling pathway activity can trigger EndMT. By contrast, laminar blood flow, hypoxia, and VEGF signaling can inhibit full EndMT. Other molecular mechanisms that have been reported to regulate EndMT include the autocrine TGF activation by ET-1, the most potent known endogenous vasoconstrictor polypeptide that triggers EndMT. CAV-1, the main protein component of caveolae, is an important inhibitor of EndMT, by means of the internalization, trafficking, and degradation of TGF receptors. H₂O₂-induced oxidative stress, NOX2 and NOX4 can induce EndMT *via* TGF signaling. Fatty acid oxidation inhibits EndMT by activating SMAD7 and inhibiting TGF signaling. Hyperglycemia leads to EndMT through increased phosphorylation of ERK1/2, Angiotensin II synthesis, miR-200b and miR-328 upregulation, and ROCK1 activation (Piera-Velazquez and Jimenez, 2019). The wide variety in the patterns of expression that represent each cell type or behavior prevents the specification of the molecules that regulate EndMT. However, if the initial cell type or behavior is known, our model allows the specification of all possible perturbations that might cause a partial or full EndMT. This information is available in the **Supplementary Files** in the folder https://github.com/NathanWeinstein/EndMT/T_Results.zip in the format used to export R objects (.RData), and summarized in **Table 6** and **Figure 7**.

What controls whether cells undergo a full or partial EndMT? Many of the molecular mechanisms involved in the regulation of EndMT and angiogenesis remain unknown. Nevertheless, we know that the activity of several molecules, including NRP1 (Oh et al., 2002; Matkar et al., 2016), SNAI1 (Sun et al., 2018), SNAI2 (Welch-Reardon et al., 2015), WNT5b (Wang et al., 2017a), and WNT7a (Howe et al., 2003; Pahnke et al., 2016) induce both EC activation and EndMT. Furthermore, TWIST1 (Mammoto et al., 2018), ZEB1 (Sanchez-Tillo et al., 2010), and ZEB2 (DaSilva-Arnold et al., 2018) induce EndMT and are not known to be involved in EC activation during angiogenesis. Finally, the activity of FGF2 (Ichise et al., 2014; Yang et al., 2015), and VEGFA (Paruchuri et al., 2006) induce angiogenesis and inhibit full EndMT. According to our model, all the cases that achieve a full EndMT with the loss of EC identity require FLI1, GATA2, HIF1 α , as well as VEGFA inactivity. These molecules, as a group, have not been involved in this process up to now. In this case, our model serves as a guide to study the role of specific molecules, while at the same time providing a hypothesis of its role in the regulatory network.

The Endothelial-to-Mesenchymal Transition in Medicine

EndMT is necessary during embryonic development for heart septation and heart valve morphogenesis. During the span of human life, EndMT is required to maintain heart valve homeostasis and adapt to hemodynamical changes. EndMT deregulation is involved in the pathophysiology of vascular malformation, vascular calcification, pulmonary arterial hypertension, and organ fibrosis (Medici, 2016; Sánchez-Duffhues et al., 2018).

EndMT is critical during the formation of the heart. Human heart development begins with the aggregation of splanchnopleuric mesenchyme cells that form part of the mesoderm into two endocardial tubes in the cardiogenic area of the embryo. Then, the two endocardial tubes fuse to form the primitive heart tube, which then begins to beat. After that, cardiac looping occurs. Next, septation and valve formation transpires (Moorman et al., 2003). Heart valves develop from endocardial cushions through two processes: the deposition of a special kind of ECM called cardiac jelly, and the arrival of mesenchymal cells that are the precursors to valve cells. Most cushion mesenchymal cells are derived from endocardial cells that undergo EndMT, while the rest originate from epicardium and cardiac neural crest cells that undergo EMT (MacGrogan et al., 2014). EndMT is also involved in adult valve homeostasis and disease. Adult heart valves are covered by a layer of ECs that undergo EndMT to replenish valve interstitial cells. Further, mechanical stretch-induced EndMT allows heart valves to adapt to changes in blood flow within the heart. However, excessive EndMT causes heart valve dysfunction thorough thickening or calcification. For instance, excessive EndMT after myocardial infarction can lead to mitral valve leaflet thickening and mitral regurgitation (Bischoff, 2019).

The formation and progression of *arteriovenous malformations and cerebral cavernous malformations* involves EndMT. CCM1, CCM2, and CCM3 loss-of-function mutations cause the formation of cerebral cavernous malformations. EC-specific disruption of the *Ccm1* causes TGF-mediated EndMT. Inhibiting TGF signaling reduces the number and size of vascular lesions caused by CCM1- deficiency (Maddaluno et al., 2013). Arteriovenous malformations are shunts that directly connect the afferent arteries to efferent veins, bypassing the usual capillary network. In addition to the fact that they take a large volume of space and prevent normal tissue perfusion, the nidus of arteriovenous malformations is prone to leaking or bursting, often causing unbearable pain and serious damage. ECs within brain arteriovenous malformations in mice undergo SOX2, and JMJD5-mediated EndMT that can be suppressed using *Pronethalol hydrochloride* (Yao et al., 2019).

Fibrosis is a wound healing process that involves the synthesis and accumulation of ECM proteins. Excessive fibrosis can cause functional organ failure. Myofibroblasts are the essential cell type in the pathogenesis of fibrotic disorders. In systemic sclerosis, cardiac fibrosis, renal fibrosis, idiopathic portal hypertension, colitis, and inflammatory bowel disease, some myofibroblasts express EC markers, suggesting that they originate from ECs that underwent TGF-induced EndMT (Pardali et al., 2017; Sánchez-Duffhues et al., 2018).

ECs can be found in every major organ in the body, and thorough EndMT ECs can become MCs that are capable of differentiating into pericytes, smooth muscle cells, skeletal muscle cells, cardiomyocytes, myofibroblasts, chondrocytes, osteocytes, adipocytes, hematopoietic stem cells, and other organ-specific cell types. Therefore, EndMT has vast potential

for tissue engineering and regenerative medicine (Medici, 2016; Man et al., 2019). Currently, EndMT is harnessed to manage ECM production and remodeling during cardiovascular tissue graft engineering (Muylaert et al., 2015).

Beyond a Synchronous BN

Despite the valuable insights provided by a Boolean model into the molecular mechanisms behind EndMT, it is evident that the complexity of the biological systems requires the incorporation of several characteristics. These constitute a set of improvements that will be incorporated into future versions of the model. The first improvement would be to convert the model into a continuous dynamical system, which will allow us to explore the biological relevance of the cyclic attractors reached by model, thus eliminating possible methodological artifacts caused by the synchronous update. Specifically, it is possible that some cycles found in the Boolean models might correspond to fixed point attractors with intermediate values when modeled as a continuous system. Furthermore, another important improvement would be the explicit modeling of the three-dimensional shape of the cells by specifying the cytoskeleton and cellular matrix. This information would allow the analysis of those signals that trigger EC cytoskeleton and ECM remodeling. This characteristic is important to understand the mechanism by which the shear stress caused by blood flow causes actin fibers within an EC to align with the flow (Kroon et al., 2017).

Conclusion

We found sufficient information obtained from published experimental results to assemble a functional model of the molecular regulatory network involved in EndMT regulation. Therefore, everything indicates that sufficient main signaling pathways that regulate EndMT are already characterized. The next logical step is to unravel the operation of the molecular regulatory network involved in EndMT control at a systemic level. The model we describe in the present manuscript constitutes an initial qualitative analysis in that direction. EndMT is required for heart valve formation during embryonic development and is an important component in the pathophysiology of cardiovascular and fibrotic diseases. Understanding how to regulate EndMT has vast applications in the treatment of disease and regenerative medicine. The simulated dynamic behavior of our model recovers fixed and cyclic patterns of molecular activation that correspond to the main cell types and behaviors involved in EndMT. Further, the simulated effect of most single gain and loss-of-function mutations of the molecules included in our model corresponds to the experimentally observed effect of the same mutations. Additionally, we used all possible perturbation patterns for 10 molecules to explore the conditions that cause EC activation, EndMT, and the reverse transitions. Based on the results of the perturbation analysis, we infer that the Phalanx and nonmesenchymal Stalk EC behaviors can only be reached from a few initial EC

behaviors, and also that the Tip EC behavior prevents direct full EndMT. Tip ECs may undergo indirect full EndMT only by previously transforming into nonphalanx, nonstalk, and nontip ECs or into mesenchymal stalk cells. Therefore, our model constitutes a theoretical framework that enables hypothesis generation, and illuminates and restricts the possible paths for future experimental EndMT research and the pharmacological control of EndMT.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the GitHub <https://github.com/NathanWeinstein/EndMT>.

AUTHOR CONTRIBUTIONS

NW, LM, and EÁ-B planned the research, wrote the article, and analyzed and discussed the results. NW reviewed the literature, composed the model, wrote the update rule, wrote the required scripts and made the tables and figures. NW and LM carried out the simulations. LM and EÁ-B obtained funding for this project.

REFERENCES

- Abaci, H. E., Coffman, A., Doucet, Y., Chen, J., Jacków, J., Wang, E., et al. (2018). Tissue engineering of human hair follicles using a biomimetic developmental approach. *Nat. Commun.* 9, 5301. doi: 10.1038/s41467-018-07579-y
- Abedin, M. J., Nguyen, A., Jiang, N., Perry, C. E., Shelton, J. M., Watson, D. K., et al. (2014). Fli1 acts downstream of ETV2 to govern cell survival and vascular homeostasis via positive autoregulation. *Circ. Res.* 114 (11), 1690–1699. doi: 10.1161/CIRCRESAHA.1134303145
- Álvarez-Buylla Rocas, M. E., Martínez-García, J. C., Dávila-Velderrain, J., Domínguez-Hüttinger, E., and Martínez-Sánchez, M. E. (2018). *Modeling Procedures* (Cham: Springer International Publishing), 35–134. doi: 10.1007/978-3-319-89354-9_2
- Aspalter, I. M., Gordon, E., Dubrac, A., Ragab, A., Narloch, J., Vizán, P., et al. (2015). Alk1 and Alk5 inhibition by Nrp1 controls vascular sprouting downstream of Notch. *Nat. Commun.* 6, 7264. doi: 10.1038/ncomms8264
- Azpeitia, E., Muñoz, S., González-Tokman, D., Martínez-Sánchez, M. E., Weinstein, N., Naldi, A., et al. (2017). The combination of the functionalities of feedback circuits is determinant for the attractors number and size in pathway-like boolean networks. *Sci. Rep.* 7, 42023. doi: 10.1038/srep42023
- Baldea, I., Teacoe, I., Olteanu, D. E., Vaida-Voievod, C., Clichici, A., Sirbu, A., et al. (2018). Effects of different hypoxia degrees on endothelial cell cultures-time course study. *Mech. Ageing Dev.* 172, 45–50.
- Bazzoni, G., and Dejana, E. (2004). Endothelial cell-to-cell junctions: molecular organization and role in vascular homeostasis. *Physiol. Rev.* 84, 869–901. doi: 10.1152/physrev.00035.2003
- Betz, C., Lenard, A., Belting, H.-G., and Affolter, M. (2016). Cell behaviors and dynamics during angiogenesis. *Development* 143, 2249–2260. doi: 10.1242/dev.135616
- Biere, A. (2008). Picosat essentials. *J. Satisfiability Boolean Modeling Comput.* 4, 75–97.
- Birbrair, A., Zhang, T., Wang, Z.-M., Messi, M. L., Mintz, A., and Delbono, O. (2015). Pericytes at the intersection between tissue regeneration and pathology. *Clin. Sci.* 128, 81–93. doi: 10.1089/scd.2017.0044
- Birbrair, A., Borges, I., d., T., Gilson Sena, I. F., Almeida, G. G., da Silva Meirelles, L., et al. (2017). How plastic are pericytes? *Stem Cells Dev.* 26, 1013–1019. doi: 10.1042/CS20140278

FUNDING

This work was funded by the UNAM-DGAPA postdoctoral scholarship to NW; grant UNAM-DGAPA-PAPIIT IN200918 to LM; and grants CONACYT CB 2014-240180-B, CONACYT PN 2015-01-687, and UNAM-DGAPA-PAPIIT IN208517 to EÁ-B.

ACKNOWLEDGMENTS

We would like to thank Eugenio Martín Azpeitia, Elisa Domínguez Hüttinger, Stalin Muñoz, and Juan Carlos Martínez García for their valuable comments. We would also like to thank Tracy Kang for revising the use of English in the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00040/full#supplementary-material>

- Bischoff, J. (2019). Endothelial-to-mesenchymal transition: Purposeful versus maladaptive differentiation. *Circ. Res.* 124, 1163–1165. doi: 10.1161/CIRCRESAHA.119.314813
- Blancas, A. A., Wong, L. E., Glaser, D. E., and McCloskey, K. E. (2012). Specialized tip/stalk-like and phalanx-like endothelial cells from embryonic stem cells. *Stem Cells Dev.* 22, 1398–1407. doi: 10.1089/scd.2012.0376
- Brinkmann, B. F., Steinbacher, T., Hartmann, C., Kummer, D., Pajonczyk, D., Mirzapourshafiyi, F., et al. (2016). VE-cadherin interacts with cell polarity protein pals1 to regulate vascular lumen formation. *Mol. Biol. Cell* 27, 2811–2821. doi: 10.1091/mbc.E16-02-0127
- Cai, X., Zhou, X., Xiao, F., Ye, B., Huang, W., and Huang, Z. (2018). Inhibition of hsa-mir-6086 protects human umbilical vein endothelial cells against TNF α -induced proliferation inhibition and apoptosis via CDH5. *Gene* 661, 202–208. doi: 10.1016/j.gene.2018.03.091
- Casas, E., Kim, J., Bendesky, A., Ohno-Machado, L., Wolfe, C. J., and Yang, J. (2011). Snail2 is an essential mediator of Twist1-induced epithelial mesenchymal transition and metastasis. *Cancer Res.* 71, 245–254. doi: 10.1158/0008-5472.CAN-10-2330
- Chae, W.-J., and Bothwell, A. L. (2018). Canonical and non-canonical wnt signaling in immune cells. *Trends Immunol.* 39, 830–847. doi: 10.1016/j.it.2018.08.006
- Chang, C.-P., Neilson, J. R., Bayle, J. H., Gestwicki, J. E., Kuo, A., Stankunas, K., et al. (2004). A field of myocardial-endocardial nfat signaling underlies heart valve morphogenesis. *Cell* 118, 649–663. doi: 10.1016/j.cell.2004.08.010
- Chang, A. C., Fu, Y., Garside, V. C., Niessen, K., Chang, L., Fuller, M., et al. (2011). Notch initiates the endothelial-to-mesenchymal transition in the atrioventricular canal through autocrine activation of soluble guanylyl cyclase. *Dev. Cell* 21, 288–300. doi: 10.1016/j.devcel.2011.06.022
- Chen, Y., and Gridley, T. (2013a). Compensatory regulation of the Snai1 and Snai2 genes during chondrogenesis. *J. Bone Miner. Res.* 28, 1412–1421. doi: 10.1002/jbmr.1871
- Chen, Y., and Gridley, T. (2013b). The SNAI1 and SNAI2 proteins occupy their own and each other's promoter during chondrogenesis. *Biochem. Biophys. Res. Commun.* 435, 356–360. doi: 10.1016/j.bbrc.2013.04.086
- Chen, S.-H., Murphy, D., Lassoued, W., Thurston, G., Feldman, M. D., and Lee, W. M. (2008). Activated STAT3 is a mediator and biomarker of VEGF endothelial activation. *Cancer Biol. Ther.* 7 (12), 1994–2003. doi: 10.4161/cbt.7.12.6967

- Chen, P.-Y., Qin, L., Barnes, C., Charisse, K., Yi, T., Zhang, X., et al. (2012). FGF regulates TGF- β signaling and endothelial-to-mesenchymal transition *via* control of let-7 mirna expression. *Cell Rep.* 2, 1684–1696. doi: 10.1016/j.celrep.2012.10.021
- Chen, Q., Zhang, H., Liu, Y., Adams, S., Eilken, H., Stehling, M., et al. (2016). Endothelial cells are progenitors of cardiac pericytes and vascular smooth muscle cells. *Nat. Commun.* 7, 12422. doi: 10.1038/ncomms12422
- Chen, J., Fu, Y., Day, D. S., Sun, Y., Wang, S., Liang, X., et al. (2017). Vegf amplifies transcription through ets1 acetylation to enable angiogenesis. *Nat. Commun.* 8, 383. doi: 10.1038/s41467-017-00405-x
- Cheng, J.-C., Chang, H.-M., and Leung, P. C. (2013). TGF- β 1 inhibits trophoblast cell invasion by inducing Snail-mediated down-regulation of VE-cadherin. *J. Biol. Chem.* 288, 33181–33192. doi: 10.1074/jbc.M113.488866
- Chi, J.-T., Chang, H. Y., Haraldsen, G., Jahnsen, F. L., Troyanskaya, O. G., Chang, D. S., et al. (2003). Endothelial cell diversity revealed by global expression profiling. *Proc. Natl. Acad. Sci.* 100, 10623–10628. doi: 10.1073/pnas.1434429100
- Coma, S., Allard-Ratick, M., Akino, T., van Meeteren, L. A., Mammoto, A., and Klagsbrun, M. (2013). Gata2 and lmo2 control angiogenesis and lymphangiogenesis *via* direct transcriptional regulation of neuropilin-2. *Angiogenesis* 16, 939–952. doi: 10.1007/s10456-013-9370-9
- Conacci-Sorrell, M., Simcha, I., Ben-Yedidia, T., Blechman, J., Savagner, P., and Ben-Ze'ev, A. (2003). Autoregulation of E-cadherin expression by cadherin-cadherin interactions: the roles of β -catenin signaling, slug, and mapk. *J. Cell Biol.* 163, 847–857. doi: 10.1083/jcb.200308162
- Conway, D. E., Breckenridge, M. T., Hinde, E., Gratton, E., Chen, C. S., and Schwartz, M. A. (2013). Fluid shear stress on endothelial cells modulates mechanical tension across ve-cadherin and pecam-1. *Curr. Biol.* 23, 1024–1030. doi: 10.1016/j.cub.2013.04.049
- Cooley, B. C., Nevado, J., Mellad, J., Yang, D., Hilaire, C. S., Negro, A., et al. (2014). Tgf- β signaling mediates endothelial-to-mesenchymal transition (EndMT) during vein graft remodeling. *Sci. Trans. Med.* 6, 227ra34–227ra34. doi: 10.1126/scitranslmed.3006927
- Correia, A. C., Moonen, J.-R. A., Brinker, M. G., and Krenning, G. (2016). FGF2 inhibits endothelial–mesenchymal transition through microRNA-20a-mediated repression of canonical tgf- β signaling. *J. Cell Sci.* 129, 569–579. doi: 10.1242/jcs.176248
- Craig, M. P., and Sumanas, S. (2016). Ets transcription factors in embryonic vascular development. *Angiogenesis* 19, 275–285. doi: 10.1007/s10456-016-9511-z
- DaSilva-Arnold, S. C., Kuo, C.-Y., Davra, V., Remache, Y., Kim, P. C., Fisher, J. P., et al. (2018). Zeb2, a master regulator of the epithelial–mesenchymal transition, mediates trophoblast differentiation. *Mol. Hum. Reprod.* 25, 61–75. doi: 10.1093/molehr/gay053
- Dejana, E., Hirschi, K. K., and Simons, M. (2017). The molecular basis of endothelial cell plasticity. *Nat. Commun.* 8, 14361. doi: 10.1038/ncomms14361
- del Toro, R., Prahst, C., Mathivet, T., Siegfried, G., Kaminker, J. S., Larrivee, B., et al. (2010). Identification and functional analysis of endothelial tip cell–enriched genes. *Blood* 116, 4025–4033. doi: 10.1182/986 blood-2010-02-270819
- Dubrova, E., and Teslenko, M. (2011). A sat-based algorithm for finding attractors in synchronous boolean networks. *IEEE/ACM Trans. Comput. Biol. Bioinf.* 8, 1393–1399. doi: 10.1109/TCBB.2010.20
- Ebnet, K., Kummer, D., Steinbacher, T., Singh, A., Nakayama, M., and Matis, M. (2018). “Regulation of cell polarity by cell adhesion receptors,” in *Seminars in cell & developmental biology*, vol. 81. (Elsevier), 2–12. doi: 10.1016/j.semcdb.2017.07.032
- Eming, S. A., Martin, P., and Tomic-Canic, M. (2014). Wound repair and regeneration: mechanisms, signaling, and translation. *Sci. Trans. Med.* 6, 265sr6–265sr6. doi: 10.1126/scitranslmed.995 3009337
- Epifanova, E., Babaev, A., Newman, A. G., and Tarabykin, V. (2018). Role of zeb2/sip1 in neuronal development. *Brain Res.* 1705, 24–31. doi: 10.1016/j.brainres.2018.09.034
- Félix, M.-A., and Barkoulas, M. (2015). Pervasive robustness in biological systems. *Nat. Rev. Genet.* 16, 483. doi: 10.1038/nrg3949
- Farber, H. W., and Loscalzo, J. (2004). Pulmonary arterial hypertension. *New Engl. J. Med.* 351, 1655–1665. doi: 10.1056/NEJMra035488
- Fischer, A., and Braga, V. M. (2018). Vascular permeability: Flow-mediated, non-canonical notch signalling promotes barrier integrity. *Curr. Biol.* 28, R119–R121. doi: 10.1016/j.cub.2017.11.065
- Forghanifard, M. M., Khales, S. A., Farshchian, M., Rad, A., Homayouni-Tabrizi, M., and Abbaszadegan, M. R. (2017). Negative regulatory role of TWIST1 on SNAIL gene expression. *Pathol. Oncol. Res.* 23, 85–90. doi: 10.1007/s12253-016-0093-2
- Forsythe, J. A., Jiang, B.-H., Iyer, N. V., Agani, F., Leung, S. W., Koos, R. D., et al. (1996). Activation of vascular endothelial growth factor gene transcription by hypoxia-inducible factor 1. *Mol. Cell. Biol.* 16, 4604–4613. doi: 10.1128/MCB.16.9.4604
- Frias, A., Lambies, G., Viñas-Castells, R., Martínez-Guillamon, C., Dave, N., de Herreros, A. G., et al. (2016). A switch in akt isoforms is required for notch-induced snail1 expression and protection from cell death. *Mol. Cell. Biol.* 36 (6), 923–940. doi: 10.1128/MCB.01074-15
- Funahashi, Y., Shawber, C. J., Vorontchikhina, M., Sharma, A., Outtz, H. H., and Kitajewski, J. (2010). Notch regulates the angiogenic response *via* induction of VEGFR-1. *J. Angiogenesis. Res.* 2, 3. doi: 10.1186/2040-2384-2-3
- Gaengel, K., Genové, G., Armulik, A., and Betsholtz, C. (2009). Endothelial-mural cell signaling in vascular development and angiogenesis. *Arterioscler. Thromb. Vasc. Biol.* 29, 630–638. doi: 10.1161/ATVBAHA.107.161521
- Garrett-Sinha, L. A., Kearly, A., and Satterthwaite, A. B. (2016). The role of the transcription factor ets1 in lupus and other autoimmune diseases. *Crit. Rev. Immunol.* 36 (6), 485–910. doi: 10.1615/CritRevImmunol.2017020284
- Gebala, V., Collins, R., Geudens, I., Phng, L.-K., and Gerhardt, H. (2016). Blood flow drives lumen formation by inverse membrane blebbing during angiogenesis *in vivo*. *Nat. Cell Biol.* 18, 443. doi: 10.1038/ncb3320
- Giannotta, M., Trani, M., and Dejana, E. (2013). Ve-cadherin and endothelial adherens junctions: active guardians of vascular integrity. *Dev. Cell* 26, 441–454. doi: 10.1016/j.devcel.2013.08.020
- Gong, H., Lyu, X., Wang, Q., Hu, M., and Zhang, X. (2017). Endothelial to mesenchymal transition in the cardiovascular system. *Life Sci.* 184, 95–102. doi: 10.1016/j.lfs.2017.07.014
- González-Mariscal, L., Tapia, R., and Chamorro, D. (2008). Crosstalk of tight junction components with signaling pathways. *Biochim. Biophys. Acta* 1778, 729–756. doi: 10.1016/j.bbamem.2007.08.018
- Good, R. B., Gilbane, A. J., Trinder, S. L., Denton, C. P., Coghlan, G., Abraham, D. J., et al. (2015). Endothelial to mesenchymal transition contributes to endothelial dysfunction in pulmonary arterial hypertension. *Am. J. Pathol.* 185, 1850–1858. doi: 10.1016/j.ajpath.2015.03.019
- Goossens, S., Wang, J., Tremblay, C., De Medts, J., T'Sas, S., Nguyen, T., et al. (2019). Zeb2 and lmo2 drive immature t-cell lymphoblastic leukemia *via* distinct oncogenic mechanisms. *Haematologica*. 104 (8), 1608–1616 doi: 10.3324/haematol.2018.207837
- Grabitz, A. L., and Duncan, M. K. (2012). Focus on molecules: Smad interacting protein 1 (Sip1, ZEB2, ZFH1B). *Exp. Eye Res.* 101, 105. doi: 10.1016/j.exer.2010.09.010
- Guaita, S., Puig, I., Francí, C., Garrido, M., Domínguez, D., Batlle, E., et al. (2002). Snail induction of epithelial to mesenchymal transition in tumor cells is accompanied by mucl repression and zeb1 expression. *J. Biol. Chem.* 277, 39209–39216. doi: 10.1074/jbc.M206400200
- Harper, S. J., and Bates, D. O. (2008). VEGF-A splicing: the key to anti-angiogenic therapeutics? *Nat. Rev. Cancer* 8, 880. doi: 10.1038/nrc2505
- Heck, T., Vaeyens, M., and Van Oosterwyck, H. (2015). Computational models of sprouting angiogenesis and cell migration: towards multiscale mechanochemical models of angiogenesis. *Math. Model. Nat. Phenoms.* 10, 108–141. doi: 10.1051/mmnp/201510106
- Herbert, S. P., and Stainier, D. Y. (2011). Molecular control of endothelial cell behaviour during blood vessel morphogenesis. *Nat. Rev. Mol. Cell Biol.* 12, 551. doi: 10.1038/nrm3176
- Ho, Y.-T., Shimbo, T., Wijaya, E., Ouchi, Y., Takaki, E., Yamamoto, R., et al. (2018). Chromatin accessibility identifies diversity in mesenchymal stem cells from different tissue origins. *Sci. Rep.* 8 (17765), 1–11. doi: 10.1038/s41598-018-36057-0
- Hollenhorst, P. C. (2012). RAS/ERK pathway transcriptional regulation through ETS/AP-1 binding sites. *Small GTPases* 3, 154–158. doi: 10.4161/srgtp.19630
- Hopper, R. K., Moonen, J.-R. A., Diebold, I., Cao, A., Rhodes, C. J., Tojais, N. F., et al. (2016). In pulmonary arterial hypertension, reduced BMPR2 promotes endothelial-to-mesenchymal transition *via* HMGAI and its target slug. *Circulation* 133, 1783–1794. doi: 10.1161/CIRCULATIONAHA.115.020617
- Howe, L. R., Watanabe, O., Leonard, J., and Brown, A. M. (2003). Twist is up-regulated in response to wnt1 and inhibits mouse mammary cell differentiation. *Cancer Res.* 63, 1906–1913.
- Ichise, T., Yoshida, N., and Ichise, H. (2014). FGF2-induced Ras–MAPK signalling maintains lymphatic endothelial cell identity by upregulating endothelial-cell-

- specific gene expression and suppressing tgfb β signalling through smad2. *J. Cell Sci.* 127, 845–857. doi: 10.1242/jcs.137836
- Ishida, W., Hamamoto, T., Kusanagi, K., Yagi, K., Kawabata, M., Takehara, K., et al. (2000). Smad6 is a Smad1/5-induced smad inhibitor characterization of bone morphogenetic protein-responsive element in the mouse Smad6 promoter. *J. Biol. Chem.* 275, 6075–6079. doi: 10.1074/jbc.275.9.6075
- Ishitani, T., Matsumoto, K., Chitnis, A. B., and Itoh, M. (2005). Nrarp functions to modulate neural-crest-cell differentiation by regulating left1 protein stability. *Nat. Cell Biol.* 7, 1106. doi: 10.1038/ncb1311
- Izawa, Y., Gu, Y.-H., Osada, T., Kanazawa, M., Hawkins, B. T., Koziol, J. A., et al. (2018). β 1-integrin–matrix interactions modulate cerebral microvessel endothelial cell tight junction expression and permeability. *J. Cereb. Blood Flow Metab.* 38, 641–658. doi: 10.1177/0271678X17722108
- Jackson, A. O., Zhang, J., Jiang, Z., and Yin, K. (2017). Endothelial-to-mesenchymal transition: A novel therapeutic target for cardiovascular diseases. *Trends Cardiovasc. Med.* 27, 383–393. doi: 10.1016/j.tcm.2017.03.003
- Jakobsson, L., Franco, C. A., Bentley, K., Collins, R. T., Ponsioen, B., Aspalter, I. M., et al. (2010). Endothelial cells dynamically compete for the tip cell position during angiogenic sprouting. *Nat. Cell Biol.* 12, 943. doi: 10.1038/ncb2103
- Jianxin, A. Y., Castranova, D., Pham, V. N., and Weinstein, B. M. (2015). Single cell analysis of endothelial morphogenesis *in vivo*. *Development* 142, (17), 2951–2961. doi: 10.1242/dev.123174
- Julien, S., Puig, I., Caretti, E., Bonaventure, J., Nelles, L., Van Roy, F., et al. (2007). Activation of NF- κ B by Akt upregulates Snail expression and induces epithelium mesenchyme transition. *Oncogene* 26, 7445.
- Kamata, M., Okitsu, Y., Fujiwara, T., Kanehira, M., Nakajima, S., Takahashi, T., et al. (2014). GATA2 regulates differentiation of bone marrow-derived mesenchymal stem cells. *Haematologica* 99, 1686–1696. doi: 10.3324/haematol.2014.105692
- Kaneko, K., Li, X., Zhang, X., Lamberti, J. J., Jamieson, S. W., and Thistlethwaite, P. A. (2008). Endothelial expression of bone morphogenetic protein receptor type 1a is required for atrioventricular valve formation. *Ann. Thorac. Surg.* 85, 2090–2098. doi: 10.1016/j.athoracsurg.2008.02.027
- Kanki, Y., Kohro, T., Jiang, S., Tsutsumi, S., Mimura, I., Suehiro, J.-I., et al. (2011). Epigenetically coordinated GATA2 binding is necessary for endothelium-specific endomucin expression. *EMBO J.* 30, 2582–2595. doi: 10.1038/emboj.2011.173
- Katoh, M., and Katoh, M. (2009). Integrative genomic analyses of ZEB2: Transcriptional regulation of ZEB2 based on SMADs, ETS1, HIF1 α , POU/OCT, and NF- κ B. *Int. J. Oncol.* 34, 1737–1742. doi: 10.3892/ijo_00000304
- Kim, D. J., Norden, P. R., Salvador, J., Barry, D. M., Bowers, S. L., Cleaver, O., et al. (2017). Src-and fyn-dependent apical membrane trafficking events control endothelial lumen formation during vascular tube morphogenesis. *PLoS One* 12, e0184461. doi: 10.1371/journal.pone.0184461
- Kobayashi, W., and Ozawa, M. (2018). The epithelial-mesenchymal transition induced by transcription factor LEF-1 is independent of β -catenin. *Biochem. Biophys. Rep.* 15, 13–18. doi: 10.1016/j.bbrep.2018.06.003
- Kokudo, T., Suzuki, Y., Yoshimatsu, Y., Yamazaki, T., Watabe, T., and Miyazono, K. (2008). Snail is required for tgfb β -induced endothelial-mesenchymal transition of embryonic stem cell-derived endothelial cells. *J. Cell Sci.* 121, 3317–3324. doi: 10.1242/jcs.028282
- Korn, C., and Augustin, H. G. (2015). Mechanisms of vessel pruning and regression. *Dev. Cell* 34, 5–17. doi: 10.1016/j.devcel.2015.06.004
- Kovacic, J. C., Dimmeler, S., Harvey, R. P., Finkel, T., Aikawa, E., Krenning, G., et al. (2019). Endothelial to mesenchymal transition in cardiovascular disease: JACC state-of-the-art review. *J. Am. Coll. Cardiol.* 73, 190–209. doi: 10.1016/j.jacc.2018.09.089
- Krenning, G., Barauna, V. G., Krieger, J. E., Harmsen, M. C., and Moonen, J.-R. A. (2016). Endothelial plasticity: shifting phenotypes through force feedback. *Stem Cells Int.* 2016, 1–15. doi: 10.1155/2016/9762959
- Kroon, J., Heemskerk, N., Kalsbeek, M. J., de Waard, V., van Rijssel, J., and van Buul, J. D. (2017). Flow-induced endothelial cell alignment requires the rhogef trio as a scaffold protein to polarize active rac1 distribution. *Mol. Biol. Cell* 28, 1745–1753. doi: 10.1091/mbc.e16-06-0389
- Kulkeaw, K., Inoue, T., Iino, T., Tani, K., Akashi, K., Speck, N. A., et al. (2017). Twist1 regulates embryonic hematopoietic differentiation through binding to Myb and Gata2 promoter regions. *Blood Adv.* 1, 1672–1681. doi: 10.1182/bloodadvances.2017006056
- Kumar, V. S., Binu, S., Soumya, S., Haritha, K., and Sudhakaran, P. (2014). Regulation of vascular endothelial growth factor by metabolic context of the cell. *Glycoconj. J.* 31, 427–434. doi: 10.1182/bloodadvances.2017006056
- Kumar, B., Uppuladinne, M. V., Jani, V., Sonavane, U., Joshi, R. R., and Bapat, S. A. (2015). Auto-regulation of Slug mediates its activity during epithelial to mesenchymal transition. *Biochim. Biophys. Acta* 1849, 1209–1218. doi: 10.1016/j.bbagr.2015.07.006
- Kutys, M. L., and Chen, C. S. (2016). Forces and mechanotransduction in 3d vascular biology. *Curr. Opin. Cell Biol.* 42, 73–79. doi: 10.1016/j.ccb.2016.04.011
- Laakkonen, J. P., Lappalainen, J. P., Theelen, T. L., Toivanen, P. I., Nieminen, T., Jauhainen, S., et al. (2017). Differential regulation of angiogenic cellular processes and claudin-5 by histamine and VEGF via PI3K-signaling, transcription factor SNAI2 and interleukin-8. *Angiogenesis* 20, 109–124. doi: 10.1007/s10456-016-9532-7
- Lambertini, E., Franceschetti, T., Torreggiani, E., Penolazzi, L., Pastore, A., Pelucchi, S., et al. (2010). SLUG: a new target of lymphoid enhancer factor-1 in human osteoblasts. *BMC Mol. Biol.* 11, 13. doi: 10.1186/1471-2199-11-13
- Lamouille, S., Xu, J., and Derynck, R. (2014). Molecular mechanisms of epithelial–mesenchymal transition. *Nat. Rev. Mol. Cell Biol.* 15, 178. doi: 10.1038/nrm3758
- Lelièvre, E., Lionneton, F., Mattot, V., Spruyt, N., and Soncin, F. (2002). Ets-1 regulates fli-1 expression in endothelial cells IDENTIFICATION OF ETS BINDING SITES IN THE fli-1 GENE PROMOTER. *J. Biol. Chem.* 277, 25143–25151. doi: 10.1074/jbc.M201628200
- Li, C., Hong, T., and Nie, Q. (2016). Quantifying the landscape and kinetic paths for epithelial–mesenchymal transition from a core circuit. *Phys. Chem. Chem. Phys.* 18, 17949–17956. doi: 10.1039/C6CP03174A
- Li, Z., Xu, Z., Duan, C., Liu, W., Sun, J., and Han, B. (2018). Role of tcf/lef transcription factors in bone development and osteogenesis. *Int. J. Med. Sci.* 15, 1415. doi: 10.7150/ijms.26741
- Liu, T., Ma, W., Xu, H., Huang, M., Zhang, D., He, Z., et al. (2018). PDGF-mediated mesenchymal transformation renders endothelial resistance to anti-VEGF treatment in glioblastoma. *Nat. Commun.* 9, 3439. doi: 10.1038/s41467-018-05982-z
- Lizama, C. O., and Zovein, A. C. (2013). Polarizing pathways: balancing endothelial polarity, permeability, and lumen formation. *Exp. Cell Res.* 319, 1247. doi: 10.1016/j.yexcr.2013.03.028
- Lopez, D., Niu, G., Huber, P., and Carter, W. B. (2009). Tumor-induced upregulation of Twist, Snail, and Slug represses the activity of the human VE-cadherin promoter. *Arch. Biochem. Biophys.* 482, 77–82. doi: 10.1016/j.jabb.2008.11.016
- Lu, M., Jolly, M. K., Levine, H., Onuchic, J. N., and Ben-Jacob, E. (2013). MicroRNA-based regulation of epithelial–hybrid–mesenchymal fate determination. *Proc. Natl. Acad. Sci.* 110, 18144–18149. doi: 10.1073/pnas.1318192110
- MacGrogan, D., Luxán, G., Driessen-Mol, A., Bouten, C., Baaijens, F., and de la Pompa, J. L. (2014). How to make a heart valve: from embryonic development to bioengineering of living valve substitutes. *Cold Spring Harbor Perspect. Med.* 4, a013912. doi: 10.1101/cshperspect.a013912
- Maddaluno, L., Rudini, N., Cattanò, R., Bravi, L., Giampietro, C., Corada, M., et al. (2013). Endmt contributes to the onset and progression of cerebral cavernous malformations. *Nature* 498, 492. doi: 10.1038/nature12207
- Magenta, A., Cencioni, C., Fasanaro, P., Zaccagnini, G., Greco, S., Sarra-Ferraris, G., et al. (2011). mir-200c is upregulated by oxidative stress and induces endothelial cell apoptosis and senescence via zeb1 inhibition. *Cell Death Differ.* 18, 1628. doi: 10.1038/cdd.2011.42
- Mahmoud, M. M., Kim, H. R., Xing, R., Hsiao, S., Mammoto, A., Chen, J., et al. (2016). Twist1 integrates endothelial responses to flow in vascular dysfunction and atherosclerosis. *Circ. Res.* 119, 450–462. doi: 10.1161/CIRCRESAHA.116.308870
- Mammoto, T., Muyleart, M., Konduri, G. G., and Mammoto, A. (2018). Twist1 in hypoxia-induced pulmonary hypertension through transforming growth factor- β –Smad signaling. *Am. J. Respir. Cell Mol. Biol.* 58, 194–207. doi: 10.1165/rcmb.2016-0323OC
- Man, S., Duffhues, G. S., ten Dijke, P., and Baker, D. (2019). The therapeutic potential of targeting the endothelial-to-mesenchymal transition. *Angiogenesis* 22, 3–13. doi: 10.1007/s10456-018-9639-0

- Matkar, P. N., Singh, K. K., Rudenko, D., Kim, Y. J., Kuliszewski, M. A., Prud'homme, G. J., et al. (2016). Novel regulatory role of neuropilin-1 in endothelial-to-mesenchymal transition and fibrosis in pancreatic ductal adenocarcinoma. *Oncotarget* 7, 69489. doi: 10.18632/oncotarget.11060
- Medici, D., Hay, E. D., and Goodenough, D. A. (2006). Cooperation between snail and LEF-1 transcription factors is essential for TGF- β 1-induced epithelial-mesenchymal transition. *Mol. Biol. Cell* 17, 1871–1879. doi: 10.1091/mbc.E05-08-0767
- Medici, D., Potenta, S., and Kalluri, R. (2011). Transforming growth factor- β 2 promotes Snail-mediated endothelial-mesenchymal transition through convergence of Smad-dependent and Smad-independent signalling. *Biochem. J.* 437, 515–520. doi: 10.1042/BJ20101500
- Medici, D. (2016). Endothelial-mesenchymal transition in regenerative medicine. *Stem Cells Int.* 2016, 1–7. doi: 10.1155/2016/6962801
- Méndez-López, L. F., Davila-Velderrain, J., Domínguez-Hüttinger, E., Enríquez-Olguín, C., Martínez-García, J. C., and Alvarez-Buylla, E. R. (2017). Gene regulatory network underlying the immortalization of epithelial cells. *BMC Syst. Biol.* 11, 24. doi: 10.1186/s12918-017-0393-5
- Menezes, M. E. (2014). The wnt/ β -catenin signaling pathway in epithelial mesenchymal transition. *J. Postdoctoral Res.* 1, 12.
- Moonen, J.-R. A., Lee, E. S., Schmidt, M., Maleszewska, M., Koerts, J. A., Brouwer, L. A., et al. (2015). Endothelial-to-mesenchymal transition contributes to fibroproliferative vascular disease and is modulated by fluid shear stress. *Cardiovasc. Res.* 108, 377–386. doi: 10.1093/cvr/cvv175
- Moorman, A., Webb, S., Brown, N. A., Lamers, W., and Anderson, R. H. (2003). Development of the heart:(1) formation of the cardiac chambers and arterial trunks. *Heart* 89, 806–814. doi: 10.1136/heart.89.7.806
- Mouillessaux, K. P., Wiley, D. S., Saunders, L. M., Wylie, L. A., Kushner, E. J., Chong, D. C., et al. (2016). Notch regulates bmp responsiveness and lateral branching in vessel networks via smad6. *Nat. Commun.* 7, 13247. doi: 10.1038/ncomms13247
- Murakami, M., Nguyen, L. T., Hatanaka, K., Schachterle, W., Chen, P.-Y., Zhuang, Z. W., et al. (2011). FGF-dependent regulation of VEGF receptor 2 expression in mice. *J. Clin. Invest.* 121, 2668–2678. doi: 10.1172/JCI44762
- Müssel, C., Hopfensitz, M., and Kestler, H. A. (2010). Boolnet—an R package for generation, reconstruction and analysis of boolean networks. *Bioinformatics* 26, 1378–1380. doi: 10.1093/bioinformatics/btq124
- Muyllaert, D. E., de Jong, O. G., Slaats, G. G., Nieuweboer, F. E., Fledderus, J. O., Goumans, M.-J., et al. (2015). Environmental influences on endothelial to mesenchymal transition in developing implanted cardiovascular tissue-engineered grafts. *Tissue Eng. Part B: Rev.* 22, 58–67. doi: 10.1089/ten.teb.2015.0167
- Nagai, N., Ohguchi, H., Nakaki, R., Matsumura, Y., Kanki, Y., Sakai, J., et al. (2018). Downregulation of ERG and FLI1 expression in endothelial cells triggers endothelial-to-mesenchymal transition. *PLoS Genet.* 14, e1007826. doi: 10.1371/journal.pgen.1007826
- Nakagomi, T., Kubo, S., Nakano-Doi, A., Sakuma, R., Lu, S., Narita, A., et al. (2015). Brain vascular pericytes following ischemia have multipotential stem cell activity to differentiate into neural and vascular lineage cells. *Stem Cells* 33, 1962–1974. doi: 10.1002/stem.1977
- Naldi, A., Berenguier, D., Fauré, A., Lopez, F., Thieffry, D., and Chaouiya, C. (2009). Logical modelling of regulatory networks with ginsim 2.3. *Biosystems* 97, 134–139. doi: 10.1016/j.biosystems.2009.04.008
- Nguyen, P., Tsunematsu, T., Yanagisawa, S., Kudo, Y., Miyauchi, M., Kamata, N., et al. (2013). The fgfr1 inhibitor pd173074 induces mesenchymal-epithelial transition through the transcription factor ap-1. *Br. J. Cancer* 109, 2248. doi: 10.1038/bjc.2013.550
- Niessen, K., Fu, Y., Chang, L., Hoodless, P. A., McFadden, D., and Karsan, A. (2008). Slug is a direct notch target required for initiation of cardiac cushion cellularization. *J. Cell Biol.* 182, 315–325. doi: 10.1083/jcb.200710067
- Nishimoto, S., and Nishida, E. (2006). Mapk signalling: erk5 versus erk1/2. *EMBO Rep.* 7, 782–786. doi: 10.1038/sj.embor.7400755
- Nishita, M., Itsukushima, S., Nomachi, A., Endo, M., Wang, Z., Inaba, D., et al. (2010). Ror2/Frizzled complex mediates Wnt5a-induced AP-1 activation by regulating dishevelled polymerization. *Mol. Cell. Biol.* 30, 3610–3619. doi: 10.1128/MCB.00177-10
- Norden, P. R., Kim, D. J., Barry, D. M., Cleaver, O. B., and Davis, G. E. (2016). Cdc42 and k-ras control endothelial tubulogenesis through apical membrane and cytoskeletal polarization: novel stimulatory roles for gtpase effectors, the small gtpases, rac2 and rap1b, and inhibitory influence of arhgap31 and rasal. *PLoS One* 11, e0147758. doi: 10.1371/journal.pone.0147758
- Oh, H., Takagi, H., Otani, A., Koyama, S., Kemmochi, S., Uemura, A., et al. (2002). Selective induction of neuropilin-1 by vascular endothelial growth factor (vegfr): a mechanism contributing to vegf-induced angiogenesis. *Proc. Natl. Acad. Sci.* 99, 383–388. doi: 10.1073/pnas.012074399
- Oh, S.-Y., Kim, J. Y., and Park, C. (2015). The ETS factor, ETV2: a master regulator for vascular endothelial cell development. *Mol. Cells* 38, 1029. doi: 10.14348/molcells.2015.0331
- Pahnke, A., Conant, G., Huyer, L. D., Zhao, Y., Feric, N., and Radisic, M. (2016). The role of wnt regulation in heart development, cardiac repair and disease: A tissue engineering perspective. *Biochem. Biophys. Res. Commun.* 473, 698–703. doi: 10.1016/j.bbrc.2015.11.060
- Papadopoulos, N., and Lennartsson, J. (2018). The pdgfr/pdgfr pathway as a drug target. *Mol. Aspects Med.* 62, 75–88. doi: 10.1016/j.mam.2017.11.007
- Pardali, E., Sanchez-Duffhues, G., Gomez-Puerto, M. C., and Ten Dijke, P. (2017). Tgf- β -induced endothelial-mesenchymal transition in fibrotic diseases. *Int. J. Mol. Sci.* 18, 2157. doi: 10.3390/ijms18102157
- Park, C., Lee, T.-J., Bhang, S. H., Liu, F., Nakamura, R., Oladipupo, S. S., et al. (2016). Injury-mediated vascular regeneration requires endothelial ER71/ETV2. *Arterioscler. Thromb. Vasc. Biol.* 36, 86–96. doi: 10.1161/ATVBAHA.115.306430
- Paruchuri, S., Yang, J.-H., Aikawa, E., Melero-Martin, J. M., Khan, Z. A., Loukogeorgakis, S., et al. (2006). Human pulmonary valve progenitor cells exhibit endothelial/mesenchymal plasticity in response to vascular endothelial growth factor- α and transforming growth factor- β 2. *Circ. Res.* 99, 861–869. doi: 10.1161/01.RES.0000245188.41002.2c
- Pavillon, N., and Smith, N. I. (2019). Immune cell type, cell activation, and single cell heterogeneity revealed by label-free optical methods. *Sci. Rep.* 9, 1–13. doi: 10.1038/s41598-019-53428-3
- Peirce, S. M. (2008). Computational and mathematical modeling of angiogenesis. *Microcirculation* 15, 739–751. doi: 10.1080/10739680802220331
- Peiro, S., Escrivá, M., Puig, I., Barbera, M. J., Dave, N., Herranz, N., et al. (2006). Snail transcriptional repressor binds to its own promoter and controls its expression. *Nucleic Acids Res.* 34, 2077–2084. doi: 10.1093/nar/gkl141
- Phng, L.-K., Potente, M., Leslie, J. D., Babbage, J., Nyqvist, D., Lobov, I., et al. (2009). Nrarp coordinates endothelial Notch and Wnt signaling to control vessel density in angiogenesis. *Dev. Cell* 16, 70–82. doi: 10.1016/j.devcel.2008.12.009
- Piera-Velazquez, S., and Jimenez, S. A. (2019). Endothelial to mesenchymal transition: role in physiology and in the pathogenesis of human diseases. *Physiol. Rev.* 99, 1281–1324. doi: 10.1152/physrev.00021.2018
- Pimanda, J. E., Donaldson, I. J., De Bruijn, M. F., Kinston, S., Knezevic, K., Huckle, L., et al. (2007a). The SCL transcriptional network and BMP signaling pathway interact to regulate RUNX1 activity. *Proc. Natl. Acad. Sci.* 104, 840–845. doi: 10.1073/pnas.0607196104
- Pimanda, J. E., Ottersbach, K., Knezevic, K., Kinston, S., Chan, W. Y., Wilson, N. K., et al. (2007b). Gata2, Flt1, and Scl form a recursively wired gene-regulatory circuit during early hematopoietic development. *Proc. Natl. Acad. Sci.* 104, 17692–17697. doi: 10.1073/pnas.0707045104
- Pires, B. R., Mencalha, A. L., Ferreira, G. M., de Souza, W. F., Morgado-Díaz, J. A., Maia, A. M., et al. (2017). NF- κ B is involved in the regulation of EMT genes in breast cancer cells. *PLoS One* 12, e0169622. doi: 10.1371/journal.pone.0169622
- Qutub, A. A., Mac Gabhann, F., Karagiannis, E. D., Vempati, P., and Popel, A. S. (2009). Multiscale models of angiogenesis. *IEEE Eng. Med. Biol. Mag.* 28 (2), 14–31. doi: 10.1109/MEMB.2009.931791
- Radeva, M., and Waschke, J. (2018). Mind the gap: mechanisms regulating the endothelial barrier. *Acta Physiol.* 222, e12860. doi: 10.1111/apha.12860
- Rahimi, N. (2017). Defenders and challengers of endothelial barrier function. *Front. Immunol.* 8, 1847. doi: 10.3389/fimmu.2017.01847
- Romashkova, J. A., and Makarov, S. S. (1999). NF- κ B is a target of AKT in anti-apoptotic PDGF signalling. *Nature* 401, 86. doi: 10.1038/43474
- Rozum, J. C., and Albert, R. (2018). Self-sustaining positive feedback loops in discrete and continuous systems. *J. Theor. Biol.* 459, 36–44. doi: 10.1016/j.jtbi.2018.09.017
- Saitoh, M., Endo, K., Furuya, S., Minami, M., Fukasawa, A., Imamura, T., et al. (2016). Stat3 integrates cooperative ras and tgfr- β signals that induce snail expression. *Oncogene* 35, 1049. doi: 10.1038/nc.2015.161

- Sánchez-Duffhues, G., García de Vinuesa, A., and ten Dijke, P. (2018). Endothelial-to-mesenchymal transition in cardiovascular diseases: developmental signaling pathways gone awry. *Dev. Dyn.* 247, 492–508. doi: 10.1002/dvdy.24589
- Sánchez-Tilló, E., de Barrios, O., Siles, L., Cuatrecasas, M., Castells, A., and Postigo, A. (2011). β -catenin/tcf4 complex induces the epithelial-to-mesenchymal transition (emt)-activator zeb1 to regulate tumor invasiveness. *Proc. Natl. Acad. Sci.* 108, 19204–19209. doi: 10.1073/pnas.1108977108
- Sanchez-Tillo, E., Lazaro, A., Torrent, R., Cuatrecasas, M., Vaquero, E., Castells, A., et al. (2010). Zeb1 represses e-cadherin and induces an emt by recruiting the swi/snf chromatin-remodeling protein brg1. *Oncogene* 29, 3490. doi: 10.1038/ncr.2010.102
- Sanchez-Tillo, E., Fanlo, L., Siles, L., Montes-Moreno, S., Moros, A., Chiva-Blanch, G., et al. (2014). The emt activator zeb1 promotes tumor growth and determines differential response to chemotherapy in mantle cell lymphoma. *Cell Death Differ.* 21, 247. doi: 10.1038/cdd.2013.123
- Shimizu, A., Zankov, D., Kurokawa-Seo, M., and Ogita, H. (2018). Vascular endothelial growth factor- α exerts diverse cellular effects via small g proteins, rho and rap. *Int. J. Mol. Sci.* 19, 1203. doi: 10.3390/ijms19041203
- Simons, M., Gordon, E., and Claesson-Welsh, L. (2016). Mechanisms and regulation of endothelial vegf receptor signalling. *Nat. Rev. Mol. Cell Biol.* 17, 611. doi: 10.1038/nrm.2016.87
- Steinway, S. N., Zañudo, J. G., Ding, W., Rountree, C. B., Feith, D. J., Loughran, T. P., et al. (2014). Network modeling of tgfb signaling in hepatocellular carcinoma epithelial-to-mesenchymal transition reveals joint sonic hedgehog and wnt pathway activation. *Cancer Res.* 74, 5963–5977. doi: 10.1158/0008-5472.CAN-14-0225
- Sun, J.-X., Chang, T.-F., Li, M.-H., Sun, L.-J., Yan, X.-C., Yang, Z.-Y., et al. (2018). Snail, an endothelial-mesenchymal transition transcription factor, promotes the early phase of ocular neovascularization. *Angiogenesis* 21, 635–652. doi: 10.1007/s10456-018-9614-9
- Sundqvist, A., Zieba, A., Vasilaki, E., Hidalgo, C. H., Söderberg, O., Koinuma, D., et al. (2013). Specific interactions between smad proteins and AP-1 components determine TGF β -induced breast cancer cell invasion. *Oncogene* 32, 3606. doi: 10.1038/ncr.2012.370
- Suzuki, T., Minerva, D., Nishiyama, K., Koshikawa, N., and Chaplain, M. A. J. (2018). Study on the tumor-induced angiogenesis using mathematical models. *Cancer Sci.* 109, 15–23. doi: 10.1111/cas.13395
- Swinton, J. (2014). The xtable gallery.
- Szymborska, A., and Gerhardt, H. (2018). Hold me, but not too tight-endothelial cell-cell junctions in angiogenesis. *Cold Spring Harbor Perspect. Biol.* 10, a029223. doi: 10.1101/cshperspect.a029223
- Takkunen, M., Grenman, R., Hukkanen, M., Korhonen, M., Herreros, A. G. D., and Virtanen, I. (2006). Snail-dependent and-independent epithelial-mesenchymal transition in oral squamous carcinoma cells. *J. Histochem. Cytochem.* 54, 1263–1275. doi: 10.1369/jhc.6A6958.2006
- Teruyama, K., Abe, M., Nakano, T., Takahashi, S., Yamada, S., and Sato, Y. (2001). Neurophilin-1 is a downstream target of transcription factor Ets-1 in human umbilical vein endothelial cells. *FEBS Lett.* 504, 1–4. doi: 10.1016/S0014-5793(01)02724-7
- Tsang, K. M., Hyun, J. S., Cheng, K. T., Vargas, M., Mehta, D., Ushio-Fukai, M., et al. (2017). Embryonic stem cell differentiation to functional arterial endothelial cells through sequential activation of ETV2 and NOTCH1 signaling by HIF1 α . *Stem Cell Rep.* 9, 796–806. doi: 10.1016/j.stemcr.2017.07.001
- van Meeteren, L. A., and Ten Dijke, P. (2012). Regulation of endothelial cell plasticity by tgfb. *Cell Tissue Res.* 347, 177–186. doi: 10.1007/s00441-011-1222-6
- von Gise, A., and Pu, W. T. (2012). Endocardial and epicardial epithelial to mesenchymal transitions in heart development and disease. *Circ. Res.* 110, 1628–1645. doi: 10.1161/CIRCRESAHA.111.259960
- Wallez, Y., and Huber, P. (2008). Endothelial adherens and tight junctions in vascular homeostasis, inflammation and angiogenesis. *Biochim. Biophys. Acta* 1778, 794–809. doi: 10.1016/j.bbame.2007.09.003
- Wang, S., Chang, J., Hsiao, J., Yen, Y., Jiang, S., Liu, S., et al. (2017a). Tumour cell-derived wnt5b modulates *in vitro* lymphangiogenesis via induction of partial endothelial-mesenchymal transition of lymphatic endothelial cells. *Oncogene* 36, 1503. doi: 10.1038/ncr.2016.317
- Wang, X., Bin, Y., Tang, G., Zhao, Y., Huang, S., Sun, X., et al. (2017b). GATA2 promotes metastasis and infers poor outcomes for nasopharyngeal carcinoma patients. *Int. J. Clin. Exp. Med.* 10, 4819–4828.
- Weinstein, N., Ortiz-Gutiérrez, E., Muñoz, S., Rosenblueth, D. A., Álvarez-Buylla, E. R., and Mendoza, L. (2015). A model of the regulatory network involved in the control of the cell cycle and cell differentiation in the caenorhabditis elegans vulva. *BMC Bioinf.* 16, 81. doi: 10.1186/s12859-015-0498-z
- Weinstein, N., Mendoza, L., Gitler, I., and Klapp, J. (2017). A network model to explore the effect of the micro-environment on endothelial cell behavior during angiogenesis. *Front. Physiol.* 8, 960. doi: 10.3389/fphys.2017.00960
- Welch-Reardon, K. M., Ehsan, S. M., Wang, K., Wu, N., Newman, A. C., Romero-Lopez, M., et al. (2014). Angiogenic sprouting is regulated by endothelial cell expression of slug (snai2). *J. Cell Sci.* 127 (9), 2017–2028. doi: 10.1242/jcs.143420
- Welch-Reardon, K. M., Wu, N., and Hughes, C. C. (2015). A role for partial endothelial-mesenchymal transitions in angiogenesis? *Arterioscler. Thromb. Vasc. Biol.* 35, 303–308. doi: 10.1161/ATVBAHA.114.303220
- Wels, C., Joshi, S., Koefinger, P., Bergler, H., and Schaidler, H. (2011). Transcriptional activation of ZEB1 by Slug leads to cooperative regulation of the epithelial-mesenchymal transition-like phenotype in melanoma. *J. Invest. Dermatol.* 131, 1877–1885. doi: 10.1038/jid.2011.142
- Weston, S., and Calaway, R. (2019). Getting started with dparallel and foreach. Available at: <https://cran.r-project.org/web/packages/doParallel/vignettes/gettingstartedParallel.pdf> (Accessed July 30, 2019).
- Wickham, H. (2011). ggplot2. *Wiley Interdiscip. Rev. Comput. Stat.* 3, 180–185. doi: 10.1002/wics.147
- Wilson, C. W., and Ye, W. (2014). Regulation of vascular endothelial junction stability and remodeling through rap1-rasip1 signaling. *Cell Adhes. Migration* 8, 76–83. doi: 10.4161/cam.28115
- Wragg, J. W., Durant, S., McGettrick, H. M., Sample, K. M., Egginton, S., and Bicknell, R. (2014). Shear stress regulated gene expression and angiogenesis in vascular endothelium. *Microcirculation* 21, 290–300. doi: 10.1111/micc.12119
- Wu, Z.-Q., Li, X.-Y., Hu, C. Y., Ford, M., Kleer, C. G., and Weiss, S. J. (2012). Canonical Wnt signaling regulates slug activity and links epithelial-mesenchymal transition with epigenetic breast cancer 1, early onset (brca1) repression. *Proc. Natl. Acad. Sci.* 109, 16654–16659. doi: 10.1073/pnas.1205822109
- Wu, M. H. (2005). Endothelial focal adhesions and barrier function. *J. Physiol.* 569, 359–366. doi: 10.1113/jphysiol.2005.096537
- Wythe, J. D., Dang, L. T., Devine, W. P., Boudreau, E., Artap, S. T., He, D., et al. (2013). Ets factors regulate vegf-dependent arterial specification. *Dev. Cell* 26, 45–58. doi: 10.1016/j.devcel.2013.06.007
- Xu, X., Tan, X., Tampe, B., Sanchez, E., Zeisberg, M., and Zeisberg, E. M. (2015). Snail is a direct target of hypoxia-inducible factor 1 α (HIF1 α) in hypoxia-induced endothelial to mesenchymal transition of human coronary endothelial cells. *J. Biol. Chem.* 290, 16653–16664. doi: 10.1074/jbc.M115.636944
- Yalim-Camci, I., Balcik-Ercin, P., Cetin, M., Odabas, G., Tokay, N., Sayan, A. E., et al. (2019). Ets1 is coexpressed with zeb2 and mediates zeb2-induced epithelial-mesenchymal transition in human tumors. *Mol. Carcinog.* 58, 1068–1081. doi: 10.1002/mc.22994
- Yang, G., and Albert, R. (2019). *Modeling of molecular networks* (Cham: Springer International Publishing), 35–62. doi: 10.1007/978-3-030-22583-4_2
- Yang, M.-H., and Wu, K.-J. (2008). Twist activation by hypoxia inducible factor-1 (HIF-1): implications in metastasis and development. *Cell Cycle* 7, 2090–2096. doi: 10.4161/cc.7.14.6324
- Yang, X., Liaw, L., Prudovsky, I., Brooks, P. C., Vary, C., Oxburgh, L., et al. (2015). Fibroblast growth factor signaling in the vasculature. *Curr. Atheroscl. Rep.* 17, 31. doi: 10.1007/s11883-015-0509-6
- Yao, J., Wu, X., Zhang, D., Wang, L., Zhang, L., Reynolds, E. X., et al. (2019). Elevated endothelial sox2 causes lumen disruption and cerebral arteriovenous malformations. *J. Clin. Invest.* 129 (8), 3121–3133. doi: 10.1172/JCI125965
- Yu, W., Zhang, Y., Ruest, L. B., and Svoboda, K. K. (2013). Analysis of snail1 function and regulation by twist1 in palatal fusion. *Front. Physiol.* 4, 12. doi: 10.3389/fphys.2013.00012

- Yun, K., Choi, Y. D., Nam, J. H., Park, Z., and Im, S.-H. (2007). NF- κ B regulates Lef1 gene expression in chondrocytes. *Biochem. Biophys. Res. Commun.* 357, 589–595. doi: 10.1016/j.bbrc.2007.03.170
- Zhang, B., Niu, W., Dong, H.-Y., Liu, M.-L., Luo, Y., and Li, Z.-C. (2018). Hypoxia induces endothelial-mesenchymal transition in pulmonary vascular remodeling. *Int. J. Mol. Med.* 42, 270–278. doi: 10.3892/ijmm.2018.3584
- Zhao, C., Qiao, Y., Jonsson, P., Wang, J., Xu, L., Rouhi, P., et al. (2014). Genome-wide profiling of AP-1-regulated transcription provides insights into the invasiveness of triple-negative breast cancer. *Cancer Res.* 74, 3983–3994. doi: 10.1158/0008-5472.CAN-13-3396

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Weinstein, Mendoza and Álvarez-Buylla. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Searching for the Mechanisms of Mammalian Cellular Aging Through Underlying Gene Regulatory Networks

Wenbo Li¹, Lei Zhao¹ and Jin Wang^{2*}

¹ State Key Laboratory of Electroanalytical Chemistry, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun, China, ² Department of Chemistry and Physics, State University of New York at Stony Brook, Stony Brook, NY, United States

OPEN ACCESS

Edited by:

Tian Hong,
The University of Tennessee, Knoxville,
United States

Reviewed by:

Hong Qin,
University of Tennessee at
Chattanooga, United States
Tiejun Li,
Peking University, China

*Correspondence:

Jin Wang
jin.wang.1@stonybrook.edu

Specialty section:

This article was submitted to
Systems Biology,
a section of the journal
Frontiers in Genetics

Received: 03 March 2020

Accepted: 15 May 2020

Published: 30 June 2020

Citation:

Li W, Zhao L and Wang J (2020)
Searching for the Mechanisms of
Mammalian Cellular Aging Through
Underlying Gene Regulatory
Networks. *Front. Genet.* 11:593.
doi: 10.3389/fgene.2020.00593

Aging attracts the attention throughout the history of humankind. However, it is still challenging to understand how the internal driving forces, for example, the fundamental building blocks of life, such as genes and proteins, as well as the environments work together to determine longevity in mammals. In this study, we built a gene regulatory network for mammalian cellular aging based on the experimental literature and quantify its underlying driving force for the dynamics as potential and flux landscape. We found three steady-state attractors: a fast-aging state attractor, slow-aging state attractor, and intermediate state attractor. The system can switch from one state attractor to another driven by the intrinsic or external forces through the genetics and the environment. We identified the dominant path from the slow-aging state directly to the fast-aging state. We also identified the dominant path from slow-aging to fast-aging through an intermediate state. We quantified the evolving landscape for revealing the dynamic characteristics of aging through certain regulation changes in time. We also predicted the key genes and regulations for fast-aging and slow-aging through the analysis of the stability for landscape basins. We also found the oscillation dynamics between fast-aging and slow-aging and showed that more energy is required to sustain such oscillations. We found that the flux is the dynamic cause and the entropy production rate the thermodynamic origin of the phase transitions or the bifurcations between the three-state phase and oscillation phase. The landscape quantification provides a global and physical approach to explore the underlying mechanisms of cellular aging in mammals.

Keywords: aging, slow-aging, landscape, flux, entropy production, gene regulatory network

1. INTRODUCTION

The study of aging has been one of the most long-lasting and influential fields for both scientists and the public. Previous studies have shown that there are nine hallmarks of aging: genomic instability, telomere attrition, epigenetic alterations, loss of proteostasis, deregulated nutrient sensing, mitochondrial dysfunction, cellular senescence, stem cell exhaustion, and altered intercellular communication (López-Otín et al., 2013). In this work, we focus on studies of cellular aging based not only on the key genes but also, more importantly, on their associated gene

regulations. Thanks to the rapid development of molecular biology, researchers can manipulate certain genes and observe their effects on the aging process of a model organism (Gems and Partridge, 2013). An early breakthrough showed that the mutation of only one gene, *daf-2*, can prolong the lifespan of *Caenorhabditis elegans* by more than two times (Kenyon et al., 1993). Since then, hundreds of genes related to aging have been isolated, and evolutionarily conserved pathways like Insulin/IGF-1 signaling, TOR signaling, AMP kinase, and Sirtuins have been identified (Kenyon, 2010; Colman et al., 2014). Although great progress has been made in aging research over the last several decades, there is still a lack of a physical model to integrate these experimental observations, to quantitatively understand the mechanisms of how the internal and external elements (such as environments) work together to control aging, and to predict the key genes and regulations that significantly affect the aging process.

The landscape paradigm for development was introduced by Waddington in the 1940s (Waddington, 2014). However, the Waddington landscape initially only provided a qualitative picture and lacked physical foundation and quantification (Wang, 2015). Recently, there has been significant progress in establishing the physical theory and foundation as well as the quantification of the Waddington landscape (Wang et al., 2008; Wang J. et al., 2011; Wang, 2015; Zhou and Li, 2016). A detailed comparison and critical review of various approaches was presented in (Zhou and Li, 2016). To find the core mechanisms of the mammalian cellular aging process, we built a gene regulatory network based on the existing experimental literature (Haruta et al., 2000; Stambolic et al., 2001; Inoki et al., 2002, 2003; Ogawara et al., 2002; Kong, 2004; Lahav et al., 2004; Nemoto, 2004; You et al., 2006; Greer et al., 2007b; Okoshi et al., 2007; Budanov and Karin, 2008; Gwinn et al., 2008; Lan et al., 2008; Salih and Brunet, 2008; Cantó et al., 2009; Chen et al., 2010; Georgescu, 2010; Ghosh et al., 2010; Sengupta et al., 2010; Yi and Luo, 2010; Budanov, 2011; Dunlop et al., 2011; Gao et al., 2011; Kim et al., 2011; Löffler et al., 2011; Renault et al., 2011; Wang F. et al., 2011; Parmigiani et al., 2014). We quantified the potential landscape through analyzing the long-term dynamic trajectories. We identified the driving forces of aging dynamics as the steady-state probability landscape and the steady-state probability flux. While the landscape tends to stabilize the states of the system, the flux tends to stabilize the flow of state. The quantification of the landscape and the flux provides us with a global way to understand the functions and stabilities of and also the relationships among different functional states. Furthermore, one can detect what key elements can lead to significant changes on the system stabilities and quantify these by the landscape topography through barrier heights and switching times between states.

In the following sections, we first detail how we built an underlying gene regulatory network of mammalian cellular aging based on the existing experimental literature (Haruta et al., 2000; Stambolic et al., 2001; Inoki et al., 2002, 2003; Ogawara et al., 2002; Kong, 2004; Lahav et al., 2004; Nemoto, 2004; You et al., 2006; Greer et al., 2007b; Okoshi et al., 2007; Budanov and Karin, 2008; Gwinn et al., 2008; Lan et al., 2008; Salih and Brunet, 2008;

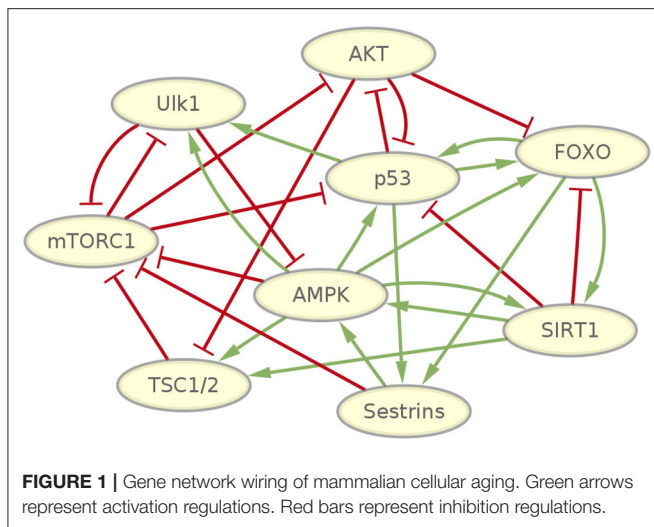
Cantó et al., 2009; Chen et al., 2010; Georgescu, 2010; Ghosh et al., 2010; Sengupta et al., 2010; Yi and Luo, 2010; Budanov, 2011; Dunlop et al., 2011; Gao et al., 2011; Kim et al., 2011; Löffler et al., 2011; Renault et al., 2011; Wang F. et al., 2011; Parmigiani et al., 2014). Based on this gene circuit, we developed a mathematical model to quantitatively describe the basic features of the mammalian cellular aging process. A landscape with three attractors that represent fast-aging, intermediate, and slow-aging, respectively, was identified. We discuss the biological functions of these three attractors and their possible effects on mammalian cellular aging. We identify the dominant paths of system switching between the fast-aging and slow-aging state attractors, giving the most likely route of how fast-aging and slow-aging processes may have occurred. Since the cellular aging process is affected by many factors from inside and outside of the system, we performed a global sensitivity analysis based on the landscape topography and kinetics to investigate how the changes of the genes and the regulations influence the fast-aging and slow-aging processes. The genes or regulations that may play key roles in controlling the mammalian cellular aging process are predicted. Finally, we also found a possible scenario of oscillation dynamics between fast-aging and slow-aging. We show the phase transition/bifurcation between a multi-stable state and oscillation of fast-aging and slow-aging. We show that the flux is the dynamic cause and entropy production rate related to the flux the thermodynamic cause for this phase transition/bifurcation process of fast-aging and slow-aging.

2. RESULTS

2.1. Network Wiring and Kinetic Equations

To investigate the fundamental dynamic features of mammalian cellular aging, we first selected genes that have been revealed to play essential roles in aging. We then gathered the regulatory information regarding these genes by mining the literature for previous relevant studies (Haruta et al., 2000; Stambolic et al., 2001; Inoki et al., 2002, 2003; Ogawara et al., 2002; Kong, 2004; Lahav et al., 2004; Nemoto, 2004; You et al., 2006; Greer et al., 2007b; Okoshi et al., 2007; Budanov and Karin, 2008; Gwinn et al., 2008; Lan et al., 2008; Salih and Brunet, 2008; Cantó et al., 2009; Chen et al., 2010; Georgescu, 2010; Ghosh et al., 2010; Sengupta et al., 2010; Yi and Luo, 2010; Budanov, 2011; Dunlop et al., 2011; Gao et al., 2011; Kim et al., 2011; Löffler et al., 2011; Renault et al., 2011; Wang F. et al., 2011; Parmigiani et al., 2014). We integrated all of this information to give rise to a gene regulatory network. This gene regulatory network of the mammalian cellular aging includes nine genes and 28 regulatory interactions, as shown in **Figure 1**.

Some well-studied genes and pathways related to mammalian cellular aging are included in the network. The PI3K/Akt signaling pathway, which inhibits FOXO transcription factors, is highly conserved across metazoans (Hay, 2011). FOXO transcription factors have consistently been revealed as important determinants in aging and longevity. In mammals, the FOXO subfamily is involved in a wide range of crucial cellular processes regulating stress resistance, metabolism, cell cycle arrest, and apoptosis (Martins et al., 2015). AMPK and



mTORC1 are important nutrient-sensing protein kinases that have antagonistic functions in regulating metabolic homeostasis. Several experiments show that inhibiting mTORC1 delays aging in yeast and invertebrates, extends lifespan in mice, and has an impact on a diverse array of age-related diseases (Johnson et al., 2013). An increase in AMPK activity extends lifespan in lower organisms (Salminen and Kaarniranta, 2012), and experiments demonstrated that AMPK together with mTORC1 and ULK1, a key protein needed in the early steps of autophagosome biogenesis, controls cell growth and autophagy in mammals (Huber et al., 2012; Dunlop and Tee, 2013). Inactivation of Sestrin genes in invertebrates resulted in diverse metabolic pathologies, including oxidative damage, fat accumulation, mitochondrial dysfunction, and muscle degeneration, which resemble accelerated tissue aging (Lee et al., 2013). SIRT1 regulates numerous processes, including inflammation and cellular senescence and aging (Rahman and Bagchi, 2013). SIRT1 is decreased in both transcriptional and post-transcriptional conditions during aging, accompanied by attenuated mitochondrial biogenesis, an important component of aging-related diseases (Yuan et al., 2016). The p53 gene is well-known as a tumor suppressor gene. Its activation also modulates cellular senescence and organismal aging. P53 also regulates aging in a complex way. It accelerates or decelerates the aging process under different circumstances (Rufini et al., 2013). Besides, a few important regulatory interactions affecting aging have been studied. It is found that an AMPK-FOXO pathway is important for mediating life span extension by caloric restriction in *C. elegans* (Greer et al., 2007a). AMPK regulation of FOXO factors may help coordinate energy metabolism with cellular responses to prevent diabetes (Greer et al., 2007b). FOXO3 and p53 are part of a common transcriptional network affecting cellular and organismal responses that is important to counter aging and cancer (Renault et al., 2011). The p53-regulated antioxidant Sestrins gene family involved in control of the AMPK-TORC1 pathway and mitochondrial function might defend against the

accumulation of detrimental damage, which potentiates aging and fuel age-associated diseases (Budanov, 2011). It has been found that SIRT2 deacetylates FOXO3 to increase the expression of its target genes, thus regulating cell proliferation, anti-oxidation, and apoptosis (Wang et al., 2007). Detailed references for each regulatory interaction in the network can be found in Table S1.

The complexity of the network wiring is reflected in two different aspects. From the molecular biological perspective, several types of regulatory interactions are present in the network, including transcriptional regulation, translational control, protein-protein interaction, and signal transduction. From the network wiring topology perspective, the intensive communications among the nine genes imply emergent biological functions as a result. The network motif includes positive and negative, feed-forward and feed-back loops. This can give rise to the possibility of generating complex dynamic features, such as forming multi-stable state attractors and oscillations.

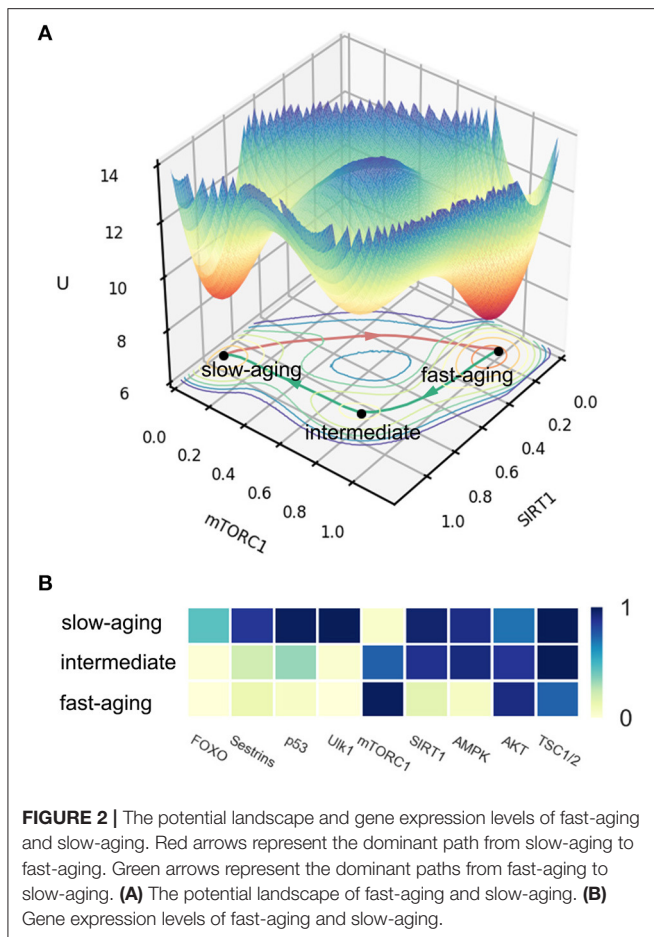
To explore the dynamics of the mammalian cellular aging network, we employ non-linear differential equations (Tyson and Novák, 2010) to describe the dynamics of each genes expression in the network. A sigmoidal function was previously used to model T-cell differentiation (Hong et al., 2012) and epithelial-mesenchymal transition (Watanabe et al., 2019) in mammalian cells and appears to be suitable for describing both gene expression and gene regulation networks (Mjolsness et al., 1991; Hong et al., 2011, 2015). There are nine genes in the network, so a total of nine equations are included in our simulation model. The form of the kinetic equation is shown as:

$$\dot{X}_i = \gamma_i [F(\sigma_i W_i) - X_i] \quad (1)$$

$$F(\sigma_i W_i) = 1/(1 + e^{-\sigma_i W_i}) \quad (2)$$

$$W_i = \omega_{i0} + \sum_j \omega_{ij} X_j \quad (3)$$

where X_i represents the expression level of the gene i , where $i = 1, \dots, 9$, in the network. The parameter γ_i denotes a reciprocal rate description of the dynamic timescale of the system. $F(\sigma_i W_i)$ denotes the regulation for gene i . It is described by a non-linear sigmoidal function that varies from 0 where $W_i \ll -1/\sigma_i$ to 1 where $W_i \gg 1/\sigma_i$. W_i denotes a combination of the effects of all input regulations to gene i . A small regulation input to gene i will lead to a weak driving force for the dynamics of gene i , while a large regulation input to gene i will give rise to a large driving force to the dynamics of gene i . The coefficient ω_{ij} indicates the regulatory strength of gene j on gene i , where $\omega_{ij} < 0$ for inhibitory interaction, $\omega_{ij} > 0$ for promoting regulation, and $\omega_{ij} = 0$ for no effect of gene j on gene i . The coefficient ω_{i0} represents the basal regulation strength. Since cellular aging is not an isolated or static process, the value of ω_{i0} can be varied under genetic changes or environmental influences. The parameter σ_i controls the steepness of the sigmoidal function at its inflection point. It provides a threshold for the onset of significant dynamics of the gene.



2.2. Potential Landscape of Aging

A biological system is naturally subject to intrinsic and extrinsic fluctuations. Therefore, we added an additional fluctuation term in the ODEs to characterize the stochastic behaviors of the mammalian cellular aging process. We use the Langevin dynamic approach to simulate the gene circuit dynamics. From the resulting dynamic trajectories of the gene expressions, we collected the statistics and quantified the underlying potential landscape (Wang J. et al., 2011). For visualization, we projected the high-dimensional state spaces into two coordinates. This choice can still distinguish the major biological functions that are reflected as attractors in the landscape.

The potential landscape of mammalian cellular aging is shown in **Figure 2A**. X and Y coordinates represent the expression levels of genes SIRT1 and mTORC1, respectively. The Z coordinate represents the landscape U . Three attractors emerge on the 3D landscape. The position of each attractor can be distinguished by the expression levels of all of the nine genes in the network. This is shown as a heatmap in **Figure 2B**. We defined the three attractors as the slow-aging state (S), fast-aging state (F), and intermediate state (I) according to the gene expression levels and their corresponding gene functions. In the slow-aging state, genes with longevity-promoting functions, such as SIRT1, AMPK, and Ulk1,

have relatively high expression levels and genes with lifespan-limiting effects, such as mTORC1 and AKT, have relatively low expression levels. In the fast-aging state, longevity-promoting genes and lifespan-limiting genes show the opposite expression patterns compared to the case of slow aging. The intermediate state is located between the fast-aging and the slow-aging state; genes, such as FOXO, Sestrins, p53, and Ulk1 show relatively low expression levels compared to the slow-aging state, while genes, such as SIRT1 and AMPK show relatively high expression levels compared to the slow-aging state. Some organisms undergo rapid aging and death, while others grow old slowly and live far longer, even within a population of isogenic organisms in identical environments (Crane et al., 2020). A previous study on an aging model of yeast cell with an intermediate state was proposed based on categorizing the age-dependent phenotypic conditions and was validated through experiment (Jin et al., 2019). The emergence of the intermediate state provides new perspectives to explain the mechanisms of the mammalian cellular aging process. The intermediate state may provide a bridge or mid-land between fast-aging and slow-aging. This can help to facilitate the fast or slow aging process through the intermediate state.

The depths of the three attractors are significantly different. A deeper attractor has lower energy U , where $U = -\log P$ and P represents the steady-state probability of the state. Thus, the system is expected to reside in a deeper attractor for a longer time, and it is harder to escape from it. The mean first passage time (MFPT) reflects the average transition time from one attractor to another. In **Figure 2A**, we can see the fast-aging state attractor is deeper than the slow-aging state attractor and the intermediate attractor. We calculate the MFPT from slow-aging to fast-aging and from fast-aging to slow-aging as 44.27 and 126.32, respectively. These quantifications indicate that under current system conditions, the system prefers to stay at the fast-aging attractor with lifespan-limiting effects, and the transition from slow-aging to fast-aging is significantly faster than that from fast-aging to slow-aging. This may explain why the fast-aging process seems more dominant, since the fast-aging state attractor is more stable and therefore has a higher chance of being observed.

The dominant path (Wang et al., 2010) is the most probable path when a system switches from one state to another. We quantify the dominant paths from slow-aging to fast-aging and from fast-aging to slow-aging, which are separately shown as a red arrow and a green arrow in **Figure 2A**. It is notable that the two dominant paths are completely different. For the fast-aging process, the red dominant path is directly from the slow-aging state to fast-aging state. The green dominant path from fast-aging to slow-aging passes through the intermediate state. This indicates that, in our mammalian cellular aging model, the slow-aging process is divided into two steps. The first step is from the fast-aging state to the intermediate state, marked by a significant increase in the expression levels of AMPK and the SIRT1. These two genes together regulate diverse processes, such as cellular fuel metabolism, inflammation, and mitochondrial function (Ruderman et al., 2010). The second step is from the intermediate state to the slow-aging state, marked by the changes of the gene expressions of the other aging-related genes,

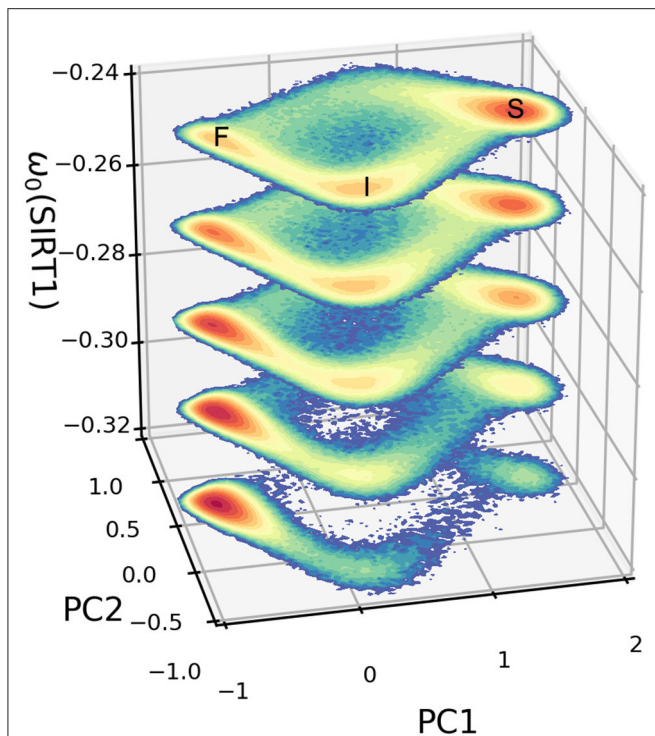


FIGURE 3 | Dynamic landscape of fast-aging and slow-aging upon changes in the basal expression level of SIRT1. The horizontal coordinates represent the top two principal components of gene expression, while the vertical axis represents changes in the basal level of expression for gene SIRT1.

such as FOXO and mTORC1 (see **Figure 2B**). It is possible to experimentally slow the rate of aging through longevity genes or dietary restriction (Rando and Chang, 2012), but further experimental verifications are needed to check the predictions of the two-step transition to slow-aging in our model.

2.3. Dynamics of Landscapes of Aging

Aging is certainly not an isolated process. Most of the aging-related genes are multi-faced, and they also play key roles in some other basic functions, such as metabolism, energy homeostasis, protein synthesis, cell growth, proliferation, autophagy, apoptosis, and senescence. Several kinds of stimulations have been found to have a great influence on the natural aging process. Genetic manipulations of certain genes have been found to significantly extend the lifespan of *C. elegans* (McCormick et al., 2011). Dietary restrictions have been found to regulate aging and increase the healthy lifespan in various model organisms (Kapahi et al., 2010; Smith-Vikos et al., 2014). The accumulation of cell damage was shown to lead to several types of degenerative diseases like cancer and Alzheimers disease (Powers et al., 2009). These examples reflect the importance of studying aging in a systematic and dynamic way.

In our mammalian cellular aging model, the parameter ω_{i0} in the ODEs represents the basal expression level for each gene i . The increase or decrease of ω_{i0} will influence the behaviors of the system. The dynamic landscapes describe the changes in the landscape topography according to the

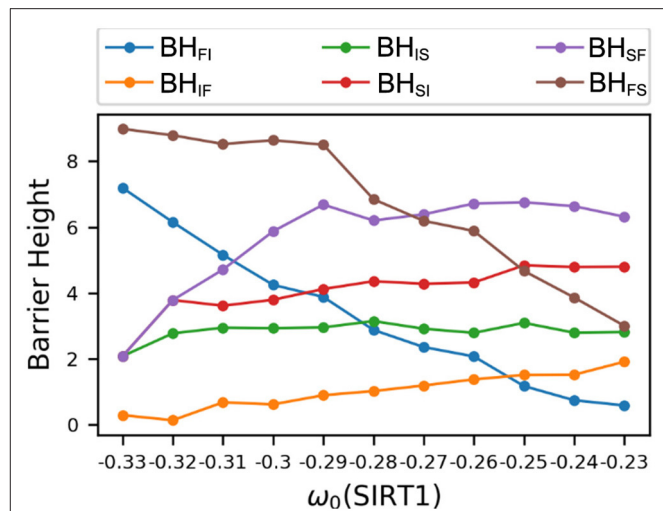


FIGURE 4 | Changes in barrier heights upon increasing the basal gene expression level of SIRT1. BH_{FI} , barrier height from the fast-aging to the intermediate attractor; BH_{IF} , barrier height from the intermediate to the fast-aging attractor; BH_{IS} , barrier height from the intermediate to the slow-aging attractor; BH_{SI} , barrier height from the slow-aging to the intermediate attractor; BH_{SF} , barrier height from the slow-aging to the fast-aging attractor; BH_{FS} , barrier height from the fast-aging to the slow-aging attractor.

changes in certain genes or regulations. The barrier height (BH) based on principal component analysis (PCA) of the landscape can be used to quantitatively measure the degree of difficulty for the system to switch from one attractor to another. BH is defined as the difference between the minimum potential in the current attractor and the potential of the saddle point from the current attractor to the other attractor. We first use the PCA method to project the nine-dimensional landscape into the top two principal components (PCs). We then calculate the BH among the attractors based on the PCA projected landscape.

The dynamical PCA landscapes according to the changes in the basal expression level of SIRT1 are shown in **Figure 3**. The X and Y coordinates represent the top two principal components, respectively. These two principal components show about 95 percent of the variance of the dynamic expression trajectory. The three attractors are labeled F (fast-aging), I (intermediate), and S (slow-aging), respectively. If we increase (decrease) the SIRT1 basal expression level, the depth of the fast-aging attractor decreases (increases) and the depths of both the intermediate and slow-aging attractors increase (decrease). **Figure 4** quantitatively shows the change of barrier height vs. the increase in basal expression level of SIRT1. There are three attractors in the PCA landscape, so a total of six barrier heights for each pair of attractors can be quantified. The line labeled BH_{SF} denotes the BH of the system switching from the slow-aging to the fast-aging attractor, while BH_{FS} represents the BH of the system switching from the fast-aging to the slow-aging attractor. Other labels have similar notations. The results clearly show an increase

(decrease) in the stability of the fast-aging attractor state and a decrease (increase) in the stability for the slow-aging and intermediate states when the SIRT1 basal expression level decreases (increases). These results are consistent with the evidence that SIRT1 plays a key role in dietary restriction-induced longevity promotion, while the activity of SIRT1 decreases with the mammalian cellular aging process (Ruderman et al., 2010).

2.4. Global Sensitivity Analysis of Aging in Mammals

Here, we use global sensitivity analysis to quantitatively identify the contributions of individual regulations on the functional behavior of mammalian cellular aging. We change the basal expression level ω_{i0} for every gene and the regulatory strength for every regulation ω_{ij} to investigate to what extent these regulations influence the functional behavior. The functional stability can be quantitatively measured by the barrier heights. BH_0 represents the barrier height with the original value of the given parameter. ΔBH represents changes in the barrier height when the regulation is changed by a constant value (0.04). Thus, $\Delta BH/BH_0$ can be used to measure the sensitivity of the barrier heights under certain regulation changes. We performed global sensitivity analysis to find the key genes or regulations by changing ω_{i0} or ω_{ij} and then finding out which genes and regulations will significantly impact the landscape stability. These predicted genes or regulations may play important roles in the mammalian cellular aging process or may even be useful in treating aging-related degenerative diseases.

We performed global sensitivity analysis on the basal expression level to quantifying the barrier height changes for every gene. The detailed results of the global sensitivity analysis are shown in **Figure 5**. For the barriers related to the slow-aging state, BH_{SF} and BH_{SI} , we can see that increasing the basal expression levels of the genes AMPK, FOXO, and Sestrins significantly enhances the stability of the slow-aging state. This indicates that it becomes harder for the system to escape from the slow-aging state. In contrast, gene AKT significantly decreases the stability of the slow-aging state. These results are consistent with previous experimental findings (Salminen and Kaarniranta, 2012; Lee et al., 2013; Gharibi et al., 2014; Martins et al., 2015). For the barrier heights related to the fast-aging state, BH_{FI} and BH_{FS} , we can clearly see that increasing the basal expression levels of the genes AMPK, SIRT1, and Sestrins significantly decreases the stability of the fast-aging state. AMPK and Sestrins play opposite roles in the slow-aging state, but the role of SIRT1 in slow-aging is not significant. For the intermediate state, the result is complex. Genes mTORC1 and p53 are only effective in the intermediate state, but not in the other two states. Although the existence of the intermediate state between fast-aging and slow-aging has not been directly verified, this study shows that different genes seem to influence different attractors. This can provide new insight for research on mammalian cellular aging mechanisms.

We also performed global sensitivity analysis on regulatory strength ω_{ij} . The bar charts shown in **Figure 6** reflect $\Delta BH =$

BH_0 vs. ω_{ij} . The most sensitive regulation from the slow-aging state to the fast-aging state is SIRT1- $\rightarrow$ AMPK, and the barrier height from the slow-aging state to the fast-aging state is increased with increasing SIRT1- $\rightarrow$ AMPK. This means that increasing the activation regulation of SIRT1- $\rightarrow$ AMPK will stabilize the slow-aging state and therefore delay the aging process. The most sensitive regulation of barrier height from the fast-aging state to the slow-aging state is AMPK- $\rightarrow$ SIRT1, and the barrier height from the fast-aging state to slow-aging state is decreased with increasing AMPK- $\rightarrow$ SIRT1. This means that increasing the activation regulation of AMPK- $\rightarrow$ SIRT1 will destabilize the fast-aging state and therefore increase the chance of slow aging, thereby delaying the aging process. The most sensitive regulation of barrier height from the intermediate state to the slow-aging state is AKT-p53, and the barrier height from the intermediate state to the slow-aging state is increased with increasing AKT-p53. This means that increasing the inhibition regulation of AKT-p53 will stabilize the intermediate state and decrease the chance of slow aging, effectively promoting the aging process. The most sensitive regulation of barrier height from the slow-aging state to the intermediate state is p53- $\rightarrow$ Sestrins, and the barrier height from the slow-aging state to the intermediate state is increased with increasing p53- $\rightarrow$ Sestrins. This means that increasing the activation regulation of p53- $\rightarrow$ Sestrins will stabilize the slow-aging state and therefore delay the aging process. The most sensitive regulation of barrier height from the fast-aging state to the intermediate state is Sestrins- $\rightarrow$ AMPK, and the barrier height from the fast-aging state to the intermediate state is decreased with increasing Sestrins- $\rightarrow$ AMPK. This means the increasing the activation regulation of Sestrins- $\rightarrow$ AMPK will destabilize the fast-aging state and therefore increase the chance of slow aging, thus effectively delaying the aging process. The most sensitive regulation of barrier height from the intermediate state to the fast-aging state is SIRT1- $\rightarrow$ AMPK, and the barrier height from the intermediate state to the fast-aging state is decreased with increasing SIRT1- $\rightarrow$ AMPK. This means that increasing the activation regulation of SIRT1- $\rightarrow$ AMPK will stabilize the intermediate state and destabilize the fast-aging state and therefore delay the aging process. We show the top three sensitive regulations for each barrier in **Table 1**. Changes in these regulatory strengths significantly change the system behavior. Further experiments are needed to validate these predictions.

2.5. Aging Oscillations Landscape

Oscillation dynamics can emerge in certain parameter regimes when the regulation strengths are varied. The transitions between the oscillation and monostable states are found to be mainly regulated by Sestrins- $\rightarrow$ AMPK. The changes in landscape topography are shown in **Figure 7**. RS represents the regulation strength of Sestrins- $\rightarrow$ AMPK. The landscape shows oscillation dynamics with a Mexican hat shape when RS is 0.76, as shown in **Figure 7B**. The two relatively deeper regions on the oscillation ring correspond to the fast-aging and slow-aging state, respectively. The states of the system rotate clockwise along the oscillation ring valley around the central hill of the Mexican hat. When the regulation strength RS is increased, the slow-aging state attractor becomes deeper. When the regulation strength

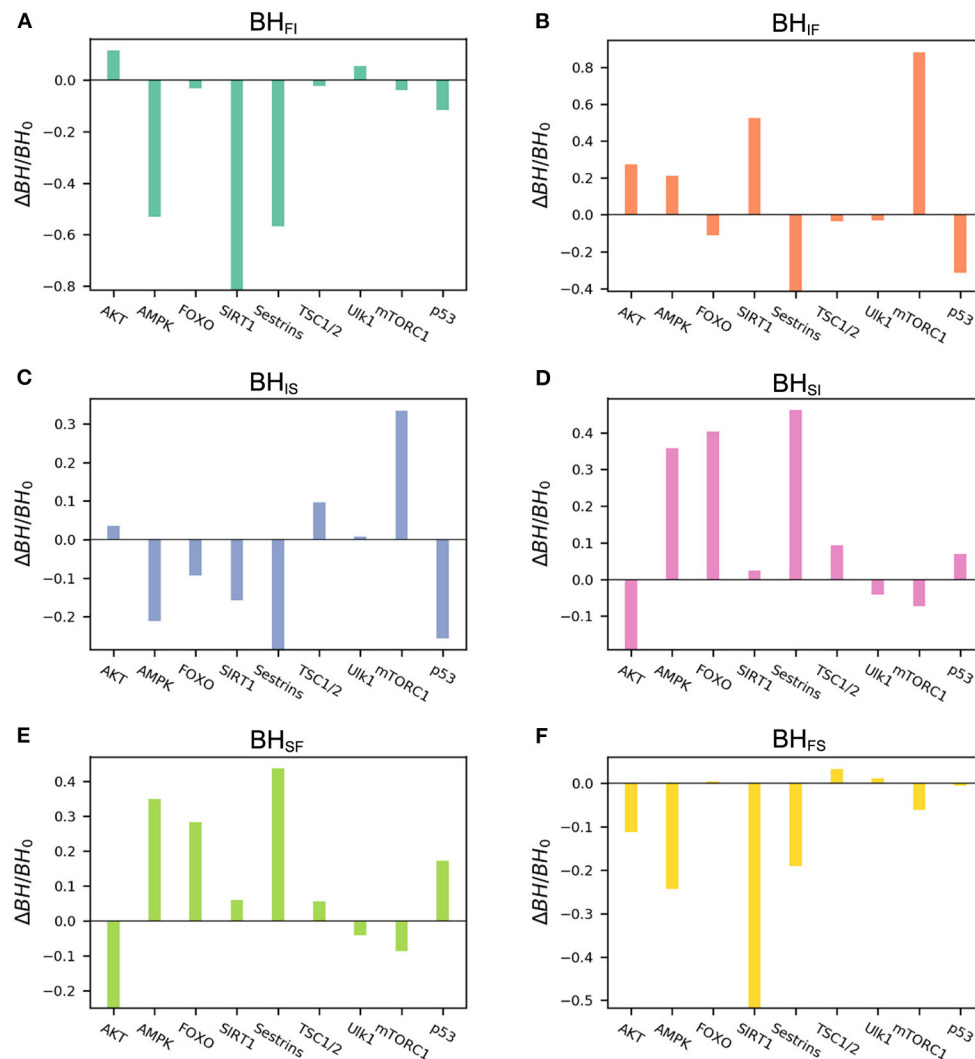


FIGURE 5 | Results of global sensitivity analysis of the barrier height upon changing the basal expression level of different genes. BH_0 , barrier height with the original value; ΔBH , change in barrier height upon changing the basal expression level. **(A)** Change in barrier height from the fast-aging to the intermediate attractor. **(B)** Change in barrier height from the intermediate to the fast-aging attractor. **(C)** Change in barrier height from the intermediate to the slow-aging attractor. **(D)** Change in barrier height from the slow-aging to the intermediate attractor. **(E)** Change in barrier height from the slow-aging to the fast-aging attractor. **(F)** Change in barrier height from the fast-aging to the slow-aging attractor.

RS is increased to 0.88, the system switches from the oscillation to a monostable state with only the slow-aging attractor state. In contrast, when the regulation strength RS is decreased, the basin at the fast-aging steady state becomes deeper. When the regulation strength RS is decreased to 0.62, the system switches from the oscillation to a monostable state with only the fast-aging steady state.

Interestingly, these oscillation dynamics were found in the previous mathematical model of *C. elegans* (Zhao and Wang, 2016). The oscillation can drive the dynamics to switch coherently (periodically) between the fast-aging state and the slow-aging state. The processes of fast-aging and slow-aging occur at different times along with the oscillation. The transitions between the fast-aging state and the slow-aging state with

the oscillation are different from the transitions in a tri-stable system. The transitions between the fast-aging state and the slow-aging state in the tri-stable regime are random and incoherent, while the transitions between the fast-aging state and the slow-aging state in the oscillation regime are periodic and coherent. In order to address the role of the flux as the driving force of the aging process in addition to the landscape, we quantified the flux integral as a measure of the magnitude of the flux and the coherence of the oscillation, as shown in **Figure 8B**. The flux integral correlates with the coherence. This indicates that higher flux leads to more stable and coherent oscillation.

We also quantified the thermodynamic cost in terms of the entropy production rate (EPR), which is related to the flux and

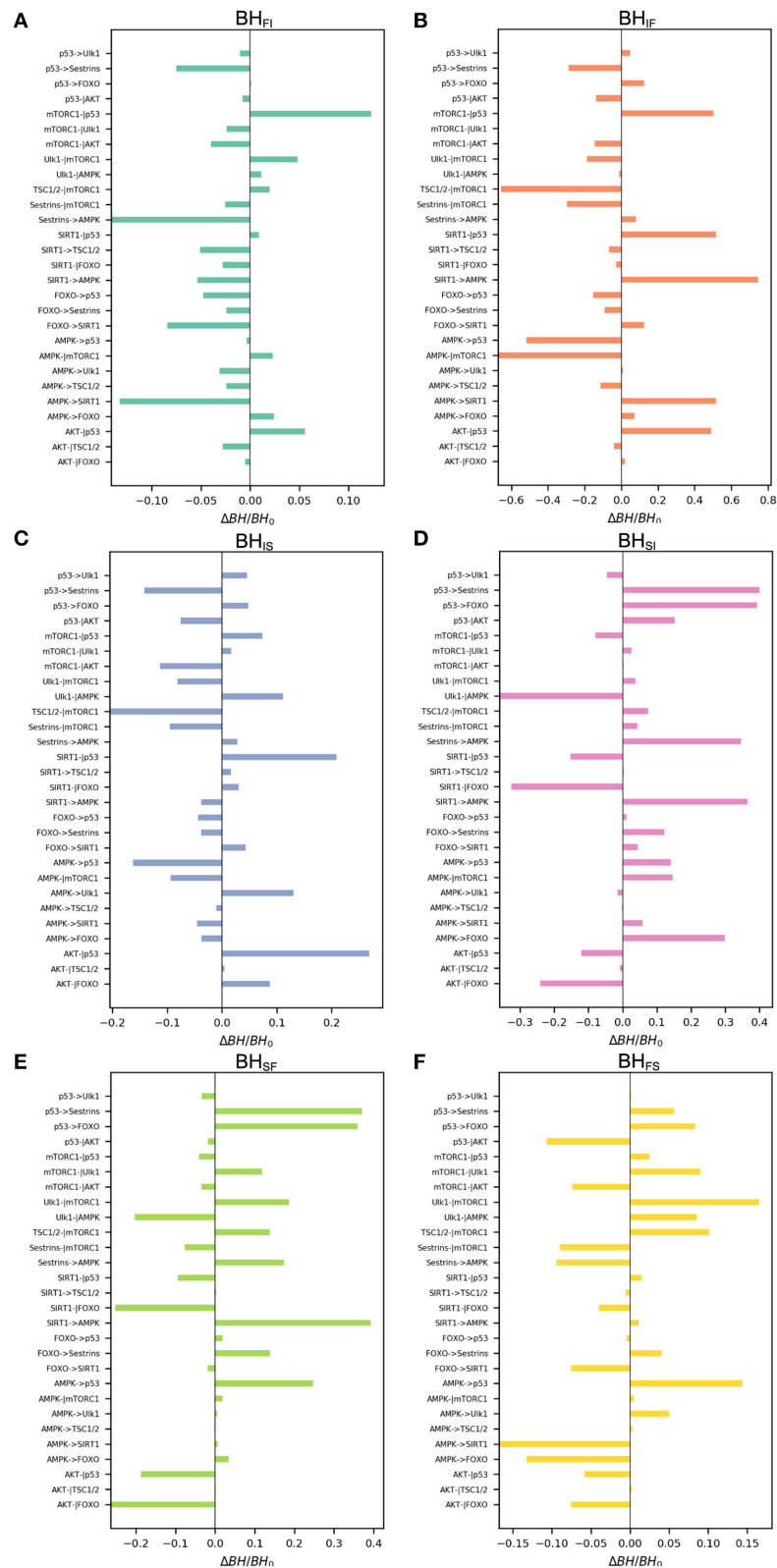


FIGURE 6 | Global sensitivity analysis of the barrier height upon changing regulatory strengths. BH_0 , barrier height with the original value; ΔBH , change in barrier height upon changing the regulation. **(A)** Change in barrier height from the fast-aging to the intermediate attractor. **(B)** Change in barrier height from the intermediate to the fast-aging attractor. **(C)** Change in barrier height from the intermediate to the slow-aging attractor. **(D)** Change in barrier height from the slow-aging to the intermediate attractor. **(E)** Change in barrier height from the slow-aging to the fast-aging attractor. **(F)** Change in barrier height from the fast-aging to the slow-aging attractor.

TABLE 1 | Top three key regulations from global sensitivity analysis.

Barrier	Sensitive regulations
BH_{SF}	SIRT1- $\rightarrow$ AMPK, p53- $\rightarrow$ Sestrins, p53- $\rightarrow$ FOXO.
BH_{FS}	AMPK- $\rightarrow$ SIRT1, Ulk1- $\rightarrow$ mTORC1, AMPK- $\rightarrow$ p53.
BH_{IS}	AKT- $\rightarrow$ p53, SIRT1- $\rightarrow$ p53, TSC1/2- $\rightarrow$ mTORC1.
BH_{SI}	p53- $\rightarrow$ Sestrins, p53- $\rightarrow$ FOXO, SIRT1- $\rightarrow$ AMPK.
BH_{FI}	Sestrins- $\rightarrow$ AMPK, AMPK- $\rightarrow$ SIRT1, mTORC1- $\rightarrow$ p53.
BH_{IF}	SIRT1- $\rightarrow$ AMPK, AMPK- $\rightarrow$ mTORC1, TSC1/2- $\rightarrow$ mTORC1.

the mean flux, for the phase transition/bifurcation from the monostability of fast-aging to oscillation and from the oscillation to monostability of slow-aging by increasing the regulation strength of Sestrins- $\rightarrow$ AMPK. An increase in the EPR indicates that the system costs more energy to maintain. The mean flux correlates with the EPR. As shown in **Figure 8A**, the EPR is low when the system stays in the phase of the fast-aging state. When the strength of Sestrins- $\rightarrow$ AMPK increases, the EPR increases sharply at the phase where the transition from the stable fast-aging state to oscillation occurs. When the system switches from oscillation to the monostable slow-aging state, the EPR sharply decreases and then stays at a low level. This demonstrates that the oscillation costs more energy to maintain than either the fast-aging or slow-aging state. Through the oscillation, the dynamic process of switching between fast-aging and slow-aging achieves functional switching, which can cost more energy. Therefore, there can be direct and indirect pathways for aging. The direct pathway is the one directly from the slow-aging state to the fast-aging state. The indirect pathways can be from the slow-aging state to the fast-aging state through either the intermediate state or oscillation.

3. DISCUSSION

In this study, we presented a mathematical model to describe the dynamic features of the mammalian cellular aging process. We built the underlying gene regulatory network by integrating the information from previous experimental studies. The genes and wirings in the gene regulatory network were formed, and the dynamics of gene expression was described by nine non-linear ordinary differential equations. Based on these equations, we quantified the potential landscape of the mammalian cellular aging process. Three attractors emerged on the landscape: the fast-aging, intermediate, and slow-aging states. When the system resides in one of the three attractors, the escape time is determined by the depth of the attractor. The system can also switch from one attractor to another, and the transition needs to overcome the barriers between the attractors. We integrated the previous studies and analyzed the mammalian cellular aging process from a systemic and network perspective.

The aging process is not only a spontaneous biological process but also can be significantly altered by interventions, such as genetic manipulations and dietary restrictions. Thus, the potential landscape of aging is not invariant. We changed

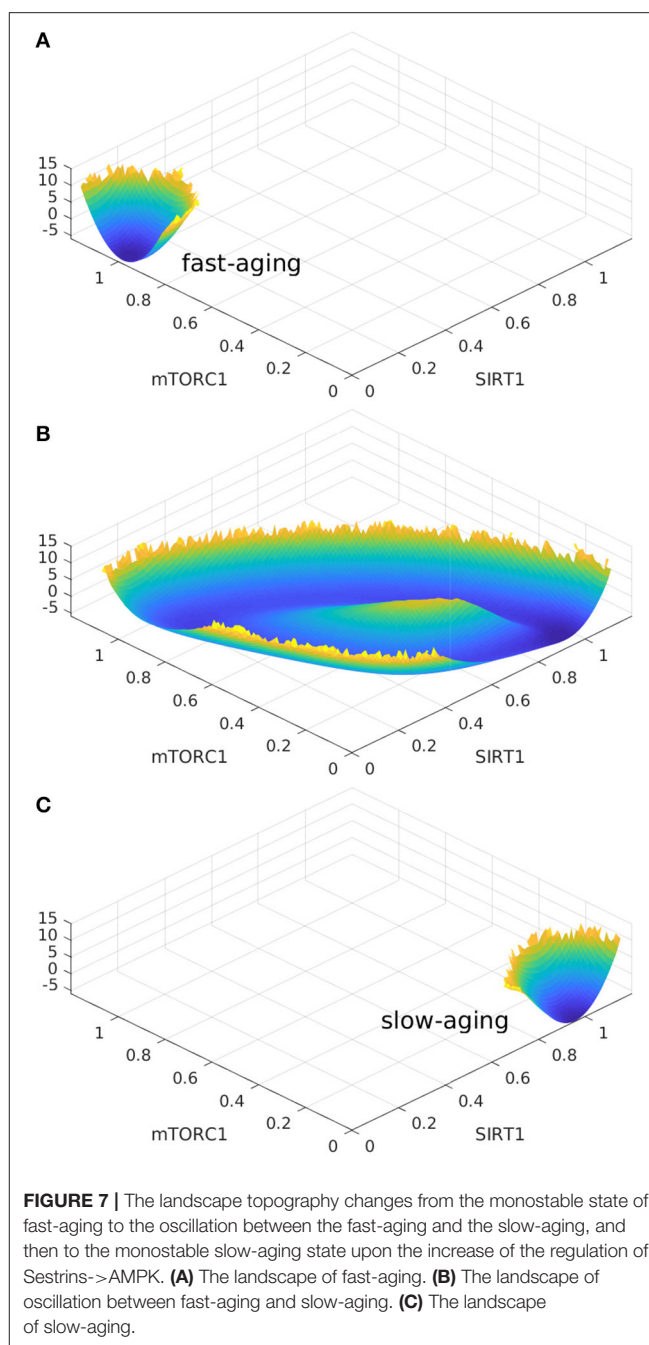
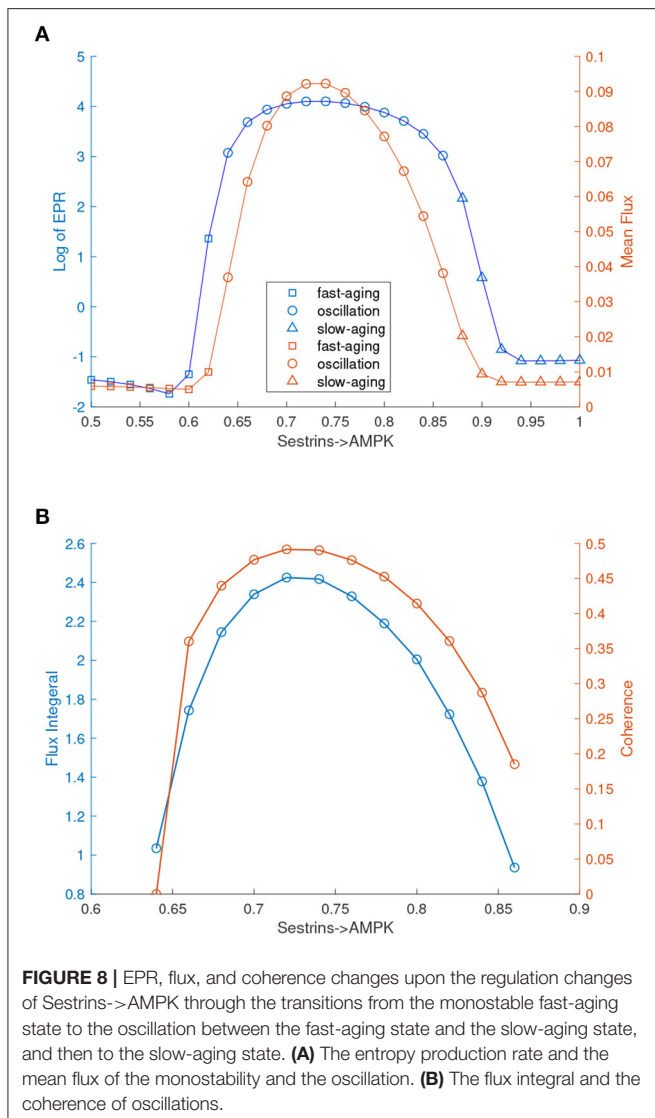


FIGURE 7 | The landscape topography changes from the monostable state of fast-aging to the oscillation between the fast-aging and the slow-aging, and then to the monostable slow-aging state upon the increase of the regulation of Sestrins- $\rightarrow$ AMPK. **(A)** The landscape of fast-aging. **(B)** The landscape of oscillation between fast-aging and slow-aging. **(C)** The landscape of slow-aging.

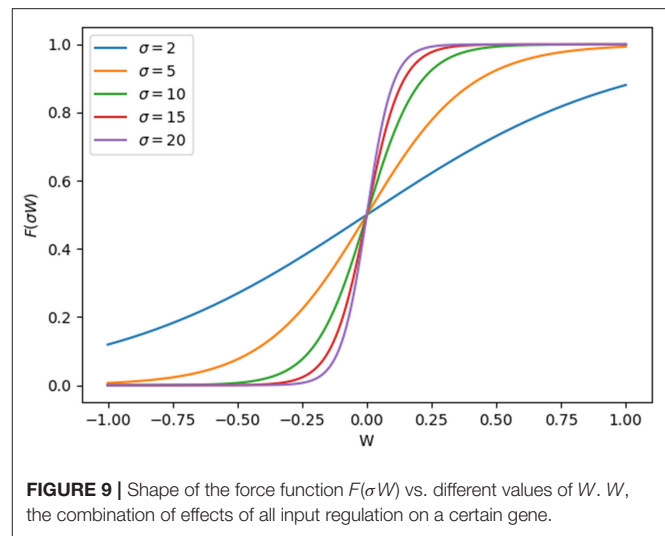
certain regulations in our model in order to perform quantitative analysis and investigate the changes in aging functions through the changes in the landscape. The stabilities of attractors can be significantly changed by the basal strength of certain genes and the regulatory strengths of gene-gene regulations. We believe that these genes or regulations may play key roles in the mammalian cellular aging process. Further experiments are needed to validate these predictions.

Oscillations emerge in certain regulation regimes. The oscillation leads to switching between the processes of fast-aging



and slow-aging. This is different from switching between the fast-aging state and the slow-aging state through the stochastic trajectories in the tri-stable regime. The switching between the fast-aging state and the slow-aging state in the oscillation regime is periodic and coherent. In contrast, the switching between the fast-aging state and the slow-aging state in the tri-stable regime is random and incoherent. Through the analysis of the flux integral and coherence as well as the mean flux and entropy production rate, it is suggested that more energy is required to sustain oscillations.

In this work, we have provided a framework to reveal the underlying mechanism of fast-aging and slow-aging in mammals based on landscape and flux theory. We predict the key genes and interactions in the fast-aging and slow-aging processes. This approach may be helpful for studying strategies for expanding lifespan in mammals or humans.



4. MATERIALS AND METHODS

4.1. Kinetic Equations

The shape of the term $F(\sigma W)$ in the non-linear ODEs used in our model is intuitively similar to Hill equations. The sigmoidal shape and the steepness of $F(\sigma W)$ can be altered by varying certain parameters, as shown in **Figure 9**.

The form of summing Hill equations as the regulation force for ODEs is used in other studies (Li, 2018; Li and Balazsi, 2018). It is shown in Equation (4).

$$\frac{dx_k}{dt} = \sum_{i \in \text{activators}} \frac{w_k x_i^n}{s_{ik}^n + x_i^n} + \sum_{j \in \text{inhibitors}} \frac{w_k s_{jk}^n}{s_{jk}^n + x_j^n} - \mu_k x_k \quad (4)$$

where x_k represents the k th gene expression, while w_k represents the relative strength of every regulation of gene k . Parameter μ_k is defined as the self-degradation rate. Parameter n is the Hill coefficient. Parameters s_{ik} and s_{jk} represent the inflection points of the activation or inhibition regulation terms.

However, the Hill equations have an inherent defect that the value of w cannot be negative. This leads to defects in the case of the presence of both activation and inhibition regulations. For example, under the additive rule, when adding a negative regulation or increasing the weight of a negative regulation, the expression changes, and dx/dt may increase, while in fact it should decrease. This problem also emerges when using the multiplicative rule.

In the equations in our model, the regulations of activation and inhibition have the same form, $\omega_{ij} X_j$, as shown in Equations (1–3). The coefficient ω_{ij} indicates the regulatory strength from gene j to gene i , where $\omega_{ij} < 0$ for inhibitory interaction and $\omega_{ij} > 0$ for promoting regulation. The value of $\dot{X}_i$ is increased when W_i is increased. W_i is increased or decreased when an activation term or inhibition term is added. Thus, $\dot{X}_i$ is increased when an activation term is added and is decreased when an inhibition

term is added. This overcomes the defects of the form representing the regulation in Equation (4). Therefore, we can directly add the regulation term, $\omega_{ij}X_j$, together to quantify the force. The calculation is logically reasonable and less time-consuming.

4.2. Parameter Setting

We assume the following restrictions on the regulation parameters. For all equations, the steepness parameter is set to $\sigma = 10$, and the time scale parameter is set to $\gamma = 1$. The basal weight parameter is set to $-1 < \omega_{i0} < 1$, and the regulation weight parameter is set to $-1 < \omega_{ij} < 1$. This equation has the great advantage that it is subject to all the powerful analytical and simulation tools of non-linear ODEs. This is because in the limit of large σ_i , it behaves like a discrete Boolean network. When $\sigma \gg 1$, X_i tends to flip between 0 and 1, and the dynamic system describes a Boolean network.

There is a question about why and how to set the basal weight parameter ω_{i0} . Technically, ω_{i0} has to be present because if only activation or inhibition on a target node X exists, the expression level of X will eventually reach the boundary values, 1 and 0. This can make it hard for multi-stability to emerge. Biologically, the model we developed can be influenced by the environment, and many conditions and molecular signals from outside can change the state of the system and affect the basal level of the expressions. The parameter settings of ω_{ij} and ω_{i0} are shown in Table S2.

4.3. Langevin Method

The aging process in real life is influenced by the intrinsic or external fluctuations of the system. Langevin equation is appropriate to describe the stochastic time evolution of gene expression dynamics. These stochastic differential equations for describing to gene regulatory network dynamics are as follows:

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}) + \boldsymbol{\eta} \quad (5)$$

where $\mathbf{x}$ is a vector of the gene expressions and $\mathbf{F}(\mathbf{x})$ is the driving force of the gene-regulating network dynamics. The term $\boldsymbol{\eta}$ represents fluctuation or noise force, which has a Gaussian probability distribution with correlation function $\langle \eta_i(t), \eta_j(t') \rangle = 2D_{ij}\delta_{ij}\delta(t - t')$, where $\mathbf{D}$ is the diffusion coefficient matrix characterizing the strength of the fluctuations. The global steady-state probability distribution P for the state space can be quantified through the statistics by collecting the time evolution trajectories of the expression dynamics from long-duration simulations.

4.4. Landscape and Flux

The individual stochastic trajectory is unpredictable due to its random nature. However, the evolution of the probability distribution is predictable and can be used to describe the probabilistic behaviors and patterns of the aging process.

The evolution of the probability distribution is governed by the Fokker-Planck equation (Wang et al., 2008; Wang, 2015) as follows:

$$\frac{\partial P(\mathbf{x}, t)}{\partial t} = -\nabla \cdot \mathbf{J}(\mathbf{x}, t) \quad (6)$$

$$\mathbf{J}(\mathbf{x}, t) = \mathbf{F}(\mathbf{x})P(\mathbf{x}, t) - \mathbf{D} \cdot \nabla P(\mathbf{x}, t) \quad (7)$$

which presents that the change in the probability $P(\mathbf{x}, t)$ in time at state $\mathbf{x}$ and time t is equal to the probability flux $\mathbf{J}(\mathbf{x}, t)$ in or out of this state at time t characterized by its divergence. In the steady state, the divergence of probability flux is equal to zero. However, the probability flux is not necessarily equal to zero. The steady-state probability flux, due to its divergent free nature, is rotational as a curl. The steady-state probability flux at the steady state (long time limit) is given in Equation (7). The steady-state probability flux being not equal to zero represents net flow to or from the system. The non-zero net flow breaks the detailed balance. Therefore, the steady-state probability flux quantifies the degree of non-equilibrium away from the equilibrium when it has deviated from zero. For non-equilibrium systems, the driving force $\mathbf{F}$ for the dynamics can be decomposed to a gradient of the potential landscape and a curl flux force under constant fluctuations (Wang et al., 2008): $\mathbf{F} = -\mathbf{D} \cdot \nabla U + \mathbf{J}_{ss}/P_{ss}$, where $U = -\ln P_{ss}$ is the potential landscape, while P_{ss} is the steady-state probability distribution.

4.5. Dominant Path

The dominant paths are the most probable paths when the system switches from one state to another. The quantification of the dominant paths is important for uncovering how the biological processes have actually occurred and is therefore the key for understanding the underlying physical mechanism and function. The dominant path can be quantified by the path integral approach (Wang et al., 2010; Wang J. et al., 2011). The probability of switching from the initial $\mathbf{x}$ at time 0 to the final $\mathbf{x}$ at time t with the path integral is given as:

$$\begin{aligned} P(\mathbf{x}_{\text{final}}, t, \mathbf{x}_{\text{initial}}, 0) &= \int D\mathbf{x} \exp\left[-\int_0^t dt \left(\frac{1}{2}\nabla \cdot \mathbf{F}(\mathbf{x})\right.\right. \\ &\quad \left.\left. + \frac{1}{4}(\dot{\mathbf{x}} - \mathbf{F}(\mathbf{x})) \cdot \mathbf{D}^{-1} \cdot (\dot{\mathbf{x}} - \mathbf{F}(\mathbf{x}))\right)\right] \\ &= \int D\mathbf{x} \exp[-S(\mathbf{x})] \\ &= \int D\mathbf{x} \exp\left[-\int_0^t L(\mathbf{x}(t))dt\right] \end{aligned} \quad (8)$$

The integral over $D\mathbf{x}$ represents the sum over all the possible trajectories from the state $\mathbf{x}_{\text{initial}}$ at time 0 to the state $\mathbf{x}_{\text{final}}$ at time t . $\mathbf{F}(\mathbf{x})$ represents the driving force of the gene regulatory network dynamics. $\mathbf{D}$ represents the strength of the diffusion coefficient matrix. $S(\mathbf{x})$ and $L(\mathbf{x}(t))$ represent the action and the Lagrangian of the associated path. Each path is assigned with a probability

weight, $\exp[-S(\mathbf{x})]$, associated with the action of that path. The dominant path is the path with the largest weight. Therefore, the dominant paths can be identified through minimizing the action.

4.6. Entropy Production Rate

A non-equilibrium system often exchanges energy, matter, and information with the environment. This leads to thermodynamic dissipation. The change in the system entropy in the non-equilibrium system can be divided into two parts (Wang et al., 2008; Wang, 2015) as:

$$\dot{S} = \dot{S}_t - \dot{S}_e \quad (9)$$

$$\dot{S}_t = \int d\mathbf{x} (\mathbf{J} \cdot \mathbf{D}^{-1} \cdot \mathbf{J}) / P \quad (10)$$

$$\dot{S}_e = \int d\mathbf{x} (\mathbf{J} \cdot \mathbf{D}^{-1}) \cdot \mathbf{F}' \quad (11)$$

where $\dot{S}_t$ represents the entropy production rate (EPR) or the total entropy rate of the system and environment and $\dot{S}_e$ represents the heat dissipation rate of the environments. The effective force $\mathbf{F}'$ is defined as $\mathbf{F}' = \mathbf{F} - \nabla \cdot \mathbf{D}$.

REFERENCES

- Budanov, A. V. (2011). Stress-responsive sestrins link p53 with redox regulation and mammalian target of rapamycin signaling. *Antioxid. Redox Signal.* 15, 1679–1690. doi: 10.1089/ars.2010.3530
- Budanov, A. V., and Karin, M. (2008). p53 target genes sestrin1 and sestrin2 connect genotoxic stress and mTOR signaling. *Cell* 134, 451–460. doi: 10.1016/j.cell.2008.06.028
- Cantó, C., Gerhart-Hines, Z., Feige, J. N., Lagouge, M., Noriega, L., Milne, J. C., et al. (2009). AMPK regulates energy expenditure by modulating NAD⁺ metabolism and SIRT1 activity. *Nature* 458, 1056–1060. doi: 10.1038/nature07813
- Chen, C.-C., Jeon, S.-M., Bhaskar, P. T., Nogueira, V., Sundararajan, D., Tonic, I., et al. (2010). FoxOs inhibit mTORC1 and activate akt by inducing the expression of sestrin3 and rictor. *Dev. Cell* 18, 592–604. doi: 10.1016/j.devcel.2010.03.008
- Colman, R. J., Beasley, T. M., Kemnitz, J. W., Johnson, S. C., Weindruch, R., and Anderson, R. M. (2014). Caloric restriction reduces age-related and all-cause mortality in rhesus monkeys. *Nat. Commun.* 5:3557. doi: 10.1038/ncomms4557
- Crane, M. M., Chen, K. L., Blue, B. W., and Kaeblerlein, M. (2020). Trajectories of aging: how systems biology in yeast can illuminate mechanisms of personalized aging. *Proteomics* 20:1800420. doi: 10.1002/pmic.201800420
- Dunlop, E. A., Hunt, D. K., Acosta-Jaquez, H. A., Fingar, D. C., and Tee, A. R. (2011). ULK1 inhibits mTORC1 signaling, promotes multisite raptor phosphorylation and hinders substrate binding. *Autophagy* 7, 737–747. doi: 10.4161/auto.7.7.15491
- Dunlop, E. A., and Tee, A. R. (2013). The kinase triad, AMPK, mTORC1 and ULK1, maintains energy and nutrient homeostasis. *Biochem. Soc. Trans.* 41, 939–943. doi: 10.1042/BST20130030
- Gao, W., Shen, Z., Shang, L., and Wang, X. (2011). Upregulation of human autophagy-initiation kinase ULK1 by tumor suppressor p53 contributes to DNA-damage-induced cell death. *Cell Death Differ.* 18, 1598–1607. doi: 10.1038/cdd.2011.33
- Gems, D., and Partridge, L. (2013). Genetics of longevity in model organisms: debates and paradigm shifts. *Annu. Rev. Physiol.* 75, 621–644. doi: 10.1146/annurev-physiol-030212-183712

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

JW designed the research. WL, LZ, and JW performed the research, analyzed the data, and wrote the paper. All authors contributed to the article and approved the submitted version.

FUNDING

WL was supported by NSFC grant no. 21721003 and MOST China grant no. 2016YFA0203200.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00593/full#supplementary-material>

- Georgescu, M.-M. (2010). PTEN tumor suppressor network in PI3k-akt pathway control. *Genes Cancer* 1, 1170–1177. doi: 10.1177/1947601911407325
- Gharibi, B., Farzadi, S., Ghuman, M., and Hughes, F. J. (2014). Inhibition of akt/mTOR attenuates age-related changes in mesenchymal stem cells. *Stem Cells* 32, 2256–2266. doi: 10.1002/stem.1709
- Ghosh, H. S., McBurney, M., and Robbins, P. D. (2010). SIRT1 negatively regulates the mammalian target of rapamycin. *PLoS ONE* 5:e9199. doi: 10.1371/journal.pone.0009199
- Greer, E. L., Dowlathshahi, D., Banko, M. R., Villen, J., Hoang, K., Blanchard, D., et al. (2007a). An AMPK-FOXO pathway mediates longevity induced by a novel method of dietary restriction in *C. elegans*. *Curr. Biol.* 17, 1646–1656. doi: 10.1016/j.cub.2007.08.047
- Greer, E. L., Oskoui, P. R., Banko, M. R., Maniar, J. M., Gygi, M. P., Gygi, S. P., et al. (2007b). The energy sensor AMP-activated protein kinase directly regulates the mammalian FOXO3 transcription factor. *J. Biol. Chem.* 282, 30107–30119. doi: 10.1074/jbc.M705325200
- Gwinn, D. M., Shackelford, D. B., Egan, D. F., Mihaylova, M. M., Mery, A., Vasquez, D. S., et al. (2008). AMPK phosphorylation of raptor mediates a metabolic checkpoint. *Mol. Cell* 30, 214–226. doi: 10.1016/j.molcel.2008.03.003
- Haruta, T., Uno, T., Kawahara, J., Takano, A., Egawa, K., Sharma, P. M., et al. (2000). A rapamycin-sensitive pathway down-regulates insulin signaling via phosphorylation and proteasomal degradation of insulin receptor substrate-1. *Mol. Endocrinol.* 14, 783–794. doi: 10.1210/mend.14.6.0446
- Hay, N. (2011). Interplay between FOXO, TOR, and akt. *Biochim. Biophys. Acta* 1813, 1965–1970. doi: 10.1016/j.bbamcr.2011.03.013
- Hong, T., Oguz, C., and Tyson, J. J. (2015). A mathematical framework for understanding four-dimensional heterogeneous differentiation of CD4⁺ t cells. *Bull. Math. Biol.* 77, 1046–1064. doi: 10.1007/s11538-015-0076-6
- Hong, T., Xing, J., Li, L., and Tyson, J. J. (2011). A mathematical model for the reciprocal differentiation of t helper 17 cells and induced regulatory t cells. *PLoS Comput. Biol.* 7:e1002122. doi: 10.1371/journal.pcbi.1002122
- Hong, T., Xing, J., Li, L., and Tyson, J. J. (2012). A simple theoretical framework for understanding heterogeneous differentiation of CD4⁺ t cells. *BMC Syst. Biol.* 6:66. doi: 10.1186/1752-0509-6-66

- Huber, T. B., Edelstein, C. L., Hartleben, B., Inoki, K., Jiang, M., Koya, D., et al. (2012). Emerging role of autophagy in kidney function, diseases and aging. *Autophagy* 8, 1009–1031. doi: 10.4161/auto.19821
- Inoki, K., Li, Y., Zhu, T., Wu, J., and Guan, K.-L. (2002). TSC2 is phosphorylated and inhibited by akt and suppresses mTOR signalling. *Nat. Cell Biol.* 4, 648–657. doi: 10.1038/ncb839
- Inoki, K., Zhu, T., and Guan, K.-L. (2003). TSC2 mediates cellular energy response to control cell growth and survival. *Cell* 115, 577–590. doi: 10.1016/S0092-8674(03)00929-2
- Jin, M., Li, Y., O’Laughlin, R., Bittihn, P., Pillus, L., Tsimring, L. S., et al. (2019). Divergent aging of isogenic yeast cells revealed through single-cell phenotypic dynamics. *Cell Syst.* 8, 242–253.e3. doi: 10.1016/j.cels.2019.02.002
- Johnson, S. C., Rabinovitch, P. S., and Kaerberlein, M. (2013). mTOR is a key modulator of ageing and age-related disease. *Nature* 493, 338–345. doi: 10.1038/nature11861
- Kapahi, P., Chen, D., Rogers, A. N., Katewa, S. D., Li, P. W.-L., Thomas, E. L., et al. (2010). With TOR, less is more: a key role for the conserved nutrient-sensing TOR pathway in aging. *Cell Metab.* 11, 453–465. doi: 10.1016/j.cmet.2010.05.001
- Kenyon, C., Chang, J., Gensch, E., Rudner, A., and Tabtiang, R. (1993). A *C. elegans* mutant that lives twice as long as wild type. *Nature* 366, 461–464. doi: 10.1038/366461a0
- Kenyon, C. J. (2010). The genetics of ageing. *Nature* 464, 504–512. doi: 10.1038/nature08980
- Kim, J., Kundu, M., Viollet, B., and Guan, K.-L. (2011). AMPK and mTOR regulate autophagy through direct phosphorylation of ulk1. *Nat. Cell Biol.* 13, 132–141. doi: 10.1038/ncb2152
- Kong, M. (2004). The PP2a-associated protein 4 is an essential inhibitor of apoptosis. *Science* 306, 695–698. doi: 10.1126/science.1100537
- Lahav, G., Rosenfeld, N., Sigal, A., Geva-Zatorsky, N., Levine, A. J., Elowitz, M. B., et al. (2004). Dynamics of the p53-mdm2 feedback loop in individual cells. *Nat. Genet.* 36, 147–150. doi: 10.1038/ng1293
- Lan, F., Cacicado, J. M., Ruderman, N., and Ido, Y. (2008). SIRT1 modulation of the acetylation status, cytosolic localization, and activity of LKB1. *J. Biol. Chem.* 283, 27628–27635. doi: 10.1074/jbc.M805711200
- Lee, J. H., Budanov, A. V., and Karin, M. (2013). Sestrins orchestrate cellular metabolism to attenuate aging. *Cell Metab.* 18, 792–801. doi: 10.1016/j.cmet.2013.08.018
- Li, C. (2018). Landscape of gene networks for random parameter perturbation. *Integr. Biol.* 10, 92–99. doi: 10.1039/C7IB00198C
- Li, C., and Balazsi, G. (2018). A landscape view on the interplay between EMT and cancer metastasis. *NPJ Syst. Biol. Appl.* 4:34. doi: 10.1038/s41540-018-0068-x
- Löffler, A. S., Alers, S., Dieterle, A. M., Keppeler, H., Franz-Wachtel, M., Kundu, M., et al. (2011). Ulk1-mediated phosphorylation of AMPK constitutes a negative regulatory feedback loop. *Autophagy* 7, 696–706. doi: 10.4161/auto.7.7.15451
- López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M., and Kroemer, G. (2013). The hallmarks of aging. *Cell* 153, 1194–1217. doi: 10.1016/j.cell.2013.05.039
- Martins, R., Lithgow, G. J., and Link, W. (2015). Long live FOXO: unraveling the role of FOXO proteins in aging and longevity. *Aging Cell* 15, 196–207. doi: 10.1111/accel.12427
- McCormick, M., Chen, K., Ramaswamy, P., and Kenyon, C. (2011). New genes that extend *Caenorhabditis elegans* lifespan in response to reproductive signals. *Aging Cell* 11, 192–202. doi: 10.1111/j.1474-9726.2011.00768.x
- Mjølness, E., Sharp, D. H., and Reinitz, J. (1991). A connectionist model of development. *J. Theor. Biol.* 152, 429–453. doi: 10.1016/S0022-5193(05)80391-1
- Nemoto, S. (2004). Nutrient availability regulates SIRT1 through a forkhead-dependent pathway. *Science* 306, 2105–2108. doi: 10.1126/science.1101731
- Ogawara, Y., Kishishita, S., Obata, T., Isazawa, Y., Suzuki, T., Tanaka, K., et al. (2002). Akt enhances mdm2-mediated ubiquitination and degradation of p53. *J. Biol. Chem.* 277, 21843–21850. doi: 10.1074/jbc.M109745200
- Okoshi, R., Ozaki, T., Yamamoto, H., Ando, K., Koida, N., Ono, S., et al. (2007). Activation of AMP-activated protein kinase induces p53-dependent apoptotic cell death in response to energetic stress. *J. Biol. Chem.* 283, 3979–3987. doi: 10.1074/jbc.M705232200
- Parmigiani, A., Nourbakhsh, A., Ding, B., Wang, W., Kim, Y. C., Akopiants, K., et al. (2014). Sestrins inhibit mTORC1 kinase activation through the GATOR complex. *Cell Rep.* 9, 1281–1291. doi: 10.1016/j.celrep.2014.10.019
- Powers, E. T., Morimoto, R. I., Dillin, A., Kelly, J. W., and Balch, W. E. (2009). Biological and chemical approaches to diseases of proteostasis deficiency. *Annu. Rev. Biochem.* 78, 959–991. doi: 10.1146/annurev.biochem.052308.114844
- Rahman, I., and Bagchi, D. (2013). *Inflammation, Advancing Age and Nutrition: Research and Clinical Interventions*. London: Elsevier Science.
- Rando, T. A., and Chang, H. Y. (2012). Aging, rejuvenation, and epigenetic reprogramming: resetting the aging clock. *Cell* 148, 46–57. doi: 10.1016/j.cell.2012.01.003
- Renault, V. M., Thekkat, P. U., Hoang, K. L., White, J. L., Brady, C. A., Broz, D. K., et al. (2011). The pro-longevity gene FoxO3 is a direct target of the p53 tumor suppressor. *Oncogene* 30, 3207–3221. doi: 10.1038/onc.2011.35
- Ruderman, N. B., Xu, X. J., Nelson, L., Cacicado, J. M., Saha, A. K., Lan, F., et al. (2010). AMPK and SIRT1: a long-standing partnership? *Am. J. Physiol. Endocrinol. Metab.* 298, E751–E760. doi: 10.1152/ajpendo.00745.2009
- Rufini, A., Tucci, P., Celardo, I., and Melino, G. (2013). Senescence and aging: the critical roles of p53. *Oncogene* 32, 5129–5143. doi: 10.1038/onc.2012.640
- Salih, D. A., and Brunet, A. (2008). FoxO transcription factors in the maintenance of cellular homeostasis during aging. *Curr. Opin. Cell Biol.* 20, 126–136. doi: 10.1016/j.ceb.2008.02.005
- Salminen, A., and Kaarniranta, K. (2012). AMP-activated protein kinase (AMPK) controls the aging process via an integrated signaling network. *Ageing Res. Rev.* 11, 230–241. doi: 10.1016/j.arr.2011.12.005
- Sengupta, S., Peterson, T. R., and Sabatini, D. M. (2010). Regulation of the mTOR complex 1 pathway by nutrients, growth factors, and stress. *Mol. Cell* 40, 310–322. doi: 10.1016/j.molcel.2010.09.026
- Smith-Vikos, T., de Lencastre, A., Inukai, S., Shlomchik, M., Holtrup, B., and Slack, F. J. (2014). MicroRNAs mediate dietary-restriction-induced longevity through PHA-4/FOXO and SKN-1/nrf transcription factors. *Curr. Biol.* 24, 2238–2246. doi: 10.1016/j.cub.2014.08.013
- Stambolic, V., MacPherson, D., Sas, D., Lin, Y., Snow, B., Jang, Y., et al. (2001). Regulation of PTEN transcription by p53. *Mol. Cell* 8, 317–325. doi: 10.1016/S1097-2765(01)00323-9
- Tyson, J. J., and Novák, B. (2010). Functional motifs in biochemical reaction networks. *Annu. Rev. Phys. Chem.* 61, 219–240. doi: 10.1146/annurev.physchem.012809.103457
- Waddington, C. (2014). *The Strategy of the Genes*. London: Routledge.
- Wang, F., Chan, C.-H., Chen, K., Guan, X., Lin, H.-K., and Tong, Q. (2011). Deacetylation of FOXO3 by SIRT1 or SIRT2 leads to skp2-mediated FOXO3 ubiquitination and degradation. *Oncogene* 31, 1546–1557. doi: 10.1038/onc.2011.347
- Wang, F., Nguyen, M., Qin, F. X.-F., and Tong, Q. (2007). SIRT2 deacetylates FOXO3a in response to oxidative stress and caloric restriction. *Aging Cell* 6, 505–514. doi: 10.1111/j.1474-9726.2007.00304.x
- Wang, J. (2015). Landscape and flux theory of non-equilibrium dynamical systems with application to biology. *Adv. Phys.* 64, 1–137. doi: 10.1080/00018732.2015.1037068
- Wang, J., Xu, L., and Wang, E. (2008). Potential landscape and flux framework of nonequilibrium networks: robustness, dissipation, and coherence of biochemical oscillations. *Proc. Natl. Acad. Sci. U.S.A.* 105, 12271–12276. doi: 10.1073/pnas.0800579105
- Wang, J., Zhang, K., and Wang, E. (2010). Kinetic paths, time scale, and underlying landscapes: a path integral framework to study global natures of nonequilibrium systems and networks. *J. Chem. Phys.* 133:125103. doi: 10.1063/1.3478547
- Wang, J., Zhang, K., Xu, L., and Wang, E. (2011). Quantifying the waddington landscape and biological paths for development and differentiation. *Proc. Natl. Acad. Sci. U.S.A.* 108, 8257–8262. doi: 10.1073/pnas.1017017108
- Watanabe, K., Panchy, N., Noguchi, S., Suzuki, H., and Hong, T. (2019). Combinatorial perturbation analysis reveals divergent regulations of mesenchymal genes during epithelial-to-mesenchymal transition. *NPJ Syst. Biol. Appl.* 5:21. doi: 10.1038/s41540-019-0097-0
- Yi, J., and Luo, J. (2010). SIRT1 and p53, effect on cancer, senescence and beyond. *Biochim. Biophys. Acta* 1804, 1684–1689. doi: 10.1016/j.bbapap.2010.05.002
- You, H., Yamamoto, K., and Mak, T. W. (2006). Regulation of transactivation-independent proapoptotic activity of p53 by FOXO3a. *Proc. Natl. Acad. Sci. U.S.A.* 103, 9051–9056. doi: 10.1073/pnas.0600889103

- Yuan, Y., Cruzat, V. F., Newsholme, P., Cheng, J., Chen, Y., and Lu, Y. (2016). Regulation of SIRT1 in aging: roles in mitochondrial function and biogenesis. *Mech. Ageing Dev.* 155, 10–21. doi: 10.1016/j.mad.2016.02.003
- Zhao, L., and Wang, J. (2016). Uncovering the mechanisms of caenorhabditis elegans ageing from global quantification of the underlying landscape. *J. R. Soc. Interface* 13:20160421. doi: 10.1098/rsif.2016.0421
- Zhou, P., and Li, T. (2016). Construction of the landscape for multi-stable systems: potential landscape, quasi-potential, a-type integral and beyond. *J. Chem. Phys.* 144:094109. doi: 10.1063/1.4964682

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Li, Zhao and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Stochastic Dynamics of Gene Switching and Energy Dissipation for Gene Expression

Quan Liu¹, FengZhen Yu², Liang Yi³, Yijun Gao¹, Rong Gui¹, Ming Yi⁴ and Jianqiang Sun^{5*}

¹ Department of Physics, College of Science, Huazhong Agricultural University, Wuhan, China, ² Department of Chemistry, College of Science, Huazhong Agricultural University, Wuhan, China, ³ College of Animal Science and Technology and College of Veterinary Medicine, Huazhong Agricultural University, Wuhan, China, ⁴ School of Mathematics and Physics, China University of Geosciences, Wuhan, China, ⁵ School of Automation and Electrical Engineering, Linyi University, Linyi, China

OPEN ACCESS

Edited by:

Chunhe Li,
Fudan University, China

Reviewed by:

Jiajun Zhang,
Sun Yat-sen University, China

Bin Zhang,
Massachusetts Institute of
Technology, United States

*Correspondence:

Jianqiang Sun
sunjianqiang@lyu.edu.cn

Specialty section:

This article was submitted to
Systems Biology,
a section of the journal
Frontiers in Genetics

Received: 11 March 2020

Accepted: 02 June 2020

Published: 02 July 2020

Citation:

Liu Q, Yu F, Yi L, Gao Y, Gui R, Yi M
and Sun J (2020) Stochastic
Dynamics of Gene Switching and
Energy Dissipation for Gene
Expression. *Front. Genet.* 11:676.
doi: 10.3389/fgene.2020.00676

Stochastic dynamics of gene switching and energy dissipation for gene expression are largely unknown, mainly due to the complexity of non-equilibrium mechanisms. Here, based on an important double-deck loop model, the stochastic mechanisms of gene switching and energy dissipation are studied. First, the probability distributions of steady states are calculated theoretically. It is found that the signal can strengthen the choice of gene switching between the “off” and “on” states. Our analysis of energy consumption illustrates that, compared with the synthesis and degradation of proteins, the process of gene switching costs little energy. Our theoretical analysis reveals some interesting insights into the determination of cell state and energy dissipation for gene expression.

Keywords: non-equilibrium mechanisms, stochastic dynamics, gene switching, energy dissipation, chemical master equations

1. INTRODUCTION

Signal pathways play vital roles in life by cooperating to control more than one biochemical process while consuming free energy supplied by ATP or high-energy bonds to carry out different vital functions. Based on the core negative feedback control loop shared by various adaption biological systems, Lan et al. show that energy dissipation is indicated to stabilize the adapted state against noise (Lan et al., 2012). Further study explores the present analytic results on the non-equilibrium steady-state (NESS) of the model through mapping to a one-dimensional birth-death process, and the result suggests that the adaptation error can be reduced exponentially as the methylation range increases (Wang et al., 2015). In recent research, the number of phase coherent periods is found to be proportional to the free energy consumed per period (Cao et al., 2015). Increasingly numerous theoretical studies focus on the role of energy in biological information processes and biochemical signal transduction (Lan and Tu, 2013; Endres, 2017).

Biological information processes are complex. In the process of skeletal development, extracellular signals activate RhoA, and control the state of downstream genes mainly through two pathways: RhoA/SRF and RhoA/ROCK (Charrasse et al., 2002; Sordella et al., 2003; Tsai et al., 2013; Matsuoka and Yashiro, 2014). The marvelous phenomenon in this signal cascade is that those two pathways exert the opposite effects, as shown in **Figure 1**. Hence, such a signal cascade is composed of two competitive pathways and possesses the ability to accurately control the vital bio-processes (Wei et al., 1998, 2000, 2001; Meriane et al., 2000; Beqaj et al., 2002; Castellani et al., 2006; Charrasse et al., 2006). The study of stochastic dynamics and energy dissipation in this biological system represents an interesting topic. The competitive networks may present two

kinetic characteristics: oscillation and bistable state. Ouyang et al. show a series of works on the non-equilibrium thermodynamics of oscillations within cells and their results revealed that the critical energy dissipation per period depends on both the frequency and strength of the exchange reaction which gave an optimal design for achieving maximum synchronization with a fixed energy budget (Cao et al., 2015; Fei et al., 2018; Zhang et al., 2020). Gene switch as another kind of competitive networks is a representative bistable state system and will be the focus of our study.

In this paper, we propose a double-deck loop (DDL) model to describe signal cascades which are similar to those found in the development of skeletal muscle, as shown in **Figure 2**. By virtue of non-equilibrium statistical physics theory and stochastic dynamics, exact analytical solutions to the steady-state probability distribution are obtained and energy dissipation in the DDL is derived, which allows for many deeper discussions. Our aims are to reveal (i) the crucial factors that determine the state of gene switching in our model and (ii) the energy dissipation in biochemical reactions. We expect that these

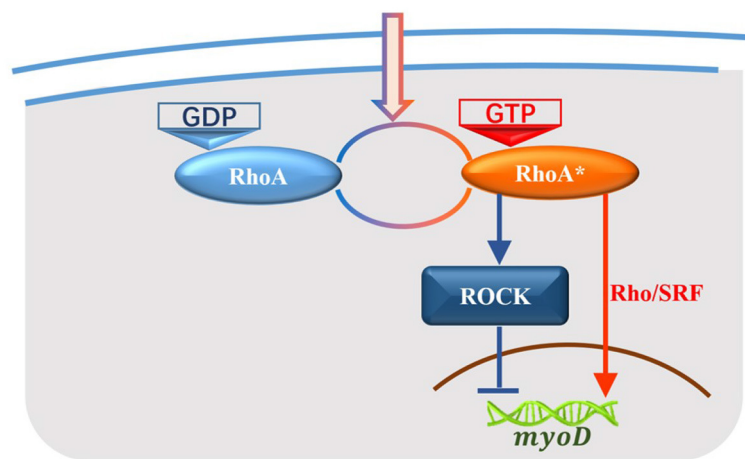


FIGURE 1 | Signal pathways from RhoA to the target gene in the development of skeletal muscle. RhoA-GDP is activated by extracellular signals to RhoA-GTP (RhoA*) while RhoA* is deactivated back to RhoA-GDP as well. RhoA* contributes to the development of skeletal muscles through two signal pathways: RhoA/ROCK regulates the target gene *myoD* negatively which is corresponding to the blue cascade, and RhoA/SRF regulates *myoD* positively which is corresponding to the red cascade.

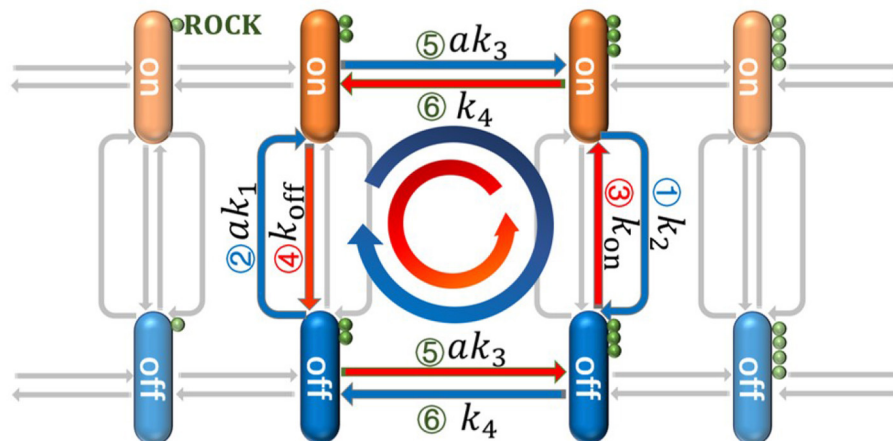


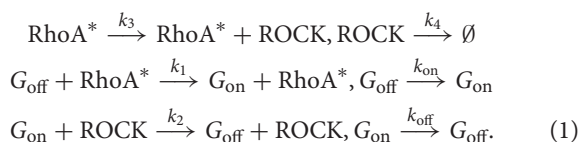
FIGURE 2 | The detailed cascades in the double-deck loop (DDL) which is a simplified model based on the biochemical reaction networks in **Figure 1**. The orange stick is the gene in the "on" state and the blue one is the gene in the "off" state; ROCK is represented as the small green ball; Arrow 1 corresponds to the Rho/ROCK pathway in **Figure 1** and Arrow 2 corresponds to the Rho/SRF pathway. The negative regulation from ROCK can push down the state of gene from the "on" state to the "off" state along Arrow 1 with rate k_2 . The positive regulation from RhoA* can also push up the state of gene along Arrow 2 with rate ak_1 . The Arrow 3 and 4 represent the basic switching of gene's state with rates k_{on} and k_{off} respectively. The synthesis and degradation of ROCK will drive the system move along the horizontal directions (i.e., Arrow 5 and 6) with rates ak_3 and k_4 , respectively. Since two parallel loops (i.e., the blue and red loops) can be found in our model, it is so called DDL.

theoretical results could help us to understand the general principles of signaling selectivity and energy dissipation in gene regulation networks.

2. THE MODEL

In the process of skeletal development, RhoA is activated by extracellular signals from RhoA-GDP to RhoA-GTP (RhoA*), and RhoA* is deactivated back to RhoA-GDP as well. RhoA* plays a key role in process of skeletal development and contributes to the regulation of the expression of muscle-specific genes both through RhoA/SRF and RhoA/ROCK signal pathways. RhoA/ROCK pathway triggers a negative control function on the target gene while RhoA/SRF pathway induces a positive regulation. The relevant biochemical processes are shown in **Figure 1**. The red signal cascade represents RhoA/SRF pathway and the blue one represents RhoA/ROCK pathway.

To derive our mathematical model, some assumptions are put forward. (i) The switching dynamics between GTP and GDP have been discussed thoroughly (Lan and Tu, 2013) which has the same dynamic mechanism with the switching between RhoA* and RhoA-GDP. This aspect is neglected in our model, since we focus on the selectivity between different gene's state modes and energy dissipation of RhoA/ROCK and RhoA/SRF signal pathways. These two pathways are adopted to regulate the development of skeletal muscle. (ii) It is well-known that a large number of genes are involved in this biological process (Matsui et al., 1996). If all genes are considered, it will complicate the modeling and theoretical analysis. We hypothesize that the development state of skeletal muscle can be represented with the state of gene *myoD*. Muscle-specific genes begin to be expressed when the state of gene *myoD* is “on”, otherwise these genes are closed when the state of gene *myoD* is “off”. (iii) Since the role of ROCK in biological activities is vital (Leung et al., 1995; Aelst and D'Souza-Schorey, 1997; Kaibuchi et al., 1999; Cloutier et al., 2010), the detailed biochemical process of RhoA/SRF pathway is neglected and the regulation of RhoA/SRF pathway is simplified to the direct regulation of RhoA* on gene *myoD* as shown in **Figure 1**. Therefore, we mainly aim to discuss the biochemical reactions including the switch of *myoD* state and synthesis/degradation of ROCK. Based on the above assumptions, the detailed biochemical equations are as follows:



The “on” state of *myoD* is indicated as G_{on} in Equation (1), while the “off” one is G_{off} . k_1 is the transition rate of gene state under the positive control and k_2 is the transition rate of gene under the negative control. The basic switching rates between G_{on} and G_{off} are k_{off} and k_{on} . k_3 is

the synthesis rate of ROCK and k_4 is the degradation rate of it.

As shown in Equation (1), every state of gene can be achieved through two ways: G_{on} can be achieved both through the promotion of RhoA* and the basic switching; G_{off} can be achieved both through the repressive control of ROCK and the basic switching. Considering the synthesis and degradation of ROCK which correspond to the increasing and decreasing of the small green balls in **Figure 2**, two parallel loops can be found. One is clockwise [i.e., $G_{\text{on}}(m) \xrightarrow{ak_3} G_{\text{on}}(m+1) \xrightarrow{k_2} G_{\text{off}}(m+1) \xrightarrow{k_4} G_{\text{off}}(m) \xrightarrow{ak_1} G_{\text{on}}(m)$] which is represented as a blue loop in **Figure 2**. The other one is anticlockwise [i.e., $G_{\text{on}}(m) \xrightarrow{k_{\text{off}}} G_{\text{off}}(m) \xrightarrow{ak_3} G_{\text{off}}(m+1) \xrightarrow{k_{\text{on}}} G_{\text{on}}(m+1) \xrightarrow{k_4} G_{\text{on}}(m)$] which is represented as red loop in **Figure 2**. Since these two loops are two parallel decks between G_{on} and G_{off} , this theoretical model is called “double-deck loop (DDL)” in this paper. Moreover, m is the number of ROCK, and a is the concentration of RhoA*. Based on above biochemical equations, the chemical master equations (i.e., CME) can be presented as

$$\begin{aligned} \frac{dP_0(m, t)}{dt} &= k_{\text{off}}P_1(m, t) - k_{\text{on}}P_0(m, t) + mk_2P_1(m, t) \\ &\quad - ak_1P_0(m, t) + ak_3P_0(m-1, t) - ak_3P_0(m, t) \\ &\quad + (m+1)k_4P_0(m+1, t) - mk_4P_0(m, t); \\ \frac{dP_1(m, t)}{dt} &= -k_{\text{off}}P_1(m, t) + k_{\text{on}}P_0(m, t) - mk_2P_1(m, t) \\ &\quad + ak_1P_0(m, t) + ak_3P_1(m-1, t) - ak_3P_1(m, t) \\ &\quad + (m+1)k_4P_1(m+1, t) - mk_4P_1(m, t) \quad (2) \end{aligned}$$

where $P_0(m, t)$ is the probability of the gene “off” state and $P_1(m, t)$ are the probability of the gene “on” states.

3. RESULTS

3.1. The Analytical Solutions of the DDL Model

Regulatory networks generally consist of interactional signal pathways. Different signal pathways may dominate cell fate in different circumstances. Based on the biology processes in the development of skeletal (Wei et al., 1998, 2000, 2001; Meriane et al., 2000; Beqaj et al., 2002; Castellani et al., 2006; Charrasse et al., 2006), the DDL model has both activation and inhibitory signal pathways originating from the same input. Therefore, it is interesting to identify the crucial factors in the selection of different signaling pathways. To address this problem, the probability distributions under different a are derived. In response to a persistent input, the synthesis and degradation of ROCK induce the evolution of ROCK, resulting in a steady state of the number of ROCK m . Using the method of probability-generating functions (Qian, 2007; Huang et al., 2015), the analytical expressions of steady-state probability distributions

$P_0(m)$ and $P_1(m)$ are obtained

$$\begin{aligned}
 P_0(m) &= \frac{1}{k_{\text{on}} + ak_1} \left(\frac{ak_2k_3}{k_2 + 1} + k_{\text{off}} \right) A_0 \frac{1}{m!} \left(\frac{ak_3}{k_2 + 1} \right)^m \\
 &\quad \times \sum_{l=0}^m k_2^l C_m^l \frac{(\alpha)_l}{(\beta)_l} {}_1F_1(\alpha + l, \beta + l; \omega_2) \\
 &\quad - \frac{1}{k_{\text{on}} + ak_1} \frac{ak_2k_3}{k_2 + 1} \frac{\alpha}{\beta} A_0 \frac{1}{m!} \left(\frac{ak_3}{k_2 + 1} \right)^m \\
 &\quad \times \sum_{l=0}^m k_2^l C_m^l \frac{(\alpha + 1)_l}{(\beta + 1)_l} {}_1F_1(\alpha + 1 + l, \beta + 1 + l; \omega_2) \\
 &\quad + \frac{ak_2k_3}{k_{\text{on}} + ak_1} \frac{\alpha}{\beta} A_0 \frac{1}{(m-1)!} \left(\frac{ak_3}{k_2 + 1} \right)^{m-1} \\
 &\quad \times \sum_{l=0}^{m-1} k_2^l C_{m-1}^l \frac{(\alpha + 1)_l}{(\beta + 1)_l} {}_1F_1(\alpha + 1 + l, \beta + 1 + l; \omega_2); \\
 P_1(m) &= A_0 \frac{1}{m!} \left(\frac{ak_3}{k_2 + 1} \right)^m \\
 &\quad \times \sum_{l=0}^m k_2^l C_m^l \frac{(\alpha)_l}{(\beta)_l} {}_1F_1(\alpha + l, \beta + l; \omega_2) \quad (3)
 \end{aligned}$$

where $\alpha = ak_2k_3/(k_2 + 1)^2 + (k_{\text{on}} + k_{\text{off}} + ak_1)/(k_2 + 1)$, $\beta = \alpha + 1$, and $\omega_2 = -ak_2k_3/(k_2 + 1)^2$. Here, all the parameters are normalized by k_4 , i.e., $k_i = k_i/k_4$ which is different with Equation (2). C_m^l is the binomial coefficient in Equation (3), $(\gamma)_l$ is the Pochhammer symbol defined as $(\gamma)_l = \Gamma(\gamma + l)/\Gamma(\gamma)$ with $\Gamma(\gamma)$ being the Gamma function, and ${}_1F_1(\alpha, \beta; \omega_2)$ is a confluent hypergeometric function (Huang et al., 2015). A_0 is the normalization constant as follows

$$\begin{aligned}
 A_0^{-1} &= e^{\frac{ak_3}{k_2+1}} \left[\frac{1}{k_{\text{on}} + ak_1} \left(\frac{ak_2k_3}{k_2 + 1} + k_{\text{off}} \right) {}_1F_1(\alpha, \beta; \omega_1) \right. \\
 &\quad - \frac{1}{k_{\text{on}} + ak_1} \frac{ak_2k_3}{k_2 + 1} \frac{\alpha}{\beta} {}_1F_1(\alpha + 1, \beta + 1; \omega_1) \\
 &\quad \left. + \frac{ak_2k_3}{k_{\text{on}} + ak_1} \frac{\alpha}{\beta} {}_1F_1(\alpha + 1, \beta + 1; \omega_1) \right] + e^{\frac{ak_3}{k_2+1}} {}_1F_1(\alpha, \beta; \omega_1) \quad (4)
 \end{aligned}$$

where ω_1 is a constant with the expression $\omega_1 = ak_2k_3/(k_2 + 1) - ak_2k_3/(k_2 + 1)^2$. The details of analysis for chemical master equations are presented in the **Supplementary Material** of this paper. The above results will be checked through the structure of our model in the following part.

The character of signaling cascades in the development of skeletal muscle can provide us with information to verify our analytical solutions. The changes of ROCK, as an upstream component of the signaling cascades, follow a basic process of birth and death. The statistical law of a birth and death process is that the probability distribution about m is a standard Poisson distribution. In order to test this, we calculate the total probability $P(m)$ which is provided by $P_0(m) + P_1(m)$. This represents the statistical law of ROCK and can be simplified from Equation (3) as:

$$P(m) = \frac{1}{m!} (ak_3)^m e^{-ak_3}. \quad (5)$$

This is a standard Poisson distribution. Furthermore, ROCK should be in its steady state $m_s = ak_3$ most of the time. This means that the peaks of $P_0(m)$ and $P_1(m)$ focus on m_s . According to Equation (3), the values of $P_0(m)$ and $P_1(m)$ with different parameters are computed through “Mathematica” which are shown in **Figure 3**. It is obvious that the peaks of distributions occur at the steady state value of m_s . It's worth mentioning that the roughness of the curves shown in **Figures 3a,c,d** is caused by the computational accuracy of “Mathematica” when it is used to calculate the confluent hypergeometric function rather than biological or physical factors. We can also get these curves through “Matlab” which appear very smooth. However, compared with “Mathematica”, “Matlab” fails to calculate confluent hypergeometric function when m is too large. In order to verify these curves, the corresponding curves with the same parameters obtained by Monte Carlo simulation are shown in **Figures S1a,b**. It is obviously that the curves in **Figures S1a,b** closely resemble the ones in **Figures 3a,b**, respectively.

The above discussion confirms the reliability of our theoretical results in Equation (3). Based on those results, we will try to explore the selectivity of different pathways and energy consumption in the following section.

3.2. Stochastic Dynamics of Gene Switching in DDL

The steady-state probability distributions with different stimulation strength are displayed in **Figure 3**. $\sum_m P_0(m)$ and $\sum_m P_1(m)$ which correspond to the areas under the curves of $P_0(m)$ and $P_1(m)$ are respectively the probabilities of the gene's “off” state and “on” state. We use them to define two gene modes: Mode I and Mode II. Mode I denotes that the probability of the gene's “off” state is larger than the one of the “on” state, and Mode II denotes that the probability of the gene's “on” state is larger than the one of the “off” state. As shown in **Figures 3a,b**, when $k_3 = 5$, the gene's state is Mode I. Conversely, if $k_3 = 1.5$, the gene's state becomes Mode II shown in **Figures 3c,d**. Compared with k_3 , even if the parameter a varies widely, the gene's state mode is not changed. **Figure 4A** shows the areas of these two modes. It is obviously that the boundary between them is almost a horizontal line where $k_3 \approx 2.1$. This means that the mode of the gene's state is determined primarily by k_3 and the parameter a has little effect on the selection of gene's state modes. Since k_3 and a represent the synthesis rate of the negative controller ROCK and the strength of external stimulations, respectively, the gene's state mode depends almost exclusively on the synthesis rate of the negative controller ROCK rather than the external stimulations.

By the definition of Mode I, the gene is more likely to be in the “off” state than the “on” state. In other words, the “off” state of gene is dominant in Mode I. Similarly, the “on” state of gene is dominant in Mode II. To quantify the dominance of the gene's state, we define a gene's state dominance factor $\delta = |P_{0,\text{max}} - P_{1,\text{max}}|/P_{1,\text{max}}$. The larger peak in $P_i(m)$ ($i = 1, 2$) means more dominance as shown in **Figure 3**. Therefore, δ can be used to measure the dominance of the gene's state. Next, we

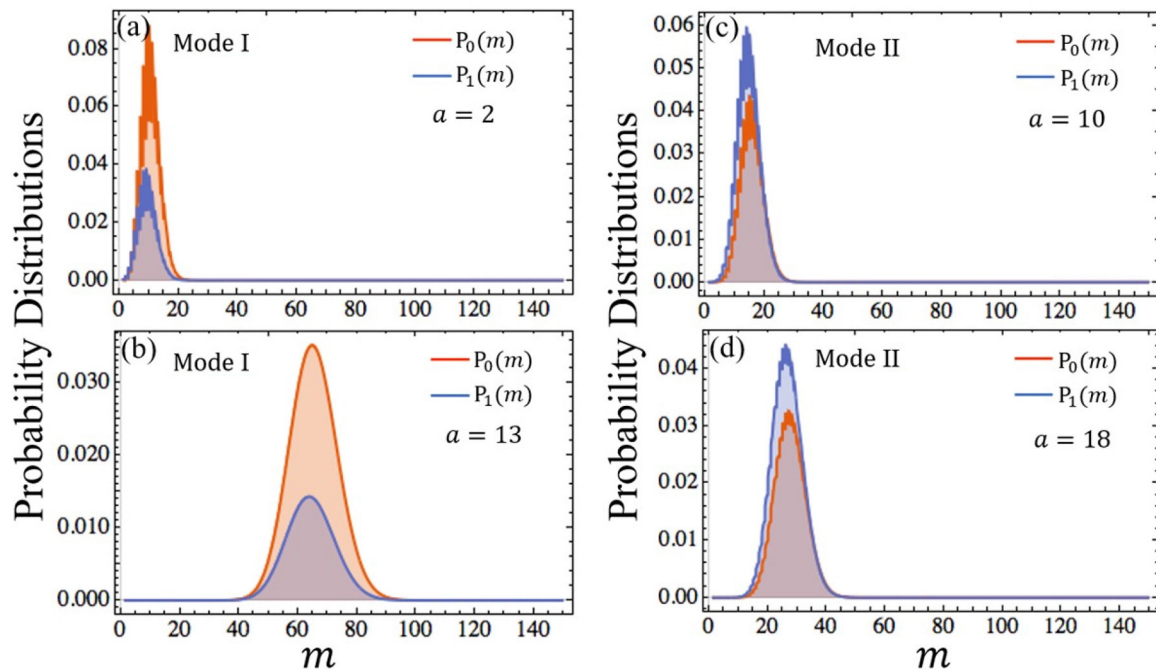


FIGURE 3 | Distributions of probability as a function of m . The synthetic rate of ROCK $k_3 = 5$ in (a,b) and $k_3 = 1.5$ in (c,d). The value of other parameters can be found in Table S1. $P_0(m)$ is the probability of the gene “off” state indicated with orange curve, $P_1(m)$ is the probability of gene “on” state indicated with blue curve.

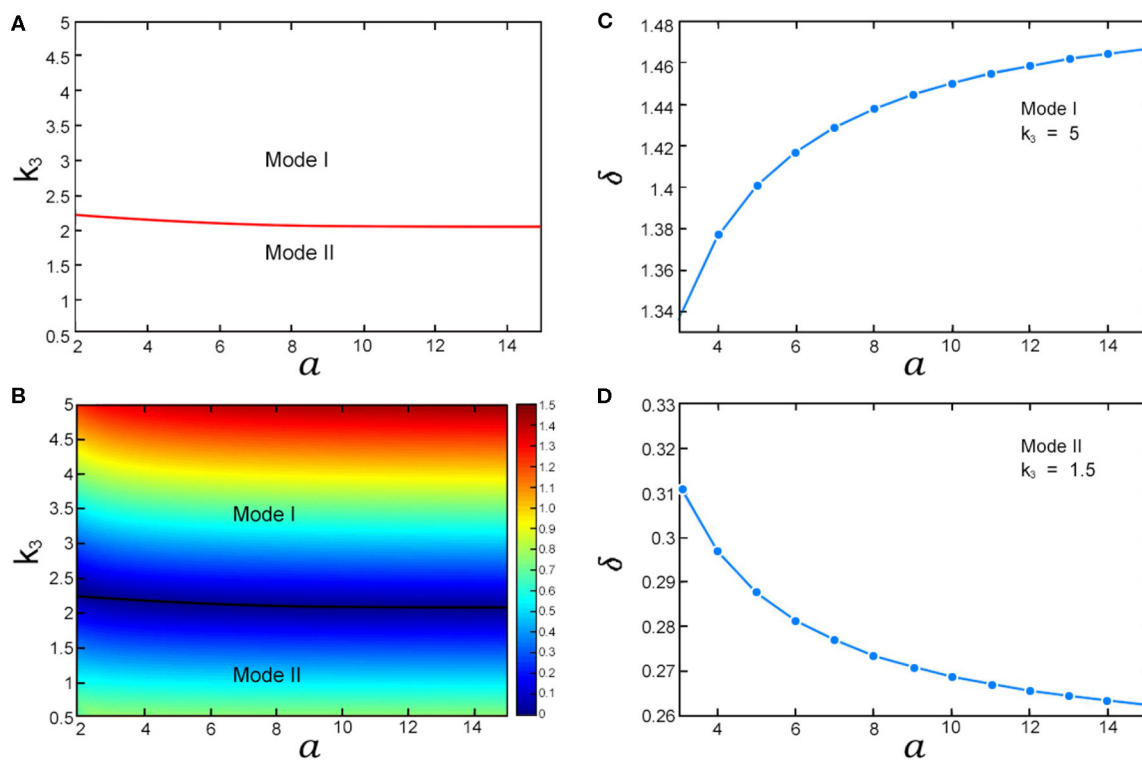


FIGURE 4 | The influences of k_3 and a on gene's “off” and “on” states. (A) The areas of Mode I and Mode II in a - k_3 plane; (B) The heat map of gene's state dominance factor δ in a - k_3 plane. The black line is the boundary between Mode I and Mode II; (C) The curve between δ and a in Mode I; (D) The curve between δ and a in Mode II. The values of other parameters are listed in Table S1.

will discuss the influences of the synthesis rate of the negative controller ROCK and the external stimulations on gene's state dominance. The values of δ in a - k_3 plane correspond to the scale values of the color bar in **Figure 4B**. As shown in **Figure 4B**, with increasing k_3 (i.e., the synthesis rate of the negative controller ROCK), δ (i.e., the dominance of the gene's "off" state) increases in Mode I. Conversely, as k_3 increases, δ decreases in Mode II. That suggests that when the synthesis rate of the negative controller ROCK increases, the dominance of the gene's "off" state is gradually weakening in Mode II, until the mode of the gene's state changes to Mode I and the dominance of the gene's "on" state increases gradually. Compared with k_3 , the influence of the strength of external stimulations RhoA* (i.e., a) on the dominance of the gene's state (i.e., δ) is not obvious enough in **Figure 4B**. Therefore, the relation between a and δ with different parameters k_3 which correspond respectively to Mode I and Mode II is shown in **Figures 4C,D** (Note that we also use Monte Carlo simulation to verify the trend of δ with a in Mode I which is shown in **Figure S2**). Similar to the case of k_3 , with increasing the strength of external stimulations RhoA* (i.e., a), the dominance of the gene's state (i.e., δ) increases in Mode I and decreases in Mode II. That means that although external stimulations RhoA* has little effect on the selection of gene's state modes, it can fine-tune the dominance of the gene's state in its respective modes.

In summary, the core factor for the stochastic dynamics of gene switching in DDL-type biochemical networks is identified in our work. Reaction rates are responsible for the selectivity between different gene's states, while the external signal stimulation fine-tunes the choice in its respective modes. The cooperation between signals and network maintains the vital process in an orderly manner.

3.3. Energy Dissipation in DDL

It is intuitively obvious that living biochemical systems need free energy (Gui et al., 2016, 2018). From the viewpoint of thermodynamics, gene expression is essentially a non-equilibrium process due to feedback or feedforward regulation that breaks detailed balance and thus necessarily consumes energy (Lu et al., 2017). But how is energy actually utilized during the regulation of gene expression in the development of skeletal muscle? To our knowledge, few works have touched upon this point. The composition of total energy consumption may help us grasp the selection mechanism between different biochemical processes.

From the definition of entropy $S(P_i) = k_B T \sum_i P_i \ln P_i$, the entropy production rate $\varepsilon_p(t)$ is given as follows (Ge and Qian, 2010):

$$\varepsilon_p(t) = \sum_{i,j} (P_i(t)q_{ij} - P_j(t)q_{ji}) \ln \left(\frac{q_{ij}}{q_{ji}} \right). \quad (6)$$

$P_i(t)$ is the probability of the system in state i at time t , while q_{ji} is the transport rate from state j to i . $k_B T$ is set to be 1 for convenience in our work. $\varepsilon_p(t)$ is the sum of energy dissipated in the biochemical network. Furthermore, the entropy

production rate for a non-equilibrium steady-state system can be calculated as

$$EP = \sum_{(\sigma, \sigma')} P(\sigma) k(\sigma, \sigma') \log \left(\frac{k(\sigma, \sigma')}{k(\sigma', \sigma)} \right) \quad (7)$$

where $k(\sigma, \sigma')$ is the transition probability from state σ to σ' .

Considering the detailed biochemical reactions in our model, we derive the EP of the DDL network as

$$\begin{aligned} EP = & \sum_m [P_0(m)ak_3 \ln \left(\frac{ak_3}{m+1} \right) + P_0(m)m \ln \left(\frac{m}{ak_3} \right) \\ & + P_0(m)(ak_1 + k_{on}) \ln \left(\frac{ak_1 + k_{on}}{k_{off} + mk_2} \right) \\ & + P_1(m)ak_3 \ln \left(\frac{ak_3}{m+1} \right) + P_1(m)m \ln \left(\frac{m}{ak_3} \right) \\ & + P_1(m)(k_{off} + mk_2) \ln \left(\frac{k_{off} + mk_2}{ak_1 + k_{on}} \right)]. \end{aligned} \quad (8)$$

Next, we discuss the influences of the strength of external stimulation on the total energy dissipation EP in Mode I and Mode II respectively. As shown in **Figure 5A**, EP increases and its rate of increase decreases with increasing a in Mode I. This suggests that when the strength of external stimulation increases, the system consumes more and more energy to response it. Compared with the small strengthen of external stimulation, there is less growth of energy for the system corresponding to the large one in Mode I. In contrast to Mode I, both EP and its reduction rate decrease in Mode II as a increases (**Figure 5B**). This means although the system requires less and less energy with the increase of the strengthen of external stimulation, a small amount of energy is still needed to sustain it in Mode II.

Since the control of *MyoD* gene expression can be divided into two parts: one is the synthesis and degradation of ROCK and the second is the switching of gene state, the process of energy dissipation can be decomposed into three state transitions: $(m-1, \text{off}) \rightleftharpoons (m, \text{off}) \rightleftharpoons (m+1, \text{off})$, $(m-1, \text{on}) \rightleftharpoons (m, \text{on}) \rightleftharpoons (m+1, \text{on})$ and $(m, \text{on}) \rightleftharpoons (m, \text{off})$. Here, (m, on) represents the state when the gene is "on" and the level of ROCK is m . Note that the first two formulas are related to the synthesis and degradation of ROCK. According to these three state transitions, the total energy dissipation EP is decomposed into three terms as follows:

$$\begin{aligned} EP_1 = & \sum_m [P_0(m)ak_3 \log \left(\frac{ak_3}{m+1} \right) + P_0(m)m \log \left(\frac{m}{ak_3} \right)] \\ EP_2 = & \sum_m [P_1(m)ak_3 \log \left(\frac{ak_3}{m+1} \right) + P_1(m)m \log \left(\frac{m}{ak_3} \right)] \\ EP_3 = & \sum_m [P_0(m)(ak_1 + k_{on}) \log \left(\frac{ak_1 + k_{on}}{k_{off} + mk_2} \right) \\ & + P_1(m)(k_{off} + mk_2) \log \left(\frac{k_{off} + mk_2}{ak_1 + k_{on}} \right)]. \end{aligned} \quad (9)$$

Since the synthetic rate of ROCK k_3 which has been normalized by its degradation rate k_4 just appears in the formulas of EP_1 and EP_2 , we define $EP_m = EP_1 + EP_2$ and use it to represent the energy dissipation in the synthesis-degradation process of

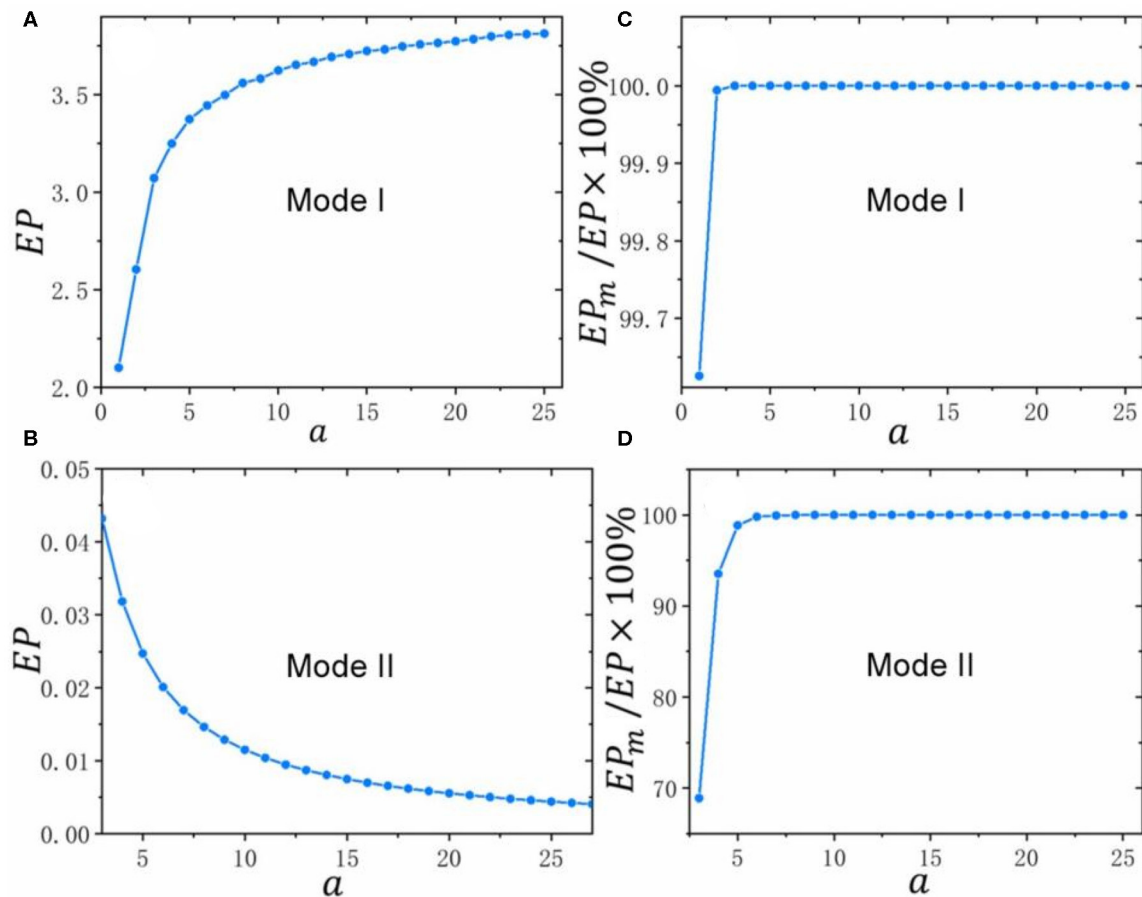


FIGURE 5 | The influences of the strength of external stimulation on energy dissipation in Mode I and II. The values of all parameters are listed in **Table S1**. **(A,B)** The relation curves between total energy dissipation EP and a in Mode I and II; **(C,D)** The relation curves between the percentages of energy dissipation in the synthesis-degradation process of ROCK (i.e., EP_m) in total energy dissipation (i.e., EP) and a in Mode I and II.

ROCK. In the following part, we will study the allocation of total energy dissipation corresponding to different strengthen of external stimulations by the comparison between EP and EP_m . The **Figure S3** shows the trends of the total energy dissipation (i.e., EP) and the energy dissipation in the synthesis-degradation process of ROCK (i.e., EP_m) with the increase of the strengthen of external stimulations (i.e., a) in Mode I and II. It is obviously that their trends are consistent as a increases in its respective modes. Specifically, EP and EP_m increase simultaneously in Mode I and decrease simultaneously in Mode II with the enhancing of the strengthen of external stimulations (i.e., a). Moreover, the difference between EP and EP_m diminishes both in Mode I and Mode II when a increases. The percentages of energy dissipation in the synthesis-degradation process of ROCK (i.e., EP_m) in total energy dissipation (i.e., EP) shown in **Figures 5C,D** are more than 60% both in Mode I and Mode II with different strengthen of external stimulations. In other words, the synthesis-degradation process of ROCK consume more energy than the third state transitions (i.e., EP_3) with different strengthen of external stimulations. Furthermore, **Figures 5C,D** also show that the percentage of energy which is consumed by the synthesis-degradation process of ROCK increases as the strengthen of external stimulations increases until almost no

energy is consumed in process of the third state transitions both in Mode I and Mode II.

4. DISCUSSION AND CONCLUSION

In our work, a double-deck loop model is constructed. Due to the stochastic nature of bio-processes (Wang et al., 2017; Yao et al., 2018a,b), we have calculated the steady-state probability distributions of ROCK protein through the method of probability-generating functions for chemical master equations. The crucial factors in the stochastic dynamics of gene switching are identified. It is found that the weights between different pathways (i.e., the internal reaction rates) in DDL are the key point governing the state of gene switching, while an external stimulus fine-tunes this choice preference. Furthermore, the energy consumption in DDL is also discussed. Our results show that most of the energy is required for synthesis and degradation of ROCK, however, a very small amount of energy consumption is required for the basic transition processes of downstream genes between “on” and “off” states. This is because the ROCK processes are not in equilibrium and do not follow detailed balance. But the inter-conversion between “on” state and “off”

state is indeed in equilibrium and follows detailed balance. In other words, the two terms in EP_3 defined in Equation (8) cancel each other out because of detailed balance. The theoretical findings about selectivity between different gene states and energy dissipation will be advantageous for our understanding of cell fate determination. Our next steps are to conduct closely related experiments about the development of skeletal muscle and to combine our theoretical study with experimental observations and data.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

QL, MY, and JS conceived and designed the work. QL, JS, FY, LY, YG, and RG carried out computer implementation and

theoretical analysis. QL, JS, FY, LY, YG, and RG interpreted the simulation results. MY supervised the project. QL, FY, LY, YG, JS, and RG wrote the original manuscript. QL, MY, and JS contributed to the writing of final manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China (Grant Nos. 11675060, 91730301, 11804111, and 11805091).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00676/full#supplementary-material>

REFERENCES

- Aelst, L. V., and D'Souza-Schorey, C. (1997). Rho GTPases and signaling networks. *Genes Dev.* 11, 2295–2322. doi: 10.1101/gad.11.18.2295
- Beqaj, S., Jakkaraju, S., Mattingly, R. R., Pan, D., and Schuger, L. (2002). High RhoA activity maintains the undifferentiated mesenchymal cell phenotype, whereas RhoA down-regulation by laminin-2 induces smooth muscle myogenesis. *J. Cell Biol.* 156, 893–903. doi: 10.1083/jcb.200107049
- Cao, Y., Wang, H., Ouyang, Q., and Tu, Y. (2015). The free-energy cost of accurate biochemical oscillations. *Nat. Phys.* 11, 772–778. doi: 10.1038/nphys3412
- Castellani, L., Salvati, E., Alema, S., and Falcone, G. (2006). Fine regulation of RhoA and rock is required for skeletal muscle differentiation. *J. Biol. Chem.* 281, 15249–15257. doi: 10.1074/jbc.M601390200
- Charrasse, S., Comunale, F., Grumbach, Y., Poulat, F., Blangy, A., and Gauthier-Rouviere, C. (2006). RhoA GTPase regulates M-cadherin activity and myoblast fusion. *Mol. Biol. Cell* 17, 749–759. doi: 10.1091/mbc.e05-04-0284
- Charrasse, S., Meriane, M., Comunale, F., Blangy, A., and Gauthier-Rouviere, C. (2002). N-cadherin-dependent cell-cell contact regulates rho GTPases and β -catenin localization in mouse C2C12 myoblasts. *J. Cell Biol.* 158, 953–965. doi: 10.1083/jcb.200202034
- Cloutier, M., Tremblay, M., and Piedboeuf, B. (2010). ROCK2 is involved in accelerated fetal lung development induced by *in vivo* lung distension. *Pediatr. Pulmonol.* 45, 966–976. doi: 10.1002/ppul.21266
- Endres, R. G. (2017). Entropy production selects nonequilibrium states in multistable systems. *Sci. Rep.* 7:14437. doi: 10.1038/s41598-017-14485-8
- Fei C., Cao Y., Ouyang, Q., and Tu, Y. (2018). Design principles for enhancing phase sensitivity and suppressing phase fluctuations simultaneously in biochemical oscillatory systems. *Nat. Commun.* 9:1434. doi: 10.1038/s41467-018-03826-4
- Ge, H., and Qian, H. (2010). Physical origins of entropy production, free energy dissipation, and their mathematical representations. *Phys. Rev. E* 81:051133. doi: 10.1103/PhysRevE.81.051133
- Gui, R., Li, Z., Hu, L., Cheng, G., Liu, Q., Xiong, J., et al. (2018). Noise decomposition algorithm and propagation mechanism in feed-forward gene transcriptional regulatory loop. *Chin. Phys. B* 27, 92–103. doi: 10.1088/1674-1056/27/2/028706
- Gui, R., Liu, Q., Yao, Y., Deng, H., Ma, C., Jia, Y., et al. (2016). Noise decomposition principle in a coherent feed-forward transcriptional regulatory loop. *Front. Physiol.* 30:600. doi: 10.3389/fphys.2016.00600
- Huang, L., Yuan, Z., Yu, J., and Zhou, T. (2015). Fundamental principles of energy consumption for gene expression. *Chaos* 25:123101. doi: 10.1063/1.4936670
- Kaibuchi, K., Kuroda, S., and Amano, M. (1999). Regulation of the cytoskeleton and cell adhesion by the rho family GTPases in mammalian cells. *Annu. Rev. Biochem.* 68, 459–486. doi: 10.1146/annurev.biochem.68.1.459
- Lan, G., Sartori, P., Neumann, S., Sourjik, V., and Tu, Y. (2012). The energy-speed-accuracy trade-off in sensory adaptation. *Nat. Phys.* 8, 422–428. doi: 10.1038/nphys2276
- Lan, G., and Tu, Y. (2013). The cost of sensitive response and accurate adaptation in networks with an incoherent type-1 feed-forward loop. *J. R. Soc. Interface* 10:20130489. doi: 10.1098/rsif.2013.0489
- Leung, T., Manser, E., Tan, L., and Lim, L. (1995). A novel serine/threonine kinase binding the Ras-related RhoA GTPase which translocates the kinase to peripheral membranes. *J. Biol. Chem.* 270, 29051–29054. doi: 10.1074/jbc.270.49.29051
- Lu, L., Jia, Y., Liu, W., and Yang, L. (2017). Mixed stimulus-induced mode selection in neural activity driven by high and low frequency current under electromagnetic radiation. *Complexity*. 2017:7628537. doi: 10.1155/2017/7628537
- Matsui, T., Amano, M., Yamamoto, T., Chihara, K., Nakafuku, M., Ito, M., et al. (1996). Rho-associated kinase, a novel serine/threonine kinase, as a putative target for small GTP binding protein rho. *EMBO J.* 15, 2208–2216. doi: 10.1002/j.1460-2075.1996.tb00574.x
- Matsuoka, T., and Yashiro, M. (2014). Rho/rock signaling in motility and metastasis of gastric cancer. *World J. Gastroenterol.* 20, 13756–13766. doi: 10.3748/wjg.v20.i38.13756
- Meriane, M., Roux, P., Primig, M., Fort, P., and Gauthier-Rouviere, C. (2000). Critical activities of rac1 and Cdc42Hs in skeletal myogenesis: antagonistic effects of JNK and p38 pathways. *Mol. Biol. Cell* 11, 2513–2831. doi: 10.1091/mbc.11.8.2513
- Qian, H. (2007). Phosphorylation energy hypothesis: open chemical systems and their biological functions. *Annu. Rev. Phys. Chem.* 58, 113–142. doi: 10.1146/annurev.physchem.58.032806.104550
- Sordella, R., Jiang, W., Chen, G.-C., Curto, M., and Settleman, J. (2003). Modulation of rho gtpase signaling regulates a switch between adipogenesis and myogenesis. *Cell* 113, 147–158. doi: 10.1016/S0092-8674(03)00271-X
- Tsai, S. H., Huang, P. H., Peng, Y. J., Chang, W. C., Tsai, H. Y., Leu, H. B., et al. (2013). Zoledronate attenuates angiotensin ii-induced abdominal aortic aneurysm through inactivation of Rho/ROCK-dependent JNK and NF- κ B pathway. *Cardiovasc. Res.* 100, 501–510. doi: 10.1093/cvr/cvt230

- Wang, L.-F., Qiu, K., and Jia, Y. (2017). Effects of time delays in a mathematical bone model. *Chin. Phys. B* 26, 251–257. doi: 10.1088/1674-1056/26/3/030503
- Wang, S., Lan, Y., and Tang, L. (2015). Energy dissipation in an adaptive molecular circuit. *J. Stat. Mech.* 2015:P07025. doi: 10.1088/1742-5468/2015/00/P07025
- Wei, L., Roberts, W., Wang, L., Yamada, M., Zhang, S., Zhao, Z., et al. (2001). Rho kinases play an obligatory role in vertebrate embryonic organogenesis. *Development* 128, 2953–2962.
- Wei, L., Zhou, W., Croissant, J. D., Johansen, F.-E., Prywes, R., Balasubramanyam, A., et al. (1998). RhoA signaling via serum response factor plays an obligatory role in myogenic differentiation. *J. Biol. Chem.* 273, 30287–30294. doi: 10.1074/jbc.273.46.30287
- Wei, L., Zhou, W., Wang, L., and Schwartz, R. J. (2000). β_1 -integrin and pi 3-kinase regulate rhoA-dependent activation of skeletal α -actin promoter in myoblasts. *J. Physiol. Heart Circ. Physiol.* 278, H1736–H1743. doi: 10.1152/ajpheart.2000.278.6.H1736
- Yao, Y., Cao, W., Pei, Q., Ma, C., and Yi, M. (2018a). Break up of spiral wave and order-disorder spatial pattern transition induced by spatially uniform cross-correlated sine-wiener noises in a regular network of Hodgkin-Huxley neurons. *Complexity* 2018:8793928. doi: 10.1155/2018/8793928
- Yao, Y., Yang, L., Wang, C., Liu, Q., Gui, R., Xiong, J., et al. (2018b). Subthreshold periodic signal detection by bounded noise-induced resonance in the Fitzhugh–Nagumo neuron. *Complexity* 2018:5632650. doi: 10.1155/2018/5632650
- Zhang, D., Cao, Y., Ouyang, Q., and Yi, T. (2020). The energy cost and optimal design for synchronization of coupled molecular oscillators. *Nat. Phys.* 16, 95–100. doi: 10.1038/s41567-019-0701-7

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Liu, Yu, Yi, Gao, Gui, Yi and Sun. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



A Probabilistic Approach to Explore Signal Execution Mechanisms With Limited Experimental Data

Michael A. Kochen¹ and Carlos F. Lopez^{1,2*}

¹ Department of Biomedical Informatics, Vanderbilt University, Nashville, TN, United States, ² Department of Biochemistry, Vanderbilt University, Nashville, TN, United States

OPEN ACCESS

Edited by:

Tian Hong,
The University of Tennessee,
Knoxville, United States

Reviewed by:

Richard Posner,
Northern Arizona University,
United States
Alexey Goltsov,
Abertay University, United Kingdom

*Correspondence:

Carlos F. Lopez
c.lopez@vanderbilt.edu

Specialty section:

This article was submitted to
Systems Biology,
a section of the journal
Frontiers in Genetics

Received: 05 March 2020

Accepted: 04 June 2020

Published: 10 July 2020

Citation:

Kochen MA and Lopez CF (2020)
A Probabilistic Approach to Explore
Signal Execution Mechanisms With
Limited Experimental Data.
Front. Genet. 11:686.
doi: 10.3389/fgene.2020.00686

Mathematical models of biochemical reaction networks are central to the study of dynamic cellular processes and hypothesis generation that informs experimentation and validation. Unfortunately, model parameters are often not available and sparse experimental data leads to challenges in model calibration and parameter estimation. This can in turn lead to unreliable mechanistic interpretations of experimental data and the generation of poorly conceived hypotheses for experimental validation. To address this challenge, we evaluate whether a Bayesian-inspired probability-based approach, that relies on expected values for quantities of interest calculated from available information regarding the reaction network topology and parameters can be used to qualitatively explore hypothetical biochemical network execution mechanisms in the context of limited available data. We test our approach on a model of extrinsic apoptosis execution to identify preferred signal execution modes across varying conditions. Apoptosis signal processing can take place either through a mitochondria independent (Type I) mode or a mitochondria dependent (Type II) mode. We first show that *in silico* knockouts, represented by model subnetworks, successfully identify the most likely execution mode for specific concentrations of key molecular regulators. We then show that changes in molecular regulator concentrations alter the overall reaction flux through the network by shifting the primary route of signal flow between the direct caspase and mitochondrial pathways. Our work thus demonstrates that probabilistic approaches can be used to explore the qualitative dynamic behavior of model biochemical systems even with missing or sparse data.

Keywords: systems biology, limited data, apoptosis, probabilistic, mechanism, inference, high performance computing

INTRODUCTION

The complex dynamics of biochemical networks, stemming from numerous interactions and pathway crosstalk, render signal execution mechanisms difficult to characterize (Bhalla and Iyengar, 1999; Kitano, 2002; Loscalzo and Barabasi, 2011). Mathematical modeling of biochemical networks has become a powerful complement to experimentation for generating hypotheses regarding the underlying mechanisms that govern signal processing and suggesting targets for further experimental examination (Aldridge et al., 2006; Le Novère, 2015). Models of biochemical reaction networks, often based on a mass action kinetics formalism, are built to represent known pathway

mechanics with knowledge garnered from years or even decades of experimentation (Albeck et al., 2008; Lopez et al., 2013). Although these models have yielded important predictions and insights about biochemical network processes, they also depend on kinetic rate parameters and protein concentrations that are often poorly characterized or simply unavailable. A typical workaround is to employ model calibration methods to estimate suitable parameter values via optimization to protein concentration time course data (van Riel, 2006; Shockley et al., 2018; Mitra et al., 2019). However, the data needed for parameter optimization is often scarce, leading to the possibility of multiple parameter sets that fit the model to that data equally well but exhibit different dynamics (Lopez et al., 2013; Shockley et al., 2018). This poses a challenge for the study of dynamic network processes as the mode of signal execution can be highly dependent on a specific parameter set and could in turn lead to inadequate model-based interpretation. A computational approach that enables the exploration of biochemical signal execution mechanisms from a probabilistic perspective, constrained only by available data, would facilitate a rigorous exploration of network dynamics and accelerate the generation of testable mechanistic hypotheses (Wrede and Hellander, 2018).

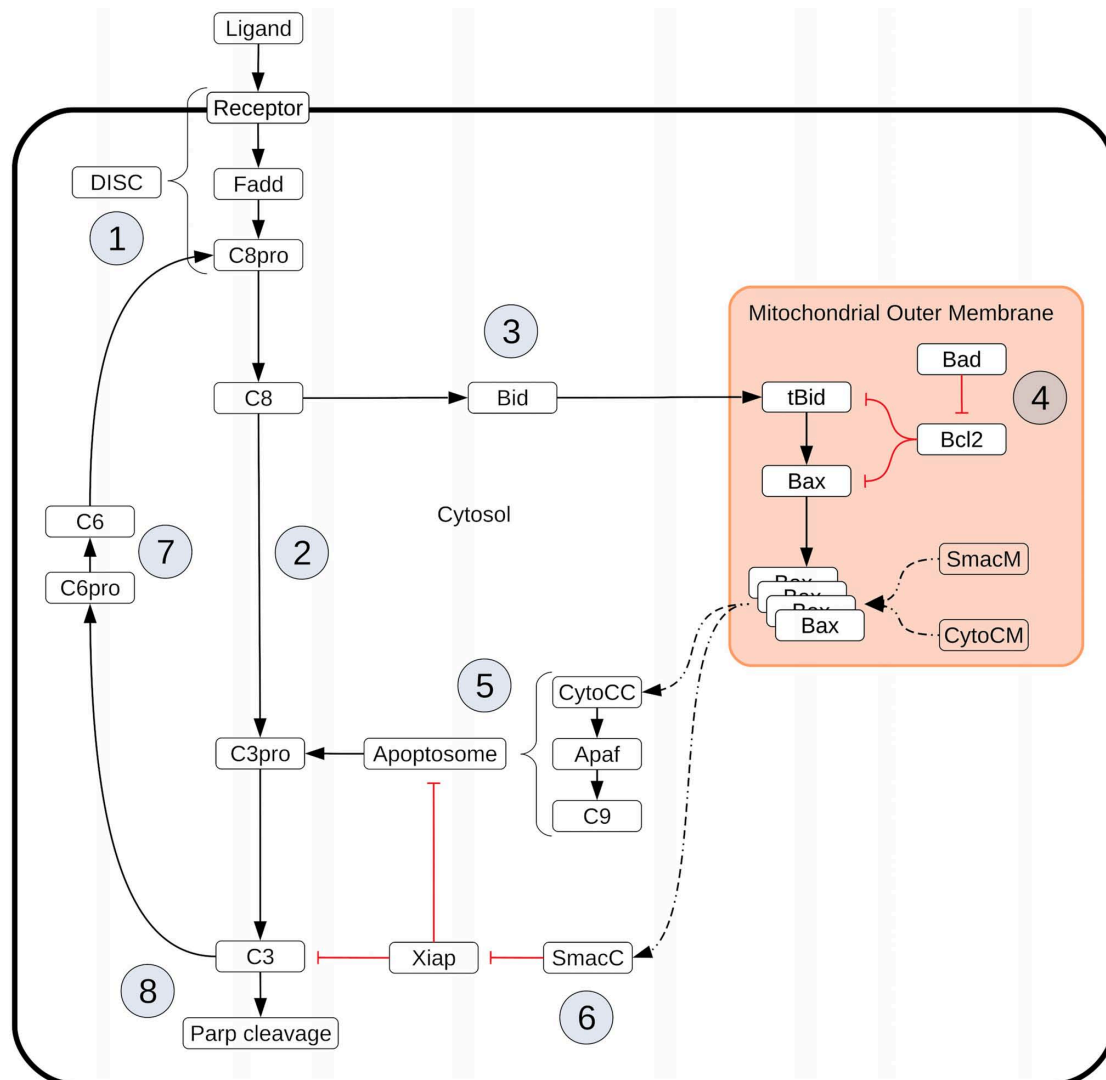
In this work, we investigate whether a Bayesian-inspired probabilistic approach can identify network signal execution mechanisms in extrinsic apoptosis restricted only by experimental observations. Two execution phenotypes have been identified for extrinsic apoptosis signaling: a mitochondria independent (Type I) phenotype, whereby initiator caspases directly activate effector caspases and induce cell death, and a mitochondria dependent (Type II) phenotype whereby initiator caspases engage the Bcl-2 family of proteins, which ultimately lead to effector caspase activation (see **Box 1** for biology details). Most mammalian cells execute apoptosis via the Type II mechanism, yet the Type I mechanism plays a central role in specific cell types, particularly certain types of lymphocytes (Scaffidi et al., 1999). A significant body of experimental and modeling work has identified key regulators for Type I vs. Type II execution. Computational approaches to study apoptosis network dynamics are numerous and range from simple dynamic Boolean networks to deterministic and stochastic kinetic models (Bentele et al., 2004; Albeck et al., 2008; Schlatter et al., 2009; Spencer and Sorger, 2011; Schleich and Lavrik, 2013; Würstle et al., 2014; Anderson et al., 2019). Aspects of apoptosis dynamics, like bistability (Eissing et al., 2004; Bagci et al., 2006; Legewie et al., 2006; Ho and Harrington, 2010) are often targets of analysis, and the structure of the apoptosis network has been examined via Bayesian model selection methods (Eydgahi et al., 2013). To specifically study phenotypic regulation of the extrinsic apoptosis network Aldridge et al. (2011) used a kinetic model in conjunction with Lyapunov exponent based bifurcation diagrams to define a boundary between phenotypes on the space of regulatory element concentrations. Raychaudhuri et al. (2008) also focused on the Type I/II phenotypes and used Monte Carlo simulations of an extrinsic apoptosis model to study stochastic fluctuations through the network.

Despite these efforts, it is still unclear how network structure and the interplay among multiple regulators can modulate signal execution for either cell type. A more traditional approach would prescribe intricate and detailed experimental measurements of cellular response to yield the desired data and improve our understanding of signal execution. However, the time and cost associated with such experiments makes it unlikely, and at times infeasible, to obtain said data. It is here that we see probabilistic inference approaches as complementary to experimentation, providing qualitative insights about signal execution mechanisms by integrating the expected parameter space subject only to available computer time. Here, we demonstrate that a probabilistic approach, constrained by network structure or molecular concentrations, can identify the dominant signal execution modes in a reaction network. Specifically, we demonstrate the dependence of Type I or a Type II cellular apoptosis execution on network structure and chemical-species concentrations. We use existing tools designed for the calculation of Bayesian evidence and repurpose them for the calculation of expected values for quantifiable *in silico* experimental outcomes. These expected values are then used as metrics for comparisons of signal flow through different pathways of the network and subnetworks in order to identify how regulators affect execution modes. We introduce two complementary approaches that can be used in tandem to explore signal execution modulation. We first define a *multimodel exploration method* to explore multiple hypothesis about apoptosis execution by deconstructing an established apoptosis network model into functional subnetworks that effectively represent *in silico* knockout experiments. We also define a *pathway flux method* to characterize the signal flux through specific network pathways within the chosen canonical network. Combined, these two approaches enable us to qualitatively identify key network components and molecular regulator combinations that yield mechanistic insights about apoptosis execution. Our approach is generalizable to other mass action kinetics-based networks where signal execution modes play important roles in cellular outcomes. This work leverages Nested Sampling algorithm methods to efficiently calculate expected values on high performance computing (HPC) platforms, both of which are seldom used in biological applications. In this manner we are able to carry out the necessary calculations to consider the entirety of the proposed parameter space and estimate expected values within the timespan of hours to days.

METHODS

Apoptosis Model and Simulations

The base model used in this work is a modified version of the Extrinsic Apoptosis Reaction Model (EARM) from Lopez et al. (2013) (EARM v2.1). The original EARM was simplified to reduce complexity and lower the number of parameters, but still retains the key features of the network for apoptosis execution. Specifically, we reduced the molecular complexity of mitochondrial outer membrane permeabilization (MOMP) down to a representative set of Bcl-2 proteins that capture



BOX 1 | Extrinsic apoptosis execution. Extrinsic apoptosis is a receptor mediated process for programmed cell death. The Type I/II phenotypes for the extrinsic apoptosis system were first described by Scaffidi et al. (1998). In that work they examined several cell lines and classified them into those that required the mitochondrial pathway to achieve apoptosis (Type II) and those that do not (Type I). They made several interesting conclusions. They found that Type II cells had relatively weak DISC formation, that both phenotypes responded equally well to receptor mediated cell death, that there was a delay in caspase activation in Type II cells, and that caspase activation happened upstream of mitochondrial activation in Type I cells and downstream in Type II. More recently, XIAP has also been put forth as a critical regulator in the choice of apoptotic phenotype. In Jost et al. (2009) they examined hepatocytes (Type II cells) and lymphocytes (Type I cells) under different conditions to examine the role XIAP plays in Type I/II determination. They made several observations upon Fas ligand or Fas-antibody induced apoptosis such as higher levels of XIAP in Type II cells and higher caspase effector activity in XIAP/Bid deficient mice versus apoptosis resistant Bid-only knockouts. In all, they concluded that XIAP is the key regulator that determines the choice of pathway. Extrinsic apoptosis is initiated when a death inducing member of the tumor necrosis factor (TNF) superfamily of receptors (FasR, TNFR1, etc.) is bound by its respective ligand (FasL, TNF- α , etc.), setting off a sequence biochemical events that result in the orderly deconstruction of the cell (Ashkenazi and Dixit, 1998). The first stage of this sequence is the assembly of the DISC at the cell membrane ① and the subsequent activation of Caspase-8. Upon ligand binding and oligomerization of a receptor such as FasR or TRAIL, an adapter protein, like FADD (Fas-associated protein with death domain), is recruited to the receptors cytoplasmic tail (Boldin et al., 1995; Kischkel et al., 2000; Sprick et al., 2000). FADD, in turn, recruits Caspase-8 via their respective death effector domains (DEDs), thus completing DISC formation (Kischkel et al., 2000; Sprick et al., 2000). Other DISC components could also be included here, such as the regulator cFlip (Krueger et al., 2001). Once recruited, proximal Procaspase-8 monomers dimerize, inducing their autoproteolytic activity and producing active Caspase-8 (Martin et al., 1998; Salvesen and Dixit, 1999; Boatright and Salvesen, 2003). After Caspase-8 activation the apoptotic signal can progress down two distinct pathways that both lead to the activation of Caspase-3 and the ensuing proteolysis of downstream targets. One pathway consists of a caspase cascade in which active Caspase-8 directly cleaves and activates Caspase-3 ② (Stennicke et al., 1998), while another, more complex pathway is routed through the mitochondria. In the mitochondrial pathway Caspase-8 cleaves the pro-apoptotic Bcl-2 family protein Bid in the cytosol, which then migrates to the mitochondria ③ where it initiates mitochondrial outer membrane permeabilization (MOMP) and the release of pro-apoptotic factors that lead to

(Continued)

BOX 1 | Continued

Caspase-3 activation (Li et al., 1998; Luo et al., 1998). MOMP has its own set of regulators that govern the strength of apoptotic signaling through the mitochondria ④. After Caspase-8 activated Bid, (tBid), migrates to the mitochondria it activates proteins in the outer mitochondrial membrane, such as Bax, that subsequently self-aggregate into membrane pores and allow exportation of Cytochrome-c and Smac/DIABLO to the cytosol (Desagher et al., 1999). Bid and Bax are examples of pro-apoptotic proteins from the Bcl-2 family, all of which share BH domain homology (Kelekar and Thompson, 1998). Other members of this family act as MOMP regulators; the anti-apoptotic Bcl-2, for example, binds and inhibits both Bid and Bax while the pro-apoptotic Bad similarly binds and inhibits its target, Bcl-2 (Oltval et al., 1993; Yang et al., 1995; Letai et al., 2002; Leber et al., 2007). Many other pro- and anti-apoptotic members of the Bcl-2 family have been discovered and together regulate MOMP (Kale et al., 2018). Regardless of which pathway is chosen, the intermediate results are Caspase-3 activation and subsequent cleavage of PARP ⑤, a proxy for cell death in the analyses here (Nicholson et al., 1995; Tewari et al., 1995). XIAP (X-linked inhibitor of apoptosis protein) is an inhibitor of Caspase-3 and has been proposed to be a key regulator in determining the Type I/II apoptotic phenotype of a cell (Jost et al., 2009). XIAP sequesters Caspase-3 but also contains a ubiquitin ligase domain that directly targets Caspase-3 for degradation. The inhibitor also sequesters and inhibits the Caspase-3 activating Caspase-9 residing within the apoptosome complex (Huang et al., 2001; Suzuki et al., 2001; Shiozaki et al., 2003). Apoptosome formation is initiated by Cytochrome-c exported from the mitochondria during MOMP ⑤. Cytochrome-c induces the protein APAF-1 to oligomerize and subsequently recruit and activate Caspase-9, thus forming the complex (Zou et al., 1999). Another MOMP export, the protein Smac/DIABLO ⑥, binds and inhibits XIAP, working in tandem with Cytochrome-c to oppose XIAP and carry out the apoptosis inducing activity of the Type II pathway (Adrain et al., 2001). Finally, Procaspase/Caspase-6 constitutes a feed forward loop between Caspase-3 and Caspase-8 ⑦ (Cowling and Downward, 2002).

the behavior of activators, inhibitors, effectors, and sensitizers. We also eliminated intermediate states for Cytochrome c and Smac to streamline effector caspase activation, and we added an explicit FADD molecule, an adapter protein in the death-inducing signaling complex (DISC), to achieve a more realistic representation of signal initiation. Overall, EARM v2.1 is comprised of 16 chemical species at non-zero initial concentrations, 50 total chemical species, 62 reactions, and 62 kinetic parameters. The modified model was recalibrated to recapitulate the time-dependent concentration trajectories of truncated Bid, Smac release from the mitochondria, and cleaved PARP analogous to the approach reported previously (Spencer et al., 2009) (**Supplementary Figure S1**). The modified EARM, and all derivative models, were encoded in PySB. All simulations were run using the mass action kinetics formalism as a system of ordinary differential equations (ODEs) using the VODE integrator in SciPy within the PySB modeling framework. All data results, representative models, and software are distributed with open-source licensing and can be found in the GitHub repository <https://github.com/LoLab-VU/BIND>.

Expected Value Estimation

The expected value for a quantifiable outcome is, by definition, the integral of an objective function that represents that outcome over the normalized distribution of parameters. This is analogous to the estimation of Bayesian evidence where a likelihood function is likewise integrated over a normalized distribution. We can thus use existing, established, Bayesian evidence estimation methods and software to estimate expected values by simply substituting the objective function for the likelihood function in the integral calculation. The remainder of this section and the next provide an overview of the evidence estimation methods and tools that we have repurposed for expected value calculations.

Bayesian evidence is the normalizing term in a Bayesian calculation and typically provides a measure for model comparison with regard to their fit to experimental data. It is expressed as:

$$P(D|M) = \int L(D|\theta, M) P(\theta|M) d\theta \quad (1)$$

Where M is the model under consideration, D is the experimental data, θ is a specific set of parameter values, $L(D|\theta, M)$ is the likelihood function describing the fit of the data to the model under those parameter values, and $P(\theta|M)$ is the prior distribution of parameters. An efficient method for evidence calculation is nested sampling (Skilling, 2006). This method simplifies the evidence calculation by introducing a prior mass element $dX = P(\theta|M)d\theta$ that is estimated by $(X_{i-1} - X_i)$ where $X_i = e^{-i/N}$, i is the current iteration of the algorithm, and N is the total number of live points. The evidence is then written as:

$$Z = \int_0^1 L dX \approx \sum_{i=1}^N L_i (X_{i-1} - X_i) \quad (2)$$

Initialization of the algorithm is carried out by randomly selecting an initial population of parameter sets (points in parameter space) from the prior distribution, scoring each one with the likelihood function, and ranking them from L_{high} to L_{low} . At each iteration of the algorithm a new set of parameter values is selected and scored. If that score is higher than L_{low} , then it is added to the population, at the appropriate rank, and L_{low} is removed from the population and added to the evidence sum (2).

Nested Sampling Software

All expected value estimates in this work are calculated with MultiNest, a nested sampling-based algorithm designed for efficient evidence calculation on highly multimodal posterior distributions (Feroz et al., 2009, 2013). MultiNest works by clustering the live points (population of parameter sets) and enclosing them in ellipsoids at each iteration. The enclosed space then constitutes a reduced space of admissible parameter sets. This lowers the probability of sampling from low likelihood areas and evaluating points that will only be discarded. The evidence estimate is accompanied by an estimate of the evidence error. The algorithm terminates when the presumed contribution of the highest likelihood member of the current set of live points, $L_{\text{high}}X_i$ is below a threshold. Here, we use a threshold of 0.0001 and a population size and 16,000 unless otherwise noted. The population size of 16,000 was found to be an acceptable compromise between precision and computational

austerity for the model sizes and *in silico* experiments performed in this study. See (Feroz et al., 2009, 2013), for more details on the MultiNest algorithm. We use MultiNest with the Python wrapper PyMultiNest (Buchner et al., 2014), which facilitates the integration with PySB into the parameter sampling pipeline.

Multimodel Exploration Analysis

We carried out an analysis analogous to knockout experiments to investigate the contribution of different network components to the overall dynamics of the apoptosis execution network.

We broke down the EARM network into six subnetworks and compared their likelihood of achieving apoptosis across increasing concentrations of the regulator XIAP. A standard proxy for apoptosis execution is cleavage of the protein PARP. We therefore define the proportion of cleaved PARP, relative to total PARP, as a metric for effective apoptosis execution. We defined the objective function that represents the amount of cleaved PARP as:

$$Obj_{multimodel} = \frac{cPARP}{tPARP} \quad (3)$$

where $cPARP$ is the amount of PARP that has been cleaved and $tPARP$ is the total amount of PARP in the system. When this objective function is substituted into Eq. (1) in place of the likelihood function, we obtain the expected value, the average over the chosen prior parameter range, for the proportion of PARP that has been cleaved at the end of the *in silico* experimental simulation. We compare PARP cleavage for different subnetworks and regulatory conditions only in qualitative terms and as a *relative* measure of the expected outcome.

Pathway Flux Analysis

We also explored the effect of molecular regulators of Type I vs. Type II execution relative to the apoptosis signal flux through the network, as we have done in previous work (Shockley et al., 2019). Briefly, signal flux is defined as the chemical reaction flux in units of molecules per unit time, that traverses through a given pathway. In the apoptosis network there are two potential pathways that can lead to Caspase-3 activation and subsequently PARP cleavage. In the direct caspase pathway initiator caspases, like Caspase-8, directly cleave and activate effector caspases, like Caspase-3. By contrast, in the mitochondrial pathway, effector caspases are activated via the apoptosome, and are dependent on MOMP. Therefore, the dominant pathway responsible for Caspase-3 activation defines the route of the signal. To estimate the flux through one of these pathways, we define the objective function as:

$$Obj_{pathway} = \sum_{t=0}^T \frac{\sum_0^t C3_{pathway}}{\sum_0^t C3_{total}} \times (cParp_t - cParp_{t-1}) \quad (4)$$

where t represents time in seconds, $\sum_0^t C3_{pathway}$ is the amount of Caspase-3 activated via the target pathway up to time t , $\sum_0^t C3_{total}$ is the total Caspase-3 activated up to time

t , and $\sum_0^t C3_{caspase} / \sum_0^t C3_{total}$ is the proportion of activated Caspase-3 that was produced via the target pathway up to time t . $(cParp_t - cParp_{t-1})$ is the total PARP that has been cleaved, and activated, by Caspase-3 from time $t - 1$ to time t . Thus, at any given time t we can estimate the amount of Caspase-3 that has been activated through a specific pathway. Multiplication of these two terms returns an estimate for the amount of PARP cleaved via the specific pathway at time t . Summing over T then returns an estimate for the total apoptosis signal flowing through the target pathway. Like the PARP cleavage objective function, the signal flux objective substituted into Eq. (1) produces an estimate of the average flux over a defined prior distribution. We estimated this quantity over increasing concentrations of the molecular regulator XIAP, but also at high and low levels of the DISC components FADD and Caspase-8. The total signal flux was estimated by summing the flux estimate for both the direct caspase and mitochondrial pathways.

Parameter Ranges and Initial Conditions

The prior distribution takes the form of a set of parameter ranges, one for each reaction rate parameter. The ranges used here span four orders of magnitude around generic reaction rates deemed plausible (Aldridge et al., 2006) and are specific to the type of reaction taking place. The ranges of reaction rate parameters, in Log_{10} space, are 1st order forward: $[-4.0, 0.0]$, 2nd order forward: $[-8.0, -4.0]$, 1st order reverse: $[-4.0, 0.0]$, catalysis: $[-1.0, 3.0]$. These ranges were also used in the calibration of the base model. Where possible, initial conditions were either collected from the literature (Eissing et al., 2004; Dai et al., 2018) or taken from a previous model of extrinsic apoptosis (Aldridge et al., 2011; Lopez et al., 2013). Because the baseline model was designed to concur with Type II apoptotic data (see above), literature derived initial conditions were based on Type II Jurkat or Hela cell lines (Supplementary Table S1).

Expected Value Ratios

Evidence estimates are often used to select between two competing models by calculating the Bayes factor (i.e., the ratio of their evidence values). This provides a measure of confidence for choosing one model over another. We can likewise use the ratios of expected values to gain additional insights into the dynamical relationship between network components. To facilitate construction of expected value ratios (EVR) with a continuous and symmetric range, we define them as:

$$EVR = \begin{cases} -\frac{Z_2}{Z_1} + 1 & \text{if } Z_1 < Z_2 \\ \frac{Z_1}{Z_2} - 1 & \text{if } Z_1 > Z_2 \end{cases} \quad (5)$$

where Z_1 and Z_2 are the expected value estimates for two networks under comparison.

Computational Resources

Because of the high computational workload necessary for this analysis, a wide range of computational resources were used. The bulk of the work was done on the ACCRE cluster

at Vanderbilt University which has more than 600 compute nodes running Intel Xeon processors and a Linux OS. As many as 300 evidence estimates were run in parallel on this system. Additional resources included two local servers, also running Intel processors and a Linux OS, as well as a small local four node cluster running Linux and AMD Ryzen 1700 processors. A detailed breakdown of CPU time can be found in the results section. In all, expected value estimates for 14 different networks/initial conditions were made across the range of XIAP concentrations. We estimate all 14 runs would take ~9 days each on a typical university server with 32 cores/64 threads.

RESULTS

Overview: A Bayesian-Inspired Approach to Explore Mechanistic Hypotheses

Our overarching goal is to understand the mechanisms and dynamics of biochemical networks responsible for cellular commitment to fate, given incomplete or unavailable data. We take a probabilistic approach, similar to those used in Bayesian evidence-based model selection and multimodel inference, to compare model subnetworks and pathways with respect to apoptotic signal execution under various *in silico* experimental conditions and enable the generation of hypotheses regarding the underlying mechanisms of signal processing. Using this approach, we've employed two distinct but complimentary strategies as displayed in **Figure 1** (Note that the base network in **Figure 1** is a simplified version of the network used for demonstration in the results. From top to bottom the four nodes correspond to signal initiation at the death inducing signaling complex (DISC), export of proapoptotic factors from the mitochondria, inhibition of the antiapoptotic protein XIAP, and catalysis/inhibition of PARP. See **Box 1** for a detailed description of the model used in this work.)

The first is *Multimodel Exploration Analysis* (**Figure 1**, left path), wherein the network model is deconstructed into biologically relevant subnetworks and the probability of each subnetwork achieving apoptosis, under various regulatory conditions, is estimated via the calculation of an expected value for a quantifiable proxy of apoptosis. This differs from traditional model selection and multimodel inference applications where models are typically ranked based on their fit to experimental data and high-ranking models may be averaged to obtain a composite model (Burnham and Anderson, 2002; Xu et al., 2010; Symonds and Moussalli, 2011; Aitken and Akman, 2013; Eydgahi et al., 2013; Pullen and Morris, 2014). Here, we already have a model that captures key features of programmed cell death execution. Instead, we use the differences in expected values for a quantity that is representative of apoptosis to construct a composite picture of mechanistic evidence for apoptosis execution. To achieve this, we first tailor the objective function to represent signal execution strength, as measured by cleaved PARP concentration at the end of the simulation.

The expected value derived from this objective function therefore describes the likelihood that the signal is effectively transmitted through a given network. It should be noted that Bayesian evidence, and by extension our expected value calculation, inherently incorporates model complexity as the objectives are integrated over normalized prior distributions (MacKay and Kay, 2003; Feroz et al., 2009). As we will see, comparison of changes in signal strength through relevant subnetworks allows inferences to be made on the effect of the perturbed network regulator as well as various network components on the overall dynamics of the system. We focus primarily on understanding how Bayesian evidence for the caspase pathway compares to that of the complete network as these are most relevant for the analysis of Type I/II execution modes. This analysis will inform on how network components contribute to overall signal execution and provide mechanistic insights about the sensitivity of PARP cleavage to subnetwork components.

The second strategy is *Pathway Flux Analysis* (**Figure 1**, right path), where we retain the complete network structure but instead tailor the objective functions to measure biochemical reaction flux through either the direct caspase or mitochondrial pathways. We primarily consider the influence of the apoptosis inhibitor XIAP on regulatory dynamics and phenotypic fate but also consider the regulatory effect of the death inducing signaling complex (DISC) and the anti-apoptotic protein Bcl-2, all of which have been found to be relevant to Type I vs. Type II execution in different cell types (Scaffidi et al., 1998; Jost et al., 2009). This analysis will inform on how molecular regulators modulate biochemical flux through the network and their influence on apoptosis completion as measured by PARP cleavage.

Decomposition of the Extrinsic Apoptosis Network and Reductive Analysis of the Effects of XIAP

To investigate the effect of network substructures on apoptosis signaling, we build a composite description of system dynamics by observing variations in signal throughput, represented by expected values of PARP cleavage, between subnetworks (**Figures 2A–F**) relative to changes in regulatory conditions. We consider relative changes in expected PARP cleavage as the number of XIAP molecules is increased where a higher value indicates a stronger average signal over the prior range of parameter values. XIAP was varied from 0 to 200,000 molecules per cell in increments of 250 to explore how changes in XIAP affect the likelihood of apoptosis execution. For subnetworks that include the mitochondrial pathway, Bcl-2 (an anti-apoptotic protein) was eliminated, to explore Type I vs. Type II activity independent of inhibitors that could confound signal throughput, and more closely simulate a cell that is “primed” for death (Certo et al., 2006). All other initial values were fixed at the levels shown in **Supplementary Table S1**. In the absence of XIAP all six subnetworks have PARP cleavage estimates greater than 0.98 (**Figure 2A**: 0.993, **Figure 2B**: 0.998,

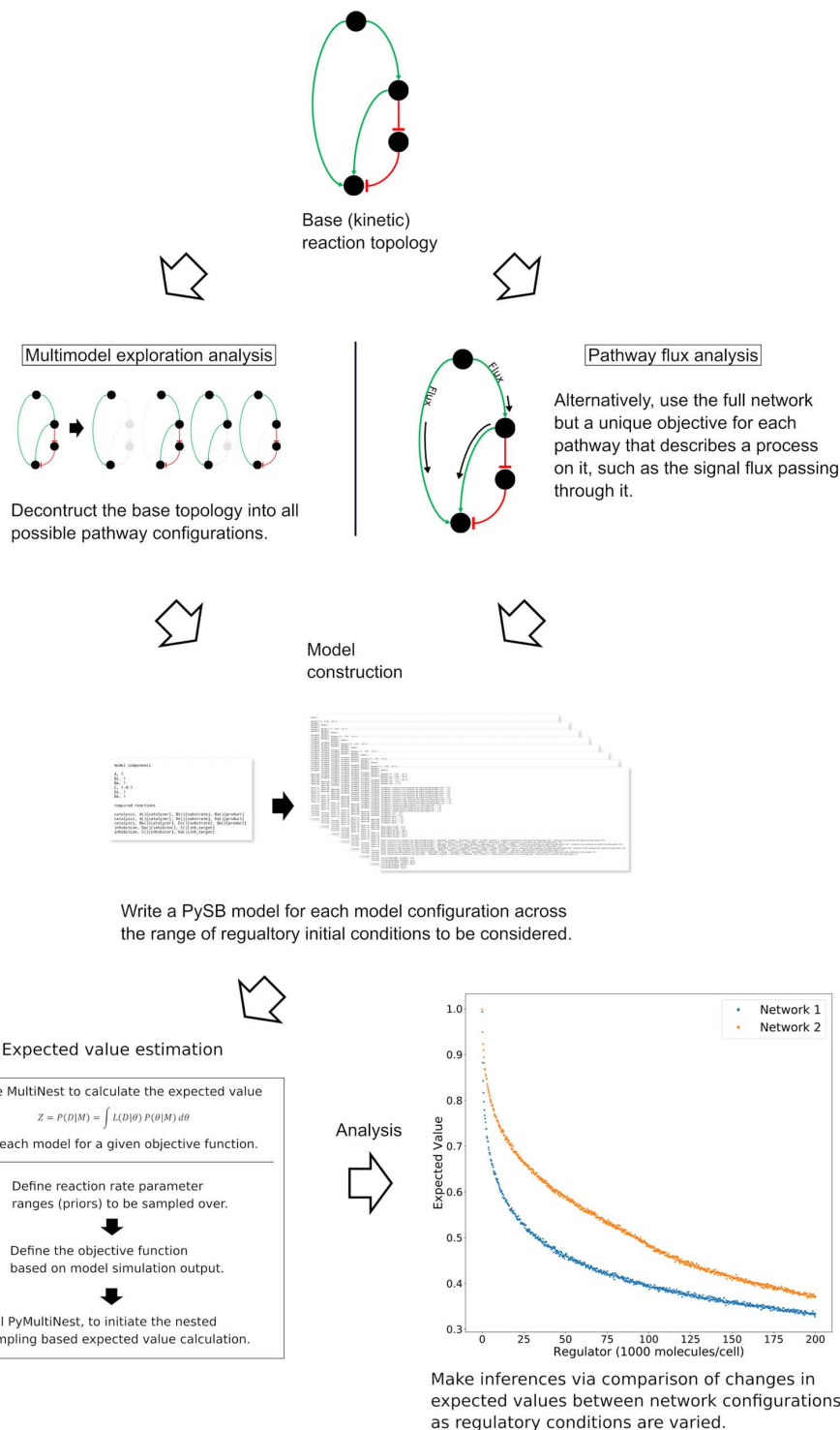


Figure 1 | General workflow for the analysis of network dynamics using trends in expected values. The target network is first deconstructed into subnetworks that effectively represent *in silico* knockouts (Note that the base network here is a simplified version of the network used for demonstration of the methodology. Briefly, the four nodes from top to bottom represent the death inducing signaling complex, the mitochondria, XIAP and PARP.) A model for each subnetwork and each incremental set of regulatory conditions is then created and passed to an algorithm for estimation of the expected value for an aspect of signal transduction. The expected value is calculated via integration of a user-defined objective function that quantifies that aspect of signal transduction over a range of parameter values (the prior). The trends in the expected values over changing regulatory conditions are then compared to make qualitative inferences regarding network dynamics. In a complimentary method, the full model is retained but the objective function is targeted to different pathways. Inferences on network dynamics can again be made via comparison of the trends in the expected values.

Figure 2C: 0.992, **Figure 2D:** 0.981, **Figure 2E:** 0.998, **Figure 2F:** 0.981, **Supplementary Table S2**) indicating a robust apoptotic signal for each across the allowed range of parameters. The log-expected value version of **Figure 2G** along with estimated errors generated by MultiNest are displayed in **Supplementary Figure S2**.

The results in Jost et al. (2009) imply that the cellular level of XIAP determines the preferred apoptosis pathway with higher levels specific to Type II cells and lower levels specific to Type I. To hypothesize a possible mechanistic explanation for this behavior we compared the expected PARP cleavage, over increasing concentrations of XIAP, for the direct caspase activation network against both the complete network and the isolated mitochondrial pathway network (**Figures 2A,G** green; **Figures 2E,G** orange; **Figures 2F,G** blue, respectively). This mimics reported experimental strategies to study Type I/II phenotypes and allows us to gauge the effect of XIAP on networks with and without a mitochondrial component (Scaffidi et al., 1998; Jost et al., 2009).

As XIAP levels increase we see differential effects on all subnetworks in the form of diverging expected value estimates, indicating differences in the efficacy of XIAP induced apoptotic inhibition. PARP cleavage values for the isolated caspase pathway (**Figure 2G** green) diverge from the complete network (**Figure 2G** orange) and mitochondrial pathway (**Figure 2G** blue) showing a steeper initial decline that diminishes as XIAP continues to increase. PARP cleavage values for the caspase pathway falls to 0.5 at an XIAP level of roughly 32,000. However, the complete network and mitochondrial pathways require XIAP levels nearly threefold higher with PARP cleavage reaching 0.5 at around 92,000 and 95,000, respectively.

Because the direct caspase activation pathway (**Figure 2G** green) is representative of the Type I phenotype, the disproportionate drop in its expected PARP cleavage as XIAP concentration increases is consistent with experimental evidence showing XIAP-induced transition from a Type I to a Type II execution mode (Jost et al., 2009). The complete network, containing the full mitochondrial subnetwork, and mitochondrial only pathway are also affected by XIAP but exhibit resistance to its anti-apoptotic effects, a difference that is most prominent at moderate levels of the inhibitor. This suggests a dependence on mitochondrial amplification for effective apoptosis as XIAP increases from low to moderate levels. At higher levels of XIAP the PARP cleavage for the caspase pathway level off and the gaps between it and the two mitochondrial containing networks narrow. The disproportionate effect of XIAP inhibition of apoptosis on the caspase pathway suggests that the mechanism for XIAP induced transition to a Type II pathway can be attributed to differential inhibition of the apoptotic signal through the isolated caspase pathway vs. a network with mitochondrial involvement.

The next two highest trends in expected values after that of the direct caspase network belong to the networks representing direct caspase activation plus mitochondrial activation and mitochondrial activation alone (**Figure 2G** purple and brown). For most of the range with XIAP below 100,000 these two

networks have largely overlapping PARP cleavage trajectories, despite the fact that the former has twice as many paths carrying the apoptotic signal. Near an XIAP level of 100,000 the two trends diverge as the decrease in PARP cleavage for the mitochondrial activation only network accelerates. This could be explained by XIAP overwhelming the apoptosome at these higher levels. The apoptosome is an apoptosis inducing complex (via Caspase-3 cleavage) consisting of Cytochrome c, APAF-1, and Caspase-9, and is an inhibitory target of XIAP. As XIAP increases past 125,000 the mitochondrial activation only PARP cleavage values fall below even the solo direct caspase values, possibly due to the two-pronged inhibitory action of XIAP at both the apoptosome and Caspase-3. An interesting observation here is that the addition of the direct caspase pathway to the mitochondrial activation pathway does not appear to increase the likelihood of achieving apoptosis for lower values of XIAP.

PARP cleavage values for the network representing direct caspase activation plus mitochondrial inhibition of XIAP are in red in **Figure 2G**. Below an XIAP level of 100,000 these values are consistently above the PARP cleavage values for the network representing direct caspase plus mitochondrial activation. Note that while direct caspase activation does not appear to increase the likelihood of achieving apoptosis when added to the mitochondrial activation pathway (**Figure 2G** purple) the amplification of the direct caspase activation via mitochondrial inhibition of XIAP leads to a higher likelihood than solo activation through the mitochondria. This suggests the possibility that the primary mechanism for mitochondrial apoptotic signal amplification, under some conditions, may be inhibition of XIAP, with direct signal transduction a secondary mechanism. Above an XIAP level of 100,000, the direct caspase with XIAP inhibition PARP cleavage values drop to levels roughly in line with the values for direct caspase activation plus mitochondrial activation, possibly due to the fact that Smac, the mitochondrial export that inhibits XIAP, is also set to 100,000 molecules per cell. Both, however, remain more likely to attain apoptosis than direct caspase activation alone.

The two subnetworks with the highest expected values for apoptotic signal execution are the complete network and the isolated mitochondrial pathway (**Figure 2E** orange and **Figure 2F** blue). As previously mentioned, both of these networks contain the full mitochondrial pathway implying that this pathway supports resistance to XIAP inhibition of apoptosis. Between XIAP levels of 0 to 100,000 the two trends track very closely, with the mitochondrial only pathway showing a slight but consistent advantage for apoptosis execution. The average difference between an XIAP level of 20,000 and 80,000 is roughly 0.014, meaning we expect the average PARP cleavage to favor the mitochondrial only pathway by about 1.4 percentage points, which may seem unremarkable. Context matters however, and the context here is that the complete network has potentially twice the bandwidth for the apoptotic signal, namely the addition of the more direct caspase pathway. Together, this raises the possibility that under some conditions the caspase pathway is not a pathway but a sink for the apoptotic signal. In such

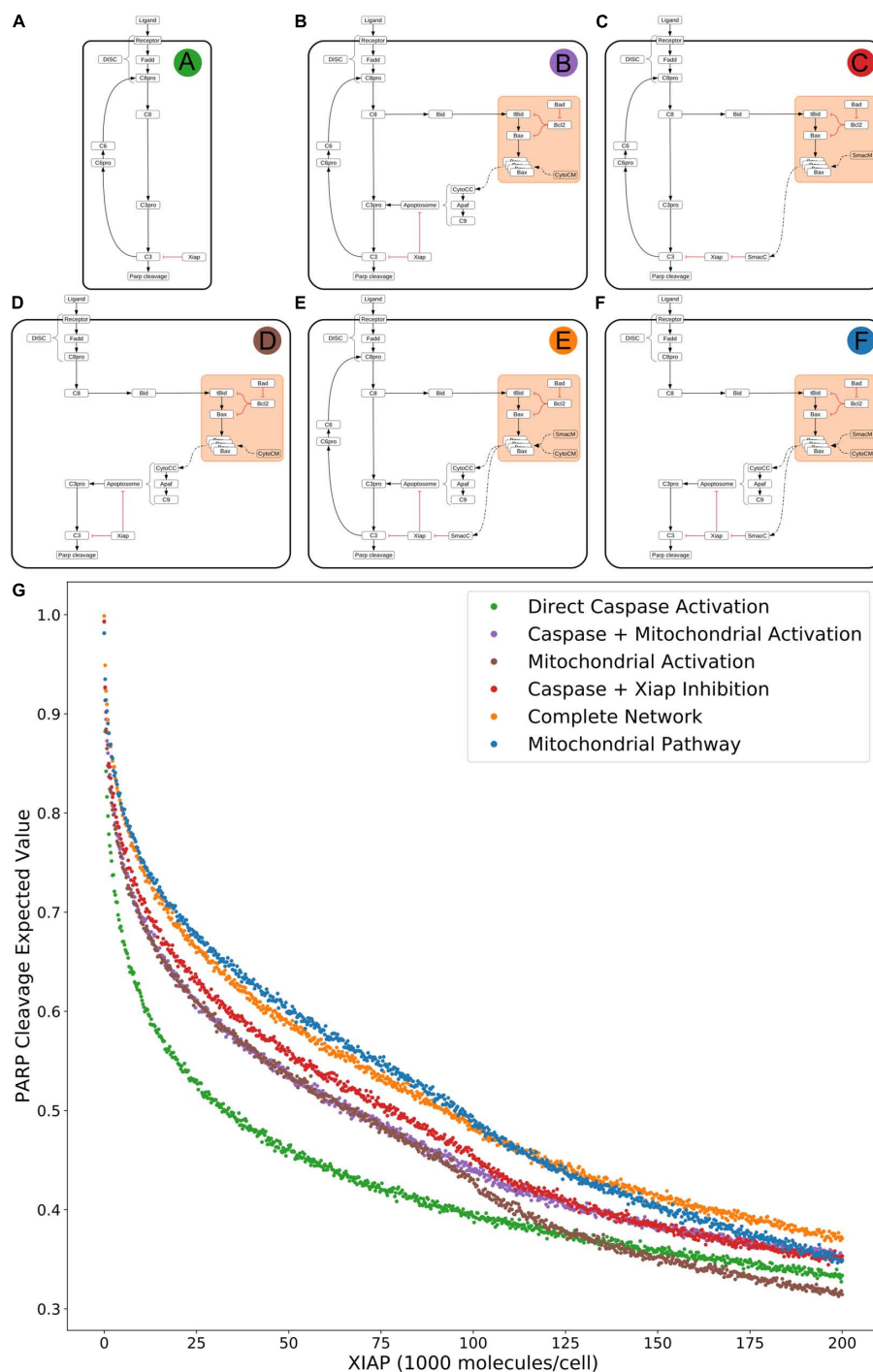


Figure 2 | Extrinsic apoptosis subnetworks and the likelihood of achieving apoptosis. **(A)** The direct caspase subnetwork. **(B)** The direct caspase + mitochondrial activation subnetwork. **(C)** The direct caspase + mitochondrial inhibition of XIAP subnetwork. **(D)** The mitochondrial activation subnetwork. **(E)** The complete network. **(F)** the mitochondrial subnetwork. **(G)** Trends in expected values for each of the networks in panels **(A–F)** over a range of values for the apoptosis inhibitor XIAP and for an objective function that computes the proportion of PARP cleavage (a proxy for cell death) at the end of the *in silico* experimental simulation.

a scenario, the signal through the caspase pathway would get lost as Caspase-3 is degraded by XIAP. Not until the signal through the mitochondrial pathway begins inhibiting XIAP could the signal proceed. Around the 100,000 level of XIAP

the PARP cleavage trend for the mitochondrial pathway crosses below that for the complete network. This could be due to the parity with Smac, components of the apoptosome, or a combination of the two.

Apoptosis Signal Strength Dictates the Signal Route Through the Network

The results in Scaffidi et al. (1998) indicate a strong phenotypic dependence on the strength of the apoptosis signal. Here, we examine hypotheses made in that work and the interplay between the DISC and XIAP regulatory axes. We again increase XIAP from 0 to 200,000 molecules in increments of 250, but this time at a low number of DISC complexes by lowering the initial values of both the scaffold protein FADD and the initiator Caspase-8, from 130,000 to 100 molecules per cell. In addition to the *Multimodel Exploration Analysis* approach used in the previous section, we also use the *Pathway Flux Analysis* approach using the signal flux objective function (see section “Methods”). In this way we attain a holistic view of network dynamics that incorporates both network structure and signal flux crosstalk from all possible pathways. Additional analysis of caspase and mitochondrial pathway signal flux over a range of values for both XIAP and Bcl-2 is displayed in **Supplementary Figure S3** and interpreted in **Supplementary Text S1**.

Figure 3A displays the PARP cleavage expected values for the direct caspase activation pathway and complete network (from **Figure 2G**) along with their low DISC counterparts. Two things are immediately apparent. PARP cleavage for the caspase pathway with a low number of DISC molecular components is lower across the entire range of XIAP concentrations. The complete network, on the other hand, shows almost no difference under low DISC conditions at lower values of XIAP. This supports the hypothesis that mitochondrial involvement is necessary to overcome weak DISC formation and that weak signal initiation constitutes a Type II trait (Scaffidi et al., 1998).

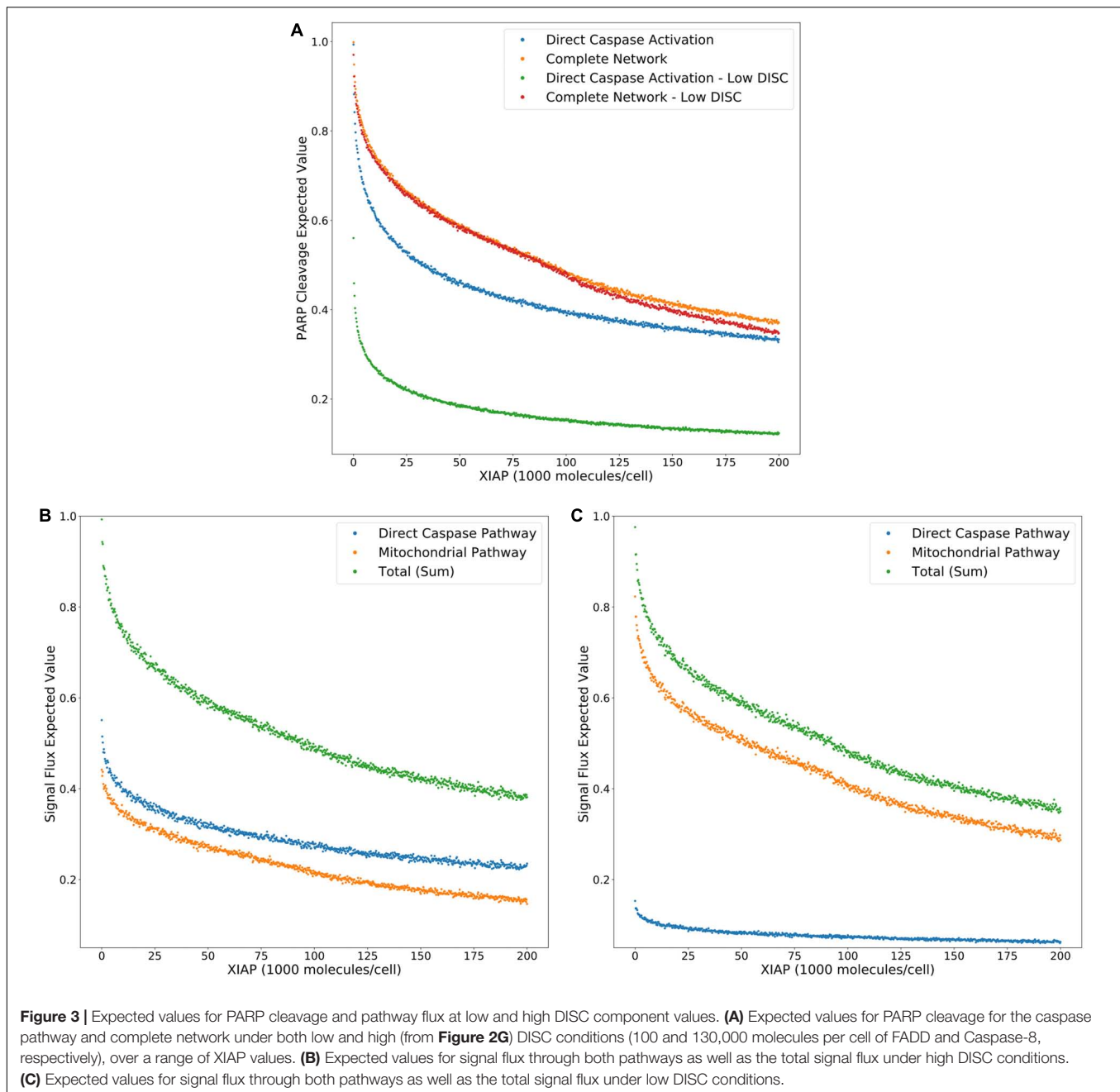
Figures 3B,C show expected values for signal flux through the caspase pathway and complete network, for high and low numbers of DISC components, respectively. At higher DISC values, signal flux through the caspase pathway is consistently higher than the flux through the mitochondrial pathway. At lower DISC values the signal flux through the mitochondrial pathway exceeds the flux through the caspase pathway. These results shed interesting mechanistic observations in the context of a previously proposed hypothesis stating that mitochondrial activation is downstream of Caspase-8 activation in Type I cells and upstream in Type II cells. If a weaker initial apoptosis cue does indeed push the signal through the mitochondrial pathway the initial activation of Caspase-8 would be weak and the amplifying activity of the mitochondria would ramp up the signal before Caspase-8 could directly activate Caspase-3. On the other hand, strong initial activation that pushes the signal through the caspase pathway would activate both Caspase-8 and Caspase-3 before MOMP becomes fully active. Also notable is the nearly identical trajectories of the total signal flux through the low and high DISC models. The average difference over the range of XIAP was only 0.011 (**Supplementary Table S3**). This is consistent with observations that both Type I and Type II cells respond equally well to receptor mediated apoptosis (Scaffidi et al., 1998).

Overall these results set up three mechanistic explanations for apoptosis execution and the signal flux schematic for each is displayed in **Figures 4A–C**, respectively. On one end, strong

signal initiation and low XIAP results in the independence of apoptosis from the mitochondrial pathway. This behavior is consistent with Type I cells like the SKW6.4 cell lines (Scaffidi et al., 1998). Under this scenario our results imply that most of the signal flux is carried through the caspase pathway and we hypothesize that control of apoptosis is dominated by that pathway. On the other end of the spectrum weak signal initiation and moderate to high levels of XIAP result in a dependence on the mitochondrial pathway. Such behavior is consistent with Type II cells like Jurkat (Scaffidi et al., 1998). In this case our results strongly indicate that most of the signal flux is carried through the mitochondrial pathway and we hypothesize that apoptosis execution is dominated by that pathway. In between these two extremes is the case with strong signal initiation, and moderate to high levels of XIAP levels with increased apoptotic dependence on mitochondrial activity versus the low XIAP case. Such a scenario that is consistent with MCF-7 cell that are known to have traits of both phenotypes (Scaffidi et al., 1998). In this case, we found that most of the apoptotic signal is carried through the caspase pathway despite the dependence on the mitochondria and we hypothesize that the mitochondrial pathway acts to allow the apoptotic signal through the caspase pathway.

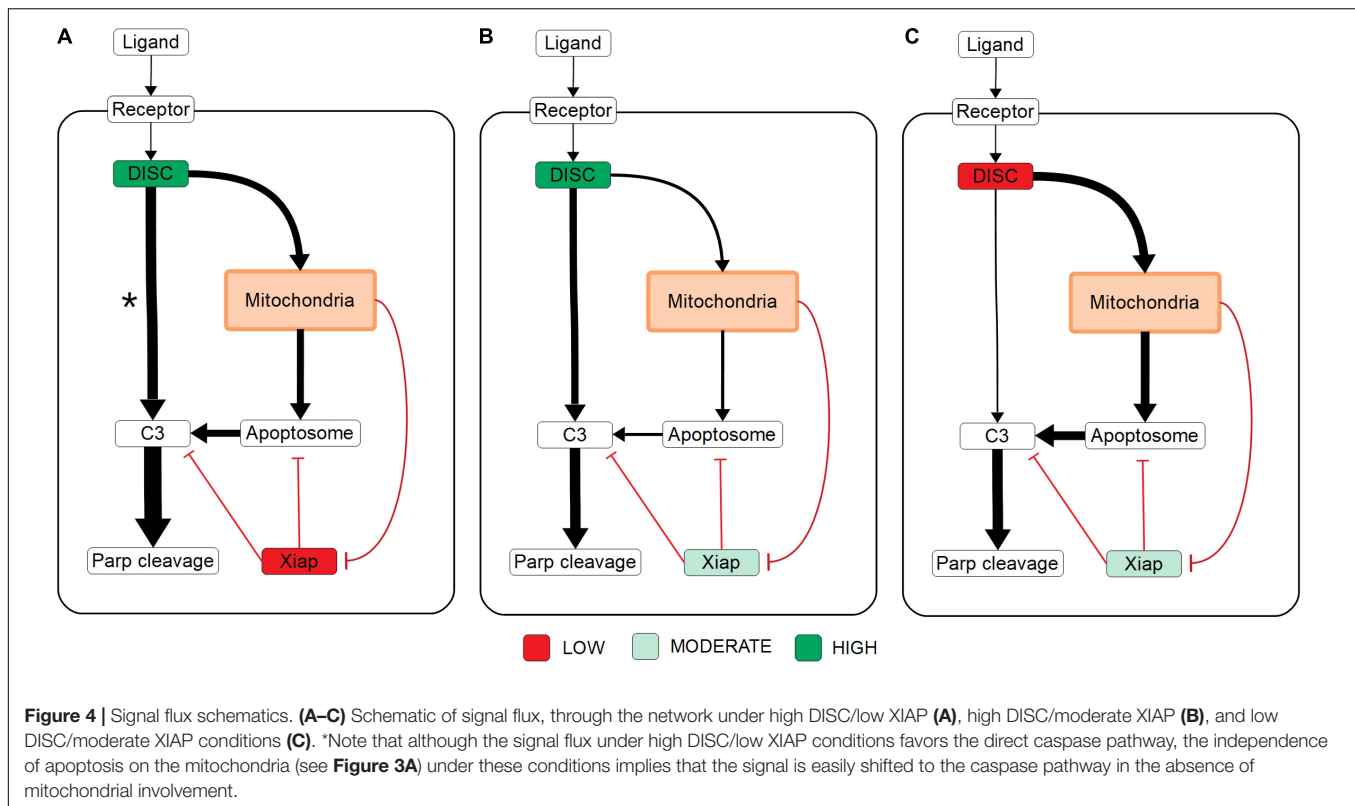
Expected Value Ratios and XIAP Influence on Type I/II Apoptosis Phenotype

Model selection methods typically calculate the evidence ratios, or Bayes factors, to choose a preferred model and estimate the confidence of that choice (Burnham and Anderson, 2002; Symonds and Moussalli, 2011). When comparing changes in likelihood of an outcome as regulatory conditions are altered we can similarly use ratios of expected values to provide additional information about evolving network dynamics under regulatory perturbations. To characterize the effect of XIAP on the choice of Type I or II apoptotic phenotype we calculated the expected value ratios (**Figure 5A**), for each value of XIAP between the caspase pathway and both the complete network and mitochondrial pathway (from **Figure 2G**). In these calculations, the denominator represents the caspase pathway so that higher values favor a need for mitochondrial involvement. An interesting feature of both the complete and mitochondrial expected value ratios is the peak and reversal at a moderate level XIAP (**Figure 5B**). This reflects the initially successful inhibition of the caspase pathway that decelerates relatively quickly as XIAP increases, and a steadier rate of increased inhibition on networks that incorporate the mitochondrial pathway. The ratios peak between 45,000 and 50,000 molecules of XIAP (more than double the value of its target molecule Caspase-3 at 21,000) and represent the optimal level of XIAP for the requirement of the mitochondrial pathway and attainment of a Type II execution. Given the near monotonic decline of the expected values for both pathways, representing increasing suppression of apoptosis, the peak and decline in the expected value ratios could represent a shift toward complete apoptotic resistance. Our results therefore complement the observations in Aldridge et al. (2011) where a similar outcome was observed experimentally.



A common technique to study apoptosis is to knockdown Bid, overexpress Bcl-2, or otherwise shut down MOMP induced apoptosis through mitochondrial regulation. This strategy was used in Ashkenazi and Dixit (1998), Jost et al. (2009), to study the role of XIAP in apoptosis and in the work of Aldridge et al. (2011) to explore Type I vs. Type II execution in different cell lines. Taking a similar approach, we set Bcl-2 levels to 328,000 molecules per cell, in line with experimental findings (Dai et al., 2018), to suppress MOMP activity and recalculated the PARP cleavage expected values and their ratios (**Figures 5C,D, Supplementary Table S5**). Under these conditions PARP cleavage for the mitochondrial pathway drop

well below that of the direct caspase pathway, which is reflected in the expected value ratios trend as a shift into negative territory and indicate that the caspase pathway is favored. PARP cleavage for the complete network under MOMP inhibition is shifted closer to that for the caspase pathway at higher concentrations of XIAP but is still higher throughout the full range of XIAP. The peak in the associated expected value ratios is flattened as the level of XIAP increases from low levels, suggesting that increasing XIAP is less likely to induce a transition to a Type II phenotype in a system with an already hampered mitochondrial pathway. We note that complete inhibition of MOMP would result in uninformative mitochondrial pathway results. PARP



cleavage expected values for the complete network would be indistinguishable from those for the direct caspase pathway and the complete/caspase ratios would simply flatline. However, our analysis shows that isolation of active biologically relevant subnetworks and direct comparison under changing molecular regulatory conditions, using trends in expected values, enables the extraction of information regarding pathway interactions and differential network dynamics.

Precision vs. Computational Cost

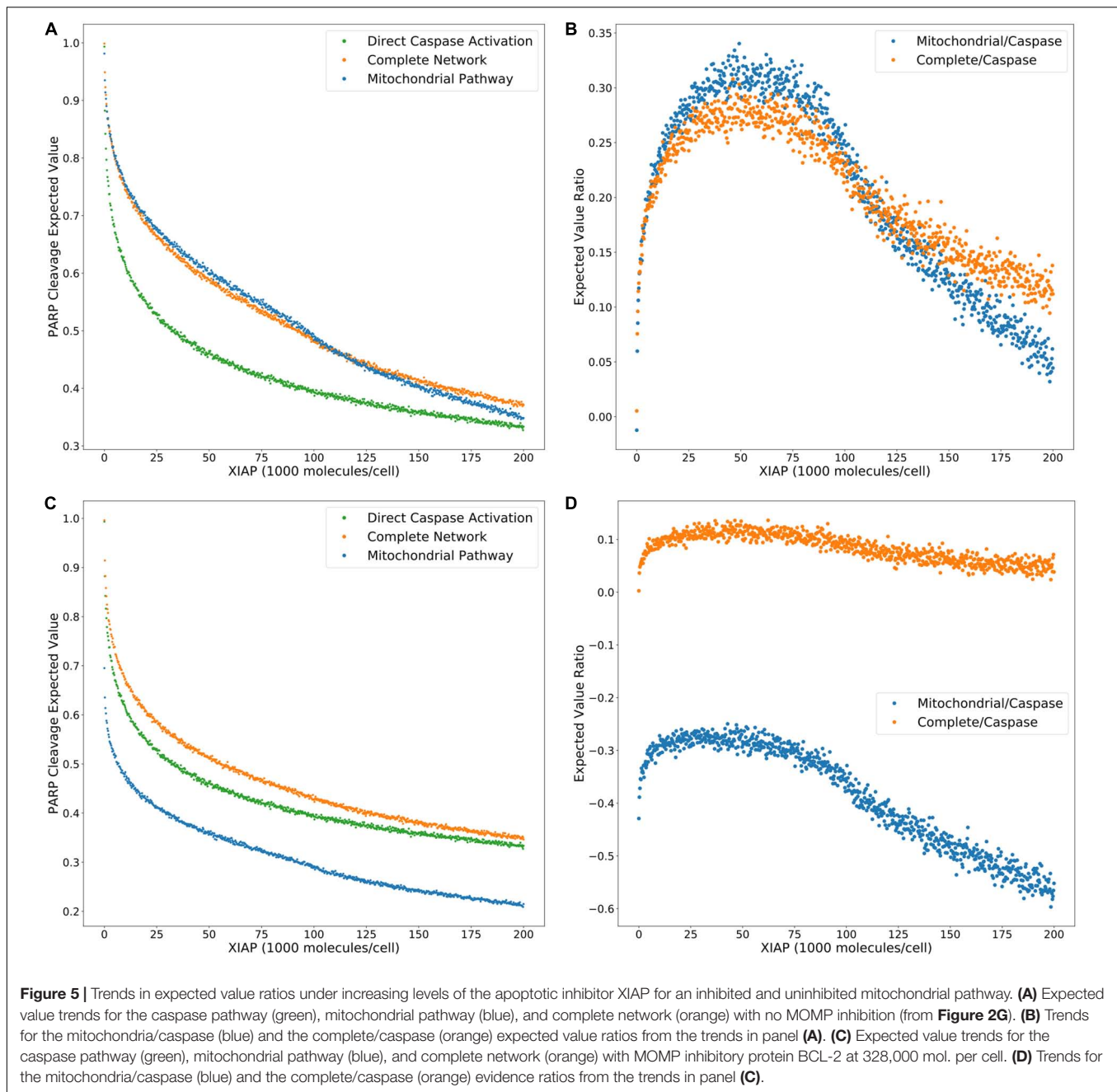
Increasing the precision of the expected value estimates and tightening their trendlines, is accomplished by increasing the number of live points in the nested sampling algorithm. The trade-off is an increase in the number of evaluations required to reach the termination of the algorithm and an accompanying increase in total computation time. **Figures 6A,B** display the required number of evaluations for the direct caspase and complete network at population sizes of 500, 1000, 2000, 4000, 8000, and 16,000, when run with the PARP cleavage objective function. For both models the number of evaluations roughly doubles for every doubling in population size. **Figures 6C,D** are the average estimated errors calculated by the MultiNest algorithm over each population size for the direct caspase and complete networks, respectively. As expected, error estimates fall roughly as $n^{-1/2}$ (Handley et al., 2015), signifying clear diminishing returns as the number of live points is increased. The average CPU process times, as estimated by Python's `time.clock()` method, are given in **Figures 6E,F** for the direct caspase and complete networks, respectively. Despite the greater number of

required evaluations for the direct caspase network the average clock times for the complete network is significantly higher. At a population of 16,000 the caspase network had an average clock time of 11,964 s compared to 76,981 for the complete network. Data for **Figure 6** can be found in **Supplementary Table S6**.

Ultimately, the choice of population size for the methods we have laid out here will depend on the networks to be compared, the objective function, and how well the trends in the expected values must be resolved in order to make inferences about network dynamics. For example, at a population size of 500 the trend in the PARP cleavage expected values for the direct caspase pathway is clearly discernable from that for the mitochondrial pathway and the complete network, but the latter two are largely overlapping (**Supplementary Figure S4A**). At higher population levels, however, two distinct mitochondrial and complete PARP cleavage trends become apparent (**Supplementary Figure S4K**). If expected value ratio trends are desired then the choice of population size must take into consideration the amplification of the noise from both expected value estimates (see **Supplementary Figures S4B,D,E,H,J,L** for complete/caspase PARP cleavage expected value trends).

DISCUSSION

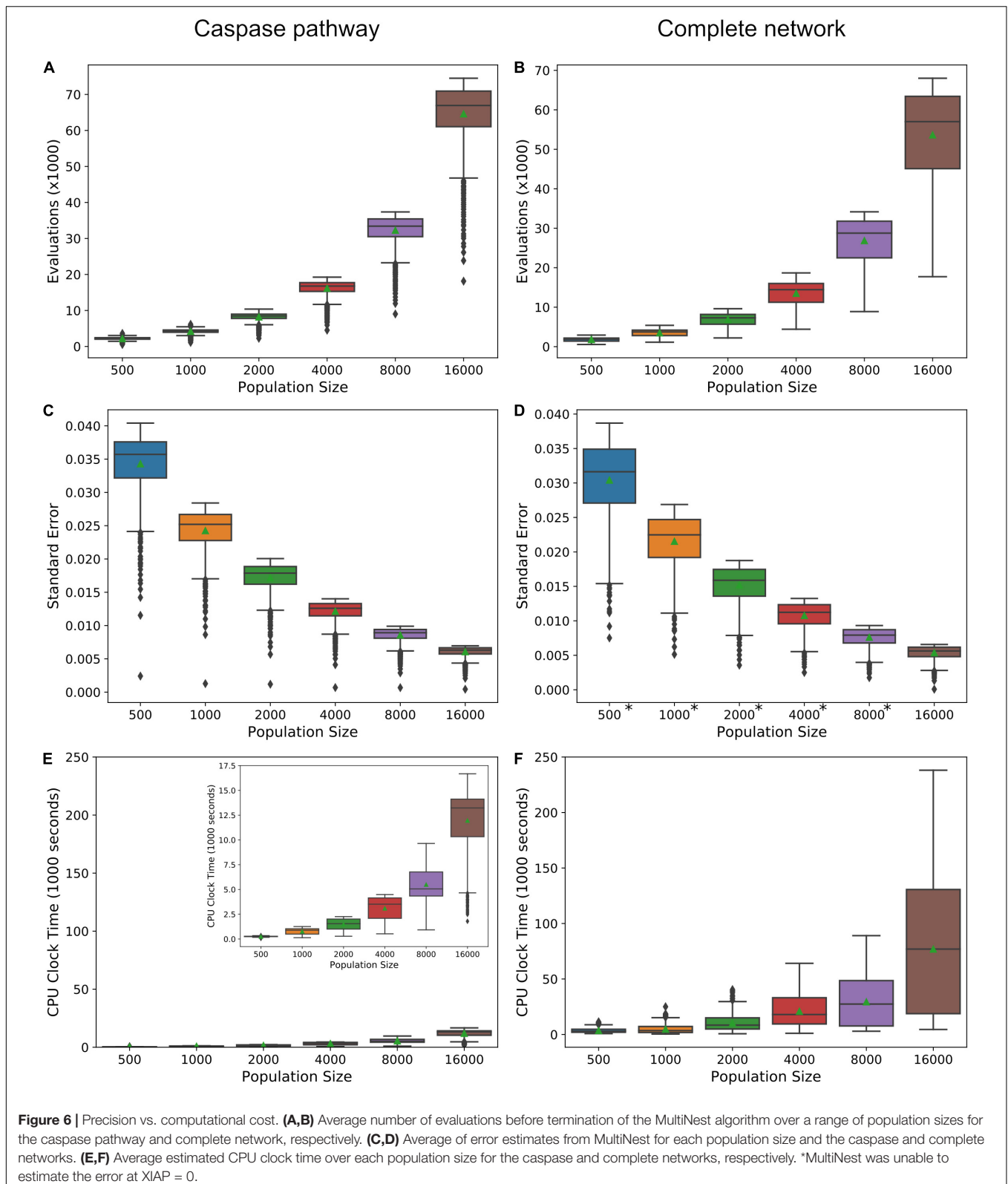
Characterizing information flow in biological networks, the interactions between various pathways or network components, and shifts in phenotype upon regulatory perturbations is a standing challenge in molecular biology. Although comparative



analysis of signal flow within a network is possible with current computational methods, the dependence of physicochemical models on unknown parameters makes the computational examination of each network component highly dependent on costly experimentation.

To take advantage of the enormous amount of existing knowledge encoded in these physicochemical networks without the dependence on explicit parameter values we take a probabilistic approach to the exploration of changes in network dynamics. By integrating an objective function that represents a simulated outcome over parameter distributions derived from existing data we obtain the likelihood of attaining that outcome

given the available information about the signaling pathways. The qualitative exploration of network behavior for various *in silico* experimental setups and regulatory conditions is then attainable without explicit knowledge of the parameter values. Although this probabilistic modeling approach is Bayesian inspired, it is a departure from strictly Bayesian methodologies. Evidence values are a relative measure of how well a model explains the data and are used as a comparative metric for model selection (Burnham and Anderson, 2002; Skilling, 2006; Feroz et al., 2009; Symonds and Moussalli, 2011; Feroz et al., 2013). The expected values calculated in this work are based solely on a given network and prior distribution; data does not *directly* come into play. There



is of course a place for data, if it exists, in the estimation of the prior parameter distributions used to calculate the expected values. Approximate Bayesian Computation, for example, can

estimate parameter distributions when a given model is too complex to be analyzed analytically, as is typical for complex biological systems (Toni et al., 2009; Toni and Stumpf, 2010). We

demonstrate the utility of the probabilistic modeling approach when applied to the regulation of extrinsic apoptosis. Networks that incorporate an active mitochondrial pathway displayed a higher resistance to apoptotic inhibition from increasing levels of XIAP, consistent with experimental evidence that XIAP induces a Type II phenotype (Jost et al., 2009). Also in line with experimental evidence (Scaffidi et al., 1998) are the results that suggest low/high signal initiation is consistent with Type II/I phenotype, respectively, and that both types achieve apoptosis equally well. The probabilistic methodology presented here has the potential to predict which proteins are potentially relevant to phenotypic outcomes and reduce the set of candidates for further perturbation experiments. Such a workflow would ultimately result in a mapping of relevant protein concentrations to those phenotypic outcomes. Moreover, by using objective functions that represent various quantitative aspects of network dynamics a more complete picture of the causal mechanisms for phenotypic outcomes can be hypothesized. For example, combining the end-product formation of cleaved PARP with the pathway flux of the apoptotic signal we hypothesized not only the conditions (regarding DISC component and XIAP concentrations) for which Type I/II or a combination of phenotypes exist, but also the roles played by both the proteins and the pathways to elicit those phenotypic responses.

A potential limitation of this probabilistic approach to the study network dynamics is the computational cost. Several factors affect the run time of the algorithm including the size of the model, the objective function, and the desired precision. Fortunately, reducing the resolution (the number of *in silico* experiments for which an expected value is estimated) and the precision (the population size) can drastically reduce the cost and in many cases the method will still be viable. One aspect of the method that is severely restrictive is the number of model components that can be varied in the same run since the computational cost increases exponentially with each additional variable. Reasonable parameter distributions must also be chosen, preferably based on existing data. Here, we were able to use generic but biologically plausible ranges with uniform distributions to produce results that were qualitatively consistent with previous experimental results. These *in silico* generated qualitative results allow us to make mechanistic hypotheses from existing data over a period of weeks rather than the months or years that would be required to attain this information with experimental approaches. Our results therefore support probabilistic approaches as a suitable complement to experimentation and a shift from purely deterministic models with a single optimum parameter set to a probabilistic understanding of mechanistic models of cellular processes.

CONCLUSION

In this paper, we have developed a probabilistic approach to the qualitative analysis of the network dynamics of physicochemical

models. It is designed to incorporate all available knowledge of the reaction topology, and the parameters on that topology, and calculate the likelihood of achieving an outcome of interest. Inferences on network dynamics are then made by repeating this calculation under changing regulatory conditions and various *in silico* experiments. We tested the method against a model of the extrinsic apoptosis system and produced qualitative results that were consistent with several lines of experimental research. To our knowledge this is the first attempt at a probabilistic analysis of network dynamics for physicochemical models and we believe this method will prove valuable for the large-scale exploration of those dynamics, particularly when parameter knowledge and data are scarce.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

MK and CL designed the project and edited the final version. MK carried out the simulations, modeling, and analysis, and wrote the first draft of the manuscript. Both authors contributed to the article and approved the submitted version.

FUNDING

This manuscript has been released as a pre-print at <https://doi.org/10.1101/732396> Kochen and Lopez (2020). This work was supported by the NIH NCI U01CA215845 to CL, as well as NSF MCB 1411482 to CL. MK was also supported by NIH NLM T15LM007450.

ACKNOWLEDGMENTS

We thank the Incyte-Vanderbilt Alliance to CL for their support of this project. We also thank the Advanced Computing Center for Research and Education (ACCRE) for computational resources and support needed to complete this work. We also extend our thanks to Dr. Blake Wilson for his effort in reviewing this work and providing feedback.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00686/full#supplementary-material>

REFERENCES

- Adrain, C., Creagh, E. M., and Martin, S. J. (2001). Apoptosis-associated release of Smac/DIABLO from mitochondria requires active caspases and is blocked by Bcl-2. *EMBO J.* 20, 6627–6636. doi: 10.1093/emboj/20.23.6627
- Aitken, S., and Akman, O. E. (2013). Nested sampling for parameter inference in systems biology: application to an exemplar circadian model. *BMC Syst. Biol.* 7:72. doi: 10.1186/1752-0509-7-72
- Albeck, J. G., Burke, J. M., Spencer, S. L., Lauffenburger, D. A., and Sorger, P. K. (2008). Modeling a snap-action, variable-delay switch controlling extrinsic cell death. *PLoS Biol.* 6:e299. doi: 10.1371/journal.pbio.0060299
- Aldridge, B. B., Burke, J. M., Lauffenburger, D. A., and Sorger, P. K. (2006). Physicochemical modelling of cell signalling pathways. *Nat. Cell Biol.* 8, 1195–1203. doi: 10.1038/ncb1497
- Aldridge, B. B., Gaudet, S., Lauffenburger, D. A., and Sorger, P. K. (2011). Lyapunov exponents and phase diagrams reveal multi-factorial control over TRAIL-induced apoptosis. *Mol. Syst. Biol.* 7:553. doi: 10.1038/msb.2011.85
- Anderson, M. W., Moss, J. J., Szalai, R., and Lane, J. D. (2019). Mathematical modeling highlights the complex role of BCL2 in TRAIL-induced apoptosis of colorectal carcinoma cells. *iScience* 12, 182–193. doi: 10.1016/j.isci.2019.01.015
- Ashkenazi, A., and Dixit, V. M. (1998). Death receptors: signaling and modulation. *Science* 281, 1305–1308. doi: 10.1126/science.281.5381.1305
- Bagci, E. Z., Vodovotz, Y., Billiar, T. R., Ermentrout, G. B., and Bahar, I. (2006). Bistability in apoptosis: roles of bax, Bcl-2, and mitochondrial permeability transition pores. *Biophys. J.* 90, 1546–1559. doi: 10.1529/biophysj.105.068122
- Bentele, M., Lavrik, I., Ulrich, M., Stöfer, S., Heermann, D. W., Kalthoff, H., et al. (2004). Mathematical modeling reveals threshold mechanism in CD95-induced apoptosis. *J. Cell Biol.* 166, 839–851. doi: 10.1083/jcb.200404158
- Bhalla, U. S., and Iyengar, R. (1999). Emergent properties of networks of biological signaling pathways. *Science* 283, 381–387. doi: 10.1126/science.283.5400.381
- Boatright, K. M., and Salvesen, G. S. (2003). Mechanisms of caspase activation. *Curr. Opin. Cell Biol.* 15, 725–731. doi: 10.1016/j.ceb.2003.10.009
- Boldin, M. P., Varfolomeev, E. E., Pancer, Z., Mett, I. L., Camonis, J. H., and Wallach, D. (1995). A novel protein that interacts with the death domain of Fas/APO1 contains a sequence motif related to the death domain. *J. Biol. Chem.* 270, 7795–7798. doi: 10.1074/jbc.270.14.7795
- Buchner, J., Georgakakis, A., Nandra, K., Hsu, L., Rangel, C., Brightman, M., et al. (2014). X-ray spectral modelling of the AGN obscuring region in the CDFS: bayesian model selection and catalogue. *Astron. Astrophys.* 564:A125.
- Burnham, K. P., and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach [Internet]*, 2nd Edn. New York, NY: Springer-Verlag.
- Certo, M., Moore, V. D. G., Nishino, M., Wei, G., Korsmeyer, S., Armstrong, S. A., et al. (2006). Mitochondria primed by death signals determine cellular addiction to antiapoptotic BCL-2 family members. *Cancer Cell* 9, 351–365. doi: 10.1016/j.ccr.2006.03.027
- Cowling, V., and Downward, J. (2002). Caspase-6 is the direct activator of caspase-8 in the cytochrome c-induced apoptosis pathway: absolute requirement for removal of caspase-6 prodomain. *Cell Death Differ.* 9, 1046–1056. doi: 10.1038/sj.cdd.4401065
- Dai, H., Ding, H., Peterson, K. L., Meng, X. W., Schneider, P. A., Knorr, K. L. B., et al. (2018). Measurement of BH3-only protein tolerance. *Cell Death Differ.* 25, 282–293. doi: 10.1038/cdd.2017.156
- Desagher, S., Osen-Sand, A., Nichols, A., Eskes, R., Montessuit, S., Lauper, S., et al. (1999). Bid-induced conformational change of bax is responsible for mitochondrial cytochrome c release during apoptosis. *J. Cell Biol.* 144, 891–901. doi: 10.1083/jcb.144.5.891
- Eissing, T., Conzelmann, H., Gilles, E. D., Allgöwer, F., Bullinger, E., and Scheurich, P. (2004). Bistability analyses of a caspase activation model for receptor-induced apoptosis. *J. Biol. Chem.* 279, 36892–36897. doi: 10.1074/jbc.m404893200
- Eydghi, H., Chen, W. W., Muhlich, J. L., Vitkup, D., Tsitsiklis, J. N., and Sorger, P. K. (2013). Properties of cell death models calibrated and compared using Bayesian approaches. *Mol. Syst. Biol.* 9:644. doi: 10.1038/msb.2012.69
- Feroz, F., Hobson, M. P., and Bridges, M. (2009). MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics. *Mon. Not. R. Astron. Soc.* 398, 1601–1614. doi: 10.1111/j.1365-2966.2009.14548.x
- Feroz, F., Hobson, M. P., Cameron, E., and Pettitt, A. N. (2013). Importance nested sampling and the multinest algorithm. *arXiv [Preprint]*
- Handley, W. J., Hobson, M. P., and Lasenby, A. N. (2015). PolyChord: next-generation nested sampling. *Mon. Not. R. Astron. Soc.* 453, 4385–4399. doi: 10.1093/mnras/stv1911
- Ho, K. L., and Harrington, H. A. (2010). Bistability in apoptosis by receptor clustering. *PLoS Comput. Biol.* 6:e1000956. doi: 10.1371/journal.pcbi.1000956
- Huang, Y., Park, Y. C., Rich, R. L., Segal, D., Myszk, D. G., and Wu, H. (2001). Structural basis of caspase inhibition by XIAP: differential roles of the linker versus the BIR domain. *Cell* 104, 781–790. doi: 10.1016/s0092-8674(02)02075-5
- Jost, P. J., Grabow, S., Gray, D., McKenzie, M. D., Nachbur, U., Huang, D. C. S., et al. (2009). XIAP discriminates between type I and type II FAS-induced apoptosis. *Nature* 460, 1035–1039. doi: 10.1038/nature08229
- Kale, J., Osterlund, E. J., and Andrews, D. W. (2018). BCL-2 family proteins: changing partners in the dance towards death. *Cell Death Differ.* 25, 65–80. doi: 10.1038/cdd.2017.186
- Kelekar, A., and Thompson, C. B. (1998). Bcl-2-family proteins: the role of the BH3 domain in apoptosis. *Trends Cell Biol.* 8, 324–330. doi: 10.1016/s0962-8924(98)01321-x
- Kischkel, F. C., Lawrence, D. A., Chuntharapai, A., Schow, P., Kim, K. J., and Ashkenazi, A. (2000). Apo2L/TRAIL-dependent recruitment of endogenous FADD and caspase-8 to death receptors 4 and 5. *Immunity* 12, 611–620. doi: 10.1016/s1074-7613(00)80212-5
- Kitano, H. (2002). Computational systems biology. *Nature* 420, 206–210.
- Kochen, M. A., and Lopez, C. F. (2020). A probabilistic approach to explore signal execution mechanisms with limited experimental data. *bioRxiv [Preprint]*
- Krueger, A., Schmitz, I., Baumann, S., Krammer, P. H., and Kirchhoff, S. (2001). Cellular FLICE-inhibitory protein splice variants inhibit different steps of caspase-8 activation at the CD95 death-inducing signaling complex. *J. Biol. Chem.* 276, 20633–20640. doi: 10.1074/jbc.m101780200
- Le Novère, N. (2015). Quantitative and logic modelling of molecular and gene networks. *Nat. Rev. Genet.* 16, 146–158. doi: 10.1038/nrg3885
- Leber, B., Lin, J., and Andrews, D. W. (2007). Embedded together: the life and death consequences of interaction of the Bcl-2 family with membranes. *Apoptosis* 12, 897–911. doi: 10.1007/s10495-007-0746-4
- Legewie, S., Blüthgen, N., and Herzog, H. (2006). Mathematical modeling identifies inhibitors of apoptosis as mediators of positive feedback and bistability. *PLoS Comput. Biol.* 2:e120. doi: 10.1371/journal.pcbi.0020120
- Letai, A., Bassik, M. C., Walensky, L. D., Sorcinelli, M. D., Weiler, S., and Korsmeyer, S. J. (2002). Distinct BH3 domains either sensitize or activate mitochondrial apoptosis, serving as prototype cancer therapeutics. *Cancer Cell* 2, 183–192. doi: 10.1016/s1535-6108(02)00127-7
- Li, H., Zhu, H., Xu, C., and Yuan, J. (1998). Cleavage of BID by caspase 8 mediates the mitochondrial damage in the Fas pathway of apoptosis. *Cell* 94, 491–501. doi: 10.1016/s0092-8674(00)81590-1
- Lopez, C. F., Muhlich, J. L., Bachman, J. A., and Sorger, P. K. (2013). Programming biological models in Python using PySB. *Mol. Syst. Biol.* 9:646. doi: 10.1038/msb.2013.1
- Loscalzo, J., and Barabasi, A.-L. (2011). Systems biology and the future of medicine. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 3, 619–627. doi: 10.1002/wsbm.144
- Luo, X., Budihardjo, I., Zou, H., Slaughter, C., and Wang, X. (1998). Bid, a Bcl2 interacting protein, mediates cytochrome c release from mitochondria in response to activation of cell surface death receptors. *Cell* 94, 481–490. doi: 10.1016/s0092-8674(00)81589-5
- MacKay, D. J. C., and Kay, D. J. C. M. (2003). *Information Theory, Inference and Learning Algorithms*. Cambridge: Cambridge University Press, 696.
- Martin, D. A., Siegel, R. M., Zheng, L., and Lenardo, M. J. (1998). Membrane oligomerization and cleavage activates the caspase-8 (FLICE/MACHo1) death signal. *J. Biol. Chem.* 273, 4345–4349. doi: 10.1074/jbc.273.8.4345
- Mitra, E. D., Suderman, R., Colvin, J., Ionkov, A., Hu, A., Sauro, H. M., et al. (2019). PyBioNetFit and the biological property specification language. *iScience* 19, 1012–1036. doi: 10.1016/j.isci.2019.08.045
- Nicholson, D. W., Ali, A., Thornberry, N. A., Vaillancourt, J. P., Ding, C. K., Gallant, M., et al. (1995). Identification and inhibition of the ICE/CED-3 protease necessary for mammalian apoptosis. *Nature* 376, 37–43. doi: 10.1038/376037a0

- Oltval, Z. N., Milliman, C. L., and Korsmeyer, S. J. (1993). Bcl-2 heterodimerizes in vivo with a conserved homolog, Bax, that accelerates programmed cell death. *Cell* 74, 609–619. doi: 10.1016/0092-8674(93)90509-o
- Pullen, N., and Morris, R. J. (2014). Bayesian model comparison and parameter inference in systems biology using nested sampling. *PLoS One* 9:e88419. doi: 10.1371/journal.pone.0088419
- Raychaudhuri, S., Willgoos, E., Nguyen, T.-N., Khan, E. M., and Goldkorn, T. (2008). Monte carlo simulation of cell death signaling predicts large cell-to-cell stochastic fluctuations through the type 2 pathway of apoptosis. *Biophys. J.* 95, 3559–3562. doi: 10.1529/biophysj.108.135483
- Salvesen, G. S., and Dixit, V. M. (1999). Caspase activation: the induced-proximity model. *Proc. Natl. Acad. Sci. U.S.A.* 96, 10964–10967. doi: 10.1073/pnas.96.20.10964
- Scaffidi, C., Fulda, S., Srinivasan, A., Friesen, C., Li, F., Tomaselli, K. J., et al. (1998). Two CD95 (APO-1/Fas) signaling pathways. *EMBO J.* 17, 1675–1687. doi: 10.1093/emboj/17.6.1675
- Scaffidi, C., Schmitz, I., Zha, J., Korsmeyer, S. J., Krammer, P. H., and Peter, M. E. (1999). Differential modulation of apoptosis sensitivity in CD95 type I and type II cells. *J. Biol. Chem.* 274, 22532–22538. doi: 10.1074/jbc.274.32.22532
- Schlatter, R., Schmich, K., Vizcarra, I. A., Scheurich, P., Sauter, T., Borner, C., et al. (2009). ON/OFF and beyond - a Boolean model of apoptosis. *PLoS Comput. Biol.* 5:e1000595. doi: 10.1371/journal.pcbi.1000595
- Schleich, K., and Lavrik, I. N. (2013). Mathematical modeling of apoptosis. *Cell Commun. Signal.* 11:44. doi: 10.1186/1478-811x-11-44
- Shiozaki, E. N., Chai, J., Rigotti, D. J., Riedl, S. J., Li, P., Srinivasula, S. M., et al. (2003). Mechanism of XIAP-mediated inhibition of caspase-9. *Mol. Cell* 11, 519–527. doi: 10.1016/s1097-2765(03)00054-6
- Shockley, E. M., Rouzer, C. A., Marnett, L. J., Deeds, E. J., and Lopez, C. F. (2019). Signal integration and information transfer in an allosterically regulated network. *NPJ Syst. Biol. Appl.* 5, 1–9.
- Shockley, E. M., Vrugt, J. A., and Lopez, C. F. (2018). PyDREAM: high-dimensional parameter inference for biological models in python. *Bioinformatics* 34, 695–697. doi: 10.1093/bioinformatics/btx626
- Skilling, J. (2006). Nested sampling for general Bayesian computation. *Bayesian Anal.* 1, 833–859. doi: 10.1214/06-ba127
- Spencer, S. L., Gaudet, S., Albeck, J. G., Burke, J. M., and Sorger, P. K. (2009). Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature* 459, 428–432. doi: 10.1038/nature08012
- Spencer, S. L., and Sorger, P. K. (2011). Measuring and modeling apoptosis in single cells. *Cell* 144, 926–939. doi: 10.1016/j.cell.2011.03.002
- Sprick, M. R., Weigand, M. A., Rieser, E., Rauch, C. T., Juo, P., Blenis, J., et al. (2000). FADD/MORT1 and caspase-8 are recruited to TRAIL receptors 1 and 2 and are essential for apoptosis mediated by TRAIL receptor 2. *Immunity* 12, 599–609. doi: 10.1016/s1074-7613(00)80211-3
- Stennicke, H. R., Jürgensmeier, J. M., Shin, H., Deveraux, Q., Wolf, B. B., Yang, X., et al. (1998). Pro-caspase-3 is a major physiologic target of caspase-8. *J. Biol. Chem.* 273, 27084–27090. doi: 10.1074/jbc.273.42.27084
- Suzuki, Y., Nakabayashi, Y., and Takahashi, R. (2001). Ubiquitin-protein ligase activity of X-linked inhibitor of apoptosis protein promotes proteasomal degradation of caspase-3 and enhances its anti-apoptotic effect in Fas-induced cell death. *Proc. Natl. Acad. Sci. U.S.A.* 98, 8662–8667. doi: 10.1073/pnas.161506698
- Symonds, M. R. E., and Moussalli, A. (2011). A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behav. Ecol. Sociobiol.* 65, 13–21. doi: 10.1007/s00265-010-1037-6
- Tewari, M., Quan, L. T., O'Rourke, K., Desnoyers, S., Zeng, Z., Beidler, D. R., et al. (1995). Yama/CPP32 β , a mammalian homolog of CED-3, is a CrmA-inhibitable protease that cleaves the death substrate poly(ADP-ribose) polymerase. *Cell* 81, 801–809. doi: 10.1016/0092-8674(95)90541-3
- Toni, T., and Stumpf, M. P. H. (2010). Simulation-based model selection for dynamical systems in systems and population biology. *Bioinformatics* 26, 104–110. doi: 10.1093/bioinformatics/btp619
- Toni, T., Welch, D., Strelkowa, N., Ipsen, A., and Stumpf, M. P. H. (2009). Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J. R. Soc. Interface* 6, 187–202. doi: 10.1098/rsif.2008.0172
- van Riel, N. A. W. (2006). Dynamic modelling and analysis of biochemical networks: mechanism-based models and model-based experiments. *Brief. Bioinform.* 7, 364–374. doi: 10.1093/bib/bbl040
- Wrede, F., and Hellander, A. (2018). Smart computational exploration of stochastic gene regulatory network models using human-in-the-loop semi-supervised learning. *Bioinformatics* 35, 5199–5206. doi: 10.1093/bioinformatics/btz420
- Würstle, M. L., Zink, E., Prehn, J. H. M., and Rehm, M. (2014). From computational modelling of the intrinsic apoptosis pathway to a systems-based analysis of chemotherapy resistance: achievements, perspectives and challenges in systems medicine. *Cell Death Dis.* 5:e1258.
- Xu, T.-R., Vysheirsky, V., Gormand, A., von Kriegsheim, A., Girolami, M., Baillie, G. S., et al. (2010). Inferring signaling pathway topologies from multiple perturbation measurements of specific biochemical species. *Sci. Signal.* 3:ra20. doi: 10.1126/scisignal.2000517
- Yang, E., Zha, J., Jockel, J., Boise, L. H., Thompson, C. B., and Korsmeyer, S. J. (1995). Bad, a heterodimeric partner for Bcl-xL and Bcl-2, displaces bax and promotes cell death. *Cell* 80, 285–291. doi: 10.1016/0092-8674(95)90411-5
- Zou, H., Li, Y., Liu, X., and Wang, X. (1999). An APAF-1-cytochrome c multimeric complex is a functional apoptosome that activates procaspase-9. *J. Biol. Chem.* 274, 11549–11556.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Kochen and Lopez. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Understanding the Principles of Pattern Formation Driven by Notch Signaling by Integrating Experiments and Theoretical Models

Federico Bocci¹, José Nelson Onuchic^{1,2,3,4*} and Mohit Kumar Jolly^{5*}

¹ Center for Theoretical Biological Physics, Rice University, Houston, TX, United States, ² Department of Physics and Astronomy, Rice University, Houston, TX, United States, ³ Department of Chemistry, Rice University, Houston, TX, United States, ⁴ Department of Biosciences, Rice University, Houston, TX, United States, ⁵ Centre for BioSystems Science and Engineering, Indian Institute of Science, Bengaluru, India

OPEN ACCESS

Edited by:

Chunhe Li,
Fudan University, China

Reviewed by:

Huijing Du,
University of Nebraska System,
United States
Nathan Weinstein,
Universidad Nacional Autónoma
de México, Mexico

*Correspondence:

José Nelson Onuchic
jonuchic@rice.edu
Mohit Kumar Jolly
mkjolly@iisc.ac.in;
mkjolly.15@gmail.com

Specialty section:

This article was submitted to
Systems Biology,
a section of the journal
Frontiers in Physiology

Received: 08 May 2020

Accepted: 10 July 2020

Published: 31 July 2020

Citation:

Bocci F, Onuchic JN and Jolly MK
(2020) Understanding the Principles
of Pattern Formation Driven by Notch
Signaling by Integrating Experiments
and Theoretical Models.
Front. Physiol. 11:929.
doi: 10.3389/fphys.2020.00929

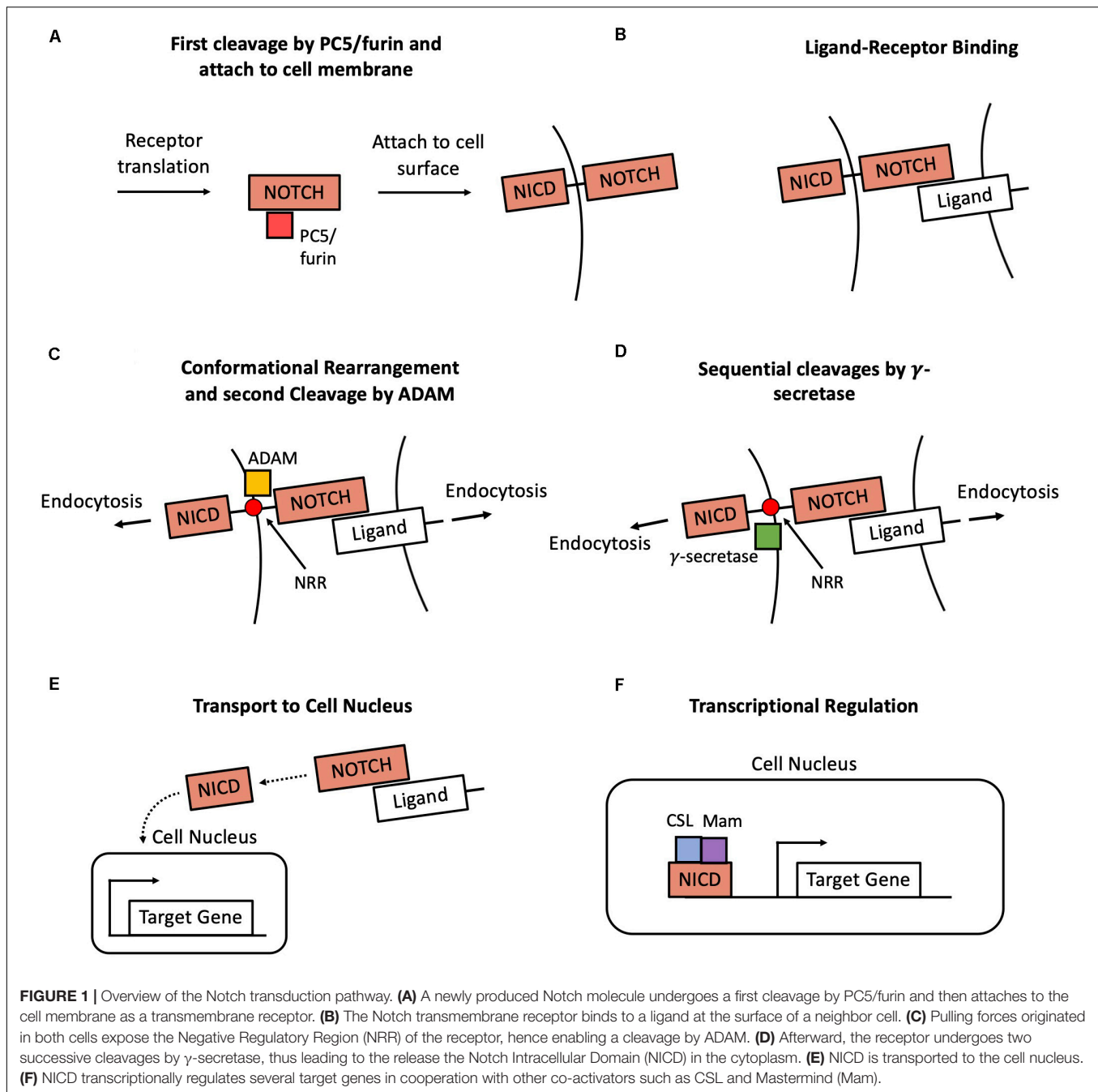
Notch signaling is an evolutionary conserved cell-cell communication pathway. Besides regulating cell-fate decisions at an individual cell level, Notch signaling coordinates the emergent spatiotemporal patterning in a tissue through ligand-receptor interactions among transmembrane molecules of neighboring cells, as seen in embryonic development, angiogenesis, or wound healing. Due to its ubiquitous nature, Notch signaling is also implicated in several aspects of cancer progression, including tumor angiogenesis, stemness of cancer cells and cellular invasion. Here, we review experimental and computational models that help understand the operating principles of cell patterning driven by Notch signaling. First, we discuss the basic mechanisms of spatial patterning via canonical lateral inhibition and lateral induction mechanisms, including examples from angiogenesis, inner ear development and cancer metastasis. Next, we analyze additional layers of complexity in the Notch pathway, including the effect of varying cell sizes and shapes, ligand-receptor binding within the same cell, variable binding affinity of different ligand/receptor subtypes, and filopodia. Finally, we discuss some recent evidence of mechanosensitivity in the Notch pathway in driving collective epithelial cell migration and cardiovascular morphogenesis.

Keywords: Notch, Delta, Jagged, lateral inhibition, lateral induction, spatial pattern, mathematical modeling, cell-cell signaling

INTRODUCTION

Notch signaling is central in cell fate decisions and therefore it is one of the most well-conserved transduction pathways in metazoans (Bray, 2016). In its simpler form, the signaling cascade includes only a limited number of well-conserved steps, including ligand binding to the Notch transmembrane receptor, release of the Notch intracellular domain (NICD) in the cytoplasm, and downstream regulation of NICD on its target genes (**Figure 1**; Bray, 2016; Kovall et al., 2017; Sjöqvist and Andersson, 2019). Despite this simplicity, Notch regulates a multitude of different biological processes including cell differentiation, proliferation and death (Bray, 2016).

Each of the abovementioned steps in this cascade raises unanswered questions that would improve our understanding of several developmental processes and may also provide key insights



to alleviate many pathological conditions, including cancer (Li et al., 2014; Bray, 2016; Kovall et al., 2017; Siebel and Lendahl, 2017; Sjöqvist and Andersson, 2019). Here, we explicitly focus on the role of Notch signaling in coordinating cell fate decisions and patterning at a multicellular level, and how various experimental and computational models can be integrated to elucidate the underlying dynamical principles of pattern formation. Due to its multi-cellular nature, Notch signaling offers an opportunity to understand how cell-fate decision in individual cells may be relayed to generate emergent multi-cellular dynamics. Different Notch ligands can orchestrate different principles of multicellular

spatial patterning via different positive and negative feedback regulation between NICD and its transcriptional targets (Bray, 2006). For instance, Notch signaling can coordinate a divergent cell fate between two neighboring cells, a process known as lateral inhibition (Bray, 2016). Moreover, Notch can modulate the opposite process, the lateral induction (Hartman et al., 2010; Petrovic et al., 2014), by coordinating a similar cell state among neighbors.

In this review, we offer a bird's eye view on how to interpret cell-level and tissue-level dynamics with simple concepts such as lateral inhibition and lateral induction, discuss the limitations

of these models, and highlight a novel set of questions that require integrating experimental investigation with concepts from quantitative mechanistic modeling. In doing so, we bring together the analysis of several biological systems as well as theoretical modeling approaches that highlight the emergence of common themes in the Notch pathway. For the sake of simplicity, technical details of the underlying biology and mathematical models have been occasionally omitted, and relevant literature has been suggested. Furthermore, given the extensive set of topics covered in this review, we have focused on certain experimental and/or theoretical models that are representative of a particular system, and pointed the interested readers to relevant reviews for in-depth discussions of specific areas of research.

First, we review some aspects of the Notch signaling cascade that are necessary to understand Notch-driven pattern formation. It is followed by a discussion of various modeling approaches that can be used to understand the operating principles of Notch. After these two introductory sections, we discuss the principles of Notch-driven patterning. We analyze how Notch signaling gives rise to divergent cell fate – lateral inhibition – and convergent cell fate – lateral induction – among neighboring cells. Experimental evidence and theoretical modeling have contributed to understanding the competition and synergy between these patterning mechanisms in various physiological and pathological systems, including angiogenesis, inner ear development and cancer metastasis. Moreover, we review the oscillatory dynamics of Notch signaling that can arise due to coupling with other signaling pathways, for instance, during somitogenesis. Further, we examine the role of various molecular and morphological features that introduce additional layers of complexity to the canonical Notch signaling outcomes. The scenarios discussed here include the role of cell shape and packing geometry, cis-interactions between molecules within the same cell, mechanisms that alter the binding affinity between ligand and receptor paralogs, and beyond-nearest neighbor signaling through filopodia. In the final section, we review evidence pointing to a role for mechanosensitivity in assisting Notch-driven cell-fate decision. Relevant examples discussed here include collective epithelial cell migration and cardiovascular morphogenesis.

OVERVIEW OF NOTCH SIGNALING

In this section, we discuss the main components and steps of the Notch signaling cascade. We will avoid excessive details on the molecular structure of the Notch receptor and ligands that are not required for the topics discussed in this review.

The main steps of the Notch signaling cascade are very well conserved across several organisms and include production and targeting of the Notch receptor to the cell membrane, ligand-receptor binding, conformational rearrangement of the receptor, release of the intracellular domain (NICD) and downstream transcriptional regulation (**Figure 1**). First, a newly produced Notch receptor molecule is glycosylated by the enzymes O-fut and Rumi, and successively subjected to proteolytic cleavage by the PC5/furin at site 1 (S1) (Kopan and Ilagan, 2009).

Afterward, the mature Notch molecule attaches to the cell surface as a transmembrane receptor (**Figure 1A**). The signaling is initiated with the binding of an extracellular ligand to the transmembrane Notch receptor (**Figure 1B**). Typically, the ligand is a transmembrane protein at the surface of a neighboring cell (juxtacrine signaling), but it can occasionally be a soluble ligand in the extracellular microenvironment (paracrine signaling) (D'Souza et al., 2010). In particular, two classes of ligands, referred to as Delta-like and Jagged-like, can bind to the Notch receptors. The ligand-receptor binding and forces originated by endocytosis induce a conformational change in the structure of the Notch receptor. This modification exposes a previously shielded region of the receptor, the Negative regulatory region (NRR). Following this conformational change, the receptor sequentially undergoes a cleavage by the enzymes ADAM at site 2 (S2, **Figure 1C**) and two cleavages by γ -secretase at sites 3 and 4 (S3–S4, **Figure 1D**; Kopan and Ilagan, 2009), resulting in the release of the NICD in the cytoplasm (**Figures 1B,C**). The NICD translocates to the cell nucleus, where it regulates several target genes together with cooperating transcriptional cofactors such as CSL and Mastermind (Mam) (**Figures 1E,F**; Bray, 2016).

Notably, NICD regulates the transcription of the Notch receptor and its ligands, either in a direct or indirect manner. Specifically, NICD promotes the transcription of Hey/Hes1 (Shimojo et al., 2011) – an inhibitor of Delta – while directly activating Notch and Jagged (Manderfield et al., 2012). Therefore, Notch signaling introduces a biochemical feedback between neighboring cells that coordinates their cell fate decision (Shaya and Sprinzak, 2011; Sjöqvist and Andersson, 2019). The implications of these biochemical feedbacks in multicellular patterning are the subject of the section about “Spatiotemporal Patterning Guided by Notch Signaling.”

Although the main steps of the signaling are quite general, there are specific aspects that differ from one organism to another, or even from cell to cell – these will be the focus of the section on “Non-canonical Modulation of Notch Signaling.” First, the signaling depends on cell-cell contact area because physical contact is required for juxtacrine signaling (Shaya et al., 2017). We will discuss how cell size modulates Notch signaling and plays a role in determining cell fate. Furthermore, ligands and receptors can bind within the same cell, thus leading to degradation of the ligand-receptor complex without release of NICD (referred as cis-inhibition). Despite not leading to NICD release, cis-inhibition plays a pivotal role by sequestering ligands and receptors that would otherwise contribute to active signaling (Celis and de Bray, 1997; Sprinzak et al., 2010). Third, different model organisms have different number of ligand and receptor subtypes. *Drosophila melanogaster* – where Notch was firstly extensively characterized – has one type of Notch receptor and two types of ligands (Delta and Serrate, equivalent of Jagged). Conversely, most mammalian organisms have four Notch paralogs (Notch1, Notch2, Notch3, Notch4), three Delta-like ligands (Dll1, Dll3, Dll4) and two Jagged-like ligands (Jagged1, Jagged2). **Table 1** offers a comparison of the main components of the signaling between several popular model organisms. Different pairs of ligand and receptor subtypes possess different binding affinities and have been even associated

with different biological functions (Bray, 2016; Sjöqvist and Andersson, 2019). Finally, the signaling can be occasionally extended beyond nearest neighbors via filopodia that introduce transient contacts between second or third-nearest neighbor cells, as seen for instance in hair cell patterning during *Drosophila* wing development (Cohen et al., 2010).

Further details on the signaling cascade will not be considered here; additional information can be found in several excellent reviews (Kopan and Ilagan, 2009; Bray, 2016; Kovall et al., 2017; Sjöqvist and Andersson, 2019). Here, we focus on generic principles of multicellular patterning obtained via Notch signaling as a whole.

MATHEMATICAL FORMALISM TO DESCRIBE NOTCH SIGNALING

In this section, we briefly overview different classes of theoretical models that have been applied to Notch signaling.

Many mathematical models aim at reconstructing the dynamics of mutually interacting biochemical species and/or genes in the Notch pathway with ordinary differential equations (ODEs). In these models, each chemical species/gene is described by a variable (X), which can either represent a concentration or copy number. In many cases, since molecular copy numbers are large, X is treated as a continuous variable that obeys an ODE of the form:

$$\frac{dX}{dt} = K_{\text{prod}} - \Gamma_{\text{Degr}} \quad (1)$$

In this equation, K_{prod} represents any biochemical process that regulates the production of X , potentially including constitutive transcription, transcriptional activation or inhibition, translation and any other post-translational interaction that might be relevant in a specific system. Transcriptional regulation of NICD on the Notch receptors and ligands is typically described with Hill functions:

$$K_{\text{prod}} = K_0 \frac{1 + \lambda \left(\frac{\text{NICD}}{S_0} \right)^n}{1 + \left(\frac{\text{NICD}}{S_0} \right)^n} \quad (2)$$

In this expression, K_0 is the basal transcription rate in absence of NICD, S_0 is a threshold concentration of NICD, λ is a fold-change and n is a coefficient that regulates how steeply transcription changes as a function of NICD. At low NICD ($\text{NICD} \ll S_0$), there is only constitutive production ($K_{\text{prod}} = K_0$). Conversely, at high NICD ($\text{NICD} \gg S_0$), the transcription rate is scaled by a fold-change ($K_{\text{prod}} = K_0 \lambda$). Therefore, $\lambda < 1$ implies a decrease of transcription rate (inhibition), while $\lambda > 1$ (activation) implies an increase of transcription rate (Figure 2A). Depending on the model, slightly different mathematical definitions might be found for this function.

Γ_{Degr} generically represents a loss term. Loss due to molecule degradation and dilution is usually modeled with a linear function $\Gamma_{\text{Degr}} = \gamma X$, where γ is the inverse half-life of X . In the specific case of Notch signaling, intracellular signaling is coupled with ligand-receptor binding (Figure 2B). Therefore, X can represent a receptor or ligand that binds to another

ligand/receptor and degrades after NICD release. This is often modeled with a chemical reaction term, thus $\Gamma_{\text{Degr}} = kXY + \gamma X$, where Y represents the concentration (or copy number) of a ligand or receptor that binds to X , and k is the ligand-receptor binding rate constant.

Therefore, a network of N interacting biochemical species or genes, such as the intracellular signaling network sketched in Figure 2B, can be described by a collection of variables ($X_1, X_2, \dots, X_N$) and a set of N ODEs of the form of Eq. 1. In such system of equations, the production term for X_i ($K_{\text{prod}}^{(i)}$) describes the regulation on X_i due to interactions with all other species in the network.

It is worth mentioning that biochemical and gene regulatory networks are sometimes modeled with Boolean, rather than continuous, variables. A Boolean variable can only assume two states $X = 0, 1$ corresponding to an inactive or active chemical species/gene, respectively. At any given time, the state of a variable (X) is determined by the incoming signal from all other chemical species/genes that interact with X .

These models of intracellular dynamics can be generalized to a multicellular scenario by arranging cells in a discrete lattice (Figure 2C). In these models, the intracellular signaling dynamics is still described by a set of ODEs. In the specific case of Notch signaling, the biochemical circuits within each cell are coupled by ligands-receptors binding between neighbors (see again Figure 2B). These lattices can have different geometries (square, hexagonal) or can be disordered to study the effect of cell size and shape (Formosa-Jordan and Ibáñez, 2009; Boareto et al., 2015a; Shaya et al., 2017).

Lattice model, however, assume a rigid arrangement of cells on a grid, and cannot take into account biophysical processes including cell migration and cell growth. Biochemical signaling and cell-level behavior can be integrated in agent-based models (Figure 2D). In agent-based models, space is discretized into small volumes, and each cell is represented by the collection of grid points sharing the same kind. This modeling approach has been successfully applied, for instance, in the context of sprouting angiogenesis where cells modify their morphology and migrate to give rise to new blood vessels (Bentley et al., 2008).

Finally, a second possibility to couple biochemical signaling with cell-level dynamics is provided by off-lattice models (Figure 2E). In off-lattice models, the membrane of a cell is described by a set of N points connected together according to a pre-defined rule, such as elastic springs (Du et al., 2015). Therefore, the motion of these connected membrane points defines the volume occupied by a cell. In the context of Notch signaling, off-lattice model must further include ligand-receptor binding between neighbors. Stopka et al. (2019) recently developed an off-lattice, multicell model of Notch signaling where membrane points of neighboring cells share adhesion junctions (modeled as elastic springs). Therefore, the number of shared junctions between neighbors modulates the amount of signaling between cells (Stopka et al., 2019).

In both agent-based and off lattice models, the signaling dynamics within each cell can still be described by a set of ODEs. One important difference is that “static” lattice models

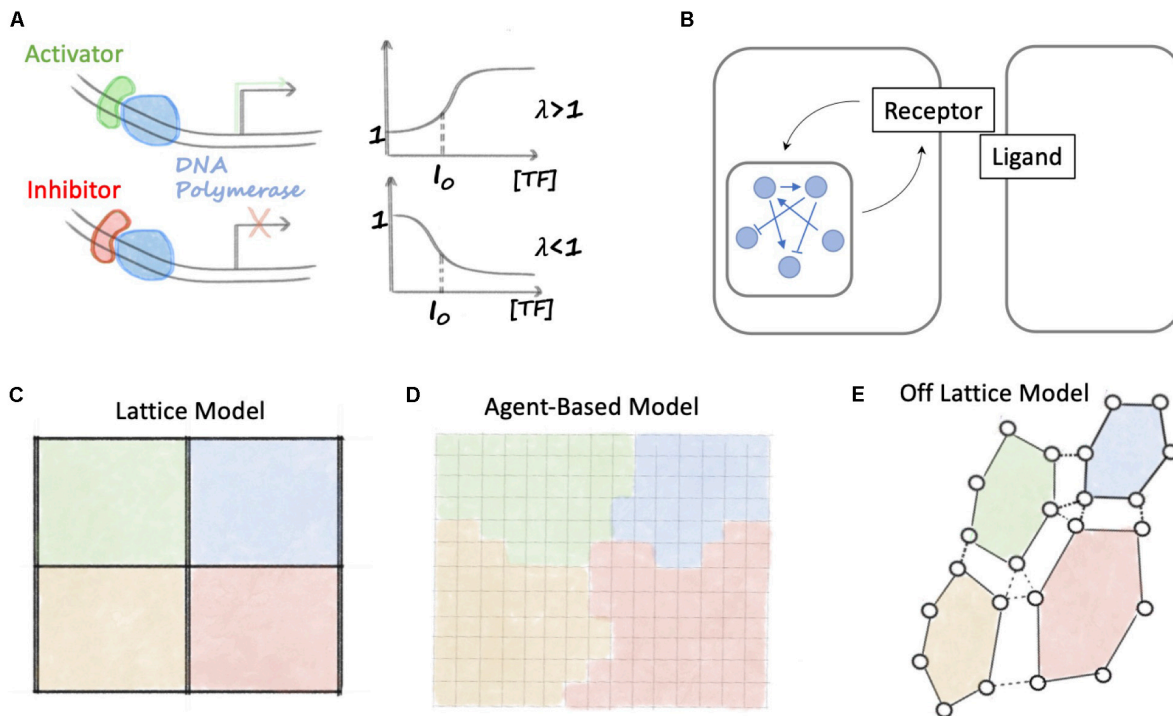


FIGURE 2 | Overview of mathematical methods to study Notch signaling. **(A)** A positive Hill function ($\lambda > 1$) describes transcription in presence of an activator that binds to DNA. When the concentration of transcription factor is high, the Hill function relaxes to a constant larger than 1. Conversely, for transcriptional inhibition, a negative Hill function ($\lambda < 1$) relaxes to a constant smaller than 1. **(B)** Models of Notch signaling dynamics integrate intracellular signaling (indicated by the blue network and interconnections) with ligand-receptor binding. **(C)** In a lattice model, cells are arranged in a fixed grid. Each position in the grid is identified as a cell, and ligands and receptors belonging to neighboring cells can bind. **(D)** In an agent-based model, space is divided into small fixed regions, and a cell is described by a set of contiguous space regions with the same cell identity (represented here as the color). **(E)** In an off-lattice model, a cell is described by the position coordinates of a set of membrane points. Membrane points of a cell are connected, for instance with elastic springs (continuous black lines). Cell-cell junctions are modeled as binds between pairs of membrane points of neighboring cells (dashed black lines).

assume fixed cell volumes; therefore, molecule concentration and copy number are equivalent descriptions. Conversely, Agent-based and off-lattice models allow changes in cell volume, thus requiring adjustment of molecular concentrations.

SPATIOTEMPORAL PATTERNING GUIDED BY NOTCH SIGNALING

In this section, we review experimental systems that exemplify two well-known patterning mechanisms enabled by Notch signaling: lateral inhibition and lateral induction. While lateral inhibition promotes opposite cell fates via biochemical negative feedbacks between the Notch receptor and Delta ligands, lateral induction promotes similar cell fates by positive feedback between Notch and Jagged ligands. Moreover, we review mathematical models that elucidate these patterning mechanisms on idealized, ordered lattices. Experiments and theoretical models help decoding the emergent outcomes of interactions between lateral inhibition and lateral induction mechanisms; specifically, we examine three biological processes that exhibit various degrees of patterning: angiogenesis, inner ear development and epithelial-mesenchymal transition in

cancer metastasis. Lastly, we discuss temporal oscillations of Notch observed during somitogenesis as an example of spatiotemporal patterning.

Biochemical Mechanisms of Lateral Inhibition and Lateral Induction

Historically, Notch signaling has been first characterized in *Drosophila melanogaster* as a mechanism that induces opposite cell fates among nearest neighbors (Heitzler and Simpson, 1991; Celis and de Garcia-Bellido, 1994; Celis and de Bray, 1997; Huppert et al., 1997; Simpson, 1997; Buceta et al., 2007). The establishment of divergent phenotypes among two neighboring cells, or lateral inhibition, relies on binding of the Notch receptor to ligands of the Delta-like family (Delta in *Drosophila*; Dll1, Dll3 and Dll4 in mammals – see **Table 1**) presented at the cell surface of a neighboring cell (Bray, 2006; Andersson et al., 2011). Upon engaging of Delta with the transmembrane Notch receptor, the intracellular domain of Notch (NICD) is cleaved by enzymes and translocates to the cell nucleus. Here, NICD activates Hey/Hes1, which in turn inhibits Delta (Shimojo et al., 2011; Bray, 2016; Sjöqvist and Andersson, 2019; **Figure 3A**). This negative feedback amplifies small initial differences in ligand and receptor concentrations among nearly equivalent neighbors

TABLE 1 | List of Notch molecular components, examples of biological processes regulated by Notch signaling, and references to examine in depth Notch signaling for various organisms.

Organism	Main components	Example of biological process	References
<i>Drosophila Melanogaster</i> (fruitfly)	Notch Delta Serrate	Wing disk formation, bristle patterning	Bray, 2016; Sjöqvist and Andersson, 2019
<i>Caenorhabditis elegans</i> (roundworm)	Lin-12, glp-1 Apx-1 Lag-2	Vulval precursor cell specification	Greenwald, 1998
<i>Danio rerio</i> (Zebrafish)	Notch 1, 2 Delta A, B, C, D Jagged 1, 2	Somitogenesis, artery and vein specification	Lawson et al., 2001; Venzin and Oates, 2019
<i>Gallus gallus domesticus</i> (chicken)	Notch 1, 2 Delta-like 1, 4 Jagged/Serrate 1, 2	Inner ear development	Neves et al., 2013
<i>Mus musculus</i> (house mouse)	Notch 1, 2, 3, 4 Delta 1, 3, 4 Jagged 1, 2	Inner ear development, vascular smooth muscle cell development	Bray, 2016; Sjöqvist and Andersson, 2019
<i>Homo sapiens</i>	Notch 1, 2, 3, 4 Delta 1, 3, 4 Jagged 1, 2	Inner ear development, vascular smooth muscle cell development	Bray, 2016; Sjöqvist and Andersson, 2019

to establish opposite cell states. The cell with higher levels of Delta can more effectively inhibit Delta in its neighbor, hence assuming a (low Notch, high Delta) or Sender phenotype, while forcing the neighbor to an opposite (high Notch, low Delta) or Receiver phenotype (Collier et al., 1996; Shaya and Sprinzak, 2011) (the green and orange cells in **Figure 3A**). This basic principle of differentiation regulates cell fate in several developmental and physiological processes. Interesting examples besides *Drosophila*'s development include angiogenesis (Benedito et al., 2009; Benedito and Hellström, 2013), spinal cord patterning in zebrafish (Appel and Eisen, 1998; Givan et al., 2001; Huang et al., 2012), development of neuroblast cells in early neurogenesis (Skeath and Carroll, 1992; Campos-Ortega, 1993; Homem and Knoblich, 2012), and vulval development in *C. elegans* (Fisher et al., 2007; Louisa et al., 2020). Thus, the Notch-Delta system can be regarded as a two-cell 'toggle switch' (Gardner et al., 2000) that enables opposite cell fates and possible switching among them under the influence of biological noise.

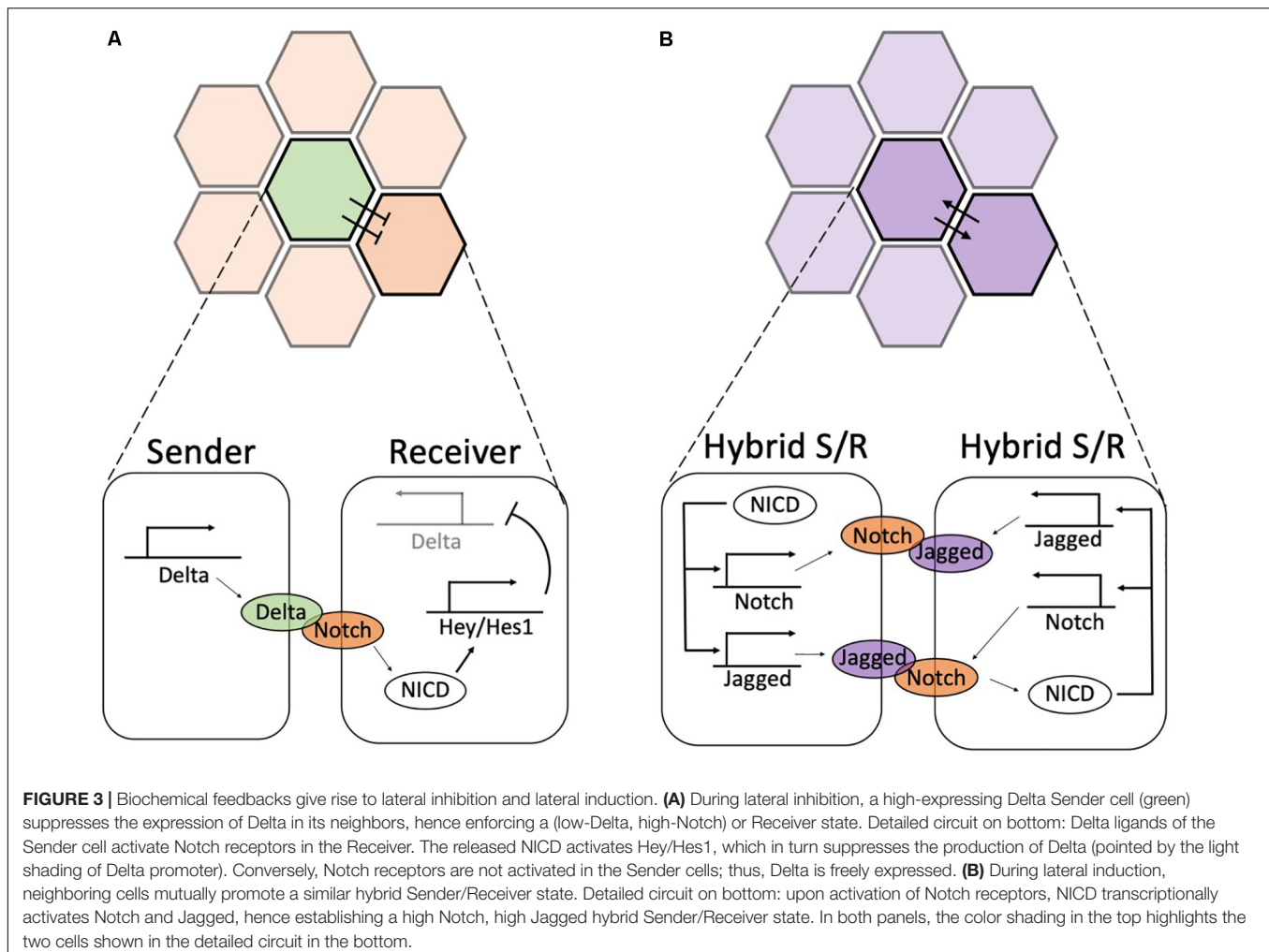
Despite being initially characterized as a driver of cell differentiation, Notch signaling can induce a convergent cell phenotype among neighbors through lateral induction (Bray, 2016; Sjöqvist and Andersson, 2019). A positive biochemical feedback between the Notch receptor and ligands of the Jagged/Serrate family establishes similar cell phenotypes that are spatially propagated to neighbors during the development of the inner ear (Lewis et al., 1998; Kiernan et al., 2001, 2006) and vascular smooth muscle cell (Manderfield et al., 2012). The Jagged family in mammals includes two paralogs (Jag1, Jag2), while *Drosophila* presents a single Serrate subtype (Bray, 2016; Sjöqvist and Andersson, 2019; see **Table 1**). Ligands of the Jagged/Serrate family are directly activated by NICD (Manderfield et al., 2012). Therefore, Notch-Jagged signaling between neighbors activates a positive feedback that establishes phenotypes with (high Notch, high Jagged) (the purple cells in **Figure 3B**), occasionally referred to as hybrid Sender/Receiver

phenotypes to highlight that both cells send and receive signals (Boareto et al., 2015a). Unless otherwise stated, green and orange colors denote high-Delta (Sender) and high-Notch (Receiver) phenotypes, respectively. Conversely, purple coloring indicates high-Jagged (hybrid Sender/Receiver) cells.

It is important to stress that positive and negative biochemical feedbacks that minimize or amplify initial differences are often assisted by a spatial and/or temporal regulation of Notch ligands and receptors (discussed in more detail by Bray, 2016). For instance, in the development of the *D. melanogaster* wing imaginal disc, the ligand Serrate is expressed only by cells on the dorsal side due to spatial confinement of the upstream transcription factor Apterous (Kim et al., 1995). This sharp boundary creates a stripe of Notch-active cells on the ventral side that leads to tissue growth thereafter (Kim et al., 1995).

Theoretical Exploration of the Notch-Delta-Jagged Circuit

Over the last two decades, theoretical models helped understanding the biochemical dynamics leading to lateral inhibition and lateral induction as well as the consequences of these signaling modes at the cell population level. In the first model of Notch-Delta lateral inhibition, Collier et al. (1996) hypothesized that activation of Delta stimulates Notch in the neighboring cells, while activation of Notch restricts Delta within the same cell (**Figure 4A**). In this model, the homogeneous state where neighbors express the same levels of Notch and Delta is stable for weak biochemical feedback, while cells differentiate into a Sender and a Receiver for strong feedbacks (Collier et al., 1996). When generalized to a spatial distribution of cells, cells tend to arrange in a 'salt-and-pepper' pattern where Senders are surrounded by Receivers and *vice versa* (Collier et al., 1996). Therefore, cell patterning in the model depends on the geometric arrangement of cells. While Senders and Receivers can perfectly



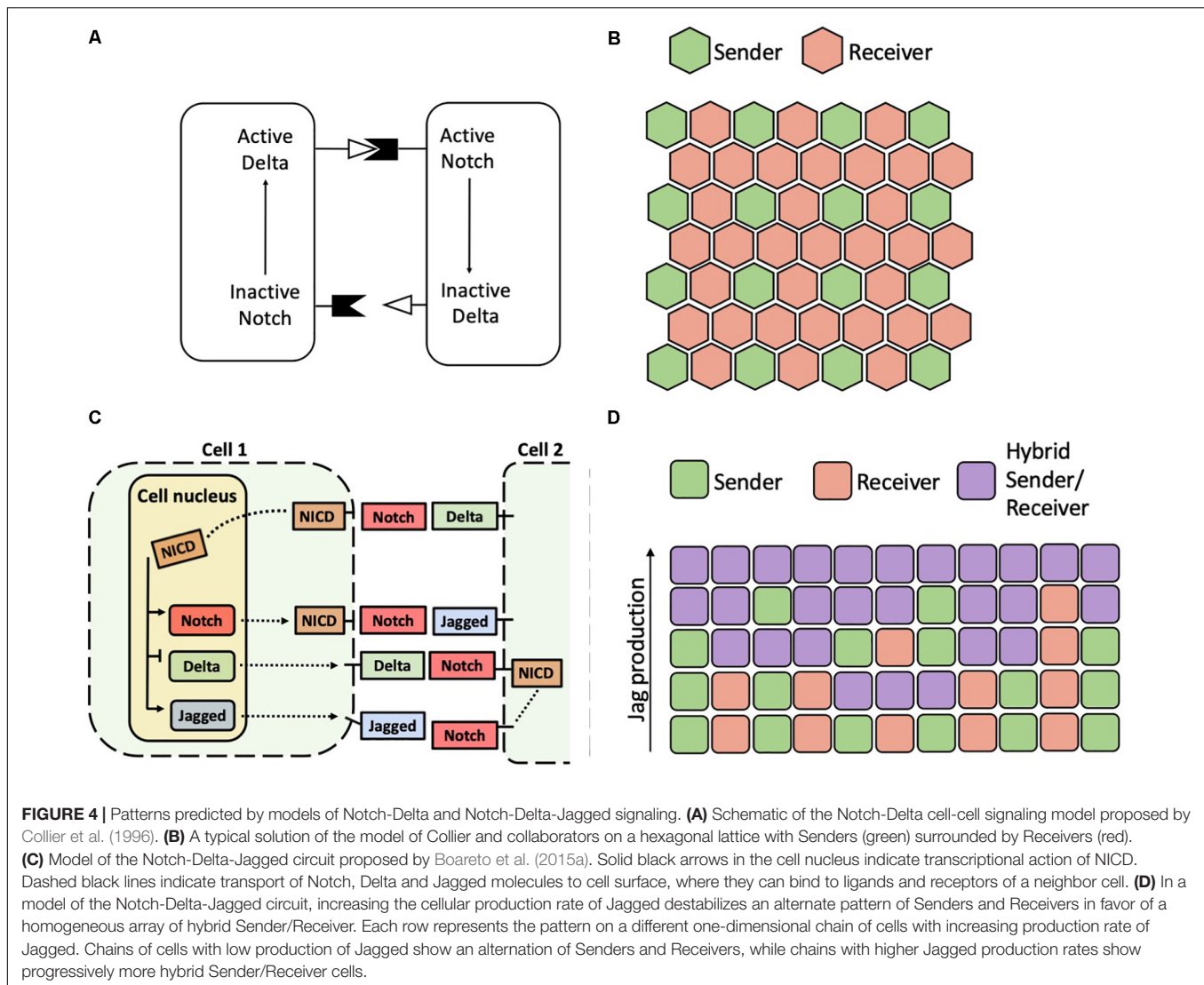
alternate on a square lattice, patterns on hexagonal lattices typically feature Senders surrounded by six Receivers, hence leading to a 3-to-1 Receiver/Sender ratio (**Figure 4B**). This patterning arises because contacts between Senders represent a more pronounced instability (Teomy et al., 2019). While contacts between Receiver cells results in the absence of signaling, two Sender cells dynamically compete until one of them eventually become a Receiver (Teomy et al., 2019). This arrangement is well reflected, for example, in the avian inner ear, where high-Delta hair cells are completely surrounded by low-Delta supporting cells (Goodyear and Richardson, 1997).

Further, some mathematical models have encapsulated the ability of Notch signaling to drive both divergent and convergent cell fates. A model developed by Boareto and colleagues considers the transcriptional activity of NICD that inhibits Delta and activates Jagged (**Figure 4C**). In this simplified representation, Delta and Jagged generically represent the two classes of ligands (Boareto et al., 2015a). In this model, the positive feedback between Notch and Jagged can drive the cells away from lateral inhibition, instead promoting a convergent hybrid Sender/Receiver state. Therefore, if the relative contribution of Notch-Delta signaling is large as compared to that of

Notch-Jagged, two neighboring cells fall into lateral inhibition. If Notch-Jagged signaling is dominant, however, the cells fall into the convergent 'hybrid Sender/Receiver' configuration with similarly high levels of Notch and Jagged (Boareto et al., 2015a). Therefore, modulating the balance between Notch-Delta and Notch-Jagged signaling in the model leads to transition between salt-and-pepper patterns and homogeneous patterns (**Figure 4D**). This trend is reminiscent of the dynamical behavior of an intracellular "toggle switch" coupled with self-activation, where the relative strengths of mutual inhibition and self-activation can drive different cell fates (Jolly et al., 2015).

Interplay of Lateral Inhibition and Lateral Induction Described by Experiments and Mathematical Models

Despite leading to opposite outcomes, lateral inhibition and lateral induction can take place at consecutive developmental steps, such as during inner ear development. Alternatively, they represent different outcomes that are selected based on signaling cues in the extracellular environment, such as during angiogenesis or tumor progression. In this section, we review



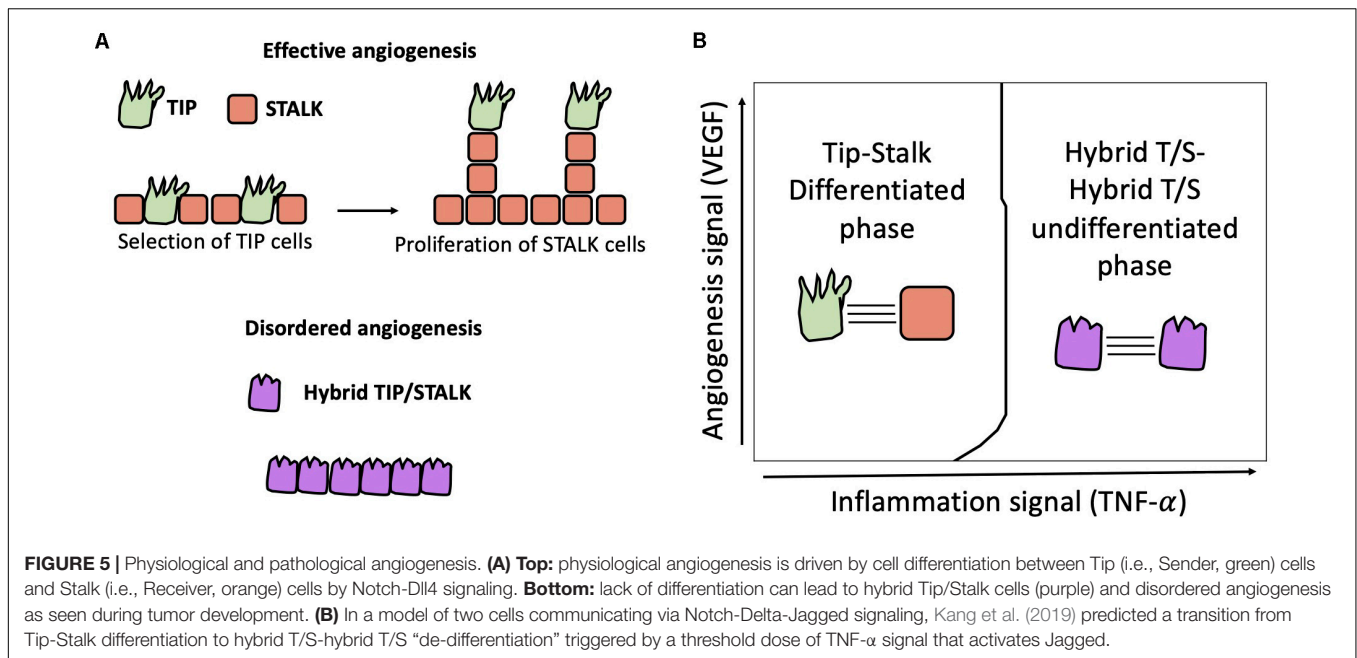
experiments and mathematical models that raise interesting questions about the interplay between lateral inhibition and lateral induction in three specific contexts: angiogenesis, inner ear development, and epithelial-mesenchymal transition during cancer metastasis.

Angiogenesis

Angiogenesis – the growth of new blood vessels from existing ones – is triggered by the hypoxia-induced signal VEGF (Vascular Endothelial Growth Factor). Secreted VEGF molecules bind to VEGF receptors (VEGFR) in the endothelial cells at the boundary of an existing blood vessel (Benedito and Hellström, 2013). Activation of VEGFRs in turn leads to transcriptional activation of the Delta subtype DLL4, hence inducing differentiation between a tip cell with high DLL4, and a stalk cells with low DLL4 by lateral inhibition (Holger et al., 2003; Benedito and Hellström, 2013). Subsequently, tip cells develop filopodia and migrate toward the VEGF gradient, while stalk cells proliferate to support the formation of the new vessel (**Figure 5A**, top).

Mathematical models suggest that tip-stalk differentiation can be understood as an example of lateral inhibition where external VEGF inputs activate Notch-DLL4 signaling (Bentley et al., 2008; Katie et al., 2014). Moreover, computational models suggest that tip-stalk selection is highly kinetic, and the typical timescale to commit to a specific cell fate varies considerably based on conditions in the extracellular environment as well as intracellular signaling dynamics (Venkatraman et al., 2016).

A binary model of Notch-Delta driven tip-stalk differentiation, however, cannot fully explain some experimental observations (Benedito and Hellström, 2013). For instance, DLL4 can occasionally act as a brake on sprouting angiogenesis by inhibiting endothelial tip formation (Suchting et al., 2007). Conversely, Jagged1 – that usually promotes lateral induction – assists vessel development in mouse models where Notch-DLL4 signaling is antagonized by the glycosylation of Notch by Fringe (Benedito et al., 2009). In addition, lateral inhibition typically leads to patterns with alternate cell fates, while



tip cells are typically separated by more than one stalk cell (Benedito et al., 2009).

Various models have been developed to explain deviations from classical Notch-Delta driven angiogenesis. Venkatraman et al. (2016) showed that regulators of Notch signaling such as lunatic fringe can slow down the Tip-Stalk differentiation process, hence giving rise to metastable partial Tip/Stalk states (Venkatraman et al., 2016). To explain sparse patterns where Tips are separated by multiple Stalks, Koon et al., integrated a standard model of Notch-Delta lateral inhibition with intracellular heterogeneity of Notch concentration and tension-dependent binding rate of the Notch-Delta complex (Koon et al., 2018). Interestingly, the addition of intracellular heterogeneity introduces states with intermediate levels of Notch and Delta, thus giving rise to pattern with multiple stalks separating consecutive Tips. Boareto et al. generalized their earlier computational model of the Notch-Delta-Jagged signaling to include VEGF-driven activation of Delta. This model predicts bistability between a Tip phenotype (i.e., Sender) and the Stalk phenotype (i.e., Receiver) when Jagged is weakly expressed (Boareto et al., 2015b). High expression of Jagged, however, stabilizes a homogeneous solution with hybrid Tip/Stalk (i.e., hybrid Sender/Receiver) cells (Figure 5A, bottom). In this model's interpretation, lateral induction between hybrid Tip/Stalk cells can prevent a binary categorization of migrating and proliferating cells, thus potentially disrupting vessel development (Boareto et al., 2015b).

To elucidate the interplay between Dll4 and Jag1 during angiogenesis experimentally, Kang and colleagues exposed human endothelial cells to both VEGF signal and the pro-inflammatory cytokine Tumor Necrosis Factor (TNF) that activates Jag1 *in vitro* (Kang et al., 2019). Strikingly, the combination of VEGF and low TNF dosage gives rise to

longer vessels. At a critical threshold of TNF dosage, however, opposite outcomes (i.e., either robust vessel formation or no vessel formation) were observed in experimental replicates. Finally, TNF dosages above the critical dosage consistently prevented vessel formation (Kang et al., 2019). Mathematical model focusing on the activation of Notch-Delta and Notch-Jagged signaling driven by VEGF and TNF, respectively, suggests a dose-dependent role for Jagged (Kang et al., 2019). While high levels of Jagged can lead to hybrid Tip/Stalk cells and disruption of angiogenesis, low Jagged activity acts synergistically with Delta to refine the alternate pattern of tips and stalks, hence contributing to more robust angiogenesis (Figure 5B). Therefore, increasing TNF dosage can lead to a switch in the role of Jagged from pro-angiogenesis to anti-angiogenesis (Kang et al., 2019).

The dynamics of Tip-Stalk differentiation receives several signaling inputs besides VEGF and TNF. Weinstein et al. (2017) developed a Boolean model of a large regulatory network governing endothelial cell behavior during angiogenesis. This model explores the crosstalk between Notch and several other signaling pathways in the cell as well as the cell microenvironment. It correctly recapitulates the molecular signatures of Tip and Stalk endothelial cells, and offers a platform to integrate signaling crosstalk in a large circuit with the simplification of a Boolean model (Weinstein et al., 2017).

In a pathological context, cancer cells can stimulate the sprouting of new blood vessels in the tumor microenvironment to supplement tumor growth (Kerbel, 2008; Weis and Cheresh, 2011). Typically, tumors exhibit irregular vascular networks that prevent efficient drug delivery (Koganehira et al., 2003; Jain, 2005), and even facilitate passive metastasis by engulfing cancer cells (Bockhorn et al., 2007; Fang et al., 2015). The ability of cancer to induce vasculature makes tumor angiogenesis a potential therapeutic target to halt tumor progression. Strikingly,

antitumor drugs that target Dll4, however, do not reduce tumor angiogenesis overall. Instead, anti-Dll4 drugs may result in a higher number of newly formed blood vessels with reduced functionality and chaotic architecture (Kerbel, 2008). Lateral induction of the hybrid tip/stalk phenotype has been proposed as a potential explanation to this paradoxical finding. As anti-Dll4 drugs tilt the balance toward Notch-Jagged signaling, the lack of tip-stalk differentiation amplifies promiscuous cell differentiation and leaky angiogenesis (Boareto et al., 2015b).

As we gain a better understanding of the complex spatiotemporal dynamics of normal and tumor angiogenesis, the advantages and disadvantages of combining drugs targeting angiogenesis with other standard-of-care therapies demand further investigation. Limited exposure to vasculature potentially protects the tumor from therapeutic agents that directly target cancer cells. Thus, perhaps counterintuitively, a transient renormalization of the tumor vasculature, timely synchronized with antitumor drugs, could serve as a potential strategy to alleviate tumor progression (Thurston et al., 2007).

Due to the strong coupling between signaling and cell mechanics observed during angiogenesis, several mathematical models have explored the connection between molecular mechanisms and cell- and organ-level behaviors. Further information on these models, which are not discussed here, are reviewed by Qutub et al. (2009).

Inner Ear Development

Lateral induction and lateral inhibition operate progressively at different stages of the inner ear development to turn an initially homogeneous population of non-sensory cells into a refined mosaic of cells with specific phenotypes. The inner ear is composed of hair cells that convert external stimuli into electrical signals, and supporting cells that provide tissue scaffolding, maintain a stable electrochemical environment, and occasionally differentiate to replenish the hair cell population after an injury (Kiernan et al., 2005; Neves et al., 2013). During the prosensory cell specification phase, Notch activates Jag1, which in turn sustains Notch in prosensory cells via lateral induction (Eddison et al., 2002; Daudet and Lewis, 2005; Daudet et al., 2007). Thus, the activation of Notch and Jag1 not only establishes the hair cell phenotype, but also propagates it through lateral induction up to several cell diameters (Hartman et al., 2010). Later, in the hair cell differentiation phase, Notch-Dll1 signaling establishes the final pattern where hair cells (i.e., the Senders) are surrounded by supporting cells (i.e., the Receivers) (Eddison et al., 2002; Daudet and Lewis, 2005; Daudet et al., 2007). For further insights on the role of Notch signaling in the inner ear development, a thorough review is offered by Neves et al. (2013). Interestingly, Petrovic et al. (2014) argued with experiments and mathematical modeling that Jag1 acts synergistically with Dll1 during the hair cell differentiation phase in enforcing a robust lateral inhibition by acting as a competitive inhibitor for Dll1. Similar to the model of Notch-driven angiogenesis proposed by Kang et al. (2019), a dose-dependent role for Jagged is suggested in inner ear development. While high levels of Jagged lead to a homogeneous state where cells attain a hybrid Sender/Receiver fate, a weak expression of Jagged can act synergistically with

Dll1 to refine the alternate pattern of Sender and Receivers. In the presence of a dominant Notch-Delta signaling, additional Jagged tends to compete with Delta over binding Notch receptors, resulting in a greater activation of NICD, and thus suppression of Delta, in Receiver cells (Petrovic et al., 2014). In this case, the ability of Jag1 to establish a convergent cell fate is negligible as compared to the cell differentiation promoted by Delta. When the signaling through the Notch-Jagged “branch” of the pathway becomes too strong, however, lateral induction dominates the patterning (Figure 6). Interestingly, the dose-dependent role of Jagged is only observed in mathematical models of extended two-dimensional lattices. For instance, Boareto et al. (2015a) showed that Notch-Delta signaling robustly give rise to salt-and-pepper patterns of Sender and Receivers on a one-dimensional chain (see Figure 4D again). In the two-dimensional lattice cells have a higher number of nearest neighbors – and thus potentially contradictory external inputs to process – hence increasing the probability of mistakes, or Sender-Sender contacts, in the pattern.

Epithelial-Mesenchymal Transition and Cancer Metastasis

Metastases represents the most critical step during tumor progression. Typically, cancer cells invade the circulatory system, reach anatomically distant sites and give rise to a secondary tumor (Gupta and Massagué, 2006). These cells can migrate individually as well as collectively as multi-cellular clusters with varying sizes depending on cancer type, stage and patient individualities (Cheung and Ewald, 2016; Jolly et al., 2017; Bocci et al., 2019b).

Generally, epithelial cancer cells partially or completely lose their cell-cell adhesion and acquire motility by undergoing the epithelial-mesenchymal transition (EMT) (Nieto et al., 2016). EMT can be activated by signaling cues in the tumor microenvironment in a cell autonomous manner as well as by Notch signaling. Activation of Notch signaling can be suppressed by EMT-inhibiting microRNAs such as miR-34 and miR-200 (Brabletz et al., 2011; de Antonellis et al., 2011; Bu et al., 2013; Bocci et al., 2019d). Notch signaling, however, can induce EMT by activating the EMT-inducing transcription factor SNAIL (Niessen et al., 2008; Sahlgren et al., 2008; Figure 7A).

An effort to elucidate the coupled dynamics of Notch signaling with the EMT gene regulatory network (Boareto et al., 2016) suggests that Delta-driven and Jagged-driven EMT can have different consequences at the level of multi-cellular patterning in a cancer tissue. While cells undergoing Notch-Delta-driven EMT are typically surrounded by epithelial cells, Notch-Jagged-driven EMT enables clustering among cells undergoing EMT (Figure 7B), hence potentially facilitating the formation of migrating multi-cellular cohorts in a tissue (Boareto et al., 2016).

Besides, Jag1 can also stabilize a hybrid epithelial/mesenchymal (E/M) cell phenotype (Boareto et al., 2016). Such hybrid E/M phenotype(s) can partially maintain cell-cell adhesion while gaining motility, and can invade as circulating tumor cell clusters (CTC clusters) that have elevated metastatic potential (Cheung and Ewald, 2016; Pastushenko and Blanpain, 2018; Sha et al., 2018; Jia D. et al., 2019). Experimental observations support this proposed role of Notch-Jagged

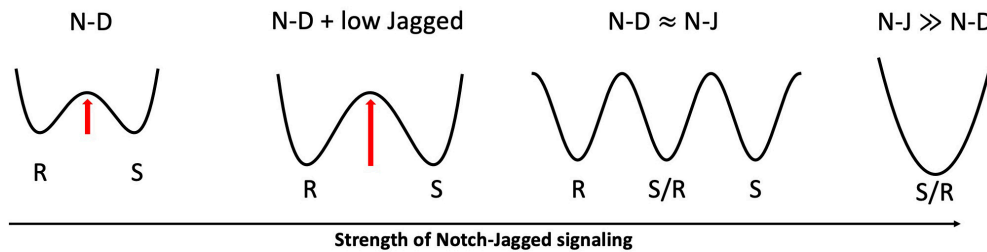


FIGURE 6 | Proposed role of Jagged dosage in Notch-driven cell fate. From left to right: in absence of Jagged (N-D), Sender and Receiver are the only accessible states in an abstract phenotypic landscape; a low Jagged dosage (N-D + low Jagged) increases the stability of Sender and Receiver states (indicated by the higher barrier in the landscape), as seen in inner ear development and angiogenesis; when both Notch-Delta and Notch-Jagged signaling are active (N-D $\approx$ N-J), a third hybrid Sender/Receiver state becomes accessible; an overwhelmingly strong Notch-Jagged signaling (N-J $\gg$ N-D) stabilizes the hybrid Sender/Receiver as the only accessible state.

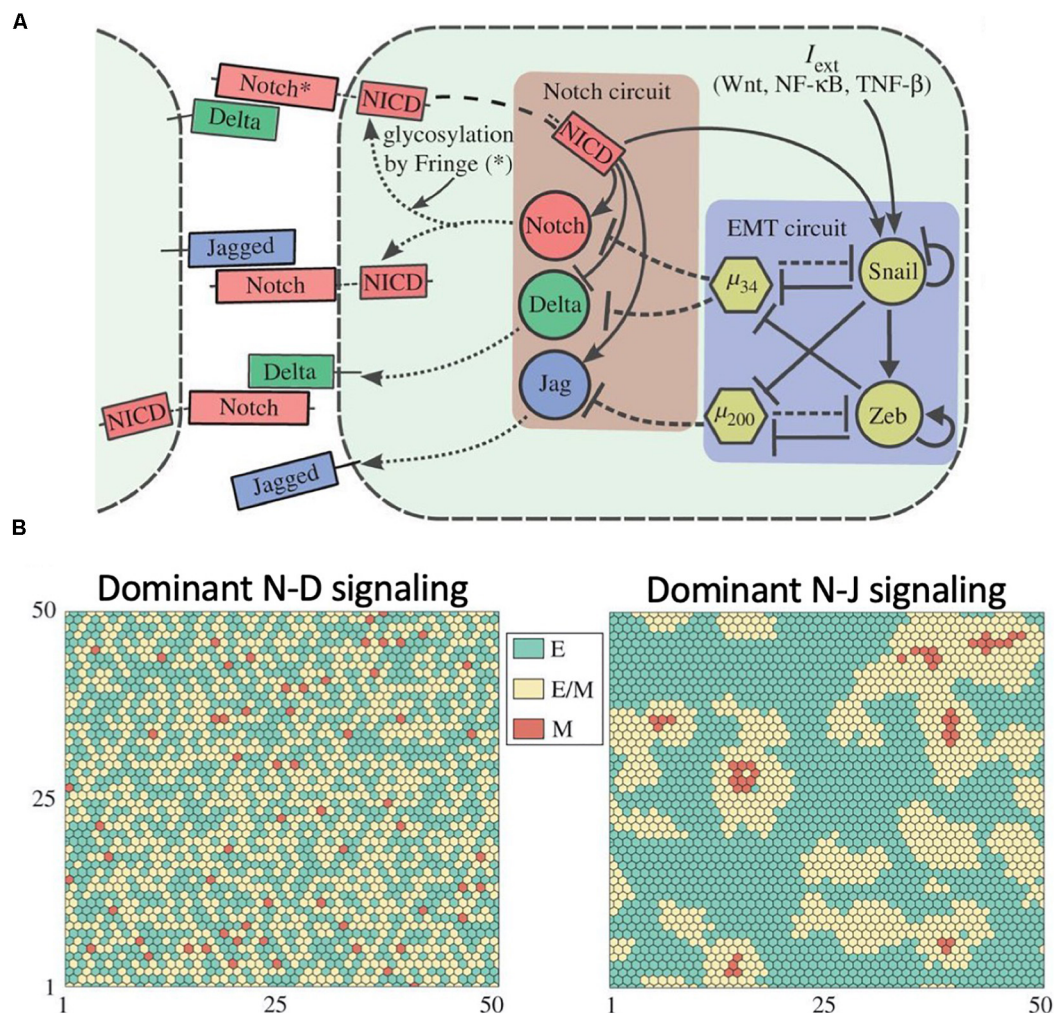


FIGURE 7 | Role of Notch signaling during Epithelial-Mesenchymal Transition. **(A)** Proposed coupling between the Notch-Delta-Jagged circuit and the core EMT regulatory network proposed by Boareto et al. (2016). **(B)** Mathematical modeling of the Notch-EMT circuit predicts patterns where hybrid epithelial/mesenchymal and mesenchymal cells are mostly surrounded by epithelial cells in presence of dominant Notch-Delta signaling (left) and patterns with clusters of hybrid E/M cells in presence of dominant Notch-Jagged signaling (right). In this figure, green, yellow and red represent epithelial, hybrid epithelial/mesenchymal and mesenchymal cells, respectively. The figure is adapted from Boareto and collaborators with permission from the published under a creative common license (Boareto et al., 2016).

signaling, although mostly through indirect evidence. First, CTC clusters from patients have a high expression of Jagged and co-express epithelial and mesenchymal markers, indicative of a hybrid epithelial/mesenchymal phenotype (Aceto et al., 2014; Jolly et al., 2017). Conversely, single CTCs mostly lack Jagged expression (Jolly et al., 2017). Second, Jag1 was identified as among top 5 differentially expressed genes in cells positive for K14, a marker for cluster-based migration (Cheung et al., 2013). Generalizations of this framework identified additional biochemical pathways that act as “phenotypic stability factors” (PSFs) and stabilize hybrid E/M phenotype by coupling to the core Notch-EMT circuit. Examples include NUMB, NF- κ B and IL-6 (Bocci et al., 2017, 2019a). Consistently, overexpression of PSFs such as NUMB correlates with a worse patient survival in various cancer types (Jia et al., 2015; Bocci et al., 2017, 2019a,e). Recently, the epigenetic landscape and transition dynamics during EMT have been unraveled with a stochastic dynamical modeling approach (Li et al., 2016; Li and Balazsi, 2018; Jia W. et al., 2019). These models suggest the presence of multiple intermediate hybrid E/M states and indicate plausible transition routes between EMT phenotypes in the noisy cellular epigenetic landscape (Li et al., 2016; Li and Balazsi, 2018). Certainly, understanding how Notch signaling affects the stability and transitions between EMT phenotypes from a landscape perspective is an exciting future direction for theoretical modeling.

To metastasize, migrating cancer cells need the proliferation potential and resistance to therapies typical of cancer stem cells (CSCs). Typically, cells undergoing a partial or complete EMT also show traits of CSCs (Mani et al., 2008; Grosse-Wilde et al., 2015; Pastushenko et al., 2018; Bocci et al., 2019c). Mathematical modeling of the gene regulatory networks underlying EMT, Notch and stemness suggests that Notch-Jagged signaling can promote a “window of opportunity” where cancer cells exist in a hybrid E/M, stem-like phenotype with aggravated metastatic potential (Bocci et al., 2018a; Nie, 2018). Consistent with this prediction, CSCs display enhanced levels of Notch and Jagged across several cancer types including glioblastoma, pancreatic cancer, colon cancer and breast cancer (Wang et al., 2009; Sikandar et al., 2010; Zhu et al., 2011; Yamamoto et al., 2013). Moreover, the glycosyltransferase Fringe which promotes Notch-Delta interactions over Notch-Jagged is reported as a tumor suppressor in multiple cancers (Xu et al., 2012; Yi et al., 2013; Zhang et al., 2014). Furthermore, it was recently shown *in vitro* that knockdown of Jag1 inhibits the formation of tumor emboli in hybrid E/M inflammatory breast cancer (IBC) – a rare but highly aggressive form of breast cancer that moves largely collectively through clusters (Jolly et al., 2017) – cells SUM149 (Bocci et al., 2019a).

Notch signaling can also regulate spatiotemporal pattern formation at the level of a tumor tissue. Analysis of breast cancer tissues highlighted subsets of mesenchymal CSCs at the tumor invasive edge, while subsets of hybrid E/M CSCs were largely localized in tumor interior (Liu et al., 2014). A recent computational model developed by Bocci et al. suggests that Notch-Jagged signaling may contribute to generating this spatial heterogeneity. In the presence of a diffusive EMT-inducing signal

such as TGF- β , Notch-Jagged signaling, but not Notch-Delta signaling, can give rise to large populations of CSCs. CSCs subsets at the tumor invasive edge are highly exposed to EMT-inducing signals and have a higher likelihood of undergoing EMT, whereas CSCs in the tumor interior are less exposed to EMT-inducing signals and hence retain a hybrid E/M phenotype (Bocci et al., 2019a). Given the varying metabolic profiles of these CSC subsets (Luo et al., 2018), such patterning is reminiscent of spatial self-organization of metabolically diverse phenotypes in other contexts such as bacterial colonies (Bocci et al., 2018b; Varahan et al., 2019).

Finally, the transition to a mesenchymal phenotype is not exclusive to epithelial cells. Besides undergoing tip-stalk differentiation in sprouting angiogenesis, endothelial cells can alternatively undergo Endothelial-to-Mesenchymal Transition (EndMT) (Lamouille et al., 2014). While tip-stalk differentiation maintains cell-cell adhesion, EndMT leads to the detachment of endothelial cells. The underlying circuitry associated with these different transition routes involves Notch, the EMT network, and other pathways such as HIF1- α and TGF. Recently, this large circuit has been modeled as a Boolean network, offering suggestions about the specific signaling features that distinguish the two transitions (Weinstein et al., 2020).

Oscillations and Synchronization as Seen in the Somite Segmentation Clock

So far, we discussed mechanisms of spatial patterning. Due to its crosstalk with other signaling pathways, however, Notch can exhibit non-trivial temporal patterns. As an example, here we discuss somite segmentation, a well-known example of Notch oscillatory dynamics. During somite segmentation, the embryo's body axis is segmented into somites – blocks of epithelial cells that later give rise to vertebrae and tissues in the adult body (Andrew et al., 2012). Segmentation is organized by a precise spatiotemporal clock. Traveling waves of gene expression move along the body axis and stop at the location of a following segmentation event (Andrew et al., 2012).

Oscillations in gene expression are generated in a cell autonomous manner via an autoregulatory negative feedback by Hes/Her proteins. Upon protein productions, Hes/Her molecules dimerize and suppress their own transcription (Lewis, 2003; Monk, 2003). The delay between transcription and protein synthesis gives rise to oscillations in Hes/Her gene expression (**Figure 8**) with a period of about 2–3 h (Hirata et al., 2002; Shimojo et al., 2008). This model, however, is not sufficient to explain how oscillations maintain a precise cell to cell synchronization in time and space.

Several experimental observations suggest a role for Notch-Delta signaling in synchronizing oscillations in neighboring cells, due to the biochemical coupling between the Notch and Hes/Her pathways. As previously discussed, NICD transcriptionally activates the family of Hes/Her molecules, which in turn, represses Delta (Shimojo et al., 2011; Bray, 2016; Sjöqvist and Andersson, 2019). Therefore, self-sustained oscillations of Hes/Her can potentially propagate to Notch (**Figure 8**). Zebrafish models indicate a periodic expression of Delta ligands during

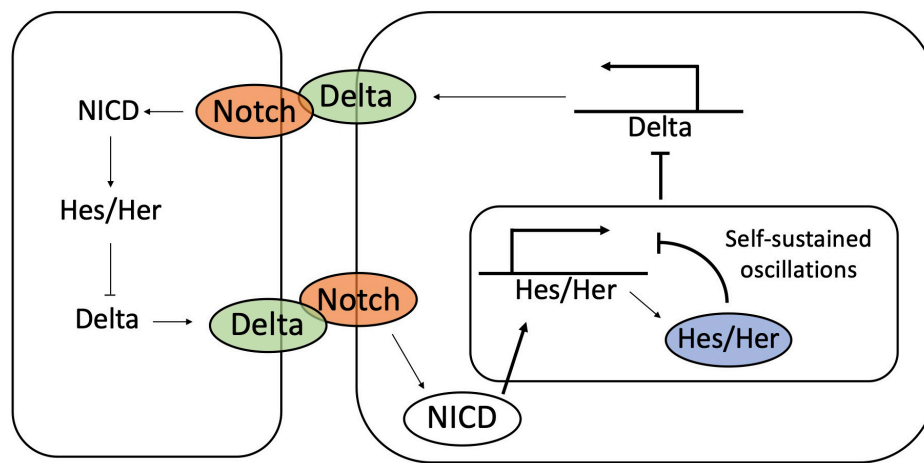


FIGURE 8 | The coupling between Notch-Delta and Hes/Her signaling synchronize temporal oscillations during somitogenesis. Hes/Her can autonomously give rise to sustained oscillations by self-inhibition of Hes/Her protein. The coupling between Hes/Her and Notch-Delta signaling synchronizes oscillations between neighbors.

somite segmentation (Jiang et al., 2000), while mouse models show oscillations of Notch, Delta and NICD (Huppert et al., 2005; Bone et al., 2014; Shimojo et al., 2016).

Notch-Delta binding potentially provides information about the phase of the Hes/Her clock in neighbors. Mathematical modeling of the Notch-Hes/Her circuit developed by Lewis and colleagues (Lewis, 2003; Ozbudak and Lewis, 2008) suggests that (i) oscillation can be self-sustained by the autoregulatory Hes/Her feedback loop, but (ii) Notch-Delta progressively couples and eventually synchronizes the clocks of neighboring cells (Lewis, 2003; Ozbudak and Lewis, 2008). In other words, each cell can be viewed as an independent biochemical oscillator, and the exchange of ligands through the Notch receptor synchronizes the oscillations of the different cells (Shimojo and Kageyama, 2016; **Figure 8**). This model is supported by observation in Zebrafish mutants that do not express Notch and Delta. In these mutants, segmentation is defective, and cells are arranged in heterogeneous patterns of high Hes/Her and low Hes/Her indicative of asynchrony in the cell population (Riedel-Kruse et al., 2007; Delaune et al., 2012).

It remains unclear whether Notch's unique role is to ensure robust temporal correlation among neighbors. While it is generally accepted that Hes/Her self-inhibition is sufficient to generate temporal patterns, a number of studies in mouse models suggest that Notch might be required for oscillations. For further details, a comprehensive review on the role of Notch signaling in the somite segmentation clock is offered by Venzin and Oates (2019).

NON-CANONICAL MODULATION OF NOTCH SIGNALING

In the previous section, we discussed mechanisms of lateral inhibition and lateral induction guided by biochemical feedbacks between Notch and its ligands. In this section, we review

mechanisms that modulate Notch signaling besides canonical positive and negative transcriptional feedbacks. These include dependence on cell-cell contact area and cell packing geometry, binding between receptors and ligands within the same cell, specificity in the affinity between receptor and ligand paralogs, and mechanisms enabling signaling beyond nearest neighbor. From a phenomenological standpoint, these mechanisms can be viewed as additional features beyond the simple nearest neighbor signaling mechanism.

Variability of Cell Packing and Contact Area

In the previous section, we developed a geometrical intuition on lateral inhibition that is based on alternate arrangement of Sender and Receiver cells. Mathematical modeling of Notch-Delta signaling helps understand these patterning dynamics on idealized ordered lattices. For instance, Notch-Delta signaling leads to a very specific pattern where Senders are surrounded by six Receivers on a perfect hexagonal lattice (see **Figure 4B**). Disordered lattices with variable cell sizes and number of nearest neighbors can lead to deviations from the standard “salt-and-pepper” pattern.

The development of the basilar papilla, the avian equivalent of the mammals' organ of Corti, exemplifies how fluctuations in cell arrangement modulate lateral inhibition. The fully developed basilar papilla consists of a hexagonal mosaic where Sender cells (i.e., hair cells) are surrounded by six Receiver cells (i.e., supporting cells) (Goodyear and Richardson, 1997). Goodyear and Richardson found experimental evidence of dynamic cell rearrangement in the early development of the basilar papilla in a seminal study (Goodyear and Richardson, 1997). At earlier developmental stages (6–7 days), cell packing in the papilla is irregular and features cells with variable size and shape. Consequently, the number of nearest neighbors fluctuate between 3 and 8 cells (Goodyear and Richardson, 1997). This underdeveloped mosaic allows occasional contacts between hair

cells. Later on, cell packing relaxes toward a precise hexagonal mosaic and the “mistakes” in the patterning are corrected (Goodyear and Richardson, 1997).

The size of shared contact area between neighbors is expected to fine-tune Notch signaling. Shaya and collaborators investigated the relation between cell size and cell fate by integrating experimental and computational methods (Shaya et al., 2017). By incorporating live-cell imaging reports to track the activity of Notch and Delta, they showed that signaling between pairs of nearest neighbors correlates with their cell-cell contact area. Smaller cells produced Delta at a higher rate and eventually became hair cells, while larger cells generally committed to a non-hair, supporting phenotype (Shaya et al., 2017). This result was reproduced by a mathematical model that generalized the seminal Notch-Delta model of Collier et al. (1996) to a disordered lattice with variable cell size (Shaya et al., 2017; **Figure 9**). In the simplest model of lateral inhibition, Senders are selected from a homogeneous population by spontaneous breaking of symmetry and amplification of initial differences in protein levels (Collier et al., 1996). Instead, this experiment shows that the fluctuations of cell size contribute to cell fate selection by introducing a weightage factor in the extent of Notch signaling between neighbors (Shaya et al., 2017).

Cis-Interactions

Although Notch has evolved as a cell-cell signaling mechanism, receptors and ligands can bind within the same cell. Ligand-receptor binding within the same cell, or cis-interaction, does not lead to downstream signaling, but rather to ligand-receptor complex degradation (cis-inhibition) (Celis and de Bray, 1997; Micchelli et al., 1997; Del Alamo et al., 2011). Despite not contributing to signaling, cis-inhibition can compete with the canonical Notch pathway by sequestering Notch receptors and ligands (**Figure 10A**).

Sprinzak et al. (2010) used time-lapse microscopy to evaluate Notch activation in response to external Delta ligands (standard trans-activation) and endogenous Delta (cis-interaction). While Notch receptors trans-activate gradually in response to external Delta, the response to indigenous, cis-Delta is sharp (**Figure 10B**). Therefore, cis-inhibition silences Notch signaling when the intracellular Delta exceeds a threshold concentration (Sprinzak et al., 2010). This mechanism improves the robustness of lateral inhibition by further inactivating Notch in Sender cells. The authors further employed mathematical modeling to evaluate the behavior of an ensemble of kinetic models of Notch-Delta signaling with randomized parameters. Compared to a control model lacking cis-inhibition, models with cis-interactions yield lateral inhibition over a much broader parameter range by further refining defects in the patterning of Sender and Receiver cells (Sprinzak et al., 2011). The role of cis-inhibition, however, is not just restricted to proof-reading, but can rather be pivotal for cell-fate decision. For instance, loss of cis-inhibition compromises cell fate specification during the development of photoreceptors in *Drosophila* (Miller et al., 2009).

Although cis-interactions are mostly known to degrade Notch signaling without any contribution to signaling, experiments recently reported cell autonomous activation

of Notch, such as in the cases of *Drosophila* bristle precursor cells and cell cycle regulation in T cells (Coumilleau et al., 2009; Guy et al., 2013). These experiments raise interesting questions about the competition between intracellular and intercellular signaling in modulating cell fate decisions. Nandagopal and colleagues engineered a synthetic system where cells constitutively express Notch while production of Delta is controlled experimentally (Nandagopal et al., 2019). Interestingly, extremal expression of Delta silenced Notch activity, whereas intermediate Delta expression maximized cis-activation (Nandagopal et al., 2019). To rationalize these observations, the authors developed various classes of mathematical models where cis-interactions can lead to either cis-activation or cis-inhibition with different rates. Interestingly, the non-monotonic response of Notch as a function of Delta concentration could only be reproduced by models with higher-order interactions and formation of clusters with multiple ligands and receptors (Nandagopal et al., 2019). Indeed, oligomerization of Notch receptors and ligands has been reported in the Notch pathway (Bardot et al., 2005; Nichols et al., 2007; Nandagopal et al., 2018).

Given the role of cis-inhibition in enforcing robust lateral inhibition, it can be postulated that a switch from cis-inhibition to cis-activation would compromise precise cell patterns of Sender and Receiver cells. Formosa-Jordan and Ibanes (2014) investigated the implication of Notch-Delta cis-activation in a disordered multicellular lattice model with variable cell size and shape. Compared to the mathematical model by Shaya et al. (2017) discussed in the previous section, the authors did not focused explicitly on the correlation between cell size and cell fate, but rather on how cis-activation biases patterns of Senders and Receivers. Their mathematical model confirms that cis-activation prevents robust lateral inhibition and instead introduces disordered patterns (Formosa-Jordan and Ibanes, 2014). Specifically, in presence of strong cis-activation, cell dynamics is predominantly cell-autonomous, rather than driven by nearest neighbors. Hence, cis-activation progressively increase the fraction of high-Delta Sender cells in the lattice model (**Figure 10C**). Indeed, cis-activation introduces a negative intracellular feedback where Delta ligands in the Sender cell promote their own inhibition by activating Notch receptors, hence driving the system away from the target Sender state with (low Notch, high Delta).

Specificity in Ligand-Receptor Binding Affinity

The number of Notch receptor and ligand subtypes varies considerably in different species (see **Table 1**). Typically, mammals have four different paralogs of the Notch receptor (Notch1–4), three Delta-like ligands (Dll1, Dll3, Dll4), and two Jagged ligands (Jag1, Jag2). Although the effect on the receiving cell is identical (i.e., NICD release), interactions through different ligand-receptor pairs can lead to differences in the downstream signaling cascade (Bray, 2016; Sjöqvist and Andersson, 2019).

First, binding affinities depend on the molecular structure. For instance, Notch1 has a greater affinity to Dll4 than to Dll1

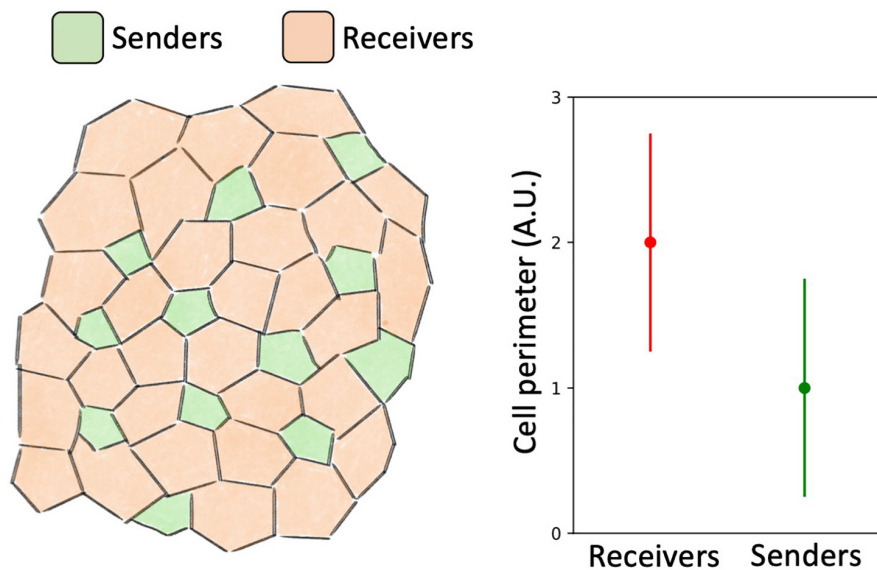


FIGURE 9 | Mathematical modeling predicts a correlation between cell size and fate. Mathematical model of Notch-Delta signaling on a disordered lattice developed by Shaya et al. (2017) suggests that larger cells assume a Receiver phenotype and smaller cells assume a Sender phenotype. **Left:** a typical spatial patterning of Senders (green) and Receivers (red) predicted by mathematical modeling. **Right:** cells with large perimeter tend to become Receivers while cells with smaller perimeter tend to become Senders.

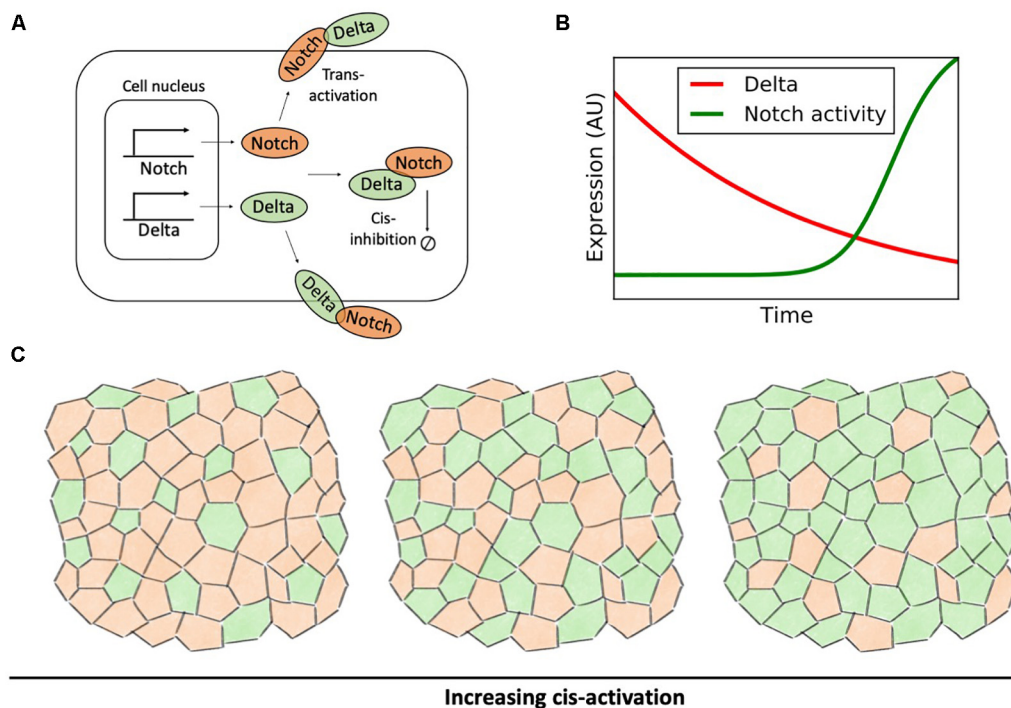


FIGURE 10 | Cis-activation destabilizes the ordered lateral inhibition pattern. **(A)** Binding of Notch and Delta molecules within the same cell leads to the degradation of the receptor-ligand complex without downstream signaling. **(B)** In a time-lapse microscopy experiment by Sprinzak et al. (2010), the concentration of Delta (red) gradually decays exponentially due to dilution and cell division. Conversely, the activity of Notch (green) is turned on sharply when the concentration of Delta decreases below a threshold. This panel is adapted from Sprinzak et al. (2010). **(C)** In a model of Notch-Delta signaling on a disordered lattice developed by Formosa-Jordan and Ibanes (2014), increasing the rate of cis-activation progressively disrupts lateral inhibition patterns. Left: in absence of cis-activation, Notch-Delta signaling gives rise to a pattern where Senders (green) are surrounded by Receivers (orange). For increasing levels of cis-activation, cell fate becomes cell autonomous and the fraction of Senders progressively increases (rightmost plots).

and Jag1 (Luca et al., 2017a). Moreover, different ligand-receptor pairings can lead to different dynamical responses in the receiving cell. For instance, Nandagopal and colleagues proposed that Notch1 can dynamically discriminate the ligands Dll1 and Dll4 in mouse and hamster cells (Nandagopal et al., 2018). Namely, while Dll4 activates Notch1 in a sustained manner, Dll1 gives rise to pulses of Notch1 activity (Nandagopal et al., 2018). Differences arise also in the ligand ability to cis-inhibit Notch receptors. For instance, Dll4 but not Dll1, can efficiently cis-inhibit Notch1 in mice cells (Preuß et al., 2015), reminiscent of the greater Notch1-Dll4 affinity observed in trans-activation (Luca et al., 2017a). Moreover, the ligand Dll3 typically does not trans-activate any of the four Notch subtypes but only contributes to cis-inhibition (Ladi et al., 2005; Chapman et al., 2011).

Mechanisms that modify the binding affinity between the various subtypes of receptor and ligand can potentially result in a shift in cell fate by introducing an asymmetry between Delta and Jagged ligands. One such well-characterized mechanism is the glycosylation by Fringe proteins that results in a conformational change in the extra cellular domain of the Notch receptor (Jane and Wu, 1999; Nadia and Rana, 2011). Glycosylation typically decreases the binding affinity of Notch with Jagged ligands both in trans- and cis-interactions (Hicks et al., 2000; Ladi et al., 2005; Hou et al., 2012; LeBon et al., 2014). Mathematical modeling of the Notch-Delta-Jagged signaling suggests that Fringe can stabilize the Sender and Receiver cell states by restricting the binding between Notch and Jagged, while loss of Fringe may tilt the balance toward Notch-Jagged signaling and lateral induction (Jolly et al., 2015; **Figure 11A**).

Interactions Beyond Nearest Neighbor Through Filopodia

Although the Notch pathway is primarily designed as a pairwise signaling mechanism among nearest neighbors, beyond nearest neighbors' interactions are occasionally enabled by filopodia.

Filopodia can extend up to several cell diameters and thus introduce contacts beyond nearest neighbor (Joussineau et al., 2003; Eom et al., 2015; Huang and Kornberg, 2015; **Figure 11B**). For instance, in the bristle patterning of *Drosophila*, the sensory organ precursor cells (SOPs) with high Delta (i.e., the Sender cells) are separated by 4–5 receiver cells. This spacing, much larger than typically observed in lateral inhibition systems, is explained by dynamically rearranging filopodia that can give rise to transient contacts among non-neighbor cells (Cohen et al., 2010). This signaling between cells that are not adjacent to one another has been interpreted as a source of noise that refines the patterning (Cohen et al., 2010).

Filopodia-driven signaling raises questions on how Notch can be effective when cells communicate through a small contact area. Khait et al. (2016) reported that the diffusion coefficient of Dll1 can vary over an order of magnitude ($0.003\text{--}0.03\mu\text{m}^2/\text{s}$) from cell to cell in hamster ovary cells (Khait et al., 2016). Based on this experimental finding, the authors developed a kinetic theoretical model including ligand-receptor binding at cell surface and lateral diffusion of Notch and Delta molecules across the cell surface. This framework highlights opposite regimes of signaling.

When the radius of the shared contact area between cells (b) is larger than the typical diffusion length scale (λ), diffusion effects are negligible and the signaling depends on only the contact area. In the opposite regime ($\lambda > b$), however, the signaling strongly depends on the influx of Delta ligands in the contact area but only weakly on the size of the contact area (Khait et al., 2016; **Figure 11C**). Diffusion coefficients in filopodia are larger by up to a 10-fold than in bulk membrane, possibly explaining how thin filopodia can still play an important role in Notch signaling (Khait et al., 2016).

INDICATIONS OF A ROLE FOR MECHANOSENSITIVITY IN NOTCH SIGNALING

Activating Notch signaling requires mechanical pulling on the ligand-receptor complex leading to NICD cleavage. Therefore, the signaling operates optimally within a certain range of mechanical constraints (Meloty-Kapella et al., 2012; Wang and Ha, 2013; Chowdhury et al., 2016). In contexts such as collective epithelial migration and cardiovascular morphogenesis, cells continuously adapt their shape, tensions and stresses. It can be speculated that these biophysical factors add a further layer of regulation on Notch-driven patterning. While the role of mechanosensitivity is more quantitatively understood at the molecular scale of ligand-receptor interaction, its consequences on multicellular patterning are still largely unexplored. The following two sections offer recent evidence suggesting a role for mechanosensitivity in leader-follower differentiation during collective epithelial cell migration and cardiovascular morphogenesis.

Lateral Inhibition and Mechanics Select Leader and Follower Cells During Collective Epithelial Cell Migration

Collective cell migration is commonly observed in physiological and pathological processes, including morphogenesis, wound healing and cancer metastasis. Collectively migrating cells conserve their cell-cell adhesion through several mechanisms, such as adherens junctions (Friedl and Mayor, 2017; Barriga et al., 2018). Typically, some cells at the front of the migrating cell layer assume a distinct morphology characterized by an enlarged size and ruffling lamellopodia, and are labeled as “leaders” at the migration (Yang et al., 2016). In a typical scratch assay that mimics wound healing, the mechanical injury at the boundary can generate a gradient of activation of several signaling pathways, with the strongest response in cells adjacent to the boundary and gradually decreasing in the inner region (Riahi et al., 2014; **Figure 12A**). Reminiscent of branching angiogenesis, the differentiation between leader and follower cells is regulated by the Notch-Delta pathway. Specifically, approximately 25% of the cells at the leading edge are leaders with high expression of Dll1. Conversely, cells with low Dll1 and high Notch1 become followers (Riahi et al., 2015). Interestingly, approximately 10% of cells transiently increase

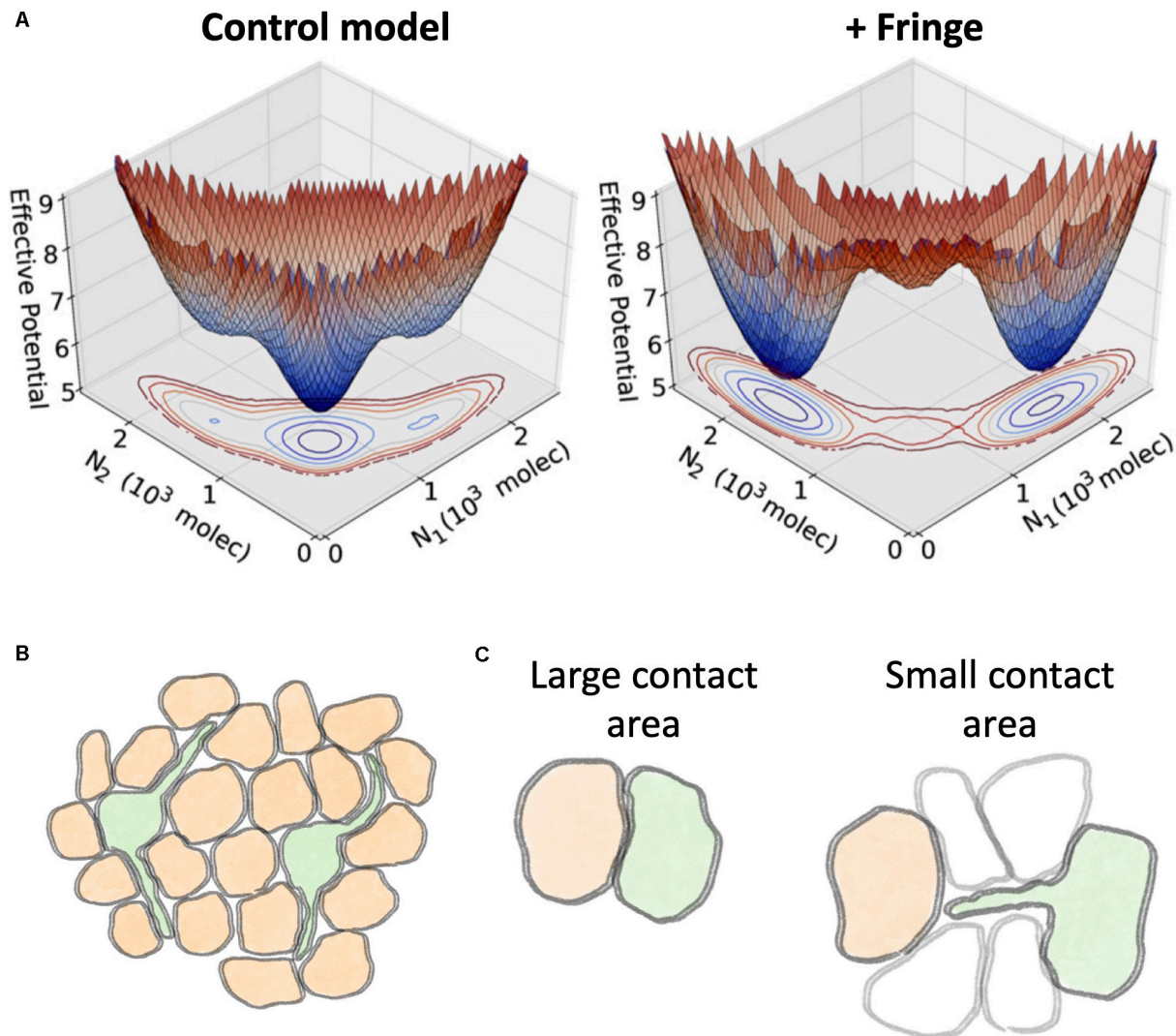


FIGURE 11 | Effect of Fringe glycosylation and filopodia on Notch signaling. **(A)** A mathematical model of Notch-Delta-Jagged signaling by Jolly et al. (2015) predicts a switch in cell fate due to Fringe glycosylation. The effective potential of a two-cell model depicts the probability for the two cells to assume specific levels of Notch (N_1 and N_2 , respectively). A control model without the effect of Fringe glycosylation (left) exhibits a single dominant minimum where both cells are hybrid Sender/Receiver with same Notch levels. Conversely, a model with Fringe modifies the landscape and introduces two separate states corresponding to Receiver-Sender ($N_1 \ll N_2$) and Sender-Receiver ($N_1 \gg N_2$). **(B)** Through filopodia, Sender cells (green) can potentially inhibit the Sender state in cells beyond nearest neighbors. **(C)** Schematic representation of the regimes of Notch-Delta signaling predicted by mathematical modeling by Khait et al. (2016). **Left:** when cells share a large contact area, diffusion of Delta ligands is negligible. **Right:** when the contact area is small, such as in the case of contact through filopodia, the signaling depends crucially on the diffusion of Delta ligands. Panel **(A)** is adapted from Jolly et al., with permission from the published under a creative common license (Jolly et al., 2015).

Dll1 after wounding but ultimately become followers, showing that the leader-follower differentiation is regulated in a highly dynamical manner by the Notch1-Dll1 pathway (Riahi et al., 2015), similar to the dynamical balance of tip-stalk decision-making in angiogenesis (Jakobsson et al., 2010).

Notably, leader-follower selection depends on feedback loops among Notch signaling and mechanical stresses. Indeed, receptor-ligand binding and the conformational change in the Notch1 domain thereafter require maintaining the receptor-ligand bond for enough time, which might be jeopardized by

forces applied to the receptor or ligand (Luca et al., 2017a,b), as can happen in the presence of mechanical injury during wound healing. Mechanical stresses inhibit the expression of Dll1 and prevent the selection of leader cells. Comparing the spatial distribution of mechanical forces and Dll1 expression suggests that the reduction of cellular stress at the boundary allows an effective Notch1-Dll1 signaling and leader-follower selection via lateral inhibition and gives rise to the observed gradient of Notch activation (Riahi et al., 2015). In the classic lateral inhibition scenario, Senders and Receivers are selected by

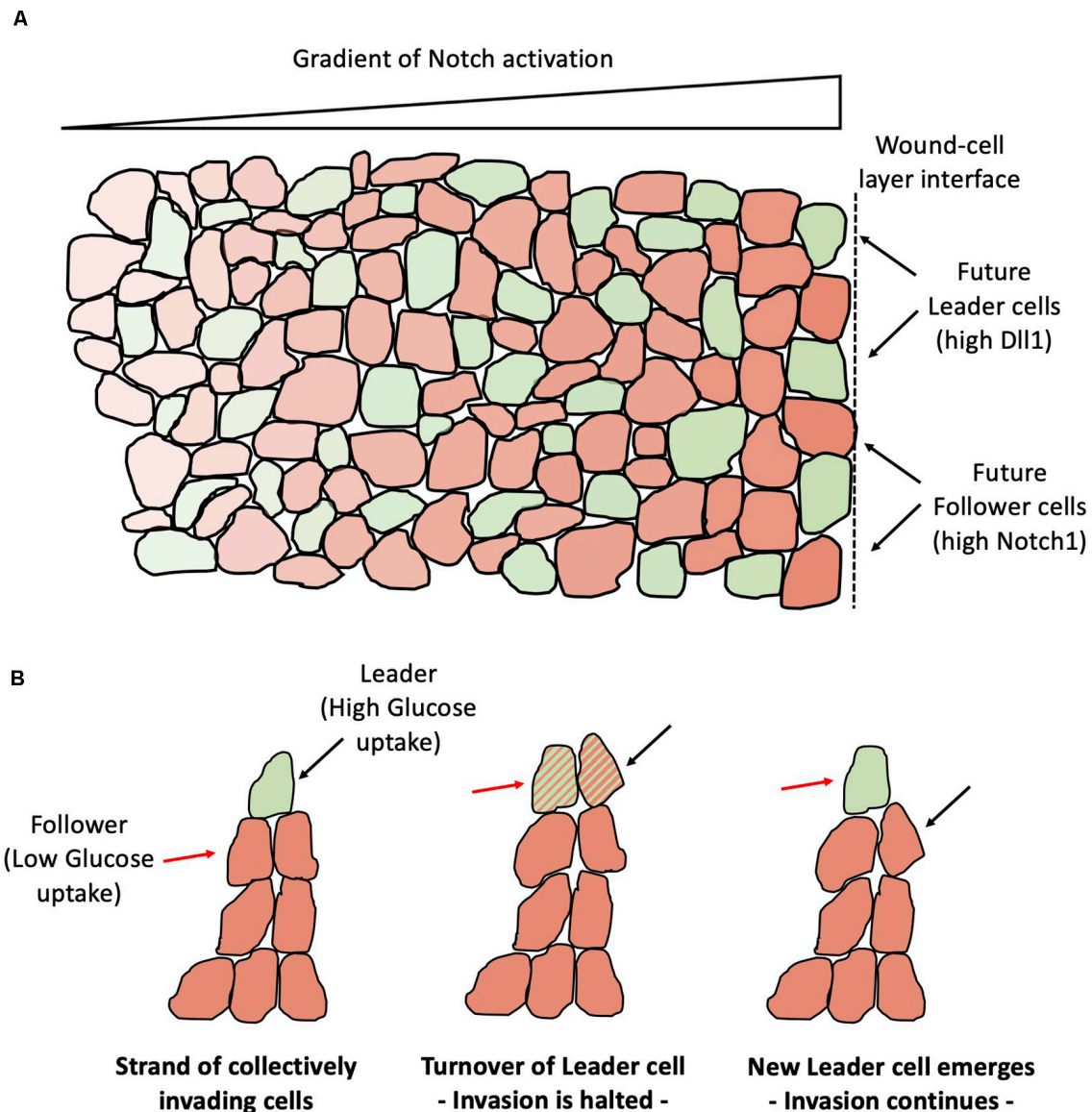


FIGURE 12 | Leader-Follower differentiation and turnover during collective epithelial migration. **(A)** Notch-Dll1 signaling differentiates cells that become leaders of the migration (green) and cells that become followers (orange). Notch-Dll1 signaling is more active toward the wound-cell layer interface (indicated by the black dashed line at the right end) and progressively inactivates far from the interface. **(B)** In strands of cells that migrate collectively, leaders have higher glucose uptake (left). Invasion halts in absence of a clear leader (center). The invasion continues after replacement of the leader cell (right). The red arrow points to a cell that is initially a follower and eventually emerges as the new migration leader. The black arrow points to a leader cell that is later substituted by a new leader.

stochastic fluctuations from competing cells that are initially in a similar cell state. A recent experiment showed via monolayer stress microscopy that mechanical interactions among follower cells behind the leading edge determine the selection and emergence of the leader cells at the leading edge (Vishwakarma et al., 2018). In other words, this finding suggests that follower cells decide the leader, not the other way around as has been a long-held belief. Another recent study shows that a leader cell maintains its foremost spatial position for only a finite period of time; later, some followers can replace the leader cells that have consumed most of their energy, indicating a dynamic turnover or

relay mechanism (Figure 12B; Zhang et al., 2019). Such metabolic regulation is likely to be connected to Notch signaling; future investigations addressing the coupling between signaling, energy consumption and mechanics will be crucial to elucidate the dynamical principles of collective cell migration.

Mechanosensitivity of Notch Signaling in Cardiovascular Morphogenesis

Evidence of Notch mechanosensitivity in leader-follower cell specification has been observed in a mouse model

of retinal angiogenesis, where the Notch1-Dll4 pathway regulates the density of tip cells that give rise to new capillaries from the existing vasculature (Wang et al., 2017). Although lateral inhibition is known to regulate tip/stalk differentiation during branching morphogenesis, this study showed that the tip/stalk differentiation heavily relies on the intercellular tension between cells in the blood vessel (Wang et al., 2017). Similarly to observations in collective epithelial cell migration, tension between cells restricts Notch1-Dll4 signaling and compromises tip cell selection (Wang et al., 2017). Overall, the density of tip cells and new branches was found to negatively correlate with the degree of mechanical stress, suggesting that Notch signaling might be tuned optimally at an intermediate range of intracellular tension that guarantees a proper angiogenic response, but limits the number of new branches (Wang et al., 2017). Interestingly, intercellular tension regulates the Notch-Delta and Notch-Jagged pathways differently in the context of human cardiovascular morphogenesis. Laminar shear stress decreases the expression of Dll4 in human umbilical vein endothelial cells (HUVEC) – as observed in mouse angiogenesis – but also increases the expression of Jagged1, and overall potentiates the signaling between endothelial cells (Driessen et al., 2018).

In the context of cardiovascular morphogenesis, the expression of Notch3, Jagged2 and multiple Notch targets decrease when a higher strain is imposed to vascular smooth muscle cells (VSMCs) (Loerakker et al., 2018). Incorporating the dependence of Notch expression on strain into a computational model shows that the mechanosensitivity of Notch signaling is key in regulating the thickness of the vascular wall. In fact, a switch in cell patterning was observed in a model with an increasing number of VSMCs corresponding to the wall thickness (Loerakker et al., 2018). For a short chain of cells (i.e., thin wall), most cells assumed a Sender state with high Delta. Conversely, thick walls exhibited a chain of cells in a Sender/Receiver state with high Notch and Jagged levels (Loerakker et al., 2018).

The coupling between Notch signaling and mechanical forces is not unidirectional: Notch signaling can, in turn, regulate the function of vascular barriers that separate blood from tissues. For instance, Notch drives the assembly of adherens junctions in a non-canonical mechanism (i.e., not via transcriptional regulation of E-Cadherin levels) (Polacheck et al., 2017). Consistently, reduction of Notch1 due to shear stresses leads to destabilization of adhesion junctions and proliferation of endothelial cells (Mack et al., 2017). Therefore, Notch1 can potentially act as a mechanosensor by regulating the response of endothelial cells based on intercellular stresses, mechanical injuries, and angiogenic signals (Mack et al., 2017). Therefore, while intercellular stresses might fine-tune Notch-Delta/Jagged signaling leading to new vessels, Notch signaling can, in turn, influence the defects in the structure of the vascular barrier by coordinating cell-cell adhesion. Future investigations integrating the interplay between Notch signaling, biomechanical aspects of mechanosensitivity, and the role of cell packing geometry will be valuable in elucidating the

emergent dynamics of tissue-level pattern formation in different biological contexts.

OPEN QUESTIONS AND FUTURE DIRECTIONS

Notch signaling is one of the most ubiquitous transduction pathways in vertebrates. Despite the variety of biological systems and processes, both physiological and pathological, that Notch signaling regulates, its structure and function are incredibly well-conserved.

Notch signaling has drawn incredible attention from the physics and mathematics community because, besides regulating cell-fate at a single cell level, it offers fertile ground to dissect the principles of spatiotemporal pattern formation in a tissue. To the eye of a physicist/mathematician, Notch signaling gives rise to the modes of lateral inhibition and lateral induction similarly to a system of spins that align together or in opposition in a magnet. However, unlike magnetism, these different outcomes of cell states emerge from underlying molecular interactions that are often non-linear and can be separated in time-scale as well. The geometrical intuition about Notch patterning via lateral inhibition and lateral induction provides a key to interpret experimental observations in physiological processes such as embryonic development and angiogenesis (Shaya and Sprinzak, 2011; Bray, 2016; Sjöqvist and Andersson, 2019). For example, lateral inhibition correctly predicts alternate patterning where hair cells (i.e., Senders) are surrounded by supporting cells (i.e., Receivers) and make up about 25% of the total cell fraction, such as in the cases of inner ear development and collective epithelial cell migration (Goodyear and Richardson, 1997; Riahi et al., 2014). Likewise, lateral induction describes well the propagation of similar cell fate observed, for instance, during inner ear development (Petrovic et al., 2014). More investigations, however, will be needed to truly test how well these simple models of biochemical kinetics and feedback loops capture the signaling and patterning dynamics emerging from Notch at a quantitative level.

Moreover, most of the theoretical efforts toward understanding the operating principles of Notch have focused on deterministic models. Cell-to-cell variability, however, can arise due to both stochasticity in the intracellular biochemical signaling (intrinsic noise) and fluctuations of other cellular components and/or in the extracellular environment (extrinsic noise) (Swain et al., 2002). Following a parallel between Notch and other patterning mechanisms driven by nearest neighbor signaling, such as the Ising model for a magnet, we speculate that stochastic fluctuation could play a relevant role in guiding, accelerating and/or disrupting ordered patterns (Rudge and Burrage, 2008).

Additional factors such as cell size and shape, affinity of ligand subtypes, molecular interactions within the same cell, and filopodia modulate the signaling. These mechanisms can be generally seen as details that add further complexity to the simple nearest neighbor's communication mechanism. For example, it is still not completely understood how trans- and cis-interactions

integrate to establish cell fate. Cis-interactions between receptor and ligands of the same cell can typically lead to mutual degradation (Celis and de Bray, 1997; Micchelli et al., 1997; Del Alamo et al., 2011). Recent evidence, however, suggests a role for cis-activation in the Notch pathway for multiple pairs of receptor and ligand subtypes (Nandagopal et al., 2019). Therefore, many context-specific signaling differences and their possible impact on spatiotemporal tissue dynamics deserve finer attention.

Moreover, early experimental findings suggest a role for mechanosensitivity in modulating Notch. The effects of extracellular forces on Notch activation are more quantified at the single molecule level (Meloty-Kapella et al., 2012; Wang and Ha, 2013; Chowdhury et al., 2016); it remains unclear, however, how these effects propagate at the level of multicellular patterning. On the experimental side, novel technologies that allow to probe the spatiotemporal Notch dynamics are starting to provide quantitative insights on the mechanochemical feedbacks between cell-cell signaling and cell mechanics (Riahi et al., 2015; Vishwakarma et al., 2018). On the other hand, integrating aspects of biochemical signaling, mechanical regulation and their interconnections is an important future challenge where theoretical and computational models can assist experimental design and *vice versa*.

Notch signaling has also received attention as a therapeutic target to curb cancer progression (Li et al., 2014; Siebel and Lendahl, 2017). While theoretical modeling of signaling and regulatory dynamics typically adopts modular approaches that treat different signaling modules as independent blocks, Notch seems to be implicated in several hallmarks of cancer progression, including drug-resistance, leaky/chaotic angiogenesis and enhanced invasion and metastasis (Li et al., 2014; Siebel and Lendahl, 2017). Jag1 is highly expressed in circulating tumor cell clusters with higher metastatic potential (Jolly et al., 2017) and by cancer cells that resist to drugs (Boareto et al., 2016; Yang et al., 2019). Generally speaking, cells that highly express Jagged seem to be associated with a more plastic and undifferentiated state such as hybrid epithelial/mesenchymal and/or a stem-like phenotype (Wang et al., 2009; Sikandar et al., 2010; Zhu et al., 2011; Yamamoto et al., 2013; Bocci et al., 2019a). Therefore, quantifying the role of interconnections between Notch and other hallmarks of cancer invasion will be a crucial challenge at the crossing point between theoretical modeling, biology and data science.

REFERENCES

- Aceto, N., Bardia, A., Miyamoto, D. T., Donaldson, M. C., Wittner, B. S., Spencer, J. A., et al. (2014). Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. *Cell* 158, 1110–1122. doi: 10.1016/j.cell.2014.07.013
- Andersson, E. R., Sandberg, R., and Lendahl, U. (2011). Notch signaling: simplicity in design, versatility in function. *Development* 138, 3593–3612. doi: 10.1242/dev.063610
- Andrew, C. O., Luis, G., and Morelli, S. A. (2012). Patterning embryos with oscillations: structure, function and dynamics of the vertebrate segmentation clock. *Development* 139, 625–639. doi: 10.1242/dev.063735

Another exciting direction concerns the widespread development of single cell sequencing techniques. Recently, Notch signaling has been studied at the single cell resolution in Zebrafish development, hematopoiesis and cancer stem cells (Mark et al., 2019; Tikhonova et al., 2019; Annika et al., 2020). For instance, it has been observed that downregulation of Notch ligands Dll1 and Dll4 in the bone marrow correlates with premature activation of a myeloid transcriptional program in hemopoietic stem cells (Tikhonova et al., 2019). While these studies provide detailed information on the transcriptional dynamics of Notch, they still lack the spatial resolution necessary to elucidate the underlying patterning mechanisms. Certainly, our understanding of Notch signaling will benefit from future developments in the field of single cell sequencing to account for spatial patterning.

Overall, insights from experimental and theoretical models continue to unravel the operating principles of Notch signaling, a master regulator of spatiotemporal cell patterning in development and tumor progression.

AUTHOR CONTRIBUTIONS

All authors discussed the content of the manuscript, agreed to be accountable for the content of the work, and discussed and edited the manuscript. FB wrote the manuscript. JO and MJ led the research.

FUNDING

The Onuchic Lab was supported by the National Science Foundation (NSF) grant for the Center for Theoretical Biological Physics NSF PHY-1427654 and CHE-1614101. JO is a CPRIT Scholar in Cancer Research sponsored by the Cancer Prevention and Research Institute of Texas. MJ was supported by the Ramanujan Fellowship awarded by SERB, DST, Government of India (SB/S2/RJN-049/2018).

ACKNOWLEDGMENTS

The authors express their gratitude to Herbert Levine and Marcelo Boareto for sharing useful comments on the manuscript.

- Annika, F., Daniel, B., Jonas, B., Klaus, J., Annika, W.-G., Severine, K., et al. (2020). Inhibiting WNT and NOTCH in renal cancer stem cells and the implications for human patients. *Nat. Commun.* 11:929.
- Appel, B., and Eisen, J. S. (1998). Regulation of neuronal specification in the zebrafish spinal cord by delta function. *Development* 125, 371–380.
- Bardot, B., Mok, L. P., Thayer, T., Ahimou, F., and Wesley, C. (2005). The Notch amino terminus regulates protein levels and Delta-induced clustering of *Drosophila* notch receptors. *Exp. Cell Res.* 304, 202–223. doi: 10.1016/j.yexcr.2004.10.030
- Barriga, E. H., Franze, K., Charras, G., and Mayor, R. (2018). Tissue stiffening coordinates morphogenesis by triggering collective cell migration in vivo. *Nature* 554, 523–527. doi: 10.1038/nature25742

- Benedito, R., and Hellström, M. (2013). Notch as a hub for signaling in angiogenesis. *Exp. Cell Res.* 319, 1281–1288. doi: 10.1016/j.yexcr.2013.01.010
- Benedito, R., Roca, C., Sörensen, I., Adams, S., Gossler, A., and Fruttiger, M. A. R. (2009). The notch ligands Dll4 and Jagged1 have opposing effects on angiogenesis. *Cell* 137, 1124–1135. doi: 10.1016/j.cell.2009.03.025
- Bentley, K., Gerhardt, H., and Bates, P. A. (2008). Agent-based simulation of notch-mediated tip cell selection in angiogenic sprout initialisation. *J. Theor. Biol.* 250, 25–36. doi: 10.1016/j.jtbi.2007.09.015
- Boareto, M., Jolly, M. K., Goldman, A., Pietilä, M., Mani, S. A., Sengupta, S., et al. (2016). Notch-Jagged signalling can give rise to clusters of cells exhibiting a hybrid epithelial/mesenchymal phenotype. *J. R. Soc. Interf.* 13:20151106. doi: 10.1098/rsif.2015.1106
- Boareto, M., Jolly, M. K., Lu, M., Onuchic, J. N., Clementi, C., and Ben-Jacob, E. (2015a). Jagged-Delta asymmetry in Notch signaling can give rise to a sender/receiver hybrid phenotype. *Proc. Natl. Acad. Sci. U.S.A.* 112, 402–409.
- Boareto, M., Kumar, M., Ben-jacob, E., and Onuchic, J. N. (2015b). Jagged mediates differences in normal, and tumor angiogenesis by affecting tip-stalk fate decision. *Proc. Natl. Acad. Sci. U.S.A.* 112, E3836–E3844.
- Bocci, F., Gaerhart-Serna, L., Ribeiro, M., Boareto, M., Ben-Jacob, E., Devi, G., et al. (2019a). Towards understanding Cancer Stem Cell heterogeneity in the tumor microenvironment. *Proc. Natl. Acad. Sci. U.S.A.* 116, 148–157.
- Bocci, F., Jolly, M. K., and Onuchic, J. N. (2019b). A biophysical model uncovers the size distribution of circulating Tumor Cell clusters across cancer types. *Cancer Res.* 79, 5535.
- Bocci, F., Levine, H., Onuchic, J. N., and Jolly, M. K. (2019c). Deciphering the dynamics of epithelial-mesenchymal transition, and cancer stem cells in (tumor). progression. *Curr. Stem Cell Rep.* 5, 11–21. doi: 10.1007/s40778-019-0150-3
- Bocci, F., Levine, H., Onuchic, J. N., and Jolly, M. K. (2019d). “Quantitative characteristic of ncRNA regulation in gene regulatory networks,” in *Computational Biology of Non-Coding RNA. Methods in Molecular Biology*, Vol. 1912, eds X. Lai, S. Gupta, and J. Vera (New York, NY: Humana Press), 341–366. doi: 10.1007/978-1-4939-8982-9_14
- Bocci, F., Tripathi, S. C., Vilchez, M. S. A., George, J. T., Casabar, J. P., Wong, P. K., et al. (2019e). NRF2 activates a partial epithelial-mesenchymal transition and is maximally present in a hybrid epithelial/mesenchymal phenotype. *Integr. Biol.* 11, 251–263. doi: 10.1093/intbio/zyz021
- Bocci, F., Jolly, M. K., George, J. T., Levine, H., and Onuchic, J. N. (2018a). A mechanism-based computational model to capture the interconnections among epithelial-mesenchymal transition, cancer stem cells, and Notch-Jagged signaling. *Oncotarget* 9, 29906–29920. doi: 10.18632/oncotarget.25692
- Bocci, F., Suzuki, Y., Lu, M., and Onuchic, J. N. (2018b). Role of metabolic spatiotemporal dynamics in regulating biofilm colony expansion. *Proc. Natl. Acad. Sci. U.S.A.* 115, 4288–4293. doi: 10.1073/pnas.1706920115
- Bocci, F., Jolly, M. K., Tripathi, S. C., Aguilar, M., Onuchic, N., Hanash, S. M., et al. (2017). Numb prevents a complete epithelial - mesenchymal transition by modulating Notch signalling. *J. R. Soc. Interf.* 14:20170512. doi: 10.1098/rsif.2017.0512
- Bockhorn, M., Jain, R. K., and Munn, L. L. (2007). Active versus passive mechanisms in metastasis: do cancer cells crawl into vessels, or are they pushed? *Lancet Oncol.* 8, 444–448. doi: 10.1016/s1470-2045(07)70140-7
- Bone, R. A., Bailey, C. S., Wiedermann, G., Ferjentsik, Z., Appleton, P. L., Murray, P. J., et al. (2014). Spatiotemporal oscillations of Notch1, Dll1 and are coordinated across the mouse PSM. *Development* 141, 4806–4816. doi: 10.1242/dev.115535
- Brabletz, S., Bajdak, K., Meidhof, S., Burk, U., Niedermann, G., Firat, E., et al. (2011). The ZEB1/miR-200 feedback loop controls Notch signalling in cancer cells. *EMBO J.* 30, 770–782. doi: 10.1038/emboj.2010.349
- Bray, S. J. (2006). Notch signalling: a simple pathway becomes complex. *Nat. Rev. Mol. Cell Biol.* 7, 678–689. doi: 10.1038/nrm2009
- Bray, S. J. (2016). Notch signaling in context. *Nat. Rev. Mol. Cell Biol.* 17, 722–735.
- Bu, P., Chen, K. Y., Chen, J. H., Wang, L., Walters, J., Shin, Y. J., et al. (2013). A microRNA miR-34a-regulated bimodal switch targets Notch in colon cancer stem cells. *Cell Stem Cell* 12, 602–615. doi: 10.1016/j.stem.2013.03.002
- Buceta, J., Herranz, H., Canela-Xandri, O., Reigada, R., Sagués, F., and Milán, M. (2007). Robustness and stability of the gene regulatory network involved in DV boundary formation in the *Drosophila* wing. *PLoS One* 2:e602. doi: 10.1371/journal.pcbi.00602
- Campos-Ortega, J. A. (1993). Mechanisms of early neurogenesis in *Drosophila melanogaster*. *Dev. Neurobiol.* 24, 1305–1327. doi: 10.1002/neu.480241005
- Celis, J. F., and de Bray, S. (1997). Feed-back mechanisms affecting notch activation at the dorsoventral boundary in the *Drosophila* wing. *Development* 124, 3241–3251.
- Celis, J. F., and de Garcia-Bellido, A. (1994). Roles of the Notch gene in *Drosophila* wing morphogenesis. *Mech. Dev.* 46, 109–122. doi: 10.1016/0925-4773(94)90080-9
- Chapman, G., Sparrow, D. B., Kremmer, E., and Dunwoodie, S. L. (2011). Notch inhibition by the ligand DELTA-LIKE 3 defines the mechanism of abnormal vertebral segmentation in *Spondylocostal dysostosis*. *Hum. Mol. Genet.* 20, 905–916. doi: 10.1093/hmg/ddq529
- Cheung, K. J., and Ewald, A. J. (2016). A collective route to metastasis: seeding by tumor cell clusters. *Science* 352, 167–169. doi: 10.1126/science.aaf6546
- Cheung, K. J., Gabrielson, E., Werb, Z., and Ewald, A. J. (2013). Collective invasion in breast cancer requires a conserved basal epithelial program. *Cell* 155, 1639–1651. doi: 10.1016/j.cell.2013.11.029
- Chowdhury, F., Li, I. T., Ngo, T. T., Leslie, B. J., Kim, B. C., Sokoloski, J. E., et al. (2016). Defining single molecular forces required for notch activation using nano yoyo. *Nano Lett.* 16, 3892–3897. doi: 10.1021/acs.nanolett.6b01403
- Cohen, M., Georgiou, M., Stevenson, N. L., Miodownik, M., and Baum, B. (2010). Dynamic filopodia transmit intermittent delta-notch signaling to drive pattern refinement during lateral inhibition. *Dev. Cell* 19, 78–89. doi: 10.1016/j.devcel.2010.06.006
- Collier, J. R., Monk, N. A., Maini, P. K., and Lewis, J. H. (1996). Pattern formation by lateral inhibition with feedback: a mathematical model of delta-notch intercellular signalling. *J. Theor. Biol.* 183, 429–446. doi: 10.1006/jtbi.1996.0233
- Coumilleau, F., Fürthauer, M., Knoblich, J. A., and González-Gaitán, M. (2009). Directional Delta and Notch trafficking in *Sara* endosomes during asymmetric cell division. *Nature* 458:10.
- Daudet, N., Ariza-McNaughton, L., and Lewis, J. (2007). Notch signalling is needed to maintain, but not to initiate, the formation of prosensory patches in the chick inner ear. *Development* 134, 2369–2378. doi: 10.1242/dev.001842
- Daudet, N., and Lewis, J. (2005). Two contrasting roles for Notch activity in chick inner ear development: specification of prosensory patches and lateral inhibition of hair-cell differentiation. *Development* 132, 541–551. doi: 10.1242/dev.01589
- de Antonellis, P., Medaglia, C., Cusanelli, E., Andolfo, I., Liguori, L., De Vita, G., et al. (2011). MiR-34a targeting of notch ligand delta-like 1 impairs CD15 +/CD133 + tumor-propagating cells and supports neural differentiation in medulloblastoma. *PLoS One* 6:e24584. doi: 10.1371/journal.pone.0024584
- Del Alamo, D., Rouault, H. i., and Schweisguth, F. (2011). Mechanism and significance of cis-inhibition in notch signalling review. *Curr. Biol.* 21, R40–R47.
- Delaune, E. A., François, P., Shih, N. P., and Amacher, S. L. (2012). Single-cell-resolution imaging of the impact of Notch signaling and mitosis on segmentation clock dynamics. *Dev. Cell* 23, 995–1005. doi: 10.1016/j.devcel.2012.09.009
- Driessen, R. C. H., Stassen, O. M. J. A., Sjöqvist, M., Suarez Rodriguez, F., Grolleman, J., Bouten, C. V. C., et al. (2018). Shear stress induces expression, intracellular reorganization and enhanced Notch activation potential of Jagged1. *Integr. Biol.* 10, 719–726. doi: 10.1039/c8ib00036k
- D'Souza, B., Meloty-Kapella, L., and Weinmaster, G. (2010). Canonical and non-canonical notch ligands. *Curr. Top. Dev. Biol.* 92, 73–129. doi: 10.1016/s0070-2153(10)92003-6
- Du, H., Nie, Q., and Holmes, W. R. (2015). The interplay between Wnt mediated expansion and negative regulation of growth promotes robust intestinal crypt structure and homeostasis. *PLoS Comput. Biol.* 11:e1004285. doi: 10.1371/journal.pcbi.1004285
- Eddison, M., Le Roux, I., and Lewis, J. (2002). Notch signaling in the development of the inner ear: lessons from *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11692–11699. doi: 10.1073/pnas.97.22.11692
- Eom, D. S., Bain, E. J., Patterson, L. B., Grout, M. E., and Parichy, D. M. (2015). Long-distance communication by specialized cellular projections during pigment pattern development and evolution. *eLife* 4:e12401.
- Fang, J. H., Zhou, H. C., Zhang, C., Shang, L. R., Zhang, L., Xu, J., et al. (2015). A novel vascular pattern promotes metastasis of hepatocellular carcinoma

- in an epithelial-mesenchymal transition-independent manner. *Hepatology* 62, 452–465. doi: 10.1002/hep.27760
- Fisher, J., Piterman, N., Hajnal, A., and Henzinger, T. A. (2007). Predictive modeling of signaling crosstalk during *C. elegans* vulval development. *PLoS Comput. Biol.* 3:e92. doi: 10.1371/journal.pcbi.100092
- Formosa-Jordan, P., and Ibáñez, M. (2009). Diffusible ligand and lateral inhibition dynamics for pattern formation. *J. Stat. Mech. Theory Exp.* 2009:03019.
- Formosa-Jordan, P., and Ibanes, M. (2014). Competition in notch signaling with cis enriches cell fate decisions. *PLoS One* 9:e95744. doi: 10.1371/journal.pcbi.10095744
- Friedl, P., and Mayor, R. (2017). Tuning collective cell migration by cell-cell junction regulation. *Cold Spring Harb. Perspect. Biol.* 9, 1–19.
- Gardner, T. S., Cantor, C. R., and Collins, J. J. (2000). Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 403, 339–342. doi: 10.1038/35002131
- Givan, L. A., Eisen, J. S., and Appel, B. (2001). Delta-Notch signaling and lateral inhibition in zebrafish spinal cord development. *BMC Dev. Biol.* 1:13. doi: 10.1186/1471-213X-1-13
- Goodyear, R., and Richardson, G. (1997). Pattern formation in the basilar papilla: evidence for cell rearrangement. *J. Neurosci.* 17, 6289–6301. doi: 10.1523/jneurosci.17-16-06289.1997
- Greenwald, I. (1998). LIN-12/Notch signaling: lessons from worms and flies. *Genes Dev.* 12, 1751–1762. doi: 10.1101/gad.12.12.1751
- Grosse-Wilde, A., d'Hérouël, A. F., McIntosh, E., Ertaylan, G., Skupin, A., Kuestner, R. E., et al. (2015). Stemness of the hybrid epithelial/mesenchymal state in breast cancer and its association with poor survival. *PLoS One* 10:e0126522. doi: 10.1371/journal.pcbi.100126522
- Gupta, G. P., and Massagué, J. (2006). Cancer metastasis: building a framework. *Cell* 127, 679–695. doi: 10.1016/j.cell.2006.11.001
- Guy, C. S., Vignali, K. M., Temirov, J., Bettini, M. L., Overacre, A. E., Smeltzer, M., et al. (2013). Distinct TCR signaling pathways drive proliferation and cytokine production in T cells. *Nat. Rev. Immunol.* 14, 262–270. doi: 10.1038/ni.2538
- Hartman, B. H., Reh, T. A., and Bermingham-McDonogh, O. (2010). Notch signaling specifies prosensory domains via lateral induction in the developing mammalian inner ear. *Proc. Natl. Acad. Sci. U.S.A.* 107, 15792–15797. doi: 10.1073/pnas.1002827107
- Heitzler, P., and Simpson, P. (1991). The choice of cell fate in the epidermis of *Drosophila melanogaster*. *Cell* 64, 1083–1092. doi: 10.1016/0092-8674(91)90263-x
- Hicks, C., Johnston, S. H., Disibio, G., Collazo, A., Vogt, T. F., and Weinmaster, G. (2000). Fringe differentially modulates Jagged1 and Delta1 signalling through Notch1 and Notch2. *Nat. Cell Biol.* 2, 515–520. doi: 10.1038/35019553
- Hirata, H., Yoshiura, S., Ohtsuka, T., Bessho, Y., Harada, T., Yoshikawa, K., et al. (2002). Oscillatory expression of the bHLH factor *hes1* regulated by a negative feedback loop. *Science* 298, 840–843. doi: 10.1126/science.1074560
- Holger, G., Matthew, G., Marcus, F., Christiana, R., Andrea, L., Alexandra, A., et al. (2003). VEGF guides angiogenic sprouting utilizing endothelial tip cell filopodia. *J. Cell Biol.* 161:1163. doi: 10.1083/jcb.200302047
- Homem, C. C. F., and Knoblich, J. A. (2012). *Drosophila* neuroblasts: a model for stem cell biology. *Development* 139, 4297–4310. doi: 10.1242/dev.080515
- Hou, X., Tashima, Y., and Stanley, P. (2012). Galactose differentially modulates lunatic and manic fringe effects on delta1-induced NOTCH signaling. *J. Biol. Chem.* 287, 474–483. doi: 10.1074/jbc.m111.317578
- Huang, H., and Kornberg, T. B. (2015). Myoblast cytonemes mediate Wg signaling from the wing imaginal disc and Delta-Notch signaling to the air sac primordium. *eLife* 4:e06114.
- Huang, P., Xiong, F., Megason, S. G., and Schier, A. F. (2012). Attenuation of Notch and hedgehog signaling is required for fate specification in the spinal cord. *PLoS Genet.* 8:e1002762. doi: 10.1371/journal.pcbi.1002762
- Huppert, S. S., Ilagan, M. X., De Strooper, B., and Kopan, R. (2005). Analysis of Notch function in presomitic mesoderm suggests a gamma-secretase-independent role for presenilins in somite differentiation. *Dev. Cell* 8, 677–688. doi: 10.1016/j.devcel.2005.02.019
- Huppert, S. S., Jacobsen, T. L., and Muskavitch, M. A. T. (1997). Feedback regulation is central to notch-delta signaling required for *Drosophila* wings. *Vein/intervein cell fate in Drosophila wings. Development* 124, 3283–3291.
- Jain, R. (2005). Normalization of tumor vasculature: an emerging concept in antiangiogenic therapy. *Science* 307, 58–62. doi: 10.1126/science.1104819
- Jakobsson, L., Franco, C. A., Bentley, K., Collins, R. T., Ponsioen, B., Aspalter, I. M., et al. (2010). Endothelial cells dynamically compete for the tip cell position during angiogenic sprouting. *Nat. Cell Biol.* 12, 943–953. doi: 10.1038/ncb2103
- Jane, Y., and Wu, Y. R. (1999). Fringe: defining borders by regulating the Notch pathway. *Curr. Opin. Neurobiol.* 9, 537–543. doi: 10.1016/S0959-4388(99)00020-3
- Jia, D., Jolly, M. K., Boareto, M., Parsana, P., Mooney, S. M., Pienta, K. J., et al. (2015). OVOL guides the epithelial-hybrid-mesenchymal transition. *Oncotarget* 6, 15436–15448. doi: 10.18632/oncotarget.3623
- Jia, D., Li, X., Bocci, F., Tripathi, S., Deng, Y., Jolly, M. K., et al. (2019). Quantifying cancer epithelial-mesenchymal plasticity and its association with stemness and immune response. *J. Clin. Med.* 8:725. doi: 10.3390/jcm8050725
- Jia, W., Deshmukh, A., Mani, S. A., Jolly, M. K., and Levine, H. (2019). A possible role for epigenetic feedback regulation in the dynamics of the epithelial-mesenchymal transition (EMT). *Phys. Biol.* 16:066004. doi: 10.1088/1478-3975/ab34df
- Jiang, Y.-J., Aerne, B. L., Smithers, L., Haddon, C., Ish-Horowicz, D., and Lewis, J. (2000). Notch signalling and the synchronization of the somite segmentation clock. *Nature* 408, 475–479. doi: 10.1038/35044091
- Jolly, M. K., Boareto, M., Debeb, B. G., Aceto, N., Farach-Carson, M. C., Woodward, W. A., et al. (2017). Inflammatory breast cancer: a model for investigating cluster-based dissemination. *NPJ Breast Cancer* 3:21.
- Jolly, M. K., Boareto, M., Lu, M., Onuchic, J. N., Clementi, C., Ben-Jacob, E., et al. (2015). Operating principles of Notch-Delta-Jagged module of cell-cell communication. *New J. Phys.* 17:55021.
- Joussineau, C., de Soule, J., Martin, M., Anguille, C., Montcourrier, P., and Alexandre, D. (2003). Delta-promoted filopodia mediate long-range lateral inhibition in *Drosophila*. *Nature* 426, 555–559. doi: 10.1038/nature02157
- Kang, T.-Y., Bocci, F., Jolly, M. K., Levine, H., Onuchic, J. N., and Levchenko, A. (2019). Pericytes enable effective angiogenesis in the presence of pro-inflammatory signals. *Proc. Natl. Acad. Sci. U.S.A.* 116, 23551–23561. doi: 10.1073/pnas.1913373116
- Katie, B., Franco, C. A., Philippides, A., Blanco, R., Dierkes, M., Gebala, V., et al. (2014). The role of differential VE-cadherin dynamics in cell rearrangement during angiogenesis. *Nat. Cell Biol.* 16, 309–321. doi: 10.1038/ncb2926
- Kerbel, R. S. (2008). Tumor angiogenesis. *N. Engl. J. Med.* 358, 2039–2049.
- Khait, I., Orsher, Y., Golan, O., Binshtok, U., Gordon-Bar, N., and Amir-Zilberstein, L. (2016). Quantitative analysis of delta-like 1 membrane dynamics elucidates the role of contact geometry on notch signaling. *Cell Rep.* 14, 225–233. doi: 10.1016/j.celrep.2015.12.040
- Kiernan, A. E., Ahituv, N., Fuchs, H., Balling, R., Avraham, K. B., and Steel, K. P. (2001). The Notch ligand Jagged1 is required for inner ear sensory development. *Proc. Natl. Acad. Sci. U.S.A.* 98, 3873–3878. doi: 10.1073/pnas.071496998
- Kiernan, A. E., Cordes, R., Kopan, R., Gossler, A., and Gridley, T. (2005). The Notch ligands DLL1 and JAG2 act synergistically to regulate hair cell development in the mammalian inner ear. *Development* 132, 4353–4362. doi: 10.1242/dev.02002
- Kiernan, A. E., Xu, J., and Gridley, T. (2006). The Notch ligand JAG1 is required for sensory progenitor development in the mammalian inner ear. *PLoS Genet.* 2:e4. doi: 10.1371/journal.pcbi.000004
- Kim, J., Irvine, K. D., and Carroll, S. B. (1995). Cell recognition, signal induction, and symmetrical gene activation at the dorsal-ventral boundary of the developing *Drosophila* wing. *Cell* 82, 795–802. doi: 10.1016/0092-8674(95)90476-x
- Koganehira, Y., Takeoka, M., Ehara, T., Sasaki, K., Murata, H., Saida, T., et al. (2003). Reduced expression of actin-binding proteins, h-caldesmon and calponin h1, in the vascular smooth muscle inside melanoma lesions: an adverse prognostic factor for malignant melanoma. *Br. J. Dermatol.* 148, 971–980. doi: 10.1046/j.1365-2133.2003.05238.x
- Koon, Y. L., Zhang, S., Rahmat, B. M., Koh, C. G., and Chiam, K.-H. (2018). Enhanced delta-notch lateral inhibition model incorporating intracellular notch heterogeneity and tension-dependent rate of delta-notch binding that reproduces sprouting angiogenesis patterns. *Sci. Rep.* 8:9519.
- Kopan, R., and Ilagan, M. X. G. (2009). The canonical notch signaling pathway: unfolding the activation mechanism. *Cell* 137, 216–233. doi: 10.1016/j.cell.2009.03.045

- Kovall, R. A., Gebelein, B., Sprinzak, D., and Kopan, R. (2017). The canonical notch signaling pathway: structural and biochemical insights into shape, sugar, and force. *Dev. Cell* 41, 228–241. doi: 10.1016/j.devcel.2017.04.001
- Ladi, E., Nichols, J. T., Ge, W., Miyamoto, A., Yao, C., Yang, L. T., et al. (2005). The divergent DSL ligand Dll3 does not activate Notch signaling but cell autonomously attenuates signaling induced by other DSL ligands. *J. Cell Biol.* 170, 983–992. doi: 10.1083/jcb.200503113
- Lamouille, S., Xu, J., and Derynck, R. (2014). Molecular mechanisms of epithelial-mesenchymal transition. *Nat. Rev. Mol. Cell Biol.* 15, 178–196. doi: 10.1038/nrm3758
- Lawson, N. D., Scheer, N., Pham, V. N., Kim, C. H., Chitnis, A. B., and Campos-Ortega, J. A. (2001). Notch signaling is required for arterial-venous differentiation during embryonic vascular development. *Development* 128, 3675–3683.
- LeBon, L., Lee, T. V., Jafar-Nejad, H., Sprinzak, D., and Elowitz, M. B. (2014). Fringe proteins modulate Notch-ligand cis and trans interactions to specify signaling states. *eLife* 3:2950. doi: 10.7554/eLife.02950
- Lewis, A. K., Frantz, G. D., Carpenter, D. A., de Sauvage, F. J., and Gao, W. Q. (1998). Distinct expression patterns of notch family receptors and ligands during development of the mammalian inner ear. *Mech. Dev.* 78, 159–163. doi: 10.1016/s0925-4773(98)00165-8
- Lewis, J. (2003). Autoinhibition with transcriptional delay: a simple mechanism for the zebrafish somitogenesis oscillator. *Curr. Biol.* 13, 1398–1408. doi: 10.1016/s0960-9822(03)00534-7
- Li, C., and Balazsi, G. (2018). A landscape view on the interplay between EMT and cancer metastasis. *npj Syst. Biol. Appl.* 4:38.
- Li, C., Hong, T., and Nie, Q. (2016). Quantifying the landscape and kinetic paths for epithelial-mesenchymal transition from a core circuit. *Phys. Chem. Chem. Phys.* 18, 17949–17956. doi: 10.1039/c6cp03174a
- Li, D., Masiero, M., Banham, A. H., and Harris, A. L. (2014). The notch ligand JAGGED1 as a target for anti-tumor therapy. *Front. Oncol.* 4:254. doi: 10.3389/fonc.2014.00254
- Liu, S., Cong, Y., Wang, D., Sun, Y., Deng, L., Liu, Y., et al. (2014). Breast cancer stem cells transition between epithelial and mesenchymal states reflective of their normal counterparts. *Stem Cell Rep.* 2, 78–91. doi: 10.1016/j.stemcr.2013.11.009
- Loerakker, S., Stassen, O. M. J. A., ter Huurne, F. M., Boareto, M., Bouten, C. V. C., and Sahlgren, C. M. (2018). Mechanosensitivity of Jagged-Notch signaling can induce a switch-type behavior in vascular homeostasis. *Proc. Natl. Acad. Sci. U.S.A.* 115, E3682–E3691.
- Louisa, M., Morf, M. K., Spiri, S., Gutierrez, P., Escobar-Restrepo, J. M., Daube, M., et al. (2020). Polarized epidermal growth factor secretion ensures robust vulval cell fate specification in *Caenorhabditis elegans*. *Development* 147:dev175760. doi: 10.1242/dev.175760
- Luca, V. C., Jude, K. M., Pierce, N. W., Nachury, M. V., Fischer, S., and Garcia, K. C. (2017a). Structural basis for Notch1 engagement of Delta-like 4. *Science* 347, 847–853. doi: 10.1126/science.1261093
- Luca, V. C., Kim, B. C., Ge, C., Kakuda, S., Wu, D., Roein-Peikar, M., et al. (2017b). Notch-Jagged complex structure implicates a catch bond in tuning ligand sensitivity. *Science* 355, 1320–1324. doi: 10.1126/science.aaf9739
- Luo, M., Shang, L., Brooks, M. D., Jiagge, E., Zhu, Y., Buschhaus, J. M., et al. (2018). Targeting breast cancer stem cell state equilibrium through modulation of redox signaling. *Cell Metab.* 28, 69–86.
- Mack, J. J., Mosquero, T. S., Archer, B. J., Jones, W. M., Sunshine, H., Faas, G. C., et al. (2017). NOTCH1 is a mechanosensor in adult arteries. *Nat. Commun.* 8:1620.
- Manderfield, L. J., High, F. A., Engleka, K. A., Liu, F., Li, L., Rentschler, S., et al. (2012). Notch activation of jagged1 contributes to the assembly of the arterial wall. *Circulation* 125, 314–323. doi: 10.1161/circulationaha.111.047159
- Mani, S. A., Guo, W., Liao, M. J., Eaton, E. N., Ayyanan, A., Zhou, A. Y., et al. (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* 133, 704–715.
- Mark, E. L., Daniel, C. D., Nina, K., Sungmin, B., Helena, B., and St Peter, M. K. (2019). scRNA-Seq reveals distinct stem cell populations that drive hair cell regeneration after loss of Fgf and Notch signaling. *eLife* 8:e44431.
- Meloty-Kapella, L., Shergill, B., Kuon, J., Botvinick, E., and Weinmaster, G. (2012). Notch ligand endocytosis generates mechanical pulling force dependent on dynamin, epsins, and actin. *Dev. Cell* 22, 1299–1312. doi: 10.1016/j.devcel.2012.04.005
- Michelli, C. A., Rulifson, E. J., and Blair, S. S. (1997). The function and regulation of cut expression on the wing margin of *Drosophila*: notch, Wingless and a dominant negative role for Delta and Serrate. *Development* 124, 1485–1495.
- Miller, A. C., Lyons, E. L., and Herman, T. G. (2009). cis-Inhibition of notch by endogenous delta biases the outcome of lateral inhibition. *Curr. Biol.* 19, 1378–1383. doi: 10.1016/j.cub.2009.06.042
- Monk, N. A. M. (2003). Oscillatory expression of Hes1, p53, and NF- κ B driven by transcriptional time delays. *Curr. Biol.* 13, 1409–1413. doi: 10.1016/s0960-9822(03)00494-9
- Nadia, A., and Rana, R. S. H. (2011). Fringe benefits: functional and structural impacts of O-glycosylation on the extracellular domain of Notch receptors. *Curr. Opin. Struct. Biol.* 21, 583–589. doi: 10.1016/j.sbi.2011.08.008
- Nandagopal, N., Santat, L. A., and Elowitz, M. B. (2019). Cis-activation in the Notch signaling pathway. *eLife* 8:e37880.
- Nandagopal, N., Santat, L. A., LeBon, L., Sprinzak, D., Bronner, M. E., and Elowitz, M. B. (2018). Dynamic ligand discrimination in the notch signaling pathway. *Cell* 172, 869–880.
- Neves, J., Abelló, G., Petrovic, J., and Giraldez, F. (2013). Patterning and cell fate in the inner ear: a case for Notch in the chicken embryo. *Dev. Growth Differ.* 55, 96–112. doi: 10.1111/dgd.12016
- Nichols, J. T., Miyamoto, A., Olsen, S. L., D'Souza, B., Yao, C., and Weinmaster, G. (2007). DSL ligand endocytosis physically dissociates Notch1 heterodimers before activating proteolysis can occur. *J. Cell Biol.* 176, 445–458. doi: 10.1083/jcb.200609014
- Nie, Q. (2018). Stem cells? a window of opportunity in low-dimensional EMT space. *Oncotarget* 9, 31790–31791. doi: 10.18632/oncotarget.25852
- Niessen, K., Fu, Y., Chang, L., Hoodless, P. A., McFadden, D., and Karsan, A. (2008). Slug is a direct Notch target required for initiation of cardiac cushion cellularization. *J. Cell Biol.* 182, 315–325. doi: 10.1083/jcb.200710067
- Nieto, M. A., Huang, R. Y., Jackson, R. A., and Thiery, J. P. (2016). EMT: 2016. *Cell* 166, 21–45.
- OZbudak, E. M., and Lewis, J. (2008). Notch signalling synchronizes the zebrafish segmentation clock but is not needed to create somite boundaries. *PLoS Genet.* 4:e15. doi: 10.1371/journal.pcbi.1000015
- Pastushenko, I., and Blanpain, C. (2018). EMT transition states during tumor progression and metastasis. *Trends Cell Biol.* 29, 212–226. doi: 10.1016/j.tcb.2018.12.001
- Pastushenko, I., Brisebarre, A., Sifrim, A., Fioramonti, M., Revenco, T., Boumahdi, S., et al. (2018). Identification of the tumour transition states occurring during EMT. *Nature* 556, 463–468. doi: 10.1038/s41586-018-0040-3
- Petrovic, J., Formosa-Jordan, P., Luna-Escalante, J. C., Abello, G., Ibanes, M., Neves, J., et al. (2014). Ligand-dependent Notch signaling strength orchestrates lateral induction and lateral inhibition in the developing inner ear. *Development* 141, 2313–2324. doi: 10.1242/dev.108100
- Polacheck, W. J., Kutys, M. L., Yang, J., Eyckmans, J., Wu, Y., Vasavada, H., et al. (2017). A non-canonical Notch complex regulates adherens junctions and vascular barrier function. *Nature* 552, 258–262. doi: 10.1038/nature24998
- Preuß, K., Tverikhina, L., Schuster-Gossler, K., Gaspar, C., Rosa, A. I., and Henrique, D. (2015). Context-dependent functional divergence of the notch ligands DLL1 and DLL4 in vivo. *PLoS Genet.* 11:e1005328. doi: 10.1371/journal.pcbi.1005328
- Qutub, A. A., Gabhann, F. M. A. C., Karagiannis, E. D., Vempati, P., and Popel, A. S. (2009). Multiscale models of angiogenesis: integration of molecular mechanisms with cell- and organ-level models. *IEEE Eng. Med. Mag.* 28, 14–31.
- Riahi, R., Long, M., Yang, Y., Dean, Z., Zhang, D. D., Slepian, M. J., et al. (2014). Single cell gene expression analysis in injury-induced collective cell migration. *Integr. Biol.* 6, 192–202. doi: 10.1039/c3ib40095f
- Riahi, R., Sun, J., Wang, S., Long, M., Zhang, D. D., and Wong, P. K. (2015). Notch1-Dll4 signalling and mechanical force regulate leader cell formation during collective cell migration. *Nat. Commun.* 6, 1–11.
- Riedel-Kruse, I. H., Müller, C., and Oates, A. C. (2007). Synchrony dynamics during initiation, failure, and rescue of the segmentation clock. *Science* 317, 1911–1915. doi: 10.1126/science.1142538
- Rudge, T., and Burrage, K. (2008). Effects of intrinsic and extrinsic noise can accelerate juxtacrine pattern formation. *Bull. Math. Biol.* 70, 971–991. doi: 10.1007/s11538-007-9286-x

- Sahlgren, C., Gustafsson, M. V., Jin, S., Poellinger, L., and Lendahl, U. (2008). Notch signaling mediates hypoxia-induced tumor cell migration and invasion. *Proc. Natl. Acad. Sci. U.S.A.* 105, 6392–6397. doi: 10.1073/pnas.0802047105
- Sha, Y., Haensel, D., Gutierrez, G., Du, H., Dai, X., and Nie, Q. (2018). Intermediate cell states in epithelial-to-mesenchymal transition. *Phys. Biol.* 16:21.
- Shaya, O., Binstok, U., Hersch, M., Rivkin, D., Weinreb, S., Amir-Zilberstein, L., et al. (2017). Cell-cell contact area affects Notch signaling and Notch-dependent patterning. *Dev. Cell* 40, 505–511.
- Shaya, O., and Sprinzak, D. (2011). From Notch signaling to fine-grained patterning: Modeling meets experiments. *Curr. Opin. Genet. Dev.* 21, 732–739. doi: 10.1016/j.gde.2011.07.007
- Shimojo, H., Isomura, A., Ohtsuka, T., Kori, H., Miyachi, H., and Kageyama, R. (2016). Oscillatory control of Delta-like1 in cell interactions regulates dynamic gene expression and tissue morphogenesis. *Genes Dev.* 30, 102–116.
- Shimojo, H., and Kageyama, R. (2016). Oscillatory control of Delta-like1 in somitogenesis and neurogenesis: a unified model for different oscillatory dynamics. *Semin. Cell Dev. Biol.* 49, 76–82.
- Shimojo, H., Ohtsuka, T., and Kageyama, R. (2008). Oscillations in notch signaling regulate maintenance of neural progenitors. *Neuron* 58, 52–64.
- Shimojo, H., Ohtsuka, T., and Kageyama, R. (2011). Dynamic expression of notch signaling genes in neural stem/progenitor cells. *Front. Neurosci.* 5:78. doi: 10.3389/fnec.2014.0078
- Siebel, C., and Lendahl, U. (2017). Notch signaling in development, tissue homeostasis, and disease. *Physiol. Rev.* 97, 1235–1294.
- Sikandar, S. S., Pate, K. T., Anderson, S., Dizon, D., Edwards, R. A., Waterman, M. L., et al. (2010). NOTCH signaling is required for formation and self-renewal of tumor-initiating cells and for repression of secretory cell differentiation in colon cancer. *Cancer Res.* 70, 1469–1478.
- Simpson, P. (1997). Notch signalling in development: on equivalence groups and asymmetric developmental potential. *Curr. Opin. Genet. Dev.* 7, 537–542.
- Sjöqvist, M., and Andersson, E. R. (2019). Do as I say, Not (ch). as I do? lateral control of cell fate. *Dev. Biol.* 447, 58–70.
- Skeath, J. B., and Carroll, S. B. (1992). Regulation of proneural gene expression and cell fate during neuroblast segregation in the *Drosophila* embryo. *Development* 114, 939–946.
- Sprinzak, D., Lakhanpal, A., Lebon, L., Garcia-ojalvo, J., and Elowitz, M. B. (2011). Mutual inactivation of notch receptors and ligands facilitates developmental patterning. *PLoS Comput. Biol.* 7:e1002069. doi: 10.1371/journal.pcbi.1002069
- Sprinzak, D., Lakhanpal, A., Lebon, L., Santat, L. A., Fontes, M. E., Anderson, G. A., et al. (2010). Cis-interactions between Notch and Delta generate mutually exclusive signalling states. *Nature* 465, 86–90.
- Stopka, A., Boaretto, M., and Iber, D. (2019). Cell-based simulations of Notch-dependent cell differentiation on growing domains. *bioRxiv* [Preprint]. doi: 10.1101/859363
- Suchting, S., Freitas, C., Le Noble, F., Benedito, R., Bréant, C., and Duarte, A. E. A. (2007). The Notch ligand Delta-like 4 negatively regulates endothelial tip cell formation and vessel branching. *Proc. Natl. Acad. Sci. U.S.A.* 104, 3225–3230.
- Swain, P. S., Elowitz, M. B., and Siggia, E. D. (2002). Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 99, 12795–12800. doi: 10.1073/pnas.162041399
- Teomy, E., Levine, H., and Kessler, D. A. (2019). Ordered hexagonal patterns via notch-delta signaling. *bioRxiv* [Preprint]. Available online at: <https://arxiv.org/abs/1902.04917> (accessed February 15, 2019).
- Thurston, G., Noguera-Troise, I., and Yancopoulos, G. D. (2007). The Delta paradox: DLL4 blockade leads to more tumour vessels but less tumour growth. *Nat. Rev. Cancer* 7, 327–331.
- Tikhonova, A. N., Dolgalev, I., Hu, H., Sivaraj, K., Hoxha, E., and Cuesta-Domínguez, A. (2019). The bone marrow microenvironment at single-cell resolution. *Nature* 569, 222–228. doi: 10.1038/s41586-019-1104-8
- Varahan, S., Walvekar, A., Sinha, V., Krishna, S., and Laxman, S. (2019). Metabolic constraints drive self-organization of specialized cell groups. *eLife* 8:e46735.
- Venkatraman, L., Regan, E. R., and Bentley, K. (2016). Time to decide? dynamical analysis predicts partial tip/stalk patterning states arise during angiogenesis. *PLoS One* 11:e0166489. doi: 10.1371/journal.pcbi.10066489
- Venzin, O. F., and Oates, A. (2019). What are you synching about? emerging complexity of notch signaling in the segmentation clock. *Dev. Biol.* 1606, 30444–30445.
- Vishwakarma, M., Di Russo, J., Probst, D., Schwarz, U. S., Das, T., and Spatz, J. P. (2018). Mechanical interactions among followers determine the emergence of leaders in migrating epithelial cell collectives. *Nat. Commun.* 9:3469.
- Wang, S., Sun, J., Xiao, Y., Lu, Y., Zhang, D. D., and Wong, P. K. (2017). Intercellular tension negatively regulates angiogenic sprouting of endothelial tip cells via Notch1-Dll4 signaling. *Adv. Biosyst.* 1:1600019.
- Wang, X., and Ha, T. (2013). Defining single molecular forces required to activate integrin and notch signaling. *Science* 340, 991–994.
- Wang, Z., Li, Y., Kong, D., Banerjee, S., Ahmad, A., Azmi, A. S., et al. (2009). Acquisition of epithelial-mesenchymal transition phenotype of gemcitabine-resistant pancreatic cancer cells is linked with activation of the notch signaling pathway. *Cancer Res.* 69, 2400–2407.
- Weinstein, N., Mendoza, L., and Álvarez-Buylla, E. R. (2020). A computational model of the endothelial to mesenchymal transition. *Front. Genet.* 11:40. doi: 10.3389/fnec.2014.0040
- Weinstein, N., Mendoza, L., Gitler, S., and Klapp, J. (2017). A network model to explore the effect of the micro-environment on endothelial cell behavior during angiogenesis. *Front. Physiol.* 8:960. doi: 10.3389/fnec.2014.00960
- Weis, S. M., and Cheres, D. A. (2011). Tumor angiogenesis: molecular pathways and therapeutic targets. *Nat. Med.* 17, 1359–1370.
- Xu, K., Usary, J., Kousis, P. C., Prat, A., Wang, D. Y., Adams, J. R., et al. (2012). Lunatic fringe deficiency cooperates with the Met/Caveolin gene amplicon to induce basal-like breast cancer. *Cancer Cell* 21, 626–641. doi: 10.1016/j.ccr.2012.03.041
- Yamamoto, M., Taguchi, Y., Ito-Kureha, T., Semba, K., Yamaguchi, N., and Inoue, J. (2013). NF- κ B non-cell-autonomously regulates cancer stem cell populations in the basal-like breast cancer subtype. *Nat. Commun.* 4, 1–13.
- Yang, J., Xing, H., Lu, D., Wang, J., Li, B., Tang, J., et al. (2019). Role of Jagged1/STAT3 signalling in platinum-resistant ovarian cancer. *J. Cell. Mol. Med.* 23, 4005–4018.
- Yang, Y., Jamilpour, N., Yao, B., Dean, Z. S., Riahi, R., and Wong, P. K. (2016). Probing leader cells in endothelial collective migration by plasma lithography geometric confinement. *Sci. Rep.* 6, 1–11.
- Yi, F., Amarasinghe, B., and Dang, T. P. (2013). Manic fringe inhibits tumor growth by suppressing Notch3 degradation in lung cancer. *Am. J. Cancer Res.* 3, 490–499.
- Zhang, J., Goliwas, K. F., Wang, W., Taufalele, P. V., Bordeleau, F., and Reinhart-King, C. A. (2019). Energetic regulation of coordinated leader-follower dynamics during collective invasion of breast cancer cells. *Proc. Natl. Acad. Sci. U.S.A.* 116, 7867–7872.
- Zhang, S., Chung, W., Wu, G., Egan, S. E., and Xu, K. (2014). Tumor-suppressive activity of lunatic fringe in prostate through differential modulation of notch receptor activation. *Neoplasia* 16, 158–167.
- Zhu, T. S., Costello, M. A., Talsma, C. E., Flack, C. G., Crowley, J. G., Hamm, L. L., et al. (2011). Endothelial cells create a stem cell niche in glioblastoma by providing NOTCH ligands that nurture self-renewal of cancer stem-like cells. *Cancer Res.* 71, 6061–6072.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Bocci, Onuchic and Jolly. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



A Guard Cell Abscissic Acid (ABA) Network Model That Captures the Stomatal Resting State

Parul Maheshwari^{1*}, Sarah M. Assmann² and Reka Albert^{1,2*}

¹ Department of Physics, Penn State University, University Park, PA, United States, ² Department of Biology, Penn State University, University Park, PA, United States

OPEN ACCESS

Edited by:

Chunhe Li,
Fudan University, China

Reviewed by:

Jie Zheng,
ShanghaiTech University, China
Mohit Kumar Jolly,
Indian Institute of Science, India

*Correspondence:

Parul Maheshwari
ppm5104@psu.edu
Reka Albert
rza1@psu.edu

Specialty section:

This article was submitted to
Systems Biology,
a section of the journal
Frontiers in Physiology

Received: 11 May 2020

Accepted: 09 July 2020

Published: 13 August 2020

Citation:

Maheshwari P, Assmann SM and
Albert R (2020) A Guard Cell Abscissic
Acid (ABA) Network Model That
Captures the Stomatal Resting State.
Front. Physiol. 11:927.
doi: 10.3389/fphys.2020.00927

Stomatal pores play a central role in the control of carbon assimilation and plant water status. The guard cell pair that borders each pore integrates information from environmental and endogenous signals and accordingly swells or deflates, thereby increasing or decreasing the stomatal aperture. Prior research shows that there is a complex cellular network underlying this process. We have previously constructed a signal transduction network and a Boolean dynamic model describing stomatal closure in response to signals including the plant hormone abscissic acid (ABA), calcium or reactive oxygen species (ROS). Here, we improve the Boolean network model such that it captures the biologically expected response of the guard cell in the absence or following the removal of a closure-inducing signal such as ABA or external Ca^{2+} . The expectation from the biological system is reversibility, i.e., the stomata should reopen after the closing signal is removed. We find that the model's reversibility is obstructed by the previously assumed persistent activity of four nodes. By introducing time-dependent Boolean functions for these nodes, the model recapitulates stomatal reopening following the removal of a signal. The previous version of the model predicts ~20% closure in the absence of any signal due to uncertainty regarding the initial conditions of multiple network nodes. We systematically test and adjust these initial conditions to find the minimally restrictive combinations that appropriately result in open stomata in the absence of a closure signal. We support these results by an analysis of the successive stabilization of feedback motifs in the network, illuminating the system's dynamic progression toward the open or closed stomata state. This analysis particularly highlights the role of cytosolic calcium oscillations in causing and maintaining stomatal closure. Overall, we illustrate the strength of the Boolean network modeling framework to efficiently capture cellular phenotypes as emergent outcomes of intracellular biological processes.

Keywords: stomatal closure, guard cell, Boolean network, Boolean model, memory, signal transduction

INTRODUCTION

Stomatal pores on the surfaces of leaves play an important role in allowing uptake of CO_2 for photosynthesis and water vapor loss in transpiration. Guard cell pairs surround these stomatal pores and control their aperture by dynamic shrinking and swelling. Guard cells respond to numerous environmental signals, including light of different wavelengths and ambient CO_2 . In

response to drought and other desiccating stresses, the plant produces the hormone abscisic acid (ABA), which induces stomatal closure. The process of stomatal closure in response to ABA involves the activity of many guard cell ion transport proteins, enzymes, and small molecules. There has been significant previous research involving experiments and theoretical and computational analyses to understand this complex process (Li et al., 2006; Ma et al., 2009; Sun et al., 2014).

Given the complex interactions among the many intracellular components of guard cells, network-based dynamic models constitute an efficient method for understanding the system. Dynamic models represent each intracellular component by a qualitative (discrete) or quantitative state variable, which describes the component's abundance or activity. Dynamic models also describe how the interactions and regulatory relationships among components (represented as directed edges in the network) change each component's state variable. Discrete dynamic models are simpler to create than quantitative models (which need extensive parameterization) yet they are rich enough to recapitulate and predict cellular behavior. The simplest type of discrete dynamic model is the Boolean model, wherein each node in the network is assumed to be either in the ON (1) or the OFF (0) state and each node state evolves in time according to its update function. The update function of a node is a logic function of the states of the regulators of the node and is expressed with the AND, OR, and NOT logic operators.

We have previously published network-based discrete dynamic models elucidating the process of stomatal response to various signals, including ABA and light of different wavelengths (Li et al., 2006; Sun et al., 2014; Gan and Albert, 2016; Albert et al., 2017; Maheshwari et al., 2019). These models use extensive iterations between experiments and model simulations. They capture the state corresponding to open or closed stomata as stable final states (attractors) of the dynamical system and they recapitulate almost all experimental results regarding knockout and constitutive activation of various nodes of the network. These models identify important mediators of stomatal closure induced by ABA or by external supply of mediators of ABA-induced closure (such as Ca^{2+}). They also identify the various subnetworks that determine the different attractors of the network. These subnetworks, called stable motifs, form generalized positive feedback loops that once stabilized, maintain the constituent nodes in a fixed state (Zañudo and Albert, 2013). The most recent model (Maheshwari et al., 2019) identifies single nodes and combinations of nodes that are sufficient to drive the system to a particular attractor. The model also elucidates various crucial feedback loops that ensure the coordination between different components of the network.

The model of Maheshwari et al. (2019) however, does not recapitulate a few cases where the biological response is open stomata. In the model, stomatal closure is an attractor that is not reversible by the removal of the closure-inducing signal. In contrast, the biological reality in a situation of providing a closure-inducing signal for a limited period is gradual opening of the stomata following removal of the signal (Cummins et al., 1971). Furthermore, this model predicts ~20% closed stomata in the absence of any signal (see Figure 3 of Maheshwari et al., 2019)

while the biological reality is that the stomata remain open in the absence of any closure-inducing signal. Here, we present two different modifications of the network model of Maheshwari et al. (2019) such that it recapitulates the expected stomatal behavior in these situations. The first modification corresponds to revising certain assumptions of sustained activities for four nodes in the model. Instead of assuming persistent activity, we incorporate a short-term memory effect for these four nodes, where the node's update function also considers the previous states of its regulator node. To understand the trajectories that lead to the open or closed stomata attractors, we construct the motif succession diagrams for the model version with persistent activity of these nodes as well as for the model version with short-term memory. The second modification corresponds to narrowing down the initial conditions used in the model simulations to ensure that the model recapitulates the open stomata state in the absence of any signal. This analysis elucidates the sensitive dependence of stomatal closure on the initial condition of six nodes whose pre-stimulus states are currently unknown. We use network analysis and causal logic (Maheshwari and Albert, 2017) to reveal the pathways by which these six nodes can lead to stomatal closure.

MATERIALS AND METHODS

Background Information on the Boolean Model of ABA-Induced Stomatal Closure

Multiple iterations of experimental and computational research have led to a successful Boolean network model of ABA induced stomatal closure (Li et al., 2006; Albert et al., 2017; Maheshwari et al., 2019). The model relies on extensive literature curation to integrate the signaling components, interactions and mechanisms that underlie ABA-induced stomatal closure. The first version of the network model, containing 40 signaling components (Li et al., 2006), successfully recapitulated knockout phenotypes observed at that time and predicted many new phenotypes. One such prediction, regarding the importance of pH changes in ABA induced closure, was experimentally tested and validated in Li et al. (2006). The model later was expanded to 84 nodes based on new experimental results, for example concerning the identity of ABA receptors (Albert et al., 2017). The model was compared to a full complement of phenotypes observed by wet-bench experimentation and achieved a high degree of agreement. Several predictions of the model with regard to reactive oxygen species, cytosolic Ca^{2+} , and heterotrimeric G-protein signaling were confirmed experimentally in Albert et al. (2017). For ease of simulation and understanding, this larger network model was later reduced to 49 signaling components in a way that preserves the outcomes of the model (Maheshwari et al., 2019).

A significant insight obtained via this model was the recognition of a feedback-rich core that complements a canonical linear ABA signaling pathway. Sustained presence of ABA leads to ABA binding to RCAR/PYR1/PYL receptors (short for soluble pyrabactin resistance 1/pyrabactin resistance 1-like regulatory component of ABA receptor). This leads to the inhibition of the clade-A protein phosphatase 2Cs (PP2Cs), which relieves

inhibition of the serine-threonine kinase OPEN STOMATA 1 (OST1). Kinase activity of OST1 results in reactive oxygen species (ROS) production, which in turn enhances Ca^{2+} uptake through the membrane (represented as the node CaIM in the model). This Ca^{2+} uptake, combined with Ca^{2+} release from intracellular stores (CIS), leads to cytosolic Ca^{2+} oscillations, which, together with the sustained presence of ABA, lead to production of phosphatidic acid (PA), ROS, and activation of phospholipase D delta (PLD δ). These three components form a positive feedback loop and hence they maintain their activation (Sierla et al., 2016). The cytosolic Ca^{2+} oscillations directly or indirectly also lead to activation of mitogen-activated protein kinases 9 and 12 (MPK9/12), calcium-dependent protein kinases 3 and 21 (CPK3/21), and depolymerization of microtubules. The ABA signal propagates through these various nodes and feedback loops to ultimately trigger opening of anion channels at the guard cell membrane (Kollist et al., 2014), leading to anion efflux. Anion efflux induces membrane depolarization, which in turn drives K^{+} efflux through depolarization-activated outward K^{+} channels (Ache et al., 2000; Hosy et al., 2003). Solute loss drives the osmotic efflux of water through aquaporins (Grondin et al., 2015), resulting in guard cell deflation and stomatal closure (Roelfsema and Hedrich, 2005). The 84-node network of ABA induced closure can be found as **Figure 1** of R. Albert et al. (2017) and the reduced, 49-node version as Figure 1 in Maheshwari et al. (2019). In the version presented here in **Figure 1**, we incorporate further visual simplifications to ease understanding; we note that our analyses were done on the 49-node network of Maheshwari et al. (2019). The node names used in this paper are the abbreviated node names used in Maheshwari et al. (2019); the full names are indicated in **Supplementary Table S1**.

Multiple methods of analysis of the Boolean model have been used to understand the dynamics and outcomes of the system under different scenarios. One method is to specify the signal (usually, ABA) and the initial states of various components of the network and simulate the trajectory of the system over a given number of time-steps (~ 50) using the software library BooleanNet (Albert et al., 2008). The initial condition corresponds to the information available on open stomata (see **Supplementary Table S2**). As there is no biological information available from the experimental literature about the pre-stimulus state of multiple nodes [there are 17 such nodes in the 49-node model of Maheshwari et al. (2019)], these nodes were started in a randomly selected initial condition. At each time-step, the nodes of the system are updated (i.e., their states are recalculated) in a randomly selected order (Wang et al., 2012). This is an appropriate update scheme for the stomatal closure network since the timescales of the internal processes are largely unknown. During each simulation, each node in the network changes state one or multiple times and after sufficient time, the system settles down into an attractor. Generally, an attractor can be a fixed point (steady state) or a set of states that repeat indefinitely (a complex attractor, corresponding to a sustained oscillation). The Boolean model of ABA-induced closure yields a complex attractor in which most of the nodes (40 out of 49) have a stationary state. Due to the stochasticity introduced by the update method and by the initial node states, a large

number (~ 2000) of replicate simulations are run. The outcome of the model is summarized as the percentage of simulations in which the node Closure is in the state 1 (ON) at each time step, which we refer to as the percentage of closure. To characterize a whole time-course of closure in response to ABA or another signal, the cumulative percentage of closure (CPC) is defined as the sum of the percentage of closure over the course of the simulation (usually 50 time steps). The external supply or constitutive activation of a node is implemented as a sustained ON state, the knockout of a node as a sustained OFF state, and the time-course of a thus-perturbed system is compared to that of the wild-type system (see Section “Simulation of Node Constitutive Activation”).

A useful analytical tool for obtaining the system trajectories and attractors of the ABA network is to identify stable motifs, which are generalized positive feedback loops that maintain an associated state of their constituent nodes (see Section “Stable Motifs and Oscillating Motifs” and Zaňudo and Albert, 2013). For example, the positive feedback between PA, ROS and PLD δ determines a stable motif that ensures the sustained ON state of these nodes. Each stable motif can yield the stabilization of further nodes in the network, trapping the system’s state into a subset of the state space. One can follow an iterative procedure where the influence of each stable motif is used to reduce the network and one then finds the stable motifs in the reduced network. Some of these secondary stable motifs are dependent on certain conditions such as the prior stabilization of a node. For this reason, they are called conditionally stable motifs (see Section “Stable Motifs and Oscillating Motifs” and Deritei et al., 2019). The ABA-induced closure network also contains a negative feedback loop formed by the nodes Ca^{2+}_c and Ca^{2+} ATPase (see **Figure 1**), which under certain conditions (e.g., in the presence of ABA) induces the sustained oscillation of these nodes and of a few nodes regulated by them. This negative feedback loop is an example of a conditional oscillating motif (see Section “Stable Motifs and Oscillating Motifs”).

Starting from a signal and obtaining the consecutive (conditionally) stable motifs and (conditional) oscillating motifs gives us the motif succession diagram, which reflects the system’s trajectories and identifies the system’s attractors (see Section “Stable Motif Succession Diagrams of the Stomatal Closure Model Versions”). Another complementary approach is to use causal logic analysis to find long-range causal dependencies (i.e., sufficient or necessary subgraphs) or cyclic causal dependencies, which correspond to stable motifs (Maheshwari and Albert, 2017). This method was used to identify the logic backbone of the ABA-induced closure process. For example, ABA is sufficient for the activation of the node pH $_c$ (meaning an increase in the cytosolic pH level) and also sufficient for deactivation of the ATRAC1 protein (Maheshwari and Albert, 2017). Causal logic analysis can also be used to find the causal effect of a node perturbation and to identify certain interventions that can independently drive the network into a certain attractor. For example, causal logic indicates that the sustained ON state of the node ROS is a driver of a stable motif corresponding to stomatal

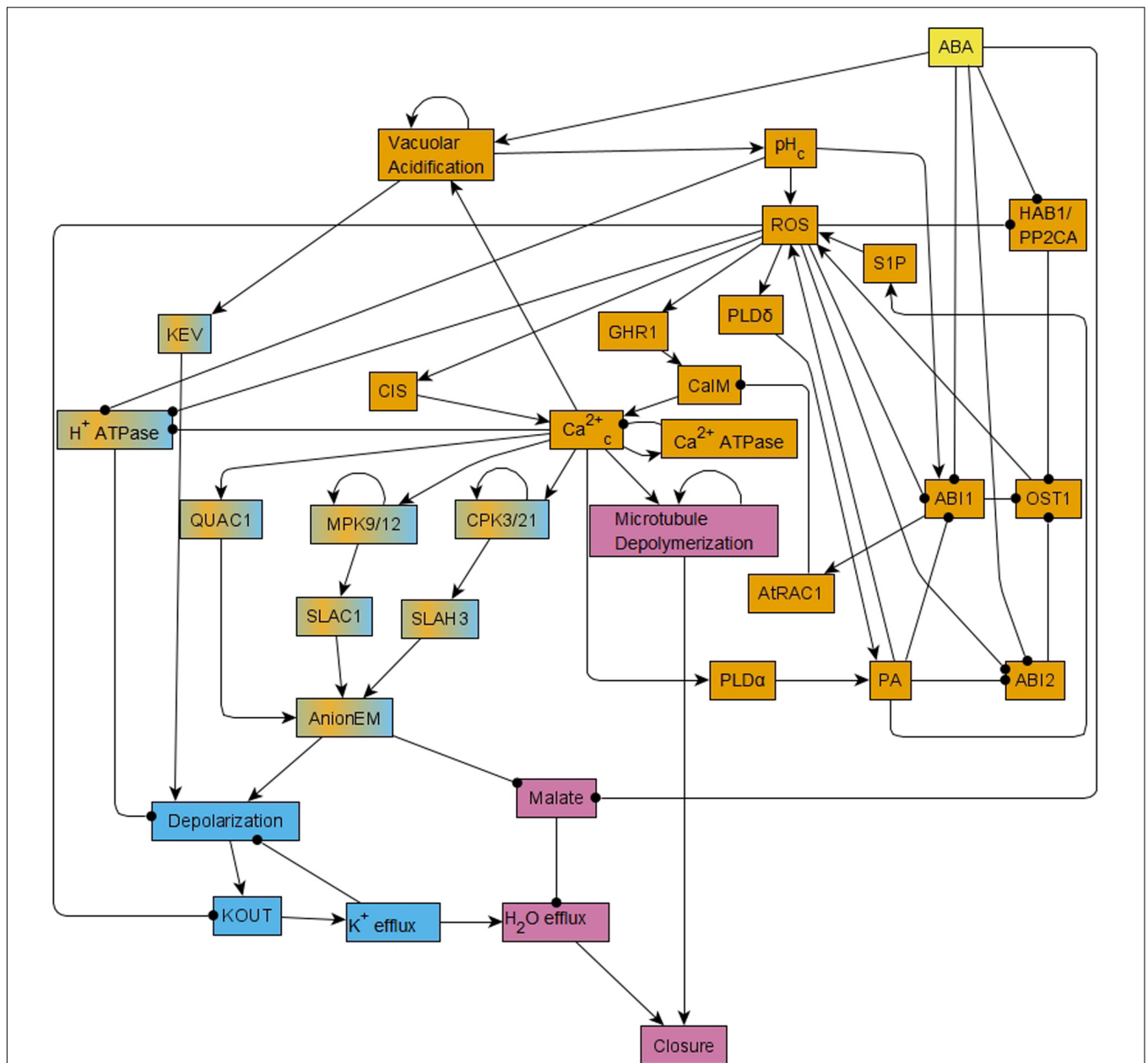


FIGURE 1 | Simplified version of the network that forms the basis of the Boolean model of ABA induced stomatal closure. This network is reduced from the 49-node network in Maheshwari et al. (2019) using methods of causal logic reduction (Maheshwari and Albert, 2017) and binary transitive reduction (Albert et al., 2007); it preserves all the relationships among nodes of the 49-node network via edges or paths. Each edge that terminates in an arrow indicates an activating relationship and each edge that terminates in a black circle indicates an inhibitory relationship. This reduced network is presented here just for ease in visualization; all the analysis in this work was conducted on the 49-node network presented in Figure 1 of Maheshwari et al. (2019). This network contains positive feedback loops, i.e., cycles of directed edges that contain no or an even number of inhibitory edges, for example $ROS \rightarrow ABI1 \rightarrow OST1 \rightarrow ROS$. It also contains negative feedback loops, i.e., cycles of directed edges that contain an odd number of inhibitory edges, for example $Depolarization \rightarrow KOUT \rightarrow K^+ \text{ efflux} \rightarrow H_2O \text{ efflux} \rightarrow Closure \rightarrow Depolarization$. This network has two strongly connected components (SCCs) i.e., subnetworks in which every pair of nodes is connected by at least two paths of opposite direction. These two SCCs are represented in orange and light blue colors. Nodes of the orange SCC can reach the nodes of the light blue SCC via paths; the nodes linking these two SCCs are shown in a mix of orange and light blue colors. The signal ABA reaches all the nodes of the network. The out-component of this network leads from the two SCCs to the Closure node and is represented in pink color.

closure in the absence of ABA (Maheshwari and Albert, 2017). Indeed, experimental results demonstrate that providing H_2O_2 leads to stomatal closure (Zhang et al., 2001; Kwak et al., 2003).

Overall, the results in Albert et al. (2017) illustrate a comprehensive process of model construction from causal relationships, model analysis through simulations and

network-based methods, model validation, and use of the model to make novel predictions. Most of the discrepancies between model results in Albert et al. (2017) and experimental data involve the model's inability to recapitulate closure in response to the constitutive activation of a node that is causally sufficient for Ca^{2+}_c oscillations. This observation inspired the prediction of a possible Ca^{2+} regulation of PP2C activity. Following a systematic analysis, Maheshwari et al. (2019) confirm, both *in silico* and experimentally, that Ca^{2+} directly or indirectly (via the activation of PA production) inhibits the PP2C ABSCISIC ACID INSENSITIVE 2 (ABI2). The thus-updated model resolves the previous discrepancies. The analysis in Maheshwari et al. (2019) also identifies the stable motif associated with stomatal closure in the absence of ABA, the drivers of this motif, and the causal relationships between various non-canonical closure signals. The basis of our present analysis is the two versions of the model reported in Maheshwari et al. (2019), which we will refer to as Model1 and Model2. The difference between the two models is that in Model1 PA is assumed to inhibit ABI2 while in Model2 Ca^{2+}_c is assumed to inhibit ABI2.

Methods

Boolean Update Functions

In a Boolean model the future state of a node is determined by the current state of its upstream regulators and is expressed as a Boolean update function. In the following we will use a simple notation convention for Boolean update functions: represent the state of each node by the name of the node and mark the future state with an asterisk. A node with a single regulator is characterized by one of two types of single variable Boolean functions: identity and negation. Identity is used for positive regulators. Denoting the regulator (and its state) by X and the target node by Y , the Boolean update function of Y is $Y^* = X$, meaning that the target Y is adopting the state of the regulator X as its future state. Negation, expressed using the Boolean operator “not,” is used for negative regulators. The Boolean update function of Y is $Y^* = \text{not } X$, meaning that the target Y adopts the opposite state as the regulator. Notice that in both cases the future state of Y depends only on the state of its regulator X .

For nodes that have more than one regulator, the “or” operator is used if a particular state of any of the regulators can independently activate the target; the “and” operator is used if a particular state of each of the regulators is needed for activation. The choice of the most appropriate operator to use is determined by experimental evidence. Our analysis uses the Boolean update functions constructed in Albert et al. (2017), and then simplified in Maheshwari et al. (2019). **Supplementary Table S3** indicates the simplified update functions of each node (as in Maheshwari et al., 2019) and experimental literature supporting it.

Including Memory in Boolean Update Functions

In most of the Boolean update functions used in models of biological systems the regulators of the target node do not include the node itself, which means that the future state of the node does not depend on its current state. In contrast to differential equation-based models, for which the change in a molecule's

concentration is described by explicit synthesis term(s) and degradation term(s) (Tyson and Novak, 2020), the degradation in Boolean models is implicit. For example, the update function $Y^* = X$ implies that the off state of X triggers the off state of Y regardless of whether Y was on before. In cases where there is a known mechanism for positive self-regulation of a target node and/or there is evidence that the target node's activity is longer-lived than the activity of its activators, it is justified to include the target node as its own self-activator. Applying this modification to the previous example yields $X^* = Y$ or X , which implies that X will stay in the on state after achieving it for the first time.

There were four such cases in the model of ABA-induced closure (Albert et al., 2017; Maheshwari et al., 2019), namely CPK3/21, MPK9/12, vacuolar acidification and the depolymerization of microtubules. All four of these nodes are activated by an increase in the cytosolic calcium level (Ca^{2+}_c). However, Ca^{2+}_c elevation cannot be maintained indefinitely because it would lead to precipitation of calcium-phosphates, which would be toxic to the cell. It is also well-documented that Ca^{2+}_c oscillates during the process of stomatal closure (Staxén et al., 1999). Yet, there is no evidence that these four nodes would have an oscillatory pattern, or decrease when the cytosolic Ca^{2+}_c level decreases, during ABA induced stomatal closure. On the contrary, their sustained activation is necessary to achieve sustained anion efflux and stomatal closure. Thus, in Albert et al. (2017) it was assumed that the activity of these four nodes decays very slowly or not at all. This assumption was supported by a specific mechanism in case of three of the four nodes (see **Supplementary Text S1**). The assumption was implemented by including the state of the node in its own update function, connected by the “or” operator to the state of Ca^{2+}_c . For example, the update function of CPK3/21 is $\text{CPK3/21}^* = \text{Ca}^{2+}_c$ or CPK3/21. According to this function, the next state of CPK3/21 equals 1 (ON) if Ca^{2+}_c is currently 1, or if the current state of CPK3/21 is 1. Once CPK3/21 turns on, it will stay on the whole duration of validity of the model, which represents the 30–60 min needed to achieve closure in response to ABA. We refer to this assumption as the assumed persistent activity of these four nodes.

Here, we study a longer process that consists of stomatal closure induced by ABA, then reopening following the removal of the signal. We therefore introduce an alternative implementation of slow decay of the activity of these four nodes: we include past states of Ca^{2+}_c in their update function. For example, $\text{CPK3/21}^* = \text{Ca}^{2+}_{c(t)}$ or $\text{Ca}^{2+}_{c(t-1)}$ or $\text{Ca}^{2+}_{c(t-2)}$ where t is the current time step. This function indicates that the next state of CPK3/21 equals 1 if Ca^{2+}_c is currently ON or has been ON in the last two time steps. We refer to this implementation as short-term memory. The duration of the short-term memory (in units of a time step) is an adjustable parameter.

Simulation of Node Constitutive Activation

The constitutive activation (CA) of a node is simulated by setting its state to ON and keeping it fixed to ON throughout the entire simulation – this can be thought of as equivalent

to providing a biomolecule in non-limiting quantity or, for an enzyme, providing the enzyme in non-limiting quantity and in the active state, during an experiment. These *in silico* mutations are categorized into different response categories (defined in Maheshwari et al., 2019) by their steady state percentage of closure and by their cumulative percentage of closure (CPC) values. In the simulations in the absence of ABA, we refer to the scenario where no node is kept constitutively active as the baseline. Significantly increased response is the case when the percentage of closure reaches 100%. Slightly increased response is the case when the final percentage of closure is higher than the baseline but less than 100%. Close to baseline is the case when the percentage of closure and CPC are within two standard deviations of the respective baseline values. As described in Section “Recapitulating the Open Stomatal State in the Absence of a Signal,” baseline percentages of closure can be non-zero.

Evaluation of Consistency Between Simulation and Experiment

We group the experimental response categories into two broad classes: close to baseline response, which indicates a stomatal aperture that is not statistically significantly different from wild type in the absence of a signal, and increased response (decreased aperture compared to wild type). Slightly or significantly increased response compared to baseline are considered consistent with experimental observation of increased response (decreased aperture compared to the wild type) in the absence of ABA. Close to baseline response is considered consistent with a stomatal aperture measured in the absence of ABA that is not statistically significantly different from or is greater than that of the wild type.

Stable Motifs and Oscillating Motifs

A stable motif of a Boolean dynamical system is a generalized feedback loop that maintains an associated state of its constituent nodes regardless of the state of nodes outside of the feedback loop. Stable motifs were first defined using the expanded network, a network that encodes the causal relationships between node states (as incorporated in the update functions). The expanded network consists of two virtual nodes for each node (one for each of the two possible states) and composite nodes that embody AND gates among two or more node states. A stable motif is a subgraph of an expanded network that satisfies four properties: (1) it is strongly connected (there is a path between every pair of nodes in the subgraph), (2) it is consistent (all represented states can be simultaneously satisfied), (3) it is composite-closed (if a composite node is in the subgraph, so too are all its virtual node parents), and (4) it is minimal (it contains no non-trivial subgraphs satisfying the first three properties). Stable motifs can also be defined using the causal logic formalism of Maheshwari and Albert (2017) i.e., a cyclic sufficient subgraph, necessary subgraph, or both sufficient and necessary subgraph determines a stable motif. After the nodes of a stable motif adopt the associated state, they will stay in that state. This may induce other nodes to adopt fixed states as well. Thus, a stable motif determines a region of the state space from which dynamical trajectories

cannot escape. In this so-called trap space, a subset of the nodes have a fixed state; for this reason this trap space is also referred to as a partial fixed point.

Another important class of expanded network subgraph is the oscillating motif. Like stable motifs, oscillating motifs are strongly connected, composite-closed subgraphs of the expanded network. Unlike stable motifs, however, oscillating motifs violate the consistency criterion in that every virtual node in the subgraph has its negation in the subgraph as well. Intuitively, oscillating motifs arise from negative feedback loops, while stable motifs arise from positive feedback loops. An oscillating motif will likely (but not always) correspond to a complex attractor in which the nodes represented in the oscillating motif oscillate.

A weaker version of stable motifs was defined by Deritei et al. (2019) and named conditionally stable motif. A conditionally stable motif is a consistent, strongly connected component of the expanded network that is not composite closed. The virtual nodes that are parent nodes of composite nodes included in the conditionally stable motif serve as the conditions. In other words, the conditionally stable motif is a generalized positive feedback loop that maintains an associated state as long as one or more nodes outside of the feedback loop are in a specific state. As long as these parent nodes are in that specific state, the conditionally stable motif behaves like a stable motif. Conditional oscillating motifs can be defined analogously as strongly connected, consistency-violating, not composite-closed subgraphs of the expanded network.

Motif Succession Diagram Analysis

In general, the stable motif succession diagram is determined by following these steps. First, the system's stable motifs are determined. For every stable motif, the corresponding node states are substituted into the Boolean regulatory functions and the network is iteratively reduced. Then, the stable motifs of the reduced network are determined; these may be stable motifs or conditionally stable motifs of the original network. This process is repeated until the network cannot be reduced anymore. Nodes remaining in the network may oscillate in the attractor(s), while the nodes that were reduced due to the influence of a stable motif will take the associated fixed values. The stable motif succession diagram encapsulates all successions of stable motifs found by this process. For more details, see Zañudo and Albert (2013), where a Java implementation is also provided.

In this paper, we perform motif succession diagram analysis for two assumptions regarding the nodes CPK3/21, MPK9/12, Microtubule Depolymerization and Vacuolar Acidification. The first assumption is persistence of these four nodes, as described in Section “Including Memory in Boolean Update Functions.” The second assumption is short term memory for these four nodes; their corresponding motif succession diagrams are detailed in Section “Stable Motif Succession Diagram Analysis of the Network with Short-term Memory.” We generate the stable motif succession diagrams for these networks primarily using the Stable Motif code provided in Zañudo and Albert (2013) and supplementing those results by theoretical analysis. The different signal conditions are set by setting the state of the signal node (e.g., ABA or Ca^{2+}) to ON or OFF and reducing the network

with that setting. We augmented these diagrams with our analysis of each conditional oscillatory motif, including identification of its condition and of the nodes whose state stabilizes due to the establishment of this motif. We use the causal logic framework from Maheshwari and Albert (2017) to aid in this analysis.

RESULTS

Stable Motif Succession Diagrams of the Stomatal Closure Model Versions

As described in Section “Including Memory in Boolean Update Functions” and **Supplementary Text S1**, the original model of ABA-induced closure (Albert et al., 2017; Maheshwari et al., 2019) includes the assumption that the activity of four nodes regulated by the cytosolic calcium level (Ca^{2+}_c), namely CPK3/21, MPK9/12, Vacuolar Acidification and Microtubule Depolymerization, decays very slowly or not at all. The assumption was implemented by introducing their current state in the update function of these nodes. For example, the update function of CPK3/21 is $\text{CPK3/21}^* = \text{Ca}^{2+}_c$ or CPK3/21. The assumed persistence of the activity of CPK3/21 constitutes a stable motif: once activated, CPK3/21 maintains its ON state regardless of the rest of the nodes. An analogous stable motif also exists for each of the other three nodes. These stable motifs appear as self-loops on the four nodes in **Figure 1**. We determined the succession diagrams of stable and oscillating motifs in three cases: absence of ABA, presence of ABA, and externally provided Ca^{2+} . In this section, we present examples of stable motif succession diagrams of Model1 and describe the changes observed in Model2.

ABA OFF Case

As reported in Maheshwari et al. (2019), in the absence of ABA there are 17 attractors. Of these attractors, only one, denoted A0, corresponds to closed stomata. The attractors A1 to A16 are highly similar and are consistent with open stomata (see **Supplementary Table S4**). These attractors differ in the state of a few nodes that can be stabilized in either the ON or OFF state or can oscillate, while still corresponding to the biologically known information on open stomata.

The network in the ABA = 0 case has 6 stable motifs, 4 of which correspond to the persistent ON states of CPK3/21, MPK9/12, Microtubule Depolymerization, and Vacuolar Acidification. The remaining 2 stable motifs, which we denote *openM1* and *openM2*, represent the simultaneous OFF state of more than 10 nodes (see **Figures 2A,B**).

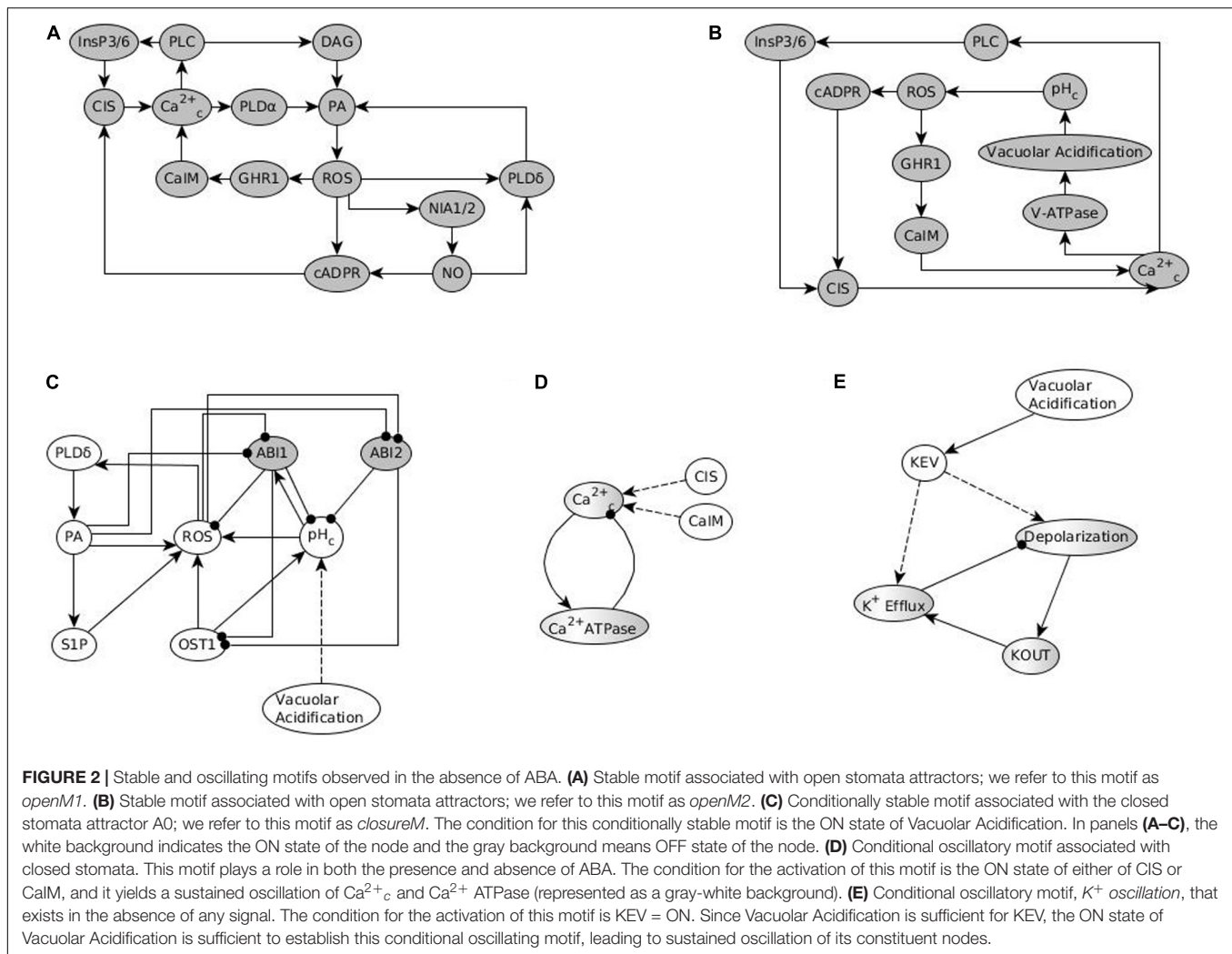
There also are 5 conditionally stable motifs (CSMs). Four of these represent the persistent OFF state of each of the four self-regulating nodes. The OFF state of CPK3/21, MPK9/12, and Microtubule Depolymerization are conditioned on either *openM1* or *openM2*. The OFF state of Vacuolar Acidification is conditioned on *openM1* (it is part of *openM2*). The fifth CSM, shown in **Figure 1C**, expresses the self-sustained activity of PLD δ , PA, ROS, S1P, OST1, and pH_c together with the sustained inactivity of ABI1 and ABI2. This CSM is dependent on the condition that the Vacuolar Acidification = 1 stable motif

is activated. There are two conditional oscillating motifs. The first is made up by the nodes Ca^{2+}_c and Ca^{2+} ATPase, both of which oscillate as long as CIS = 1 or CaIM = 1. The second motif comprises the oscillations of the K^+ efflux, KOUT and Depolarization nodes, conditioned on KEV = 1. We refer to this latter motif as *K⁺ oscillation*. These motifs are presented in **Figure 2**.

The stabilization of the CSM shown in **Figure 1C** ensures the establishment of sustained Ca^{2+}_c oscillations. Indeed, if either of the positive regulators of Ca^{2+}_c is stabilized in its active state, the negative feedback between Ca^{2+}_c and Ca^{2+} ATPase makes both of them oscillate (as characterized in detail in **Supplementary Text S2** of Maheshwari et al., 2019). The CSM shown in **Figure 1C** includes the sustained ON state of the node ROS, which is sufficient for the node CIS. The CSM also includes the sustained OFF state of ABI2, which in combination with the ON state of ROS is sufficient for the node CaIM. Thus, the stabilization of the CSM leads to Ca^{2+}_c – Ca^{2+} ATPase oscillations. As a result of these oscillations, most nodes directly downstream of Ca^{2+}_c , i.e., PLC, PLD δ , QUAC1, TCTP, V-ATPase, DAG, and InsP3/6, also oscillate. Their oscillation periods are given in **Supplementary Table S4** and explained in Maheshwari et al. (2019). The remaining two nodes directly regulated by Ca^{2+}_c , namely CPK3/21 and MPK9/12, persist in their ON state after first turning on. The oscillations of TCTP stabilize the Microtubule Depolymerization = 1 motif. Altogether, the stabilization of the CSM leads indirectly to the sustained ON state of Microtubule Depolymerization (via sustained Ca^{2+}_c oscillation), and to the sustained ON state of H_2O Efflux (via the sustained ON state of multiple nodes downstream of the CSM). The activation of H_2O Efflux and Microtubule Depolymerization, in turn, leads to stomatal closure. Since this CSM drives the network into an attractor corresponding to closed stomata (see **Supplementary Table S4**), we will refer to it as *closureM*.

In conclusion, any system trajectory that (i) involves the stabilization of the stable motifs Vacuolar Acidification = ON, CPK3/21 = ON, MPK9/12 = ON, Microtubule Depolymerization = ON, *closureM*, and the Ca^{2+}_c – Ca^{2+} ATPase conditional oscillating motif, and (ii) satisfies the restrictions that Vacuolar Acidification = ON establishes before *closureM* and *closureM* stabilizes before the Ca^{2+}_c – Ca^{2+} ATPase motif, reaches the closure attractor A0. **Figure 3A** shows two of the possible trajectories to the closed stomata attractor A0 when the first motif to stabilize is Vacuolar Acidification = ON, the second motif to stabilize is MPK9/12 = ON and the third motif to stabilize is CPK3/21 = ON. **Figure 3A** also illustrates that if the Vacuolar Acidification stable motif is already locked in then the system's trajectories may bifurcate. In one branch the *closure* motif stabilizes, leading to the closure attractor A0. In the other branch the *openM1* stable motif stabilizes, which leads to the open stomata attractor A1.

Figure 3B denotes two possible motif successions that contain the stable motif *openM1* and lead to an attractor that corresponds to open stomata. In both successions, the first motif to stabilize is Microtubule Depolymerization = ON, the second motif to stabilize is MPK9/12 = ON, and the third motif is CPK3/21. *openM1* is the condition for the CSMs



Vacuolar Acidification = OFF, CPK3/21 = OFF, MPK9/12 = OFF, and Microtubule Depolymerization = OFF. Hence, a possible system trajectory involving *openM1* can have both the OFF state and ON state of the four self-regulating nodes after *openM1* is established, while it can have only their ON state before *openM1*. Any trajectory that contains *openM1* and the Vacuolar Acidification = ON motif leads to the establishment of the conditional oscillating motif K^+ oscillation. These cases give rise to open stomata attractors that have the K^+ efflux, KOUT and Depolarization nodes oscillating. Trajectories that contain *openM1* and the Vacuolar Acidification = OFF motif lead to open stomata attractors in which K^+ efflux, KOUT and Depolarization are off. As a result, this case has many possible trajectories and hence leads to a complex succession diagram, one branch of which is displayed in Figure 3B.

Figure 3C denotes a possible succession diagram leading to an open stomata attractor that involve the stabilization of *openM2*. The Vacuolar Acidification = ON motif is mutually exclusive with *openM2*. Similar to the case of *openM1*, the OFF states of CPK3/21, MPK9/12 and Microtubule Depolymerization are CSMs with *openM2* as the condition. Hence, a possible system

trajectory involving *openM2* can have either the ON or OFF state of the three self-regulating nodes after *openM2* but it can have only their ON state before *openM2*. Figure 3C shows the case when the first motif to stabilize is MPK9/12 = ON and the second is CPK3/21.

ABA ON Case

In the presence of ABA, the network has four stable motifs. Three of these express the self-sustained activity of CPK3/21, MPK9/12, and Microtubule Depolymerization, respectively; these were stable motifs in the absence of ABA as well. The self-loop of Vacuolar Acidification does not appear as a stable motif because the ON state of Vacuolar Acidification is indirectly determined by ABA and does not need self-stabilization. The fourth motif is a cycle that sustains the ON state of PLDδ, PA, and ROS. This cycle is a subset of the *closureM* motif; the rest of the nodes of this motif are stabilized by ABA. There is also an oscillatory motif comprised of Ca^{2+}_c and Ca^{2+} ATPase. It is the same as the conditional oscillating motif in the ABA = OFF case, however, in the ABA = ON case, it is not conditional anymore, because the sustained presence of ABA leads to the ON state

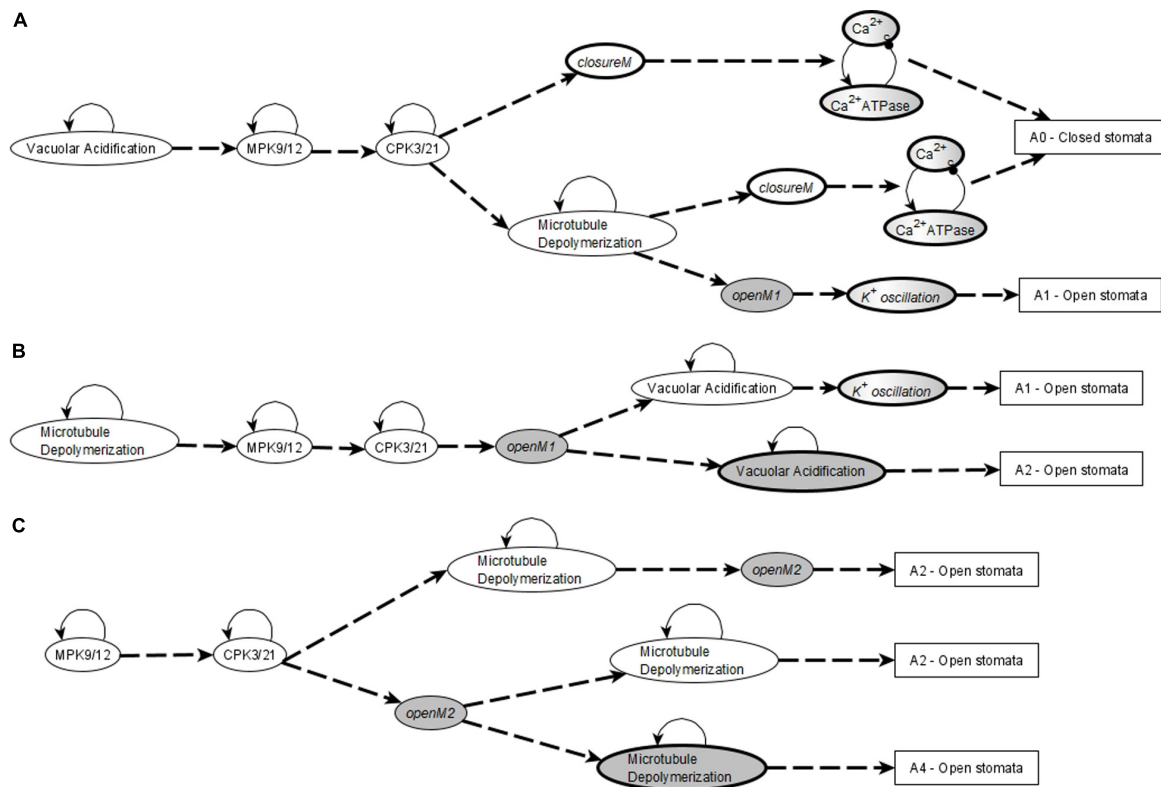


FIGURE 3 | Motif succession diagrams of Model1 in the absence of ABA and any other closure signal. Stable motifs are shown with oval symbols and attractors are indicated by rectangles. Each dashed directed line between two motifs indicates that the system states in which the first motif has established and any nodes driven by it have stabilized admit the second motif as next to activate. The dashed directed line converging into an attractor symbol indicates that the succession of stable motifs ensures the system's convergence into the respective attractor. **(A)** shows a subset of the succession diagram that converges to the attractor A0, corresponding to closure of the stomata in the absence of ABA (see **Supplementary Table S4**). The stable motif characteristic to this attractor is the conditionally stable motif *closureM*. **(B,C)** show a subset of the succession diagram corresponding to a sample of the 16 attractors that describe open stomata in the absence of ABA (see **Supplementary Table S4**). Each possible trajectory in this case contains exactly one of the stable motifs *openM1* and *openM2*. **(B)** describes some of the trajectories containing *openM1* while C describes some of the trajectories containing *openM2*. **(A)** also indicates the existence of bifurcations in the system's trajectory due to the mutually exclusive motifs *closureM* and *openM1*. If *closureM* stabilizes the system converges into the attractor A0, and if *openM1* stabilizes the system converges into attractor A1. The diagram encodes node states into the background color of the stable motif symbols. When referring to single nodes, white background indicates the ON state of the node, gray background means the OFF state of the node, and gray-white background represents oscillating nodes. Since in the *openM1* and *openM2* stable motifs all nodes are OFF (see **Figure 2**), we use a gray background color for these stable motifs. We use white background to represent the locking in of the *closureM* motif and represent the oscillating nature of the K^+ oscillation motif by a gray-white background. The conditionally stable motifs are marked by thick boundaries.

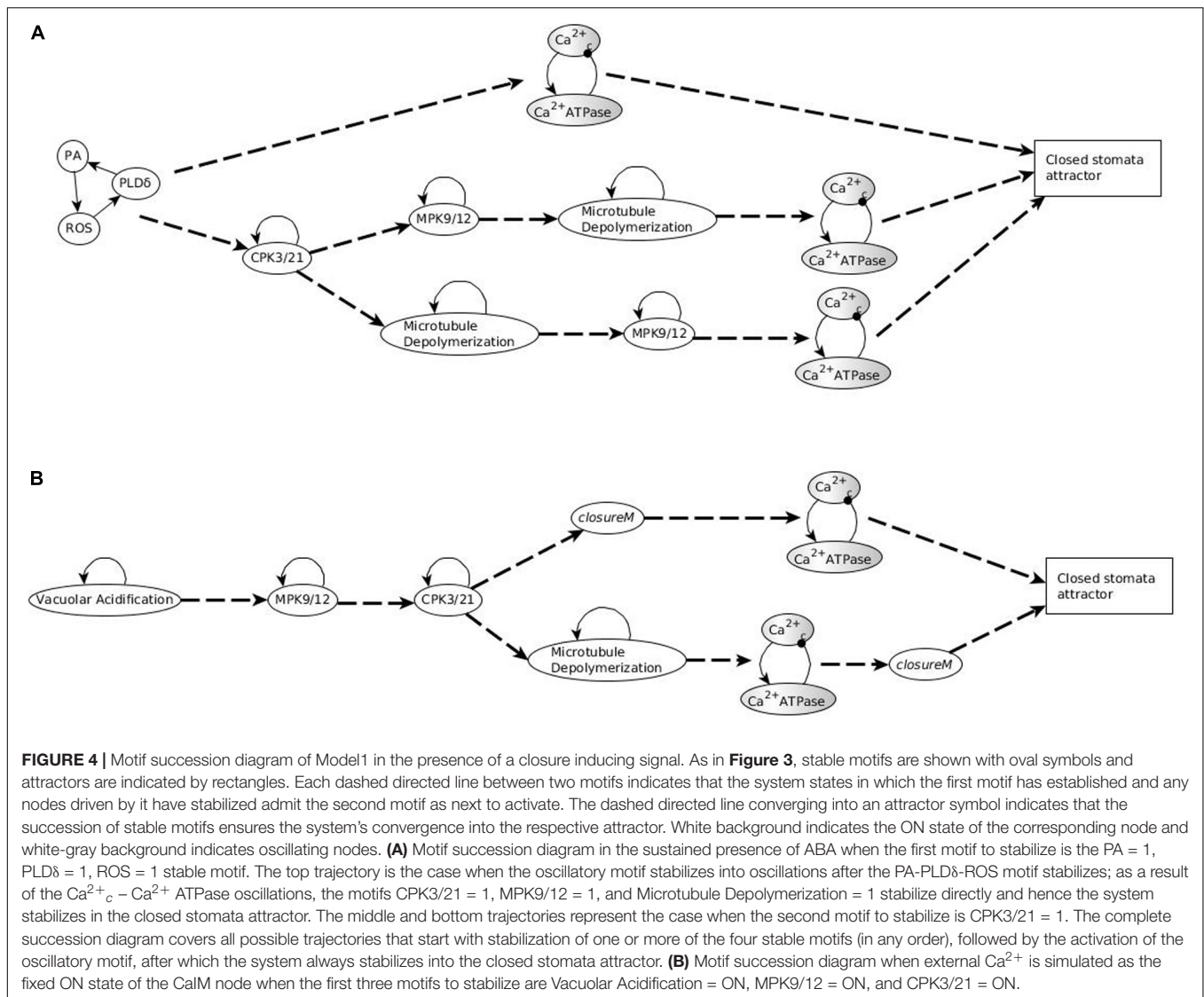
of CaIM, which fulfills the condition for $Ca^{2+}_c - Ca^{2+}$ ATPase oscillations. The stabilization of these four motifs, in any order, leads to the sole attractor reachable in this case. As expected, and consistent with the knowledge that ABA is sufficient for stomatal closure, this attractor corresponds to closed stomata. It differs from the closure attractor A0 reached in the absence of ABA in the state of five nodes whose state is determined solely by ABA: RCARs, V-PPase, Actin reorganization (which are ON in the presence of ABA and OFF in its absence) and PEPC, ATRAC1 (which are OFF in the presence of ABA and ON in its absence). A subset of the motif succession diagram for this case is presented in **Figure 4A**.

Externally Provided Ca^{2+} as Signal

In the absence of ABA, external Ca^{2+} can be simulated either as the fixed ON state of the Ca^{2+}_c node or as the fixed ON state

of the CaIM node, which represents calcium influx through the membrane. In the case of fixed ON state of the Ca^{2+}_c node the stable motifs in the network corresponding to closed stomata (i.e., the four self-regulating nodes and the *closureM* motif) are quickly stabilized, because Ca^{2+}_c is a driver node for all of them. Thus, any initial condition leads to an attractor corresponding to closed stomata. This attractor is slightly different from the attractor corresponding to closure induced by ABA in that there are no oscillating nodes (see **Supplementary Table S4**). In fact, the effect of fixed ON state of Ca^{2+}_c is slightly stronger than the effect of the sustained presence of ABA, which leads to sustained Ca^{2+}_c oscillations. This is because the fixed ON state of Ca^{2+}_c drives the network into a closed stomata attractor in fewer time steps than the sustained presence of ABA.

When the external Ca^{2+} is simulated as the fixed ON state of the CaIM node, the network has all of the motifs



associated with the closed stomata attractor in the absence of ABA, i.e., the four stable motifs corresponding to persistent activity, the *ClosureM* motif and the oscillatory motif containing Ca $^{2+}_c$ and Ca $^{2+}$ ATPase (whose condition is now satisfied). The succession diagram of this case, shown in **Figure 4B**, is very similar to the one shown in **Figure 3A** except that the Ca $^{2+}_c$ - Ca $^{2+}$ ATPase oscillating motif is not a conditional motif anymore and can hence form a trajectory by appearing in any order with the four stable motifs. The *closureM* motif is still a conditionally stable motif. **Figures 3B,C** are not possible in this case since the ON state of CalM is incompatible with the *openM1* and *openM2* motifs. In other words, the only attractor possible in this case is the closed stomata attractor A0.

The stable motifs and the succession diagram of Model2, i.e., the network model in which Ca $^{2+}_c$ directly inhibits ABI2, are the same as those of Model1 in the presence of ABA and in the presence of external calcium. There are some small differences

between the models in the absence of ABA, which are described in **Supplementary Text S2**.

Relaxation to Resting State After Removing the Signal

According to our analysis of the model in Maheshwari et al. (2019), stomatal closure involves the stabilization of four stable motifs, the *ClosureM* conditionally stable motif (whose condition is the Vacuolar Acidification stable motif) and a Ca $^{2+}_c$ - Ca $^{2+}$ ATPase conditional oscillating motif (whose condition is satisfied by the *ClosureM* motif, see Section "Stable Motif Succession Diagrams of the Stomatal Closure Model Versions"). The four stable motifs stay in their associated state once stabilized. Even if the signal is taken away in the model, these stable motifs still stay stabilized, and so do the conditional motifs (since their condition is still satisfied), keeping the simulated stomata closed. **Figure 5** shows a simulation of Model1,

indicating that removal of the signal (ABA) does not lead to re-opening of the simulated stomata. This is not accurate since the biological reality is the reopening of the stomata after ABA is removed (Cummins et al., 1971). Here, we resolve this discrepancy by making the activity of certain nodes less persistent than assumed in Albert et al. (2017) and Maheshwari et al. (2019).

As described in Section “Including Memory in Boolean Update Functions,” the model currently assumes persistent activity of CPK3/21, MPK9/12, Microtubule Depolymerization and Vacuolar Acidification, which form stable motifs (see **Figure 3**). These stable motifs make the closure attractor irreversible, which is not biologically accurate. Hence, we considered the possibility of decreasing the persistence of these nodes, i.e., we assumed that the state of the target node is sustained only for a few time steps after its regulator (e.g., Ca^{2+}_c) turns off, after which the target node goes back to the OFF state. We implement this short-term memory by storing the state of the regulator in the past few timesteps (see **Figure 6**). Specifically, we use auxiliary nodes to remember the past states of the regulator (e.g., Ca^{2+}_c); these auxiliary nodes are always updated before the rest of the nodes are updated.

We use the ranked asynchronous update of the BooleanNet software library, in which the nodes are classified as rank 1, 2, ... and during each timestep, nodes of rank 1 are updated first (according to a random order among these nodes), followed by nodes of rank 2, and so on. We designate the auxiliary nodes corresponding to the largest memory as rank 1, the auxiliary nodes with next largest memory as rank 2, and so on. The regular nodes of the network have the lowest rank (highest numerical value). For example, when Microtubule Depolymerization considers the last five timesteps of TCPT, CPK3/21 and MPK 9/12 consider the last three timesteps of Ca^{2+}_c and Vacuolar Acidification considers the last two timesteps of V-ATPase (as in **Figure 7**), the auxiliary nodes TCTP_5, Ca_3 and V-ATPase_2 have rank 1, the auxiliary nodes TCTP_4, Ca_2 and V-ATPase_1 have rank 2, the auxiliary nodes TCTP_3 and Ca_1 have rank 3, the auxiliary node TCTP_4 has rank 4, the auxiliary node TCTP_5 has rank 5 and all the regular nodes have rank 6. The Boolean update functions for this case are given in **Supplementary Text S3**.

We implemented short-term memory effect for the four nodes and explored various durations for each of them to identify cases which show efficient closure in response to ABA and efficient re-opening of the stomata after ABA is removed. The analysis of these durations, described in **Supplementary Text S4**, indicates that a memory of three timesteps ensures the persistent ON state of CPK3/21 and MPK9/12, while the persistence of Microtubule Depolymerization or Vacuolar Acidification would be ensured for a memory of six timesteps. We found that the best result is exhibited by the model version where Vacuolar Acidification considers the last 2 time-steps of V-ATPase, CPK3/21 and MPK9/12 each consider the last 3 time-steps of Ca^{2+}_c , and Microtubule Depolymerization considers the last 5 time-steps of TCTP. **Figure 7** shows the re-opening of stomata after the removal of ABA with these parameters. The assumed memory effect is weakest for the node Vacuolar Acidification; for two timesteps memory the probability of the ON state of this node

is 85% (see **Supplementary Text S4**). This node is required to be in the ON state for the stabilization of the *closureM* motif and the turning off of this node contributes to reopening of the stomata. The larger memory duration of three timesteps ensures the persistence of CPK3/21 and MPK9/12 activity. The five timestep memory of Microtubule Depolymerization yields a 97% probability of its ON state. Overall, these memory durations ensure close to 100% closure in response to ABA, followed by reopening after the loss of the ABA signal.

Stable Motif Succession Diagram Analysis of the Network With Short-Term Memory

In the updated model with short term memory, the nodes CPK3/21, MPK9/12, Microtubule Depolymerization and Vacuolar Acidification do not form stable motifs anymore. Indeed, their future state is strictly independent of their own current state and is solely determined by their regulators' current and past states. Since these nodes no longer form motifs, the stable motif succession diagram is modified. However, the short-term memory effect needs to be strong enough to ensure that in the presence of a signal (for example, ABA or external calcium) there is a system trajectory in the succession diagram that leads to the closed stomata attractor. This is possible as ABA or external calcium causes Ca^{2+}_c oscillations, which are sufficient to maintain these four nodes in their ON state if the memory duration is sufficiently large (see **Supplementary Text S4** and **Figure 8**).

ABA OFF Case

In the absence of ABA, similar to the case described in Section “Stable Motif Succession Diagrams of the Stomatal Closure Model Versions,” Model1 has two stable motifs, *openM1* and *openM2*; these motifs are given in **Figure 2**. The stabilization of either of these two stable motifs causes a fixed OFF state of Ca^{2+}_c , which in turn leads to a fixed OFF state of CPK3/21, MPK9/12, Microtubule Depolymerization and Vacuolar Acidification. This leads to an open stomata attractor (see **Supplementary Table S4**). The presence of only two motifs leads to the possibility of just two trajectories, depicted in **Figure 8A**.

ABA ON Case

In the presence of ABA, Model1 has two motifs, one of which is a stable motif comprised of $\text{PA} = 1$, $\text{PLD}\delta = 1$ and, $\text{ROS} = 1$, and other is the oscillating motif consisting of Ca^{2+}_c and Ca^{2+} ATPase. This oscillating motif is the same as the one described in the ABA = ON case in Section “Stable Motif Succession Diagrams of the Stomatal Closure Model Versions,” and as in that case the condition of this motif is satisfied since ABA indirectly activates CalM. The sustained oscillations of Ca^{2+}_c are sufficient to fix the ON state of CPK3/21 and MPK9/12 if their memory is three or more time steps; they will fix Microtubule Depolymerization and Vacuolar Acidification if their memory is six or more time steps (see **Supplementary Text S4**). Stabilization of these two motifs (in either order) leads to the closed stomata attractor when the memory duration is large enough (see **Figure 8B**). If the memory duration is smaller, the system instead converges into

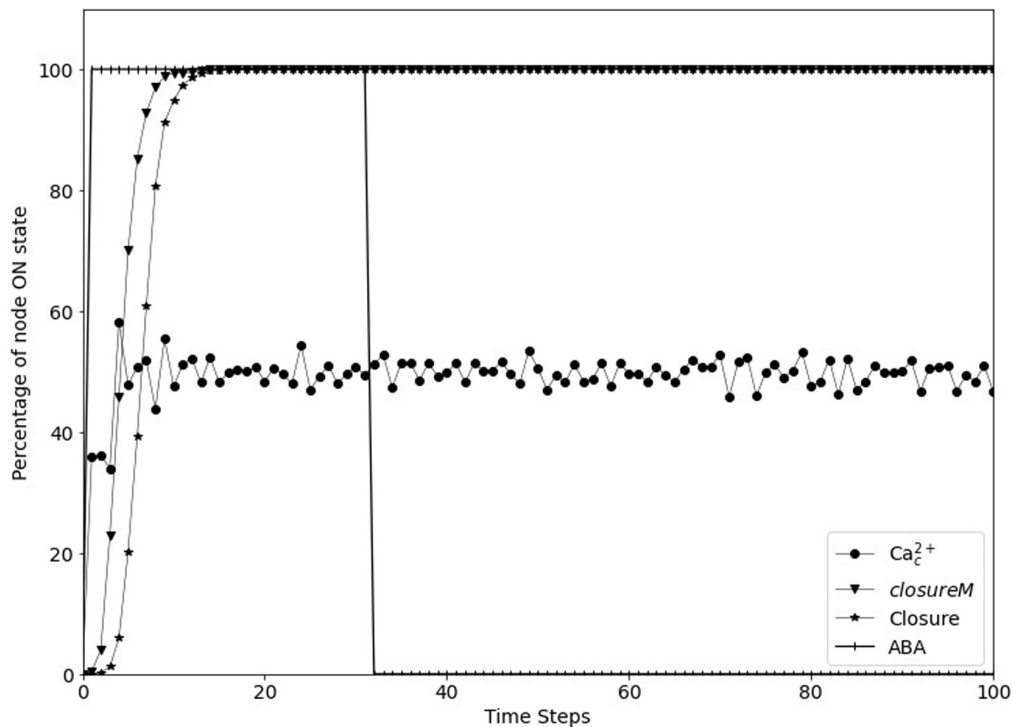


FIGURE 5 | Simulated stomatal closure is maintained after removal of ABA. The state of the node ABA (vertical line symbols) is ON for 30 time-steps and is then OFF. The *closureM* conditionally stable motif (downward pointing triangles) establishes in less than 15 timesteps and remains stable despite the loss of ABA. The percentage of closure (star symbols) increases to 100% and stays at this value even after the signal is removed. The circles represent the percentage of the ON state of Ca^{2+}_c which after a fast increase fluctuates around 50% since Ca^{2+}_c oscillates with approximately equal ON and OFF time periods. The biological expectation is that after the signal is removed the states of ROS, Closure, and Ca^{2+}_c should go to OFF eventually.

an attractor where many nodes, including Closure, oscillate (see **Supplementary Table S4**). After ABA is removed, the condition for the conditional oscillating motif of Ca^{2+}_c and Ca^{2+} ATPase is no longer satisfied. Hence, the oscillating motif ceases to exist and the Ca^{2+}_c oscillations decay, which leads to the fixed OFF state of CPK3/21, MPK9/12, Microtubule Depolymerization and Vacuolar Acidification. This in turn leads to an open stomata attractor (see **Supplementary Table S4**).

Externally Provided Ca^{2+} as the Signal

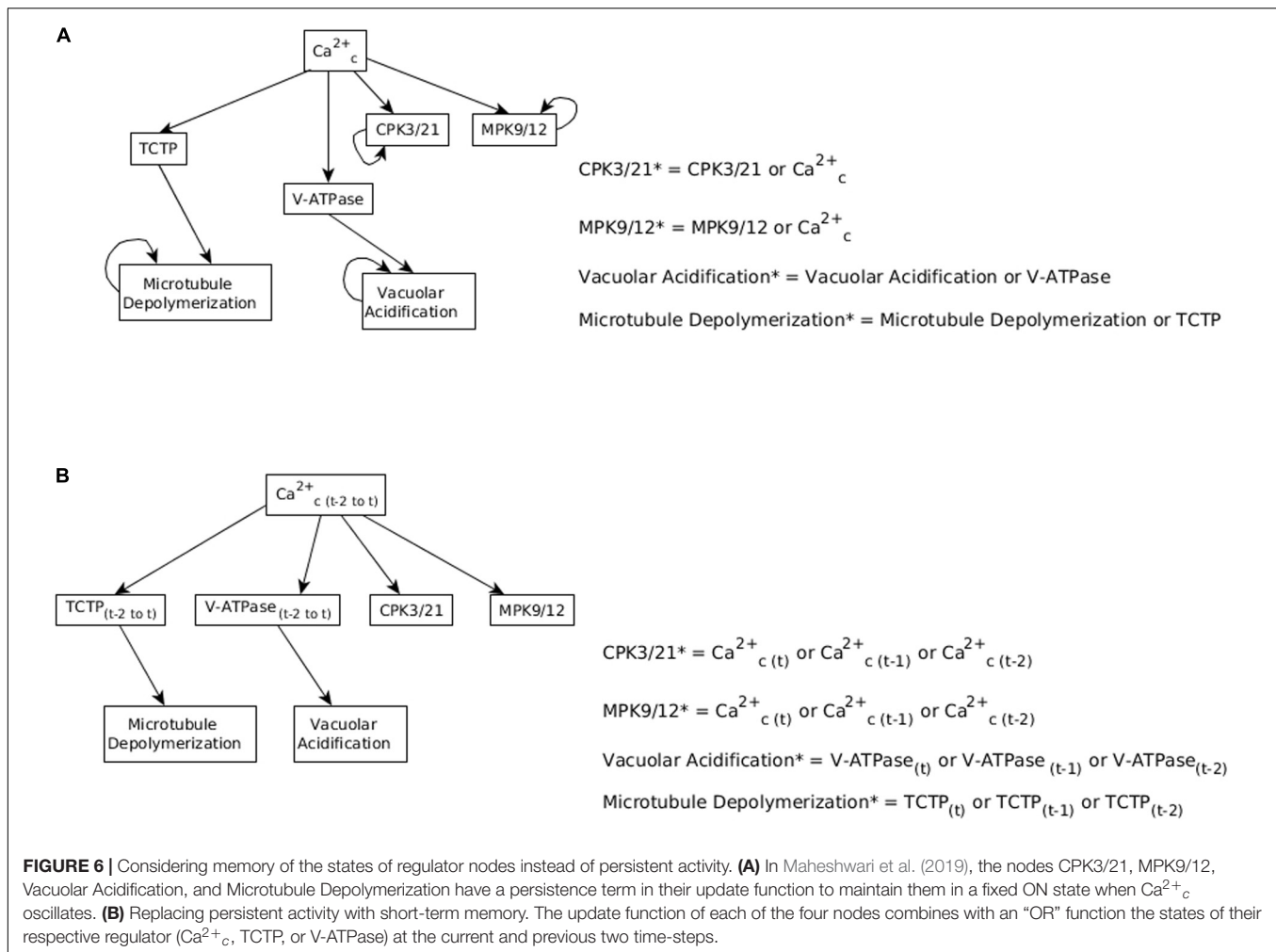
Similar to Section “Stable Motif Succession Diagrams of the Stomatal Closure Model Versions,” we considered two methods of simulating external Ca^{2+} . With Ca^{2+}_c fixed in the ON state, the states of CPK3/21, MPK9/12, Microtubule Depolymerization and Vacuolar Acidification are fixed in the ON state too. This leads to activation of the *closureM* motif, as described in Section “Relaxation to Resting State After Removing the Signal,” and hence the network stabilizes in a closed stomata attractor. When external Ca^{2+} is simulated as fixed ON state of CaIM, the Ca^{2+}_c – Ca^{2+} ATPase oscillating motif establishes. Indeed, experiments confirm that high external Ca^{2+} leads to sustained oscillations in Ca^{2+}_c (Jeon et al., 2019). The sustained oscillations of Ca^{2+}_c in the model lead to sustained ON state of CPK3/21, MPK9/12 if their memory is three or more time steps, and to sustained ON state of Microtubule Depolymerization and

Vacuolar Acidification if their memory is six or more time steps. The sustained ON state of Vacuolar Acidification leads to establishment of the *closureM* motif. The percentage of simulations in which the *closureM* motif stabilizes is always less than the percentage of ON state of Vacuolar Acidification and it increases as the memory duration is increased – see **Supplementary Figure S7**. Once the *closureM* motif stabilizes, it leads the system to the closed stomata attractor.

The stable motifs and the succession diagram of Model2 with the short-term memory effect are the same as those of Model1 in the presence of ABA or in the presence of external calcium. Similar to the case described in Section “Stable Motif Succession Diagrams of the Stomatal Closure Model Versions,” there are some small differences between the models in the absence of ABA, which are described in **Supplementary Text S2**.

Recapitulating the Open Stomatal State in the Absence of a Signal

The model in Maheshwari et al. (2019) yields ~20% closed stomata in simulated wild type guard cells that did not receive any closure signal. We refer to the cumulative percentage of closure (CPC) obtained in the simulation of unstimulated wild type cells as baseline. The biological expectation is that the baseline percentage of closure should be 0. We determined the baseline percentage of closure in the updated Model1 that has short term



memory instead of persistent activity of CPK3/21, MPK9/12, Microtubule Depolymerization, and Vacuolar Acidification. We found that the percentage of closure approaches 0 after a long time; however, there is a non-zero transient level of closure (see **Figure 9**), and thus a non-zero baseline CPC. While this is a significant improvement compared to the final percentage of ~20% reported in Maheshwari et al. (2019) and further supports the replacement of persistent activity with short-term memory, this is still not an accurate recapitulation of the biological expectation. Hence, in this section, we explore ways to achieve zero cumulative percentage of closure to better recapitulate the biological expectation.

In Maheshwari et al. (2019), seventeen nodes of the network were initialized randomly because of the lack of experimental evidence regarding their state in open stomata. These nodes are indicated in **Supplementary Table S2**. We reported in Maheshwari et al. (2019) that the non-zero baseline percentage of stomatal closure can be attributed to the initial condition. When certain nodes are initialized randomly and by chance their initial state is the same as their state in the closed stomata attractor, then the initial state essentially acts as a transient (single timestep) closure signal, which may in certain cases lead to the activation

of the *closureM* motif. As we see in **Figure 7**, after a closure signal is removed, the percentage of closure, activation of the *closureM* motif, and Ca^{2+}_c oscillations reduce over time. We observe a similar behavior in **Figure 9** where, as a result of the initial condition the percentage of closure, motif stabilization and Ca^{2+}_c oscillations initially increase, and then reduce to less than 5% after 100 timesteps. The peak of transient percentage of closure, motif stabilization or Ca^{2+}_c oscillations is much lower in **Figure 9** than **Figure 7** because the probability of activation of the *ClosureM* motif due to the initial condition is much lower than the probability of activation due to an ABA signal sustained for 30 timesteps. Hence, we hypothesized that the model would yield zero cumulative percentage of closure (CPC) if all of these seventeen nodes were initiated in the state opposite to their state corresponding to closed stomata. This “furthest from closure” initial condition indeed resulted in zero CPC in both Model1 and Model 2 (that is, both when PA inhibits ABI2 and when Ca^{2+}_c is assumed to directly inhibit ABI2).

The “furthest from closure” initial condition assumes a particular state for each of the 17 nodes that were initialized randomly (Maheshwari et al., 2019). Since the pre-stimulus state of none of these nodes was measured experimentally, the validity

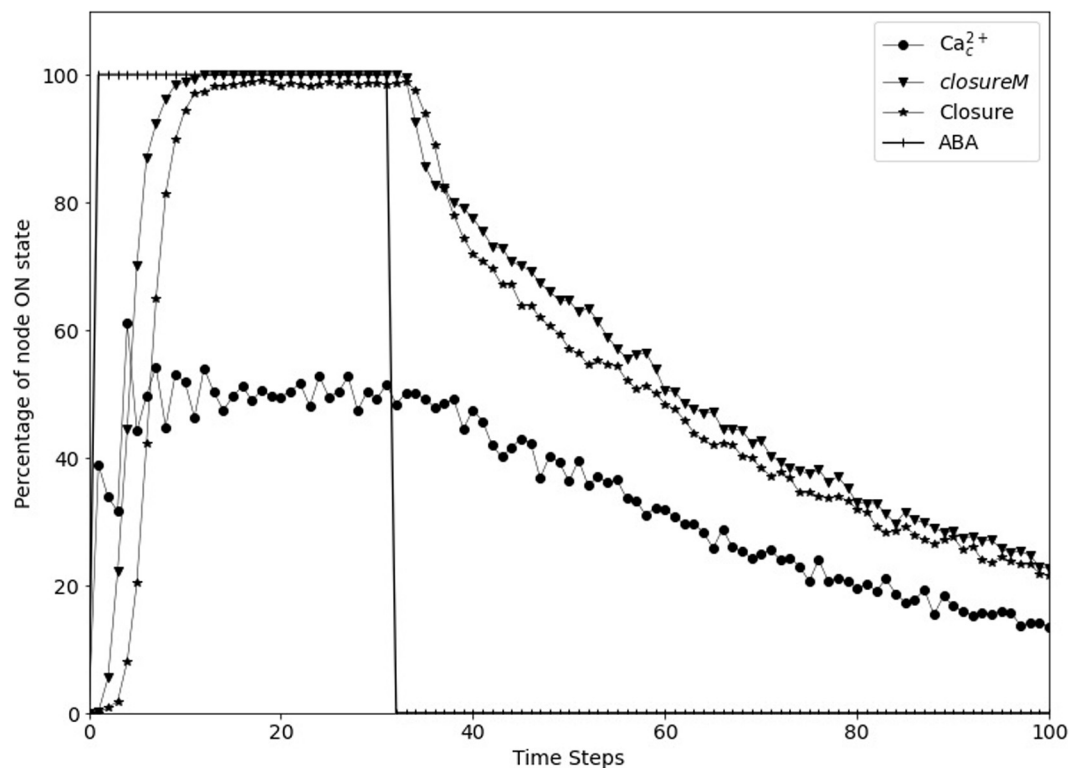


FIGURE 7 | Stomatal re-opening after removal of the signal in the updated Model1. The signal, ABA (vertical lines) is set to fixed ON state for the first 30 time-steps and then set to fixed OFF state for 70 time-steps. The *closureM* conditionally stable motif (downward triangles) is stabilized within the first 10 time-steps and it slowly destabilizes after ABA is set to the OFF state. The state of the node Closure, depicted by the star symbols, shows a similar behavior i.e., it starts decreasing after the signal is removed. The state of the node Ca^{2+}_c , depicted by the circle symbols, is oscillating in the presence of ABA and it slowly transitions to the OFF state after ABA is fixed to OFF state.

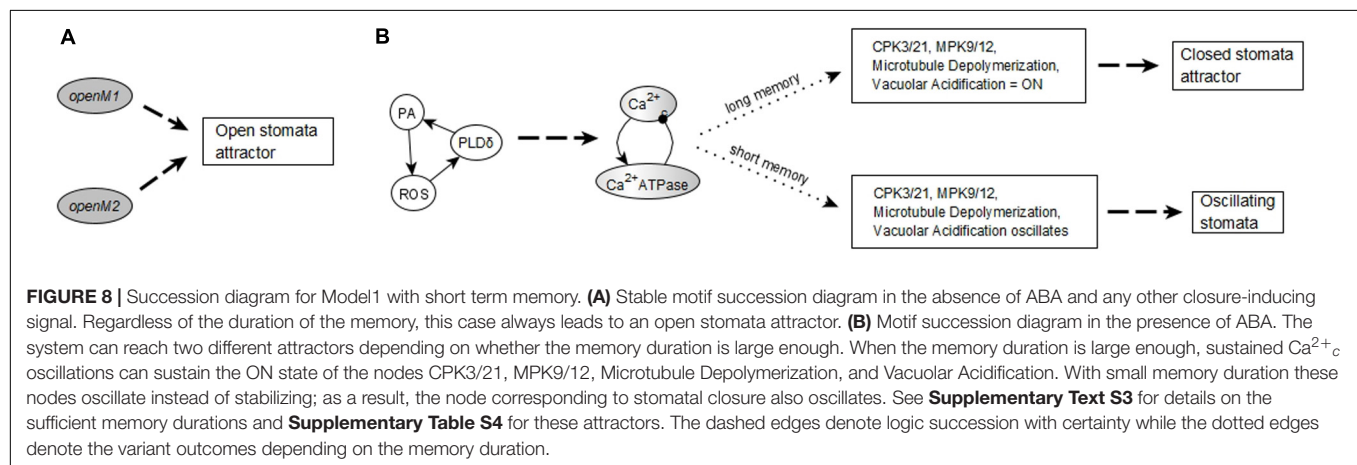


FIGURE 8 | Succession diagram for Model1 with short term memory. **(A)** Stable motif succession diagram in the absence of ABA and any other closure-inducing signal. Regardless of the duration of the memory, this case always leads to an open stomata attractor. **(B)** Motif succession diagram in the presence of ABA. The system can reach two different attractors depending on whether the memory duration is large enough. When the memory duration is large enough, sustained Ca^{2+}_c oscillations can sustain the ON state of the nodes CPK3/21, MPK9/12, Microtubule Depolymerization, and Vacuolar Acidification. With small memory duration these nodes oscillate instead of stabilizing; as a result, the node corresponding to stomatal closure also oscillates. See **Supplementary Text S3** for details on the sufficient memory durations and **Supplementary Table S4** for these attractors. The dashed edges denote logic succession with certainty while the dotted edges denote the variant outcomes depending on the memory duration.

of each of these assumptions is unknown. To reduce the chance of incorrect assumptions, we next explore the possibility of minimizing the number of nodes for which such assumptions are made while still ensuring a zero baseline CPC. We use logical analysis and simulations to find this minimal restriction on initial conditions. In Model1, i.e., the model version where PA inhibits ABI2 through an edge, we find that in order to obtain a baseline CPC of zero, the initial states of 6 of the

17 randomly initialized nodes need to be the opposite of their state corresponding to closed stomata. These six nodes and their corresponding states are cADPR = OFF, GHR1 = OFF, AtRAC1 = ON, PLC = OFF, $\text{PLD}\delta$ = OFF, and DAG = OFF. All of these nodes affect the *closureM* conditionally stable motif directly or indirectly. The nodes $\text{PLD}\delta$ and DAG are sufficient activators of the node PA, which is an internal driver node of *closureM* (if Vacuolar Acidification has already stabilized in the ON state). The

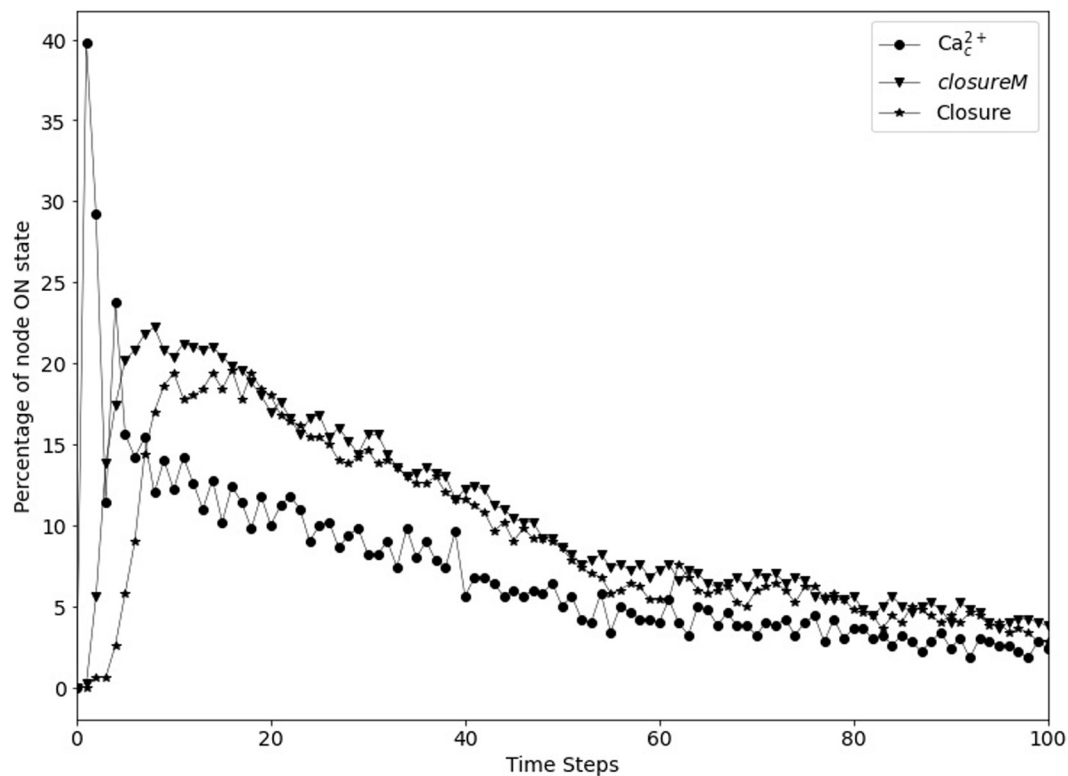


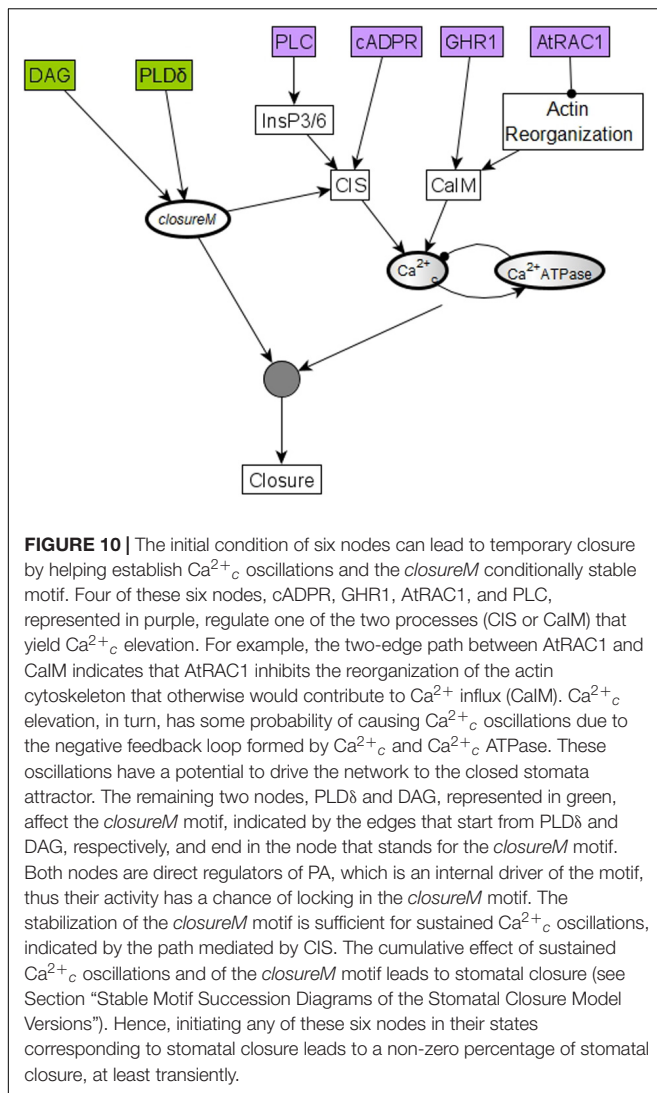
FIGURE 9 | Transient closure observed in Model1 in the absence of ABA when using short-term memory instead of node persistence. Due to the random initial conditions of the 17 nodes, the *closureM* motif temporarily stabilizes in ~29% of the simulations (downward triangles) and there is a non-zero level of Ca^{2+}_c oscillations, which in turn lead to up to ~20% transient closure. Eventually, the percentage of closure reduces to less than 5%.

remaining four nodes, i.e., cADPR, GHR1, AtRAC1, and PLC, promote either Ca^{2+} influx through the membrane (CaIM) or Ca^{2+} release from intracellular stores (CIS), hence promoting Ca^{2+}_c oscillations – see the purple nodes in **Figure 10**. Ca^{2+}_c is an external driver of *closureM* (Maheshwari et al., 2019), as it can induce the ON state of the Vacuolar Acidification node (the condition of the *closureM* motif) and induce the ON state of PA, the ON state of pH_c and the OFF state of ABI2. Since these 3 nodes form a 3-node driver of this conditionally stable motif, the ON state of Ca^{2+}_c leads to stabilization of *closureM* (Maheshwari et al., 2019). Sustained oscillations of Ca^{2+}_c can also lead to the stabilization of the *closureM* motif if the relationship between the $\text{Ca}^{2+}_c = \text{ON}$ period and the short-term memory of Vacuolar Acidification is such that the Vacuolar Acidification = ON condition is satisfied (see Maheshwari et al., 2019, **Supplementary Text S3**, and **Supplementary Figure S7**). Hence if any of these four nodes are initialized in their state corresponding to stomatal closure, there is a chance of indirectly stabilizing the *closureM* motif and to transiently driving the system to closure.

In Model2, i.e., the model version where the inhibition of ABI2 happens through a direct edge from Ca^{2+}_c , we found that a baseline CPC of zero is achieved when the initial state of just four of the 17 nodes is specified and the remaining 13 are initialized randomly. These four nodes are shown in

purple background in **Figure 10**; their corresponding states are cADPR = OFF, GHR1 = OFF, AtRAC1 = ON and PLC = OFF. These four nodes are a subset of the six nodes we identify in the case of Model1; specifically, they are the nodes that can promote Ca^{2+}_c oscillations. Since in Model2 Ca^{2+}_c directly inhibits ABI2, the initial conditions that promote Ca^{2+}_c have a slightly higher likelihood of leading to closure. In Model2 PA is not an internal driver of the *closureM* motif (Maheshwari et al., 2019), thus the initial state of nodes DAG or PLD δ does not play a role in promoting the stabilization of this motif and hence promoting closure. In summary, Model2 requires fewer restrictions than Model1 to avoid the activation, in the absence of ABA, of the *closureM* conditionally stable motif.

We evaluated the agreement of these model versions, sampling their respective sets of initial conditions that correspond to zero baseline closure percentage, with the extant experimental evidence for the effects of constitutive activation of nodes (e.g., whether constitutive activation of a node induces a significant decrease in stomatal aperture, see Section “Evaluation of Consistency Between Simulation and Experiment”). The model-indicated effects of constitutive activation of single nodes fall into three categories: close to baseline response (which in this case means a CPC of zero), slightly increased response, and significantly increased response (which in this case leads to a final closure percentage of 100%). In Model1, i.e., the model



version where PA directly inhibits ABI2, the simulations agree with experimental observations in 14 instances and they disagree in 3 instances (see **Table 1** and **Supplementary Table S5**). The instances of agreement include six cases in which the model and experiments agree in observing closure (decreased stomatal aperture) in case of constitutive activation of the corresponding node and 8 cases in which neither the model nor experiments observe closure. The instances of agreement include the constitutive activation of CalM, which happens in experiments where Ca^{2+} is provided externally (see last entry of **Table 1**). The experimental observation of $\text{Ca}^{2+}_{\text{cyt}}$ oscillations, as well as stomatal closure, in response to the presence of external Ca^{2+} (Jeon et al., 2019) also supports the model prediction that the sustained $\text{Ca}^{2+}_{\text{cyt}}$ oscillations are sufficient for stomatal closure.

Conversely, Model1 yields a zero percentage of closure in case of constitutive activity of S1P and pH_c (as neither S1P nor pH_c can independently stabilize the *closureM* motif) while a decreased aperture was observed experimentally. In case of

PLD α constitutive activity Model1 yields slightly higher than baseline closure (as PLD α leads to PA production, which is an internal driver of the *closureM* motif) while experiments did not observe a statistically significant decrease in stomatal aperture. In the “furthest from closure” initial condition (when nodes are initialized in the state opposite their state corresponding to stomatal closure) there are 13 cases of agreement and four cases of disagreement. The 13 cases of agreement include four cases of significantly increased closure and nine cases of close to baseline closure. In addition to S1P and pH_c , there is also disagreement for external supply of PA and NO, where the model yields close to baseline response but a higher than baseline degree of closure was observed experimentally. Similar to the case of S1P and pH_c , PA and NO cannot independently stabilize the *closureM* motif. They can however stabilize the motif in a fraction of cases when initial conditions are favorable.

In Model2, i.e., the model version where $\text{Ca}^{2+}_{\text{cyt}}$ directly inhibits ABI2, we found the same instances of agreement (13) and disagreement (4) as for the “furthest from closure” initial conditions of Model1 (see **Supplementary Table S6**). The reason for the cases of disagreement between experimental results and simulation results remains the same as for Model1: none of S1P, NO, PA, or pH_c can stabilize the *closureM* motif. Each of them can only activate the motif as part of a two- or three-node driver set (e.g., the two-node driver set of PA and Vacuolar Acidification and the three-node driver set of S1P, Vacuolar Acidification, and PLD δ). The same categories of responses to node constitutive activations, and thus the same cases of disagreement, were found for the “furthest from closure” initial condition and for the least restrictive initial condition that yields a baseline percentage of closure of zero.

We next sought to determine whether initiating one or more nodes of the previously restricted sets, i.e., (cADPR, GHR1, AtRAC1, PLC) or (cADPR, GHR1, AtRAC1, PLC, PLD δ , DAG), respectively, in a random state would still maintain near-zero baseline closure while increasing the percentage of closure when S1P, NO, PA or pH_c is constitutively activated. We indeed found this to be the case. A few slightly less restricted initial conditions in Model1 lead to a non-zero but small value of transient baseline closure (see **Supplementary Table S7**). These cases show improved agreement with experiments regarding node constitutive activation compared to the least restricted initial condition summarized in column 2 of **Supplementary Table S6**. Specifically, the constitutive activation of pH_c is now in the slightly increased category, in agreement with experimental results, while its categorization in **Supplementary Table S6** as yielding close to baseline response disagreed with experiments. The node pH_c alone is not sufficient to be a driver of the *closureM* motif; however, it can contribute to 2-node or 3-node drivers when combined with the initial activity of DAG or active PLD δ there is a non-zero probability of the *closureM* motif stabilizing and leading to the closed stomata attractor.

In Model2, we found initial conditions where the simulations of node constitutive activations include fewer cases of disagreement with experimental results than the initial conditions that ensure a CPC of zero, but these initial conditions

TABLE 1 | Comparison between experimental results and simulation results in Model1 with short-term memory for constitutive activation of nodes in the absence of ABA.

Node that is constitutively active	<i>In silico</i> response when using the initial condition furthest from closure	<i>In silico</i> response when using the least restricted initial condition	Experimentally observed response
TCTP	Close to baseline	Close to baseline	Close to baseline (Kim et al., 2012)
ROP11	Close to baseline	Close to baseline	Close to baseline (Li et al., 2012)
Microtubule Depolymerization	Close to baseline	Close to baseline	Close to baseline (Jiang et al., 2014)
PLD α	Close to baseline	Slightly increased	Close to baseline (Mishra et al., 2006)
PA	Close to baseline	Slightly increased	Increased (Jacob et al., 1999)
NO	Close to baseline	Slightly increased	Increased (Desikan et al., 2002)
S1P	Close to baseline	Close to baseline	Increased (Ng et al., 2001; Coursol et al., 2003)
AtRAC1	Close to baseline	Close to baseline	Close to baseline (Lemichez et al., 2001)
H ⁺ ATPase	Close to baseline	Close to baseline	Close to baseline (Wang et al., 2014)
PP2CA	Close to baseline	Close to baseline	Close to baseline (Kuhn et al., 2006)
ABI1	Close to baseline	Close to baseline	Close to baseline (Allen et al., 1999)
ABI2	Close to baseline	Close to baseline	Close to baseline (Allen et al., 1999)
cADPR	Significantly increased	Significantly increased	Increased (Joudoi et al., 2013)
InsP3/6	Significantly increased	Significantly increased	Increased (Gilroy et al., 1990)
ROS	Significantly increased	Significantly increased	Increased (Zhang et al., 2001; Kwak et al., 2003)
CaIM	Significantly increased	Significantly increased	Increased (Jeon et al., 2019)

We performed 500 simulations over 50 time-steps in each setting. The cumulative percentage of closure (CPC) for the WT simulation (where no node was constitutively activated) was 0.0. Each row of the table indicates the node that is constitutively active, the response category obtained from the model for two different initial conditions, and the experimentally observed response. There are three response categories: close to baseline, slightly increased compared to baseline, and significantly increased compared to baseline (see Section “Evaluation of Consistency Between Simulation and Experiment”). The second column corresponds to the furthest from closure initial condition, where the 17 nodes with unknown pre-stimulus state are initialized in the opposite of the state they achieve in the closure attractor. The third column corresponds to the least restricted initial condition that gives zero baseline closure percentage and CPC, in which six nodes, namely cADPR, GHR1, AtRAC1, PLC, PLD δ , DAG are initiated in the opposite of the state they achieve in the closure attractor and 11 nodes are initialized randomly. The fourth column lists the experimentally observed responses for each of the node constitutive activation with the corresponding literature references. The experimental responses have two categories: close to baseline and increased response (i.e., decreased aperture). When using the furthest from closure initial condition there are 13 cases of agreement with experiments and 4 cases of disagreement. When using the least restricted initial condition there are 14 cases of agreement and 3 cases of disagreement. All the differences between results lie in the constitutive activation of PLD α , PA, S1P and NO. **Supplementary Table S6** indicates the CPC values of each case.

also result in a CPC that is significantly higher than zero. For example, random initialization of PLC yields a higher than baseline closure in case of constitutive activation of each of S1P, PA, and pH_c, with the only remaining disagreement being NO. The trade-off for this regained agreement of the simulated node constitutive activation with experiments is an increase in the baseline percentage of closure and CPC: the peak percentage of closure is 3% and the CPC is 1.91. In conclusion, it is not possible to simultaneously ensure a zero baseline percentage of closure and recapitulate closure in response to constitutive activity of S1P or pH_c given our current knowledge of the biological resting/open stomata state. The node initializations that help achieve better agreement with experimental data on node constitutive activations also increase the baseline percentage of closure.

DISCUSSION

We present ways to improve the Boolean model of ABA-induced stomatal closure by Maheshwari et al. (2019) to recapitulate biological expectations regarding stomata remaining in an open state in the absence of closure signals, or relaxing back to the open stomata state following the loss of closure signals. We find that modifying the assumed persistent activity of four nodes to a short-term memory effect helps recapitulate re-opening of the stomata after the closure signal is removed. The implementation

of such short-term memory still yields transient closure in the absence of any closure signal; thus we also identify different combinations of initial conditions that minimize this transient closure. We find that the percentage of stomatal closure is sensitive to the initial states of certain nodes. This highlights the significance of these internal nodes and the importance of experimentally determining the resting states (or open stomata states) of these nodes.

Motivated by the incorporation of timing in Boolean modeling in Thakar et al. (2007), we replaced the assumed persistent activity of four nodes, i.e., CPK3/21, MPK9/12, Microtubule Depolymerization, and Vacuolar Acidification, with a function that considers the cumulative effect of the current and the past states of their regulator. We show that this short-term memory effect of an oscillating regulator, for example, Ca²⁺_c, helps these four nodes maintain the persistent activity necessary to ensure stomatal closure in response to closure-inducing signals. This assumption brings the model closer to biological reality by exhibiting reopening of the stomata after the closure-inducing signal is removed. We perform a systematic study of the effect of varying memory durations on the extent of stomatal closure and the rate of reopening upon removal of the signal. Our analysis of the increase in the percentage of ON state of a node for larger memory duration can be extended to other patterns of oscillations of the regulator node and hence it will be useful in various other Boolean models of biological networks.

Our analysis demonstrates that motif succession diagrams provide a powerful means to present and understand the system trajectories, highlighting the points of no return in the system's dynamics and identifying the various attractors the system can lead to. In this work, we present the motif succession diagram of the stomatal closure network, which integrates and summarizes previous research on this network (Li et al., 2006; Albert et al., 2017; Maheshwari et al., 2019). This succession diagram highlights the key role of oscillating motifs, drawing attention to the significance of oscillations in this network. This advances our understanding of node oscillations in the context of a biological system modeled as a Boolean network. Seeking motivation from recent work (Deritei et al., 2019), we also identify conditionally stable and conditional oscillating motifs and differentiate them from their condition-free counterparts.

The motif succession diagram is also an effective measure of the consequences of short-term versus long-term memory. Comparing **Figure 8A** with **Figure 3**, we can see that short-term memory eliminates the trajectories that would yield a closure attractor in the absence of ABA and reduces the variability of the open stomatal attractors. Another way to illustrate the qualitative difference between short-term and long-term memory is through a bifurcation diagram (Tyson and Novak, 2020), which indicates the steady state value(s) of the node Closure for different values of ABA. As described in **Supplementary Text S5**, an on-off-on sequence of ABA yields an irreversible switch in the status of closure in case of persistent activity of CPK3/21, MPK9/12, Microtubule Depolymerization and Vacuolar Acidification, while short-term memory of the same nodes yields reversible closure. Comparing **Figure 8B** with **Figure 4**, we can see that short-term memory leads to an attractor in which the Closure node oscillates in the presence of ABA. The fact that oscillating stomatal apertures have not been observed experimentally in response to ABA suggests that the biological persistence of vacuolar acidification, microtubule depolymerization and of the activity of CPK3/21 and MPK9/12 is essential to the control of oscillations. This also highlights the importance of characterizing and quantifying the biological mechanisms underlying the persistence of these nodes (see **Supplementary Text S1**).

This work and our previous research on stable motifs (Zañudo and Albert, 2013; Steinway et al., 2014; Albert et al., 2017; Zanudo and Albert, 2015; Maheshwari et al., 2019; Rozum and Albert, 2018; Gan and Albert, 2018; Deritei et al., 2019) contributes to the broader field of investigation that connects positive feedback loops, multistability, cell fates and phenotypes (Thomas and Ari, 1990; Huang, 2007; Hari et al., 2020). Specifically, single or intersecting positive feedback loops form stable motifs or conditionally stable motifs. Mutually exclusive stable motifs determine distinct attractors (distinct phenotypes). For example, in the model studied here the *openM1* stable motif and *closureM* conditionally stable motif are mutually exclusive (see **Figure 3A**). The stabilization of one of two mutually exclusive stable motifs

at the expense of the other represents a bifurcation in the system's trajectory toward a specific phenotype. The specific example in this work is the possibility of a trajectory toward closed stomata in the absence of ABA. Such a bifurcation can be viewed as a cellular decision point. Our work suggests a mechanism of cellular plasticity (phenotype switching): destabilization of the conditionally stable motif that underlies the phenotype by deactivating its condition. As seen in previous work (Deritei et al., 2019), stable motifs that are condition-free within the context of one model can become conditionally stable motifs in a broader model that encompasses the original model but includes more regulators and processes.

Our analysis of initial conditions of 17 nodes with uncertain states found that it is not possible to simultaneously have a baseline cumulative percentage of closure of 0 and also recapitulate the experimentally observed closure for constitutive activation of pH_c, PA, NO or S1P (Jacob et al., 1999; Ng et al., 2001; Desikan et al., 2002; Coursol et al., 2003; Mishra et al., 2006; Gonugunta et al., 2008). This discrepancy suggests that the actual guard cell resting/open state does *not* correspond to the state farthest from the state associated with stomatal closure. In order to ensure optimal response over a range of conditions, certain nodes have to already be "primed" (be in their state associated with stomatal closure) prior to receiving the closure stimulus. The effectiveness of such priming has been documented in the case of Ca²⁺_c stimulus: when guard cells were pre-exposed to ABA or CO₂, elevated Ca²⁺_c strongly activated S-type anion channels by shifting their Ca²⁺_c sensitivity to lower levels (Israelsson et al., 2006; Hubbard et al., 2012; Laanemets et al., 2013). This suggestion is also in accordance with previous experimental observations that the cellular changes underlying stomatal closure (e.g., induced by ABA) are not simply the reverse of the processes underlying stomatal opening (e.g., induced by light) (Assmann, 1993; Wang et al., 2001; Yin et al., 2013; He et al., 2018). Such "flexible" nodes may provide important portals for regulation by other stimuli to which multisensory guard cells also respond, including not only CO₂ concentrations but also blue and red light, humidity, and pathogens (Sun et al., 2014; Murata et al., 2015; Assmann and Jegla, 2016; Engineer et al., 2016).

DATA AVAILABILITY STATEMENT

The original and updated network versions analyzed for this study can be found in the following github repository: https://github.com/parulm/Stomata_PP2Cs.

AUTHOR CONTRIBUTIONS

PM, SA, and RA designed the research and methodology, and wrote the manuscript. PM and RA performed the analyses. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by NSF grant MCB-1715826 to SA and RA and NSF grant IIS-1814405 to RA. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGMENTS

We thank Jorge G. T. Zanudo, Xiao Gan, and Jordan Rozum for their helpful discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphys.2020.00927/full#supplementary-material>

REFERENCES

- Ache, P., Becker, D., Ivashikina, N., Dietrich, P., Roelfsema, M. R. G., and Hedrich, R. (2000). GORK, a delayed outward rectifier expressed in guard Cells of *Arabidopsis thaliana*, Is a K⁺-Selective, K⁺-Sensing ion channel. *Febs Lett.* 486, 93–98. doi: 10.1016/S0014-5793(00)02248-1
- Albert, I., Thakar, J., Li, S., Zhang, R., and Albert, R. (2008). Boolean network simulations for life scientists. *Source Code Biol. Med.* 3:16.
- Albert, R., Acharya, B. R., Jeon, B. W., Zañudo, J. G., Zhu, M., Osman, K., et al. (2017). A new discrete dynamic model of ABA-induced stomatal closure predicts key feedback loops. *PLoS Biol.* 15:e2003451. doi: 10.1371/journal.pbio.2003451
- Albert, R., DasGupta, B., Dondi, R., Kachalo, S., Sontag, E., Zelikovsky, A., et al. (2007). A novel method for signal transduction network inference from indirect experimental evidence. *J. Comput. Biol.* 14, 927–949. doi: 10.1089/cmb.2007.0015
- Allen, G. J., Kuchitsu, K., Chu, S. P., Murata, Y., and Schroeder, J. I. (1999). Arabidopsis ABI1-1 and ABI2-1 phosphatase mutations reduce abscisic acid-induced cytoplasmic calcium rises in guard cells. *Plant Cell* 11, 1785–1798. doi: 10.1105/tpc.11.9.1785
- Assmann, S. M. (1993). Signal transduction in guard cells. *Annu. Rev. Cell Biol.* 9, 345–375. doi: 10.1146/annurev.cb.09.110193.002021
- Assmann, S. M., and Jegla, T. (2016). Guard Cell Sensory systems: recent insights on stomatal responses to light, abscisic acid, and CO₂. *Curr. Opin. Plant Biol.* 33, 157–167. doi: 10.1016/j.pbi.2016.07.003
- Coursol, S., Fan, L.-M., Le Stunff, H., Spiegel, S., Gilroy, S., and Assmann, S. M. (2003). Sphingolipid signalling in arabidopsis guard cells involves heterotrimeric G proteins. *Nature* 423, 651–654. doi: 10.1038/nature01643
- Cummins, W. R., Kende, H., and Raschke, K. (1971). Specificity and reversibility of the rapid stomatal response to abscisic acid. *Planta* 99, 347–351. doi: 10.1007/bf00385826
- Deritei, D., Rozum, J., Regan, E. R., and Albert, R. (2019). A feedback loop of conditionally stable circuits drives the cell cycle from checkpoint to checkpoint. *Sci. Rep.* 9, 1–19.
- Desikan, R., Griffiths, R., Hancock, J., and Neill, S. (2002). A new role for an old enzyme: nitrate reductase-mediated nitric oxide generation is required for abscisic acid-induced stomatal closure in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 99, 16314–16318. doi: 10.1073/pnas.252461999
- Engineer, C. B., Hashimoto-Sugimoto, M., Negi, J., Israelsson-Nordström, M., Azoulay-Shemer, T., Rappel, W.-J., et al. (2016). CO₂ sensing and CO₂ regulation of stomatal conductance: advances and open questions. *Trends Plant Sci.* 21, 16–30. doi: 10.1016/j.tplants.2015.08.014
- Gan, X., and Albert, R. (2016). Analysis of a dynamic model of guard cell signaling reveals the stability of signal propagation. *BMC Syst. Biol.* 10:78.
- Gan, X., and Albert, R. (2018). General method to find the attractors of discrete dynamic models of biological systems. *Phys. Rev. E* 97:042308.
- Gilroy, S., Read, N., and Trewavas, A. J. (1990). Elevation of cytoplasmic calcium by caged calcium or caged inositol trisphosphate initiates stomatal closure. *Nature* 346, 769–771. doi: 10.1038/346769a0
- Gonugunta, V. K., Srivastava, N., Puli, M. R., and Raghavendra, A. S. (2008). Nitric oxide production occurs after cytosolic alkalinization during stomatal closure induced by abscisic acid. *Plant Cell Environ.* 31, 1717–1724. doi: 10.1111/j.1365-3040.2008.01872.x
- Grondin, A., Rodrigues, O., Verdoucq, L., Merlot, S., Leonhardt, N., and Maurel, C. (2015). Aquaporins contribute to ABA-triggered stomatal closure through OST1-mediated phosphorylation. *Plant Cell* 27, 1945–1954. doi: 10.1105/tpc.15.00421
- Hari, K., Sabuwala, B., Subramani, B. V., La Porta, C. A., Zapperi, S., Font-Clos, F., et al. (2020). Identifying inhibitors of epithelial-mesenchymal plasticity using a network topology-based approach. *NPJ Syst. Biol. Appl.* 6, 1–12.
- He, J., Zhang, R.-X., Peng, K., Tagliavia, C., Li, S., Xue, S., et al. (2018). The BIG protein distinguishes the process of CO₂-induced stomatal closure from the inhibition of stomatal opening by CO₂. *New Phytol.* 218, 232–241. doi: 10.1111/nph.14957
- Hosy, E., Vavasseur, A., Mouline, K., Dreyer, I., Gaymard, F., Porée, F., et al. (2003). The Arabidopsis Outward K⁺ channel GORK is involved in regulation of stomatal movements and plant transpiration. *Proc. Natl. Acad. Sci. U.S.A.* 100, 5549–5554. doi: 10.1073/pnas.0733970100
- Huang, S. (2007). *Cell Fates as Attractors-Stability and Flexibility of Cellular Phenotype: Endothelial Biomedicine*, 1 Edn. New York, NY: Cambridge University Press, 1761–1779.
- Hubbard, K. E., Siegel, R. S., Valerio, G., Brandt, B., and Schroeder, J. I. (2012). Abscisic acid and CO₂ signalling via calcium sensitivity priming in guard cells, new CDPK mutant phenotypes and a method for improved resolution of stomatal stimulus-response analyses. *Ann. Bot.* 109, 5–17. doi: 10.1093/aob/mcr252
- Israelsson, M., Siegel, R. S., Young, J., Hashimoto, M., Iba, K., and Schroeder, J. I. (2006). Guard cell ABA and CO₂ signaling network updates and Ca²⁺ sensor priming hypothesis. *Curr. Opin. Plant Biol.* 9, 654–663. doi: 10.1016/j.pbi.2006.09.006
- Jacob, T., Ritchie, S., Assmann, S. M., and Gilroy, S. (1999). Abscisic acid signal transduction in guard cells is mediated by phospholipase D activity. *Proc. Natl. Acad. Sci.* 96, 12192–12197. doi: 10.1073/pnas.96.21.12192
- Jeon, B. W., Acharya, B. R., and Assmann, S. M. (2019). The Arabidopsis heterotrimeric G-protein β subunit, AGB 1, is required for guard cell calcium sensing and calcium-induced calcium release. *Plant J.* 99, 231–244.

TABLE S1 | The full names corresponding to the abbreviated node names in the 49-node ABA induced closure network.

TABLE S2 | Initial conditions of all nodes in the model, reproduced from the 49-node model.

TABLE S3 | The Boolean update function of each node in Model1.

TABLE S4 | Attractors of Model1 and Model2.

TABLE S5 | Attractors of Model1 with short-term memory.

TABLE S6 | Simulation results from Model1 and Model2 with short-term memory for constitutive activation of nodes in the absence of ABA for initial conditions that ensure the absence of baseline closure.

TABLE S7 | Simulation results for constitutive activation of each node in the absence of ABA in Model1 with short term memory for initial conditions that result in near zero baseline percentage of closure.

TEXT S1 | Biological reasoning for the assumed slow decay of the activity of four network elements.

TEXT S2 | Succession diagram for Model2.

- Jiang, Y., Wu, K., Lin, F., Qu, Y., Liu, X., and Zhang, Q. (2014). Phosphatidic acid integrates calcium signaling and microtubule dynamics into regulating aba-induced stomatal closure in *Arabidopsis*. *Planta* 239, 565–575. doi: 10.1007/s00425-013-1999-5
- Joudoi, T., Shichiri, Y., Kamizono, N., Akaike, T., Sawa, T., Yoshitake, J., et al. (2013). Nitrated cyclic GMP modulates guard cell signaling in *Arabidopsis*. *Plant Cell* 25, 558–571. doi: 10.1105/tpc.112.105049
- Kim, Y.-M., Han, Y.-J., Hwang, O.-J., Lee, S.-S., Shin, A.-Y., Kim, S. Y., et al. (2012). Overexpression of arabidopsis translationally controlled tumor protein gene AtTCTP enhances drought tolerance with rapid aba-induced stomatal closure. *Mol. Cells* 33, 617–626. doi: 10.1007/s10059-012-0080-8
- Kollist, H., Nuhkat, M., and Roelfsema, M. R. G. (2014). Closing Gaps: linking elements that control stomatal movement. *New Phytol.* 203, 44–62. doi: 10.1111/nph.12832
- Kuhn, J. M., Boisson-Dernier, A., Dizon, M. B., Maktabi, M. H., and Schroeder, J. I. (2006). The protein phosphatase AtPP2CA negatively regulates abscisic acid signal transduction in arabidopsis, and effects of Abh1 on AtPP2CA mRNA. *Plant Physiol.* 140, 127–139. doi: 10.1104/pp.105.070318
- Kwak, J. M., Mori, I. C., Pei, Z.-M., Leonhardt, N., Torres, M. A., Dangl, J. L., et al. (2003). NADPH oxidase AtrbohD and AtrbohF genes function in ROS-dependent ABA signaling in *Arabidopsis*. *EMBO J.* 22, 2623–2633. doi: 10.1093/emboj/cdg277
- Laanemets, K., Brandt, B., Li, J., Merilo, E., Wang, Y.-F., Keshwani, M. M., et al. (2013). Calcium-dependent and-independent stomatal signaling network and compensatory feedback control of stomatal opening via Ca^{2+} sensitivity priming. *Plant Physiol.* 163, 504–513. doi: 10.1104/pp.113.22.0343
- Lemichez, E., Wu, Y., Sanchez, J.-P., Mettouchi, A., Mathur, J., and Chua, N.-H. (2001). Inactivation of AtRac1 by abscisic acid is essential for stomatal closure. *Genes Dev.* 15, 1808–1816. doi: 10.1101/gad.900401
- Li, S., Assmann, S. M., and Albert, R. (2006). Predicting essential components of signal transduction networks: a dynamic model of guard cell abscisic acid signaling. *PLoS Biol.* 4:e312. doi: 10.1371/journal.pbio.0040312
- Li, Z., Kang, J., Sui, N., and Liu, D. (2012). ROP11 GTPase is a negative regulator of multiple aba responses in *Arabidopsis*. *J. Integrat. Plant Biol.* 54, 169–179. doi: 10.1111/j.1744-7909.2012.01100.x
- Ma, Y., Szostkiewicz, I., Korte, A., Moes, D., Yang, Y., Christmann, A., et al. (2009). Regulators of PP2C phosphatase activity function as abscisic acid sensors. *Science* 324, 1064–1068.
- Maheshwari, P., and Albert, R. (2017). A framework to find the logic backbone of a biological network. *BMC Syst. Biol.* 11:122.
- Maheshwari, P., Du, H., Sheen, J., Assmann, S. M., and Albert, R. (2019). Model-driven discovery of calcium-related protein-phosphatase inhibition in plant guard cell signaling. *PLoS Comput. Biol.* 15:e1007429. doi: 10.1371/journal.pcbi.1007429
- Mishra, G., Zhang, W., Deng, F., Zhao, J., and Wang, X. (2006). A bifurcating pathway directs abscisic acid effects on stomatal closure and opening in *Arabidopsis*. *Science* 312, 264–266. doi: 10.1126/science.1123769
- Murata, Y., Mori, I. C., and Munemasa, S. (2015). Diverse stomatal signaling and the signal integration mechanism. *Ann. Rev. Plant Biol.* 66, 369–392. doi: 10.1146/annurev-arplant-043014-114707
- Ng, C. K.-Y., Carr, K., McAinsh, M. R., Powell, B., and Hetherington, A. M. (2001). Drought-induced guard cell signal transduction involves sphingosine-1-phosphate. *Nature* 410, 596–599. doi: 10.1038/35069092
- Roelfsema, M. R. G., and Hedrich, R. (2005). In the light of stomatal opening: new insights into 'the watergate'. *New Phytol.* 167, 665–691. doi: 10.1111/j.1469-8137.2005.01460.x
- Rozum, J. C., and Albert, R. (2018). Self-sustaining positive feedback loops in discrete and continuous systems. *J. Theoret. Biol.* 459, 36–44. doi: 10.1016/j.jtbi.2018.09.017
- Sierla, M., Waszczak, C., Vahisalu, T., and Kangasjärvi, J. (2016). Reactive oxygen species in the regulation of stomatal movements. *Plant Physiol.* 171, 1569–1580. doi: 10.1104/pp.16.00328
- Staxén, I., Pical, C., Montgomery, L. T., Gray, J. E., Hetherington, A. M., and McAinsh, M. R. (1999). Abscisic acid induces oscillations in guard-cell cytosolic free calcium that involve phosphoinositide-specific Phospholipase C. *Proc. Natl. Acad. Sci. U.S.A.* 96, 1779–1784. doi: 10.1073/pnas.96.4.1779
- Steinway, S. N., Zañudo, J. G., Ding, W., Rountree, C. B., Feith, D. J., Loughran, T. P., et al. (2014). Network modeling of TGF β signaling in hepatocellular carcinoma epithelial-to-mesenchymal transition reveals joint sonic hedgehog and wnt pathway activation. *Cancer Res.* 74, 5963–5977. doi: 10.1158/0008-5472.can-14-0225
- Sun, Z., Jin, X., Albert, R., and Assmann, S. M. (2014). Multi-level modeling of light-induced stomatal opening offers new insights into its regulation by drought. *PLoS Comput. Biol.* 10:e1003930. doi: 10.1371/journal.pcbi.1003930
- Thakar, J., Pilione, M., Kirimanjeswara, G., Harvill, E. T., and Albert, R. (2007). Modeling systems-level regulation of host immune responses. *PLoS Comput. Biol.* 3:e109. doi: 10.1371/journal.pcbi.0030109
- Thomas, R., and Ari, R. D. (1990). *Biological Feedback*. Boca Raton, FL: CRC press.
- Tyson, J. J., and Novak, B. (2020). A dynamical paradigm for molecular cell biology. *Trends Cell Biol.* 504–515. doi: 10.1016/j.tcb.2020.04.002
- Wang, R.-S., Saadatpour, A., and Albert, R. (2012). Boolean modeling in systems biology: an overview of methodology and applications. *Phys. Biol.* 9:055001. doi: 10.1088/1478-3975/9/5/055001
- Wang, X.-Q., Ullah, H., Jones, A. M., and Assmann, S. M. (2001). G protein regulation of ion channels and abscisic acid signaling in *Arabidopsis* guard cells. *Science* 292, 2070–2072. doi: 10.1126/science.1059046
- Wang, Y., Noguchi, K., Ono, N., Inoue, S., Terashima, I., and Kinoshita, T. (2014). Overexpression of plasma membrane H^{+} -ATPase in guard cells promotes light-induced stomatal opening and enhances plant growth. *Proc. Natl. Acad. Sci. U.S.A.* 111, 533–538. doi: 10.1073/pnas.1305438111
- Yin, Y., Adachi, Y., Ye, W., Hayashi, M., Nakamura, Y., Kinoshita, T., et al. (2013). Difference in abscisic acid perception mechanisms between closure induction and opening inhibition of stomata. *Plant Physiol.* 163, 600–610. doi: 10.1104/pp.113.223826
- Zañudo, J. G., and Albert, R. (2013). An effective network reduction approach to find the dynamical repertoire of discrete dynamic networks. *Chaos* 23:025111. doi: 10.1063/1.4809777
- Zanudo, J. G., and Albert, R. (2015). Cell fate reprogramming by control of intracellular network dynamics. *PLoS Comput. Biol.* 11:e1004193. doi: 10.1371/journal.pcbi.1004193
- Zhang, X., Zhang, L., Dong, F., Gao, J., Galbraith, D. W., and Song, C.-P. (2001). Hydrogen peroxide is involved in abscisic acid-induced stomatal closure in *vicia faba*. *Plant Physiol.* 126, 1438–1448. doi: 10.1104/pp.126.4.1438

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Maheshwari, Assmann and Albert. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Orchestration of lincRNA-p21 and miR-155 in Modulating the Adaptive Dynamics of HIF-1 α

Cheng-Yuan Sun¹, Xiao-Peng Zhang^{2,3*}, Feng Liu^{1,3*} and Wei Wang^{1,3}

¹ National Laboratory of Solid State Microstructures, Department of Physics, Nanjing University, Nanjing, China, ² Kuang Yaming Honors School, Nanjing University, Nanjing, China, ³ Institute for Brain Sciences, Nanjing University, Nanjing, China

Hypoxia-inducible factor-1 (HIF-1) is the key regulator of cellular adaptive response to hypoxia. Accumulating evidence shows that HIF-1 induces some non-coding RNAs (ncRNAs) including lincRNAs and miRNAs to modulate its own activity, enclosing several feedback loops. How the two classes of ncRNAs are orchestrated in the HIF-1-dependent adaptive response to hypoxia is poorly understood. By selecting lincRNA-p21 and miR-155 as the representatives, we develop an integrated model of the HIF-1 network comprising interlinked positive and negative feedback loops to clarify the interplay between the two ncRNAs in the hypoxic response. By numerical simulations, we find that coordination of lincRNA-p21 and miR-155 shapes the adaptive dynamics of HIF-1 α : lincRNA-p21 induction in the early phase stimulates the upregulation of HIF-1 α via stabilizing it, while miR-155 induction in the late phase promotes the recovery of HIF-1 α via enhancing the degradation of its mRNA. Moreover, HIF-1 α -induced PHD2 plays an auxiliary role in the decline of HIF-1 α . In addition, lincRNA-p21 and miR-155 modulate each other via regulating HIF-1 α activity. Together, lincRNA-p21 and miR-155 coordinate in modulating HIF-1 α dynamics, and our work may shed light on the role for ncRNAs in the cellular adaptation to hypoxia.

Keywords: hypoxia, HIF-1 α , lincRNA-p21, miR-155, adaptive dynamics, feedback loop

OPEN ACCESS

Edited by:

Chunhe Li,
Fudan University, China

Reviewed by:

Xiaojun Tian,
Arizona State University, United States
Tianhai Tian,
Monash University, Australia

*Correspondence:

Xiao-Peng Zhang
zhangxp@nju.edu.cn
Feng Liu
fliu@nju.edu.cn

Specialty section:

This article was submitted to
Systems Biology,
a section of the journal
Frontiers in Genetics

Received: 02 June 2020

Accepted: 16 July 2020

Published: 18 August 2020

Citation:

Sun C-Y, Zhang X-P, Liu F and
Wang W (2020) Orchestration of
lincRNA-p21 and miR-155 in
Modulating the Adaptive Dynamics of
HIF-1 α . *Front. Genet.* 11:871.
doi: 10.3389/fgene.2020.00871

1. INTRODUCTION

Hypoxia plays significant roles in human physiology and diseases including cancer (Koh and Powis, 2012). Hypoxia-inducible factor-1 (HIF-1) is the key mediator of the cellular adaption to hypoxia (Schofield and Ratcliffe, 2004). HIF-1 is a heterodimer composed of an oxygen-dependent α -subunit (HIF-1 α) and a constitutively expressed nuclear β -subunit (HIF-1 β) (Wang et al., 1995). Under normoxia, HIF-1 α is hydroxylated by prolyl hydroxylases (PHDs) on Pro402 and Pro564, and these modifications facilitate the binding of HIF-1 α to VHL (von Hippel-Lindau), promoting the ubiquitin-dependent proteasomal degradation of HIF-1 α (Ohh et al., 2000; Jaakkola et al., 2001). In addition, the hydroxylase factor inhibiting HIF-1 (FIH-1) hydroxylates HIF-1 α on Asn803 to repress its transcriptional activity via preventing the recruitment of coactivator p300/CBP (Mahon et al., 2001). Upon hypoxia, PHDs and FIH-1 are deactivated so that HIF-1 α is stabilized and translocates to the nucleus to form a transcriptional complex with HIF-1 β (Maxwell et al., 2001). Activated HIF-1 induces hundreds of genes involved in glycolysis, angiogenesis, cell survival, and metastasis (Harada et al., 2007; Semenza, 2009, 2012; Zeng et al., 2015). Moreover, HIF-1 α itself shows adaptive dynamics in the hypoxic response (Stiehl et al., 2006; Minamishima et al., 2009). It

has been reported that tight control of transient HIF-1 α dynamics is essential for cell survival in hypoxia (Ginouvé et al., 2008; Henze et al., 2010; Bagnall et al., 2014). The detailed mechanism underlying the adaptive dynamics of HIF-1 α in hypoxia is unclear.

MicroRNAs (miRNAs), especially HIF-1-inducible miRNAs, also play significant roles in cellular response to hypoxia (Serocki et al., 2018). For example, HIF-1-targeted miR-210 is shown to regulate cellular metabolism or angiogenesis during hypoxia (Chan et al., 2009; Li et al., 2016). miR-155 is induced by HIF-1 in multiple cell lines (Xie et al., 2015). It was found that miR-155 contributes to the descending of HIF-1 α in the late phase by enhancing the degradation of HIF-1 α mRNA, enclosing a negative feedback loop (Bruning et al., 2011). Moreover, HIF-1 upregulates the expression of PHD2 or PHD3 to promote HIF-1 α degradation, compensating for repression of PHD activity in hypoxia (Minamishima et al., 2009; Bagnall et al., 2014). An intriguing question is whether miR-155 and PHDs play distinct roles in the downregulation of HIF-1 α .

Long non-coding RNAs (lncRNAs) are also involved in the hypoxic response by regulating HIF-1 activity (Chang et al., 2016). LincRNA-p21 represses the degradation of HIF-1 α by blocking the VHL-HIF-1 α interaction, enclosing a positive feedback loop (Yang et al., 2014). As a result, HIF-1 α amplifies its own activation and induces Glut1 and LDHA to facilitate glycolysis in hypoxic cells (Yang et al., 2014). Thus, there exist several HIF-1 α -centered negative and positive feedback loops involving PHDs, miR-155, and lincRNA-p21. It is a challenge to clarify how these interlinked feedback loops interplay in shaping HIF-1 α dynamics under distinct hypoxic conditions.

A series of theoretical models have been developed to explore the mechanism for the regulation of HIF-1 α dynamics (Kohn et al., 2004; Qutub and Popel, 2006; Dayan et al., 2009; Nguyen et al., 2013). Kohn et al. explored the mechanism for the switch-like response of HIF-1 to hypoxia (Kohn et al., 2004). Qutub et al. characterized the effects of micro-environmental factors, such as ascorbate, iron, and PHD, on the hydroxylation of HIF-1 α (Qutub and Popel, 2006). Nguyen et al. clarified the regulation of HIF-1 α stability and activity by FIH-1 (Nguyen et al., 2013). We have explored the interplay between HIF-1 α and p53 upon hypoxia in several models (Zhou et al., 2015; Wang et al., 2019; Ye et al., 2019). Although miRNAs-mediated HIF-1 α regulation has been involved in some modeling studies (Bruning et al., 2011; Fábíán et al., 2016), how HIF-1-targeted miRNAs and lncRNAs are orchestrated to regulate HIF-1 α is less understood. It is feasible to select miR-155 and lincRNA-p21 as the representatives since they are both expressed at least in HeLa cells (Bruning et al., 2011; Yang et al., 2014). It is promising to clarify how the crosstalk between miR-155 and lincRNA-p21 modulates HIF-1 α dynamics upon hypoxia by modeling.

Here, we develop a model of the HIF-1 signaling network including lincRNA-p21 and miR-155 to explore how different ncRNAs coordinate to mediate the adaptive response of HIF-1 α to hypoxia. Our results show that lincRNA-p21 and miR-155 are induced in different phases of the response to shape the adaptative dynamics of HIF-1 α . LincRNA-p21 induction in the early phase stabilizes HIF-1 α by blocking its degradation,

whereas miR-155 is induced in the late phase to downregulate HIF-1 α via enhancing HIF-1 α mRNA degradation. Moreover, miR-155 and PHD2 cooperate to facilitate the recovery of HIF-1 α . We found that lincRNA-p21 and miR-155 compete with each other to modulate HIF-1 α dynamics. Together, HIF-1 sequentially induces lincRNA-p21 and miR-155 to facilitate the cellular adaption to hypoxia.

2. MODELS AND METHODS

2.1. Overview of the Model

We built an integrated model of the HIF-1 network in response to hypoxia, focusing on the role of lincRNA-p21 and miR-155 in shaping HIF-1 dynamics (**Figure 1**). For simplicity, subcellular compartmentalization is not considered. Given the constitutive expression of HIF-1 β in the hypoxic response (Wang et al., 1995), the dimerization of HIF-1 α and HIF-1 β is ignored, and HIF-1 heterodimer is not distinguished from HIF-1 α thereafter. The model is mainly composed of two modules responsible for oxygen sensing and feedback regulation of HIF-1 α .

The stability and activity of HIF-1 α is controlled by the sensors of oxygen, PHDs, and FIH-1, respectively. In our model, PHD2 is considered the representative of PHDs as it is the primary oxygen sensor among the three PHD isoforms (Berra et al., 2003; Takeda et al., 2006). We consider two forms of PHD2: PHD2 (inactive form) and PHD2_{ac} (active form); FIH-1 is divided into FIH-1 (inactive form) and FIH-1_{ac} (active form). Upon hypoxia, PHD2 and FIH-1 are deactivated, leading to the stabilization and activation of HIF-1 α (Jaakkola et al., 2001; Mahon et al., 2001). It is assumed that the total amount of PHD-2, PHD2_{tot}, is HIF-1 α -dependent (Stiehl et al., 2006), whereas that of FIH-1, FIH-1_{tot}, is supposed to be a constant since its expression is independent of HIF-1 α .

HIF-1 α protein is produced by the translation of HIF-1 α mRNA (HIF-1 α m). Three forms of HIF-1 α protein are considered, i.e., HIF-1 α (unhydroxylated), HIF-1 α -aOH (asparagine-hydroxylated), and HIF-1 α -aOHpOH (hydroxylated at both proline and asparagine sites). In other words, HIF-1 α -aOHpOH is proline-hydroxylated form while HIF-1 α and HIF-1 α -aOH are proline-unhydroxylated forms. Of note, the hydroxylation steps are supposed to be irreversible (Schofield and Ratcliffe, 2004; Chan et al., 2005). Although the hydroxylation of HIF-1 α by FIH-1 represses its transcriptional activity via preventing the recruitment of co-activator p300 (Lando et al., 2002), we assume that HIF-1 α with asparagine-hydroxylation alone is of partial transcriptional activity (Dayan et al., 2006; Chan et al., 2016). Thus, it is assumed both HIF-1 α and HIF-1 α -aOH can induce miR-155, lincRNA-p21, and PHD2. miR-155 regulates HIF-1 α posttranscriptionally by promoting the degradation of HIF-1 α mRNA (Bruning et al., 2011), while lincRNA-p21 can promote the stabilization of HIF-1 α (Yang et al., 2014). Together, two negative and one positive feedback loops are interlinked to regulate HIF-1 α .

2.2. Details of the Model

The network model is described by a set of ordinary differential equations. The key points for the equations are listed as follows.

not significantly affected by hydroxylation (Yang et al., 2014). Thus, the Michaelis constants of lincRNA-p21 for repressing the degradation of proline-hydroxylated HIF-1 α (j_{dhifpoh}) and proline-unhydroxylated HIF-1 α (j_{dhif}) are assumed to be equal.

2.3. Methods

The concentration of each species is represented by $[\cdot]$, corresponding to a state variable in rate equations in **Supporting Material**. The relative value of the oxygen level is adopted, and 1 represents 1% O₂ in the model. The reactions concerned with hydroxylation or activation are described by Michaelis-Menten kinetics. The depiction of variables and their initial values are listed in **Table S1**. All the initial values of the variables are set to be their steady states under normoxia. The standard values of the parameters are listed in **Table S2**. The unit of time is minutes and the units of parameters are decided so that the concentration of proteins or RNAs is dimensionless. The ordinary differential equations were solved numerically by Oscill8 (<http://oscill8.sourceforge.net/>) with adaptive time steps. The bifurcation diagrams were also plotted using Oscill8.

3. RESULTS

3.1. Overview of HIF-1 α Dynamics Upon Hypoxia

We first display the dependence of the total level of HIF-1 α , [HIF-1 α_{tot}], on O₂ level by bifurcation diagram (**Figure 2A**). The response curve is divided into several parts by four bifurcation points including two saddle-node bifurcation points (S1 and S2) and two Hopf bifurcation points (H1 and H2). The two branches separated by S1 and S2 correspond to the low and high states of [HIF-1 α_{tot}]. With decreasing O₂ levels, the stable level of HIF-1 α_{tot} increases slowly. The steady states of [HIF-1 α_{tot}] become unstable for O₂ levels between H1 and S1. [HIF-1 α_{tot}] switches to high stable states in this regime and its stable-state level rises continuously until it reaches the maxima around 0.1% O₂, then drops to low levels under severe hypoxia, which is qualitatively in accordance with the experimental data (Jiang et al., 1996). In addition, the high steady states of [HIF-1 α_{tot}] exhibit instability for O₂ levels between H2 and S2.

Under different O₂ levels, the temporal dynamics of [HIF-1 α_{tot}] are shown in **Figure 2B**. For mild hypoxia (3% O₂), [HIF-1 α_{tot}] keeps at low levels; for moderate hypoxia (1% O₂), [HIF-1 α_{tot}] exhibits adaptive dynamics, which is related to the existence of Hopf bifurcation points; for severe hypoxia (0.05% O₂), [HIF-1 α_{tot}] eventually reaches a high level; for anoxia (0% O₂), [HIF-1 α_{tot}] shows a smaller pulse and drops to lower levels finally. Of note, when O₂ level is between H1 and S1 (0.3% O₂), [HIF-1 α_{tot}] first climbs to a very high level and then settles down to a fairly high level (**Figure S1**), consistent with the instability of low steady states (see **Figure 2A**).

As mentioned above, FIH-1 preferentially hydroxylates HIF-1 α and can maintain its activity at lower oxygen levels than PHD2 (Koivunen et al., 2004). As a result, PHD2 and FIH-1 are deactivated sequentially under aggravating hypoxia. Our results show that [HIF-1 α -aOH] and [HIF-1 α] are predominant under moderate and severe hypoxia, respectively (**Figure 2C**). For

moderate hypoxia, [HIF-1 α -aOH] exhibits pulsatile dynamics and is much higher than [HIF-1 α], while fully activated HIF-1 α becomes dominant under severe hypoxia. Therefore, HIF-1 α is progressively activated in response to hypoxia.

As shown in **Figure 2A**, [HIF-1 α] drops markedly under anoxia, consistent with experimental results (Jiang et al., 1996). **Figure 2D** shows the bifurcation diagrams of [HIF-1 α_{tot}] vs. O₂% with or without FIH-1-mediated degradation. The two diagrams are separable only in severe hypoxia and anoxia, which means that FIH-1 protects HIF-1 α from degradation only under such conditions. HIF-1 α accumulates markedly and its level decreases mildly in the absence of FIH-1-mediated degradation. The marked decline of [HIF-1 α] should result from FIH-1 deactivation that facilitates HIF-1 α degradation under severe hypoxia or anoxia. Our results may provide a plausible mechanism for the regulation of HIF-1 α degradation independent of PHDs.

3.2. HIF-1 α -Induced lincRNA-p21 Modulates the Adaptive Dynamics of HIF-1 α Through a Positive Feedback Loop

HIF-1 α induces lincRNA-p21 to promote its own stabilization (Yang et al., 2014), and how the latter modulates the adaptive dynamics of HIF-1 α is investigated in the following. To further verify our model, we compare the simulation results for [lincRNA-p21] at 24h with the experimental results for different O₂ levels (Yang et al., 2014). LincRNA-p21 is indeed markedly evoked in hypoxia, showing good agreements with the experimental data (**Figure 3A**). Both [HIF-1 α_{tot}] and [lincRNA-p21] exhibit adaptive dynamics at 1% O₂ in the standard parameter setting, and the results are well consistent with experimental data (Yang et al., 2014) (**Figures 3B,C**). When HIF-1 α -dependent lincRNA-p21 expression is removed, [HIF-1 α_{tot}] remains at rather low levels, meaning that lincRNA-p21 induction is crucial for the accumulation of HIF-1 α in the ascending phase. In addition, given translation inhibition and the high initial level of HIF-1 α_{tot} , lincRNA-p21 knockout makes [HIF-1 α_{tot}] decay much faster compared with the normal case at 1% O₂ (**Figure 3D**). These results show that lincRNA-p21 upregulates HIF-1 α by repressing its degradation (Yang et al., 2014; Meng et al., 2018). Together, lincRNA-p21 is required for the adaptive dynamics of HIF-1 α .

Given lincRNA-p21 is induced by HIF-1 α , we further explore the effect of lincRNA-p21 induction rate on HIF-1 α adaptation to hypoxia. Since HIF-1 α -aOH is predominant under moderate hypoxia (**Figure 2C**), we only consider the effect of HIF-1 α -aOH-dependent induction rate of lincRNA-p21 (k_{sLAP21a}) on HIF-1 α dynamics (**Figures 4A,B**). At 1% O₂, as mentioned above, [HIF-1 α_{tot}] displays adaptive dynamics; for $k_{\text{sLAP21a}} = 0$, HIF-1 α cannot be evoked; for rather large k_{sLAP21a} (0.08), although [HIF-1 α_{tot}] rises more sharply in the early phase, it settles at fairly high levels instead of dropping to low levels (**Figure 4A**). Therefore, proper expression of lincRNA-p21 is required for the perfect adaptation of HIF-1 α to hypoxia.

The peak during the temporal evolution of [HIF-1 α_{tot}], HIF-1 α_{peak} , is selected as another indicator to show the effect of

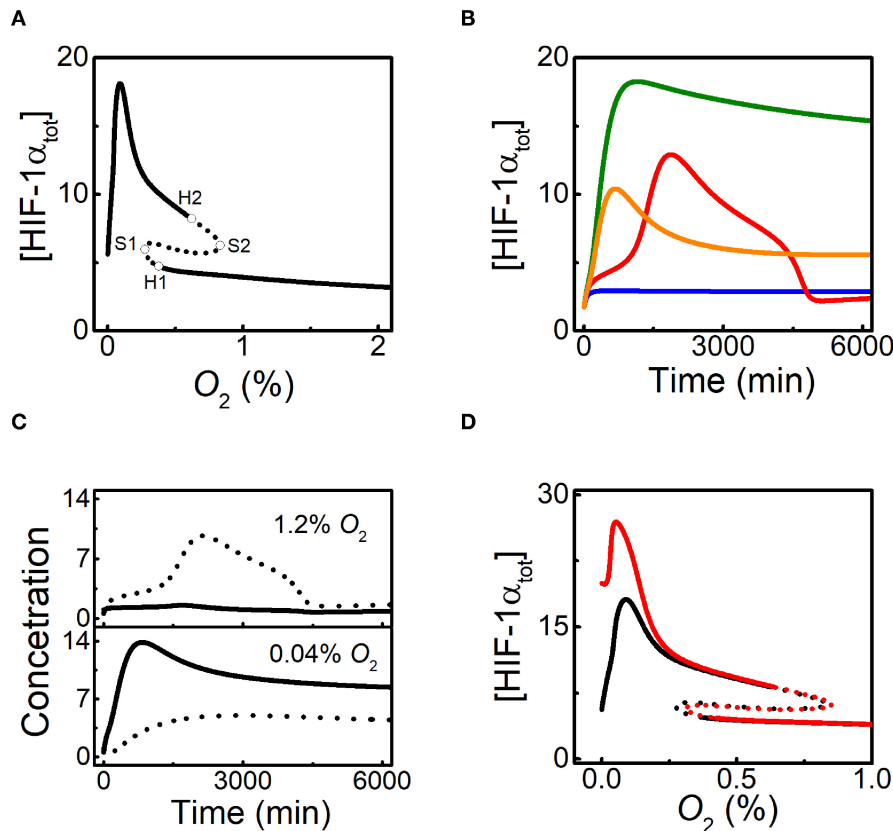


FIGURE 2 | HIF-1 α dynamics under different hypoxic conditions. **(A)** Bifurcation diagrams of $[HIF-1\alpha_{tot}]$ vs. $O_2\%$, in the default parameters setting. The stable and unstable steady states are indicated by solid and dashed lines, respectively. The saddle-node bifurcation points are marked with S1 and S2, while the Hopf bifurcation points are labeled with H1 and H2. **(B)** Time courses of $[HIF-1\alpha_{tot}]$ for 3% (blue), 1% (red), 0.05% (green), and 0% O_2 (orange). The initial values of all the species are set to be their steady states at 21% O_2 in the simulation of the dynamics of the species (the same below). **(C)** Time courses of $[HIF-1\alpha]$ (solid) and $[HIF-1\alpha-aOH]$ (dashed) in moderate hypoxia (1.2% O_2) or severe hypoxia (0.04% O_2). **(D)** Bifurcation diagrams of $[HIF-1\alpha_{tot}]$ vs. O_2 level for k_{dhifH} = 0.0032 (black) or 0 (red). Notably, k_{dhifH} is designated as FIH-1-related degradation rate of HIF-1 α . The types of bifurcation points are similar to **Figure 2B**.

lincRNA-p21 abundance on HIF-1 α dynamics. We find that the induction rate of lincRNA-p21, $k_{sLAP21a}$, affects HIF-1 induction in a switch-like way (**Figure 4B**). For small $k_{sLAP21a}$, HIF-1 α_{peak} keeps rather small. But when $k_{sLAP21a}$ is increased and exceeds some threshold, HIF-1 α_{peak} rises sharply. With further increasing $k_{sLAP21a}$, HIF-1 α_{peak} rises continuously. Thus, the peak of $[HIF-1\alpha_{tot}]$ is remarkably modulated by lincRNA-p21 abundance. To explain these results, we plot the bifurcation diagram of $[HIF-1\alpha_{tot}]$ vs. $k_{sLAP21a}$ with miR-155 knockout at 1% O_2 (**Figure 4B**). Due to the HIF-1 α -lincRNA-p21 positive feedback loop, the steady state of $[HIF-1\alpha_{tot}]$ exhibits bistability with varying $k_{sLAP21a}$ and the threshold of $k_{sLAP21a}$ is very close to that in the curve of HIF-1 α_{peak} . For $k_{sLAP21a}$ exceeding the threshold, the steady state of $[HIF-1\alpha_{tot}]$ is close to HIF-1 α_{peak} and increases monotonically with increasing $k_{sLAP21a}$. Therefore, HIF-1 α is augmented by its target lincRNA-p21 in the rising phase in response to hypoxia.

We further analyze how lincRNA-p21-mediated HIF-1 α stabilization affects its dynamics. The thresholds of lincRNA-p21 for repressing the degradation of proline-hydroxylated and

-unhydroxylated HIF-1 α are represented by $j_{dhifpoh}$ and j_{dhif} , respectively. $[HIF-1\alpha_{tot}]$ remains at rather low levels when j_{dhif} is enlarged markedly (**Figure 4C**). Inhibiting the stabilization of proline-unhydroxylated HIF-1 α by lincRNA-p21 markedly influences $[HIF-1\alpha_{tot}]$ since both HIF-1 α and HIF-1 α -aOH are destabilized remarkably in this case. In contrast, for very large $j_{dhifpoh}$, $[HIF-1\alpha_{tot}]$ still exhibits adaptive dynamics perfectly which is close to the case in the standard parameter setting. The proline-hydroxylated HIF-1 α -aOHpOH is rather unstable due to inhibition of its stabilization via lincRNA-p21. Therefore, HIF-1 α induces lincRNA-p21 to facilitate its own accumulation in the adaptive response to moderate hypoxia.

HIF-1 α adapts to moderate hypoxia while it remains at rather high levels under severe hypoxia (see **Figure 2B**). The effect of lincRNA-p21 induction rate on HIF-1 α dynamics is investigated. Since lincRNA-p21 is mainly induced by un-hydroxylated HIF-1 α in this case, we only consider the influence of the production rate of lincRNA-p21, k_{sLAP21} , on HIF-1 α dynamics (**Figure 4D**). In the default case, HIF-1 α stays at high levels after a slight decrease at 0.1% O_2 . For increased k_{sLAP21} , HIF-1 α_{tot} rises to

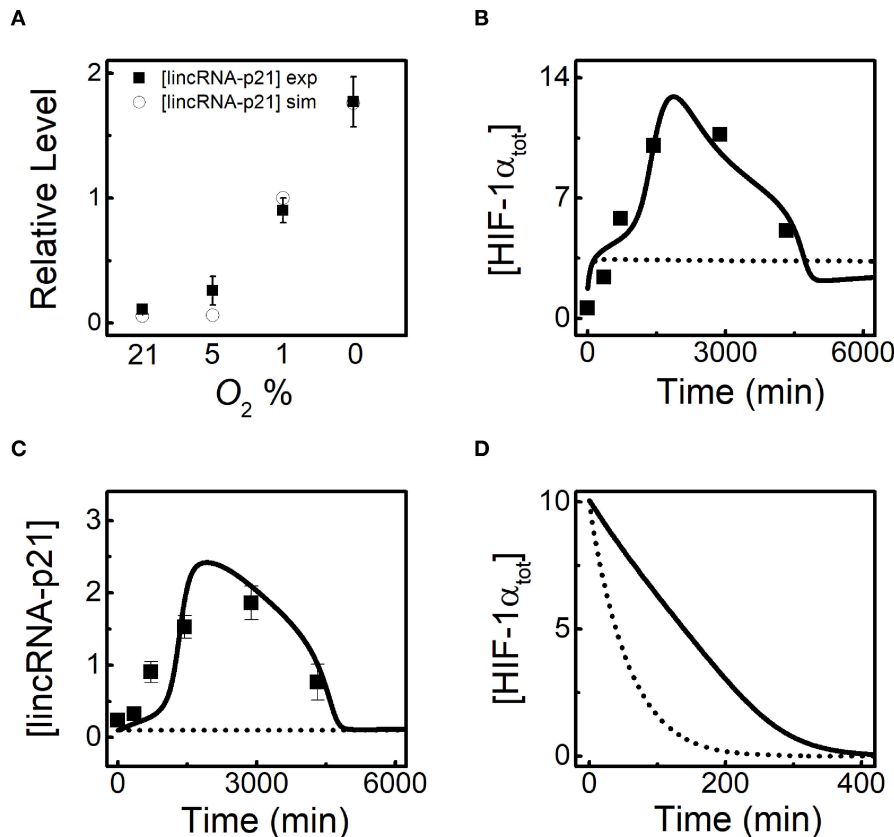


FIGURE 3 | LincRNA-p21 is required for the adaptive dynamics of HIF-1α. **(A)** Comparison of [lincRNA-p21] at 24 h for different O₂ levels between simulation (circle) and experiment (black square). The experimental data are retrieved from Yang et al. (2014). **(B,C)** Time courses of [HIF-1α_{tot}] **(B)** and [lincRNA-p21] **(C)** with normal lincRNA-p21 expression (solid, $k_{sLAP21} = 0.08$ and $k_{sLAP21a} = 0.05$) and knockout (dashed, $k_{sLAP21} = 0$ and $k_{sLAP21a} = 0$) at 1% O₂. Black squares denote the experimental data for [HIF-1α_{tot}] and [lincRNA-p21] with normal lincRNA-p21 expression, retrieved from Yang et al. (2014). **(D)** With inhibited protein synthesis ($k_{thif} = 0$), time courses of [HIF-1α_{tot}] with normal lincRNA-p21 expression (solid, $k_{sLAP21} = 0.08$ and $k_{sLAP21a} = 0.05$) and knockout (dashed, $k_{sLAP21} = 0$ and $k_{sLAP21a} = 0$) under 1% O₂. Of note, the initial state refers to the transient state at 24 h under 1% O₂ in the standard parameter setting.

higher levels. For smaller k_{sLAP21} , [HIF-1α_{tot}] can exhibit an adaptive pulse with a lower amplitude. However, for very small k_{sLAP21} , HIF-1α is hardly induced. Together, lincRNA-p21 is also required for HIF-1 accumulation under severe hypoxia and its induction rate can modulate the dynamic modes of HIF-1α.

3.3. HIF-1α Induces miR-155 to Promote Its Own Recovery in the Late Phase

It has been reported experimentally that HIF-1α-induced miR-155 can promote the recovery of HIF-1α in several cell lines (Bruning et al., 2011). The time courses of [HIF-1α_{tot}], [miR-155], and [HIF-1α_m] in hypoxia (1% O₂) are shown in Figure 5A. HIF-1α_{tot} rises and induces miR-155, which promotes the degradation of HIF-1α mRNA so that both HIF-1α mRNA and HIF-1α_{tot} exhibit adaptive dynamics with some phase difference. As a result, miR-155 also drops to basal levels in the late phase, well consistent with the experimental data (Bruning et al., 2011; Wan et al., 2014). Next we explore the effect of HIF-1α-aOH-dependent induction rate of miR-155 ($k_{smiR1551a}$) on the adaptive response of HIF-1α to hypoxia (Figure 5B). For decreased $k_{smiR1551a}$, the adaptive property of

[HIF-1α_{tot}] dynamics weakens remarkably: HIF-1α_{tot} reaches a slightly higher peak and maintains at rather high levels finally. For increased $k_{smiR1551a}$, [HIF-1α_{tot}] still exhibits adaptive dynamics while the peak and the width of the dynamic curves reduce gradually and [HIF-1α_{tot}] only rises slightly for very large $k_{smiR1551a}$ (Figure 5B). Together, miR-155 can facilitate the recovery of HIF-1α levels in the late phase.

In addition to miR-155, HIF-1α-induced PHD2 may also contribute to the adaptation of HIF-1α to hypoxia (Stiehl et al., 2006; Henze et al., 2010; Bagnall et al., 2014). We further explore the potential interplay between miR-155 and PHD2 in regulating HIF-1α dynamics (Figure 5C). Knockout of either miR-155 or PHD2 is mimicked by setting the corresponding HIF-1α-dependent induction rates to zero. At 1% O₂, [HIF-1α_{tot}] still shows adaptive dynamics with a higher peak in the case of PHD2 knockout. With miR-155 knockout, [HIF-1α_{tot}] settles at a plateau instead of showing adaptation, consistent with the experimental results (Bruning et al., 2011). In the case of both miR-155 and PHD2 knockout, [HIF-1α_{tot}] stays at higher levels persistently compared to the case of miR-155 knockout alone. Therefore, our results suggest that miR-155 is required for the

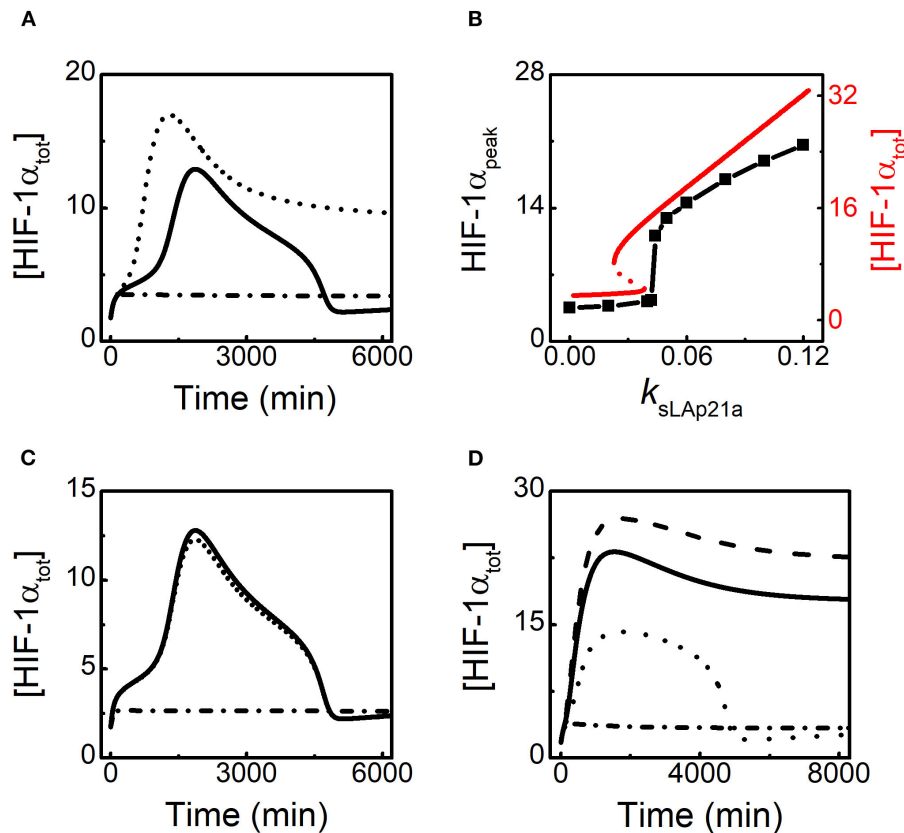


FIGURE 4 | HIF-1 α induces lincRNA-p21 to promote its own accumulation in hypoxia. **(A)** Time courses of [HIF-1 α_{tot}] at 1% O₂ for different HIF-1 α -aOH-dependent lincRNA-p21 induction rates: $k_{sLAp21a} = 0.08$ (dashed), 0.05 (solid, default), and 0 (dash-dotted). **(B)** The curve of HIF-1 α_{peak} (black) vs. $k_{sLAp21a}$ at 1% O₂; the bifurcation diagram of [HIF-1 α_{tot}] (red) vs. $k_{sLAp21a}$ at 1% O₂ with miR-155 knockout ($k_{smiR1551} = 0$ and $k_{smiR1551a} = 0$). HIF-1 α_{peak} denotes the maximal of [HIF-1 α_{tot}] on individual simulation trials. **(C)** Time courses of [HIF-1 α_{tot}] at 1% O₂ in the following cases: $j_{dhif} = 0.3$ and $j_{dhif} = 0.3$ (solid); increasing j_{dhif} to 100 (dashed); increasing j_{dhif} to 100 (dash-dotted). **(D)** Time courses of [HIF-1 α_{tot}] at 0.1% O₂ for different HIF-1 α -dependent lincRNA-p21 synthesis rates: $k_{sLAp21} = 0.12$ (dashed), 0.08 (solid, default), 0.035 (dotted), and 0 (dash-dotted).

adaptive dynamics of HIF-1 α while PHD2 mainly contributes to the suppression of HIF-1 α accumulation.

It is intriguing to investigate the effect of miR-155 abundance on HIF-1 α dynamics in severe hypoxia. We show the above effect of miR-155 by plotting the curves of [HIF-1 α_{tot}] dynamics for various $k_{smiR1551}$ (the induction rate of miR-155 by unhydroxylated HIF-1 α) at 0.1% O₂ (Figure 5D). As shown previously, HIF-1 α_{tot} stays at high levels in the default case. For smaller $k_{smiR1551}$, [HIF-1 α_{tot}] rises to higher levels; for larger $k_{smiR1551}$, [HIF-1 α_{tot}] stays at lower levels; for further increased $k_{smiR1551}$, [HIF-1 α_{tot}] cannot keep at high levels and drops to basal levels, exhibiting adaptive dynamics (Figure 5D). Together, miR-155 modulates HIF-1 α dynamic modes markedly under severe hypoxia, and its overexpression can transform HIF-1 α dynamics to adaptive mode.

3.4. Crosstalk of lincRNA-p21 and miR-155 in Shaping HIF-1 α Dynamics

Since both lincRNA-p21 and miR-155 are involved in HIF-1 α -centered feedback loops, there may exist crosstalk between them in modulating HIF-1 α dynamics. The dynamic curves of

[lincRNA-p21], [miR-155], and [HIF-1 α_{tot}] are plotted together to show their temporal evolution during the hypoxic response (Figure 6A). LincRNA-p21 and miR-155 are separately induced by HIF-1 α so that lincRNA-p21 stimulates the rising of HIF-1 α in the early phase while miR-155 renders the recovery of HIF-1 α in the late phase. With decreased $k_{sLAp21a}$, lincRNA-p21 settles at low state instead of showing adaptive dynamics (Figure 6B). As a result, [HIF-1 α_{tot}] stays at rather low levels and miR-155 is not induced markedly without enough HIF-1 α . Our results reveal that sufficient induction of lincRNA-p21 by HIF-1 α is required for subsequent induction of miR-155 and recovery of HIF-1 α .

We further explore how lincRNA-p21 and miR-155 interplay to modulate the adaptive dynamics of HIF-1 α . We compare the dynamics of [HIF-1 α_{tot}] by increasing the induction rates of lincRNA-p21 and (or) miR-155 (Figure 6C). For increased $k_{sLAp21a}$, lincRNA-p21 counteracts the recovery of HIF-1 α by miR-155, and [HIF-1 α_{tot}] maintains at moderate levels in the late phase. When $k_{smiR1551a}$ is also enlarged, [HIF-1 α_{tot}] restores perfect adaptation to hypoxia. Likewise, with increasing $k_{smiR1551a}$ alone, overexpressed miR-155 suppresses the stabilization of HIF-1 α by lincRNA-p21 and [HIF-1 α_{tot}] only

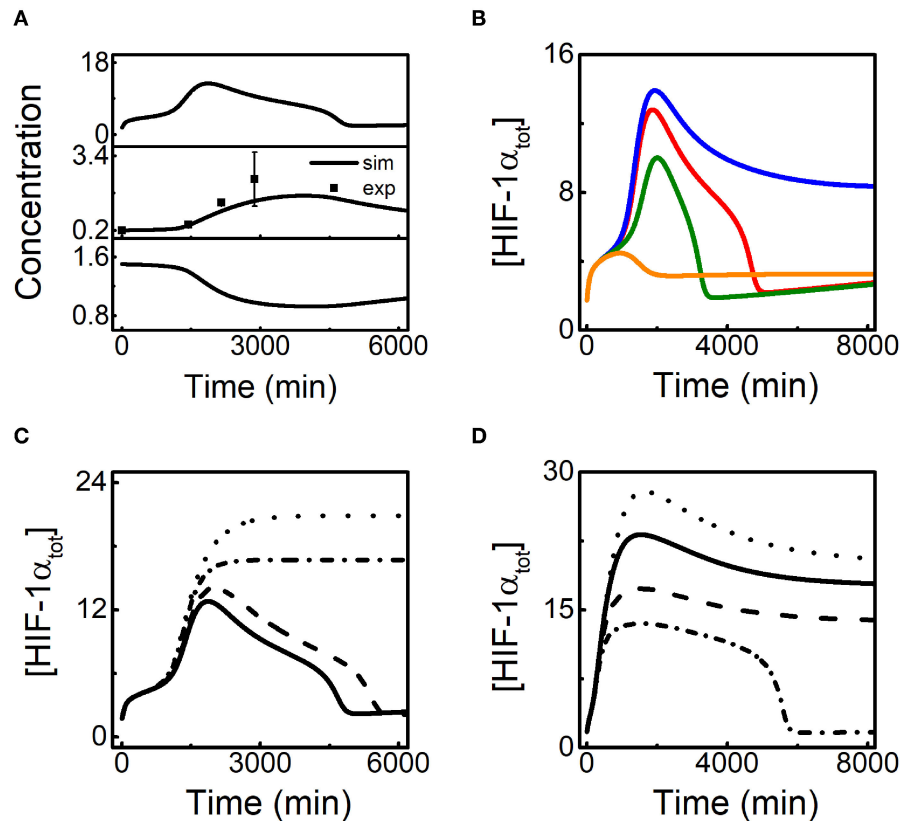


FIGURE 5 | HIF-1 α -induced miR-155 promotes the recovery of HIF-1 α in the late phase of the adaptive response. **(A)** Time courses of [HIF-1 α_{tot}], [miR-155], and [HIF-1 α_m] from top to bottom at 1% O₂, and the experimental data (retrieved from Bruning et al., 2011; Wan et al., 2014) are denoted by the black squares. **(B)** Time courses of [HIF-1 α_{tot}] at 1% O₂ for different HIF-1 α -aOH-dependent miR-155 induction rates: $k_{smiR1551a} = 0.0006$ (blue), 0.0012 (red, default), 0.0036 (green), and 0.007 (brown). **(C)** Time courses of [HIF-1 α_{tot}] at 1% O₂ with miR-155 and PHD2 double knockout (dotted, $k_{smiR1551} = 0$, $k_{smiR1551a} = 0$, $k_{sphd1} = 0$, and $k_{sphd1a} = 0$), miR-155 knockout (dash-dotted, $k_{smiR1551} = 0$ and $k_{smiR1551a} = 0$), PHD2 knockout (dashed, $k_{sphd1} = 0$ and $k_{sphd1a} = 0$), and normal miR-155 and PHD2 expression. **(D)** Time courses of [HIF-1 α_{tot}] at 0.1% O₂ for different HIF-1 α -dependent miR-155 induction rates: $k_{smiR1551} = 0.0006$ (dotted), 0.0024 (solid, default), 0.012 (dashed), and 0.04 (dash-dotted).

shows a small pulse. We also show the effects of decreasing $k_{sLap21a}$ and/or $k_{smiR1551a}$ on HIF-1 α dynamics (**Figure 6D**). When $k_{smiR1551a}$ alone is decreased, [HIF-1 α_{tot}] stays at moderate levels; while decreasing $k_{sLap21a}$ alone, it only rises to low levels; adaptive dynamics of [HIF-1 α_{tot}] reappears in the case of decreasing both $k_{sLap21a}$ and $k_{smiR1551a}$. Together, lincRNA-p21 and miR-155 cooperate in shaping HIF-1 α dynamics and their balance is critical for the perfect adaptation of HIF-1 α to hypoxia.

4. CONCLUSION AND DISCUSSION

We have built a network model to probe how HIF-1 α -targeted lincRNA-p21 and miR-155 coordinate to regulate the adaption of HIF-1 α to hypoxia. We found that lincRNA-p21 and miR-155 are sequentially induced during hypoxia. LincRNA-p21 promotes the rising of HIF-1 α by stabilizing it in the early phase, while miR-155 promotes the recovery of HIF-1 α in the late phase by facilitating the degradation of its mRNA. Moreover, there exists a delicate balance between lincRNA-p21 and miR-155 in shaping HIF-1 α dynamics: variation in the adaptive dynamics of HIF-1 α due to

changes in the expression of either ncRNA can be counteracted by changing the expression of the other.

It has been shown that both HIF-1 α and lincRNA-p21 exhibit adaptive dynamics in response to hypoxia (Yang et al., 2014). How they are down-regulated in the late phase of the response is not well-elucidated. We proposed that HIF-1 α -induced miR-155 may contribute to the recovery of HIF-1 α and lincRNA-p21 to low levels. In addition, we also reveal that lincRNA-p21 promotes the rising of HIF-1 α in the early phase via stabilizing it. Furthermore, we indicate that the stabilization of partially activated proline-unhydroxylated HIF-1 α by lincRNA-p21 plays a dominant role in promoting HIF-1 α accumulation. Therefore, coordination of HIF-1 α -centered positive and negative feedback loops ensures the adaptive adaption of HIF-1 α to hypoxia.

We assumed that lincRNA-p21 can repress the degradation of proline-unhydroxylated HIF-1 α . The assumption is supported by the experimental evidence that HIF-1 α induces lincRNA-p21 to stabilize the proline-unhydroxylated HIF-1 α by disrupting the interaction with VHL and enhances its own transcriptional

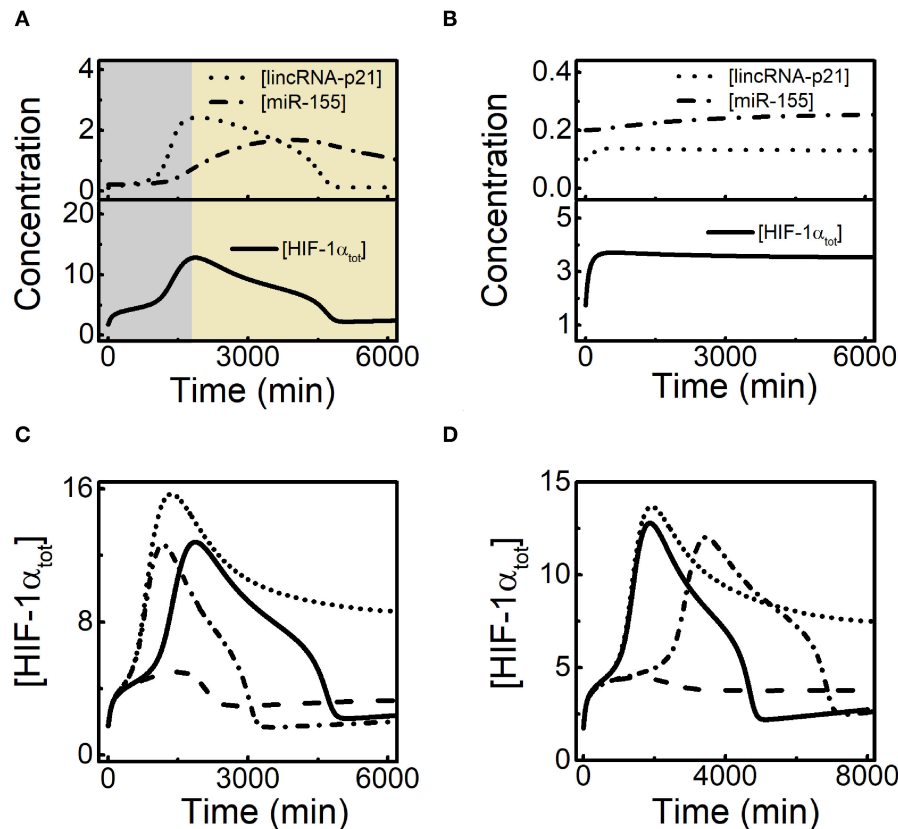


FIGURE 6 | Interplay between lincRNA-p21 and miR-155 in the adaptive dynamics of HIF-1 α . **(A)** Time courses of [lincRNA-p21] (dotted), [miR-155] (dash-dotted), and [HIF-1 α_{tot}] (solid) at 1% O_2 in the default case. **(B)** Time courses of [lincRNA-p21] (dotted), [miR-155] (dash-dotted), and [HIF-1 α_{tot}] (solid) at 1% O_2 with decreased lincRNA-p21 induction rate ($k_{sLAp21a} = 0.02$). **(C)** Time courses of [HIF-1 α_{tot}] at 1% O_2 with the standard parameter setting (solid, $k_{sLAp21a} = 0.05$ and $k_{smiR1551a} = 0.0012$), increased lincRNA-p21 induction rate alone (dotted, $k_{sLAp21a} = 0.07$, $k_{smiR1551a} = 0.0012$), increased miR-155 expression alone (dashed, $k_{sLAp21a} = 0.05$ and $k_{smiR1551a} = 0.0005$), and both increased lincRNA-p21 and miR-155 expression (dash-dotted, $k_{sLAp21a} = 0.07$ and $k_{smiR1551a} = 0.0005$). **(D)** Time courses of [HIF-1 α_{tot}] at 1% O_2 for the default case (solid, $k_{sLAp21a} = 0.05$ and $k_{smiR1551a} = 0.0012$), decreased miR-155 expression alone (dotted, $k_{sLAp21a} = 0.05$ and $k_{smiR1551a} = 0.0007$), decreased lincRNA-p21 expression alone (dashed, $k_{sLAp21a} = 0.043$ and $k_{smiR1551a} = 0.0012$), and both decreased lincRNA-p21 and miR-155 expression (dash-dotted, $k_{sLAp21a} = 0.043$ and $k_{smiR1551a} = 0.0007$).

activity remarkably (Yang et al., 2014). It has been reported that two forms of proline-unhydroxylated HIF-1 α , i.e., unhydroxylated HIF-1 α and HIF-1 α with asparaginyl-hydroxylation alone, can transactivate the target genes (Dayan et al., 2006). Moreover, given proline-hydroxylated HIF-1 α loses its transcriptional activity and prolyl-hydroxylation is considered to be irreversible (Schofield and Ratcliffe, 2004; Chan et al., 2005), lincRNA-p21-dependent accumulation of proline-hydroxylated HIF-1 α has no contribution to induction of target genes. Therefore, it is plausible to assume that lincRNA-p21 enhances the transcriptional activity of HIF-1 α by stabilizing the proline-unhydroxylated forms. It has been reported that lincRNA-p21 promotes HIF-1 α -dependent glycolysis via inducing several target genes (Yang et al., 2014). Nevertheless, the interaction between HIF-1 α and VHL is rather weak, and the detailed mechanism underlying the upregulation of HIF-1 α activity by lincRNA p21 is to be further investigated.

Our work reveals that miR-155 and PHD2 play non-redundant roles in promoting HIF-1 α recovery in hypoxia. Our results show that PHD2 mainly modulates the peak level of HIF-1 α in the adaptive dynamics, consistent with the modeling results reported by Fábíán et al. (2016). On the other hand, it has been suggested that PHD2 mainly regulates HIF-1 α in the early phase in contrast to miR-155 (Bruning et al., 2011). We proposed that HIF-1 α -miR-155 and HIF-1 α -PHD2 negative feedback loops play the main and auxiliary role respectively in adaptive response of HIF-1 α to hypoxia. Nevertheless, PHD2 may play a significant role in the adaptive dynamics of HIF-1 α in some other cell lines (Ginouvé et al., 2008; Bagnall et al., 2014). Therefore, the roles of PHD2 and miR-155 may be context-dependent or cell-type specific.

LincRNA-p21 and miR-155 regulate HIF-1 α positively and negatively in the separate phases to ensure the adaptive dynamics. They should be induced and predominate in the early and late phase of the response, respectively (Bruning et al., 2011; Yang

et al., 2014). There may exist a delicate balance between lincRNA-p21 and miR-155 in shaping HIF-1 α dynamics. For increased lincRNA-p21 expression, dominance of miR-155 is weakened in the late phase so that HIF-1 α cannot return to low levels entirely. Increased miR-155 expression also impairs the balance between it and lincRNA-p21, suppressing HIF-1 α accumulation. Therefore, when the expression of either one is changed, the other needs to be varied in the same direction to guarantee the perfect adaptation of HIF-1 α to hypoxia. In addition, under severe hypoxia, enhancing lincRNA-p21 induction impairs the balance in the regulation of HIF-1 α , so HIF-1 α settles down at relatively high levels. We propose that repressing lincRNA-p21 or increasing miR-155 expression may facilitate the adaptation of HIF-1 α under serious hypoxia.

It has been indicated that tight control of transient HIF-1 α dynamics is crucial for cell survival (Bagnall et al., 2014). LincRNA-p21 and miR-155 have the potential to modulate cellular outcome in the hypoxic response since they can modulate the dynamic modes of HIF-1 α . On one hand, knocking down lincRNA-p21 may be a rational strategy for repressing tumorigenesis as lincRNA-p21 can promote HIF-1 α accumulation and facilitate the adaptation of tumors to hypoxia (Yang et al., 2014; Koyasu et al., 2018). Indeed, lincRNA-p21 knockdown induces G2/M phase arrest and promotes apoptosis to enhance the radiosensitivity of SMMC7721 and U251MG cells in hypoxia (Shen et al., 2017). Moreover, it has been reported that lincRNA-p21 knockout abrogates the migration and survival of mesenchymal stem cells by hypoxia preconditioning (Meng et al., 2018). On the other hand, miR-155 mediates proliferation suppression of non-small cell lung cancer cells in radiotherapy via decreasing the expression of HIF-1 α (Zhu et al., 2019). miR-155 deficiency results in upregulated HIF-1 α expression, promoting the growth of solid tumors (Wang et al., 2015). In addition, a recent study showed that targeting the temporal dynamics of HIF-1 α -induced tumor-secreted microenvironmental factors can halt tumor migration (Singh et al., 2019). Lehmann et al. revealed that hypoxia can induce a HIF-1 α -dependent transition

from collective-to-amoeboid dissemination in epithelial cancer cells (Lehmann et al., 2017). Kang et al. proposed that HIF-1 α and several microRNAs including miR-34, miR-145, and miR-200, may play critical roles in epithelial to mesenchymal transition and cancer metastasis (Kang et al., 2019). Thus, we predict that modulating HIF-1 α dynamics or activity by changing lincRNA-p21 or miR-155 expression may affect cancer migration and dissemination remarkably. Together, more attention should be paid to the treatment strategy of cancer by targeting lincRNA-p21 and miR-155 in the future.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

X-PZ conceived and designed the research. C-YS carried out computer implementation and wrote the original manuscript. C-YS and X-PZ analyzed and interpreted the simulation results. X-PZ and FL supervised the project. C-YS, X-PZ, FL, and WW revised and contributed to the writing of final manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China (Nos. 11574139, 11547025, 11874209, and 11934008).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00871/full#supplementary-material>

REFERENCES

- Bagnall, J., Leedale, J., Taylor, S. E., Spiller, D. G., White, M. R., Sharkey, K. J., et al. (2014). Tight control of hypoxia-inducible factor- α transient dynamics is essential for cell survival in hypoxia. *J. Biol. Chem.* 289, 5549–5564. doi: 10.1074/jbc.M113.500405
- Berra, E., Benizri, E., Ginouvés, A., Volmat, V., Roux, D., and Pouyssegur, J. (2003). HIF prolyl-hydroxylase 2 is the key oxygen sensor setting low steady-state levels of HIF-1 α in normoxia. *EMBO J.* 22, 4082–4090. doi: 10.1093/emboj/cdg392
- Bruning, U., Cerone, L., Neufeld, Z., Fitzpatrick, S. F., Cheong, A., Scholz, C. C., et al. (2011). MicroRNA-155 promotes resolution of hypoxia-inducible factor 1 α activity during prolonged hypoxia. *Mol. Cell Biol.* 31, 4087–4096. doi: 10.1128/MCB.01276-10
- Chan, D. A., Sutphin, P. D., Yen, S. E., and Giaccia, A. J. (2005). Coordinate regulation of the oxygen-dependent degradation domains of hypoxia-inducible factor 1 α . *Mol. Cell Biol.* 25, 6415–6426. doi: 10.1128/MCB.25.15.6415-6426.2005
- Chan, M. C., Ilott, N. E., Schödel, J., Sims, D., Tumber, A., Lippl, K., et al. (2016). Tuning the transcriptional response to hypoxia by inhibiting hypoxia-inducible factor (HIF) prolyl and asparaginyl hydroxylases. *J. Biol. Chem.* 291, 20661–20673. doi: 10.1074/jbc.M116.749291
- Chan, S. Y., Zhang, Y. Y., Hemann, C., Mahoney, C. E., Zweier, J. L., and Loscalzo, J. (2009). MicroRNA-210 controls mitochondrial metabolism during hypoxia by repressing the iron-sulfur cluster assembly proteins ISCU1/2. *Cell Metab.* 10, 273–284. doi: 10.1016/j.cmet.2009.08.015
- Chang, Y. N., Zhang, K., Hu, Z. M., Qi, H. X., Shi, Z. M., Han, X. H., et al. (2016). Hypoxia-regulated lncRNAs in cancer. *Gene* 575, 1–8. doi: 10.1016/j.gene.2015.08.049
- Dayan, F., Monticelli, M., Pouyssegur, J., and Pécou, E. (2009). Gene regulation in response to graded hypoxia: the non-redundant roles of the oxygen sensors PHD and FIH in the HIF pathway. *J. Theor. Biol.* 259, 304–316. doi: 10.1016/j.jtbi.2009.03.009
- Dayan, F., Roux, D., Brahimi-Horn, M. C., Pouyssegur, J., and Mazure, N. M. (2006). The oxygen sensor factor-inhibiting hypoxia-inducible factor-1 controls expression of distinct genes through the bifunctional transcriptional character of hypoxia-inducible factor-1 α . *Cancer Res.* 66, 3688–3698. doi: 10.1158/0008-5472.CAN-05-4564
- Fábián, Z., Taylor, C. T., and Nguyen, L. K. (2016). Understanding complexity in the HIF signaling pathway using systems biology and

- mathematical modeling. *J. Mol. Med.* 94, 377–390. doi: 10.1007/s00109-016-1383-6
- Genoué, A., Ilc, K., Macías, N., Pouyssegur, J., and Berra, E. (2008). PHDs overactivation during chronic hypoxia “desensitizes” HIF and protects cells from necrosis. *Proc. Natl. Acad. Sci. U.S.A.* 105, 4745–4750. doi: 10.1073/pnas.0705680105
- Harada, H., Kizaka-Kondoh, S., Li, G., Itasaka, S., Shibuya, K., Inoue, M., et al. (2007). Significance of HIF-1-active cells in angiogenesis and radioresistance. *Oncogene* 26, 7508–7516. doi: 10.1038/sj.onc.1210556
- Henze, A. T., Riedel, J., Diem, T., Wenner, J., Flamme, I., Pouyssegur, J., et al. (2010). Prolyl hydroxylases 2 and 3 act in gliomas as protective negative feedback regulators of hypoxia-inducible factors. *Cancer Res.* 70, 357–366. doi: 10.1158/0008-5472.CAN-09-1876
- Ivan, M., Kondo, K., Yang, H., Kim, W., Valiando, J., Ohh, M., et al. (2001). HIF α targeted for VHL-mediated destruction by proline hydroxylation: implications for O₂ sensing. *Science* 292, 464–468. doi: 10.1126/science.1059817
- Jaakkola, P., Mole, D. R., Tian, Y. M., Wilson, M. I., Gielbert, J., Gaskell, S. J., et al. (2001). Targeting of HIF- α to the von Hippel-Lindau ubiquitylation complex by O₂-regulated prolyl hydroxylation. *Science* 292, 468–472. doi: 10.1126/science.1059796
- Jiang, B.-H., Semenza, G. L., Bauer, C., and Marti, H. H. (1996). Hypoxia-inducible factor 1 levels vary exponentially over a physiologically relevant range of O₂ tension. *Am. J. Physiol.* 271, C1172–C1180. doi: 10.1152/ajpcell.1996.271.4.C1172
- Kang, X., Wang, J., and Li, C. (2019). Exposing the underlying relationship of cancer metastasis to metabolism and epithelial-mesenchymal transitions. *iScience* 21, 754–772. doi: 10.1016/j.isci.2019.10.060
- Koh, M. Y., and Powis, G. (2012). Passing the baton: the HIF switch. *Trends Biochem. Sci.* 37, 364–372. doi: 10.1016/j.tibs.2012.06.004
- Kohn, K. W., Riss, J., Aprelikova, O., Weinstein, J. N., Pommier, Y., and Barrett, J. C. (2004). Properties of switch-like bioregulatory networks studied by simulation of the hypoxia response control system. *Mol. Biol. Cell* 15, 3042–3052. doi: 10.1091/mbc.e03-12-0897
- Koivunen, P., Hirsilä, M., Günzler, V., Kivirikko, K. I., and Myllyharju, J. (2004). Catalytic properties of the asparaginyl hydroxylase (FIH) in the oxygen sensing pathway are distinct from those of its prolyl 4-hydroxylases. *J. Biol. Chem.* 279, 9899–9904. doi: 10.1074/jbc.M312254200
- Koyasu, S., Kobayashi, M., Goto, Y., Hiraoka, M., and Harada, H. (2018). Regulatory mechanisms of hypoxia-inducible factor 1 activity: two decades of knowledge. *Cancer Sci.* 109, 560–571. doi: 10.1111/cas.13483
- Lando, D., Peet, D. J., Gorman, J. J., Whelan, D. A., Whitelaw, M. L., and Bruck, R. K. (2002). FIH-1 is an asparaginyl hydroxylase enzyme that regulates the transcriptional activity of hypoxia-inducible factor. *Gene Dev.* 16, 1466–1471. doi: 10.1101/gad.991402
- Lehmann, S., Te Boekhorst, V., Odenthal, J., Bianchi, R., van Helvert, S., Ikenberg, K., et al. (2017). Hypoxia induces a HIF-1-dependent transition from collective-to-amoeboid dissemination in epithelial cancer cells. *Curr. Biol.* 27, 392–400. doi: 10.1016/j.cub.2016.11.057
- Li, Z., Meng, D., Li, G., Xu, J., Tian, K., and Li, Y. (2016). Overexpression of microRNA-210 promotes chondrocyte proliferation and extracellular matrix deposition by targeting HIF-3 α in osteoarthritis. *Mol. Med. Rep.* 13, 2769–2776. doi: 10.3892/mmr.2016.4878
- Mahon, P. C., Hirota, K., and Semenza, G. L. (2001). FIH-1: a novel protein that interacts with HIF-1 α and VHL to mediate repression of HIF-1 transcriptional activity. *Gene Dev.* 15, 2675–2686. doi: 10.1101/gad.924501
- Maxwell, P. H., Pugh, C. W., and Ratcliffe, P. J. (2001). Activation of the HIF pathway in cancer. *Curr. Opin. Genet. Dev.* 11, 293–299. doi: 10.1016/S0959-437X(00)00193-3
- Meng, S.-S., Xu, X.-P., Chang, W., Lu, Z.-H., Huang, L.-L., Xu, J.-Y., et al. (2018). LincRNA-p21 promotes mesenchymal stem cell migration capacity and survival through hypoxic preconditioning. *Stem Cell Res. Ther.* 9:280. doi: 10.1186/s13287-018-1031-x
- Minamishima, Y. A., Moslehi, J., Padera, R. F., Bronson, R. T., Liao, R., and Kaelin, W. G. J. (2009). A feedback loop involving the Phd3 prolyl hydroxylase tunes the mammalian hypoxic response *in vivo*. *Mol. Cell Biol.* 29, 5729–5741. doi: 10.1128/MCB.00331-09
- Nguyen, L. K., Cavadas, M. A., Scholz, C. C., Fitzpatrick, S. F., Bruning, U., Cummins, E. P., et al. (2013). A dynamic model of the hypoxia-inducible factor 1 α (HIF-1 α) network. *J. Cell Sci.* 126(Pt 6), 1454–1463. doi: 10.1242/jcs.119974
- Ohh, M., Park, C. W., Ivan, M., Hoffman, M. A., Kim, T. Y., Huang, L. E., et al. (2000). Ubiquitination of hypoxia-inducible factor requires direct binding to the β -domain of the von Hippel-Lindau protein. *Nat. Cell Biol.* 2, 423–427. doi: 10.1038/35017054
- Qutub, A. A., and Popel, A. S. (2006). A computational model of intracellular oxygen sensing by hypoxia-inducible factor HIF1 α . *J. Cell Sci.* 119(Pt 16), 3467–3480. doi: 10.1242/jcs.03087
- Schofield, C. J., and Ratcliffe, P. J. (2004). Oxygen sensing by HIF hydroxylases. *Nat. Rev. Mol. Cell Biol.* 5, 343–354. doi: 10.1038/nrm1366
- Semenza, G. L. (2009). Regulation of cancer cell metabolism by hypoxia-inducible factor 1. *Semin. Cancer Biol.* 19, 12–16. doi: 10.1016/j.semcancer.2008.11.009
- Semenza, G. L. (2012). Molecular mechanisms mediating metastasis of hypoxic breast cancer cells. *Trends Mol. Med.* 18, 534–543. doi: 10.1016/j.molmed.2012.08.001
- Serocki, M., Bartoszewski, S., Janaszak-Jasiecka, A., Ochocka, R. J., Collawn, J. F., and Bartoszewski, R. (2018). miRNAs regulate the HIF switch during hypoxia: a novel therapeutic target. *Angiogenesis* 21, 183–202. doi: 10.1007/s10456-018-9600-2
- Shen, Y., Liu, Y., Sun, T., and Yang, W. (2017). LincRNA-p21 knockdown enhances radiosensitivity of hypoxic tumor cells by reducing autophagy through HIF-1/Akt/mTOR/P70S6K pathway. *Exp. Cell Res.* 358, 188–198. doi: 10.1016/j.yexcr.2017.06.016
- Singh, M., Tian, X. J., Donnenberg, V. S., Watson, A. M., Zhang, J., Stabile, L. P., et al. (2019). Targeting the temporal dynamics of hypoxia-induced tumor-secreted factors halts tumor migration. *Cancer Res.* 79, 2962–2977. doi: 10.1158/0008-5472.CAN-18-3151
- Stiehl, D. P., Wirthner, R., Köditz, J., Spielmann, P., Camenisch, G., and Wenger, R. H. (2006). Increased prolyl 4-hydroxylase domain proteins compensate for decreased oxygen levels. Evidence for an autoregulatory oxygen-sensing system. *J. Biol. Chem.* 281, 23482–23491. doi: 10.1074/jbc.M601719200
- Takeda, K., Ho, V. C., Takeda, H., Duan, L.-J., Nagy, A., and Fong, G.-H. (2006). Placental but not heart defects are associated with elevated hypoxia-inducible factor α levels in mice lacking prolyl hydroxylase domain protein 2. *Mol. Cell Biol.* 26, 8336–8346. doi: 10.1128/MCB.00425-06
- Wan, G., Xie, W., Liu, Z., Xu, W., Lao, Y., Huang, N., et al. (2014). Hypoxia-induced MIR155 is a potent autophagy inducer by targeting multiple players in the MTOR pathway. *Autophagy* 10, 70–79. doi: 10.4161/aut.26534
- Wang, G. L., Jiang, B.-H., Rue, E. A., and Semenza, G. L. (1995). Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O₂ tension. *Proc. Natl. Acad. Sci. U.S.A.* 92, 5510–5514. doi: 10.1073/pnas.92.12.5510
- Wang, J., Yu, F., Jia, X., Iwanowycz, S., Wang, Y., Huang, S., et al. (2015). MicroRNA-155 deficiency enhances the recruitment and functions of myeloid-derived suppressor cells in tumor microenvironment and promotes solid tumor growth. *Int. J. Cancer* 136, E602–E613. doi: 10.1002/ijc.29151
- Wang, P., Guan, D., Zhang, X. P., Liu, F., and Wang, W. (2019). Modeling the regulation of p53 activation by HIF-1 upon hypoxia. *FEBS Lett.* 593, 2596–2611. doi: 10.1002/1873-3468.13525
- Xie, S., Chen, H., Li, F., Wang, S., and Guo, J. (2015). Hypoxia-induced microRNA-155 promotes fibrosis in proximal tubule cells. *Mol. Med. Rep.* 11, 4555–4560. doi: 10.3892/mmr.2015.3327
- Yang, F., Zhang, H., Mei, Y., and Wu, M. (2014). Reciprocal regulation of HIF-1 α and LincRNA-p21 modulates the Warburg effect. *Mol. Cell* 53, 88–100. doi: 10.1016/j.molcel.2013.11.004
- Ye, X.-W., Zhang, X.-P., and Liu, F. (2019). CSB modulates the competition between HIF-1 and p53 upon hypoxia. *Math. Biosci. Eng.* 16, 5247–5262. doi: 10.3934/mbe.2019262

- Zeng, L., Morinibu, A., Kobayashi, M., Zhu, Y., Wang, X., Goto, Y., et al. (2015). Aberrant IDH3 α expression promotes malignant tumor growth by inducing HIF-1-mediated metabolic reprogramming and angiogenesis. *Oncogene* 34, 4758–4766. doi: 10.1038/onc.2014.411
- Zhou, C. H., Zhang, X. P., Liu, F., and Wang, W. (2015). Modeling the interplay between the HIF-1 and p53 pathways in hypoxia. *Sci. Rep.* 5:13834. doi: 10.1038/srep13834
- Zhu, L., Xue, F., Cui, Y., Liu, S., Li, G., Li, J., et al. (2019). miR-155-5p and miR-760 mediate radiation therapy suppressed malignancy of non-small cell lung cancer cells. *Biofactors* 45, 393–400. doi: 10.1002/biof.1500

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Sun, Zhang, Liu and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



A Systems Biology Approach to Identifying a Master Regulator That Can Transform the Fast Growing Cellular State to a Slowly Growing One in Early Colorectal Cancer Development Model

Jihye Choi, Jeong-Ryeol Gong, Chae Young Hwang, Chang Young Joung, Soobeom Lee and Kwang-Hyun Cho*

OPEN ACCESS

Edited by:

Chunhe Li,
Fudan University, China

Reviewed by:

Suoqin Jin,
University of California, Irvine,
United States
Yong Wang,
Academy of Mathematics and
Systems Science (CAS), China

*Correspondence:

Kwang-Hyun Cho
ckh@kaist.ac.kr

Specialty section:

This article was submitted to
Systems Biology,
a section of the journal
Frontiers in Genetics

Received: 08 June 2020

Accepted: 10 September 2020

Published: 08 October 2020

Citation:

Choi J, Gong J-R, Hwang CY,
Joung CY, Lee S and Cho K-H (2020)
A Systems Biology Approach to
Identifying a Master Regulator That
Can Transform the Fast Growing
Cellular State to a Slowly Growing
One in Early Colorectal Cancer
Development Model.
Front. Genet. 11:570546.
doi: 10.3389/fgene.2020.570546

Laboratory for Systems Biology and Bio-Inspired Engineering, Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea

Colorectal cancer (CRC) has been most extensively studied for characterizing genetic mutations along its development. However, we still have a poor understanding of CRC initiation due to limited measures of its observation and analysis. If we can unveil CRC initiation events, we might identify novel prognostic markers and therapeutic targets for early cancer detection and prevention. To tackle this problem, we establish the early CRC development model and perform transcriptome analysis of its single cell RNA-sequencing data. Interestingly, we find two subtypes, fast growing vs. slowly growing populations of distinct growth rate and gene signatures, and identify CCDC85B as a master regulator that can transform the cellular state of fast growing subtype cells into that of slowly growing subtype cells. We further validate this by *in vitro* experiments and suggest CCDC85B as a novel potential therapeutic target that may prevent malignant CRC development by suppressing stemness and uncontrolled cell proliferation.

Keywords: colorectal cancer, adenomatous polyposis coli, single cell transcriptomics, gene regulatory network, CCDC85B, systems biology, master regulator analysis

INTRODUCTION

Colorectal cancer (CRC) has been most extensively studied for characterizing genetic mutations along its development. Loss of adenomatous polyposis coli (APC) is considered as the first step of CRC development, which is followed by mutations of other driver genes such as KRAS and TP53 (Fearon and Vogelstein, 1990; Powell et al., 1992). Gene alterations of APC abrogate its binding with β -catenin and result in β -catenin release, which in turn brings about hyper-activity of the canonical Wnt signaling pathway and failure of the cell-cell adhesion regulation (Valenta et al., 2012). The disruption of APC

ultimately leads to dysfunction in maintaining the homeostasis of cellular regulation and results in more chances of other genetic alterations. Our understanding of CRC progression has been advanced over last few decades, but we still do not know much about its initiation process starting from APC deficiency. This is because there are limited measures for observation and analysis of cancer initiation events. If we unveil CRC initiation events, we might be able to identify novel prognostic markers and therapeutic targets for early cancer detection and prevention (Kaufman et al., 2016).

In order to investigate the cancer initiation process, we need to utilize a tool that can monitor instantaneous and delicate changes of the transcriptomic landscape during the initiating events. Single cell RNA sequencing (scRNA-seq) can fulfill this demand by dissecting gene expressions at each individual cellular resolution. Several studies on development and oncology exemplified that we can capture heterogeneity in cell fate decision or drug response using scRNA-seq (Maamar et al., 2007; Huang, 2009; Eldar and Elowitz, 2010; Petropoulos et al., 2016).

As it is not possible to analyze or understand all uncountable miniscule changes at an mRNA level captured by scRNA-seq, it is essential to find out a few key genes that might be primarily responsible for controlling the cell phenotypes. Such genes, called “master regulators,” trigger a series of gene regulation events which ultimately lead to critical changes in gene regulatory networks (Califano and Alvarez, 2017). We note that recent progresses in systems biology show the importance of unraveling the gene regulatory network and the causal relationships among the gene regulations to properly understand the complex biological phenomena (Schmidt et al., 2005; Kim and Cho, 2006; Park et al., 2006; Kim et al., 2007, 2011; Kwon and Cho, 2007; Murray et al., 2010). Previous studies report that master regulator analysis can successfully identify crucial genes for maintaining and controlling cancer gene regulatory networks (Wang et al., 2009; Campbell et al., 2016).

In this study, to understand the earliest events in CRC initiation, we establish an early CRC development model by disrupting APC in the normal human colorectal epithelial cell with shRNA and conduct scRNA-seq. Interestingly, we find two subtypes, fast growing vs. slowly growing populations of distinct growth rate and gene signatures. We focus on how they work differently at the transcriptomic level and conduct master regulator analysis. As a result, we find CCDC85B as a master regulator that can transform the cellular state of fast growing subtype cells into that of slowly growing subtype cells. We further validate this by *in vitro* experiments and suggest a novel therapeutic strategy that may prevent malignant CRC development by suppressing stemness and uncontrolled cell proliferation.

MATERIALS AND METHODS

Cell Culture

Immortalized human colon epithelial cells (HCEC), 1CT and its wild type APC depleted version, 1CT-A cells are generously provided by Jerry W. Shay (University of Texas, Dallas, TX, United States). 1CT and 1CT-A cells are cultured in basal X media (DMEM: M199, 4:1; WelGENE Inc., Gyeongsan, Korea), supplemented with epidermal growth factor (20 ng·ml⁻¹; Thermo Fisher Scientific, Waltham, MA, United States), hydrocortisone (1 mg·ml⁻¹), insulin (10 mg·ml⁻¹), transferrin (2 mg·ml⁻¹), sodium selenite (5 nM; all from Sigma, Deisenhofen, Germany), 2% FBS, and antibiotics (100 units·ml⁻¹ of penicillin, 100 µg·ml⁻¹ streptomycin, and 0.25 µg·ml⁻¹ of Fungizone; Life Technologies Corp., Carlsbad, CA, United States). Cells are cultured at 37°C in a humidified atmosphere containing 5% CO₂.

Transfection and Transduction of shRNA

For lentivirus production, HEK 293T cells are transfected with shRNA targeting APC (shAPC; TRCN0000244294, Sigma) and packaging mix (pLP1, pLP2, and pLP/VSVG) using Lipofectamine (Invitrogen, Waltham, MA, United States), according to manufacturer's protocols. Then viral supernatants are collected and applied to target cells with polybrene (4 µg·ml⁻¹; Sigma). Infected cells are selected with puromycin (500 ng·ml⁻¹; Sigma) before harvest. 1CT cells infected with scrambled shRNA (shScr) are prepared as control samples, and their culture periods are matched with the shAPC samples.

Transfection of siRNA

Control siRNA (siControl), CCDC85B siRNA (siCCDC85B-1 and siCCDC85B-2), and PTTG1 siRNA (siPTTG1-1 and siPTTG1-2) oligonucleotides (BIONEER Corporation, Daejeon, South Korea) are synthesized in a sense-antisense duplex form (Table 1). Primer sequences for ASCL2, CCDC85B, CCNE1, and CCNA2 were referred from OriGene Technologies, Inc. (Rockville, MD, United States). For siRNA transfection, mixture of siRNAs and RNAiMAX (Thermo Fisher Scientific) with final concentration of 2 µM is applied to the target cell on the 60 mm culture dish, following the manufacturer's protocol. After 24 h, transfected cells are subcultured into 24-well plates and 60 mm culture dish for growth curve check and RNA harvest, respectively.

Total RNA Extraction and qRT-PCR

Total RNA is extracted from cells by using RNA-spin™ Total RNA Extraction Kit (iNtRON Biotechnology, Gyeonggi, South Korea), according to the manufacturer's protocol, and treated

TABLE 1 | Sequences of siRNAs.

siRNA name	Target gene	Sense siRNA sequence (5'-3')	Antisense siRNA sequence (5'-3')
siCCDC85B-1	CCDC85B	GAGGUUCGAAGCUCCUAGU	ACUAGGAGCUUCGAACCUC
siCCDC85B-2	CCDC85B	GAUUGGCGUCCUCCAUUA	UAUGGAAGGACAGCCAAUC
siPTTG1-1	PTTG1	AGCACCAGAUUGCGCACCU	AGGUGCGCAAUCUGGUGCU
siPTTG1-2	PTTG1	GUUGAAUUGCCACCUUUU	AAACAGGUGGCAAUUAAC

with RNase-free DNase I (Thermo Fisher Scientific) to remove contaminating genomic DNA. cDNA is then synthesized from total RNA by reverse transcription (RT) using a DiaStar RT kit (Solgent, Daejeon, Korea) and the PCR system (Veriti 96-well Thermal Cycler; Applied Biosystems, Waltham, MA, United States). Quantitative reverse transcription PCR (qRT-PCR) analysis is performed using the QuantStudio 5 real-time PCR system (Applied Biosystems) with the corresponding primers (Table 2).

Bulk RNA Sequencing

RNA sequencing experiments are performed using tools from the commercial microarray service Ebiogen, Inc. (Seoul, Korea). Total RNA is extracted from ICT and ICT-A cells using RNA-spin™ (iNtRON) according to the manufacturer's instructions. The isolated RNA is amplified and subjected to cDNA microarray (Ebiogen).

Single Cell RNA Sequencing

HCEC-ICT cells infected with shAPC and their matched cells infected with shScr are harvested at 3- and 7-days after transduction, and stored on ice in PBS before the single cell library preparation. scRNA-seq is performed using the 10x Genomics Chromium V3 kit, following the manufacturer's protocol (Zheng et al., 2017). We align the scRNA-seq dataset along hg38 with CellRanger 3.0.0, and process with Seurat 3.1 R toolkit (Butler et al., 2018; Stuart et al., 2019). We perform initial quality control with Seurat, following the standard preprocessing workflow for scRNA-seq data (Illic et al., 2016). Dying cells and multiplets are excluded under the assumption that unhealthy cells tend to have either very few genes (<200) or low unique feature counts (<500; **Supplementary Figure S1 and Supplementary Table S1**).

For batch correction, we use ComBat from surrogate variable analysis (sva) package (Johnson et al., 2007; Leek et al., 2012) on Trimmed Mean of M-values (TMM) normalized data. In addition, we check the consistency of the batch correction results by comparing the results of ComBat and the different method called canonical correlation analysis (CCA; **Supplementary Figure S10**).

Then we additionally conduct data imputation using deep count autoencoder (DCA) in order to denoise scRNA-seq

datasets (Eraslan et al., 2019), and compare the scRNA-seq datasets before and after denoising process with DCA (**Supplementary Figure S8**). By using another data imputation tool, Adaptively-thresholded Low Rank Approximation (ALRA), we check the consistency of the imputation performance (Linderman et al., 2018; **Supplementary Figure S9**).

Single cell gene expression levels are scaled, so that the mean is equal to zero and the variance is equal to one, and the effects of cell cycle heterogeneity are ruled out by cell cycle score regression according to Seurat manual.

Clustering of Fast Growth and Slow Growth Subpopulations

We perform unsupervised clustering of single cell dataset using shared nearest neighbor (SNN) modularity optimization using FindClusters function of Seurat with resolution of one. These clusters are visualized using uniform manifold approximation and projection (UMAP) dimensionality reduction (McInnes et al., 2018).

We assign cell cycle score to each cell using the CellCycleScoring function of Seurat, which quantifies G2M and S phase scores of single cells based on the scoring strategy and the cell cycle marker genes suggested from previous studies (Kowalczyk et al., 2015; Tirosh et al., 2016). Then, each cell is classified as a cell in G2M, S, or G1 phase according to its cell cycle score.

The arrest signature score of each cell is quantified with AddModuleScore of Seurat along the cell cycle arrest related gene sets extracted from MSigDB (Subramanian et al., 2005; Liberzon et al., 2011, 2015). It is the gene set from Gene Ontology (GO) term, GO_REGULATION_OF_CELL_CYCLE_ARREST, that shows the most general coverage (Ashburner et al., 2000; The Gene Ontology Consortium, 2019). This gene set comprises 107 genes related with any process that modulates the rate, frequency, or extent of cell cycle arrest, the process in which the cell cycle is halted during one of the normal phases. Single cells are labeled as “arrested (Arr)” if their arrest signature scores are ranked higher than one-fifth of those of the whole single cells, otherwise labeled as “non-arrested (NArr).” If the ratio of Arr cells to NArr cells in a cluster is over 0.8 or less 0.2, then the cluster is labeled as “Arr” or “NArr,” respectively. A cluster with the average APC expression level

TABLE 2 | Sequences of qRT-PCR primers.

Target gene	Forward primer sequence (5'-3')	Reverse primer sequence (5'-3')
β actin	AGAGCTACGAGCTGCCTGAG	AGCACTGTGTTGGCGTACAG
APC	GCCCCACGAATTCTAAACCA	TTGTCTCTGCCTCGAGAGATT
MYC	GTCAAGAGGCGAACACAC	TTGGACGGACAGGATGTA
CCDC85B	TCATGCAGGAGGTGAATCGGCA	AGTCCAGGAAGCAGCAGAGGTC
PTTG1	GGACCCCTCAAACAAAACA	GAGAGGCACTCCACTCAAGG
CCNA2	CTCTACACAGTCACGGGACAAAG	CTGTGGTGCTTTGAGGTAGGTC
CCNB1	TTGGTGTCACTGCCATGTTT	CCGACCCAGACCAAGTTTA
CCND1	GCTGCGAAGTGGAACCATC	CCTCCTTCTGCACACATTGTA
CCNE1	TGTGTCCTGGATGTTGACTGCC	CTCTATGTCGCACCACTGATACC
LGR5	CTCCCAGGTCTGGTGTGT	GAGGTCTAGGTAGGAGGTGAAG
ASCL2	CGCCTACTCGTCGGACGA	GCCGCTCGCTCGGCTCCG

smaller than the first tertile is labeled as “low APC,” otherwise it is labeled as “high APC.” Then a cluster with both “low APC” and “Arr” is defined as slow growth subpopulation (SG) and clusters with “low APC” and “NArr” are defined as fast growth subpopulations (FG).

In addition, we investigate on differential markers of the identified clusters using FindMarkers function in Seurat package (Butler et al., 2018; Stuart et al., 2019), by setting log fold change threshold to 0.25 (**Supplementary Table S3**). We also characterize the cell type change of shAPC single cell RNA-seq samples with differentially expressed genes (DEGs) between 1CT and 1CT-A to examine whether there is any new cell type appeared, but no distinctive patterns are observed. The DEGs between 1CT and 1CT-A are determined with two-fold change and 0.01 cutoff of value of p using two-tailed t -test.

Characterization of Slow Growth Subpopulation

Apoptosis signature scores of SG cells are measured in the same way as the arrest signature score described in “Clustering of Fast Growth and Slow Growth Subpopulation” section, according to the cell apoptosis related gene set, GO_EXECUTION_PHASE_OF_APOPTOSIS (Ashburner et al., 2000; The Gene Ontology Consortium, 2019).

Stemness signature scores of SG cells are quantified using TCGAanalyze_Stemness function provided by TCGAbiolinks R toolkit, which generates mRNAi stemness index described in the previous study (Malta et al., 2018). The stemness signature used here is PCBC_stemSig which is the default stemness signature obtained using the data from Progenitor Cell Biology Consortium (PCBC). We analyze whether there are subclusters within FG or SG along with the signature score, but the groups along with signatures are not clearly discriminated in the activity inference (**Supplementary Figure S7**).

Protein Activity Inference Using VIPER

meta-Virtual Inference of Protein-activity by Enriched Regulon (metaVIPER) analysis is conducted to investigate the master regulators which control the fate determination of FG and SG (Alvarez et al., 2016). Since the single cell dataset lacks a tissue context, we use here metaVIPER which infers a regulatory network without tissue-specific regulatory information (Ding et al., 2018). First, the network of CRC cell is inferred using the patient expression dataset obtained from The Cancer Genome Atlas (TCGA) by the RTN package (Fletcher et al., 2013; Castro et al., 2016). Then, metaVIPER analysis is performed upon this CRC network with inputs composed of those genes of interest. The input gene lists used here are the list of DEGs between FG and SG (1.5-fold change, $p < 0.01$), and regulon lists generated by single-cell regulatory network inference and clustering (SCENIC).

Reconstruction of Gene Regulatory Networks Using SCENIC

Single-cell regulatory network inference and clustering analysis is performed to generate the gene regulatory networks of SG

and FG as described in the original paper using pySCENIC version 0.9.19 (Aibar et al., 2017). The corresponding auxiliary datasets used for SCENIC analysis are human cisTarget of 100 bp down, 500 bp up, and 10 kb up and down with the genome version of hg38, human TF binding motif provided by cisTargetDB of version 9, and the list of curated human TF comprising 1,390 genes. The resulting regulon lists are collected and used for the metaVIPER analysis to compute the activity difference between SG and FG.

Selection of Targets

Target candidates generated from metaVIPER are filtered by t -test between SG and FG ($p < 0.01$). Then hierarchical clustering is performed with these genes to select a gene group downregulated in SG. The candidates are ranked along with the difference of average activity and expression level between SG and FG. The candidates of the top largest difference are selected for the next step analysis. At this point, since the expression level shows less clear discrimination between SG and FG, the threshold for expression difference is set to be one-third while that for activity difference is set to be one-sixth.

Next, the filtered target candidates are screened by how they are closely related to APC under the rationale that the target should cover the effect of APC in order to keep SG. Therefore, we investigate the shortest path length between APC and the target candidate by using the input list comprising APC, CTNNB1, WNT, candidate itself and its downstream target gene list produced by SCENIC in STRING DB version 11 (Szklarczyk et al., 2019). In addition, how much target genes a candidate shares with APC is quantified from the gene regulatory network inferred from CRC patients in TCGA database described in “Protein Activity Inference Using VIPER” section.

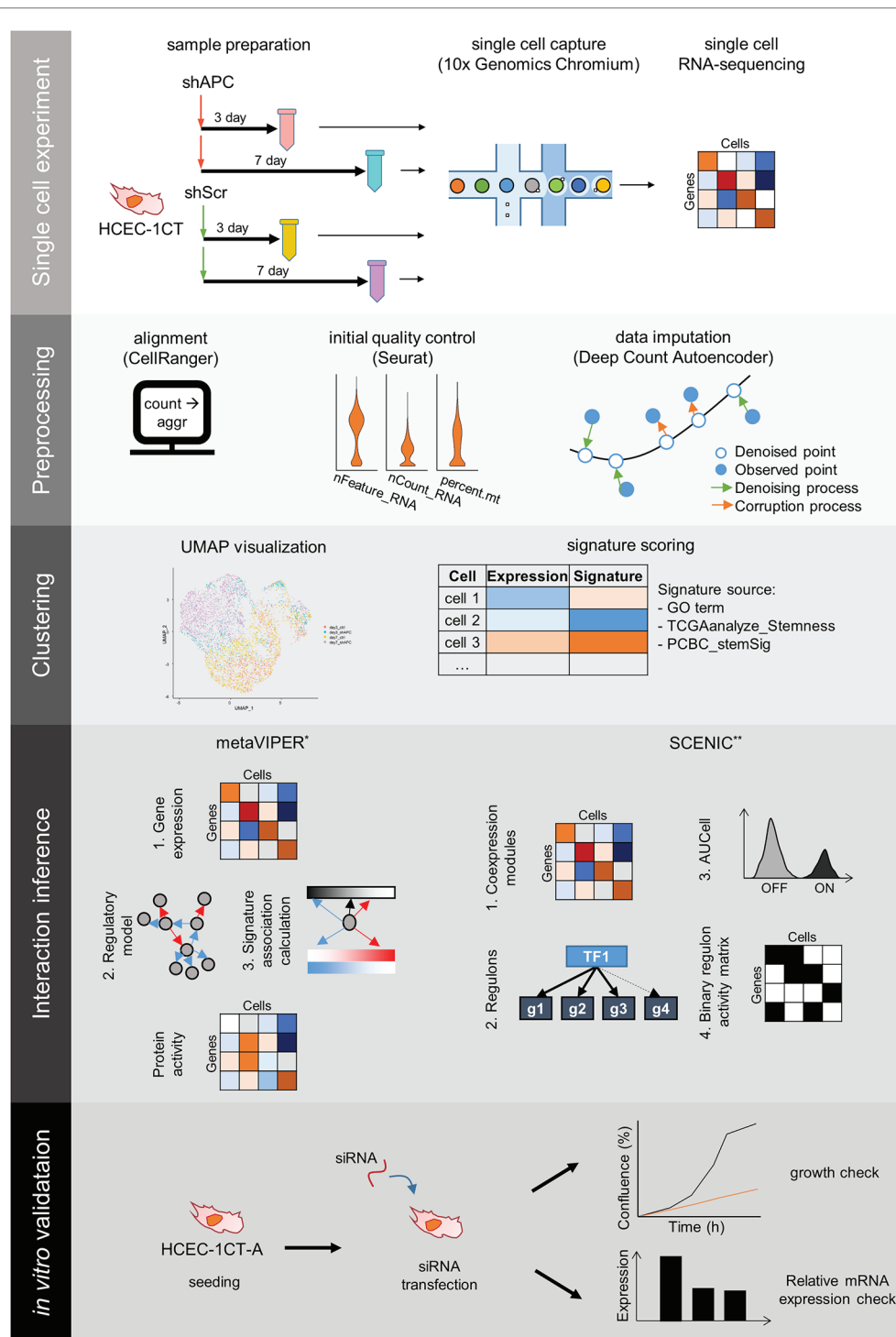
Then, interactions within target candidates are explored by MINDY to reconstruct a network composed of candidate TFs and their modulators (Wang et al., 2009). The most densely connected TF with others is considered as the most important master regulator.

Networks are visualized using Cytoscape version 3.7.1 (Shannon et al., 2003).

RESULTS

Slowly Growing and Fast Growing Subpopulations Are Found From the scRNA-seq Dataset of APC-Deficient Normal Colon Epithelial Cells

In order to investigate complex events occurring during the cancer initiation, we establish an early CRC development model, perform scRNA-seq, and analyze the scRNA-seq dataset (**Figure 1**). scRNA-seq is conducted at 3- and 7-days after transduction of shAPC or shScr on HCEC-1CT (1CT) cells. We examine the relative gene expression levels of APC and its downstream targets by performing qRT-PCR of remaining



* Redrawn from Figure 1 of the original VIPER article.

** Redrawn from Figure 1 of the original SCENIC article.

FIGURE 1 | The scheme of single cell RNA-sequencing (scRNA-seq) experiment and analysis. scRNA-seq experiment and analysis comprise five steps: single cell experiment, preprocessing, clustering, interaction inference, and *in vitro* validation. Samples for scRNA-seq are prepared by transduction of shRNA targeting APC (shAPC) or scrambled shRNA (shScr) in HCEC-1CT cells, and scRNA-seq is performed using 10x chromium platform. Then we take preprocessing steps such as alignment, initial quality control, and data imputation. The single cell data points are clustered, and each cluster is scored according to gene signatures. Then, interactions within distinct clusters are inferred using Virtual Inference of Protein-activity by Enriched Regulon (VIPER) and single-cell regulatory network inference and clustering (SCENIC) to produce master regulators for the clusters. These master regulators are validated using siRNA transfection in 1CT-A cells.

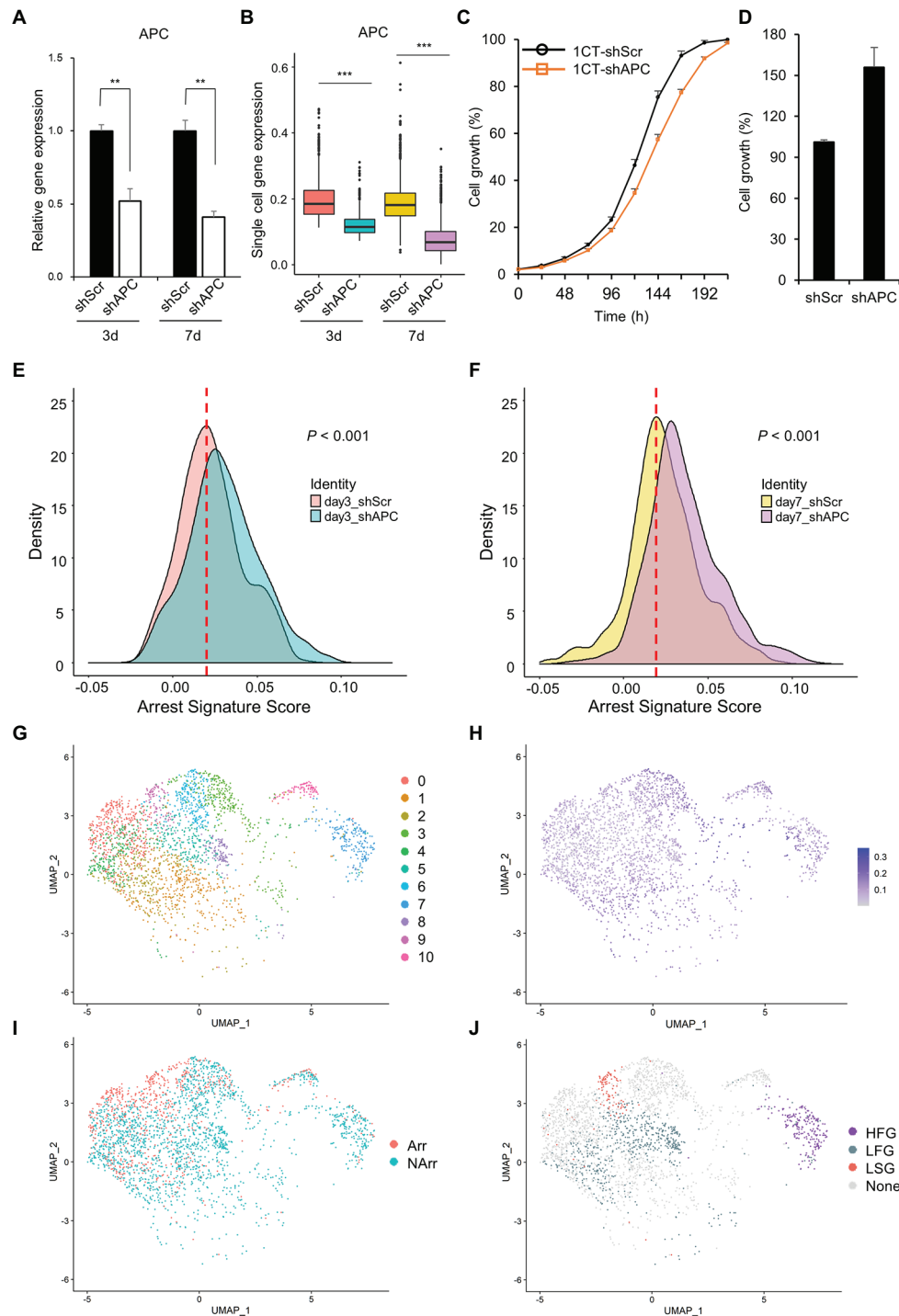


FIGURE 2 | Clustering of slow growth and fast growth subpopulation in scRNA-seq dataset. **(A)** Relative adenomatous polyposis coli (APC) gene expression of single cell samples in quantitative reverse transcription PCR (qRT-PCR). **(B)** APC gene expression of single cell samples in scRNA-seq dataset. **(C)** Growth curve of HCEC-1CT with shScr and shAPC transduction during the short initial period of time (~7 days after transduction). The shAPC samples grow slightly slower than shScr samples. **(D)** Growth rate of HCEC-1CT with shScr and shAPC transduction (16 days after transduction). The shAPC samples grow faster than shScr samples. Distribution of arrest signature after APC knockdown **(E)** at day 3 scRNA-seq dataset and **(F)** at day 7 scRNA-seq dataset. **(G)** Unsupervised clustering of scRNA-seq dataset. **(H)** APC gene expression level of scRNA-seq dataset. **(I)** Binarized arrest signature score of scRNA-seq dataset. **(J)** Cluster labels of scRNA-seq dataset. The criteria are designated to each cluster according to the combination of the APC level and arrest signature score: HFG for APC High and Fast Growth; LFG for APC Low and Fast Growth; LSG for APC Low and Slow Growth; and None for the remainders.

cells after single cell library preparation (**Figure 2A** and **Supplementary Figure S2A**), as well as by investigating the expression levels from scRNA-seq (**Figure 2B** and **Supplementary Figure S2B**). We confirm that the level of APC is dropped to at least 50% in shAPC samples compared to shScr samples in both bulk and single cell data.

Interestingly, we find that APC knockdown of 1CT decreases the cell growth mildly during a short initial period of time (about 7 days elapsed after shAPC transduction; **Figure 2C**) but eventually increases the cell growth at a later time (16 days elapsed after shAPC transduction; **Figure 2D**). This relationship between depletion of APC and the relatively slow cell growth is partially supported by a previous study reporting that APC loss drives the growth arrest or senescence program in the premalignant renal tumor (Cole et al., 2010). Since this trend is not observed in bulk qRT-PCR results (**Figure 2C**), we assume that it might be originated from rare and hard-to-observe events during CRC initiation.

To check out this assumption, we initially analyze changes in the arrest signature score between APC deficient cells and others, and find that the arrest signature score is increased in shAPC samples compared to that of shScr samples (**Figures 2E,F** and **Supplementary Figure S3**). This shift of the arrest signature score appears in both day3 and day7 samples, and becomes clearer in day 7 samples.

In order to figure out the source of driving this increased arrest signature in APC downregulated cells, we investigate the characteristics of clusters in shAPC single cell samples by assuming that there might be a subpopulation responsible for this phenomenon. Eleven clusters are identified and labeled according to four criteria (HFG for APC High and Fast Growth; LFG for APC Low and Fast Growth; LSG for APC Low and Slow Growth; and None for the remainders) based on arrest signature and APC level (**Figures 2G–J** and **Supplementary Table S2**). There are one cluster of HFG (Cluster 7), three clusters of LFG (Clusters 2, 5, and 8), and one cluster of LSG (Cluster 9). It is remarkable that the population of LSG is about one-sixth of LFG population, which is the reason why bulk analysis could not capture the characteristics originated from LSG (**Supplementary Table S4** and **Supplementary Figure S2A**). Since our interest lies on the cells affected by APC downregulation, we exclude the HFG cluster in downstream analysis and take only LFG and LSG for further analysis. The labels of LFG and LSG are shortened hereafter as FG (Fast Growth) and SG (Slow Growth), respectively.

SG and FG Have Different Phenotypical Characteristics and Gene Regulatory Networks

We further examine the characteristics of SG and FG to see whether they actually differ in phenotypical biological processes such as apoptosis and stemness besides the arrest signature. As a result, we find that SG has a higher apoptosis signature and a lower stemness signature than FG (**Figure 3A** and **Supplementary Figure S4**), implying that SG has a fate to go through apoptosis without developing further malignancy.

Since SG is assumed to eventually diminish while FG is to progress into advanced cancer, we perform master regulator analysis to identify transcription factors that can drive FG to SG such that the majority of APC deficient cells undergo apoptosis. We perform metaVIPER (Alvarez et al., 2016; Ding et al., 2018) analysis with 848 DEGs between 1CT and its wild type APC depleted version, HCEC-1CT-A (1CT-A), and, as a result, we find 412 master regulators present in either SG or FG.

For further master regulator analysis, we examine the gene regulatory networks of FG and SG using SCENIC (Aibar et al., 2017) in order to determine whether the two subpopulations have differently organized gene regulation structures. Here, we define SG regulons and FG regulons as those genes shared by both regulons inferred from SCENIC and the master regulators found from DEG metaVIPER. It turns out that FG regulons comprise seven transcription factors such as E2F7, FOXN2, TFAP4, FOXK2, NFIX, RARA, and HMGA1 (**Figure 3B**), whereas SG regulons comprise nine transcription factors such as ARNTL2, YBX1, ZNF513, HINFP, PPARG, TER, TEAD4, ZNF766, and NR1D1 (**Figure 3C**).

We expand the list of regulons up to 1,411 genes by merging SG regulons, FG regulons, and their target genes. Then, we perform metaVIPER again with this list of regulons to find out the master regulators that can drive FG into SG (**Figure 3D**). Genes downregulated in SG with statistical significance (two-tailed *t*-test, $p < 0.05$) are taken and they are filtered again according to the level of difference in their expressions and activities across SG and FG (**Figures 4A,B**).

CCDC85B and PTTG1 Are the Most Important Master Regulators Responsible for the Difference Between SG and FG

In order to narrow down the final targets, we examine the interactions between APC and target candidates or within target candidates (**Figures 4A,B**). First, the shortest path lengths between candidates and APC are investigated using STRING DB (Szklarczyk et al., 2019), resulting in three groups of genes which have any connection to APC: HDAC2, RUVBL1, and RUVBL2 for a length of two; CCDC85B, ELOB, ELOC, ILF2, PFN1, and PTTG1 for a length of three; DNTTIP2 and PA2G4 for a length of four (**Figure 4C**). In addition to the shortest path lengths, we investigate the number of genes which candidates share with APC, and CCDC85B is found to be one of the most densely APC regulon sharing genes (**Figure 4D**).

Since HDAC2 is known to have many redundant functions, it is classified as a less attractive marker for early cancer development (Jurkin et al., 2011). Considering that 1CT cell line has hTERT manipulation, genes related with telomerase such as RUVBL1 and RUVBL2 might be screened as 1CT context specific targets. Therefore, the genes with the length of three are considered as more promising targets instead of those with the length of two, and their interactions *via* modulators are probed using Modulator Inference by Network Dynamics (MINDy; Wang et al., 2009; **Figure 4E**). As a result, we find that only CCDC85B and PTTG1 have a direct cross-modulation relationship among six candidate genes.

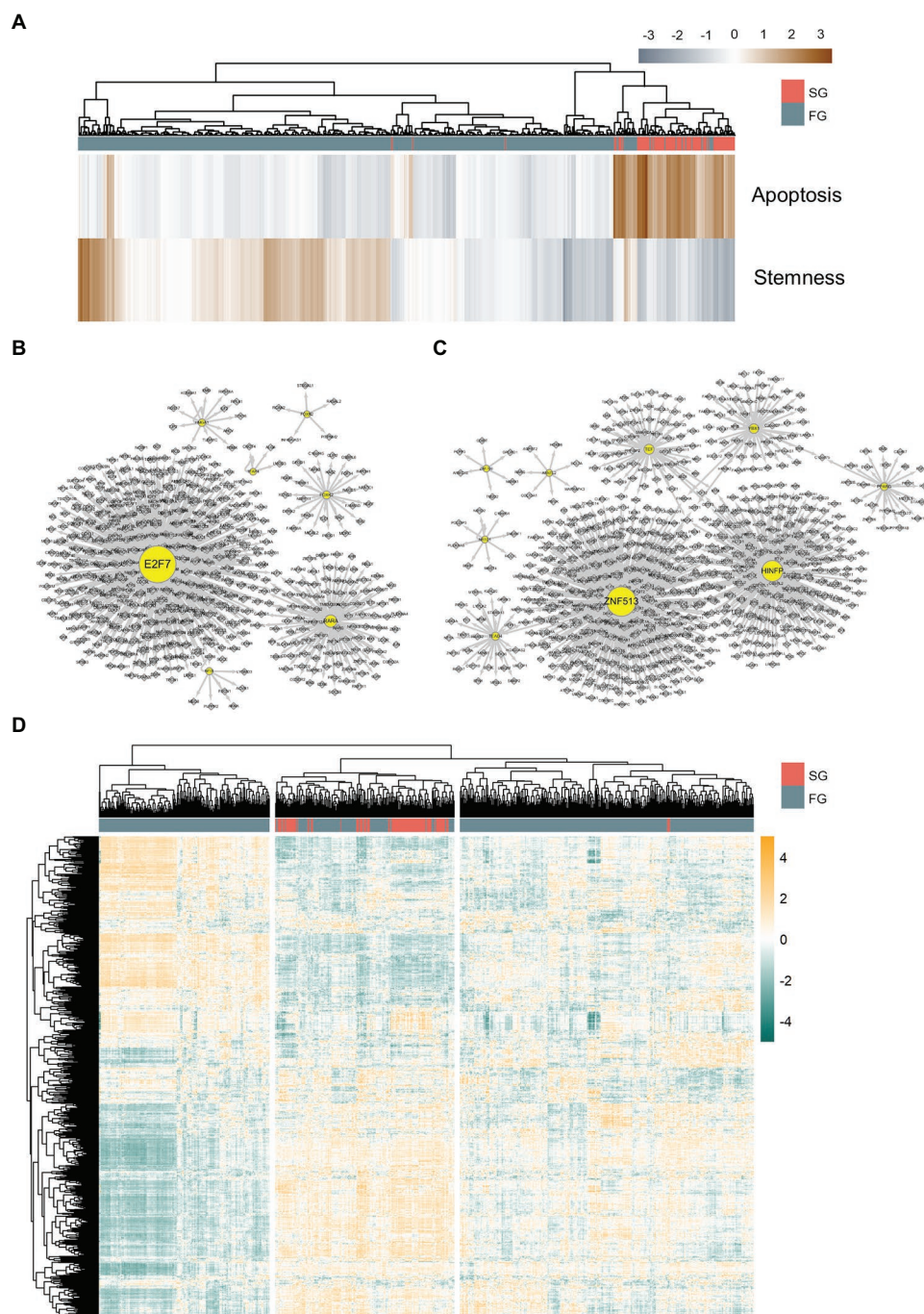


FIGURE 3 | Differently organized gene regulatory networks of slow growth subpopulation (SG) and fast growth subpopulation (FG). **(A)** Apoptosis and stemness signature scores of SG and FG. Gene regulatory networks of **(B)** FG and **(C)** SG. **(D)** Master regulator analysis heatmap for SG and FG. SG and FG have different patterns of master regulator expressions.

Assuming that tightly bound master regulators are more likely to control the biological process that is distinct in each of SG and FG, we conclude that CCDC85B and PTTG1 can be the final target candidates.

Since CCDC85B and APC share 124 genes (**Figure 4F**) and their shared genes are participants of essential biological

processes such as regulation of macromolecule biosynthetic process and regulation of RNA metabolic (**Supplementary Table S5**), we select CCDC85B as a primary target candidate.

To validate that CCDC85B and PTTG1 are relevant with the characteristics of SG, the correlation between their expressions or activities and the apoptosis or arrest signature are further

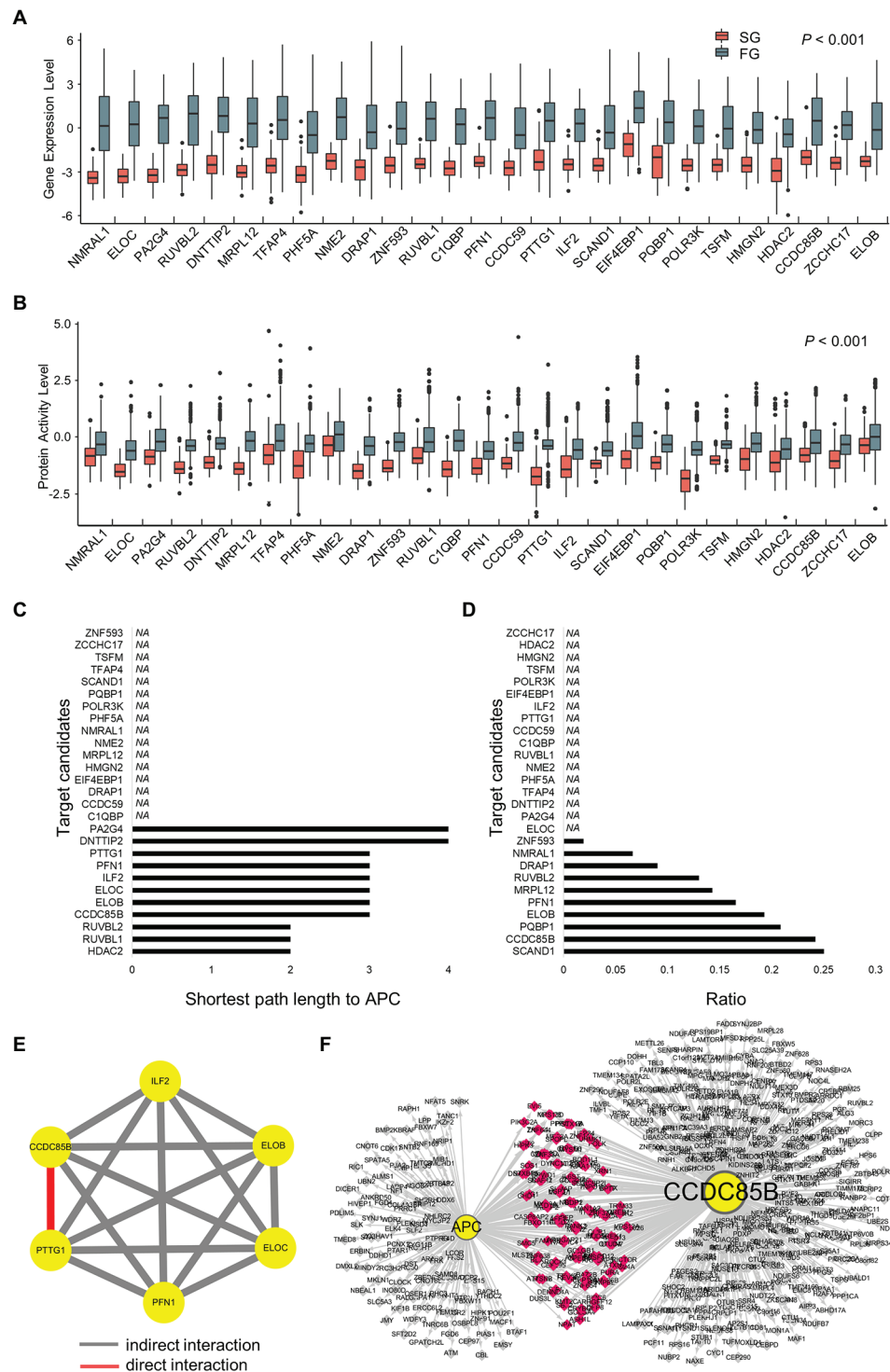


FIGURE 4 | Prioritization of target candidates according to interactions with APC and the interactions among the candidates. **(A)** Gene expression levels and **(B)** protein activity levels of target candidates. **(C)** The shortest path length from each target candidate to APC. **(D)** Ratio of shared regulons between target candidates and APC. **(E)** Modulatory interactions within six target candidates. **(F)** Shared target genes of CCDC85B and APC.

investigated (Supplementary Figures S5, S6). Both activity and expression of CCDC85B have a negative correlation with arrest and apoptosis signature scores, implying that its downregulation

might slow down the cell cycle. The relationship of the expression or activity of PTTG1 and the score of apoptosis or arrest is similar to that of CCDC85B.

***In vitro* Knockdown of CCDC85B Shows Significant Influences on Both Stemness and Cell Cycle as Predicted by Network Analysis**

To validate whether the candidate targets can actually interrupt cancer progression, we perform *in vitro* knockdown of CCDC85B and PTTG1 using siRNA and examine the changes of the growth rate and transcriptomic levels. The cell growth rate of 1CT-A is dramatically decreased with the transfection of siRNA targeting CCDC85B (siCCDC85B; **Figures 5A,C**). The relative mRNA levels of APC and MYC are not affected by siCCDC85B transfection, whereas those of CCDC85B and PTTG1 are decreased a lot (**Figure 5E**). We investigate the changes in the level of various cyclins to figure out which cyclins CCDC85B has affected. Since siCCDC85B decreased the relative mRNA levels of Cyclin A2 and Cyclin B1, we can infer that CCDC85B

might act on G2/M phase (**Figure 5E**). Interestingly, it coincides with the cell phase where the majority of SG stays in our single cell data (**Supplementary Table S2**).

Next, we examine the effects of PTTG1 interference using siRNA (siPTTG1) in 1CT-A on cell cycle and stemness since PTTG1 is considered to function with CCDC85B according to MINDy analysis. The cell cycle is arrested by siPTTG1 transfection, and the primarily affected cyclins are Cyclin A2 and Cyclin B1 as in the case with the siCCDC85B results (**Figures 5B,D**). siPTTG1 transfection reduces only PTTG1 significantly not CCDC85B, whereas siCCDC85B transfection reduces the level of both CCDC85B and PTTG1. We need to note that stemness markers for colon cells such as LGR5 and ASCL2 show a non-significant change when PTTG1 is perturbed unlike CCDC85B perturbation experimental results (**Figure 5F**).

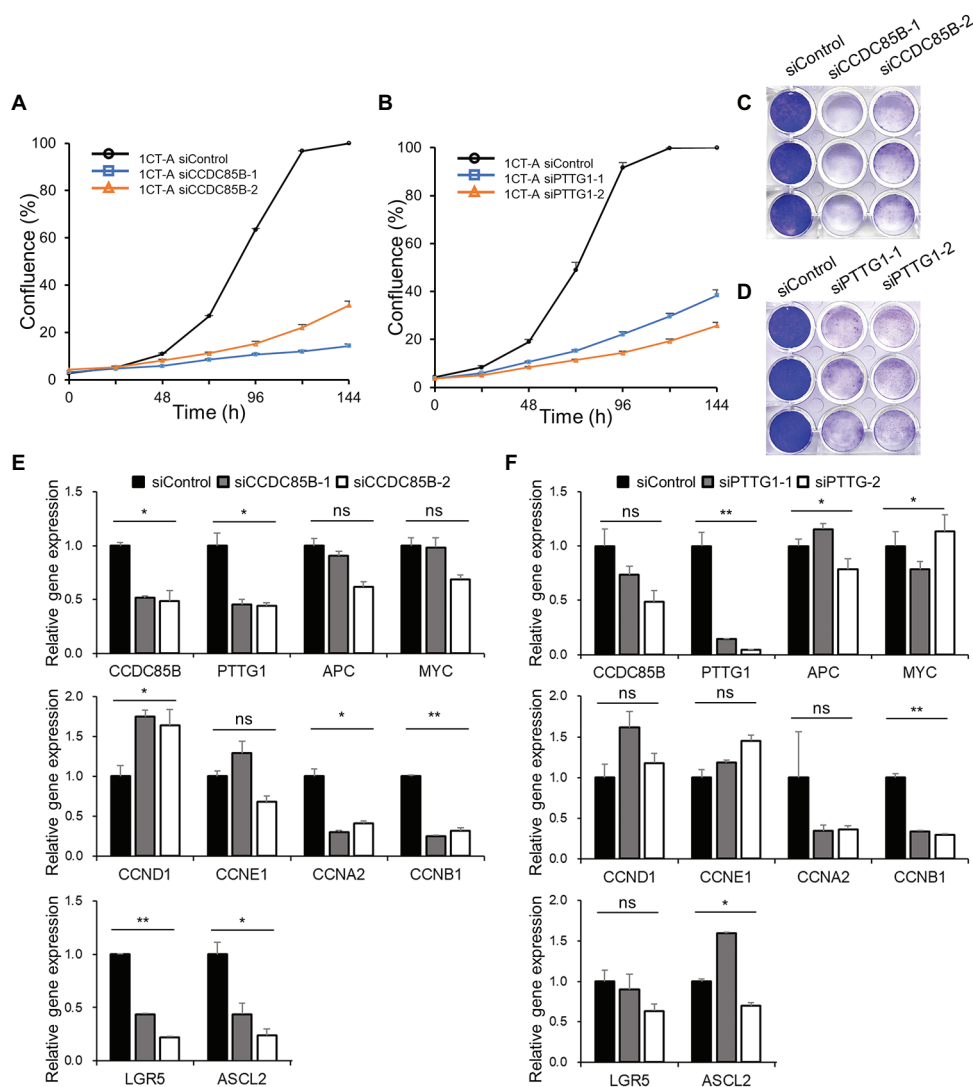


FIGURE 5 | Growth curve and gene expression level after CCDC85B or PTTG1 interference in 1CT-A cells. Growth rate of 1CT-A cells after interfering with (A,C) CCDC85B or (B,D) PTTG1. Relative gene expression levels of 1CT-A cells after interfering with (E) CCDC85B or (F) PTTG1.

DISCUSSION

In this study, we investigate master regulators the downregulation of which can lead to suppression of early CRC progression by analyzing scRNA-seq data. We establish the early CRC development model by interfering with APC using shRNA in normal colon epithelial cells, 1CT, and then we conduct scRNA-seq to capture small and heterogeneous changes that occur during the earliest events in CRC initiation. Since increment of arrest signature after APC downregulation is observed, we assume that there might be subpopulations responsible for this shift. We find out two subpopulations with different growth rates, and define one subpopulation with a relatively slow cell cycle as the slow growth subpopulation (SG) and the other with a relatively fast cell cycle as the fast growth subpopulation (FG). Through further analysis, we find that SG and FG differ in their organization of gene regulatory networks, as well as cell growth rates. Interestingly, SG has a low stemness signature and a high apoptosis signature, whereas FG has a high stemness signature and a low apoptosis signature. Although there is no direct experimental evidence presented in this study, it is highly likely that SG eventually goes through apoptosis instead of developing malignancy by acquiring stemness contrasting to the opposite fate of FG. Hence, we presume that transforming the FG into the SG might be a useful strategy of restraining early CRC development as it pursues diminishing the cell population of a malignant fate.

From the master regulator analysis, we identify CCDC85B and PTTG1 as the two most promising master regulators that can discriminate SG and FG and validate that both can lower cell growth rates by knockdown experiments using siRNA. In particular, knockdown of CCDC85B lowers the expression level of stemness markers such as ASCL2 and LGR5 in addition to the level of cyclins, whereas knockdown of PTTG1 lowers only the expression level of cyclins. Both CCDC85B and PTTG1 affect Cyclin A2 and Cyclin B1, which are known to act at G2/M phase. This might be a predictable result since PTTG1 is previously reported to act as a master regulator that controls the cell cycle at G2/M phase (Quereda and Malumbres, 2009; Liang et al., 2011). It is noteworthy that HDAC2 is one of the differential master regulators between SG and FG besides CCDC85B and PTTG1, since it implies that chromatin regulation plays a role in the discrimination of SG and FG. Considering that APC is known for its contribution in the chromosomal instability seen in many colon cancer cells, it seems natural for chromatin regulation to appear as one of the controlling mechanism of the earliest events of cancer development. A further study on this relationship between chromatin regulation and characteristics of SG and FG would add more value to the understanding on the earliest events in CRC initiation.

We infer that CCDC85B regulates stemness and cell cycle *via* β -catenin and PTTG1, respectively, based on literature survey and our own experiments. It is known that CCDC85B is overexpressed in the tumor sample of non-small cell lung cancer patients and that CCDC85B takes a crucial part in activation of β -catenin (Feng et al., 2019). Therefore, we can infer the decreased level of colon epithelial stemness markers after CCDC85B knockdown might be a result of the decreased active β -catenin induced by CCDC85B knockdown.

We show that CCDC85B knockdown decreases the relative mRNA expression levels of both CCDC85B and PTTG1, and that CCDC85B and PTTG1 has similar effects on the identical cell cyclins such as Cyclin A2 and Cyclin B1. Thus, we can infer that CCDC85B affects cell cyclins through PTTG1.

In summary, we suggest CCDC85B as a novel potential therapeutic target for restraining early CRC progression by lowering both the cell growth rate and stemness through the regulation of PTTG1 and β -catenin.

DATA AVAILABILITY STATEMENT

All data generated or analyzed during this study are included in this published article (and its **Supplementary Material** files). The sequencing data discussed in this publication are deposited in NCBI's Gene Expression Omnibus (Edgar et al., 2002) and accessible through GEO Series accession number GSE152746 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE152746>).

AUTHOR CONTRIBUTIONS

K-HC designed the project and supervised the research. SL, CYH, and JC performed single cell RNA-sequencing experiment. J-RG and CYH provided analytical and experimental guidance. CYJ provided experimental support. JC and K-HC wrote and revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Research Foundation of Korea (NRF) grants funded by the Korea Government, the Ministry of Science and ICT (2020R1A2B5B03094920), Electronics and Telecommunications Research Institute (ETRI) grant (20ZS1100, Core Technology Research for Self-Improving Integrated Artificial Intelligence System), and by KAIST Grand Challenge 30 Project.

ACKNOWLEDGMENTS

We would like to thank all the members of SBiE (Systems Biology and Bio-inspired Engineering) laboratory at KAIST for their insightful comments. In particular, we thank Dongsan Kim for preprocessing scRNA-seq dataset. We also thank Jerry W. Shay for providing 1CT and its derived cell lines.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.570546/full#supplementary-material>

REFERENCES

- Aibar, S., Gonzalez-Blas, C. B., Moerman, T., Huynh-Thu, V. A., Imrichova, H., Hulselmans, G., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* 14, 1083–1086. doi: 10.1038/nmeth.4463
- Alvarez, M. J., Shen, Y., Giorgi, F. M., Lachmann, A., Ding, B. B., Ye, B. H., et al. (2016). Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat. Genet.* 48, 838–847. doi: 10.1038/ng.3593
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556
- Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* 36, 411–420. doi: 10.1038/nbt.4096
- Califano, A., and Alvarez, M. J. (2017). The recurrent architecture of tumour initiation, progression and drug sensitivity. *Nat. Rev. Cancer* 17, 116–130. doi: 10.1038/nrc.2016.124
- Campbell, T. M., Castro, M. A., Ponder, B. A., and Meyer, K. B. (2016). Identification of post-transcriptional modulators of breast cancer transcription factor activity using MINDy. *PLoS One* 11:e0168770. doi: 10.1371/journal.pone.0168770
- Castro, M. A., de Santiago, I., Campbell, T. M., Vaughn, C., Hickey, T. E., Ross, E., et al. (2016). Regulators of genetic risk of breast cancer identified by integrative network analysis. *Nat. Genet.* 48, 12–21. doi: 10.1038/ng.3458
- Cole, A. M., Ridgway, R. A., Derkits, S. E., Parry, L., Barker, N., Clevers, H., et al. (2010). p21 loss blocks senescence following Apc loss and provokes tumorigenesis in the renal but not the intestinal epithelium. *EMBO Mol. Med.* 2, 472–486. doi: 10.1002/emmm.201000101
- Ding, H., Douglass, E. F. Jr., Sonabend, A. M., Mela, A., Bose, S., Gonzalez, C., et al. (2018). Quantitative assessment of protein activity in orphan tissues and single cells using the metaVIPER algorithm. *Nat. Commun.* 9:1471. doi: 10.1038/s41467-018-03843-3
- Edgar, R., Domrachev, M., and Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 30, 207–210. doi: 10.1093/nar/30.1.207
- Eldar, A., and Elowitz, M. B. (2010). Functional roles for noise in genetic circuits. *Nature* 467, 167–173. doi: 10.1038/nature09326
- Eraslan, G., Simon, L. M., Mircea, M., Mueller, N. S., and Theis, F. J. (2019). Single-cell RNA-seq denoising using a deep count autoencoder. *Nat. Commun.* 10:390. doi: 10.1038/s41467-018-07931-2
- Fearon, E. R., and Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. *Cell* 61, 759–767. doi: 10.1016/0092-8674(90)90186-i
- Feng, Y., Gao, Y., Yu, J., Jiang, G., Zhang, X., Lin, X., et al. (2019). CCDC85B promotes non-small cell lung cancer cell proliferation and invasion. *Mol. Carcinog.* 58, 126–134. doi: 10.1002/mc.22914
- Fletcher, M. N., Castro, M. A., Wang, X., de Santiago, I., O'Reilly, M., Chin, S. F., et al. (2013). Master regulators of FGFR2 signalling and breast cancer risk. *Nat. Commun.* 4:2464. doi: 10.1038/ncomms3464
- Huang, S. (2009). Non-genetic heterogeneity of cells in development: more than just noise. *Development* 136, 3853–3862. doi: 10.1242/dev.035139
- Ilicic, T., Kim, J. K., Kolodziejczyk, A. A., Bagger, F. O., McCarthy, D. J., Marioni, J. C., et al. (2016). Classification of low quality cells from single-cell RNA-seq data. *Genome Biol.* 17:29. doi: 10.1186/s13059-016-0888-1
- Johnson, W. E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 8, 118–127. doi: 10.1093/biostatistics/kxj037
- Jurkin, J., Zupkovitz, G., Lagger, S., Grausenburger, R., Hagelkruys, A., Kenner, L., et al. (2011). Distinct and redundant functions of histone deacetylases HDAC1 and HDAC2 in proliferation and tumorigenesis. *Cell Cycle* 10, 406–412. doi: 10.4161/cc.10.3.14712
- Kaufman, C. K., Mosimann, C., Fan, Z. P., Yang, S., Thomas, A. J., Ablain, J., et al. (2016). A zebrafish melanoma model reveals emergence of neural crest identity during melanoma initiation. *Science* 351:aad2197. doi: 10.1126/science.aad2197
- Kim, J. R., and Cho, K. H. (2006). The multi-step phosphorelay mechanism of unorthodox two-component systems in *E. coli* realizes ultrasensitivity to stimuli while maintaining robustness to noises. *Comput. Biol. Chem.* 30, 438–444. doi: 10.1016/j.compbiolchem.2006.09.004
- Kim, S., Kim, J., and Cho, K. H. (2007). Inferring gene regulatory networks from temporal expression profiles under time-delay and noise. *Comput. Biol. Chem.* 31, 239–245. doi: 10.1016/j.compbiolchem.2007.03.013
- Kim, J. R., Kim, J., Kwon, Y. K., Lee, H. Y., Heslop-Harrison, P., and Cho, K. H. (2011). Reduction of complex signaling networks to a representative kernel. *Sci. Signal.* 4:ra35. doi: 10.1126/scisignal.2001390
- Kowalczyk, M. S., Tirosh, I., Heckl, D., Rao, T. N., Dixit, A., Haas, B. J., et al. (2015). Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. *Genome Res.* 25, 1860–1872. doi: 10.1101/gr.192237.115
- Kwon, Y. K., and Cho, K. H. (2007). Boolean dynamics of biological networks with multiple coupled feedback loops. *Biophys. J.* 92, 2975–2981. doi: 10.1529/biophysj.106.097097
- Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., and Storey, J. D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 28, 882–883. doi: 10.18129/B9.bioc.sva
- Liang, M., Chen, X., Liu, W., Li, S., Li, C., Jiang, L., et al. (2011). Role of the pituitary tumor transforming gene 1 in the progression of hepatocellular carcinoma. *Cancer Biol. Ther.* 11, 337–345. doi: 10.4161/cbt.11.3.14102
- Liberzon, A., Birger, C., Thorvaldsdottir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* 1, 417–425. doi: 10.1016/j.cels.2015.12.004
- Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdottir, H., Tamayo, P., and Mesirov, J. P. (2011). Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 27, 1739–1740. doi: 10.1093/bioinformatics/btr260
- Linderman, G. C., Zhao, J., and Kluger, Y. (2018). Zero-preserving imputation of scRNA-seq data using low-rank approximation. *bioRxiv* [Preprint]. doi: 10.1101/397588
- Maamar, H., Raj, A., and Dubnau, D. (2007). Noise in gene expression determines cell fate in *Bacillus subtilis*. *Science* 317, 526–529. doi: 10.1126/science.1140818
- Malta, T. M., Sokolov, A., Gentles, A. J., Burzykowski, T., Poisson, L., Weinstein, J. N., et al. (2018). Machine learning identifies stemness features associated with oncogenic dedifferentiation. *Cell* 173, 338.e315–354.e315. doi: 10.1016/j.cell.2018.03.034
- McInnes, L., Healy, J., and Melville, J. (2018). Umap: uniform manifold approximation and projection for dimension reduction. *arXiv* [Preprint]. Available at: <https://arxiv.org/abs/1802.03426v2>
- Murray, P. J., Kang, J. W., Mirams, G. R., Shin, S. Y., Byrne, H. M., Maini, P. K., et al. (2010). Modelling spatially regulated beta-catenin dynamics and invasion in intestinal crypts. *Biophys. J.* 99, 716–725. doi: 10.1016/j.bpj.2010.05.016
- Park, S. G., Lee, T., Kang, H. Y., Park, K., Cho, K. H., and Jung, G. (2006). The influence of the signal dynamics of activated form of IKK on NF-kappaB and anti-apoptotic gene expressions: a systems biology approach. *FEBS Lett.* 580, 822–830. doi: 10.1016/j.febslet.2006.01.004
- Petropoulos, S., Edsgard, D., Reinis, B., Deng, Q., Panula, S. P., Codeluppi, S., et al. (2016). Single-cell RNA-Seq reveals lineage and X chromosome dynamics in human preimplantation embryos. *Cell* 165, 1012–1026. doi: 10.1016/j.cell.2016.03.023
- Powell, S. M., Zilz, N., Beazer-Barclay, Y., Bryan, T. M., Hamilton, S. R., Thibodeau, S. N., et al. (1992). APC mutations occur early during colorectal tumorigenesis. *Nature* 359, 235–237. doi: 10.1038/359235a0
- Quereda, V., and Malumbres, M. (2009). Cell cycle control of pituitary development and disease. *J. Mol. Endocrinol.* 42, 75–86. doi: 10.1677/jme-08-0146
- Schmidt, H., Cho, K. H., and Jacobsen, E. W. (2005). Identification of small scale biochemical networks based on general type system perturbations. *FEBS J.* 272, 2141–2151. doi: 10.1111/j.1742-4658.2005.04605.x
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M. 3rd, et al. (2019). Comprehensive integration of single-cell data. *Cell* 177, 1888.e1821–1902.e1821. doi: 10.1016/j.cell.2019.05.031
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-

- based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* 102, 15545–15550. doi: 10.1073/pnas.0506580102
- Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47, D607–D613. doi: 10.1093/nar/gky1131
- The Gene Ontology Consortium (2019). The Gene Ontology resource: 20 years and still GOing strong. *Nucleic Acids Res.* 47, D330–D338. doi: 10.1093/nar/gky1055
- Tirosh, I., Izar, B., Prakadan, S. M., Wadsworth, M. H. 2nd, Treacy, D., Trombetta, J. J., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* 352, 189–196. doi: 10.1126/science.aad0501
- Valenta, T., Hausmann, G., and Basler, K. (2012). The many faces and functions of β -catenin. *EMBO J.* 31, 2714–2736. doi: 10.1038/emboj.2012.150
- Wang, K., Saito, M., Bisikirska, B. C., Alvarez, M. J., Lim, W. K., Rajbhandari, P., et al. (2009). Genome-wide identification of post-translational modulators of transcription factor activity in human B cells. *Nat. Biotechnol.* 27, 829–839. doi: 10.1038/nbt.1563
- Zheng, G. X., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., et al. (2017). Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* 8:14049. doi: 10.1038/ncomms14049

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Choi, Gong, Hwang, Joung, Lee and Cho. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Quantifying the Stability of Coupled Genetic and Epigenetic Switches With Variational Methods

Amogh Sood and Bin Zhang*

Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA, United States

OPEN ACCESS

Edited by:

Chunhe Li,
Fudan University, China

Reviewed by:

Masaki Sasai,
Nagoya University, Japan
Yueheng Lan,
Beijing University of Posts and
Telecommunications (BUPT), China

*Correspondence:

Bin Zhang
binz@mit.edu

Specialty section:

This article was submitted to
Systems Biology,
a section of the journal
Frontiers in Genetics

Received: 01 December 2020

Accepted: 29 December 2020

Published: 22 January 2021

Citation:

Sood A and Zhang B (2021)
Quantifying the Stability of Coupled
Genetic and Epigenetic Switches With
Variational Methods.
Front. Genet. 11:636724.
doi: 10.3389/fgene.2020.636724

The Waddington landscape provides an intuitive metaphor to view development as a ball rolling down the hill, with distinct phenotypes as basins and differentiation pathways as valleys. Since, at a molecular level, cell differentiation arises from interactions among the genes, a mathematical definition for the Waddington landscape can, in principle, be obtained by studying the gene regulatory networks. For eukaryotes, gene regulation is inextricably and intimately linked to histone modifications. However, the impact of such modifications on both landscape topography and stability of attractor states is not fully understood. In this work, we introduced a minimal kinetic model for gene regulation that combines the impact of both histone modifications and transcription factors. We further developed an approximation scheme based on variational principles to solve the corresponding master equation in a second quantized framework. By analyzing the steady-state solutions at various parameter regimes, we found that histone modification kinetics can significantly alter the behavior of a genetic network, resulting in qualitative changes in gene expression profiles. The emerging epigenetic landscape captures the delicate interplay between transcription factors and histone modifications in driving cell-fate decisions.

Keywords: gene expression noise, minimum action, chromatin state, gene network, self-regulating gene

1. INTRODUCTION

A little more than five decades ago, Waddington introduced the metaphor to view cellular differentiation into distinct lineages and cell types as a sequence of transitions among basins in a landscape, wherein basins indicate stable phenotypes (Waddington and Kacser, 1957). The appeal of this metaphor to intuition has inspired efforts of theoretical formulation at the molecular level by studying genetic networks formed by transcription factors (TF) (Sasai and Wolynes, 2003; Hornos et al., 2005; Kærn et al., 2005; Walczak et al., 2005a,b; Xu and Tao, 2006; Goldberg et al., 2007; Kim and Wang, 2007; Shahrezaei and Swain, 2008; Cao et al., 2010; Venegas-Ortiz and Evans, 2011; Wang et al., 2011, 2014; Zhang et al., 2013; Zhang and Wolynes, 2014; Lv et al., 2015; Chen et al., 2016; Qiu et al., 2020). These studies highlighted the importance of gene expression noise in driving the transition among steady states. Noise is a manifestation of the inherent stochasticity of chemical reactions and arises in gene regulatory networks as a result of protein production/degradation and TF binding/unbinding. Noise, or fluctuation, is non-negligible due to the finite number of protein molecules and the single molecule nature of DNA. Stochastic noise and network topology together define the epigenetic landscape, much like the one envisioned by Waddington, that quantifies the stability of various cell-defining gene expression levels or patterns.

For eukaryotic organisms, in addition to transcription factors, epigenetic marks such as DNA methylation and histone modifications also play essential roles in regulating gene expression (Lister et al., 2009; Lu et al., 2009; Artyomov et al., 2010; Krishnakumar and Kraus, 2010; Margueron and Reinberg, 2010; Mariani et al., 2010; Andrew Angel, 2011; Miller-Jensen et al., 2011; Furey and Sethupathy, 2013). They are known to affect local chromatin packaging and global genome organization (Zhou et al., 2011; Schlick et al., 2012; Rowley and Corces, 2018; Parsons and Zhang, 2019; Qi et al., 2020; Xie et al., 2020), which in turn can regulate DNA accessibility to regulatory proteins. Furthermore, DNA methylation directly impacts the DNA binding affinity of transcription factors (Tate and Bird, 1993; Zhou et al., 2016; Flavahan et al., 2019). Importantly, the chemical modifications themselves may give rise to steady states independent of the TF-centric genetic network. For example, modification of nucleosomes recruits enzymes affecting the neighboring nucleosomes, causing them to be similarly modified (Bannister and Kouzarides, 2011). Many elegant theoretical attempts have demonstrated how such interactions can bring about collective changes of many nucleosomes and allow them to exhibit distinct multistable states (Dodd et al., 2007; Sedighi and Sengupta, 2007; David-Rus et al., 2009; Micheelsen et al., 2010; Sneppen and Mitarai, 2012; Dayarian and Sengupta, 2013; Jost, 2014; Sood and Zhang, 2020). Therefore, it is crucial to account for the dynamics and regulation of epigenetic modifications when constructing the landscape for cellular differentiation in eukaryotes.

Many research groups have studied the interplay between genetic and epigenetic switches in regulating gene expression. For instance, generalized genetic networks that couple each gene to a binary or ternary variable representing the collective histone states have been used as models for stem cells to account for epigenetic degrees of freedom, albeit in a coarse grained fashion (Artyomov et al., 2010; Binder et al., 2013; Sasai et al., 2013; Ashwin and Sasai, 2015; Huang and Lei, 2018; Folguera-Blasco et al., 2019). These studies found a significant dependence of the probability landscape of protein expression computed from stochastic simulations on chromatin state dynamics. Similarly coarse-grained treatment of epigenetic switches was shown to introduce hysteresis (Bhattacharyya et al., 2020) and homeorhesis (Matsushita and Kaneko, 2020) to the dynamics of gene regulatory networks. Notably, Zhang et al. (2019) explicitly considered the modification of individual nucleosomes and studied the impact of such modifications on the probability landscape of a single self-activating gene and a pair of mutually repressive genes. However, the lack of analytical results has made the sensitivity analysis of the computed landscape with respect to parameter values, which may vary along cell differentiation, numerically challenging.

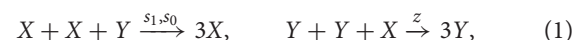
In this work, we investigate the combined impact of TF binding and epigenetic modifications in regulating the expression of a self-activating gene. Rather than coarse-graining the epigenetic switch into a binary or ternary variable, we explicitly account for the dynamical modification of individual nucleosomes. The variational approach (Eyink, 1996; Sasai and Wolynes, 2003) was used to compute steady-state

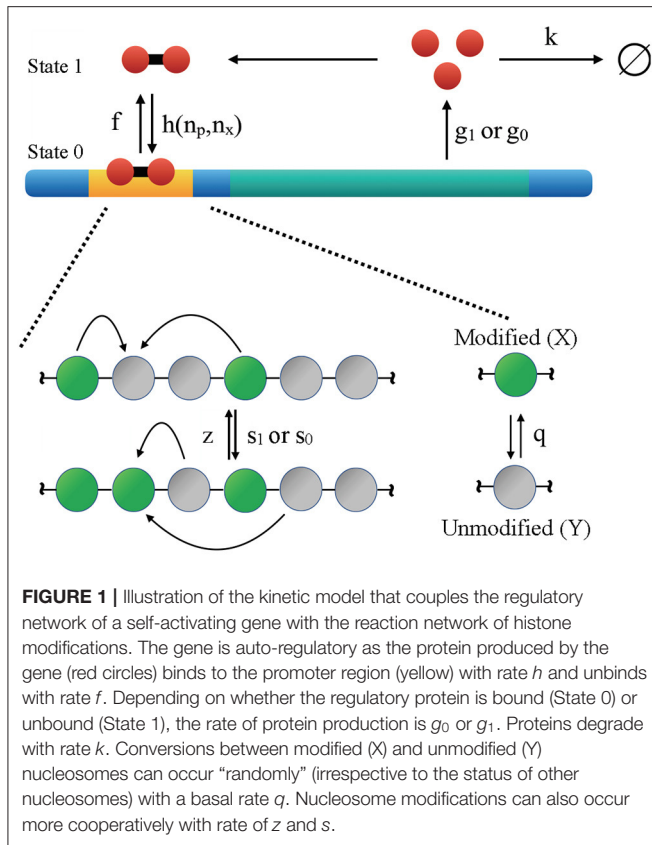
probability distributions from deterministic equations and avoid computationally intensive stochastic simulations. Moreover, we generalize the typically used Poisson ansatz to better treat systems with particle conservation constraints, such as our epigenetic switch, that are more naturally described using SU(2) than Bosonic operators (Sood and Zhang, 2020). The approach enabled a convenient exploration of the model's steady-state behavior across a wide range of parameters. Our study suggests that fast, random perturbations to individual histone modifications lead to the formation of a poised, uncommitted chromatin state, which in turn can drive noisy gene expression seen in stem cells. As the rate of such random perturbations decreases and the role of cooperative modifications of nucleosome prevails, the system transitions to a bistable regime resembling a differentiated state. The transition goes through an activated state with high gene expression, highlighting the robustness of the network in activating gene expression due to the feedback between genetic and epigenetic switches. We further compared variational results with stochastic simulations and discussed potential improvements in the accuracy of the variational method.

2. MODEL

We consider a simplified model of eukaryotic gene regulation that accounts for TF binding/unbinding as well as histone modifications. The model couples the regulatory network of a self-activating gene with an epigenetic switch that can lead to active and repressive chromatin states.

For self-activating genes, their protein products bind with the promoter to upregulate the transcription rate. As illustrated in **Figure 1**, proteins are produced and destroyed with rates of g and k , respectively. The protein production rate is further dependent on whether the gene's promoter is bound by TF (state 0) or not (state 1), and we have $g_1 < g_0$ since the proteins are activators. Here TFs correspond to gene transcription products, and they bind to the promoter with rate h as dimers. The corresponding unbinding rate is f . Binding rate depends on protein copy number n_p as well as the number of modified nucleosomes n_x as detailed in Equation (3) below. Self-activating genes are known to occur both as isolated entities (Ptashne et al., 1980; Johnson et al., 1981; Hasty et al., 2000; Rosenfeld et al., 2002) and as common motifs of larger interacting networks (Ralston and Rossant, 2005; Loh et al., 2008; Orkin and Zon, 2008). They have been the subject of extensive theoretical study as models of cellular differentiation (Sasai and Wolynes, 2003; Hornos et al., 2005; Walczak et al., 2005a,b; Xu and Tao, 2006; Goldberg et al., 2007; Kim and Wang, 2007; Shahrezaei and Swain, 2008; Venegas-Ortiz and Evans, 2011; Wang et al., 2011; Zhang et al., 2013; Zhang and Wolynes, 2014). The epigenetic switch concerns a cluster of $N = 60$ nucleosomes, each of which can exist in a modified (X) or unmodified (Y) state. The kinetics of chromatin system can be described with the non-linear dynamics given below





Furthermore, the recruited conversion rate of unmodified to modified nucleosomes depends on TF binding with $s_0 > s_1$, assuming that TFs can attract modification enzymes to chromatin. The values for the kinetic parameters were set relative to the degradation rate k as $g_1 = 4$, $g_0 = 65$, $h_0 = 1$, $f = 100$, $s_1 = 8$, $s_0 = 10s_1$, $z = 8$. The random histone modification rate, q , was varied over a wide range of values as detailed below. We used $k = 1s^{-1}$, though changing this value will not affect the steady state distributions and only renormalizes the timescale in the model.

We carried out stochastic simulations of the kinetic model using the Gillespie algorithm (Gillespie, 1977). Each plot shown in **Figure 2** was obtained from averaging over 100 independent 10^5 -second-long simulations. These trajectories were initialized with random configurations, and the number of modified nucleosomes and protein molecules along each trajectory was recorded at every second. We then combined the values from all trajectories to estimate the steady state probability distributions, P_{ss} . For the plots shown in **Figure 3** we used $q = 10$ and set $n_x = 40$ and $n_p = 20$ at $t = 0$. 200 independent trajectories were performed to produce the average numbers recorded at every 0.5 s.

3. THEORY

We reformulated the master equation describing the dynamical evolution of the kinetic network as an *imaginary* time Schrödinger equation

$$X \xrightarrow{q} Y, \quad Y \xrightarrow{q} X. \quad (2)$$

The inter-conversion between modified and unmodified nucleosomes can either proceed via Equation (1) that requires a pair of similarly modified nucleosomes to alter the state of a nucleosome, or via noisy conversion (Equation 2) with first-order kinetics. The former is meant to account for nucleosomes being actively interconverted by modifying and removing enzymes recruited by the similarly modified nucleosomes in their vicinity. It is this recruitment that forms the positive feedback in the system (Dodd et al., 2007; Micheelsen et al., 2010; Xie and Zhang, 2019; Sood and Zhang, 2020). s , z , and q are the rate constants of the corresponding reactions.

The coupling between the genetic and epigenetic switch is achieved by introducing a dependence of protein binding rate on the number of modified nucleosomes, i.e.,

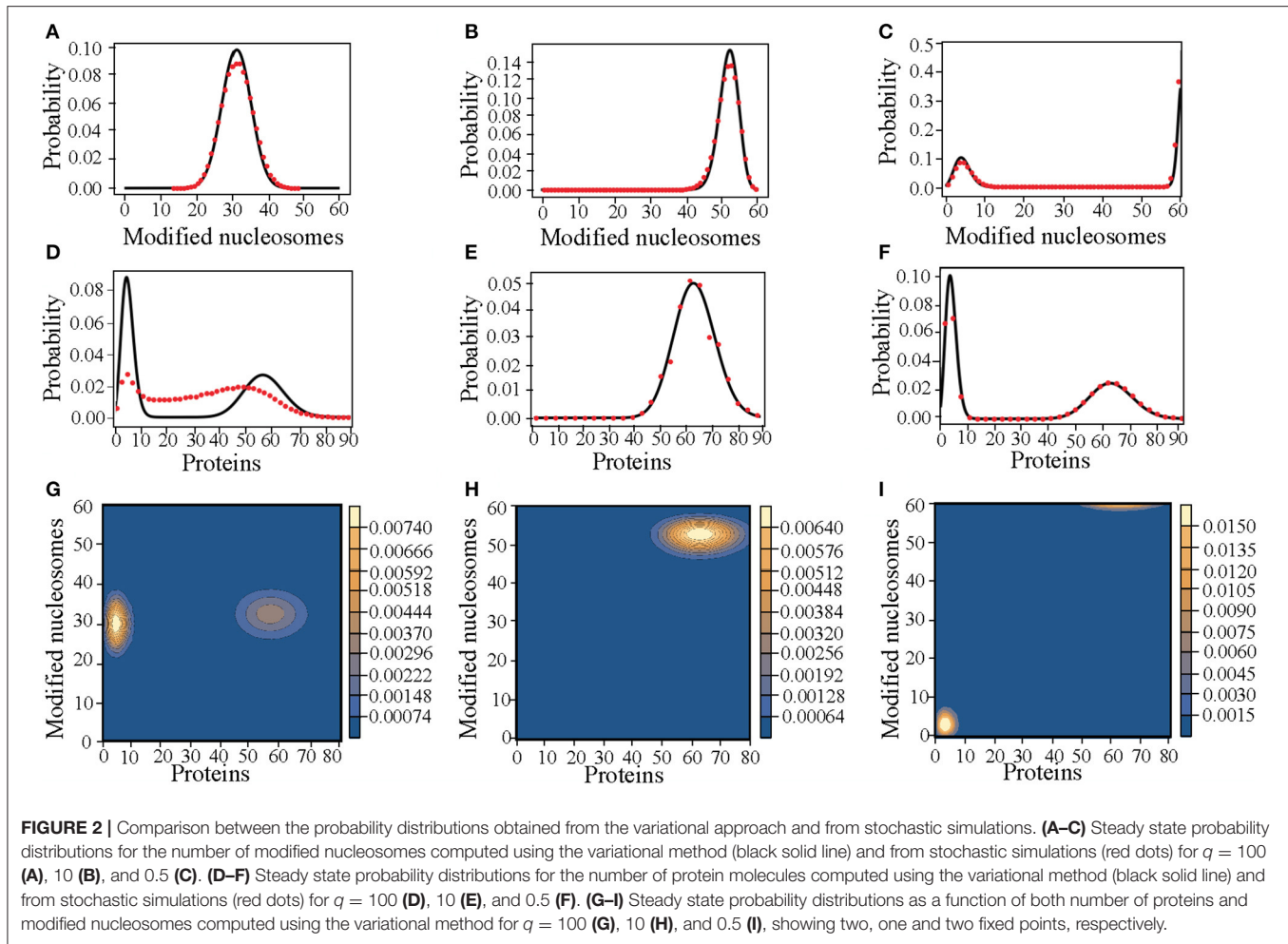
$$h(n_p, n_x) = h_0 \frac{n_p(n_p - 1)}{1 + \exp(-0.5(n_x - 35))}. \quad (3)$$

This dependence is motivated by the realization that actively modified chromatin ($n_x > 35$) exists in a more open state that is more accessible to regulatory proteins. The particular expression $[1 + \exp(-0.5(n_x - 35))]^{-1}$ as the probability for chromatin being open is typical of a two state system, assuming that the energetic difference between open and closed chromatin depends linearly on the number of modified nucleosomes.

$$\partial_t |\Psi(t)\rangle = \Omega |\Psi(t)\rangle. \quad (4)$$

The state vector $|\Psi(t)\rangle = \begin{pmatrix} \Psi_1(t) \\ \Psi_0(t) \end{pmatrix}$ is a superposition of all possible configurations weighted with their corresponding probabilities such that $\Psi_i(t) = \sum_{n_p, n_x} P_i(\{n_p, n_x\}; t) |n_p, n_x\rangle$ for $i = 0, 1$. The two components correspond to the DNA state with regulatory proteins unbound (state 1) or bound (state 0), respectively. This reformulation makes use of a second quantization based method (the Doi-Peliti approach), which has been successfully employed in the study of reaction-diffusion processes (Lee and Cardy, 1995), gene switches (Sasai and Wolynes, 2003; Zhang and Wolynes, 2014), and other systems (Täuber, 2014). In previous work, we applied the Doi-Peliti approach to the epigenetic switch using operators that are a representation of the SU(2) algebra (Sood and Zhang, 2020). The SU(2) algebra allows us to treat the constraint of conservation of particle in number in a mathematically elegant and convenient way. When coupled to the self-activating gene, the stochastic Hamiltonian for the system described in **Figure 1** is given by

$$\begin{aligned} \Omega = & g(a_p^\dagger - 1) + k(a_p - a_p^\dagger a_p) + s[J_+ \hat{n}_x^2 - \hat{n}_x^2 \hat{n}_y] \\ & + z[J_- \hat{n}_y^2 - \hat{n}_y^2 \hat{n}_x] + q[J_- - \hat{n}_x] + q[J_+ - \hat{n}_y] \\ & + \begin{pmatrix} -h(\hat{n}_p, \hat{n}_x) & f \\ h(\hat{n}_p, \hat{n}_x) & -f \end{pmatrix}, \end{aligned} \quad (5)$$



where $\mathbf{g} = \begin{pmatrix} g_1 & 0 \\ 0 & g_0 \end{pmatrix}$, $\mathbf{s} = \begin{pmatrix} s_1/N^3 & 0 \\ 0 & s_0/N^3 \end{pmatrix}$, $\mathbf{z} = \begin{pmatrix} z/N^3 & 0 \\ 0 & z/N^3 \end{pmatrix}$, $\mathbf{k} = \begin{pmatrix} k & 0 \\ 0 & k \end{pmatrix}$, $\mathbf{q} = \begin{pmatrix} q/N & 0 \\ 0 & q/N \end{pmatrix}$, and $h(\hat{n}_p, \hat{n}_x) = \frac{\hat{n}_p(\hat{n}_p-1)}{1+\exp(-0.5(\hat{n}_x-35))}$. The operator $a_p^\dagger$ creates a protein molecule when it acts on a state, $a_p^\dagger |n_p, n_x\rangle = |n_p + 1, n_x\rangle$, whereas a_p serves to remove a protein molecule when acting on the same state, $a_p |n_p, n_x\rangle = |n_p - 1, n_x\rangle$. J_+ converts an unmodified nucleosome to a modified one by acting on a state, $J_+ |n_p, n_x\rangle = (N - n_x) |n_p, n_x + 1\rangle$, while J_- acts to convert a modified nucleosome to an unmodified one, $J_- |n_p, n_x\rangle = n_x |n_p, n_x - 1\rangle$. $\hat{n}_p$ denotes the number operator, as its action on a ket gives the number of protein molecules, $\hat{n}_p |n_p, n_x\rangle = n_p |n_p, n_x\rangle$. In a similar fashion, $\hat{n}_x$ gives the number of modified nucleosomes when it acts on a ket, $\hat{n}_x |n_p, n_x\rangle = n_x |n_p, n_x\rangle$, and $\hat{n}_y$ gives the number of unmodified nucleosomes, $\hat{n}_y |n_p, n_x\rangle = (N - n_x) |n_p, n_x\rangle$. $n^2 = n(n-1)$ denotes the falling factorial.

Exact solutions to Equation (4) are difficult to obtain. Instead, we make use of an approximate, yet succinct and powerful, variational approach originally introduced by Eyink (Eyink, 1996; Alexander and Eyink, 1997). First, we realize that

the imaginary time Schrödinger equation is equivalent to the functional variation of the following action Γ with respect to Φ , i.e., $\frac{\delta \Gamma}{\delta \Phi} = 0$ for

$$\Gamma = \int dt \langle \Phi | \partial_t - \Omega | \Psi \rangle. \quad (6)$$

By designing trial functions for Φ and Ψ parameterized with $\alpha_L = \alpha_L^1, \alpha_L^2, \dots, \alpha_L^K$ and $\alpha_R = \alpha_R^1, \alpha_R^2, \dots, \alpha_R^K$, minimizing the action leads to a set of ordinary differential equations,

$$\sum_{l=1}^K \left[\left\langle \frac{\partial \Phi}{\partial \alpha_L^m} \right| \frac{\partial \Psi}{\partial \alpha_R^l} \frac{d\alpha_R^l}{dt} - \left\langle \frac{\partial \Phi}{\partial \alpha_L^m} \right| \Omega | \Psi \right]_{\alpha_L^m=0} = 0, \quad (7)$$

$$\text{for } m = 1, \dots, K. \quad (8)$$

Also, we demand (to stay true to the probabilistic interpretation) $\langle \Phi(\alpha_L = 0) | \Psi(\alpha_R) \rangle = 1$. The variational approach was first applied with great success to stochastic gene regulatory networks by Sasai and Wolynes (2003). In its original formulation, Poisson distributions were used as trial functions, with the Poisson mean

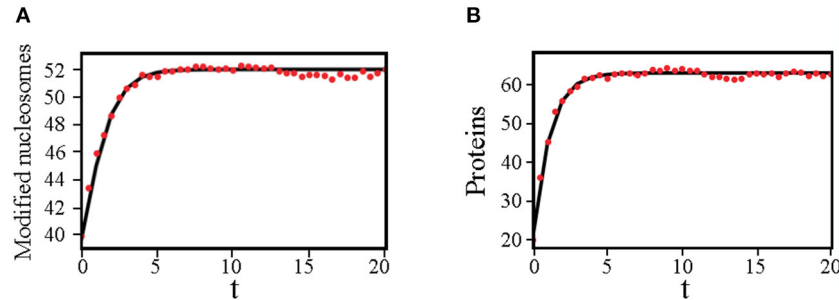


FIGURE 3 | Dynamical trajectories determined from the variational approach agree well with stochastic simulations in favorable regimes. **(A)** Time evolution of the average number of modified nucleosomes computed using the variational method (black solid line) and stochastic simulations (red dots). **(B)** Time evolution of the average number of modified nucleosomes computed using the variational method (black solid line) and stochastic simulation (red dots). We used $q = 10$, $M = 60$, and set $c_1 p_1 = 0$, $c_0 p_0 = 20$, $c_1 t_1 = 0$, $c_1 t_0 = 0.66$ as the initial values when solving the deterministic equations (Equation 11).

being the variational parameter. Since protein molecules can be approximately treated as products of a birth-death process, the probability distribution to find n_p molecules should be Poisson at large t (Sasai and Wolynes, 2003). Furthermore, the stochastic Hamiltonian for genetic networks consists of only Bosonic operators, the coherent states of which correspond to Poisson distributions. In this work, we exploit the symmetry imposed on the system by particle number constraints to derive a new variational trial function for the chromatin switch. As shown in the **Supplementary Material**, an excellent candidate is the binomial distribution function since the coherent states for the SU(2) operators in our stochastic Hamiltonian are binomial (Fu and Sasaki, 1997, 1998). Taken together, we can thus use the following ansatz as variational functions for the coupled genetic and epigenetic switch

$$|\Psi\rangle = \begin{pmatrix} c_1 \exp(p_1(a_p^\dagger - 1))(1 - \theta_1)^N \exp\left(\frac{\theta_1}{1 - \theta_1} J_+\right) |0, 0\rangle \\ c_0 \exp(p_0(a_p^\dagger - 1))(1 - \theta_0)^N \exp\left(\frac{\theta_0}{1 - \theta_0} J_+\right) |0, 0\rangle \end{pmatrix}, \quad (9)$$

and

$$\langle \Phi | = \left(\langle 0, 0 | e^{a_p} e^{J_-} \exp(\alpha_1 + \lambda_1^{(p)} a_p + \lambda_1^{(x)} J_-) \right. \\ \left. \langle 0, 0 | e^{a_p} e^{J_-} \exp(\alpha_0 + \lambda_0^{(p)} a_p + \lambda_0^{(x)} J_-) \right). \quad (10)$$

The set of variational parameters is $\alpha_R = \{c_1, c_0, p_1, p_0, \theta_1, \theta_0\}$. Here $c_1(c_0)$ represents the probability of the DNA being in state 1 (state 0), while $p_1(p_0)$ and $N\theta_1(N\theta_0)$ represent the mean number of proteins and modified nucleosomes when DNA is in state 1 (state 0). $\alpha_L = \{\alpha_1, \alpha_0, \lambda_1^{(p)}, \lambda_0^{(p)}, \lambda_1^{(x)}, \lambda_0^{(x)}\}$ are the corresponding conjugate variables.

Plugging (10) and (9) into (7), we obtain the following set of variational equations

$$\frac{dc_1}{dt} = c_0 f - c_1 \langle h(n_p, n_x) \rangle_1 \quad (11a)$$

$$\frac{dc_0}{dt} = -c_0 f + c_1 \langle h(n_p, n_x) \rangle_1 \quad (11b)$$

$$c_1 \frac{dp_1}{dt} + p_1 \frac{dc_1}{dt} = c_1 g_1 - c_1 k p_1 + c_0 f p_0 - c_1 \langle n_p h(n_p, n_x) \rangle_1 \quad (11c)$$

$$c_0 \frac{dp_0}{dt} + p_0 \frac{dc_0}{dt} = c_0 g_0 - c_0 k p_0 - c_0 f p_0 + c_1 \langle n_p h(n_p, n_x) \rangle_1 \quad (11d)$$

$$N\theta_1 \frac{d\theta_1}{dt} + Nc_1 \frac{d\theta_1}{dt} = c_1 \left(\frac{s_1}{N^3} \right) \langle n_x^2 (N - n_x) \rangle_1 \\ - c_1 \left(\frac{z_1}{N^3} \right) \langle (N - n_x)^2 (n_x) \rangle_1 \\ + c_1 \frac{q}{N} (-\langle n_x \rangle_1 + \langle N - n_x \rangle_1) \\ + c_0 f N \theta_0 - c_1 \langle n_x h(n_p, n_x) \rangle_1 \quad (11e)$$

$$N\theta_0 \frac{d\theta_0}{dt} + Nc_0 \frac{d\theta_0}{dt} = c_0 \left(\frac{s_0}{N^3} \right) \langle n_x^2 (N - n_x) \rangle_0 \\ - c_0 \left(\frac{z_0}{N^3} \right) \langle (N - n_x)^2 (n_x) \rangle_0 \\ + c_0 \frac{q}{N} (-\langle n_x \rangle_0 + \langle N - n_x \rangle_0) - c_0 f N \theta_0 \\ + c_1 \langle n_x h(n_p, n_x) \rangle_1. \quad (11f)$$

The angular brackets represent ensemble averaging over protein numbers and modified nucleosomes, i.e., $\langle \cdot \rangle_i = \sum_{n_p, n_x} \frac{e^{-p_i} p_i^{n_p}}{n_p!} \frac{N!}{n_x!} \theta_k^{n_x} (1 - \theta_k)^{N - n_x}$. We also make the simplifying approximation for the average binding rate as $\langle h(n_p, n_x) \rangle = \left\langle \frac{n_p(n_p - 1)}{1 + \exp(0.5(n_x - 35))} \right\rangle \approx \frac{\langle n_p(n_p - 1) \rangle}{1 + \exp(0.5(\langle n_x \rangle - 35))}$. Numerical integration of Equation (11) yields the time evolution of the variational parameters α_R , from which the probability distributions can be determined using Equation (9).

We solved Equation (11) using `scipy.integrate.odeint()` module in python with a time step of 0.01 s. The initial conditions were varied and individual trajectories were integrated for 10^5 s till convergence to obtain the steady state results.

4. RESULTS

Using the variational equations, we studied the dependence of steady-state solutions on the rate of noisy histone mark modification, q . For comparison, we carried out stochastic simulations of the kinetic network using the Gillespie algorithm (Gillespie, 1977) at selected q -values. The noisy modification rate and, in particular, its relative value to the rate for recruited conversions is an important parameter for cell differentiation. For example, recruited conversions arise due to the diffusion of histone-modifying enzymes from modified nucleosomes to the nearby unmodified ones. The more open chromatin conformation seen in stem cells with larger inter-nucleosome distances (Gaspar-Maia et al., 2011; Mas et al., 2018) will, therefore, suppress recruited conversions in favor of the noisy ones. As cells differentiate, chromatin will become more compact, and the importance of noisy conversions will decline. Previous studies of isolated epigenetic switches (Dodd et al., 2007; Micheelsen et al., 2010; Sood and Zhang, 2020) also found q as an important parameter that controls the onset and maintenance of bistability in the epigenetic switch.

In **Figure 2**, we show the probability distributions obtained from stochastic simulations and from the variational approach at $q = 100, 10$, and 0.5 . We notice that the Binomial ansatz introduced in the *Theory* section captures the distribution for the number of modified nucleosomes with quantitative accuracy (**Figures 2A–C**). The Poissonian ansatz also performs well for the distribution of protein numbers at small and medium q values, though deviations from stochastic simulations are apparent at large q (**Figures 2D–F**). The inconsistency between the two distributions in that regime is mainly due to underestimating the population of intermediate states that bridge the high and low gene expression values by the variation method.

In addition to steady-state solutions, the time evolution of observables, such as the mean number of proteins and modified nucleosomes, can be determined using the variational approach as well. As shown in **Figure 3**, in parameter regimes where the effect of fluctuations is not too drastic, the dynamical trajectories determined using Equation (11) are in quantitative agreements with those computed using stochastic simulations.

Given its reasonable performance, we next applied the variational approach to study the network model's steady-state behavior at a broader range of q -values. As already mentioned, q is an important variable that might be tuned along the developmental axis for cell differentiation. For large q values, chromatin stabilizes in an undecided state with roughly half the nucleosomes modified (active) and the other half carrying no modification (repressive). The corresponding protein expression is noisy with a broad probability distribution. Stochastic simulations further support a significant mixing between “on” and “off” gene states, and an unambiguous assignment of either state is not warranted (**Figure 2D**). When the value for q is quenched, we observe the emergence of a coherent epigenetic state along with coherent gene expression. Therefore, both switches are turned on and the combined system exhibits a single attractor. At even lower values of q , both the epigenetic and gene switch exhibit bistability.

We note that the chromatin state changes described above differs from that of an isolated epigenetic switch studied previously (Sood and Zhang, 2020). There, we saw a shift from a unimodal probability distribution indicating an equal admixture of modified and unmodified nucleosomes to a symmetric bimodal probability distribution as the value for q is quenched. The appearance of a single coherent epigenetic state in **Figure 4** is a result of the coupling with the gene switch in our model, which breaks the symmetry between active and repressive chromatin states. The coupling works both ways. In an isolated gene switch, a single state with high gene expression is not expected either. Modulating the kinetics of TF binding to the promoter only resolves a broad probability distribution exhibiting no coherent gene expression to a bistable state with high and low levels of gene expression (Walczak et al., 2005a).

5. DISCUSSION

We introduced a kinetic model that couples a genetic network with an epigenetic switch to study the combined role of transcription factors and histone modifications in regulating gene expression. An approximation scheme based on the variational approach was further developed to obtain steady-state solutions. This method is unencumbered by the complexity associated with numerical simulations and more detailed analytical calculations. It would be a useful tool for exploratory studies of the parameter space and identifying regions of interest. While we focused our analysis on a single gene, the variational method can be relatively easily generalized to networks with multiple interacting genetic and epigenetic switches that provide more realistic modeling of stem cell differentiation (Zhang and Wolynes, 2014).

We explored the behavior of the network model across a wide range of parameters. Our model exhibits a poised state for the gene switch at high q , where the chromatin system contains an equal admixture of modified and unmodified nucleosomes. The network in this parameter regime appears to qualitatively capture the behavior of chromatin and gene expression in undifferentiated stem cells. In particular, stem cells are known to exhibit bivalent chromatin with both activating and repressive marks (Bernstein et al., 2006; Vastenhouw and Schier, 2012) and noisy gene expression profiles (Kar et al., 2017). We point out that the exact definition of bivalent chromatin remains controversial, and multiple mechanisms have been proposed for its formation (Azuara et al., 2006; Sneppen and Ringrose, 2019; Lim and Meshorer, 2020). Additional studies are needed to determine whether the stochastic conversion observed here is the key driver for the observed chromatin bivalency.

Upon quenching q , the gene is activated along with a concomitant resolution of the chromatin state. The coupling between the two switches reinforces the stability of the active state and can lead to more robust upregulation of gene expression upon cell differentiation. It also ensures that the genetic and epigenetic switches are always in sync. We observe at most two steady states representing active chromatin with high gene expression and repressive chromatin with low gene expression. We note that the inactive state only becomes stable at minimal

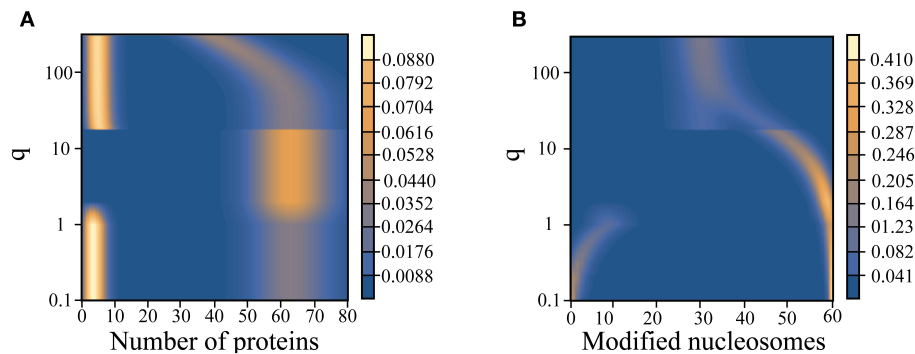


FIGURE 4 | Variation of the steady state probability distribution for the number of proteins **(A)** and modified nucleosomes **(B)** as a function of the noisy histone modification rate, q .

q values, arguing for strong noise suppression for gene silencing. Its limited stability may explain the presence of DNA methylation on top of histone modifications to safeguard the silent state against perturbations that might arise from fluctuation in protein concentration or histone marks during cell division.

The strong dependence of the landscape tomography on q shown in **Figure 4** suggests that the histone modification rate may act like a knob to be tuned along the developmental axis to facilitate cellular differentiation. Of course, the presented landscape is probably too crude a simplification to be termed the Waddington landscape since many additional factors that contribute to the stability of gene expression patterns could be varied along the developmental axis as well.

In favorable regimes, the variational approach produces results of quantitative accuracy. The discrepancy between the probability distribution obtained from stochastic simulations and the variational method in the high q region can be attributed to the fact that the Poisson ansatz does not sufficiently account for the variance and the effect of fluctuations which become increasingly important as the value for q increases. This situation can be remedied by going beyond the Poisson ansatz, and utilizing the superposition ansatz as described in Ohkubo (2008). Mathematically, this would mean to modify our ansatz as follows,

$$|\Psi\rangle = \left(\int_0^\infty dp_1 \mathcal{F}(p_1; \{\lambda_j^{(1)}\}) c_1 \exp\left(p_1(a_p^\dagger - 1)\right) (1 - \theta_1)^N \exp\left(\frac{\theta_1}{1 - \theta_1} J_+\right) |0, 0\rangle \right) + \left(\int_0^\infty dp_0 \mathcal{F}(p_0; \{\lambda_j^{(0)}\}) c_0 \exp\left(p_0(a_p^\dagger - 1)\right) (1 - \theta_0)^N \exp\left(\frac{\theta_0}{1 - \theta_0} J_+\right) |0, 0\rangle \right). \quad (12)$$

This new “superposition ansatz” is constructed by the superposition of the coherent states (i.e., Poisson distribution) as defined in (12), where now $\mathcal{F}$ serves as the variational function. Hence, the real probability distribution is obtained by the superposition of the Poisson distributions of mean p_i weighed by the distribution $\mathcal{F}$ with parameters $\{\lambda_j^{(i)}\}$. We anticipate that

doing so can not only improve the agreement between theory and simulation but can in principle allow for the computation of time evolution of other interesting quantities such as variance, and covariance in addition to means. However, in general the choice of an appropriate $\mathcal{F}$ is a non-trivial problem, and thus has been avoided in this text in favor of a clearer exposition. The choice of appropriate variational functions can be guided by the work done on exact solutions of the master equations of genetic switches (Hornos et al., 2005; Shahrezaei and Swain, 2008; Ramos et al., 2011).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

AS and BZ conceived and designed the work, interpreted the results, and wrote the manuscript. AS carried out computer implementation and data analysis. BZ supervised the project. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Institutes of Health (Grant 1R35GM133580).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.636724/full#supplementary-material>

REFERENCES

- Alexander, F. J., and Eyink, G. L. (1997). Rayleigh-ritz calculation of effective potential far from equilibrium. *Phys. Rev. Lett.* 78, 1–4. doi: 10.1103/PhysRevLett.78.1
- Angel, A., Song, J., Dean, C., and Howard, M. (2011). A polycomb-based switch underlying quantitative epigenetic memory. *Nature* 476, 105–108. doi: 10.1038/nature10241
- Artyomov, M. N., Meissner, A., and Chakraborty, A. K. (2010). A model for genetic and epigenetic regulatory networks identifies rare pathways for transcription factor induced pluripotency. *PLoS Comput. Biol.* 6:e1000785. doi: 10.1371/journal.pcbi.1000785
- Ashwin, S. S., and Sasai, M. (2015). Effects of collective histone state dynamics on epigenetic landscape and kinetics of cell reprogramming. *Sci. Rep.* 5, 1–12. doi: 10.1038/srep16746
- Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jørgensen, H. F., John, R. M., et al. (2006). Chromatin signatures of pluripotent cell lines. *Nat. Cell Biol.* 8, 532–538. doi: 10.1038/ncb1403
- Bannister, A. J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Res.* 21, 381–395. doi: 10.1038/cr.2011.22
- Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125, 315–326. doi: 10.1016/j.cell.2006.02.041
- Bhattacharyya, B., Wang, J., and Sasai, M. (2020). Stochastic epigenetic dynamics of gene switching. *Phys. Rev. E* 102:042408. doi: 10.1103/PhysRevE.102.042408
- Binder, H., Steiner, L., Przybilla, J., Rohlf, T., Prohaska, S., and Galle, J. (2013). Transcriptional regulation by histone modifications: towards a theory of chromatin re-organization during stem cell differentiation. *Phys. Biol.* 10:026006. doi: 10.1088/1478-3975/10/2/026006
- Cao, Y., Lu, H.-M., and Liang, J. (2010). Probability landscape of heritable and robust epigenetic state of lysogeny in phage lambda. *Proc. Natl. Acad. Sci. U.S.A.* 107, 18445–18450. doi: 10.1073/pnas.1001455107
- Chen, H., Thill, P., and Cao, J. (2016). Transitions in genetic toggle switches driven by dynamic disorder in rate coefficients. *J. Chem. Phys.* 144:175104. doi: 10.1063/1.4948461
- David-Rus, D., Mukhopadhyay, S., Lebowitz, J. L., and Sengupta, A. M. (2009). Inheritance of epigenetic chromatin silencing. *J. Theor. Biol.* 258, 112–120. doi: 10.1016/j.jtbi.2008.12.021
- Dayarian, A., and Sengupta, A. M. (2013). Titration and hysteresis in epigenetic chromatin silencing. *Phys. Biol.* 10:036005. doi: 10.1088/1478-3975/10/3/036005
- Dodd, I. B., Micheelsen, M. A., Sneppen, K., and Thon, G. (2007). Theoretical analysis of epigenetic cell memory by nucleosome modification. *Cell* 129:813. doi: 10.1016/j.cell.2007.02.053
- Eyink, G. L. (1996). Action principle in nonequilibrium statistical dynamics. *Phys. Rev. E* 54, 3419–3435. doi: 10.1103/PhysRevE.54.3419
- Flavahan, W. A., Drier, Y., Johnstone, S. E., Hemming, M. L., Tarjan, D. R., Hegazi, E., et al. (2019). Altered chromosomal topology drives oncogenic programs in SDH-deficient gists. *Nature* 575, 229–233. doi: 10.1038/s41586-019-1668-3
- Folguera-Blasco, N., Pérez-Carrasco, R., Cuyàs, E., Menendez, J. A., and Alarcón, T. (2019). A multiscale model of epigenetic heterogeneity-driven cell fate decision-making. *PLoS Comput. Biol.* 15:e1006592. doi: 10.1371/journal.pcbi.1006592
- Fu, H.-C., and Sasaki, R. (1997). Negative binomial and multinomial states: probability distributions and coherent states. *J. Math. Phys.* 38, 3968–3987. doi: 10.1063/1.532102
- Fu, H.-C., and Sasaki, R. (1998). Probability distributions and coherent states of, and algebras. *J. Phys. A* 31, 901–925. doi: 10.1088/0305-4470/31/3/006
- Furey, T. S., and Sethupathy, P. (2013). Genetics driving epigenetics. *Science* 342, 705–706. doi: 10.1126/science.1246755
- Gaspar-Maia, A., Alajem, A., Meshorer, E., and Ramalho-Santos, M. (2011). Open chromatin in pluripotency and reprogramming. *Nat. Rev. Mol. Cell Biol.* 12, 36–47. doi: 10.1038/nrm3036
- Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* 81, 2340–2361. doi: 10.1021/j100540a008
- Goldberg, A. D., Allis, C. D., and Bernstein, E. (2007). Epigenetics: a landscape takes shape. *Cell* 128, 635–638. doi: 10.1016/j.cell.2007.02.006
- Hasty, J., Pradines, J., Dolnik, M., and Collins, J. J. (2000). Noise-based switches and amplifiers for gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 97, 2075–2080. doi: 10.1073/pnas.040411297
- Hornos, J. E. M., Schultz, D., Innocentini, G. C. P., Wang, J., Walczak, A. M., Onuchic, J. N., et al. (2005). Self-regulating gene: An exact solution. *Phys. Rev. E* 72:051907. doi: 10.1103/PhysRevE.72.051907
- Huang, R., and Lei, J. (2018). Dynamics of gene expression with positive feedback to histone modifications at bivalent domains. *Int. J. Modern Phys. B* 32:1850075. doi: 10.1142/S0217979218500753
- Johnson, A. D., Poteete, A. R., Lauer, G., Sauer, R. T., Ackers, G. K., and Ptashne, M. (1981). λ repressor and cro-components of an efficient molecular switch. *Nature* 294, 217–223. doi: 10.1038/294217a0
- Jost, D. (2014). Bifurcation in epigenetics: Implications in development, proliferation, and diseases. *Phys. Rev. E* 89:010701. doi: 10.1103/PhysRevE.89.010701
- Kærn, M., Elston, T. C., Blake, W. J., and Collins, J. J. (2005). Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.* 6, 451–464. doi: 10.1038/nrg1615
- Kar, G., Kim, J. K., Kolodziejczyk, A. A., Natarajan, K. N., Triglia, E. T., Mifsud, B., et al. (2017). Flipping between Polycomb repressed and active transcriptional states introduces noise in gene expression. *Nat. Commun.* 8, 1–13. doi: 10.1038/s41467-017-00052-2
- Kim, K.-Y., and Wang, J. (2007). Potential energy landscape and robustness of a gene regulatory network: toggle switch. *PLoS Comput. Biol.* 3:e30060. doi: 10.1371/journal.pcbi.0030060
- Krishnakumar, R., and Kraus, W. L. (2010). Parp-1 regulates chromatin structure and transcription through a kdm5b-dependent pathway. *Mol. Cell* 39, 736–749. doi: 10.1016/j.molcel.2010.08.014
- Lee, B. P., and Cardy, J. (1995). Renormalization group study of the $a+b \rightarrow 0$ diffusion-limited reaction. *J. Stat. Phys.* 80, 971–1007. doi: 10.1007/BF02179861
- Lim, P. S. L., and Meshorer, E. (2020). Dppa2 and dppa4 safeguard bivalent chromatin in order to establish a pluripotent epigenome. *Nat. Struct. Mol. Biol.* 27, 685–686. doi: 10.1038/s41594-020-0453-1
- Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462, 315–322. doi: 10.1038/nature08514
- Loh, Y.-H., Ng, J.-H., and Ng, H.-H. (2008). Molecular framework underlying pluripotency. *Cell Cycle* 7, 885–891. doi: 10.4161/cc.7.7.5636
- Lu, R., Markowitz, F., Unwin, R. D., Leek, J. T., Airolidi, E. M., MacArthur, B. D., et al. (2009). Systems-level dynamic analyses of fate change in murine embryonic stem cells. *Nature* 462, 358–362. doi: 10.1038/nature08575
- Lv, C., Li, X., Li, F., and Li, T. (2015). Energy landscape reveals that the budding yeast cell cycle is a robust and adaptive multi-stage process. *PLoS Comput. Biol.* 11:e1004156. doi: 10.1371/journal.pcbi.1004156
- Margueron, R., and Reinberg, D. (2010). Chromatin structure and the inheritance of epigenetic information. *Nat. Rev. Genet.* 11, 285–296. doi: 10.1038/nrg2752
- Mariani, L., Schulz, E. G., Lexberg, M. H., Helmstetter, C., Radbruch, A., Löhning, M., et al. (2010). Short-term memory in gene induction reveals the regulatory principle behind stochastic il-4 expression. *Mol. Syst. Biol.* 6:359. doi: 10.1038/msb.2010.13
- Mas, G., Blanco, E., Ballaré, C., Sansó, M., Spill, Y. G., Hu, D., et al. (2018). Promoter bivalency favors an open chromatin architecture in embryonic stem cells. *Nat. Genet.* 50, 1452–1462. doi: 10.1038/s41588-018-0218-5
- Matsushita, Y., and Kaneko, K. (2020). Homeorhesis in waddington's landscape by epigenetic feedback regulation. *Phys. Rev. Res.* 2:023083. doi: 10.1103/PhysRevResearch.2.023083
- Micheelsen, M. A., Mitarai, N., Sneppen, K., and Dodd, I. B. (2010). Theory for the stability and regulation of epigenetic landscapes. *Phys. Biol.* 7:026010. doi: 10.1088/1478-3975/7/2/026010
- Miller-Jensen, K., Dey, S. S., Schaffer, D. V., and Arkin, A. P. (2011). Varying virulence: epigenetic control of expression noise and disease processes. *Trends Biotechnol.* 29, 517–525. doi: 10.1016/j.tibtech.2011.05.004
- Ohkubo, J. (2008). Approximation scheme for master equations: variational approach to multivariate case. *J. Chem. Phys.* 129:044108. doi: 10.1063/1.2957462
- Orkin, S. H., and Zon, L. I. (2008). Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* 132, 631–644. doi: 10.1016/j.cell.2008.01.025

- Parsons, T., and Zhang, B. (2019). Critical role of histone tail entropy in nucleosome unwinding. *J. Chem. Phys.* 150:185103. doi: 10.1063/1.5085663
- Ptashne, M., Jeffrey, A., Johnson, A., Maurer, R., Meyer, B., Pabo, C., et al. (1980). How the λ repressor and CRO work. *Cell* 19, 1–11. doi: 10.1016/0092-8674(80)90383-9
- Qi, Y., Reyes, A., Johnstone, S. E., Aryee, M. J., Bernstein, B. E., and Zhang, B. (2020). Data-driven polymer model for mechanistic exploration of diploid genome organization. *Biophys. J.* 119, 1905–1916. doi: 10.1016/j.bpj.2020.09.009
- Qiu, B., Zhou, T., and Zhang, J. (2020). Molecular-memory-driven phenotypic switching in a genetic toggle switch without cooperative binding. *Phys. Rev. E* 101:022409. doi: 10.1103/PhysRevE.101.022409
- Ralston, A., and Rossant, J. (2005). Genetic regulation of stem cell origins in the mouse embryo. *Clin. Genet.* 68, 106–112. doi: 10.1111/j.1399-0004.2005.00478.x
- Ramos, A. F., Innocentini, G. C. P., and Hornos, J. E. M. (2011). Exact time-dependent solutions for a self-regulating gene. *Phys. Rev. E* 83:062902. doi: 10.1103/PhysRevE.83.062902
- Rosenfeld, N., Elowitz, M. B., and Alon, U. (2002). Negative autoregulation speeds the response times of transcription networks. *J. Mol. Biol.* 323, 785–793. doi: 10.1016/S0022-2836(02)00994-4
- Rowley, M. J., and Corces, V. G. (2018). Organizational principles of 3D genome architecture. *Nat. Rev. Genet.* 19, 789–800. doi: 10.1038/s41576-018-0060-8
- Sasai, M., Kawabata, Y., Makishi, K., Itoh, K., and Terada, T. P. (2013). Time scales in epigenetic dynamics and phenotypic heterogeneity of embryonic stem cells. *PLoS Comput. Biol.* 9:e1003380. doi: 10.1371/journal.pcbi.1003380
- Sasai, M., and Wolynes, P. G. (2003). Stochastic gene expression as a many-body problem. *Proc. Natl. Acad. Sci. U.S.A.* 100, 2374–2379. doi: 10.1073/pnas.2627987100
- Schlick, T., Hayes, J., and Grigoryev, S. (2012). Toward convergence of experimental studies and theoretical modeling of the chromatin fiber. *J. Biol. Chem.* 287, 5183–5191. doi: 10.1074/jbc.R111.305763
- Sedighi, M., and Sengupta, A. M. (2007). Epigenetic chromatin silencing: bistability and front propagation. *Phys. Biol.* 4, 246–255. doi: 10.1088/1478-3975/4/4/002
- Shahrezaei, V., and Swain, P. S. (2008). Analytical distributions for stochastic gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 105, 17256–17261. doi: 10.1073/pnas.0803850105
- Sneppen, K., and Mitarai, N. (2012). Multistability with a metastable mixed state. *Phys. Rev. Lett.* 109:100602. doi: 10.1103/PhysRevLett.109.100602
- Sneppen, K., and Ringrose, L. (2019). Theoretical analysis of polycomb-trithorax systems predicts that poised chromatin is bistable and not bivalent. *Nat. Commun.* 10:2133. doi: 10.1038/s41467-019-10130-2
- Sood, A., and Zhang, B. (2020). Quantifying epigenetic stability with minimum action paths. *Phys. Rev. E* 101:062409. doi: 10.1103/PhysRevE.101.062409
- Tate, P. H., and Bird, A. P. (1993). Effects of DNA methylation on DNA-binding proteins and gene expression. *Curr. Opin. Genet. Dev.* 3, 226–231. doi: 10.1016/0959-437X(93)90027-M
- Tauber, U. C. (2014). *Critical Dynamics A Field Theory Approach to Equilibrium and Non-Equilibrium Scaling Behavior*. Cambridge: Cambridge University Press.
- Vastenhouw, N. L., and Schier, A. F. (2012). Bivalent histone modifications in early embryogenesis. *Curr. Opin. Cell Biol.* 24, 374–386. doi: 10.1016/j.ceb.2012.03.009
- Venegas-Ortiz, J. and Evans, M. R. (2011). Analytical study of an exclusive genetic switch. *J. Phys. A* 44:355001. doi: 10.1088/1751-8113/44/35/355001
- Waddington, C., and Kacser, H. (1957). *The Strategy of the Genes: A Discussion of Some Aspects of Theoretical Biology*. London: Allen & Unwin.
- Walczak, A. M., Onuchic, J. N., and Wolynes, P. G. (2005a). Absolute rate theories of epigenetic stability. *Proc. Natl. Acad. Sci. U.S.A.* 102, 18926–18931. doi: 10.1073/pnas.0509547102
- Walczak, A. M., Sasai, M., and Wolynes, P. G. (2005b). Self-consistent proteomic field theory of stochastic gene switches. *Biophys. J.* 88, 828–850. doi: 10.1529/biophysj.104.050666
- Wang, J., Zhang, K., Xu, L., and Wang, E. (2011). Quantifying the Waddington landscape and biological paths for development and differentiation. *Proc. Natl. Acad. Sci. U.S.A.* 108, 8257–8262. doi: 10.1073/pnas.1017017108
- Wang, P., Song, C., Zhang, H., Wu, Z., Tian, X.-J., and Xing, J. (2014). Epigenetic state network approach for describing cell phenotypic transitions. *Interface Focus* 4:20130068. doi: 10.1098/rsfs.2013.0068
- Xie, W. J., Qi, Y., and Zhang, B. (2020). Characterizing chromatin folding coordinate and landscape with deep learning. *PLoS Comput. Biol.* 16:e1008262. doi: 10.1371/journal.pcbi.1008262
- Xie, W. J., and Zhang, B. (2019). Learning the formation mechanism of domain-level chromatin states with epigenomics data. *Biophys. J.* 116, 2047–2056. doi: 10.1016/j.bpj.2019.04.006
- Xu, B.-L. and Tao, Y. (2006). External noise and feedback regulation: Steady-state statistics of auto-regulatory genetic network. *J. Theor. Biol.* 243, 214–221. doi: 10.1016/j.jtbi.2006.06.003
- Zhang, B., and Wolynes, P. G. (2014). Stem cell differentiation as a many-body problem. *Proc. Natl. Acad. Sci. U.S.A.* 111, 10185–10190. doi: 10.1073/pnas.1408561111
- Zhang, K., Sasai, M., and Wang, J. (2013). Eddy current and coupled landscapes for nonadiabatic and nonequilibrium complex system dynamics. *Proc. Natl. Acad. Sci. U.S.A.* 110, 14930–14935. doi: 10.1073/pnas.1305604110
- Zhang, Y., Liu, N., Lin, W., and Li, C. (2019). Quantifying the interplay between genetic and epigenetic regulations in stem cell development. *New J. Phys.* 21:103042. doi: 10.1088/1367-2630/ab4c82
- Zhou, V. W., Goren, A., and Bernstein, B. E. (2011). Charting histone modifications and the functional organization of mammalian genomes. *Nat. Rev. Genet.* 12, 7–18. doi: 10.1038/nrg2905
- Zhu, H., Wang, G., and Qian, J. (2016). Transcription factors as readers and effectors of dna methylation. *Nature Reviews Genetics*. 17, 551–565. doi: 10.1038/nrg.2016.83

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Sood and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Bistable Mathematical Model of Neutrophil Migratory Patterns After LPS-Induced Epigenetic Reprogramming

Stanca M. Ciupe^{1*}, Brittany P. Boribong², Sarah Kadelka³ and Caroline N. Jones^{4*}

¹ Department of Mathematics, Virginia Tech, Blacksburg, VA, United States, ² Division of Pediatric Pulmonology, Massachusetts General Hospital, Boston, MA, United States, ³ Department of Environmental Systems Science, ETH Zurich, Zurich, Switzerland, ⁴ Department of Bioengineering, University of Texas, Dallas, TX, United States

OPEN ACCESS

Edited by:

Tian Hong,
The University of Tennessee, Knoxville,
United States

Reviewed by:

Stephanie Evans,
University of Michigan, United States
Dongheon Lee,
Duke University, United States

*Correspondence:

Stanca M. Ciupe
stanca@vt.edu
Caroline N. Jones
caroline.jones@utdallas.edu

Specialty section:

This article was submitted to
Systems Biology,
a section of the journal
Frontiers in Genetics

Received: 26 November 2020

Accepted: 27 January 2021

Published: 23 February 2021

Citation:

Ciupe SM, Boribong BP, Kadelka S
and Jones CN (2021) Bistable
Mathematical Model of Neutrophil
Migratory Patterns After LPS-Induced
Epigenetic Reprogramming.
Front. Genet. 12:633963.
doi: 10.3389/fgene.2021.633963

The highly controlled migration of neutrophils toward the site of an infection can be altered when they are trained with lipopolysaccharides (LPS), with high dose LPS enhancing neutrophil migratory pattern toward the bacterial derived source signal and super-low dose LPS inducing either migration toward an intermediary signal or dysregulation and oscillatory movement. Empirical studies that use microfluidic chemotaxis-chip devices with two opposing chemoattractants showed differential neutrophil migration after challenge with different LPS doses. The epigenetic alterations responsible for changes in neutrophil migratory behavior are unknown. We developed two mathematical models that evaluate the mechanistic interactions responsible for neutrophil migratory decision-making when exposed to competing chemoattractants and challenged with LPS. The first model, which considers the interactions between the receptor densities of two competing chemoattractants, their kinases, and LPS, displayed bistability between high and low ratios of primary to intermediary chemoattractant receptor densities. In particular, at equilibrium, we observe equal receptor densities for low LPS ($< 15\text{ng/mL}$); and dominance of receptors for the primary chemoattractant for high LPS ($> 15\text{ng/mL}$). The second model, which included additional interactions with an extracellular signal-regulated kinase in both phosphorylated and non-phosphorylated forms, has an additional dynamic outcome, oscillatory dynamics for both receptors, as seen in the data. In particular, it found equal receptor densities in the absence of oscillation for super-low and high LPS challenge (< 0.4 and $1.1 < \text{LPS} < 375 \text{ ng/mL}$); equal receptor densities with oscillatory receptor dynamics for super-low LPS ($0.5 < \text{LPS} < 1.1\text{ng/mL}$); and dominance of receptors for the primary chemoattractant for super-high LPS ($> 376 \text{ ng/mL}$). Predicting the mechanisms and the type of external LPS challenge responsible for neutrophils migration toward pro-inflammatory chemoattractants, migration toward pro-tolerant chemoattractants, or oscillatory movement is necessary knowledge in designing interventions against immune diseases, such as sepsis.

Keywords: neutrophil migration, mathematical model, lipopolysaccharide (LPS), bistability, cellular decision-making

1. INTRODUCTION

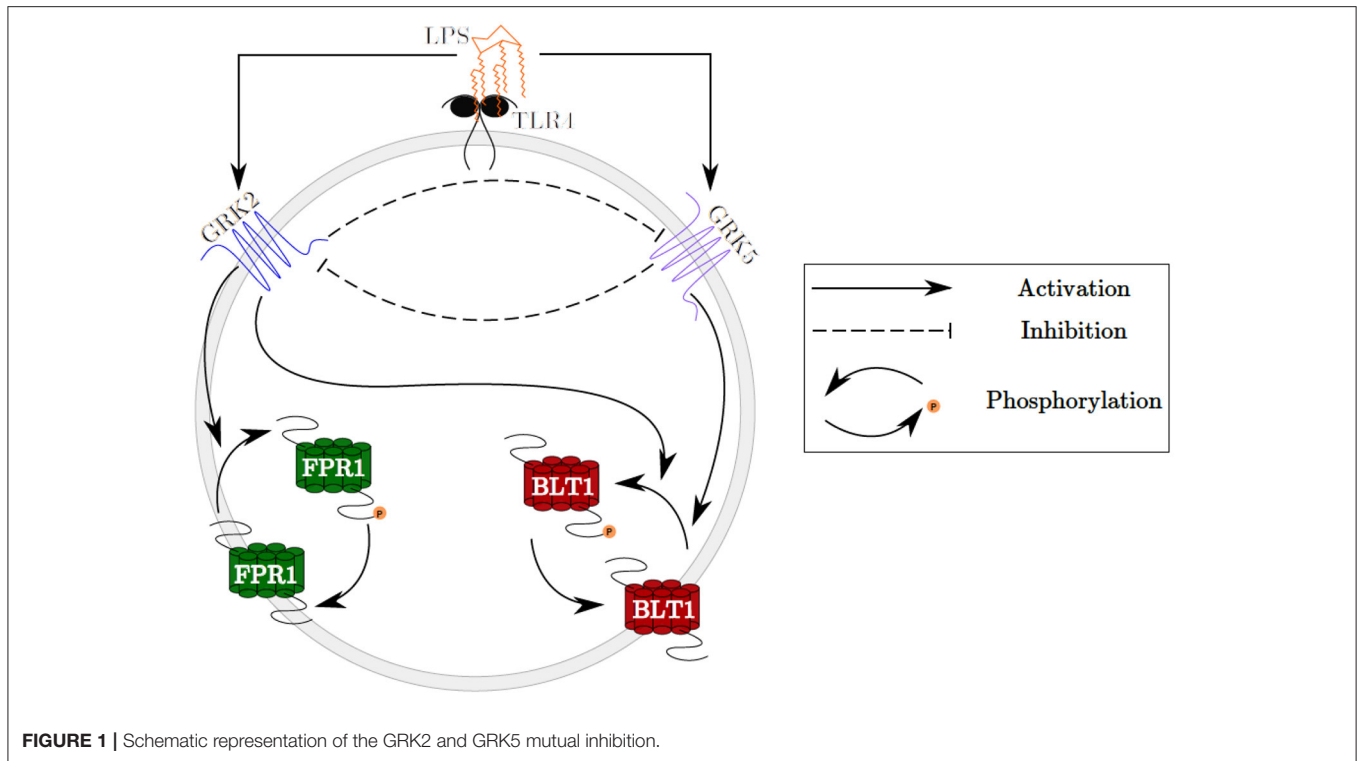
Researchers have recently challenged the dogma that innate immunity is the same at every challenge. It has been shown that macrophages are able to develop different kinds of memory depending on the type of priming they encounter via epigenetic reprogramming (Yuan et al., 2016a,b). For instance, they can develop a memory phenotype that leads them to be less reactive or even tolerant to a challenge, or they can develop a memory phenotype that leads them to have an enhanced response to a challenge. This same concept has recently been shown by us for neutrophil migratory decision-making, and it is thought that the response is influencing the outcomes of infectious diseases. For example, in sepsis or COVID-19 infection, the immune system overreacts because of underlying low-grade inflammation that primes neutrophils into choosing between tolerant and inflammatory migratory phenotypes (Alves-Filho et al., 2005, 2010). As a result, neutrophils can migrate to healthy organs and unleash their anti-microbial arsenal in healthy tissue, leading to organ failure in the lungs, kidney, or heart. The mechanisms underlying trained innate immunity have not been fully elucidated, with epigenetic modifications playing a key role in the induction of *innate memory or training* (Pillay et al., 2010; Demaret et al., 2015). In this study we investigate innate memory in the context of neutrophil migratory decision-making.

The ability of neutrophils to migrate plays a pivotal role in a cell's ability to clear infections and resolve inflammation. During infection and inflammation, chemoattractants are released, signaling and activating neutrophils in the bloodstream. Neutrophils must be able to precisely migrate within the tissue to the specific site of infection, without being diverted toward other locations, in a process called chemotaxis. Chemotaxis is a highly regulated process that involves activation of various pathways and downstream polarization of the cell (Kolaczowska and Kubes, 2013). The first step in chemotaxis is recognition of chemoattractants by the cell. Cells have specific receptors on their surface for various chemoattractants. These chemoattractant receptors are G protein-coupled receptors (GPCRs), which are regulated by a variety of G protein-coupled receptor kinases (GRKs) (Murphy, 1994; Dianqing, 2005). When bound by a specific agonist, in this case a chemoattractant, the GPCRs undergo phosphorylation, which unbinds the G proteins and desensitizes the receptor. This leads to internalization of the receptor, activation of downstream signaling pathways, and activation of cellular responses, such as cell polarization and chemotaxis (Murphy, 1994; Dianqing, 2005; Futosi et al., 2013). After internalization, receptors can be recycled back to the cell surface, where they can again be bound by the receptor's agonist. This process is crucial in chemotaxis, as it allows the cell to continue sensing the chemoattractant and migrate in its direction (Neel et al., 2005). Most chemoattractant receptors are similar in their response to ligand-binding; however there are slight differences in the activated signaling pathways (Heit et al., 2002, 2008). Within the tissue, neutrophils are exposed to several chemoattractants at once, originating from pathogens, cells within the tissue, the endothelium, and several other sources (Kolaczowska and Kubes, 2013). Cells must

prioritize these signals to properly clear the pathogen. It has been hypothesized that neutrophils have an internal hierarchy, where chemoattractants derived from bacterial sources and the complement system, such as fMLP and C5a (Heit et al., 2002; Petri and Sanz, 2018), take precedent over intermediary chemoattractants, such as LTB₄ and IL-8, which are secreted by other immune cells. This leads to neutrophils migrating toward end-target chemoattractants over intermediary chemottractants in a competitive environment (Heit et al., 2002, 2008; Wang et al., 2016b), allowing neutrophils to prioritize an invading pathogen. This hierarchy is thought to occur through the activation of differing signaling pathways, where end-target chemoattractants signal through p38 MAPK and intermediary chemoattractants signal through PI3K (Heit et al., 2002, 2008).

The highly controlled migration of neutrophils toward the site of an infection, as well as their dynamic interaction with pathogens, can be altered when they are pre-conditioned with Lipopolysaccharides (LPS) to induce endotoxin priming. In previous work, we showed that training with high dose LPS (100 ng/mL) enhances neutrophil migration toward the end-target, bacterial derived, source signal fMLP. By contrast, training with super-low dose LPS (1 ng/mL) alters neutrophil migratory phenotypes, which either migrate toward the intermediary signal LTB₄ or become dysregulated and exhibit oscillatory migratory patterns (Jones et al., 2016; Boribong et al., 2019). While the empirical data shows that neutrophils trained with LPS change migratory phenotype, it does not give information on the molecular mechanisms responsible for the difference in behavior. The migratory decision-making process is finely governed by complex signaling networks that dynamically receive and interpret molecular and cellular signals from outside and within. The intrinsic complexity of immune cell decision-making processes has created difficulty for experimental immunologists to determine the mechanisms of disease, in spite of expansive experimental studies with conventional reductionist cellular and molecular approaches. It is increasingly recognized that cross-disciplinary studies combining experimental and mathematical modeling approaches are critically required.

In this study, we investigate the molecular mechanisms of neutrophil migratory decision-making in the presence of competing chemoattractants and external challenge with LPS, by building deterministic mathematical models of interaction between two chemoattractant receptors, Formyl Peptide Receptor 1 (FPR1) and Leukotriene B4 Receptor 1 (BLT1), and key molecules involved in their regulation. We are interested in determining the relationship between the receptor dynamics and migration pattern, and in quantifying the LPS dose resulting in neutrophils migration toward a pro-inflammatory chemoattractant, toward a pro-resolution chemoattractant, or in neutrophils dysregulation and oscillation (Fan and Malik, 2003; Liu et al., 2012; Byrne et al., 2014). The model will qualitatively match the experimental results of our previous work, where stimulation with a super-low concentration of LPS will result in greater BLT1 over FPR1, and stimulation with a high concentration of LPS will result in greater FPR1 over BLT1 (Boribong et al., 2019). We construct a model with bistable behavior, with the motif for bistability coming from the

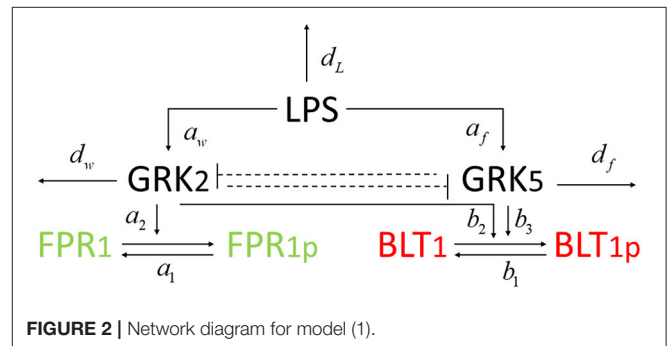


non-linear mutual inhibition of GRK2 and GRK5 (see **Figure 1**). The dual inhibition leads to the activation of different signaling pathways (p38/JNK vs. ERK), leading to differences in functional neutrophil migration (Davenport et al., 2020). Both GRK2 and GRK5 have been demonstrated to be critical mediators of the molecular alterations that occur in the inflammatory disorders, but the complex mutual inhibition interaction has largely been ignored (Philipp et al., 2014). Mathematical models have been used before to model cellular decision-making (Day et al., 2006; Kadelka et al., 2019), neutrophil chemotaxis (Ionides et al., 2004; Postma and van Haastert, 2016; Bayani et al., 2020), immune responses (Reynolds et al., 2006; Fischer, 2008; Nelson et al., 2009; Vodovotz et al., 2009) and bistable dynamics (Ciupe et al., 2007, 2018; Leber et al., 2016).

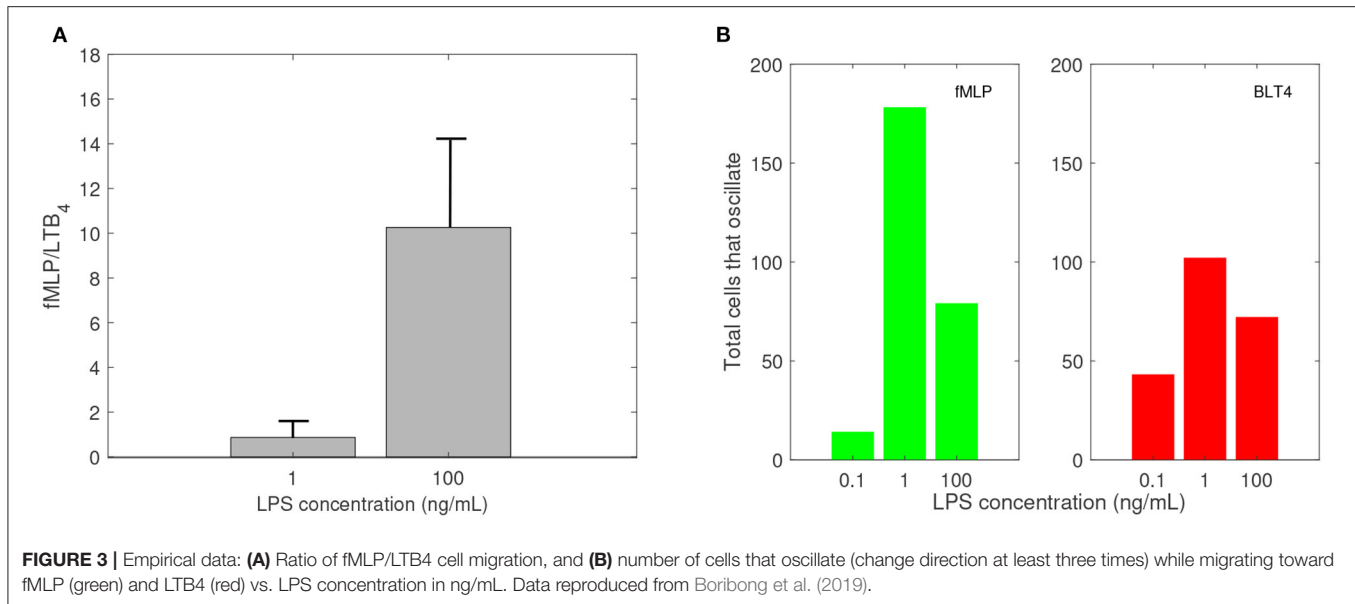
2. METHODS

2.1. Mathematical Model of Migratory Decision-Making

We developed a novel system of differential equations based on diagram in **Figure 2**, which describes the interactions between [LPS], kinases [GRK2] and [GRK5], the receptor for end-target chemoattractant fMLP, [FPR1], and the receptor for intermediary chemoattractant LTB₄, [BLT1]. Priming by LPS occurs through activation of both GRK2 and GRK5 (Prossnitz et al., 1995; Arraes et al., 2006; Sorriento et al., 2008; Wang et al., 2016a). For simplicity, we model linear effects of LPS on the kinases' activity. In particular, we assume that the GRK2 activation occurs at rate $c_w + a_w[LPS]$, with c_w and a_w being the



LPS-independent and LPS-dependent activation rates. Similarly, GRK5 activation occurs at rate $c_f + a_f[LPS]$, with c_f and a_f being the LPS-independent and LPS-dependent activation rates. The two kinases mutually inhibit one another. We model inhibition of GRK2 via GRK5 at rate $1/(b_{fw} + [GRK5]^n)$ and inhibition of GRK5 via GRK2 at rate $1/(b_{wf} + [GRK2])$, where b_{fw} and b_{wf} are the mutual inhibition rates of GRK2 by GRK5 and GRK5 by GRK1, respectively. n is the cooperativity coefficient. We assumed increased cooperativity in GRK2 inhibition by GRK5, but not the inhibition of GRK5 by GRK2. The results are preserved if the same cooperativity is included in the GRK5 inhibition by GRK2 (not shown). We assume GRK2 and GRK5 decay at per capita rates d_w and d_f , respectively, with GRK5 decay being modeled in a density dependent manner, with the



GRK5 value where the decay is half-maximal being given by parameter b_f .

The chemoattractant receptors FPR1 and BLT1 internalize from the plasma membrane into the cell via phosphorylation (Magalhaes et al., 2012; Mócsai et al., 2015). We assume that the number of receptors on a cell is conserved and, through the process of dephosphorylation, the receptors are recycled and brought back to the surface of the cell. Thus, we have conservation laws of the total number of the receptor equalling the sum of the non-phosphorylated and phosphorylated receptor, $[FPR1]_{total} = [FPR1] + [FPR1_p]$ and $[BLT1]_{total} = [BLT1] + [BLT1_p]$. The process of receptor phosphorylation and dephosphorylation is modeled using Hill-type functions. In particular, FPR1 is produced through dephosphorylation, modeled by a Michaelis-Menten term $a_1([FPR1]_{total} - [FPR1])/(J_{F1} + [FPR1]_{total} - [FPR1])$, where a_1 is maximal production and J_{F1} is the receptor quantity where dephosphorylation is half-maximal. Similarly, FPR1 is lost through phosphorylation, which is enhanced in the presence of GRK2 (Wang et al., 2016a). We model this by a Hill-type function $a_2[FPR1][GRK2]/(J_{F2} + [FPR1])$, where a_2 is the maximal rate and J_{F2} is the receptor quantity where phosphorylation is half-maximal.

BLT1 is produced through dephosphorylation, modeled by a Michaelis-Menten term $b_1([BLT1]_{total} - [BLT1])/(J_{B1} + [BLT1]_{total} - [BLT1])$, where b_1 is the maximal production rate and J_{B1} is the receptor quantity where dephosphorylation is half-maximal. BLT1 is lost through phosphorylation, which is enhanced in the presence of both GRK2 and GRK5 (Gaudreau et al., 2002; Chen et al., 2004). We model this by a Hill-type function $[BLT1](b_2[GRK2] + b_3[GRK5])/(J_{B2} + [BLT1])$, where b_2 and b_3 are maximal decay rates and J_{B2} is the receptor quantity where phosphorylation is half-maximal. We assume a single LPS dose, after which LPS decays exponentially at a rate d_L (Kadelka et al., 2019). The dynamical system describing these interactions

is given by:

$$\begin{aligned}
 \frac{d[GRK2]}{dt} &= \frac{c_w + a_w[LPS]}{b_{fw} + [GRK5]^n} - d_w[GRK2], \\
 \frac{d[GRK5]}{dt} &= \frac{c_f + a_f[LPS]}{b_{wf} + [GRK2]} - d_f \frac{[GRK5]}{b_f + [GRK5]}, \\
 \frac{d[FPR1]}{dt} &= \frac{a_1([FPR1]_{total} - [FPR1])}{J_{F1} + [FPR1]_{total} - [FPR1]} \\
 &\quad - a_2[GRK2] \frac{[FPR1]}{J_{F2} + [FPR1]}, \\
 \frac{d[BLT1]}{dt} &= \frac{b_1([BLT1]_{total} - [BLT1])}{J_{B1} + [BLT1]_{total} - [BLT1]} - (b_2[GRK2] \\
 &\quad + b_3[GRK5]) \frac{[BLT1]}{J_{B2} + [BLT1]}, \\
 \frac{d[LPS]}{dt} &= -d_L[LPS].
 \end{aligned} \tag{1}$$

We are interested in determining the ratio between the cells that migrate toward the primary and those that migrate toward the intermediary chemoattractants given, as a proxy, by the ratio of their receptors FPR1/BLT1, when initial LPS is varied.

2.1.1. Experimental Data

In previous research, we used a microfluidic competitive chemotaxis-chip device to measure the migratory decision-making process of dHL-60 cells, a model neutrophil cell line, 5 h after they were pre-challenged with super-low-dose (1 ng/mL) and high-dose (100 ng/mL) of LPS in the presence of two competing chemoattractants, LTB4 and fMLP (Boribong et al., 2019). Challenging the cells with a super-low dose of LPS resulted in fMLP/LTB4 ratio of 0.8672. Challenging the cells with a high dose of LPS

TABLE 1 | Parameters and initial conditions used in model 1.

Parameter	Description	Value	References
c_w	[LPS]-independent [GRK2] activation	15	
a_w	[LPS]-dependent [GRK2] activation	1	
d_w	[GRK2] degradation	0.69	Penela et al., 1998; Luo and Benovic, 2003
b_{fw}	[GRK2] inhibition by [GRK5]	5×10^{-4}	
c_f	[LPS]-independent [GRK5] activation	1	
a_f	[LPS]-dependent [GRK5] activation	1	
d_f	[GRK5] degradation	0.33	Wu et al., 2012
b_{wf}	[GRK5] inhibition by [GRK2]	0.13	
b_f	[GRK5] where degradation is half-maximal	1	
n	Hill coefficient	3	
$a_i, i = \{1, 2\}$	phosphorylation and dephosphorylation rates	166	Leoni et al., 2015
$b_i, i = \{1, 2, 3\}$	phosphorylation and dephosphorylation rates	20	
$J_i, i = \{F1, F2, B1, B2\}$	values where phosphorylation is half-maximal	1	Gaudreault et al., 2005
$[FRP1]_{total}$	Total [FRP1] receptors	40,000	Schneider et al., 2012
$[BLT1]_{total}$	Total [BLT1] receptors	13,333	Schneider et al., 2012
d_L	[LPS] loss	0.1	Kadelka et al., 2019
Initial conditions	Description	Value	References
$[GRK2](0)$	Initial [GRK2] value	0.75	Arraes et al., 2006
$[GRK5](0)$	Initial [GRK5] value	0.5	Arraes et al., 2006
$[FPR1](0)$	Initial [FPR1] value	40,000	Schneider et al., 2012
$[BLT1](0)$	Initial [BLT1] value	13,333	Schneider et al., 2012
$[LPS](0)$	Initial [LPS] value	Varied	Boribong et al., 2019

(100 ng/mL) resulted in fMLP/LTB4 ratio of 10.2646 (see **Figure 3A**).

2.1.2. Parameter Values

There are approximately 40,000 FPR1 and 13,333 BLT1 receptors on each neutrophil (Schneider et al., 2012). We therefore set initial conditions to $[FPR1](0) = 40,000$ and $[BLT1](0) = 13,333$. The reported GRK2/GRK5 ratio is 1.5 (Arraes et al., 2006). We choose initial conditions $[GRK2](0) = 0.75$ and $[GRK5](0) = 0.5$, to preserve this ratio. The reported GRK2 half-life varies between 60 min in HEK, COS-7, Jurkat, C6 glioma cells (Penela et al., 1998) and 20–24 h in undifferentiated HL-60 cells (Luo and Benovic, 2003). We choose a shorter half-life of 1 h, which corresponds to the GRK decay rate $d_w = \log(2)/1 = 0.69$ per hour. The reported GRK5 life-span is 3 h (Wu et al., 2012), which corresponds to the GRK5 decay rate $d_f = 1/3 = 0.33$ per hour. The FPR1's phosphorylation half-life is 15 s (Leoni et al., 2015). We choose both the phosphorylation and dephosphorylation rates based on this value, $a_1 = a_2 = \log(2) \times 3,600/15 = 166$ per hour. BLT1 phosphorylation's half-life is 120 s (Gaudreault et al., 2005). We choose both the phosphorylation and dephosphorylation rates based on this value, $b_1 = b_2 = b_3 = \log(2) \times 3,600/120 = 20$ per hour. As in our previous work (Kadelka et al., 2019), the LPS degradation rate is $d_L = 0.1$ per day. For simplicity, we fix most unknown parameters at one, $a_w = c_f = a_f = b_f = J_{F1} = J_{F2} = J_{B1} = J_{B2} = 1$. Moreover, $c_w = 15$, $b_{fw} = 5 \times 10^{-4}$, $b_{wf} = 0.13$ and $n = 3$. The parameter values are summarized in **Table 1**.

2.2. Mathematical Model of Oscillatory Movement

We coupled system (1) with an oscillator describing the dynamics of non-phosphorylated and phosphorylated extracellular signal-regulated kinases (ERK), $[ERK]$ and $[ERK_p]$, that are participating in an autocatalytic reaction with the help of intermediate non-phosphorylated and phosphorylated enzymes, $[E]$ and $[E_p]$ (see **Figure 7**). We assume that $[ERK]$ activation is LPS-dependent and occurs at rate $k_1 [LPS]$. The phosphorylated $[ERK_p]$ decays at rate k_2 . The phosphorylated enzyme $[E_p]$ follows the following reaction:

$$\frac{d[E_p]}{dt} = k_{E1}[ERK_p] \frac{[E]}{J_{E1} + [E]} - \frac{k_{E2}[E_p]}{J_{E2} + [E_p]}, \quad (2)$$

where k_{Ei} are the dephosphorylation and phosphorylation rates and J_{Ei} are the phosphorylation and dephosphorylation half-maximal rates, $i = \{1, 2\}$. If we assume chemical equilibrium, $[E_p] + [E] = 1$, and $k_3 = k_{E2}/k_{E1}$, we obtain that:

$$\begin{aligned} [E_p] &= \frac{X - \sqrt{X^2 - 4(k_3 - [ERK_p])[ERK_p]J_{E2}}}{2(k_3 - [ERK_p])} \\ &= \frac{2J_{E2}[ERK_p]}{X + \sqrt{X^2 - 4(k_3 - [ERK_p])[ERK_p]J_{E2}}} \\ &= G([ERK_p], k_3, J_{E1}, J_{E2}) \end{aligned} \quad (3)$$

where $X = k_3 - [ERK_p] + k_3 J_{E1} + J_{E2} [ERK_p]$ and $G([ERK_p], k_3, J_{E1}, J_{E2})$ is the Goldbeter-Koshland function (Goldbeter and Koshland, 1981). Hence, the phosphorylation of $[ERK]$ occurs at rate $(k_{0s} + k_0 G([ERK_p], k_3, J_{E1}, J_{E2})) [ERK]$. Lastly, the LPS is constant at all times $LPS = [LPS](0)$, to account for positive long-term $[ERK]$ levels. The model is given by the system:

$$\begin{aligned} \frac{d[ERK]}{dt} &= k_1 [LPS] - (k_{0s} + k_0 G([ERK_p], k_3, J_{E1}, J_{E2})) [ERK], \\ \frac{d[ERK_p]}{dt} &= (k_{0s} + k_0 G([ERK_p], k_3, J_{E1}, J_{E2})) [ERK] - k_2 [ERK_p], \\ \frac{d[GRK2]}{dt} &= \frac{c_w + a_w [ERK_p]}{b_{fw} + [GRK5]^n} - d_w [GRK2], \\ \frac{d[GRK5]}{dt} &= \frac{c_f + a_f [ERK_p]}{b_{wf} + [GRK2]} - d_f \frac{[GRK5]}{b_f + [GRK5]}, \\ \frac{d[FPR1]}{dt} &= \frac{a_1 ([FPR1]_{total} - [FPR1])}{J_{F1} + [FPR1]_{total} - [FPR1]} \\ &\quad - a_2 [GRK2] \frac{[FPR1]}{J_{F2} + [FPR1]}, \\ \frac{d[BLT1]}{dt} &= \frac{b_1 ([BLT1]_{total} - [BLT1])}{J_{B1} + [BLT1]_{total} - [BLT1]} - (b_2 [GRK2] \\ &\quad + b_3 [GRK5]) \frac{[BLT1]}{J_{B2} + [BLT1]}. \end{aligned} \quad (4)$$

2.2.1. Experimental Data

Experimental results reported that neutrophils treated overnight with LPS may lose their ability to move up the chemoattractant gradient, become disoriented, and display oscillatory behavior (Boribong et al., 2019). Moreover, the highest number of cells to display such oscillatory behavior occurs following LPS exposure with super-low dose (1 ng/mL) (see **Figure 3B**) (Boribong et al., 2019).

2.2.2. Parameter Values

We assume that initially $[ERK](0) = 5$ and $[ERK_p](0) = 0.1$. Kinase $[ERK]$ is produced at rate $k_1 = 0.3$ and phosphorylated at rate $k_{0s} = 0.01$. Kinase $[ERK_p]$ is lost at rate $k_2 = 1$. Enzyme $[E]$ is phosphorylated, in the presence of $[ERK]$, at rate $k_0 = 0.4$ and dephosphorylated at rate $k_3 = 0.3$. The processes are modeled using Michaelis-Menten terms, with densities where phosphorylation/ dephosphorylation are half-maximal being set to $J_{E1} = J_{E2} = 0.005$. All other parameters and initial conditions are as in model (1). The new parameter values are summarized in **Table 2**.

3. RESULTS

3.1. Bistable FPR1 and BLT1 Dynamics

We evaluated neutrophil migration between end-target chemoattractant fMLP and intermediary chemoattractant LTB₄ by developing model (1), which considers the interaction between the chemoattractants' receptors, $[FPR1]$ and $[BLT1]$, the receptors' kinases, $[GRK2]$ and $[GRK5]$, and $[LPS]$. We

TABLE 2 | Parameters and initial conditions used in model 2.

k_0	$[ERK]$ -dependent $[E]$ phosphorylation	0.4	
k_1	$[LPS]$ -dependent $[ERK]$ activation	0.3	
k_2	$[ERK_p]$ degradation rate	1	
k_3	$[ERK]$ -independent $[E]$ dephosphorylation	0.3	
k_{0s}	$[ERK]$ phosphorylation rate	0.01	
$J_i, i = [E1, E2]$	values where phosphorylation is half-maximal	0.005	
Initial conditions	Description	Value	References
$[ERK](0)$	Initial $[ERK]$ value	5	
$[ERK_p](0)$	Initial $[ERK_p]$ value	0.1	

quantified the $[FPR1]/[BLT1]$ ratio for different $[LPS]$ doses under the dynamics of system (1), and parameters/initial conditions given in **Table 1**. Since the experimental data has collected ratios of cell migration 5 h after LPS challenge, we first quantified $[FPR1]/[BLT1]$ at time $t = 5$.

Model (1) exhibits bistable behavior between high and low $[GRK2]$ concentrations (low and high $[GRK5]$ concentrations), with low $[LPS]$ priming leading to high $[GRK2]$ production and high $[LPS]$ priming leading to low $[GRK2]$ production (see **Figure 4B**). Five hours following challenge with super-low-dose (1 ng/mL) LPS, model (1) predicts the presence of a small number of receptors, which are distributed equally among FPR1 and BLT1, $[FPR1]/[BLT1](5) = 1$ (see **Figure 4A**, solid lines). Under our abstraction this means that, following challenge with 1 ng/mL LPS, an equal number of neutrophils migrated toward the fMLP and LTB₄ gradients. Conversely, 5 h following high-dose challenge (100 ng/mL) LPS, model (1) predicts the presence of a large number of receptors of both types, with $[FPR1]$ exceeding $[BLT1]$ by one fold, i.e., $[FPR1]/[BLT1](5) = 10$ (see **Figure 4A**, solid lines). Under our abstraction this means that a large number of neutrophils have migrated in both directions, with ten times more neutrophils migrating toward fMLP than LTB₄. These $[FPR1]/[BLT1]$ ratios are similar to the fMLP/LTB₄ ratios observed in the experimental data (Boribong et al., 2019) (see **Figure 3A**). We further quantified the $[FPR1]/[BLT1]$ ratio past the 5 h in the experiment. For the $[LPS](0) = 1$ ng/mL LPS challenge, $[FPR1]/[BLT1](t) = 1$ for all $t \geq 5$. By contrast, for the $[LPS](0) = 100$ ng/mL LPS challenge, the $[FPR1]/[BLT1](t)$ ratio becomes larger and larger as t increases, with the majority of cells favoring the primary fMLP gradient (not shown).

To determine the relationship between the LPS challenge dose and the FPR1/BLT1 ratio, we derived a graph that quantifies $[FPR1]/[BLT1](5)$, 5 h following cell priming, as a function of the $[LPS]$ dose, predicted by model (1) and parameter values/initial conditions in **Table 1**. We found that the experimental observation for the super-low-dose (1 ng/mL) LPS, $[FPR1]/[BLT1](5) = 1$, is preserved for all challenges with LPS values lower than 3.9 ng/mL. For 4 – 6.7 ng/mL LPS challenge, $[BLT1]$ exceeds $[FPR1]$ at $t = 5$ h, but the two receptors will eventually reach identical levels at equilibrium.

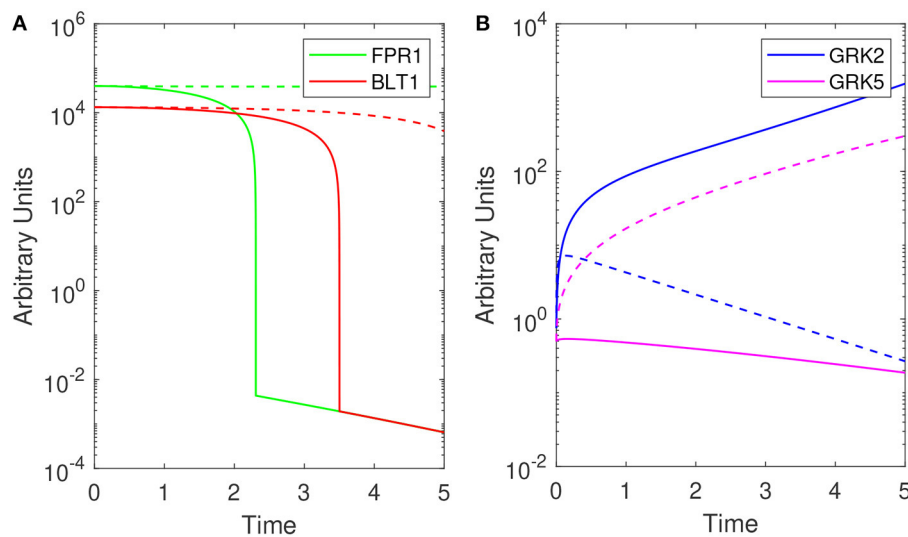


FIGURE 4 | Theoretical results: Dynamics of (A) $[FPR1]$ (green) and $[BLT1]$ (red) and (B) $[GRK2]$ (purple) and $[GRK5]$ (pink) for $[LPS](0) = 1$ ng/mL (solid lines) and $[LPS](0) = 100$ ng/mL (dashed lines) as given by model (1). Parameters and initial conditions are given in **Table 1**.

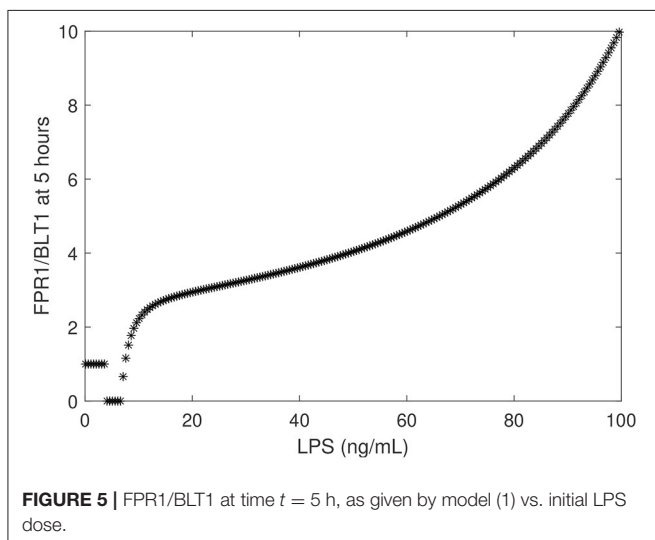


FIGURE 5 | $FPR1/BLT1$ at time $t = 5$ h, as given by model (1) vs. initial LPS dose.

Lastly, the $[FPR1]/[BLT1](5)$ ratio grows larger than one and keeps increasing for LPS dose > 6.7 ng/mL, eventually reaching the experimental prediction of ten, $[FPR1]/[BLT1](5) = 10$, for high-dose LPS challenge (100 ng/mL) (see **Figure 5**) and increasing further as time passes or for higher challenge (not shown).

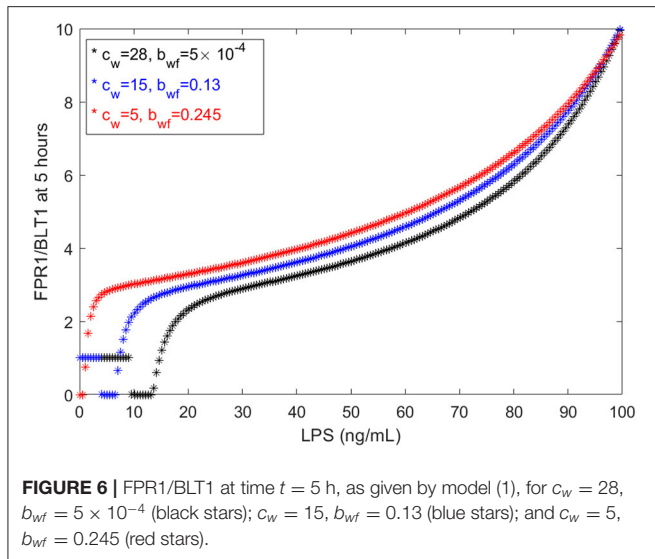
3.2. Long-Term Results and Motifs of Bistability

We have chosen the parameters in model (1) such that the $[FPR1]/[BLT1](5)$ ratio matches the observed fMLP/LTB4 data (Boribong et al., 2019). We are interested in determining how this balance can be broken and which interactions are responsible for the bimodal switch between equal $[FPR1]$ and $[BLT1]$ values and

dominant $[FPR1]$ values. The results presented at $t = 5$ h are transient results. At equilibrium, the $[FPR1]/[BLT1]$ ratio is 1 for $LPS < 15$ ng/mL and as large as 10^7 for $LPS = 100$ ng/mL. This indicates that all $[BLT1]$ molecules have been down regulated, and only $[FPR1]$ molecules remain on the surface of neutrophils. This is due to the large non-LPS activation rate of $[GRK2]$ protein, $c_w = 15$. If we either increase the c_w value to $c_w = 28$ or decrease it to $c_w = 5$, we maintain the $[FPR1]/[BLT1]$ ratio 5 h after super-low-dose (1 ng/mL) and high-dose (100 ng/mL), $[FPR1]/[BLT1](5)$, if we simultaneously decrease the inhibition rate of $[GRK5]$ to 5×10^{-4} or increase it to 0.25, respectively (see **Figure 6**). The range of LPS initial conditions that lead to identical $[FPR1]$ and $[BLT1]$ distribution decrease as c_w decreases, with the $[FPR1]/[BLT1](5) = 1$ prediction being lost for $c_w = 5$ (see **Figure 6**, red curves). The results are insensitive to the LPS decay rate, or to $[FPR1]$ and $[BLT1]$ values where phosphorylation and dephosphorylation levels are half-maximal (not shown). These results suggests that the bimodal switch between the $[FPR1]/[BLT1]$ levels is due to mutual inhibition of GRK2-GRK5 kinases. To confirm this, we removed the inhibition factors, by replacing $(c_w + a_w[LPS])/(b_{fw} + [GRK5]^n)$ with $(c_w + a_w[LPS])/b_{fw}$ and $(c_f + a_f[LPS])/(b_{wf} + [GRK2])$ with $(c_f + a_f[LPS])/b_{wf}$. When the mutual inhibition is removed, the equilibrium $[FPR1]/[BLT1]$ levels are constant, and equal to 1,500, regardless of the size of LPS stimulus.

3.3. Molecular Mechanisms of Cell Oscillatory Migration

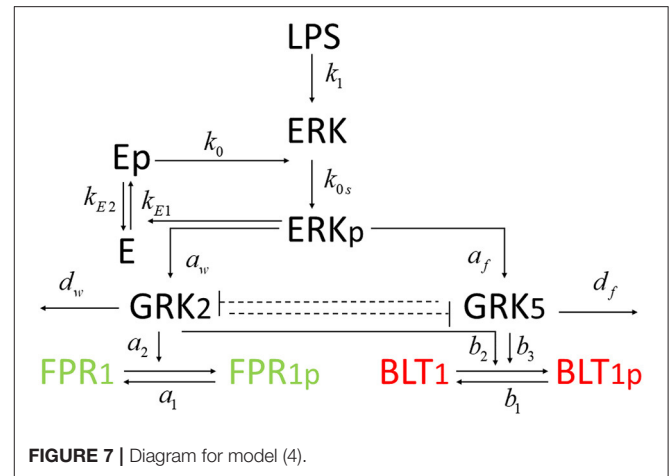
Experimental results reported that neutrophils treated overnight with LPS may lose their ability to move up the chemoattractant gradient, become disoriented, and display oscillatory behavior (Boribong et al., 2019). Moreover, the highest number of cells to display such oscillatory behavior occurs following LPS exposure with low dose of 1 mg/ml (see **Figure 3B**) (Boribong et al., 2019).



To determine the molecular mechanisms responsible for the oscillations, we extended the bistable system (1), by coupling it with an activator-inhibitor oscillatory model for the dynamics of non-phosphorylated and phosphorylated extracellular signal-regulated kinases, $[ERK]$ and $[ERK_p]$, and two auxiliary enzymes, $[E]$ and $[E_p]$ based on diagram (7), model (4) and parameter values/initial conditions in **Tables 1, 2**. Moreover LPS is fixed at its initial condition. Under the chosen parameters, we obtain long-term oscillatory movement for all populations, for constant super-low-challenge LPS (1 ng/mL), as predicted by the data (Boribong et al., 2019) (see **Figures 8A–C**). Interestingly, the oscillatory behavior is maintained for a short LPS range, (0.5–1.1) ng/mL; and corresponds to equal distribution of $[FPR1]$ and $[BLT1]$ receptors. If we either lower the constant LPS challenge to < 0.4 ng/mL or increase it to 1.2 – 375 ng/mL, we obtain equal density of $[FPR1]$ and $[BLT1]$ receptors, but no oscillations (see **Figures 8D–F**). If the LPS constant challenge is increased further, to > 376 ng/mL, $[FPR1]$ receptors dominate the outcome (see **Figures 8G–I**). While the switch between low and high $[FPR1]/[BLT1]$ ratio observed in the constant high dose LPS challenge is due to mutual inhibitions of $[GRK2]$ and $[GRK5]$ kinases, as observed in model (1), the oscillatory dynamics are due to the oscillatory dynamics of the $[ERK]$ and $[ERK_p]$ kinases. This oscillatory behavior can be broken by either increasing or decreasing the constant LPS challenge. Such information can inform interventions, as dysoriented neutrophil movement is not desirable and has been shown to have negative effects during pathogenic infections. Dysregulated neutrophil response to infection can lead to sepsis and end-organ failure and is a leading cause of death worldwide (Reddy and Standiford, 2010; Shen et al., 2017).

4. DISCUSSION

In this study, we developed compartmental mathematical models of molecular interactions that govern neutrophil migratory

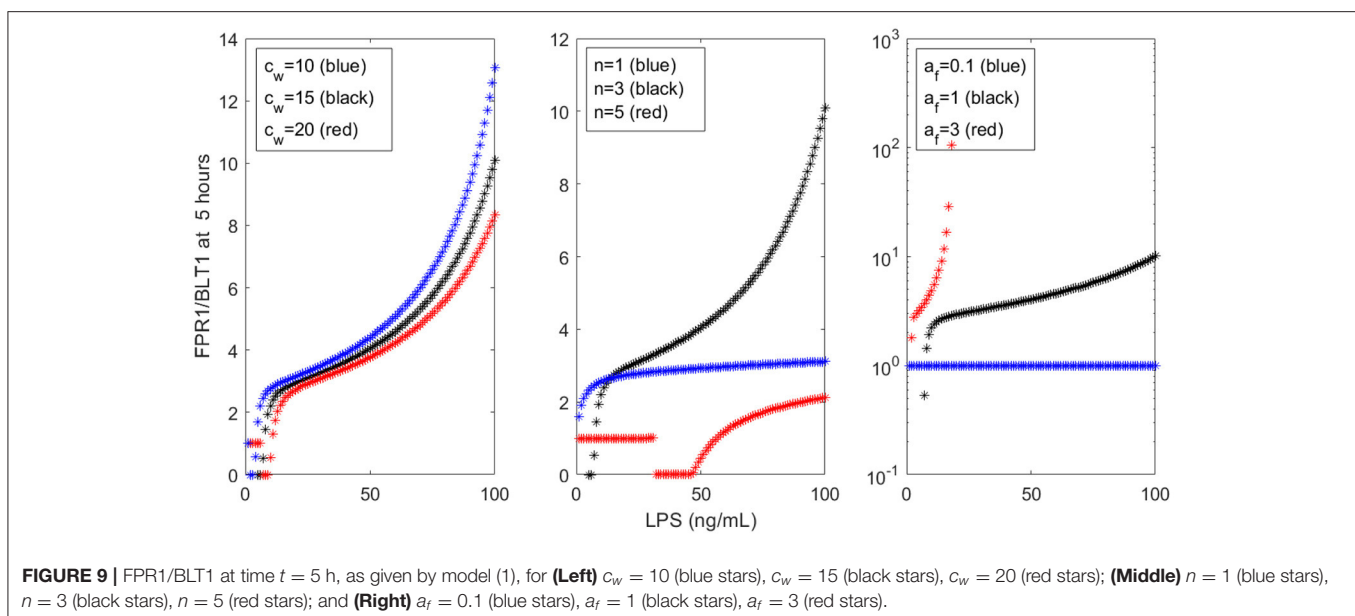
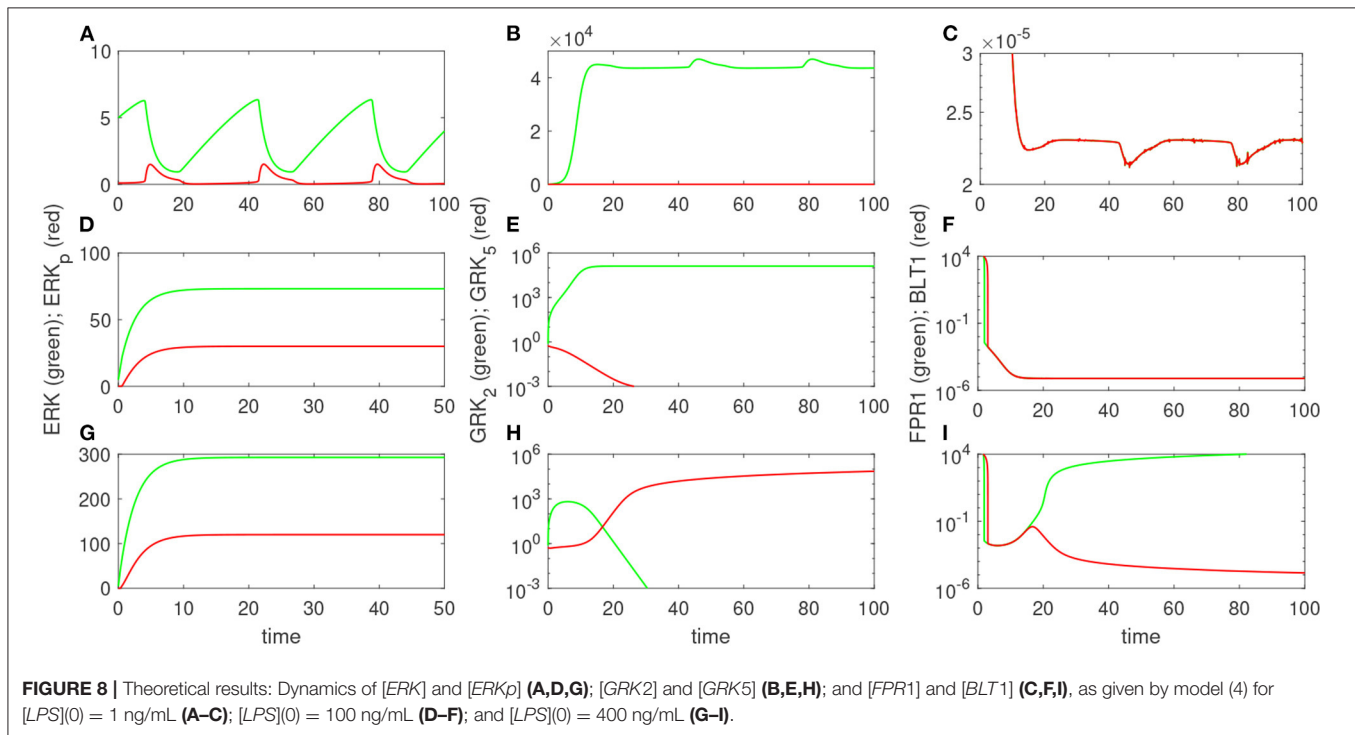


patterns when exposed to competing chemoattractants and challenged with external stimuli. When the models were restricted to the interactions between the chemoattractants' receptors, their kinases, and LPS, we predicted a bistable switch between two states: one in which the densities of the two chemoattractant receptors, FPR1 and BLT1, are equal and one in which the receptors for the primary chemoattractant, FPR1, dominate. We hypothesized that the two states correspond to two states observed experimentally: equal migration toward the primary and intermediary chemoattractants, fMLP and LTB4, and predominant migration toward the primary chemoattractant, fMLP (Boribong et al., 2019). The experimental data connected the differential migratory outcomes with the magnitude of the external LPS challenge, with super-low 1 ng/mL LPS leading to equal migration toward both chemoattractants and high 100 ng/mL LPS leading to ten times higher migration toward fMLP, 5 h after challenge. In the mathematical model, the external signal corresponds to the initial condition for variable LPS. Our model was calibrated to match the experimental data 5 h after stimuli, with $[FPR1]/[BLT1](5) = 1$ for $[LPS](0) = 1$ and $[FPR1]/[BLT1](5) = 10$ for $[LPS](0) = 100$. Furthermore, 5 h after LPS challenge, we obtain equal levels of FPR1 and BLT1 receptors for all initial conditions $[LPS](0) < 3.7$ ng/mL and increasingly more FPR1 than BLT1 receptors when the initial condition for LPS increases, with ten times more FPR1 than BLT1 receptors when $[LPS](0) = 100$ ng/mL. When we run the model to equilibrium, however, we obtain equal FPR1 and BLT1 receptors for an even larger range of LPS initial conditions, < 15 ng/mL; and dominant FPR1 levels for $LPS > 16$ ng/mL. This implies that not just the challenge dose, but the duration of the experiment may influence the quantitative outcomes.

We investigated the molecular interactions that are responsible for the bistable outcomes and found that when the mutual inhibition of GRK2 and GRK5 kinases is either removed, or the balance is broken, the model is no longer bistable. Instead, when run to equilibrium, it settles into a state where FPR1 receptors dominate the outcome, indicative of predominant migration toward the primary chemoattractant.

When the models were expanded to add the interaction with the ERK signaling pathways under constant LPS challenge, we obtained a third dynamical state, where we have equal FPR1/BLT1=1, but the equilibrium is lost, and the FPR1 and BLT1 receptors oscillate between two values. We assume this to be indicative of neutrophil oscillation which, in the experimental setup, is equivalent to cells changing direction (Boribong et al., 2019). We investigated how changes in the constant LPS dose affect the outcomes and found that, at equilibrium, FPR1 and

BLT1 receptors are equal and non-oscillating for both super-low (< 0.5 ng/mL) and high dose (1.1, 375 ng/mL) LPS; are equal and oscillating for super-low dose (0.5–1.1 ng/mL) LPS; and FPR1 outnumbers BLT1 for super-high dose (> 376 ng/mL) LPS. The non-asymptotic dynamics are due to the oscillatory behavior of phosphorylated and non-phosphorylated ERK molecules, who undergo an auto-catalytic interaction with two undefined enzymes. We have modeled the interaction using an activation-inhibition motif, with similar dynamics being



obtained if the oscillations are induced by a substrate depletion motif (Tyson et al., 2003). Further work is needed to determine the nature of enzyme or their regulation. We are currently working to validate the model by retrieving neutrophils from our microfluidic device post-migration and quantifying FPR1, BLT1, GRK2, and GRK 5 levels by Droplet DigitalTM PCR (ddPCRTM).

Our models are limited by the presence of many unknown parameters. While we strived to match the empirical data, the results are mostly qualitative. Similar results can be obtained with many different parameters sets. For example, c_w only slightly influences the FPR1/BLT1 ratio at $t = 5$ h (see Figure 9, left panel), while the cooperativity coefficient n and the LPS-dependent GRK5 production a_f have drastic effects on the size of the ration (see Figure 9, middle and right panels). The unchanging factors, however, are the motifs of bistability, which are induced by the dual inhibition of the G-protein kinases; the oscillatory motifs, which are induced by the oscillatory ERK dynamics; and the influence of external stimuli on outcomes.

In conclusion, we developed mathematical models for the molecular interactions responsible for neutrophils migratory phenotypes, calibrated them against empirical data, and used their dynamics to determine the external stimuli ranges that account for neutrophils migration toward a pro-inflammatory chemoattractant, a pro-tolerant chemoattractant, or oscillatory dynamics indicative of dysorientation and loss of function. Understanding the relationship between neutrophils' dynamics and the mechanisms responsible for their movement is important for preventing and predicting immune disorders. Activation markers in neutrophils are potential biomarkers for the diagnosis and prognosis of sepsis. Septic patients can be screened for

neutrophil markers, such as GRK 2/5, FPR1, and BLT1, and this predictive model can guide patient treatment. In the future, we plan to expand this model to include more complex signaling from the microenvironment and aim to predict not only dysfunctional migration in neutrophils, but even the probability of cells accumulating in specific organs, such as the lung or kidney. We can use these predictive models to define optimal patient treatment and to identify immunotherapeutic targets (i.e., small molecule inhibition, microRNAs, gene therapy) to promote directional neutrophil migration.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

SC, SK, and BB wrote the code. SC wrote the manuscript. All authors reviewed and revised the manuscript and conceived the study.

ACKNOWLEDGMENTS

SK and SC acknowledge funding from National Science Foundation grant No. 1813011. BB and CJ acknowledge funding from The National Institute of General Medical Sciences of the National Institutes of Health under award number R35GM133610.

REFERENCES

- Alves-Filho, J. C., Benjamim, C., Tavares-Murta, B. M., and Cunha, F. Q. (2005). Failure of neutrophil migration toward infectious focus in severe sepsis: a critical event for the outcome of this syndrome. *Memorias do Instituto Oswaldo Cruz* 100, 223–226. doi: 10.1590/S0074-02762005000900038
- Alves-Filho, J. C., Spiller, F., and Cunha, F. Q. (2010). Neutrophil paralysis in sepsis. *Shock* 34, 15–21. doi: 10.1097/SHK.0b013e3181e7e61b
- Arraes, S. M. A., Freitas, M. S., da Silva, S. V., de Paula Neto, H. A., Alves-Filho, J. C., Martins, M. A., et al. (2006). Impaired neutrophil chemotaxis in sepsis associates with grk expression and inhibition of actin assembly and tyrosine phosphorylation. *Blood* 108, 2906–2913. doi: 10.1182/blood-2006-05-024638
- Bayani, A., Dunster, J., Crofts, J., and Nelson, M. (2020). Mechanisms and points of control in the spread of inflammation: a mathematical investigation. *Bull. Math. Biol.* 82, 1–22. doi: 10.1007/s11538-020-00709-y
- Boribong, B. P., Lenzi, M. J., Li, L., and Jones, C. N. (2019). Super-low dose lipopolysaccharide dysregulates neutrophil migratory decision-making. *Front. Immunol.* 10:359. doi: 10.3389/fimmu.2019.00359
- Byrne, M. B., Kimura, Y., Kapoor, A., He, Y., Mattam, K. S., Hasan, K. M., et al. (2014). Oscillatory behavior of neutrophils under opposing chemoattractant gradients supports a winner-take-all mechanism. *PLoS ONE* 9:e85726. doi: 10.1371/journal.pone.0085726
- Chen, Z., Gaudreau, R., Le Gouill, C., Rola-Pleszczynski, M., and Stanková, J. (2004). Agonist-induced internalization of leukotriene B4 receptor 1 requires G-protein-coupled receptor kinase 2 but not arrestins. *Mol. Pharmacol.* 66, 377–386. doi: 10.1124/mol.66.3
- Ciupe, S. M., Miller, C. J., and Forde, J. E. (2018). A bistable switch in virus dynamics can explain the differences in disease outcome following SIV infections in rhesus macaques. *Front. Microbiol.* 9:1216. doi: 10.3389/fmicb.2018.01216
- Ciupe, S. M., Ribeiro, R. M., Nelson, P. W., and Perelson, A. S. (2007). Modeling the mechanisms of acute hepatitis B virus infection. *J. Theor. Biol.* 247, 23–35. doi: 10.1016/j.jtbi.2007.02.017
- Davenport, A. P., Scully, C. C., de Graaf, C., Brown, A. J., and Maguire, J. J. (2020). Advances in therapeutic peptides targeting G protein-coupled receptors. *Nat. Rev. Drug Discov.* 19, 389–413. doi: 10.1038/s41573-020-0062-z
- Day, J., Rubin, J., Vodovotz, Y., Chow, C. C., Reynolds, A., and Clermont, G. (2006). A reduced mathematical model of the acute inflammatory response II. Capturing scenarios of repeated endotoxin administration. *J. Theor. Biol.* 242, 237–256. doi: 10.1016/j.jtbi.2006.02.015
- Demaret, J., Venet, F., Friggeri, A., Cazalis, M.-A., Plassais, J., Jallades, L., et al. (2015). Marked alterations of neutrophil functions during sepsis-induced immunosuppression. *J. Leukocyte Biol.* 98, 1081–1090. doi: 10.1189/jlb.4A0415-168RR
- Dianqing, W. (2005). Signaling mechanisms for regulation of chemotaxis. *Cell Res.* 15, 52–56. doi: 10.1038/sj.cr.7290265
- Fan, J., and Malik, A. B. (2003). Toll-like receptor-4 (TLR4) signaling augments chemokine-induced neutrophil migration by modulating cell surface expression of chemokine receptors. *Nat. Med.* 9, 315–321. doi: 10.1038/nm832
- Fischer, H. P. (2008). Mathematical modeling of complex biological systems: from parts lists to understanding systems behavior. *Alcohol Res. Health* 31:49.
- Futosi, K., Fodor, S., and Mócsai, A. (2013). Neutrophil cell surface receptors and their intracellular signal transduction pathways. *Int. Immunopharmacol.* 17, 1185–1197. doi: 10.1016/j.intimp.2013.11.010
- Gaudreau, R., Le Gouill, C., Venne, M.-H., Stankova, J., and Rola-Pleszczynski, M. (2002). Threonine 308 within a putative casein kinase 2 site of the cytoplasmic

- tail of leukotriene B4 receptor (BLT1) is crucial for ligand-induced, G-protein-coupled receptor-specific kinase 6-mediated desensitization. *J. Biol. Chem.* 277, 31567–31576. doi: 10.1074/jbc.M202723200
- Gaudreault, E., Thompson, C., Stankova, J., and Rola-Pleszczynski, M. (2005). Involvement of BLT1 endocytosis and yes kinase activation in leukotriene b4-induced neutrophil degranulation. *J. Immunol.* 174, 3617–3625. doi: 10.4049/jimmunol.174.6.3617
- Goldbeter, A., and Koshland, D. E. (1981). An amplified sensitivity arising from covalent modification in biological systems. *Proc. Natl. Acad. Sci. U.S.A.* 78, 6840–6844. doi: 10.1073/pnas.78.11.6840
- Heit, B., Robbins, S. M., Downey, C. M., Guan, Z., Colarusso, P., Miller, B. J., et al. (2008). Pten functions to 'prioritize' chemotactic cues and prevent 'distraction' in migrating neutrophils. *Nat. Immunol.* 9, 743. doi: 10.1038/ni.1623
- Heit, B., Tavener, S., Raharjo, E., and Kubes, P. (2002). An intracellular signaling hierarchy determines direction of migration in opposing chemotactic gradients. *J. Cell Biol.* 159, 91–102. doi: 10.1083/jcb.200202114
- Ionides, E. L., Fang, K. S., Isseroff, R. R., and Oster, G. F. (2004). Stochastic models for cell motion and taxis. *J. Math. Biol.* 48, 23–37. doi: 10.1007/s00285-003-0220-z
- Jones, C. N., Hoang, A. N., Martel, J. M., Dimisko, L., Mikkola, A., Inoue, Y., et al. (2016). Microfluidic assay for precise measurements of mouse, rat, and human neutrophil chemotaxis in whole-blood droplets. *J. Leukocyte Biol.* 100, 241–247. doi: 10.1189/jlb.5TA0715-310RR
- Kadelka, S., Boribong, B. P., Li, L., and Ciupe, S. M. (2019). Modeling the bistable dynamics of the innate immune system. *Bull. Math. Biol.* 81, 256–276. doi: 10.1007/s11538-018-0527-y
- Kolaczowska, E., and Kubes, P. (2013). Neutrophil recruitment and function in health and inflammation. *Nat. Rev. Immunol.* 13, 159–175. doi: 10.1038/nri3399
- Leber, A., Abedi, V., Hontecillas, R., Viladomiu, M., Hoops, S., Ciupe, S., et al. (2016). Bistability analyses of CD4+ T follicular helper and regulatory cells during *Helicobacter pylori* infection. *J. Theor. Biol.* 398, 74–84. doi: 10.1016/j.jtbi.2016.02.036
- Leoni, G., Gripenrot, J., Lord, C., Riesselman, M., Sumagin, R., Parkos, C. A., et al. (2015). Human neutrophil formyl peptide receptor phosphorylation and the mucosal inflammatory response. *J. Leukocyte Biol.* 97, 87–101. doi: 10.1189/jlb.4A0314-153R
- Liu, X., Ma, B., Malik, A. B., Tang, H., Yang, T., Sun, B., et al. (2012). Bidirectional regulation of neutrophil migration by mitogen-activated protein kinases. *Nat. Immunol.* 13, 457. doi: 10.1038/ni.2258
- Luo, J., and Benovic, J. L. (2003). G protein-coupled receptor kinase interaction with Hsp90 mediates kinase maturation. *J. Biol. Chem.* 278, 50908–50914. doi: 10.1074/jbc.M307637200
- Magalhaes, A. C., Dunn, H., and Ferguson, S. S. (2012). Regulation of GPCR activity, trafficking and localization by GPCR-interacting proteins. *Brit. J. Pharmacol.* 165, 1717–1736. doi: 10.1111/j.1476-5381.2011.01552.x
- Mócsai, A., Walzog, B., and Lowell, C. A. (2015). Intracellular signalling during neutrophil recruitment. *Cardiovasc. Res.* 107, 373–385. doi: 10.1093/cvr/cvv159
- Murphy, P. M. (1994). The molecular biology of leukocyte chemoattractant receptors. *Annu. Rev. Immunol.* 12, 593–633. doi: 10.1146/annurev.iy.12.040194.003113
- Neel, N. F., Schutysse, E., Sai, J., Fan, G.-H., and Richmond, A. (2005). Chemokine receptor internalization and intracellular trafficking. *Cytokine Growth Factor Rev.* 16, 637–658. doi: 10.1016/j.cytogfr.2005.05.008
- Nelson, P., Smith, N., Ciupe, S., Zou, W., Omenn, G. S., and Pietropaolo, M. (2009). Modeling dynamic fluctuations in type 1 diabetes progression: quantifying β -cell variation after the appearance of islet-specific autoimmune responses. *Math. Biosci. Eng.* 6, 753. doi: 10.3934/mbe.2009.6.753
- Penela, P., Ruiz-Gómez, A., Castaño, J. G., and Mayor, F. (1998). Degradation of the G protein-coupled receptor kinase 2 by the proteasome pathway. *J. Biol. Chem.* 273, 35238–35244. doi: 10.1074/jbc.273.52.35238
- Petri, B., and Sanz, M.-J. (2018). Neutrophil chemotaxis. *Cell Tissue Res.* 371, 425–436. doi: 10.1007/s00441-017-2776-8
- Philipp, M., Berger, I. M., Just, S., and Caron, M. G. (2014). Overlapping and opposing functions of G protein-coupled receptor kinase 2 (GRK2) and GRK5 during heart development. *J. Biol. Chem.* 289, 26119–26130. doi: 10.1074/jbc.M114.551952
- Pillay, J., Ramakers, B. P., Kamp, V. M., Loi, A. L. T., Lam, S. W., Hietbrink, F., et al. (2010). Functional heterogeneity and differential priming of circulating neutrophils in human experimental endotoxemia. *J. Leukocyte Biol.* 88, 211–220. doi: 10.1189/jlb.1209793
- Postma, M., and van Haastert, P. J. (2016). Mathematics of experimentally generated chemoattractant gradients. *Methods Mol Biol.* 1407, 381–396. doi: 10.1007/978-1-4939-3480-5_26
- Prossnitz, E. R., Kim, C. M., Benovic, J. L., and Richard, D. Y. (1995). Phosphorylation of the N-formyl peptide receptor carboxyl terminus by the G protein-coupled receptor kinase, GRK2. *J. Biol. Chem.* 270, 1130–1137. doi: 10.1074/jbc.270.3.1130
- Reddy, R. C., and Standiford, T. J. (2010). Effects of sepsis on neutrophil chemotaxis. *Curr. Opin. Hematol.* 17, 18–24. doi: 10.1097/MOH.0b013e32833338f3
- Reynolds, A., Rubin, J., Clermont, G., Day, J., Vodovotz, Y., and Ermentrout, G. B. (2006). A reduced mathematical model of the acute inflammatory response: I. Derivation of model and analysis of anti-inflammation. *J. Theor. Biol.* 242, 220–236. doi: 10.1016/j.jtbi.2006.02.016
- Schneider, E. H., Weaver, J. D., Gaur, S. S., Tripathi, B. K., Jesaitis, A. J., Zelenka, P. S., et al. (2012). The leukocyte chemotactic receptor FPR1 is functionally expressed on human lens epithelial cells. *J. Biol. Chem.* 287, 40779–40792. doi: 10.1074/jbc.M112.411181
- Shen, X.-F., Cao, K., Jiang, J. P., Guan, W.-X., and Du, J.-F. (2017). Neutrophil dysregulation during sepsis: an overview and update. *J. Cell. Mol. Med.* 21, 1687–1697. doi: 10.1111/jcmm.13112
- Sorriento, D., Ciccarelli, M., Santulli, G., Campanile, A., Altobelli, G. G., Cimini, V., et al. (2008). The G-protein-coupled receptor kinase 5 inhibits NF- κ B transcriptional activity by inducing nuclear accumulation of κ B α . *Proc. Natl. Acad. Sci. U.S.A.* 105, 17818–17823. doi: 10.1073/pnas.0804446105
- Tyson, J. J., Chen, K. C., and Novak, B. (2003). Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr. Opin. Cell Biol.* 15, 221–231. doi: 10.1016/S0955-0674(03)00017-6
- Vodovotz, Y., Constantine, G., Rubin, J., Csete, M., Voit, E. O., and An, G. (2009). Mechanistic simulations of inflammation: current state and future prospects. *Math. Biosci.* 217, 1–10. doi: 10.1016/j.mbs.2008.07.013
- Wang, X., Qin, W., Song, M., Zhang, Y., and Sun, B. (2016a). Exogenous carbon monoxide inhibits neutrophil infiltration in LPS-induced sepsis by interfering with FPR1 via p38 MAPK but not GRK2. *Oncotarget* 7:34250. doi: 10.18632/oncotarget.9084
- Wang, X., Qin, W., Zhang, Y., Zhang, H., and Sun, B. (2016b). Endotoxin promotes neutrophil hierarchical chemotaxis via the p38-membrane receptor pathway. *Oncotarget* 7:74247. doi: 10.18632/oncotarget.12093
- Wu, Z., Chen, Y., Yang, T., Gao, Q., Yuan, M., Ma, L., et al. (2012). Targeted ubiquitination and degradation of G-protein-coupled receptor kinase 5 by the DDB1-CUL4 ubiquitin ligase complex. *PLoS ONE* 7:e43997. doi: 10.1371/journal.pone.0043997
- Yuan, R., Geng, S., Chen, K., Diao, N., Chu, H. W., and Li, L. (2016a). Low-grade inflammatory polarization of monocytes impairs wound healing. *J. Pathol.* 238, 571–583. doi: 10.1002/path.4680
- Yuan, R., Geng, S., and Li, L. (2016b). Molecular mechanisms that underlie the dynamic adaptation of innate monocyte memory to varying stimulant strength of TLR ligands. *Front. Immunol.* 7:497. doi: 10.3389/fimmu.2016.00497

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Ciupe, Boribong, Kadelka and Jones. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



DEEPsc: A Deep Learning-Based Map Connecting Single-Cell Transcriptomics and Spatial Imaging Data

Floyd Maseda¹, Zixuan Cang^{1,2*} and Qing Nie^{1,2,3*}

¹ Department of Mathematics, University of California, Irvine, Irvine, CA, United States, ² The NSF-Simons Center for Multiscale Cell Fate Research, University of California, Irvine, Irvine, CA, United States, ³ Department of Developmental and Cell Biology, University of California, Irvine, Irvine, CA, United States

OPEN ACCESS

Edited by:

Chunhe Li,
Fudan University, China

Reviewed by:

Shihua Zhang,
Academy of Mathematics
and Systems Science (CAS), China
Xiaoqiang Sun,
Sun Yat-sen University, China

*Correspondence:

Zixuan Cang
zcang@uci.edu
Qing Nie
qnie@uci.edu

Specialty section:

This article was submitted to
Systems Biology,
a section of the journal
Frontiers in Genetics

Received: 02 December 2020

Accepted: 23 February 2021

Published: 23 March 2021

Citation:

Maseda F, Cang Z and Nie Q
(2021) DEEPsc: A Deep
Learning-Based Map Connecting
Single-Cell Transcriptomics
and Spatial Imaging Data.
Front. Genet. 12:636743.
doi: 10.3389/fgene.2021.636743

Single-cell RNA sequencing (scRNA-seq) data provides unprecedented information on cell fate decisions; however, the spatial arrangement of cells is often lost. Several recent computational methods have been developed to impute spatial information onto a scRNA-seq dataset through analyzing known spatial expression patterns of a small subset of genes known as a reference atlas. However, there is a lack of comprehensive analysis of the accuracy, precision, and robustness of the mappings, along with the generalizability of these methods, which are often designed for specific systems. We present a system-adaptive deep learning-based method (DEEPsc) to impute spatial information onto a scRNA-seq dataset from a given spatial reference atlas. By introducing a comprehensive set of metrics that evaluate the spatial mapping methods, we compare DEEPsc with four existing methods on four biological systems. We find that while DEEPsc has comparable accuracy to other methods, an improved balance between precision and robustness is achieved. DEEPsc provides a data-adaptive tool to connect scRNA-seq datasets and spatial imaging datasets to analyze cell fate decisions. Our implementation with a uniform API can serve as a portal with access to all the methods investigated in this work for spatial exploration of cell fate decisions in scRNA-seq data. All methods evaluated in this work are implemented as an open-source software with a uniform interface.

Keywords: spatial gene expression atlas, scRNA-seq data, spatial information imputation, deep learning, metric learning, comprehensive evaluation metric

INTRODUCTION

While cells of a biological system have access to the same genetic blueprint, they navigate through different developmental paths toward various cell fates. These diverse fate programs of cells are controlled by their own states, interactions with spatially neighboring cells, and other environmental cues (Guo et al., 2010). To decipher the processes of cell fate acquisitions, observations of the transcriptomics with single-cell resolution in spatial context are desired. The advent of sophisticated single-cell RNA sequencing (scRNA-seq) techniques now allows investigation of the transcriptomic landscape of tens of thousands of genes across tissues at the

resolution of individual cells (Rosenberg et al., 2018; Svensson et al., 2018). However, a drawback to scRNA-seq methods is the necessity of dissociating the sample in question, thereby destroying any spatial context which can be crucial to the understanding of cellular development and dynamics (Yuan et al., 2017). In current common workflows of scRNA-seq data analysis, unsupervised clustering of cells is carried out, followed by identifying marker genes associated with each cell cluster (Luecken and Theis, 2019). While the list of marker genes for each cell cluster can be screened for genes associated with known spatial regions to estimate the spatial origin of the cluster, the spatial arrangement of individual cells remains unclear (Kiselev et al., 2019; Luecken and Theis, 2019). Several existing methods attempt to impute a pseudospacial or pseudotemporal axis onto the data (Joost et al., 2016; Puram et al., 2017; Pandey et al., 2018; Wang et al., 2019); however, little related to physical space is immediately discernible from scRNA-seq data alone.

The loss of spatial information in scRNA-seq data can be partially mitigated by referring to spatial staining data (Sprague et al., 2006; Fowlkes et al., 2008). Another promising solution is the emerging spatial transcriptomics methods such as osmFISH (Codeluppi et al., 2018), MERFISH (Moffitt et al., 2018), seqFISH (Shah et al., 2016), seqFISH+ (Eng et al., 2019), STARmap (Wang et al., 2018), and Slide-seq (Rodriques et al., 2019) that obtain *in situ* spatial expression patterns. Compared to scRNA-seq, current spatial techniques often cover fewer cells or genes or with a suboptimal resolution and depth. It is therefore a trending theme to combine the strengths of both methods to achieve a high coverage and individual-cell resolution while retaining the spatial arrangement (Yuan et al., 2017; Kiselev et al., 2019). Due to these differences among the scRNA-seq and spatial techniques, and biological systems, it is challenging to derive a generally applicable computation method to integrate the two kinds of data.

Several recent computational methods have been developed to impute spatial data onto existing scRNA-seq datasets through analyzing known spatial expression patterns of a small subset of genes, termed a “spatial reference atlas.” Seminal methods were developed independently by Achim et al. (2015) and Satija et al. (2015) and were applied to the *Platynereis dumerilii* brain and zebrafish embryo, respectively, using binarized reference atlases derived from *in situ* hybridization (ISH) images. DistMap, another method that uses a binarized ISH-based reference atlas, was developed by Karaikos et al. (2017) and applied to the *Drosophila* embryo. Achim et al. (2015) use an empirical correspondence score between each cell-location pair based on the specificity ratio of genes. Satija et al. (2015) (Seurat v1) fits a bimodal mixture model to the scRNA-seq data and then projects cells to their spatial origins using a probabilistic score. DistMap applies Matthew's correlation coefficients to the binarized spatial imaging and scRNA-seq data to assign a cell-location score (Karaikos et al., 2017). Several methods have also been developed which use spatial reference atlases directly measuring the RNA counts that are comparable to scRNA-seq data without binarization (Peng et al., 2016; Halpern et al., 2017). More recently, computational methods have been developed for imputing gene expression in spatial data (Lopez et al., 2019), transferring cell type label from scRNA-seq data

to spatial data (Zhu et al., 2018; Dries et al., 2019; Andersson et al., 2020), *de novo* spatial placement of single cells (Nitzan et al., 2019), and inferring spatial distances between single cells (Cang and Nie, 2020).

In addition to the methods designed specifically for integrating spatial data and scRNA-seq data, other computational methods have been developed recently for general data integration. Such methods focus on the general task of integrating RNA sequencing datasets obtained from the same biological system through different technologies, *in situ* data being one possibility among many, into one large dataset offering a more complete description of the system under study. These methods include newer versions of Seurat (Butler et al., 2018; Stuart et al., 2019), LIGER (Welch et al., 2019), Harmony (Korsunsky et al., 2019), and Scanorama (Hie et al., 2019) which are mainly based on correlation analyses and matrix factorizations. Another more specific task is to transfer high-level information such as cell types between datasets. Many machine learning- and deep learning-based methods have been developed for this task by formulating a supervised learning problem with the high-level information being the target (Kiselev et al., 2018; Lieberman et al., 2018; Lopez et al., 2018; Wagner and Yanai, 2018; Tan and Cahan, 2019; Boufea et al., 2020; Hu et al., 2020; Ma and Pellegrini, 2020).

Since the spatial characteristics of different biological systems could be significantly different, we aim to develop a system-adaptive method specifically designed for imputing spatial information onto scRNA-seq data. To this end, unlike other spatial integration methods that use predefined algorithms for computing scores, we learn a specialized correspondence score between cells and locations for a given biological system. This can then be regarded as a general metric learning task (Kulis, 2013). In addition to linear methods that learn a pseudometric (Weinberger and Saul, 2009), there has been increasing interest in applying deep learning to metric learning (Kaya and Bilge, 2019; Chicco, 2020). These methods are mostly designed for cases where the pair of data points to be compared are in the same space. Though the common genes from the spatial data and scRNA-seq data are used here, directly treating them as in the same space may cause inaccuracy due to differences in the original datasets such as scaling and noise.

Here we develop a system-adaptive deep learning-based method (DEEPsc) for imputing spatial data onto scRNA-seq data. A DEEPsc network accepts a low-dimensional feature vector corresponding to a single position in the spatial reference atlas along with a corresponding feature vector of the gene expression of a single cell and returns a likelihood the input cell originated from the input position. The network is trained and validated using positions in the spatial reference atlas as simulated scRNA-seq data. The network is also validated through the task of predicting the scRNA-seq data from the spatial reference atlas or the other way around. In addition, we implemented several strong baseline methods using different norms and linear metric learning for benchmark comparison. We further develop a comprehensive measure, which was previously lacking, for evaluating how well a given method maps scRNA-seq data to known spatial origins, called a performance score. This score contains three components that measure the accuracy,

precision, and robustness of a method, respectively. Using this score on four biological systems, we show that DEEPsc maintains a comparable accuracy to four existing methods while achieving a better balance between precision and robustness.

RESULTS

A Deep-Learning Based Method to Connect scRNA-seq Datasets and Spatial Imaging Data

Given any spatial reference atlas consisting of binary or continuous gene expression levels for a biological system on a set of locations with known spatial coordinates, and a scRNA-seq dataset consisting of binary or continuous gene expression levels for the same biological system, we introduce a Deep-learning based Environment for the Extraction of Positional information from scRNA-seq data (DEEPsc) to impute the spatial information onto the scRNA-seq data.

In DEEPsc, we first select a common set of genes from the reference atlas and scRNA-seq data, then perform dimensionality reduction via principal component analysis (PCA) on the reduced reference atlas to shorten training time (Figure 1A). The scRNA-seq data is then projected into the same PCA space on which we learn a metric for comparison between cells and spatial positions. The DEEPsc network accepts a concatenated feature vector for a single cell and a single position and returns a likelihood the input cell originated from the input position. The network contains two fully connected hidden layers with N nodes each, where N is the number of principal components kept from PCA, and a single node in the output layer. Sigmoid activation functions are applied to each node, including the output node, so that the resulting output is in $[0, 1]$ and can be interpreted as a likelihood that the input cell originated from the input spatial position. To train the DEEPsc network, we use the spatial position feature vectors as simulated scRNA-seq data for comparison (Figure 1B). Each simulated cell is compared pairwise with every position in the spatial reference atlas; if the simulated cell is an exact match to a given position, the target output is 1 (a high likelihood of origin), and if the simulated cell and chosen position are not an exact match, the target output is 0 (a low likelihood of origin). Training is terminated when the error on a randomly chosen validation set is no longer improving.

After training the DEEPsc network, a feature vector associated with an actual cell from the scRNA-seq data is fed in as input and compared to each position in the reference atlas individually. We display the results as a heatmap on the schematic diagram of the biological system, choosing the spatial position with the largest likelihood of origin according to DEEPsc as the determined origin of the cell. This process is repeated for each cell in the scRNA-seq dataset to assign spatial origins of all cells (Figure 1C).

Quantifying Spatial Mapping Performance

Each of the highlighted methods to impute spatial data onto scRNA-seq data, including DEEPsc, can be essentially boiled

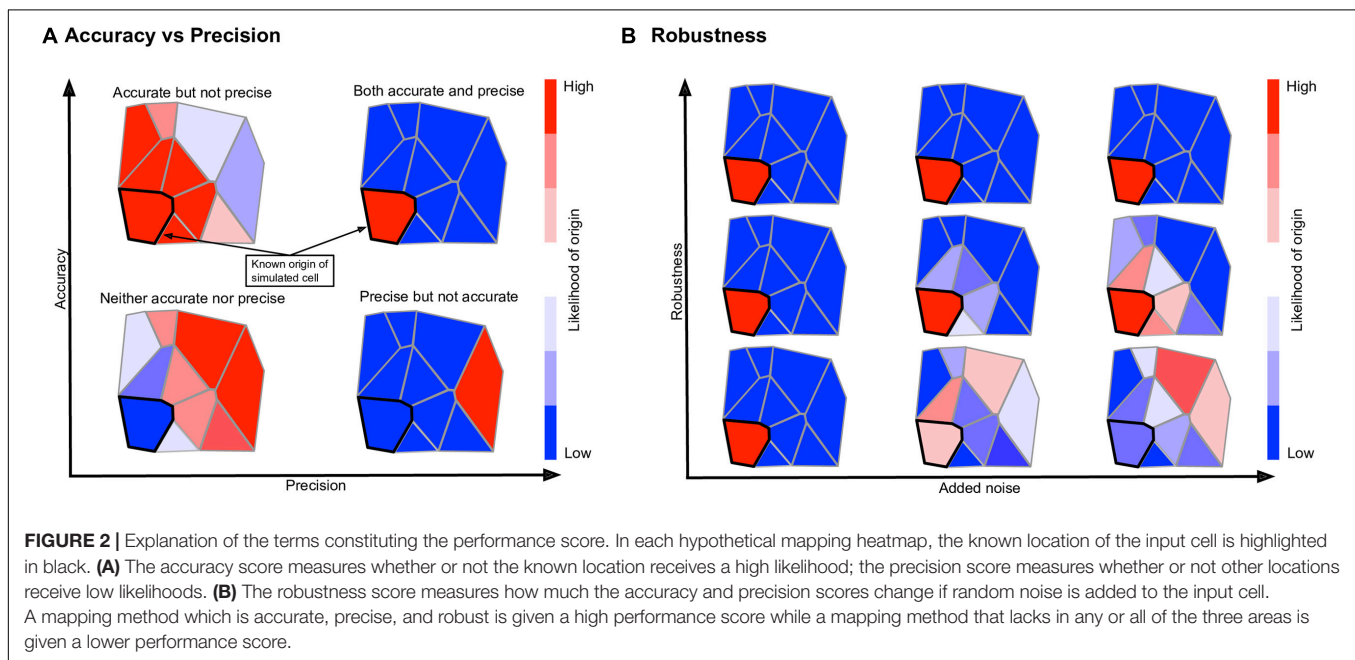
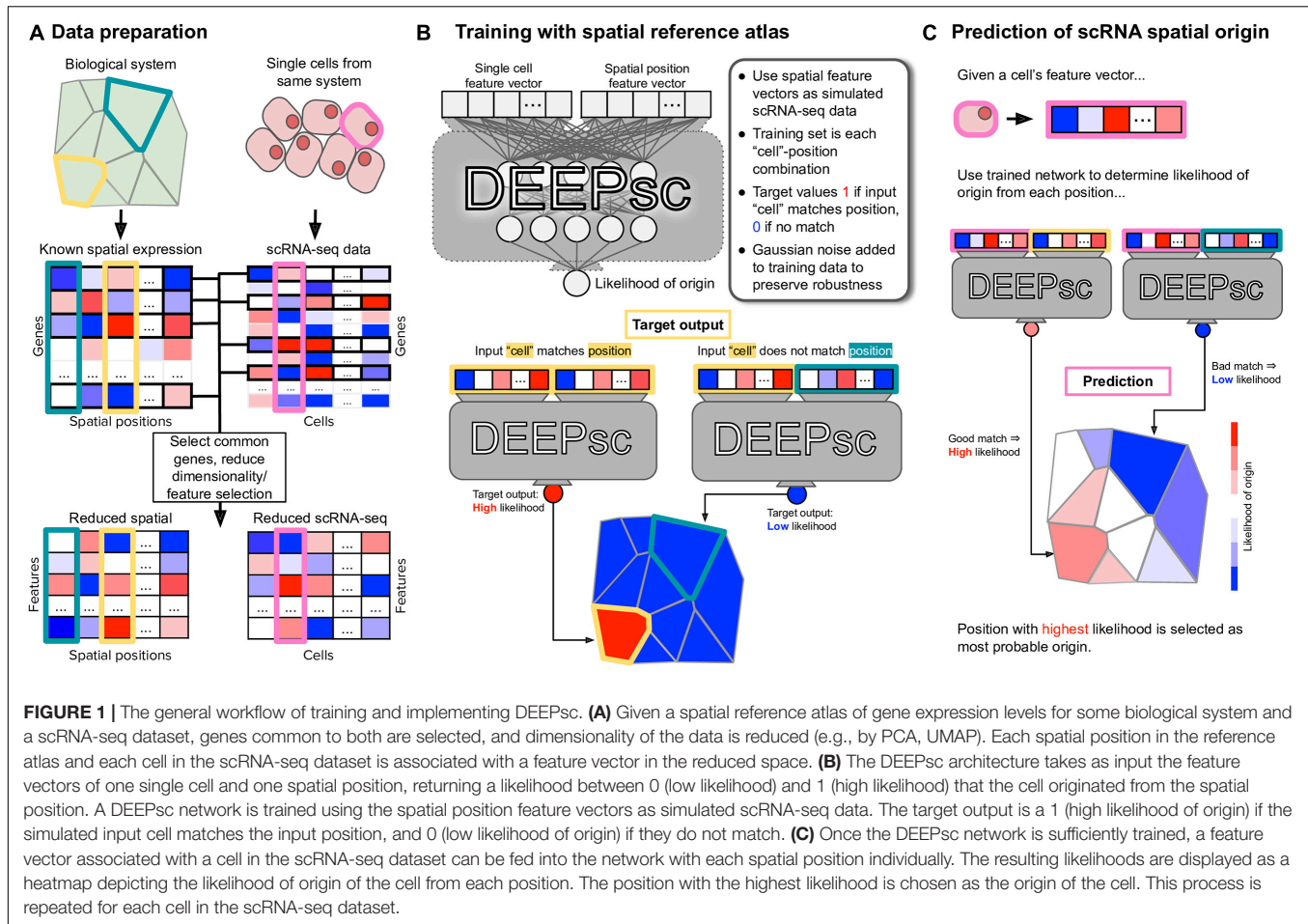
down to the following: For some tissue with a well-defined standard spatial structure, given known binary or continuous expression levels of G genes at each of P spatial locations (the reference atlas), calculate a correspondence score, S , of how similar each of C cells in an scRNA-seq dataset is to each of the P positions in the atlas. That is, define a function, $S: [0, 1]^G \times [0, 1]^G \rightarrow [0, 1]$, such that $S(c_i, p_j)$; $i = 1, 2, \dots, C$; $j = 1, 2, \dots, P$; which describes the likelihood that cell c_i originated from position p_j , based on the similarity of the expression vectors of the cell and position.

To quantify how well a given method performs for a given spatial reference atlas, we use the reference atlas itself as simulated single cell data; that is, we generate a simulated scRNA-seq dataset with $C = P$ cells, each an exact copy of a reference atlas position. This allows us to treat the simulated scRNA-seq data as having a known spatial origin, against which we can compare the output of each method. We define a system-adaptive, comprehensive performance score, consisting of three penalty terms: accuracy, which determines whether or not the known spatial origin was given a high likelihood of origin; precision, which determines whether or not other locations were given low likelihoods of origin; and robustness, which determines how sensitive a mapping method is to random noise in the input data. Each penalty term is represented by a number in $[0, 1]$, with 0 being no penalty and 1 being a worst-case scenario. The

performance score is defined as $E = \frac{1}{P} \sum_{i=1}^P E_i$, where

$$E_i = 1 - \frac{1}{3} \left(\underbrace{1 - S_{i,i}}_{\text{Accuracy}} + \underbrace{\left| \frac{1 - \sum_{j=1}^P S_{i,j}}{P - 1} \right|}_{\text{Precision}} + \underbrace{(1 - \sigma^*)^4}_{\text{Robustness}} \right),$$

$S_{i,j} = S(c_i, p_j)$ is the correspondence score of cell c_i to position p_j , and E_i is interpreted as the error in the mapping of cell c_i . The quantity σ^* in the robustness term is calculated by determining the accuracy and precision penalty terms with no Gaussian noise added to the input data, then calculating the same two penalties with various levels of Gaussian noise with standard deviation $\sigma \in [0, 1]$. The quantity σ^* is defined to be the level of Gaussian noise required to raise the mean of the accuracy and precision penalties by 0.1 from their values with no added noise, or $\sigma^* = 1$, whichever is smallest. The exponent of four in the robustness term was chosen empirically such that the robustness term does not dominate the performance score, keeping in mind that expression levels are normalized to $[0, 1]$ before calculating the correspondence scores, so e.g., $\sigma^* = 0.5$ means a method requires noise on the order of half of the expression levels to raise the precision and accuracy penalties by 0.1. The performance score has a range of $[0, 1]$, where a performance score of $E = 1$ represents an ideal mapping that maps a cell to its known location with high confidence, to all other locations with low confidence, and does so in a manner robust to noise. An illustration of each term in the performance score is shown in Figure 2.



This performance score is limited by the fact that it relies on ground truth knowledge of the spatial origin of a single cell/spot to determine the performance of a given mapping method. However, this ground truth knowledge is not available for actual scRNA-seq data. To more directly quantify the mapping performance on actual scRNA-seq datasets, we use a measure of predictive reproducibility, obtained from a k -fold cross validation scheme, in which we randomly split the common genes in the reference atlas and scRNA-seq data into k folds and calculate the correspondence score for each method using all but one fold. The correspondence scores are then used as coefficients in a weighted sum to predict the value of the dropped-out genes in each fold for each cell (scRNA-seq predictive reproducibility) or each spatial position (atlas predictive reproducibility) and determine the error in the predicted expression level. The predicted expression of gene k in cell c_i is computed as $\hat{c}_i^{(k)} = \sum_{j=1}^P S_{i,j}^{(k)} p_j^{(k)} / \sum_{j=1}^P S_{i,j}^{(k)}$ and the predicted expression of gene k in

position p_j is computed as $\hat{p}_j^{(k)} = \sum_{i=1}^C S_{i,j}^{(k)} c_i^{(k)} / \sum_{i=1}^C S_{i,j}^{(k)}$ where $S_{i,j}^{(k)}$ is the correspondence score between cell c_i and position p_j with genes in folds not containing gene k and $c_i^{(k)}$ and $p_j^{(k)}$ are the known expression values of gene k from the scRNA-seq and the spatial atlas data, respectively. To accommodate the sparsity of data, we compute the predictive reproducibility scores separately for cells or positions with zero expression values and with positive expression values. For example, we measure the predictive reproducibility for the task of reproducing gene k in scRNA-seq data on cells with zero expression using $R_{sc_zero}^{(k)} = 1 - \sum_{i \in I_{sc_zero}^{(k)}} |\hat{c}_i^{(k)} - c_i^{(k)}| / |I_{sc_zero}^{(k)}|$ where $I_{sc_zero}^{(k)} = \{i : c_i^{(k)} = 0\}$.

Taking the average over all common genes results in a single score R_{sc_zero} , and in the same manner, we define $R_{sc_nonzero}$, R_{atlas_zero} , and $R_{atlas_nonzero}$. When producing predictive reproducibility scores, we use the same k -fold split across all methods to ensure a fair comparison.

Comparisons of Multiple Methods Using Simulated scRNA-seq Data

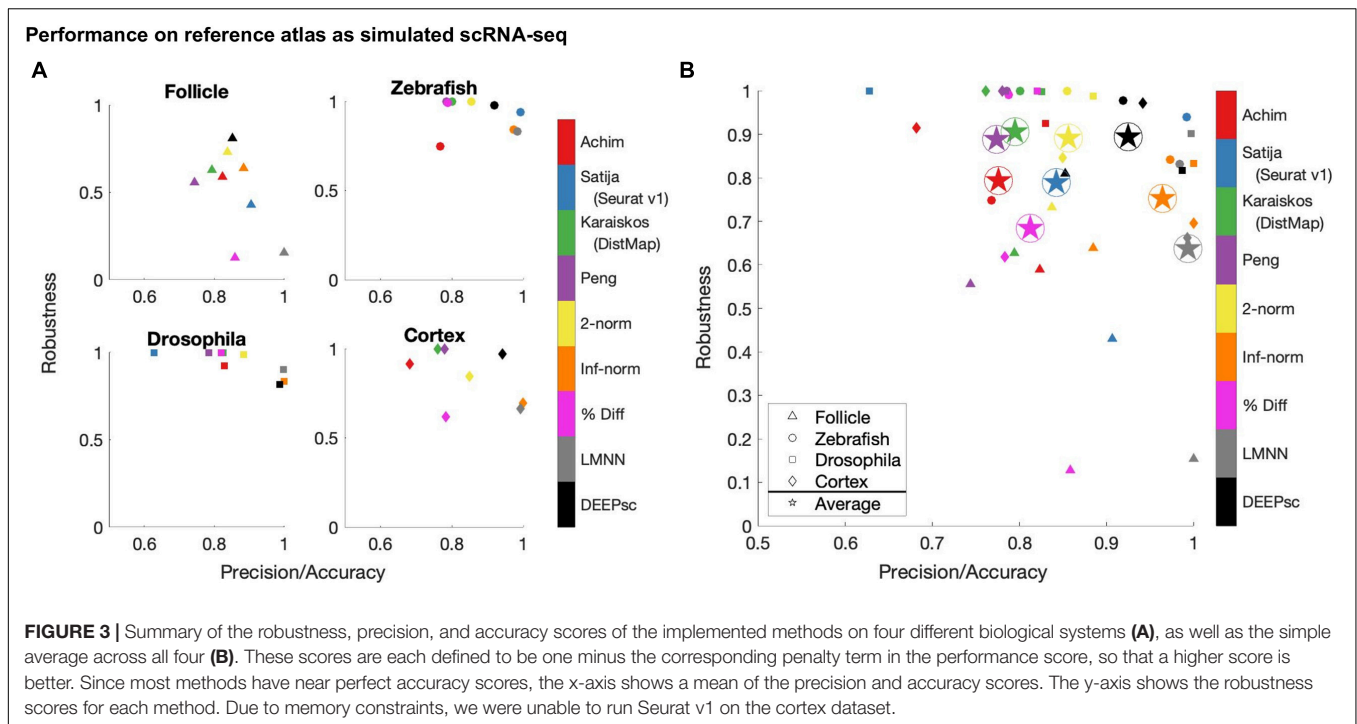
Using the performance score, we benchmarked the methods developed by Achim et al. (2015) and Satija et al. (2015) (Seurat v1), Karaiskos et al. (2017) (DistMap), and Peng et al. (2016) together with our DEEPsc method and applied them to four different biological systems: the zebrafish embryo (Satija et al., 2015), the *Drosophila* embryo (Karaiskos et al., 2017), the murine hair follicle (Joost et al., 2016), and the murine frontal cortex, downloaded from the 10x Genomics Spatial Gene Expression Datasets. The reference atlas for the zebrafish embryo contains the binarized expression of 47 genes on 64 spatial bins that assemble half of the hemisphere of the 6hpf embryo (Satija et al., 2015). The *Drosophila* embryo reference atlas contains 84 genes on 3,039 spatial positions (Karaiskos et al., 2017). The spatial reference atlas generated with the Visium technology (Ståhl et al., 2016) for the murine frontal cortex contains 32,285 genes on 961 spatial positions (a subset presenting the frontal

cortex from the original data), from which we kept 2755 genes from the 3,000 most variable genes in spatial data that are also present in scRNA-seq data. Segmenting a standard diagram of the follicle into 233 spatial positions and using FISH imaging of eight genes identified as spatially localized (Joost et al., 2016), we manually defined a continuous reference atlas for the follicle (section “Materials and Methods”). For mapping methods requiring a binary reference atlas, we defined a cutoff expression of 0.2 to be considered on in this follicle reference atlas of follicle. We further implemented several baseline methods for benchmark comparisons, including several methods using predefined metrics where the correspondence score S is defined to be the 2-norm, infinity norm, or mean percent difference in the space of common genes between the input cell and spatial position. We also implemented a large margin nearest neighbor (LMNN) method that learns a linear metric (section “Materials and Methods”). **Figure 3** shows a scatter plot of the penalty terms constituting the performance score of each implemented method on each of the four biological systems, as well as the average for each method across all four systems. **Table 1** includes the numerical values for each penalty term, as well as the calculated performance score for each method. **Figure 4** includes example heatmaps of simulated cells for each of the biological systems. The penalty terms for the individual locations are shown in **Figure 5**.

The majority of methods were able to project the simulated scRNA-seq cells to their known spatial origins with high accuracy. Specifically, Seurat v1 and DistMap achieve high performance scores in the zebrafish embryo and *Drosophila* embryo datasets that they were originally applied to, respectively. Designed to be a system-adaptive method, DEEPsc has the best average performance score across the four datasets (**Table 1**). Moreover, while some methods are stronger in terms robustness or precision, DEEPsc attains a balance between robustness and precision (**Figure 3**). This balance is especially important when investigating the impact of cellular spatial neighborhood on cell fate acquisition. It is desired to narrow down the inferred spatial neighborhood (precision) and at the same time reduce the sensitivity to noise (robustness). The high precision and robustness of DEEPsc is consistently observed across all locations in the dataset (**Figure 5**). Finally, it is worth mentioning that DEEPsc has a significantly higher robustness in the follicle dataset which has the smallest number of genes and is the noisiest among the four datasets.

Applications to Real scRNA-seq Datasets

We now map actual scRNA-seq data for each system and calculate the predictive reproducibility for each method (**Table 2** and **Figure 6**). For the follicle, the scRNA-seq data contains 1,422 cells with 26,024 genes measured containing the eight genes in the spatial atlas (Joost et al., 2016). For the *Drosophila* embryo, we used the scRNA-seq dataset with 1,297 cells and 8,924 genes among which all the 84 spatial genes are present (Karaiskos et al., 2017). For the Zebrafish embryo, there are 1,152 cells and 11,978 genes in the scRNA-seq dataset with all the 47 spatial genes



included (Satija et al., 2015). For the murine frontal cortex, we used the scRNA-seq dataset provided by the Allen Institute (Tasic et al., 2016), generated with SMART-Seq2, which contains 14,249 cells and 34,617 genes, from which a set of 2,755 genes were found to be present in the top 3,000 highly variable genes in spatial atlas. These four datasets cover different situations. The follicle data contains a moderate number of locations, and the cells form well-defined layered structures such that there could be long and thin spatial regions that contain the same cells. The zebrafish embryo spatial data has a suboptimal resolution such that each spatial location consists of multiple cells. This data helps to evaluate the methods in treating coarse spatial atlases. The *Drosophila* embryo data contains rich spatial characteristics. There is a well-defined global ventral-dorsal/anterior-posterior coordinate system. Locally, there is also a striped pattern in the lateral side of the embryo. The frontal cortex data examines spatial gene expression at the transcriptomics level, and functions as a demonstration that DEEPsc is able to maintain a high performance on high-dimensional datasets.

For the baseline models, we linearly normalized each gene in the log-normalized scRNA-seq dataset onto the interval [0, 1]. Continuous spatial atlases with expression values in the [0, 1] range were used for the follicle, *Drosophila* embryo, and murine frontal cortex systems, the latter two having been linearly normalized to [0, 1] in the same fashion as the scRNA-seq data. Since a continuous spatial atlas for Zebrafish embryo is lacking, we applied a spatial convolution to the binary atlas and added a small amount of Gaussian noise to simulate a continuous atlas. The 2-norm, Inf-norm, percent difference, and LMNN baseline models are then applied to the vectors of the commonly expressed genes in the spatial atlas and scRNA-seq data. For

DEEPsc, we first applied a PCA reduction to the spatial atlas, and then applied the same linear transformation to the normalized expression values of the common genes in the scRNA-seq data. The feature vectors for the locations in the spatial atlas and the cells in the scRNA-seq data in the PCA space were then fed to the neural network. For the four existing methods, we followed the procedure as described in the associated original publications, scaling the resulting correspondence scores to [0, 1] for direct comparison with baseline methods. For all the methods, we compute the predictive reproducibility by iterating over all common genes, attempting to reconstruct the expression of one gene using the k -fold cross validation scheme described in the previous section. We used $k = 4$ for the follicle and *Drosophila* embryo dataset, and $k = 5$ for the zebrafish embryo and cortex dataset.

DEEPsc has a comparable accuracy compared to other methods, and it also has a consistent performance across different systems (Table 2 and Figure 6). This consistent performance further demonstrates the system-adaptive advantage of DEEPsc and the benefit of using adaptive metrics over predefined ones. We also notice that similar to the simulated case, DEEPsc also achieves a balance between precision and robustness in the case of real scRNA-seq data. For example, while it exhibits high precision by mapping the example cell to a specific local spot in the Zebrafish embryo or a local strip in *Drosophila* embryo, it also robustly maps a cell to the entire outer bulge of the follicle instead of only part of it (Figure 7). The high precision ensures that we can resolve the heterogeneity in the spatial environment and further relate them to the heterogeneity in cell fates. The high robustness prevents the identification of false correlations. Overall, DEEPsc achieves a

TABLE 1 | Numerical values of each of the three constituent terms of the performance score, as determined from simulated scRNA-seq data for each biological system, as well as the average across all systems.

Method (Author)	Accuracy Term	Precision Term	Robustness Term	Performance Score
Follicle				
(Achim)	0.0043	0.3484	0.4116	0.7452
Seurat v1 (Satija)	0.0795	0.1076	0.5704	0.7475
DistMap (Karaiskos)	0.0043	0.4076	0.3723	0.7386
(Peng)	0.0000	0.5118	0.4439	0.6814
2-norm (baseline)	0.0000	0.3255	0.2686	0.8020
Inf-norm (baseline)	0.0005	0.2299	0.3613	0.8028
% difference (baseline)	0.0000	0.2829	0.8722	0.6150
LMNN (baseline)	0.0000	0.0002	0.8455	0.7181
DEEPsc (ours)	0.0272	0.2684	0.1904	0.8380
Zebrafish				
(Achim)	0.0000	0.4645	0.2516	0.7613
Seurat v1 (Satija)	0.0000	0.0156	0.0604	0.9747
DistMap (Karaiskos)	0.0000	0.3989	0.0000	0.8670
(Peng)	0.0000	0.4296	0.0000	0.8568
2-norm (baseline)	0.0000	0.2902	0.0003	0.9302
Inf-norm (baseline)	0.0000	0.0536	0.1588	0.9292
% difference (baseline)	0.0000	0.4249	0.0095	0.8552
LMNN (baseline)	0.0000	0.0315	0.1689	0.9332
DEEPsc (ours)	0.0339	0.1281	0.0230	0.9383
Drosophila				
(Achim)	0.0000	0.3407	0.0759	0.8611
Seurat v1 (Satija)	0.6605	0.0848	0.0000	0.7516
DistMap (Karaiskos)	0.0000	0.3496	0.0024	0.8827
(Peng)	0.0000	0.4313	0.0000	0.8562
2-norm (baseline)	0.0000	0.2310	0.0130	0.9186
Inf-norm (baseline)	0.0000	0.0006	0.1671	0.9441
% difference (baseline)	0.0000	0.3597	0.0013	0.8797
LMNN (baseline)	0.0000	0.0052	0.0987	0.9653
DEEPsc (ours)	0.0087	0.0179	0.1827	0.9303
Cortex				
(Achim)	0.0000	0.6357	0.0859	0.7594
Seurat v1 (Satija)	–	–	–	–
DistMap (Karaiskos)	0.0000	0.4778	0.0000	0.8407
(Peng)	0.0000	0.4400	0.0000	0.8533
2-norm (baseline)	0.0000	0.3008	0.1546	0.8482
Inf-norm (baseline)	0.0000	0.0006	0.3042	0.8984
% difference (baseline)	0.0000	0.4332	0.3817	0.7284
LMNN (baseline)	0.0000	0.0143	0.3376	0.8827
DEEPsc (ours)	0.0000	0.1167	0.0289	0.9515
Average				
(Achim)	0.0011	0.4473	0.2063	0.7818
Seurat v1 (Satija)	0.1850	0.0693	0.2103	0.8246
DistMap (Karaiskos)	0.0011	0.4085	0.0937	0.8323
(Peng)	0.0000	0.4532	0.1110	0.8119
2-norm (baseline)	0.0000	0.2869	0.1091	0.8748
Inf-norm (baseline)	0.0001	0.0712	0.2479	0.8936
% difference (baseline)	0.0000	0.3752	0.3162	0.7696
LMNN (baseline)	0.0000	0.0128	0.3627	0.8748
DEEPsc (ours)	0.0175	0.1328	0.1063	0.9145

For each term, a value closer to zero signifies lower error. For the performance score, a value closer to one indicates a better performing method. The best method for each term is bolded for each system.

high predictive reproducibility across all cells in the scRNA-seq dataset.

Comparison of Dimensionality Reduction Methods

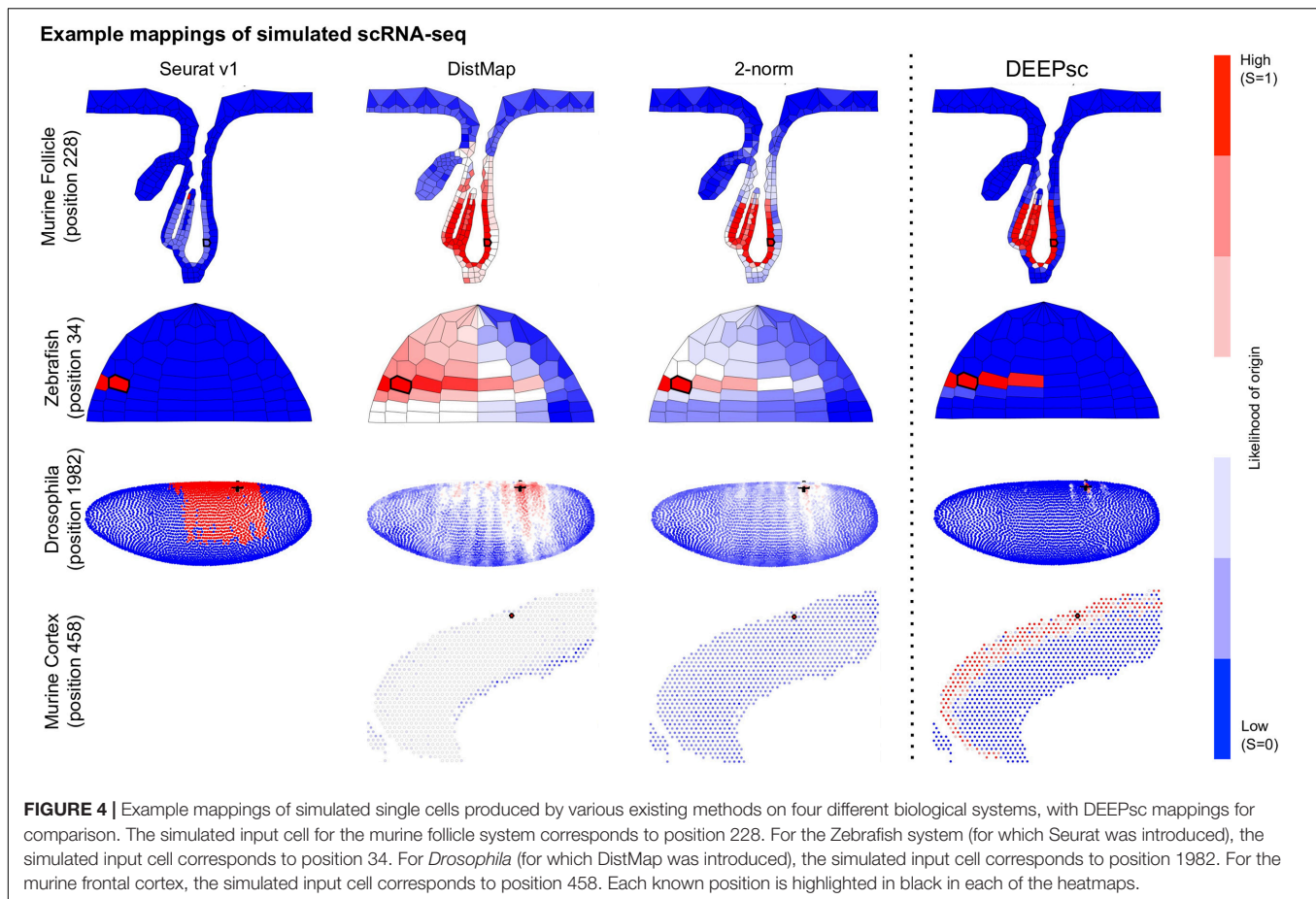
Dimension reduction is a crucial initial step of DEEPsc. A dimension reduction method that can be trained on one dataset and deterministically applied to another is needed due to the separated training and predicting steps. Here, we explore two different representative dimension reduction methods in the linear and nonlinear categories, PCA and Uniform Manifold Approximation and Projection (UMAP; McInnes et al., 2018). To compare these two methods, we trained several networks with varying amounts of added noise on the reference atlases of the four studied biological systems (Figure 8). We compared PCA (8 principal components), UMAP30 ($n_components = 8$, $n_neighbors = 30$), and UMAP5 ($n_components = 8$, $n_neighbors = 5$). While on the follicle system all three reduction methods performed virtually identically, on all three other systems PCA outperformed the other reduction methods by achieving a higher robustness score while maintaining similar accuracy.

DISCUSSION

We have developed the DEEPsc framework, which trains a deep neural network using the known expression levels of a small subset of genes in a spatial context, then imputes that spatial information onto a non-spatial scRNA-seq dataset. Instead of using a predefined metric, DEEPsc finds a metric adaptive to data. This framework is system-adaptive and designed to be robust to noise. DEEPsc consistently performs at or above the level of several existing methods across all four biological systems studied herein, including systems for which existing methods were originally developed (Figure 3 and Tables 1, 2), based on our comprehensive performance measure. While DEEPsc achieves comparable accuracy and precision to other methods, it is significantly more robust to noise.

The source of DEEPsc's ability to perform well across multiple biological systems is likely the generality of its neural network architecture and the multiple checks for robustness employed during training on the reference atlas. The various parameters for training a DEEPsc network, though chosen empirically, appear to translate to multiple systems effectively, so we expect DEEPsc to continue to perform well across more biological systems in future study.

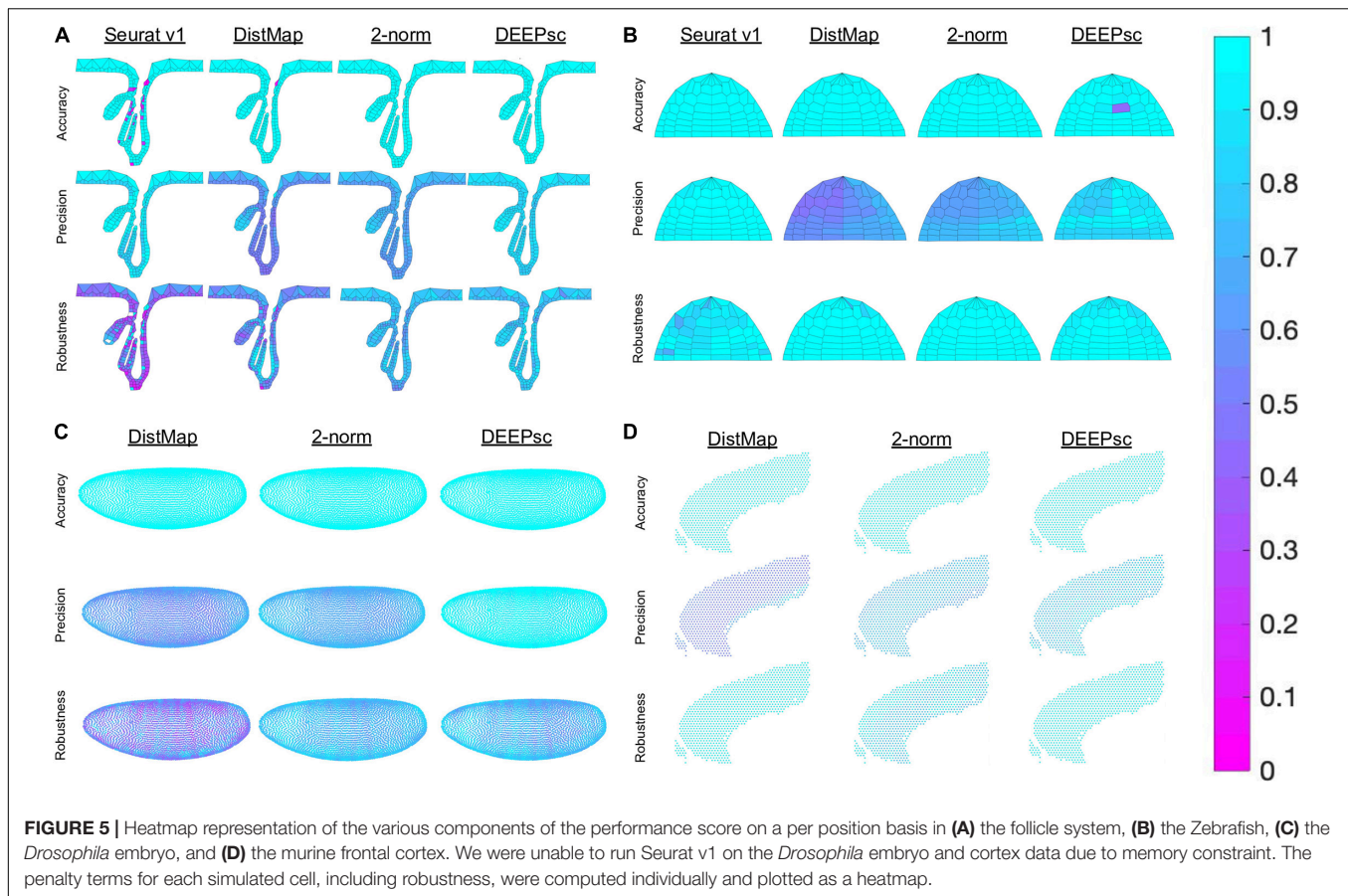
One notable weakness of DEEPsc is the significant amount of training time required to produce a final mapping. While most existing reference atlas methods simply involve a deterministic calculation to produce a mapping, DEEPsc requires an initial training, and the training time depends on the number of locations in the spatial atlas. The training process of DEEPsc can be effectively accelerated by iterating over a subset of possible location pairs. Due to the dimension reduction step, DEEPsc can still be trained efficiently on datasets with large amount of genes, for example, the spatial transcriptomics data on the



murine frontal cortex. Though the predefined metrics including the 2-norm and inf-norm perform well in terms of accuracy and precision, they are less robust to noise. This is further the case for LMNN as it tries to amplify any small variations. This drawback in robustness is mitigated by DEEPsc by controlling the balance between precision and robustness.

Learning a metric from high-dimensional datasets can be generally useful for analysis and integration of omics datasets. A future research interest is to decrease training time in such framework by developing a better method for reducing the size of the training set to a small, targeted fraction of relevant examples, particularly for very large atlases such as those derived from spatial transcriptomics assays. Since the size of the training set can increase quadratically with the number of positions in the atlas, it is beneficial to develop a more efficient training pipeline. We have developed a method of sparsifying the training set (section “Materials and Methods”), so that its size only increases linearly with the number of positions in the atlas, though further improvement may be warranted. The largest atlas studied here was that of *Drosophila* ($P = 3039$), the training of which took several hours even with the sparsified training set. Typical numbers of distinct spatial locations in a spatial transcriptomics dataset can be orders of magnitude larger.

DEEPsc aside, the performance score we have created can serve as a comprehensive measure of mapping performance for future work. The performance score is able to be calculated for any mapping method that assigns a likelihood of origin from each spatial location, particularly within the reference atlas framework. It is not dependent on any specific system or mapping method, and the individual terms which constitute it allow for a detailed analysis and comparison of various methods. Potential improvements might include incorporating some amount of spatial awareness into the calculation. Currently each spatial position is treated as completely independent from every other spatial position, so the precision term, for example, can yield unintuitive results if a method maps a cell, for example, with high probability to two positions on opposite sides of a system and low probability everywhere else, compared to a different method mapping the same cell with high probability to five positions in a tightly clustered, spatially compact region of the system. If, for example, the various correspondence scores for each position with high probability were weighted by their physical distance from other cells with high probability, this term might more accurately reflect the intuitive idea of precision. Other improvements might include simplifying the calculation of the robustness term to require fewer intensive calculations.



CONCLUSION

DEEPsc achieves an accuracy comparable to several existing methods while attaining improved precision and robustness. It also has a more consistent performance across the four different biological systems tested thanks to the system-adaptive design. As spatially resolved gene expression data becomes more readily available, our method will serve as a useful tool to infer spatial origins from non-spatial scRNA-seq data.

Additionally, our comprehensive performance score and the collection of reproductions of previously developed methods in a single software framework will serve as useful tools for future comparisons of spatial mapping methods. This systematic approach to imputing spatial information to scRNA-seq data is crucial to studying the spatial impact on cell fate dynamics.

MATERIALS AND METHODS

Data Preparation for DEEPsc

Given a matrix of scRNA-seq read counts where each row is a different gene and each column is a different cell, and a matrix representing a spatial reference atlas where each row is a different gene and each column is a different spatial position, we first select common genes by eliminating rows in each corresponding to genes not in the other matrix. Once we have eliminated genes not

in common, we are left with a number of cells (C) $\times$ number of genes (G) matrix for the scRNA-seq data and a number of positions (P) $\times$ number of genes (G) matrix for the spatial reference atlas.

We then apply dimensionality reduction to the atlas in the form of a PCA projection, selecting a user-configurable number of principal components to serve as feature vectors. We find in our analysis that keeping the top eight principal components yields satisfactory results. The same PCA coefficients are used to project the scRNA-seq matrix into these principal components. After projection, both matrices are normalized by dividing by the largest element in each, so that the elements are all in $[0, 1]$.

For the comparisons in section “Comparison of Dimensionality Reduction Methods,” we use the UMAP implementation by Meehan et al. (2021), found on the MATLAB Central File Exchange at <https://www.mathworks.com/matlabcentral/fileexchange/71902>. Specifically, we ran the `run_umap()` function on the spatial reference atlas with $n_dimensions = 8$ and $n_neighbors = 30$ or $n_neighbors = 5$ for UMAP30 and UMAP5, respectively.

Training a DEEPsc Network

To train the DEEPsc network, we use the spatial position feature vectors themselves as simulated scRNA-seq data. The training data is a set of P^2 vectors of length $2N$, where N is the reduced

TABLE 2 | Predictive reproducibility of each method for real scRNA-seq data.

Method (Author)	Follicle	Zebrafish	Drosophila	Cortex	Average
<i>R_{sc_zero}</i>					
(Achim)	0.8772	0.5537	0.7798	0.8019	0.7531
Seurat v1 (Satija)	0.8335	0.6842	–	–	0.7589
DistMap (Karaiskos)	0.8404	0.6641	0.7850	0.8055	0.7738
(Peng)	0.8219	0.6375	0.7859	0.8092	0.7636
Two-norm (baseline)	0.8017	0.6973	0.7874	0.8114	0.7745
Inf-norm (baseline)	0.8641	0.6180	0.7807	0.8141	0.7692
% difference (baseline)	0.8357	0.5657	0.7790	0.8079	0.7471
LMNN (baseline)	0.8254	0.6795	0.7917	0.8120	0.7772
DEEPsc (ours)	0.8344	0.7335	0.7961	0.8165	0.7951
<i>R_{sc_nonzero}</i>					
(Achim)	0.7495	0.7698	0.8126	0.6693	0.7503
Seurat v1 (Satija)	0.7640	0.6975	–	–	0.7308
DistMap (Karaiskos)	0.7705	0.7619	0.8103	0.6685	0.7528
(Peng)	0.7801	0.7663	0.8114	0.6680	0.7565
Two-norm (baseline)	0.7891	0.7386	0.8083	0.6667	0.7507
Inf-norm (baseline)	0.7496	0.7636	0.8128	0.6695	0.7489
% difference (baseline)	0.7740	0.7721	0.8115	0.6690	0.7567
LMNN (baseline)	0.7730	0.7477	0.8117	0.6643	0.7492
DEEPsc (ours)	0.7352	0.7026	0.8080	0.6691	0.7287
<i>R_{atlas_zero}</i>					
(Achim)	0.7680	0.9042	0.9264	0.8360	0.8587
Seurat v1 (Satija)	0.7681	0.9088	–	–	0.8385
DistMap (Karaiskos)	0.7674	0.9005	0.9259	0.8374	0.8578
(Peng)	0.7707	0.9006	0.9267	0.8406	0.8597
Two-norm (baseline)	0.7681	0.9003	0.9278	0.8411	0.8593
Inf-norm (baseline)	0.7623	0.9050	0.9259	0.8343	0.8569
% difference (baseline)	0.7714	0.9035	0.9261	0.8438	0.8612
LMNN (baseline)	0.7677	0.8937	0.9289	0.8359	0.8566
DEEPsc (ours)	0.7881	0.9148	0.9257	0.8415	0.8675
<i>R_{atlas_nonzero}</i>					
(Achim)	0.7598	0.6658	0.8523	0.5124	0.6976
Seurat v1 (Satija)	0.7570	0.6776	–	–	0.7173
DistMap (Karaiskos)	0.7584	0.6709	0.8527	0.5127	0.6987
(Peng)	0.7570	0.6682	0.8530	0.5135	0.6979
Two-norm (baseline)	0.7582	0.6755	0.8530	0.5135	0.7001
Inf-norm (baseline)	0.7583	0.6745	0.8534	0.5130	0.6998
% difference (baseline)	0.7573	0.6669	0.8524	0.5134	0.6975
LMNN (baseline)	0.7573	0.6764	0.8564	0.5129	0.7008
DEEPsc (ours)	0.7724	0.7079	0.8527	0.5125	0.7114

A value closer to one signifies higher predictive reproducibility. A missing entry signifies that we were not able to run the relevant method on the given dataset. The best method for each term is bolded for each system.

dimensionality of the reference atlas. The first N components correspond to a feature vector of one position in the reference atlas (functioning as a simulated cell) and the last N components correspond to some other position in the reference atlas. Each simulated cell is compared pairwise with every position in the spatial reference atlas; if the simulated cell is an exact match to a given position, the target output is chosen to be 1 (a high likelihood of origin), and if the simulated cell and chosen position are not an exact match, the target output is chosen to be 0 (a low likelihood of origin).

The DEEPsc architecture is an artificial neural network with $2N$ inputs, two fully connected hidden layers with N nodes each and a single node in the output layer. Sigmoid activation functions are attached to each node, including the output node, so that the resulting output is in $[0, 1]$ and can be interpreted as a likelihood that the input cell originated from the input spatial position. To preserve robustness and avoid overfitting the training data, a layer of Gaussian noise is added to the simulated cells so that the network is pushed to learn complex nonlinear relationships among the spatial positions in the reference atlas rather than simply activate when an exact match is encountered. This Gaussian noise layer allows the user to configure the standard deviation of the added noise, as well as to configure the probability that any noise will be added in a given training epoch. We find empirically that a noise level of about 0.10 and a probability of 0.5 yield reasonable robustness to noise, though this may vary from system to system.

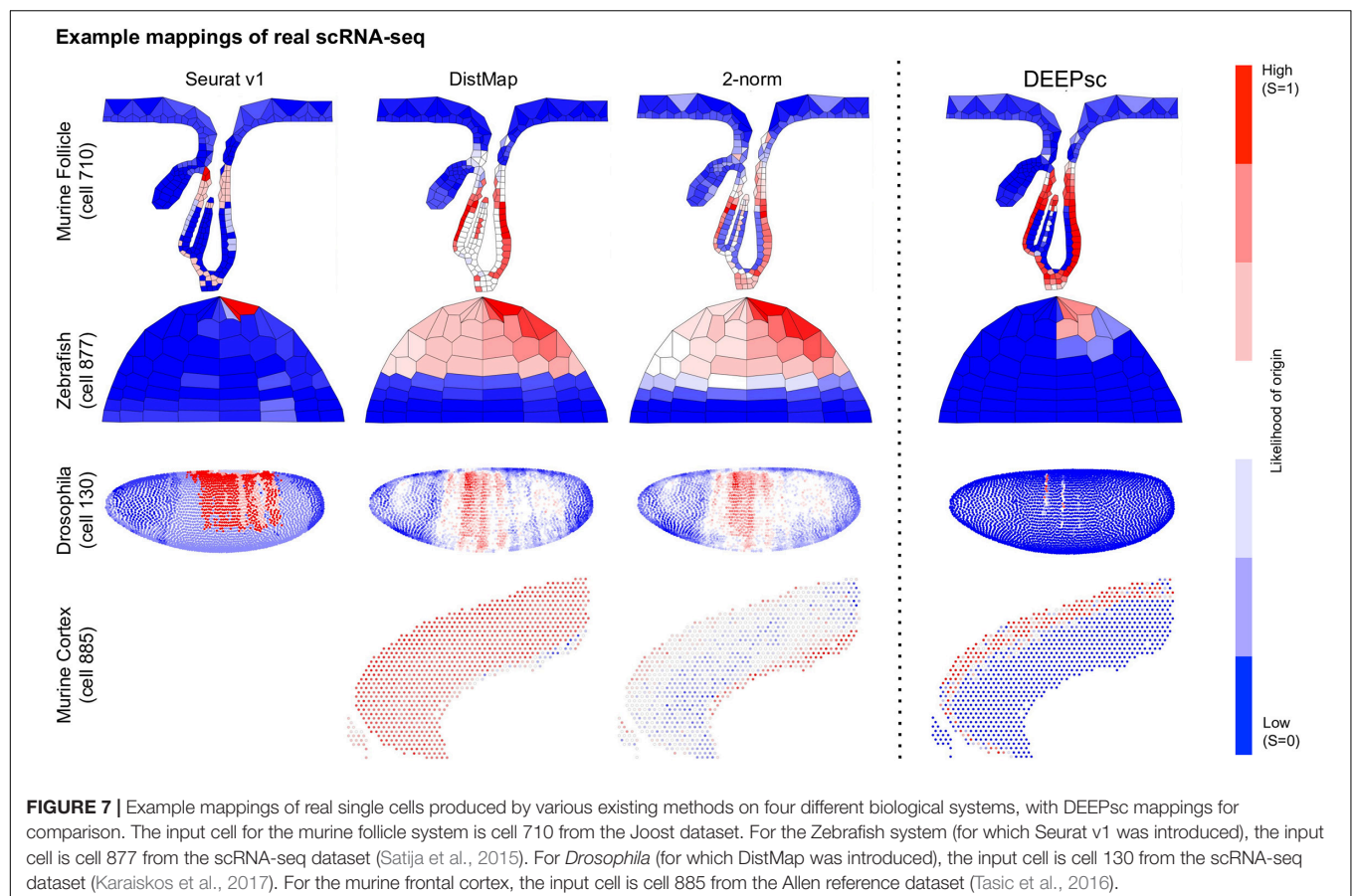
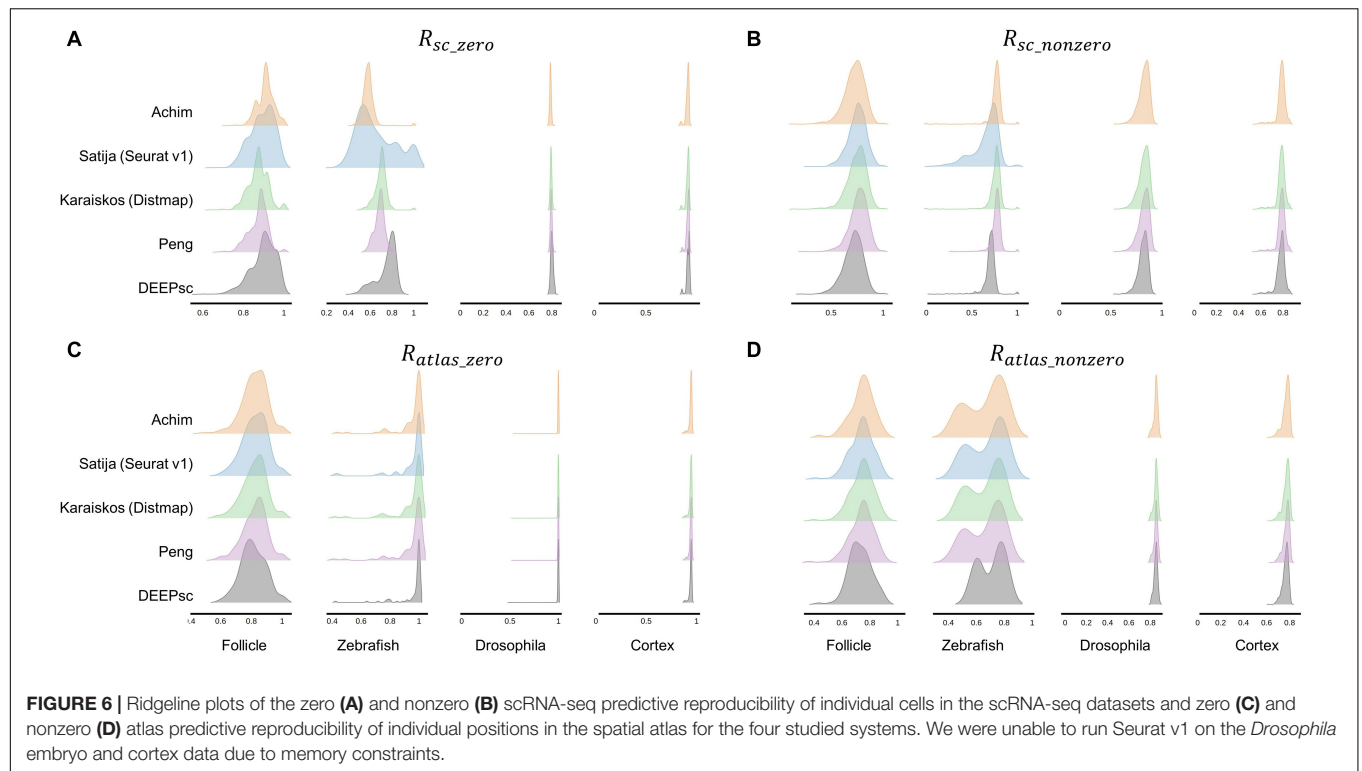
Since the training data will naturally consist of many more non-matches than matches, and thus the target data will contain many more zeros than ones, we use a novel custom objective function,

$$L(Y, T) = \sum_{i=1}^P (y_i - t_i)^2 \frac{1}{1.001 - t_i}$$

where y_i is the network's predicted output and t_i is the target output ($t_i = 1$ if exact match and $t_i = 0$ if not), to more heavily penalize the network when it gives a false negative (low likelihood when it should be high) than when it gives a false positive (high likelihood when it should be low). This acts to counteract the tendency of the network to “learn” to simply return 0 for every single input and “ignore” any comparably rare training data with $t_i = 1$.

To further account for the sparsity of exact matches in the training set, we split it into a test and validation set, the former consisting of a configurable fraction of the inputs corresponding to exact matches as well as a configurable multiple of the inputs corresponding to non-matches. If $trainFrac = 0.9$ and $trainingMultiple = 99$, for example, 90% of the exact matches will be added to the training set and 99x more non-matches will be added, so that the exact matches make up 1% of the training set. The rest of the inputs are reserved for the (generally much larger) validation set. This is beneficial in reducing training time because it allows us to train with a much smaller fraction of the P^2 input vectors, giving preference to the exact matches. Indeed, this reduces the size of the actual training set to scale linearly with the size of the atlas rather than quadratically.

Training is performed in MATLAB using the *trainNetwork()* function in the Deep Learning Toolbox (The Mathworks, Inc, 2019a), for which we implemented the above-described custom network layers. Since the input data is already normalized in preprocessing, we disable the default normalization of *trainNetwork()*. We use the default Glorot (Xavier and Yoshua, 2010) initialization of weights and biases in the fully connected layers. We then train each network for a maximum of 50,000 epochs of standard gradient descent with a learning rate of $\eta = 0.01$, shuffle the order of the data each epoch, and use



Comparison of dimensionality reduction techniques

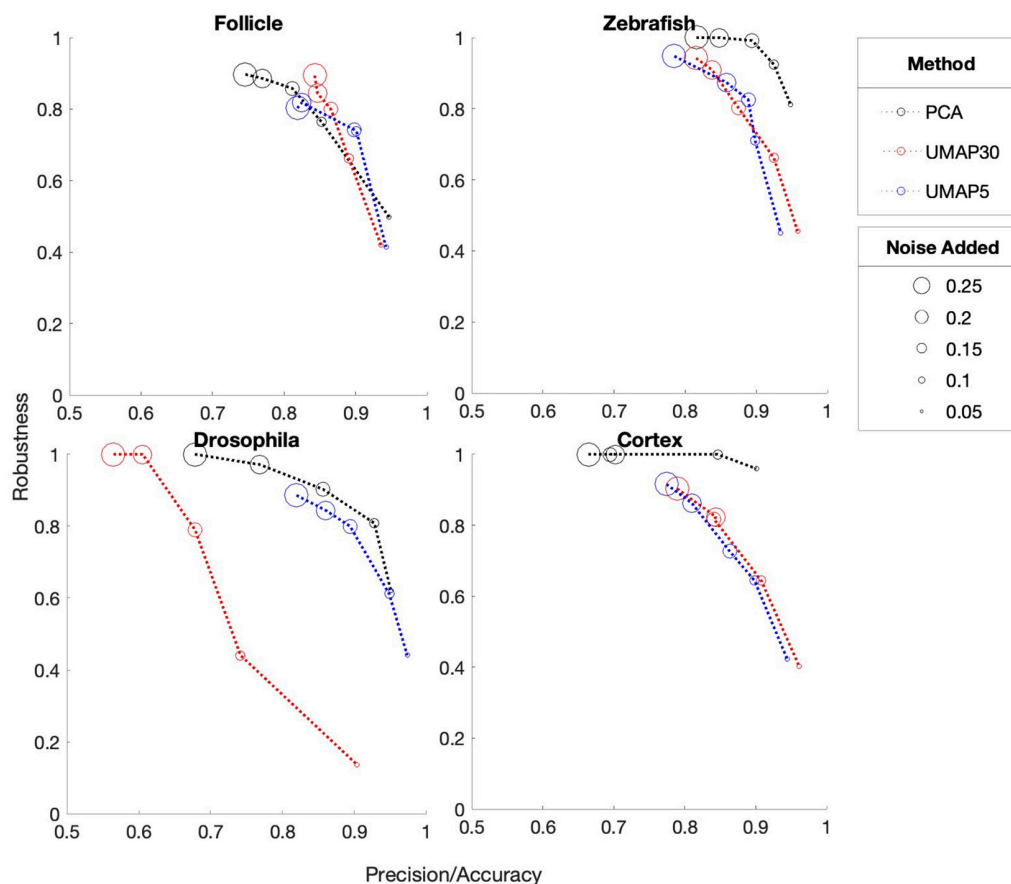


FIGURE 8 | A comparison of the performance of DEEPsc networks using different dimensionality reduction methods on each of the biological systems for various levels of added noise during training. We compare principal component analysis (PCA) to Uniform Manifold Approximation and Projection (UMAP) with $n_neighbors = 30$ (UMAP30) and $n_neighbors = 5$ (UMAP5). Each of these methods reduce the dimensionality of the initial dataset to $n_dimensions = 8$. These scores are each defined to be one minus the corresponding penalty term in the performance score, so that a higher score is better. Since most methods have near perfect accuracy scores, the x-axis shows a mean of the precision and accuracy scores. The y-axis shows the robustness scores for each method.

the ADAM optimization method (Kingma and Ba, 2014) with the default parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 10^{-8}$. In addition to the custom objective function layer we describe above, *trainNetwork()* by default adds an L^2 -regularization term to the loss with a regularization factor of $\lambda = 0.0001$. We monitor the RMSE of the validation set throughout training and manually stop training if it is no longer improving before the maximum number of epochs has been reached. The *trainNetwork()* function also allows for parallel computation via the Parallel Computing Toolbox (The Mathworks, Inc, 2019b), which is highly recommended but not strictly required for training.

Creating a Reference Atlas for the Murine Follicle

To create a spatial reference atlas for the murine follicle system, we patterned the spatial coordinates of each position in the atlas off of a standard diagram of a mouse follicle found in Figure 1 of Joost et al. (2016). We constructed a Voronoi diagram around

each of the cell centers and made manual adjustments to the vertices as we saw fit aesthetically. We then selected the eight genes in the atlas from the systematic staining catalog made available by Joost. We chose the genes based on a combination of high image quality and spatial diversity. Gene expression levels in $[0, 1]$ were chosen manually to best represent the images, though to eliminate any implicit bias we also added a small level of Gaussian noise to the atlas. For all methods requiring a binary atlas, we chose a cutoff of 0.2 to represent “on” expression in this atlas.

Large Margin Nearest Neighbor Baseline

To implement a LMNN baseline for benchmarking comparison, we used code from the MATLAB Toolbox for Dimensionality Reduction found at <https://lvdmaaten.github.io/drtoolbox/> and modified it for our uses. Specifically, we used the *lmnn()* function in the “techniques” subfolder, and modified the code to set

$\mu = 1$, i.e., to remove the “pull” term, as well as setting the number of targets to 1 (the point itself) and treating all other points as imposters. Further, we modified the slack variables to enforce a minimum separation of $\sqrt{D}$, where D is the dimensionality of the space ($D = G$ for our applications). For the numerical experiments of the LMNN method with the cortex dataset, a PCA dimension reduction (50 PCs) was performed before applying LMNN to accommodate the large number of genes.

DATA AVAILABILITY STATEMENT

The original data used in this paper can be accessed through the following links: (1) zebrafish embryo spatial data: downloaded from (https://www.dropbox.com/s/ev78jelev0jgu5s/seurat_files_zfin.zip?dl=1) (Satija et al., 2015); (2) zebrafish embryo scRNA-seq data: GEO accession codes: GSE66688 (Satija et al., 2015); (3) Drosophila embryo spatial and scRNA-seq data: accessible at the Dream Single cell Transcriptomics Challenge through Synapse ID (syn15665609) (Karaiskos et al., 2017); (4) mouse frontal cortex spatial data: downloaded from 10x Genomics Spatial Gene Expression Datasets (https://support.10xgenomics.com/spatial-gene-expression/datasets/1.1.0/V1_Mouse_Brain_Sagittal_Anterior); (5) mouse frontal cortex scRNA-seq

data: downloaded from (https://www.dropbox.com/s/cuowvm4vrf65pvq/allen_cortex.rds?dl=1) (Tasic et al., 2016); (6) follicle scRNA-seq data and spatial imaging data from which the reference atlas was derived: downloaded from the supplementary of the associated publication (Joost et al., 2016). All software developed for the purposes of this comparison are made freely available at: <https://github.com/fmaseda/DEEPsc>.

AUTHOR CONTRIBUTIONS

FM carried out computer implementation and data analysis. ZC and QN supervised the project. FM and ZC wrote the original manuscript. All authors conceived and designed the work, interpreted the simulation results, and contributed to the writing of the final manuscript.

FUNDING

The work was partially supported by the NIH grants U01AR073159 and P30AR075047, NSF grants DMS1763272 and MCB2028424, and a grant from the Simons Foundation (594598, QN).

REFERENCES

- Achim, K., Pettit, J.-B., Saraiva, L. R., Gavriouchkina, D., Larsson, T., Arendt, D., et al. (2015). High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat. Biotechnol.* 33, 503–509. doi: 10.1038/nbt.3209
- Andersson, A., Bergensträhle, J., Asp, M., Bergensträhle, L., Jurek, A., Navarro, J. F., et al. (2020). Single-cell and spatial transcriptomics enables probabilistic inference of cell type topography. *Commun. Biol.* 3:565. doi: 10.1038/s42003-020-01247-y
- Boufe, K., Seth, S., and Batada, N. N. (2020). scID uses discriminant analysis to identify transcriptionally equivalent cell types across single-Cell RNA-seq data with batch effect. *iScience* 23:100914. doi: 10.1016/j.isci.2020.100914
- Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* 36, 411–420. doi: 10.1038/nbt.4096
- Cang, Z., and Nie, Q. (2020). Inferring spatial and signaling relationships between cells from single cell transcriptomic data. *Nat. Commun.* 11:2084. doi: 10.1038/s41467-020-15968-5
- Chicco, D. (2020). Siamese neural networks: an overview. *Methods Mol. Biol.* 2190, 73–94. doi: 10.1007/978-1-0716-0826-5_3
- Codeluppi, S., Borm, L. E., Zeisel, A., La Manno, G., Van Lunteren, J. A., Svensson, C. I., et al. (2018). Spatial organization of the somatosensory cortex revealed by osmFISH. *Nat. Methods* 15, 932–935. doi: 10.1038/s41592-018-0175-z
- Dries, R., Zhu, Q., Eng, C.-H. L., Sarkar, A., Bao, F., George, R. E., et al. (2019). Giotto, a pipeline for integrative analysis and visualization of single-cell spatial transcriptomic data. *bioRxiv* [preprint]. doi: 10.1101/701680
- Eng, C.-H. L., Lawson, M., Zhu, Q., Dries, R., Koulana, N., Takei, Y., et al. (2019). Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* 568, 235–239. doi: 10.1038/s41586-019-1049-y
- Fowlkes, C. C., Hendriks, C. L. L., Keränen, S. V. E., Weber, G. H., Rübel, O., Huang, M.-Y., et al. (2008). A quantitative spatiotemporal Atlas of gene expression in the *Drosophila* blastoderm. *Cell* 133, 364–374. doi: 10.1016/j.cell.2008.01.053
- Guo, G., Huss, M., Tong, G. Q., Wang, C., Li Sun, L., Clarke, N. D., et al. (2010). Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Dev. Cell* 18, 675–685. doi: 10.1016/j.devcel.2010.02.012
- Halpern, K. B., Shenhav, R., Matcovitch-Natan, O., Tóth, B., Lemze, D., Golan, M., et al. (2017). Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. *Nature* 542, 352–356. doi: 10.1038/nature21065
- Hie, B., Bryson, B., and Berger, B. (2019). Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. *Nat. Biotechnol.* 37, 685–691. doi: 10.1038/s41587-019-0113-3
- Hu, J., Li, X., Hu, G., Lyu, Y., Susztak, K., and Li, M. (2020). Iterative transfer learning with neural network for clustering and cell type classification in single-cell RNA-seq analysis. *bioRxiv* [preprint]. doi: 10.1038/s42256-020-00233-7
- Joost, S., Zeisel, A., Jacob, T., Sun, X., La Manno, G., Lönnerberg, P., et al. (2016). Single-Cell transcriptomics reveals that differentiation and spatial signatures shape epidermal and hair follicle heterogeneity. *Cell Syst.* 3, 221–237.e229. doi: 10.1016/j.cels.2016.08.010
- Karaiskos, N., Wahle, P., Alles, J., Boltengagen, A., Ayoub, S., Kipar, C., et al. (2017). The *Drosophila* embryo at single-cell transcriptome resolution. *Science* 358, 194–199. doi: 10.1126/science.aan3235
- Kaya, M., and Bilge, H. (2019). Deep metric learning: a survey. *Symmetry* 11:1066. doi: 10.3390/sym11091066
- Kingma, D. P., and Ba, J. (2014). Adam: a method for stochastic optimization. *arXiv* [preprint]:1412.6980. arXiv e-prints
- Kiselev, V. Y., Andrews, T. S., and Hemberg, M. (2019). Challenges in unsupervised clustering of single-cell RNA-seq data. *Nat. Rev. Genet.* 20, 273–282. doi: 10.1038/s41576-018-0088-9
- Kiselev, V. Y., Yiu, A., and Hemberg, M. (2018). scmap: projection of single-cell RNA-seq data across data sets. *Nat. Methods* 15, 359–362. doi: 10.1038/nmeth.4644
- Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., et al. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* 16, 1289–1296. doi: 10.1038/s41592-019-0619-0
- Kulis, B. (2013). Metric learning: a survey. *Found. Trends® Mach. Learn.* 5, 287–364. doi: 10.1561/22000000019
- Lieberman, Y., Rokach, L., and Shay, T. (2018). CaSTLe – Classification of single cells by transfer learning: harnessing the power of publicly available single

- cell RNA sequencing experiments to annotate new experiments. *PLoS One* 13:e0205499. doi: 10.1371/journal.pone.0205499
- Lopez, R., Nazaret, A., Langevin, M., Samaran, J., Regier, J., Jordan, M. I., et al. (2019). A joint model of unpaired data from scRNA-seq and spatial transcriptomics for imputing missing gene expression measurements. *arXiv [preprint]* arXiv:1905.02269.
- Lopez, R., Regier, J., Cole, M. B., Jordan, M. I., and Yosef, N. (2018). Deep generative modeling for single-cell transcriptomics. *Nat. Methods* 15, 1053–1058. doi: 10.1038/s41592-018-0229-2
- Luecken, M. D., and Theis, F. J. (2019). Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol. Syst. Biol.* 15:e8746. doi: 10.15252/msb.2018.8746
- Ma, F., and Pellegrini, M. (2020). ACTINN: automated identification of cell types in single cell RNA sequencing. *Bioinformatics* 36, 533–538. doi: 10.1093/bioinformatics/btz592
- McInnes, L., Healy, J., and Melville, J. (2018). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*. Available online at: <https://ui.adsabs.harvard.edu/abs/2018arXiv180203426M> (accessed February 01, 2018).
- Meehan, C., Ebrahimian, J., Moore, W., and Meehan, S. (2021). *Uniform Manifold Approximation and Projection (UMAP)*. *MATLAB Central File Exchange*. Available online at: <https://www.mathworks.com/matlabcentral/fileexchange/71902>
- Moffitt, J. R., Bambach-Mukku, D., Eichhorn, S. W., Vaughn, E., Shekhar, K., Perez, J. D., et al. (2018). Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* 362:eaau5324. doi: 10.1126/science.aau5324
- Nitzan, M., Karaiskos, N., Friedman, N., and Rajewsky, N. (2019). Gene expression cartography. *Nature* 576, 132–137. doi: 10.1038/s41586-019-1773-3
- Pandey, S., Shekhar, K., Regev, A., and Schier, A. F. (2018). Comprehensive identification and spatial mapping of habenular neuronal types using single-Cell RNA-Seq. *Curr. Biol.* 28, 1052–1065.e1057. doi: 10.1016/j.cub.2018.02.040
- Peng, G., Suo, S., Chen, J., Chen, W., Liu, C., Yu, F., et al. (2016). Spatial transcriptome for the molecular annotation of lineage fates and cell identity in mid-gastrula mouse embryo. *Dev. Cell* 36, 681–697. doi: 10.1016/j.devcel.2016.02.020
- Puram, S. V., Tirosh, I., Parikh, A. S., Patel, A. P., Yizhak, K., Gillespie, S., et al. (2017). Single-Cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* 171, 1611–1624.e1624. doi: 10.1016/j.cell.2017.10.044
- Rodrigues, S. G., Stickels, R. R., Goeva, A., Martin, C. A., Murray, E., Vanderburg, C. R., et al. (2019). Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* 363, 1463–1467. doi: 10.1126/science.aaw1219
- Rosenberg, A. B., Roco, C. M., Muscat, R. A., Kuchina, A., Sample, P., Yao, Z., et al. (2018). Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* 360, 176–182. doi: 10.1126/science.aam.8999
- Satija, R., Farrell, J. A., Gennert, D., Schier, A. F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* 33, 495–502. doi: 10.1038/nbt.3192
- Shah, S., Lubeck, E., Zhou, W., and Cai, L. (2016). In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. *Neuron* 92, 342–357. doi: 10.1016/j.neuron.2016.10.001
- Sprague, J., Bayraktaroglu, L., Clements, D., Conlin, T., Fashena, D., Frazer, K., et al. (2006). The zebrafish information network: the zebrafish model organism database. *Nucleic Acids Res.* 34, D581–D585. doi: 10.1093/nar/gkj086
- Stahl, P. L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J. F., Magnusson, J., et al. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 353, 78–82. doi: 10.1126/science.aaf.2403
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., et al. (2019). Comprehensive integration of single-cell data. *Cell* 177, 1888–1902.e1821. doi: 10.1016/j.cell.2019.05.031
- Svensson, V., Vento-Tormo, R., and Teichmann, S. A. (2018). Exponential scaling of single-cell RNA-seq in the past decade. *Nat. Protoc.* 13, 599–604.
- Tan, Y., and Cahan, P. (2019). SingleCellNet: a computational tool to classify single cell RNA-Seq data across platforms and across species. *Cell Syst.* 9, 207–213.e202.
- Tasic, B., Menon, V., Nguyen, T. N., Kim, T. K., Jarsky, T., Yao, Z., et al. (2016). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosci.* 19, 335–346.
- The Mathworks, Inc (2019a). *MATLAB Deep Learning Toolbox Release 2019b*. Natick: The Mathworks, Inc.
- The Mathworks, Inc (2019b). *MATLAB Parallel Toolbox Release 2019b*. Natick: The Mathworks, Inc.
- Wagner, F., and Yanai, I. (2018). Moana: a robust and scalable cell type classification framework for single-cell RNA-Seq data. *bioRxiv [preprint]*. doi: 10.1101/456129
- Wang, S., Karikomi, M., Maclean, A. L., and Nie, Q. (2019). Cell lineage and communication network inference via optimization for single-cell transcriptomics. *Nucleic Acids Res.* 47:e66.
- Wang, X., Allen, W. E., Wright, M. A., Sylwestrak, E. L., Samusik, N., Vesuna, S., et al. (2018). Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* 361:eaat5691.
- Weinberger, K. Q., and Saul, L. K. (2009). Distance metric learning for large margin nearest neighbor classification. *J. Mach. Learn. Res.* 10, 207–244.
- Welch, J. D., Kozareva, V., Ferreira, A., Vanderburg, C., Martin, C., and Macosko, E. Z. (2019). Single-Cell multi-omic integration compares and contrasts features of brain cell identity. *Cell* 177, 1873–1887.e1817.
- Xavier, G., and Yoshua, B. (2010). “Understanding the difficulty of training deep feedforward neural networks: PMLR,” in *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics, JMLR Workshop and Conference Proceedings*, Vol. 9 (Sardinia: Chia Laguna Resort), 249–256.
- Yuan, G.-C., Cai, L., Elowitz, M., Enver, T., Fan, G., Guo, G., et al. (2017). Challenges and emerging directions in single-cell analysis. *Genome Biol.* 18:84.
- Zhu, Q., Shah, S., Dries, R., Cai, L., and Yuan, G.-C. (2018). Identification of spatially associated subpopulations by combining scRNAseq and sequential fluorescence in situ hybridization data. *Nat. Biotechnol.* 36, 1183–1190.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Maseda, Cang and Nie. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Exact Probability Landscapes of Stochastic Phenotype Switching in Feed-Forward Loops: Phase Diagrams of Multimodality

Anna Terebus^{1,2†}, Farid Manuchehrfar^{1†}, Youfang Cao^{1,3} and Jie Liang^{1*}

¹ Center for Bioinformatics and Quantitative Biology, Richard and Loan Hill Department of Bioengineering, University of Illinois at Chicago, Chicago, IL, United States, ² Constellation, Baltimore, MD, United States, ³ Merck & Co., Inc., Kenilworth, NJ, United States

OPEN ACCESS

Edited by:

Chunhe Li,
Fudan University, China

Reviewed by:

Yang Cao,
Virginia Tech, United States
Davit Potoyan,
Iowa State University, United States

*Correspondence:

Jie Liang
jliang@uic.edu

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Biology Archive,
a section of the journal
Frontiers in Genetics

Received: 23 December 2020

Accepted: 26 April 2021

Published: 08 July 2021

Citation:

Terebus A, Manuchehrfar F, Cao Y
and Liang J (2021) Exact Probability
Landscapes of Stochastic Phenotype
Switching in Feed-Forward Loops:
Phase Diagrams of Multimodality.
Front. Genet. 12:645640.
doi: 10.3389/fgene.2021.645640

Feed-forward loops (FFLs) are among the most ubiquitously found motifs of reaction networks in nature. However, little is known about their stochastic behavior and the variety of network phenotypes they can exhibit. In this study, we provide full characterizations of the properties of stochastic multimodality of FFLs, and how switching between different network phenotypes are controlled. We have computed the exact steady-state probability landscapes of all eight types of coherent and incoherent FFLs using the finite-butler Accurate Chemical Master Equation (ACME) algorithm, and quantified the exact topological features of their high-dimensional probability landscapes using persistent homology. Through analysis of the degree of multimodality for each of a set of 10,812 probability landscapes, where each landscape resides over 10^5 – 10^6 microstates, we have constructed comprehensive phase diagrams of all relevant behavior of FFL multimodality over broad ranges of input and regulation intensities, as well as different regimes of promoter binding dynamics. In addition, we have quantified the topological sensitivity of the multimodality of the landscapes to regulation intensities. Our results show that with slow binding and unbinding dynamics of transcription factor to promoter, FFLs exhibit strong stochastic behavior that is very different from what would be inferred from deterministic models. In addition, input intensity play major roles in the phenotypes of FFLs: At weak input intensity, FFL exhibit monomodality, but strong input intensity may result in up to 6 stable phenotypes. Furthermore, we found that gene duplication can enlarge stable regions of specific multimodalities and enrich the phenotypic diversity of FFL networks, providing means for cells toward better adaptation to changing environment. Our results are directly applicable to analysis of behavior of FFLs in biological processes such as stem cell differentiation and for design of synthetic networks when certain phenotypic behavior is desired.

Keywords: systems biology, feed forward loop, gene regulatory network, network motif, stochastic reaction network, persistent homology, ACME algorithm

1. INTRODUCTION

Cells with the same genetic make-ups can exhibit a variety of different behavior. They can also switch between these different phenotypes stochastically. This phenomenon has been observed in bacteria, yeast, and mammals such as neural cells (Acar et al., 2005; Choi et al., 2008; Guo and Li, 2009; Gupta et al., 2011). The ability to exhibit multiple phenotypes and switching between them is the foundation of cellular fate decision (Schultz et al., 2007; Cao et al., 2010; Ye et al., 2019), stem cell differentiation (Feng and Wang, 2012; Papatsenko et al., 2015; Zhang et al., 2019), and tumor formation (Huang et al., 2009; Shiraishi et al., 2010).

Cells exhibiting different phenotypes have different patterns of gene expression. Single-cell studies demonstrated that isogenic cells can exhibit different modes of gene expression (Shalek et al., 2013), indicating that distinct phenotypes are encoded in the wiring of the genetic regulatory networks (Liang and Qian, 2010). This phenomenon of epigenetic control of bimodality in gene expression by network architecture is well-known and has been extensively studied in earlier works of phage-lambda (Arkin et al., 1998; Ptashne, 2004; Zhu et al., 2004a,b; Cao et al., 2010).

Understanding multimodality in gene regulatory networks and its control mechanism can provide valuable insight into how different cellular phenotypes arises and how cellular programming and reprogramming proceed (Lu et al., 2007). Much of current knowledge of multimodality is derived from analysis of networks with feedback loops or cooperative interactions (Siegal-Gaskins et al., 2009). However, recent studies suggest that multimodality and phenotype switching can also arise from slow promoter binding, which may result in distinct protein expression levels of long durations (Feng and Wang, 2012; Thomas et al., 2014; Chen et al., 2015; Duncan et al., 2015; Terebus et al., 2019). Nevertheless, the nature and extent of this type of bimodality is not well-understood.

In this work, we study the network modules of feed-forward loops (FFLs) and characterize the stochastic nature of their multimodalities. FFLs are one of the most prevalent three-node network motifs in nature (Alon, 2006) and play important regulatory roles (Lee et al., 2002; Shen-Orr et al., 2002; Boyer et al., 2005; Mangan et al., 2006; Tsang et al., 2007; Ma et al., 2009; Sorrells and Johnson, 2015). They appear in stem cell pluripotency networks (Boyer et al., 2005; Papatsenko et al., 2015; Sorrells and Johnson, 2015), microRNA regulation networks (Tsang et al., 2007; Re et al., 2009; Ivey and Srivastava, 2010), and cancer networks (Re et al., 2009). The behavior of FFLs has been studied extensively using deterministic ODE models. These studies revealed important functions of FFLs in signal processing, including sign-sensitive acceleration and delay pulse generation functions, and increased cooperativity (Mangan and Alon, 2003; Ma et al., 2009). FFLs are also found to be capable of maintaining robust adaptation (François and Siggia, 2008; Ma et al., 2009) and detecting “fold-changes” (Goentoro et al., 2009).

However, analysis based on ODEs is limited in its ability to characterize probabilistic events, as they do not capture bimodality in gene expression that arises from slow promoter binding (Vellela and Qian, 2009). The stochastic behavior of

FFLs is not well-characterized: Basic properties such as the number of different phenotypes FFLs are capable of exhibiting, the conditions required for their emergency, their relative prominence, and the sensitivity of different phenotypes to perturbations are not known.

Our stochastic FFL models are based on processes of Stochastic Chemical Kinetics (SCK), which provides a general framework for understanding the stochastic behavior of reaction networks. Quantitative SCK modeling can uncover different network phenotypes, the conditions for their occurrence, and the nature of the prominence of the stability peaks. However, analysis of stochastic networks is challenging. First, models based on stochastic differential equations such as Fokker–Planck and Lagenvin models may be inadequate due to their Gaussian approximations. This is further compounded by the limited number of simulation trajectories that can be generated. These difficulties are reflected in the reported failure of a Fokker–Planck model in accounting for multimodality in the simple network model of single self-regulating gene at certain reaction rates (Duncan et al., 2015). Second, the widely used Stochastic Simulation Algorithm (Gillespie simulations) can generate SCK trajectories (Gillespie, 1977), but are challenged in capturing rare events and in computing efficiency. There are also difficulties in assessing convergency and in estimating computational errors (Cao and Liang, 2013). Third, even if the probabilistic landscape can be accurately reconstructed with acceptable accuracy, detecting topological features such as peaks in high-dimensional probability landscapes and assessing their objectively prominence at large-scale remains an unsolved problem.

To characterize the stochastic behavior of FFLs using models based on SCK processes, our approach is to solve the underlying discrete Chemical Master Equation (dCME) using the ACME (Accurate Chemical Master Equation) algorithm (Cao et al., 2016a,b), and to obtain the exact probability landscapes of all 8 varieties of FFLs.

Aided by the computational efficiency of ACME, we are able to explore the behavior FFLs under broad conditions of synthesis, degradation, binding, and unbinding rates of transcription factors genes binding. Furthermore, we analyze the topological features of the exactly constructed high-dimensional probability landscapes using persistent homology, so the number of probability peaks and the prominence measured by their persistence are quantified objectively. These techniques allow us to examine details of the number of possible phenotypic states at different conditions, as well as the ranges of conditions enabling phenotypic switching. With broad exploration of the model parameter space, we are able to construct detailed phase diagrams of multimodalities under different conditions.

Our results reveal how FFL network behaves differently under varying strengths of regulations intensities and the input. In addition, we characterize quantitatively the effects of duplication of genes in the FFL network modules. We show gene duplication can significantly affect the diversity of multimodality, and can enlarge monomodal regions so FFLs can have robust phenotypes. The results we obtained can be useful for analysis of phenotypic switching in biological networks containing the FFL modules.

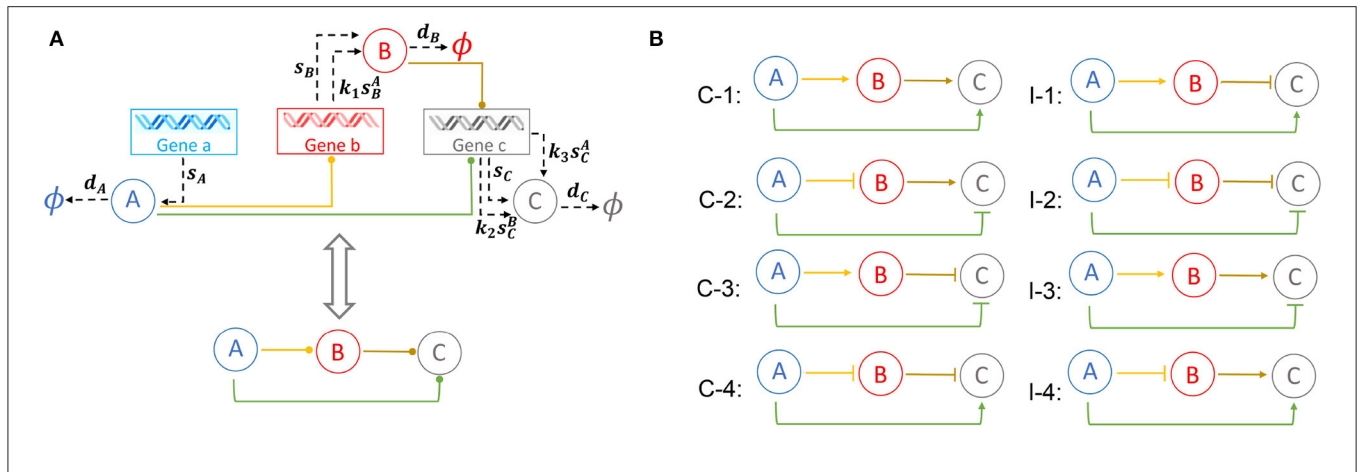


FIGURE 1 | Representation and the types of feed-forward loop (FFL) network: **(A)** General wiring and corresponding 3-node schematic representation of an FFL module containing three genes *a*, *b*, and *c* expressing three proteins *A*, *B*, and *C*. Protein *A* regulates the expressions of genes *b* and *c* through binding to their promoters. Protein *B* regulates the expression of gene *c* through promoter binding. **(B)** The FFL modules can be classified into eight different types. Coherent/incoherent FFLs are on the left/right, respectively.

They can also be used for construction of synthetic networks with the goal of generating certain desired phenotypic behavior.

2. MODELS AND METHODS

2.1. Architecture and Types of Feed-Forward Loop Network Modules

2.1.1. Overview

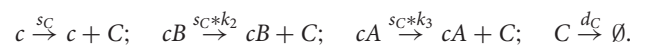
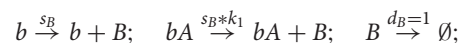
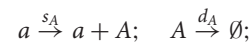
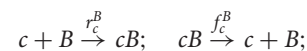
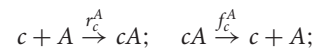
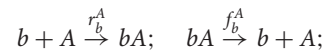
FFLs consists of three nodes representing three genes, each expresses a different protein product (Figure 1A). An FFL regulates the network output from the left input node toward the right output node via two paths; the direct path from the left node to the right node, and the indirect path from the left node to the right node via an intermediate buffer node. As each of the three regulations can be either up- or downregulation, there are altogether $2^3 = 8$ types of FFL.

2.1.2. Network Architecture

Specifically, we denote the three genes of an FFL module as *a*, *b*, and *c*, which expresses protein products *A*, *B*, and *C* at constant synthesis rate of s_A , s_B , and s_C , respectively (Figure 1A). Proteins *A*, *B*, and *C* are degraded at rate d_A , d_B , and d_C , respectively. Both proteins *A* and *B* function as transcription factors and can bind competitively to the promoter of gene *c* and regulates its expression. As the promoter of gene *c* can bind to either protein *A* or *B*, but not both, this type of regulation is known as the “OR” gate. In addition, protein *A* can bind to the promoter of gene *b* and regulate its expression. Specifically, protein *A* can bind to the promoter of gene *c* at rate r_c^A to form complex *cA*, which dissociates at rate f_c^A . *cA* expresses protein *C* at a rate k_3 -fold over the basal rate of s_C . Similarly, protein *B* can bind to the promoter of gene *c* at rate r_c^B to form complex *cB*, which dissociates at rate f_c^B . *cB* expresses protein *C* at a rate k_2 -fold over the basal rate of s_C . Furthermore, protein *A* binds to the promoter of gene *b* at rate r_b^A to form gene–protein complex *bA*, which dissociate at rate f_b^A .

Upon binding protein *A*, *bA* expresses protein *B* at a rate k_1 -fold over the basal rate of s_B .

The biochemical reactions of our FFL model are summarized below:



Here, we set $r_b^A = r_c^A = r_c^B = 0.005 \text{ s}^{-1}$, $f_b^A = f_c^A = f_c^B = 0.1 \text{ s}^{-1}$, $d_A = d_B = d_C = 1 \text{ s}^{-1}$, and $s_A = s_B = s_C = 10 \text{ s}^{-1}$. All reaction rate constants are of the unit s^{-1} , while coefficients k_1, k_2 , and k_3 are ratio of reaction rates and therefore unitless. The ratios k_1, k_2 , and k_3 can take different values so the network represents different types of FFLs.

2.1.3. Types of FFL Modules

Depending on the nature of the regulations, namely, whether each of regulation intensities k_1, k_2 , and k_3 is ≥ 1 (activating) or < 1 (inhibiting), there are $2^3 = 8$ types of FFLs. These FFLs are classified into two classes, the coherent FFLs and the incoherent FFLs (Figure 1B) (Alon, 2006). A FFL is termed coherent (C_1, C_2, C_3, C_4 in Figure 1B), if the direct effect of protein *A* on the gene *c* has the same sign (positive or negative)

TABLE 1 | Parameter ranges for eight types of feed-forward loop (FFL) model.

FFL type	k_1 range	k_2 range	k_3 range
C_1	(1.0 3.0]	(1.0 5.0]	(1.0 5.0]
C_2	[0.025 1.0)	(1.0 5.0]	(0.025 1.0]
C_3	(1.0 3.0]	[0.025 1.0)	[0.025 1.0]
C_4	[0.025 1.0)	[0.025 1.0)	(1.0 5.0]
I_1	(1.0 3.0]	[0.025 1.0)	(1.0 5.0]
I_2	[0.025 1.0)	[0.025 1.0)	[0.025 1.0]
I_3	(1.0 3.0]	(1.0 5.0]	[0.025 1.0]
I_4	[0.025 1.0)	(1.0 5.0]	(1.0 5.0]

as its net indirect effect through protein B . Taking the FFL model C_1 (Figure 1B) as an example, protein A activates gene b , and protein B activates gene c , with an overall effect of “activation.” At the same time, the direct effect of product of gene a protein A is also activation of gene c . Therefore, C_1 is a coherent FFL. When the sign of the indirect path of the regulation is opposite to that of the direct path, we have incoherent FFLs (I_1 , I_2 , I_3 , I_4 in Figure 1B). Taking the FFL model I_1 as an example, the effect of the direct path is positive, but the overall effect of the indirect path is negative. As can be seen from Figure 1B, all incoherent FFLs have an odd number of edges of inhibition.

2.1.4. Model Parameters

In order to explore broadly the behavior of all types of FFLs, we construct FFL models over the parameter space of a wide range of possible combinations of k_1 , k_2 , and k_3 , representing all 8 types of FFLs. The regulation intensity is set to values based on values reported in (Bu et al., 2016; Tej et al., 2019). We then altered the regulation intensities by about 10-fold to study the general behavior of different types of FFLs at the steady state. We take parameter values of $k_1 \in \{0.025, 0.1, 0.4, 0.8, 1.5, 2.1, 2.4, 3.0\}$, $k_2 \in [0.025, 5.0]$ with step size of 0.25, $k_3 \in [0.025, 5.0]$ with step size of 0.25. In addition, for the input intensity, the values are selected based on the analysis of abundance pattern reported in (Momin and Biswas, 2020). We take $s_A \in \{3.0, 10.0\}s^{-1}$, r_c^A and $r_c^B \in \{0.5, 2, 8, 16\}s^{-1}$ for one and two copies of genes b and c . Details of the relationship of FFL types with k_1 , k_2 , and k_3 are listed in Table 1. Over this parameter space, we study the behavior of all 8 types of FFLs. Overall, we constructed a total of 10,812 examples of FFLs and computed the steady-state probability landscape for each of them.

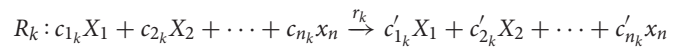
2.2. Computing Probability Landscape Using ACME

2.2.1. Exact Computation of Probability Landscape of FFLs

Consider a well mixed system of reaction with constant volume and temperature. This system has n species X_i , $i = 1, 2, \dots, n$, in which each particle can participate in m reactions R_k , $k = 1, 2, \dots, m$. A microstate of the system at time t , $\mathbf{x}(t)$ is a column vector representing the copy number of species: $\mathbf{x}(t) = (x_1(t), x_2(t), \dots, x_n(t))^T$, where the values of copy numbers are non-negative integers. The state space Ω of the system includes all the possible microstate of the system from $t = 0$ to infinity,

$\Omega = \{\mathbf{x}(t) | t \in [0, \infty)\}$. In this study, the size of the state space is $|\Omega| = 657,900$ when genes b and c are single-copy, and $|\Omega| = 686,052$ and $1,289,656$ when there are two copies of gene b and c , respectively.

The reaction R_k of the system takes the form of



which brings the system from a microstate $\mathbf{x}$ to a new microstate $\mathbf{x} + \mathbf{s}_k$, where $\mathbf{s}_k$ is the stoichiometry vector and is defined as

$$\mathbf{s}_k = (c'_{1k} - c_{1k}, c'_{2k} - c_{2k}, \dots, c'_{nk} - c_{nk}).$$

In a well mixed system, the propensity function of reaction k , $A_k(\mathbf{x})$ is given by the product of the intrinsic reaction rate constant r_k and possible combinations of the relevant reactants in the current state $\mathbf{x}$.

$$A_k(\mathbf{x}) = r_k \prod_{l=1}^n \binom{x_l}{c_{lk}}$$

With the above definitions, the dCME of a network model of the SCK processes consists of a set of linear ordinary differential equations defining the changes in the probability landscape over time at each microstate $\mathbf{x}$. Denote the probability of the system at a specific microstate $\mathbf{x}$ at time t as $p(\mathbf{x}, t) \in \mathbb{R}_{[0,1]}$, the probability landscape of the system over the whole state space Ω as $\mathbf{p}(t) = \{p(\mathbf{x}(t)) | \mathbf{x}(t) \in \Omega\}$, the dCME of the system can be written as the general form of

$$\frac{d\mathbf{p}(\mathbf{x}, t)}{dt} = \sum_{k=1}^m [A_k(\mathbf{x} - \mathbf{s}_k)p(\mathbf{x} - \mathbf{s}_k, t) - A_k(\mathbf{x})p(\mathbf{x}, t)],$$

where $\mathbf{x}$ and $\mathbf{x} - \mathbf{s}_k \in \Omega$.

The steady-state probability landscapes is obtained by solving the dCME directly. The exact solution is made possible by using the the ACME algorithm (Cao et al., 2016a,b). The ACME algorithm eliminates potential problems due to inadequate sampling, where rare events of very low probability is difficult to estimate using techniques such as the stochastic simulation algorithm (SSA) (Gillespie, 1977; Kuwahara and Mura, 2008; Daigle et al., 2011; Cao and Liang, 2013).

2.3. Identification of Multimodality by Persistent Homology

Despite its simple architecture, FFLs have a 9-dimensional probability landscape: There are three genes (a , b , and c), three proteins (A , B , and C), and three bound genes bA , cA , and cB (i.e., gene b bound to protein A , gene c bound to either protein A or protein B). Because of the high dimensionality, it is challenging to characterize the topological structures of their probability landscapes; restricting networks to only “on” and “off” state separately makes it difficult to gain insight into the overall behavior of the network.

There have been studies that analyze d -dimensional probability landscape by examining its projection onto 1-d or 2-d subspaces (e.g., 2-d heatmaps or contour plots) (Bu et al., 2016; Dey and Barik, 2021). However, projected probability

surface on lower dimensional space often no longer reflects the topology of the original space, with results and interpretations likely erroneous or misleading (Manuchehrfar et al., 2021). Finding peak states by examining distinct local maxima is equivalent to locating hypercubes that are critical points of Morse index of d in the d -dimension state space. While, local maxima may be identified by comparing its probability value with those of all of its neighbors, all peaks regardless their prominence will be identified. As numerical calculation may introduce small errors, peaks of tiny magnitude will be included. It is non-trivial to decide on a proper threshold to filter them out.

Persistent homology provides a powerful method that can characterize topological features of high-dimensional probability landscapes (Edelsbrunner et al., 2002; Carlsson, 2009). Here, we use newly developed cubic complex algorithm to compute homology groups¹ and quantitatively assess the exact topology of the 9-dimensional probability landscape.

2.3.1. Homology Groups

We use homology groups from algebraic topology to characterize the probability landscape. Homology group provides an unambiguous and quantitative description on how a space is connected. It returns a set of algebraic groups describing topological features of holes of various dimensions in the space. The rank of each i -th groups counts the number of linearly independent holes in the corresponding i th dimension. For example, $\text{Rank}(H_0)$ counts the number of connected components (0th dimensional holes).

2.3.2. Persistent Homology

Persistent homology measures the importance of these topological features (Edelsbrunner et al., 2002), and has been applied in studies of chemical compounds and biomolecules (Xia and Wei, 2014, 2015; Xia et al., 2015). Here, we focus on the topological features of probability peaks, including their appearance and disappearance. They are measured by persistent homology of the 0-th homology group. Specifically, we take the probability $p(\mathbf{x})$ as a height function, and construct a sequence of topological spaces using thresholds $\{r_i\}$ for $p(\mathbf{x})$:

$$1 = r_0 > r_1 > r_2 > \cdots > r_{i_{n-1}} > r_{i_n} = 0, \quad (1)$$

The superlevel sets $\{X_i\}$ has $X_i = \{\mathbf{x} \in X | p(\mathbf{x}) \geq r_i\}$, which corresponds to the threshold r_i . The sequence $\{X_i\}$ gives a sequence of subspaces, which is called *filtration*:

$$\emptyset \equiv X_{i_0} \subset X_{i_1} \subset X_{i_2} \subset \cdots \subset X_{i_{n-1}} \subset X_{i_n} \equiv \Omega, \quad (2)$$

As the threshold changes, the peak of a probability landscape emerges from the sea-level at a specific threshold, which is the *birth time* of the corresponding 0-homology group in the filtration. It disappears as an independent component when merged with a prior peak at a particular threshold, which is called the *death time*. When the sea-level recedes to the ground level at $p(\mathbf{x}) = 0$, only the first peak remains.

¹Tian, W., Manuchehrfar, F., Wagner, H., Edelsbrunner, H., and Liang, J. (2021). Persistent homology and moment of probability landscapes of stochastic reaction networks and their changes.

2.3.3. Persistent Diagram of Multimodality in Probability Landscape

We keep track of the probability peaks by recording the birth and death times of their corresponding 0-homology groups throughout the filtration. This relationship is depicted by the two-dimensional *persistent diagram*.

For the i th probability peak, when the threshold r reaches the value $r_b(i)$, the probability peak appears. We call this value the *birth probability* $p_b(i) = r_b(i)$ of peak i . When the threshold r is lowered to a value $r_d(i)$, this peak is merged to an existing peak. We call this value the *death probability* $p_d(i) = r_d(i)$ of peak i . The *persistence* of peak i is defined as:

$$\text{pers}(i) \equiv p_b(i) - p_d(i). \quad (3)$$

The *persistent diagram* plots peak i using the birth probability $p_b(i)$ as the y -coordinate and the death probability $p_d(i)$ as the x -coordinate. The number of dots on the persistent diagram corresponds to the number of probability peaks. Those that are further off the diagonals are the more prominent probability peaks as their persistences are larger.

3. RESULTS

3.1. Multimodality and Persistent Homology of FFLs

For each FFL network, we first compute its probability landscapes $p = p(x_A, x_B, x_C, x_a, x_b, x_c, x_{bA}, x_{cA}, x_{cB})$ at the steady-state under various conditions of model parameters. Here, x_A , x_B , and x_C are copy numbers of proteins A, B, and C, respectively; x_a , x_b , and x_c are copy numbers of genes a, b, and c, respectively; x_{bA} and x_{cA} are copy numbers of genes b and c bound by protein A; x_{cB} is the copy number of gene c bound by protein B.

Our results show that the 8 types of FFLs can exhibit up to six different phenotypes of mono- and multimodality at different conditions in the parameter spaces we investigated. An illustration of these six different types of multimodality is shown in **Figure 2**.

We further computed their 0-th homology groups at varying sea level of probability. The number of peaks, the birth, and death probability associated with each peak in **Figure 2** are shown in the persistent diagrams of **Figure 3**.

3.1.1. Behavior of FFLs From Stochastic Models Differ From Deterministic ODE Models

The behavior of FFL network modules revealed from our stochastic models are fundamentally different from that of deterministic models of ordinary differential equations (ODEs). ODE models are based on kinetics of law of mass action and are used to calculate the mean concentrations of A, B, and C at equilibrium state. However, they do not provide accurate pictures on the degree of multimodality. For example, the steady-state ODE solutions with respect to different gene occupancy for mass action kinetics show that there are at most six phenotypic states (see **Supplementary Material** for more details). However, as there are no probabilistic considerations, conclusions drawn from ODE models can be problematic.

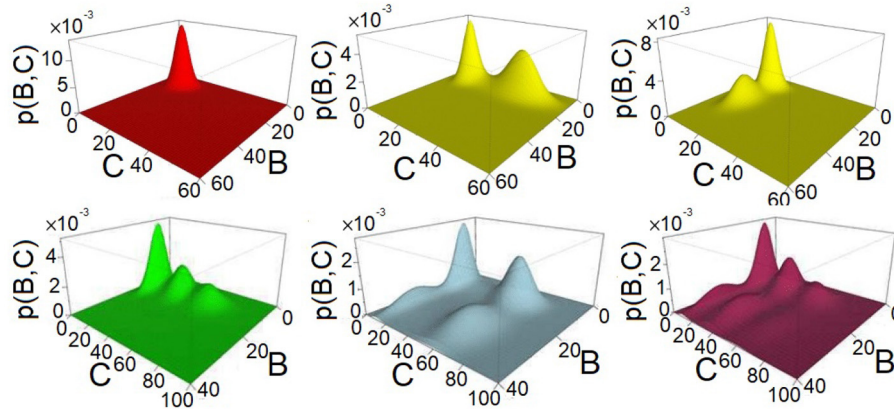


FIGURE 2 | Examples of multimodality exhibited by feed-forward loop (FFL) network motifs. The steady-state probability landscape can exhibit up to 6 different multimodes. The illustrative examples are as follows: 1 peak (red), coherent FFL of type C1 when $k_1 = 1.2$, $k_2 = 1.2$, and $k_3 = 1.2$; 2 peaks (yellow), either in protein B with coherent FFL of type C1, where $k_1 = 3.0$, $k_2 = 1.2$, and $k_3 = 1.2$, or in protein C with coherent FFL of type C1, where $k_1 = 1.2$, $k_2 = 6.0$, and $k_3 = 6.0$; 3 peaks (green), coherent FFL of type C1, where $k_1 = 1.2$, $k_2 = 6.0$, and $k_3 = 3.6$; 4 peaks (light-blue), coherent FFL of type C1 exhibits two peaks for protein B and two peaks for protein C , where $k_1 = 3.0$, $k_2 = 6.0$, and $k_3 = 6.0$; and 6 peaks (purple), coherent FFL of type C1 exhibit two peaks for B and three peaks for C , where $k_1 = 3.0$, $k_2 = 6.0$, and $k_3 = 3.6$.

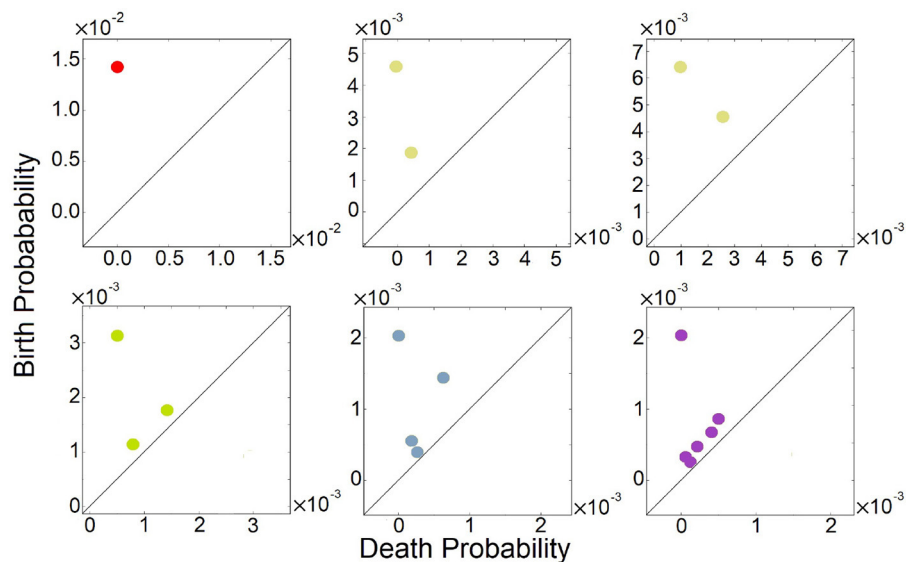


FIGURE 3 | Persistent diagrams (PDs) of feed-forward loop (FFL) network modules of **Figure 2** exhibiting different multimodalities. Red: The probability landscape with monomodality. Yellow: These two PDs depict the two steady-state landscapes exhibiting bimodality. Green, light blue, and purple: These three PDs depict the landscape exhibiting tri-modality, 4-modality, and 6-modality, respectively.

An example of the diverging results between ODE and stochastic models is shown in **Figure 4A** for an FFL of C1 type. The mean values of C obtained from the ODE model (vertical blue line) and the expectation computed from the probability landscape (vertical purple line) diverge from each other (**Figure 4A**). There are three different phenotypic states by the ODE model (green lines, **Figure 4A**), which are different from the bimodal probability distribution obtained from the SCK model (**Figure 4A**).

A further example is provided by the FFL of type I1. Here, the ODE model predicts the existence of three phenotypes at $k_1 = 2.7$, $k_2 = 0.4$, and $k_3 = 1.8$ (**Figure 4B**, green vertical lines). However, the stochastic model shows that there is only one stability peak. Although the mean value of C obtained from the ODE model and the expected C value computed from the probability landscape largely overlap, the ODE model provides no information on phenotypic variability. Overall, stochastic models provide

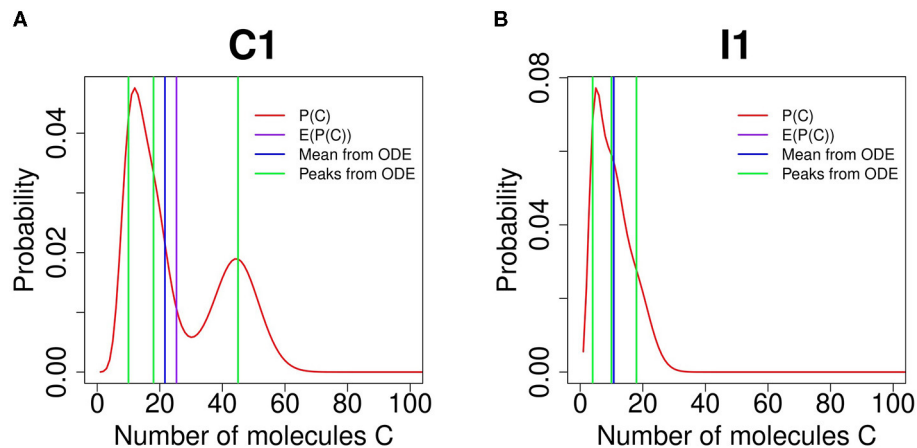


FIGURE 4 | Comparing feed-forward loop (FFL) behavior by Accurate Chemical Master Equation (ACME) and by deterministic ordinary differential equation (ODE) models. **(A)** shows the results of FFL of C1 type for $(k_1, k_2, k_3) = (2.4, 4.5, 1.8)$. The exact results obtained using ACME exhibit bimodality in protein C (red curve), while trimodality is predicted by the deterministic ODE model (green vertical lines). The mean copy number from ACME (purple vertical line) is also different from the that from ODE (blue vertical line). **(B)** shows the results of FFL of I1 type for $(k_1, k_2, k_3) = (2.4, 0.4, 1.8)$. The exact results obtained using ACME exhibit monomodality in protein C (red curve), while deterministic ODE model predicts trimodality (green vertical lines), even though the mean copy number of protein C are the same between ACME and ODE models (purple and blue vertical lines, respectively).

accurate and rich information that are not possible with ODE models.

3.1.2. Behavior of FFLs From Exact Solution to dCME by ACME Can Be Differ From That by Stochastic Simulation Algorithm

Results from simulations using SSA may differ from the exact solution to dCME obtained using ACME. We illustrate this using two incoherent FFLs, one at $(k_1, k_2, k_3) = (3.0, 0.5, 5.0)$ of I1-FFL (**Figures 5A–C**) and another at $(k_1, k_2, k_3) = (0.1, 2.75, 5.0)$ (**Figures 5D–F**) of the I4-type FFL. The exact steady-state probability landscape of the I1-FFL network computed using ACME is multimodal, exhibiting two peaks in protein B and two peaks in protein C (**Figure 5A**). However, these peaks are not definitive when 30,000 reaction trajectories up to 2,500 s are simulated using SSA (upper plots, **Figures 5B,C**). Bimodality in proteins B and C becomes only definitive when simulation time is extended to 5,000 s (lower plots, **Figures 5B,C**).

The exact steady-state probability landscape of the I4-FFL network computed using ACME exhibits tri-modality in protein C and bimodality in protein B (**Figure 5D**). However, tri-modality is not clearly captured when the reaction trajectories are $< 2,500$ s (upper plot, **Figure 5E**), and becomes definitive only after 5,000 s (lower plot, **Figure 5E**). In addition, bimodality in protein B is not captured, even when the reaction trajectories are at 5,000 s (upper and lower plot, **Figure 5F**).

3.2. Phase Diagrams of Multimodality in FFLs

Current studies of stochastic networks are limited to their behavior under a few selected conditions. Here, we explore the multimodality of all eight types of FFLs under broad conditions of synthesis, degradation, binding, and unbinding as outlined in **Table 1**. This is made possible by the efficiency of the multi-finite

buffer ACME algorithm. The analysis using persistent homology further allows us to quantitatively characterize the exact topology of the landscape. Together, we are able to obtain the full phase diagrams on the phenotype of multimodality of FFLs at different combinations of parameter values (**Figure 6**).

Altogether, we compute 10,812 probability landscapes of the 8-types of FFL modules. Depending on the values of k_1, k_2 , and k_3 , each phase diagram shown depicts the behavior of four types of FFLs, one for each of the four quadrants formed by the two straight lines of $k_2 = 1$ and $k_3 = 1$ (**Figure 6**), with the type of FFL labeled accordingly. The specific types also depend on k_1 , which is listed at the top of each plot (**Figure 6**). As a result, we have gained comprehensive and accurate characterization of the multimodality phenotypes of this type of important network modules.

3.2.1. Monomodality

As shown in **Figure 2**, the steady-state probability landscape of the FFL at $k_1 = k_2 = k_3 = 1.2$ exhibits one probability peak. At this condition, it is a coherent FFL of type C1. The projected distributions of B and C exhibit monomodality and has only one peak (**Figure 2**, red) when the values of intensities k_1, k_2 , and k_3 are close to 1.0 (**Figure 6**). Overall, there is only one phenotypic state when the regulations intensities in FFL are weak.

3.2.2. Bimodality

The steady-state probability landscape of FFLs can exhibit two types of bimodality (colored yellow in **Figure 2**). The first type occurs when $k_1 < 0.4$ or $k_1 \geq 2.4$, with bimodality in protein B while monomodality in protein C. This is illustrated as green regions in **Figure 6** shown at the two top-left and the two bottom right phase diagrams where $k_1 \in \{0.025, 0.1, 2.4, 3.0\}$. That is, if the regulation intensities of k_1 and k_2 are about two-fold different either way, bimodality in B arises.

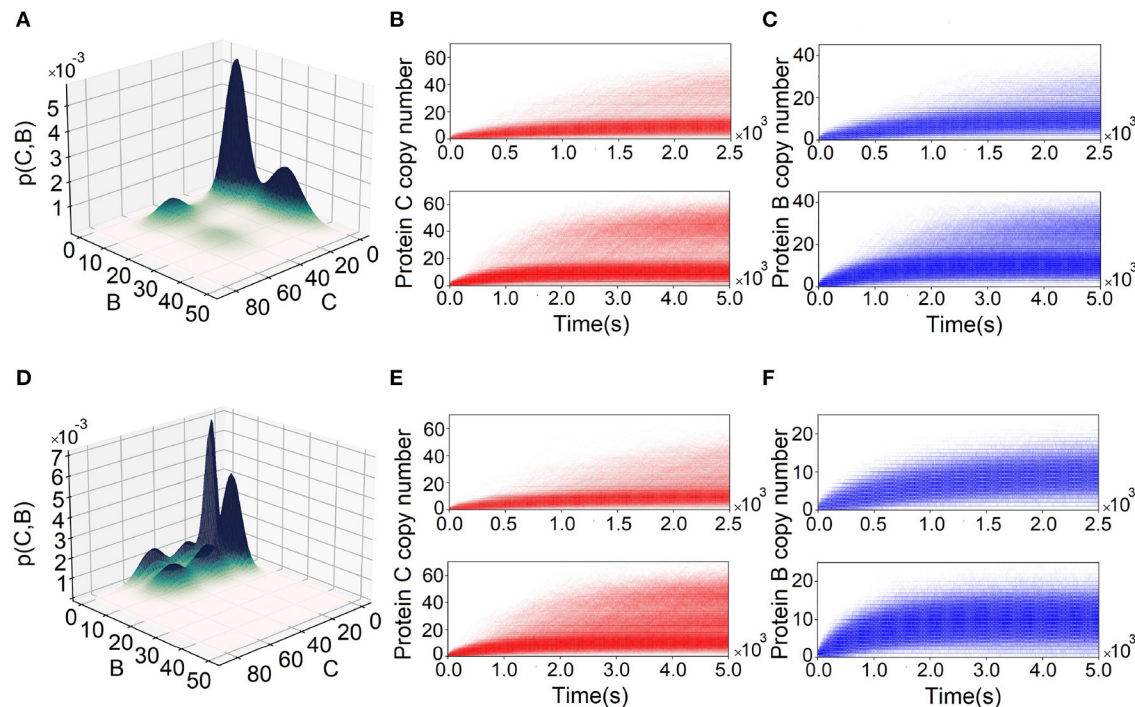


FIGURE 5 | Comparing landscapes from Accurate Chemical Master Equation (ACME) and reaction trajectories from the stochastic simulation algorithm (SSA). **(A)** Probability surface projected onto the (B, C) -plane for the feed-forward loop (FFL) with $(k_1, k_2, k_3) = (3.0, 0.5, 5.0)$. There is bimodality in both proteins B and C. **(B,C)** The reaction trajectories computed from SSA corresponding to condition in **(A)** for proteins C and B, respectively. The upper plots are for 2,500 s and lower plots are for 5,000 s. SSA does not capture the bimodality of proteins B and C until 2,500 s. **(D)** The probability surface projected onto $(B-C)$ plain for FFL with $(k_1, k_2, k_3) = (0.1, 2.75, 5.0)$. There is tri-modality in protein C and bimodality in protein B. **(E,F)** Corresponding reaction trajectories in proteins C and B, respectively. Upper plots are for the results for 2,500 s and lower plots are for 5,000 s. SSA does not capture tri-modality of protein C until 2,500 s. In addition, SSA fails to capture bimodality in protein B.

The second type of bimodality occurs when $0.4 \leq k_1 < 2.4$, where protein C exhibit bimodality while monomodality is maintained in B. This is illustrated as green regions in the remaining phase diagrams of **Figure 6**, where $k_1 \in \{0.4, 0.8, 1.5, 2.1\}$.

3.2.3. Tri-modality

The steady-state probabilistic landscape of FFL can exhibit tri-modality (green, **Figure 2**). There are three possible phenotypes in protein C while monomodality in protein B is maintained. Trimodal regions are colored red in the phase diagrams of **Figure 6**. They arise when the difference in rates k_2 and k_3 is at least about two-fold and $0.4 \leq k_1 \leq 2.1$.

3.2.4. Multimodality

The steady-state probability landscape of the FFL can exhibit 4 to 6 probability peaks (orange, purple, and green, respectively, in **Figure 2**). Landscapes with 4 modes have bimodality in both protein B and protein C. Those with 5 modes has bimodality in B and tri-modality in C. Landscapes with 6 modes exhibit bimodality in B and tri-modality in C. Inspection on the conditions indicates that when the regulations are strong; i.e., when k_1, k_2 , and $k_3 \geq 2.1$, FFLs exhibit very well-defined multimodality peaks. However, when the regulation intensity

k_1 is weak, the steady-state probability landscape exhibits multimodality only when the other two regulation intensities, namely, k_2 and k_3 are strong. As shown in **Figure 6**, there are two groups of FFLs based on the characteristics of the multimodality they exhibit: One group consists of FFLs of C_2, C_4, I_1 , and I_3 types, where tri-modality of output protein C always exists, as long as k_2 and k_3 are at least about two-fold different. The other group consists of FFLs of C_1, C_3, I_2 , and I_4 types where the signs of the regulations that the output node C receives from B and A are the same (both activation or both inhibition). Tri-modality occurs when the regulations k_2 and k_3 have very distinct values.

Overall, protein B can exhibit either mono- or bimodality, and protein C can exhibit mono-, bi-, or tri-modality on the probability landscape.

3.3. Increasing Input Intensity Amplifies Multimodality in FFL

To understand how input intensity affect the response of FFL networks, we examine their behavior under different input conditions. Specifically, we examine how different synthesis rate s_A of protein A affects the number of modes in proteins B and C.

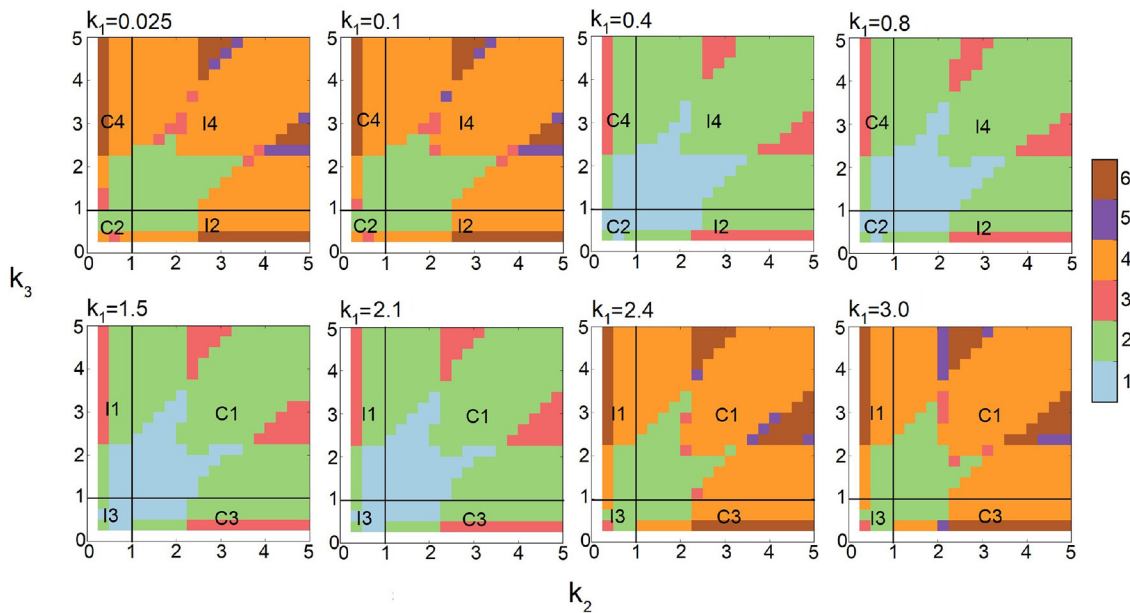


FIGURE 6 | Phase diagrams of multimodality of Feed-forward loop (FFL) network modules based on 10,812 steady-state probability landscapes at different condition of regulation intensities for all 8 types of FFL network modules. **Monomodality** occurs when $0.4 \leq k_1 \leq 2.1$ and k_2, k_3 intensities are moderate, i.e., $0.4 \leq k_1 \leq 3$ (blue region when $k_1 = 0.4, 0.8, 1.5$, and 2.1). **Bimodality** may occur for different combinations of regulation intensities. When k_1 intensity is either very high ($2.4 \leq k_1$) or very low ($k_1 \leq 0.1$), bimodality occurs when k_2, k_3 intensities are moderate, i.e., $0.4 \leq k_1 \leq 3$. When k_1 intensity is moderate ($0.4 \leq k_1 \leq 2.1$), bimodality occurs when at least one of the other regulation intensities k_2 or k_3 is high. **Tri-modality** occurs when k_1 is moderate ($0.4 \leq k_1 \leq 2.1$) and either k_2 or k_3 is moderate. **Multimodality** occurs when k_1 is low or high ($k_1 \leq 0.4$ or $k_1 \geq 2.1$), and at least either k_2 or k_3 is high. Color scheme (vertical bar): Blue, green, red, orange, purple, and brown represent regions with one, two three, four, five, and six peaks, respectively.

We first carry out computations and broadly survey the behavior of FFLs at strong input intensity, where s_A is set to 10.0. The values of k_2 and k_3 are sampled broadly, and k_1 is tested for three different values of $k_1 = 0.8, 2.1$, and 2.4 . The results are summarized in **Figure 7** (top row). We then similarly survey the behavior of FFLs at decreased synthesis intensity of protein A, with $s_A = 3.0$ (**Figure 7**, bottom row).

There are clear changes in the mode of multimodality of FFLs. At $k_1 = 0.8$ and $k_1 = 2.1$ (**Figure 7**, left and center columns), when protein A synthesis rate s_A is reduced from 10.0 (top) to 3.0 (bottom), regions with one (blue) and three (red) peaks are reduced. In addition, certain areas of the tri-stable (red) regions become bimodal (green).

At larger $k_1 = 2.4$ (**Figure 7**, right column), the FFLs exhibits dramatic changes in the modes of multimodality when synthesis rate s_A of protein A is reduced from 10.0 (top) to 3.0 (bottom). In many regions, one or more stability peaks disappear. There are regions with two peaks at $s_A = 10.0$ that become monomodal. There are also regions of six peaks that become those of four peaks. This is due to the loss of one stability peak from three in protein C. In addition, large regions with four peaks (orange) disappear and become either regions with two peaks (green) or with three peaks (red). Overall, we can conclude that high-input intensity represented by high s_A rate for protein A induces changed phenotypes of multimodality in FFLs.

3.4. Binding and Unbinding Dynamics Are Critical for Multiple Phenotypic Behavior

Results obtained so far are based on the assumption of slow binding ($r_b^A = r_c^A = r_c^B = 0.005$) and unbinding ($f_b^A = f_c^A = f_c^B = 0.1$) reactions, which we call the *generic case*. When the FFL network slowly switches between phenotypic states, the process of synthesis degradation of protein C has sufficient time to converge to equilibrium at each phenotypic state of gene c . An important question is how slow the promoter dynamics need to be for FFLs to exhibit multiple phenotypes, without feedback loops or cooperatively.

To answer this question, we explore the behavior of FFLs under different binding and unbinding dynamics of gene c for an FFL of type I1. In this case, protein A activates protein B and protein C, while protein B inhibits protein C (see **Figure 1B**). With slow binding kinetics as described above, the output C of this FFL exhibits three stability peaks. These are at the expression level of protein C of (1) $C = 0$, corresponding to the condition when gene c is inhibited by B, (2) $C = 9$, corresponding to the basal level of C expression, and (3) $C = 49$, when C expression is activated by A. We then fix the regulation intensities at $k_1 = 3.0$, $k_2 = 0.025$, and $k_3 = 5.1$, and examine how the number of phenotypic states is affected by gene c binding dynamics (**Figure 8**).

We first set the binding affinities between gene c and protein A and between gene c and protein B to the same

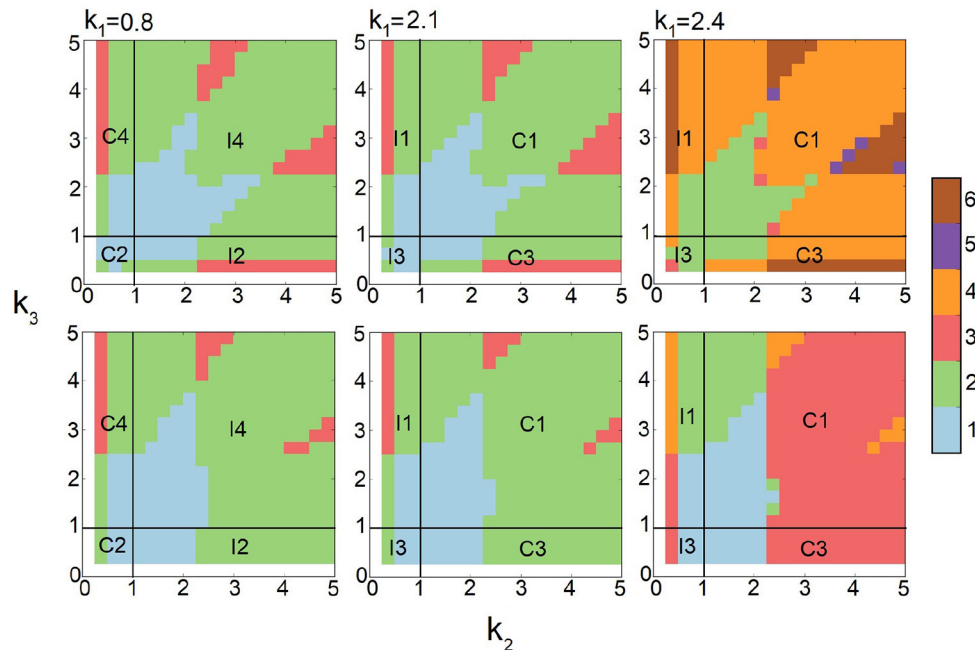


FIGURE 7 | Effects of input intensity on multimodality of Feed-forward loops (FFLs). The phase diagrams of the number of stability peaks in the steady-state probability landscapes at strong input intensity $s_A = 10.0$ (top row) and weak input intensity $s_A = 3.0$ (bottom row) for different k_2 and k_3 at three different conditions of $k_1 = 0.8, 2.1$, and 2.4 . Color scheme (vertical bar): Blue, green, red, orange, purple, and brown represent regions with one, two, three, four, five, and six peaks, respectively.

values, and change them together to n -fold of the *generic case*, where $n \in \{0.5, 2, 8, 16\}$. For slower binding and unbinding dynamics (yellow line for $n = 0.5$, **Figure 8A**), the modes of the distribution of the output of protein C are even better distinguished. However, when both binding and unbinding rates are increased to $n = 8$ fold (green line), the probability peak at $C = 9$, which corresponds to basal level of C expression, merges with the probability peak at $C = 0$. At $n = 16$, the distribution of C is bimodal.

We then keep the binding affinity between gene c and protein A unchanged and alter only the binding affinity between gene c and protein B by n -fold, where $n \in \{0.5, 2, 8, 16\}$. When the binding affinity increases (e.g., $n = 8$), the probability peak at $C = 9$ disappears, while the probability peak at high copy number of $C = 49$ robustly remains, although with less magnitude (**Figure 8B**).

When only the binding affinity between gene c and protein A is altered while that between gene c and protein B is held constant (**Figure 8C**), the probability peak at the basal level of C expression ($C = 9$) diminishes when the binding affinity increases (e.g., $n = 8$). However, the probability peak at $C = 49$ becomes more prominent. At $n = 8$, the distribution of C is tri-modal. At $n = 16$, it becomes bimodal. This indicates that multiple phenotypes arise in FFLs when the unbinding rate is about an order of magnitude smaller than the expression rate of the protein.

3.5. Gene Duplication Can Enrich Phenotypic Diversity and Enlarge Stable Regions of Specific Multimodality of FFLs

Gene duplication provides a basic route of evolution (Lynch and Conery, 2000) and is an important driver of phenotypal diversity in organisms (Conrad and Antonarakis, 2007). Here, we study how gene duplication affects the phenotypes of FFLs.

We examine how duplication of gene c and separately duplication of gene b affect the behavior of the FFL network modules. With two copies of gene c , there can be six possible states of gene c activation. Depending on whether the promoter sites of both copies of gene c are free or occupied by either protein A or protein B, we have for both c genes to have unoccupied, protein A bound, or protein B bound promoter site. This can be denoted as a triplet (c, cA, cB) , which can take any of the possible values of $(2, 0, 0)$, $(0, 2, 0)$, $(0, 0, 2)$, $(1, 1, 0)$, $(1, 0, 1)$, and $(0, 1, 1)$. For the case when there are two copy number of gene B, there are three possible states of gene b activation, depending on whether the promoter site of both copies of gene b are free or occupied by protein A. This can be denoted as a duplicate (b, bA) , which can take any of the possible values of $(2, 0)$, $(1, 1)$, or $(0, 2)$.

The phase diagrams of the number of modes of stability peaks are shown in **Figure 9**, when there is only one copy of both gene b and gene c (first row), when there are two copies of gene c but one copy of gene b (second row), and when there are two copy number of gene b but one copy of gene c (third row). The

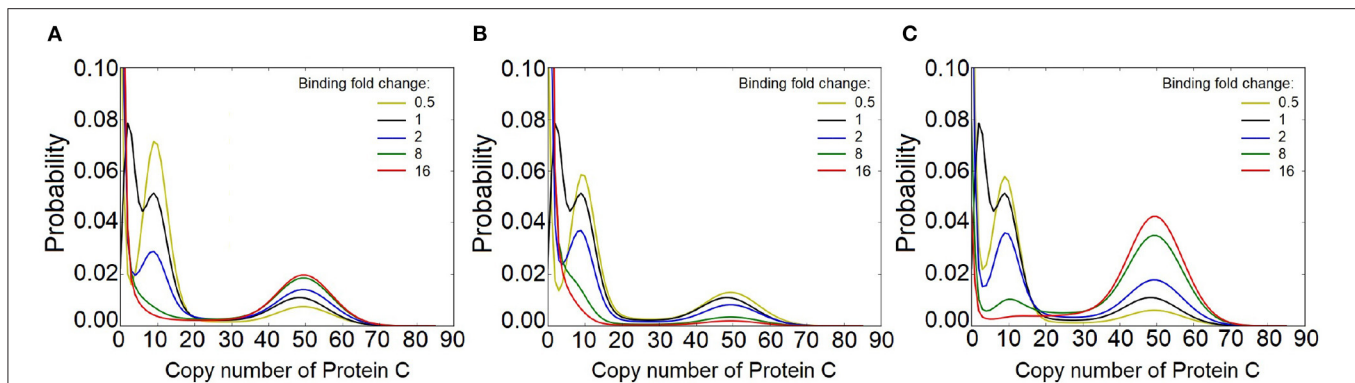


FIGURE 8 | Effect of binding dynamics on the modality of protein C in the feed-forward loop (FFL) network of type I1, with $(k_1, k_2, k_3) = (3.0, 0.025, 5.1)$. **(A)** Effects when binding affinity between gene *c* and both protein A and protein B are altered by *n*-fold, where $n \in \{0.5, 2, 8, 16\}$. At slower binding (yellow line), the modes of distribution of protein C are well-distinguished. However, when the binding and unbinding rates increased to 8 (green line), the peak at *C* = 9 disappears. At $n = 16$, bimodality is observed in protein C. **(B)** Effects when only the binding affinity of gene *c* and protein B is altered by *n*-fold, where $n \in \{0.5, 2, 8, 16\}$. When the binding affinity of gene *c* and protein B increases, the peak at *C* = 9 disappears, while the peaks at *C* = 49 robustly remains. However, the peak at *C* = 49 becomes less significant. **(C)** Effects when only the binding affinity of gene *c* and protein A is altered by *n*-fold, where $n \in \{0.5, 2, 8, 16\}$. At high binding affinity, the peak at *C* = 9 disappears while the peak at *C* = 49 becomes more prominent.

conditions are $k_1 = 0.025, 0.8, 1.5$, and 2.4 , for different values of $k_2 \in [0.1, 5]$ and $k_3 \in [0.1, 5]$, where there are slow binding and unbinding ($r_b^A = r_c^A = r_c^B = 0.005, f_b^A = f_c^A = f_c^B = 0.1$). Each phase diagram in **Figure 9** consists of 400 steady-state probability landscapes with a total $12 \times 400 = 4,800$ landscapes. This broad range of parameters allow us to study all 8 different modules of FFL network and the effects of gene *c* and gene *b* duplications.

We examine the behavior of FFL in three different regimes of k_1 : (1) When $k_1 \ll 1.0$ (**Figure 9**, first column), the bimodal regions (green) expands when there are two copies of gene *c* (second row), but there are no significant changes when there are two copies of gene *b* (third row). The overall size of multimodal regions increases in both cases. (2) When $k_1 \approx 1.0$ (**Figure 9**, second and third columns), the duplication of gene *c* (second row) expands the regions with three stability peaks and reduces regions with two peaks. In contrast, the duplication of gene *b* (third row) has no significant effects on multimodality. (3) When $k_1 = 2.4$ (fourth column), duplication of gene *c* (second row) expands regions with two and six stability peaks. Duplication of gene *b* (third row) reduces the region with four peaks and expands the region with five peaks.

These results show that introducing additional copy of gene *b* or gene *c* not only can enrich different phenotypic behavior but can also increase the stability of specific phenotypic states, namely, enlarge regions of particular phenotypes by uniting previously different phenotypic regions together. Overall, gene duplication can increase phenotypic diversity, and enlarge stability regions of specific multimodal states.

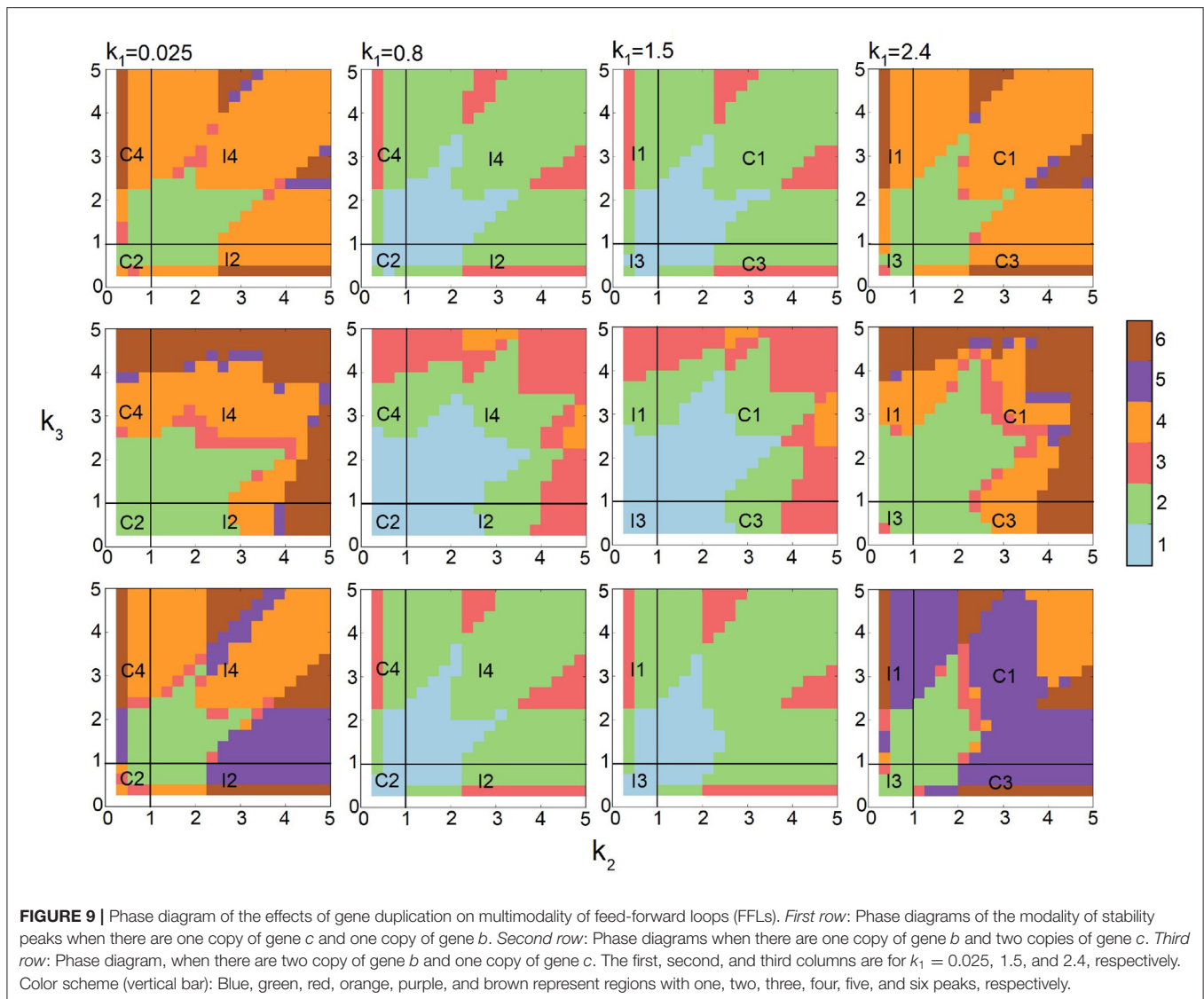
Bacterial cells have fast binding and unbinding dynamics (Ali Al-Radhawi et al., 2019), and it is unlikely that the occurrence of multiple copies of the same gene in FFLs plays significant roles in stochastic multimodality. In contrast, mammalian cells have slower promoter dynamics (Forger and Peskin, 2003). Gene duplication in FFLs may provide a natural mechanism for enriched multimodality with enhanced stochastic phenotypic

switching. This is reflected in reduced monomodal regions, and enlarged multimodal regions where there are 4 (orange), 5 (purple), and 6 (brown) phenotypic states of the output *C* (second and third row in **Figure 9**).

Assuming that initially both copies of the gene were functioning, but subsequently one gene copy lost its biochemical function due to mutations, we can expect two opposite types of scenarios to occur: If regulation intensities are strong (k_2 and k_3 are large), one of the phenotypic states becomes lost (e.g., green region becomes light blue, and orange region becomes red, **Figure 9**). If regulation intensities are weak, the duplication of gene *c* or gene *b* can lead to enlargement of the region of monomodality. It can also lead to the appearance of new regimes where there are a larger number of multimodality modes (orange, purple, and green regions in **Figure 9**). That is, gene duplication can create new stable states, leading to an enlarged number of high probability states. This, however, occurs only in FFL modules with strong regulations intensities. FFL modules with low regulation intensities instead lose phenotypical diversity and become more robust in monomodality with enlarged region in the parameter space.

4. DISCUSSION

Gene regulatory networks (GRNs) play critical roles in defining cellular phenotypes but it is challenging to characterize the behavior of GRNs. Although GRNs may consist of dozens or more of genes and proteins, their functions often can be defined by smaller sub-networks called network motifs. How small network motifs are responsible for complex properties such as the maintenance of multi-phenotypic behavior or modules is poorly understood. Current widely practiced approach is studying network motifs using deterministic models. However, this approach imposes restrictions on the types of network motifs



capable of exhibiting multimodal phenotype to mostly feedback networks.

In this study, we examined the FFL network motifs, one of the most ubiquitous three-node network motifs. Although their deterministic behavior is well-studied, with great understanding of their functions such as signal processing and adaptations gained, their stochastic behavior remains poorly characterized.

Here, we showed the direct regulation path from the input node to the output node and the indirect path through the intermediate buffer node provide the necessary architecture for distinct multiple modalities. Phase diagrams of FFL in **Figure 6** show that FFLs of various types can exhibit different multimodality. At large copy numbers and large volume, our model of stochastic reaction kinetics are the same as those based on mass action kinetics (Kurtz, 1971, 1972; Vellela and Qian, 2007), where ordinary differential equation (ODE) models are appropriate. When ODE models are applied to enzyme-substrate reactions, they can be further approximated

by Michaelis-Menten kinetics, with the additional assumption that the substrate is in instantaneous chemical equilibrium with the enzyme-substrate complex. When ODE models are applied to the reaction of one receptor and n identical simultaneously binding ligands, we arrive at the Hill equation, with the coefficient n phenomenologically characterizing cooperativity. These kinetic models based on ODE approximations, however, are not applicable to the current study, as we are examining strong stochasticity arising at low copy number of molecules, where ODE models are not valid.

FFLs play important roles in gene regulatory networks. For example, it is shown that several 11-FFL sub-networks control the process of *Bacillus subtilis* sporulation (Eichenberger et al., 2004; Mangan et al., 2006). In addition, C1-FFL network is found to be present in the *L-arabinose* (*ara*) utilization system of *E. coli*, where *araBAD* is the target (gene *c*) activated by the intermediate gene *araC* and the input gene *CRP*. Gene *araC* is also activated by *CRP*. Therefore, they form a 3-node C1 type FFL (Mangan et al.,

2003). Results in this work can help to gain understanding of the behavior of these different types of FFLs found in gene regulatory networks.

In addition, we have shown that input intensity affects the multimodal behavior of various types of FFLs. Examples shown in **Figure 7** demonstrate that at high k_1 values, input intensity dramatically changes the multimodality as shown in the phase diagrams. Our results are consistent with previous findings that input intensity is an important factor in determining output intensity of FFLs (Mangan et al., 2003; Goentoro et al., 2009; Lin et al., 2018). Here, we further demonstrated that input intensity is also important in determining the modality of the steady-state behavior of FFLs.

In mammalian cells, slow dynamics of transcription factor binding to promoter is often observed (Dermitzakis and Clark, 2002; Hager et al., 2009; Lickwar et al., 2012; Tuğrul et al., 2015; Hasegawa and Struhl, 2019). This is likely due to the complex process of chromatin regions opening up so they become accessible and the slow nature of events such as promoter, enhancer, and mediator binding. These physical processes result in highly stochastic behavior of networks. Stochastic models have demonstrated that complex multimodality phenotypes can naturally arise from stochastic fluctuations when genes have distinct expression levels, a phenomenon widely observed in mammalian cells (Cao et al., 2018). We showed that binding and unbinding dynamics are critical for multi-phenotypic behavior. For an I1-FFL with $(k_1, k_2, k_3) = (3.0, 0.025, 5.1)$, **Figure 8** highlighted that binding and unbinding rates affect multiple peaks in protein C.

Results of this study indeed showed that once stochastic fluctuations between distinct expression levels due to slow promoter dynamics are considered, FFLs can exhibit complex multimodal phenotypes. When the expression levels of the output gene (gene *c*) at the inhibited, basal, and activated states are well-separated, three distinct phenotypes arise. Combined with two additional possible phenotypes of different levels of gene *b* expression, we can have up to six modalities for FFLs. Furthermore, high intensity of input amplifies multimodality in FFLs, suggesting that the FFL architecture are favored for maintaining multiple phenotypic states. In addition, we find that regulation intensities are key determinants of specific stochastic behavior of FFLs, which could be tuned in order to obtain any desired phenotypic behavior between 1 and 6 stability modes.

Our study also revealed the roles of gene duplication. When there are two copies of gene *c*, while one in principle could expect $2 \times 6 = 12$ different phenotypes for the output protein C. This is, however, not observed, as the regulation intensities or reaction rates are not well-separated. In contrast, instead of further increase in multimodality beyond six, we observe the expansion of the area of monomodality, resulting from the connectedness of regions of expression with different rates that are merged together. Our results showed that duplication of gene *b* and gene *c* not only can enrich different phenotypic behavior but can also increase the stability of certain phenotypic states, while decreasing others (**Figure 9**). We showed that in general, gene duplication can

enrich phenotypic diversity. The presence and functional roles of gene duplication are well-known (Hurles, 2004). For example, in human-induced pluripotent stem cells (HiPSCs), chromosome 12 duplication leads to significant enrichment of cell cycle related genes (Mayshar et al., 2010), in which FFL sub-networks play important roles. This abnormality results in increase in the tumorigenicity of HiPSCs. Our findings may also shed light on how gene duplication affects cellular adaptation to changing environment (Kondrashov, 2012): As the support regions of monomodality are enlarged, smaller fluctuations in regulation intensities will not switch cells with duplicated genes to a different phenotypic state. Thus, gene duplication may help to stabilize the behavior of the network, so cells are better adapted to a changing environment.

Analysis of stochastic behavior of FFLs reported here have implications in a variety of biological problems. For example, the stem cell regulation network consisting of pluripotency transcription factors *Oct4* and *Nanog* maintain pluripotency against differentiation (Boyer et al., 2005; Chickarmane et al., 2006; Papatsenko et al., 2015; Lin et al., 2018). A component of this network can be abstracted as an FFL: *Nanog* participates as the intermediate node (gene *b*, which is activated by *Oct4* (gene *a*), and both regulate the expression of genes associated with the onset of differentiation or pluripotency (gene *cs*). In addition, regulation networks in hematopoietic stem cells are formed by two FFL networks involving β globin, *GATA-1*, *EKLF*, and *FOG-1*. In each network, *FOG-1* and *EKLF* function as the intermediate genes (gene *b*) and are activated by *GATA-1* (gene *a*), while all of them activate β globin (gene *c*) (Swiers et al., 2006). Moreover, in other stem cell differentiation networks, there are several sub-networks that exhibit behaviors of different types of FFLs. For example, *Klf4* (gene *a*) activates *Pou5f1* (gene *b*) and inhibits *Sox2* (gene *c*), while *Pou5f1* activated *Sox2* (Onichtchouk et al., 2010; Okawa et al., 2016), as in the C3-type FFL (**Figure 1**).

In summary, we have constructed and analyzed the exact high-dimensional steady-state probability landscapes of FFLs under broad conditions and have constructed their phase diagrams in multimodality. These results are based on 10,812 exactly computed probability landscapes and their topological features as measured by persistent homology. With slow binding and unbinding dynamics of transcription factor binding to promoter, FFLs exhibit strong stochastic behavior that is very different from deterministic models, and can exhibit from 1 up to 6 stability peaks. In addition, input intensity play major roles in the phenotypes of FFLs: At weak input intensity, FFLs exhibit monomodality, but strong input intensity may result in up to 6 stable phenotypes. Furthermore, we found that gene duplication can enrich the diversity of FFL network phenotypes and enlarge stable regions of specific multimodalities.

Results reported here can be useful for constructing synthetic networks, and for selecting model parameters, so a particular desirable phenotypic behavior can materialize (Jones et al., 2020). Our results can also be used for analysis of behavior of FFLs in biological processes such as stem cell differentiation and for design of synthetic networks with desired phenotype behavior. We hope results reported here for different types of FFL can be tested experimentally.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

AT and JL conceived and designed the study. AT designed and carried out analysis of the ODE model, multimodality, phase diagrams, slow dynamics, and gene duplication. FM designed and carried out analysis of persistent homology, and assisted in multimodality and phase diagram computation. YC participates in design and data analysis. AT and JL wrote the manuscript with significant input from FM. All authors have read and approved the final manuscript.

REFERENCES

- Acar, M., Becskei, A., and van Oudenaarden, A. (2005). Enhancement of cellular memory by reducing stochastic transitions. *Nature* 435, 228–232. doi: 10.1038/nature03524
- Ali Al-Radhawi, M., Del Vecchio, D., and Sontag, E. D. (2019). Multi-modality in gene regulatory networks with slow promoter kinetics. *PLoS Comput. Biol.* 15:e1006784. doi: 10.1371/journal.pcbi.1006784
- Alon, U. (2006). *An Introduction to Systems Biology: Design Principles of Biological Circuits*. New York, NY: CRC Press. doi: 10.1201/9781420011432
- Arkin, A., Ross, J., and McAdams, H. H. (1998). Stochastic kinetic analysis of developmental pathway bifurcation in phage λ -infected *Escherichia coli* cells. *Genetics* 149, 1633–1648. doi: 10.1093/genetics/149.4.1633
- Boyer, L. A., Lee, T. I., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 122, 947–956. doi: 10.1016/j.cell.2005.08.020
- Bu, P., Wang, L., Chen, K.-Y., Srinivasan, T., Murthy, P., Tung, K.-L., et al. (2016). A mir-34a-numb feedforward loop triggered by inflammation regulates asymmetric stem cell division in intestine and colon cancer. *Cell Stem Cell* 18, 189–202. doi: 10.1016/j.stem.2016.01.006
- Cao, Y., Lei, X., Ribeiro, R. M., Perelson, A. S., and Liang, J. (2018). Probabilistic control of hiv latency and transactivation by the tat gene circuit. *Proc. Natl. Acad. Sci. U.S.A.* 115, 12453–12458. doi: 10.1073/pnas.1811195115
- Cao, Y., and Liang, J. (2013). Adaptively biased sequential importance sampling for rare events in reaction networks with comparison to exact solutions from finite buffer DCME method. *J. Chem. Phys.* 139:025101. doi: 10.1063/1.4811286
- Cao, Y., Lu, H.-M., and Liang, J. (2010). Probability landscape of heritable and robust epigenetic state of lysogeny in phage lambda. *Proc. Natl. Acad. Sci. U.S.A.* 107, 18445–18450. doi: 10.1073/pnas.1001455107
- Cao, Y., Terebus, A., and Liang, J. (2016a). Accurate chemical master equation solution using multi-finite buffers. *Multiscale Model. Simul.* 14, 923–963. doi: 10.1137/15M1034180
- Cao, Y., Terebus, A., and Liang, J. (2016b). State space truncation with quantified errors for accurate solutions to discrete chemical master equation. *Bull. Math. Biol.* 78, 617–661. doi: 10.1007/s11538-016-0149-1
- Carlsson, G. (2009). Topology and data. *Bull. Am. Math. Soc.* 46, 255–308. doi: 10.1090/S0273-0979-09-01249-X
- Chen, Y., Lv, C., Li, F., and Li, T. (2015). Distinguishing the rates of gene activation from phenotypic variations. *BMC Syst. Biol.* 9:1. doi: 10.1186/s12918-015-0172-0
- Chickarmane, V., Troein, C., Nuber, U. A., Sauro, H. M., and Peterson, C. (2006). Transcriptional dynamics of the embryonic stem cell switch. *PLoS Comput. Biol.* 2:e123. doi: 10.1371/journal.pcbi.0020123
- Choi, P. J., Cai, L., Frieda, K., and Xie, X. S. (2008). A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science* 322, 442–446. doi: 10.1126/science.1161427

FUNDING

This work was supported by NIH grant R35 GM127084.

ACKNOWLEDGMENTS

We thank Wei Tian for assistance in computing persistent homology.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.645640/full#supplementary-material>

- Conrad, B., and Antonarakis, S. E. (2007). Gene duplication: a drive for phenotypic diversity and cause of human disease. *Annu. Rev. Genomics Hum. Genet.* 8, 17–35. doi: 10.1146/annurev.genom.8.021307.110233
- Daigle, B. J., Roh, M. K., Gillespie, D. T., and Petzold, L. R. (2011). Automated estimation of rare event probabilities in biochemical systems. *J. Chem. Phys.* 134:044110. doi: 10.1063/1.3522769
- Dermitzakis, E. T., and Clark, A. G. (2002). Evolution of transcription factor binding sites in mammalian gene regulatory regions: conservation and turnover. *Mol. Biol. Evol.* 19, 1114–1121. doi: 10.1093/oxfordjournals.molbev.a004169
- Dey, A., and Barik, D. (2021). Potential landscapes, bifurcations, and robustness of tristable networks. *ACS Synth. Biol.* 10, 391–401. doi: 10.1021/acssynbio.0c00570
- Duncan, A., Liao, S., Vejchodský, T., Erban, R., and Grima, R. (2015). Noise-induced multistability in chemical systems: discrete versus continuum modeling. *Phys. Rev. E* 91:042111. doi: 10.1103/PhysRevE.91.042111
- Edelsbrunner, H., Letscher, D., and Zomorodian, A. (2002). Topological persistence and simplification. *Discrete Comput. Geometry* 28, 511–533. doi: 10.1007/s00454-002-2885-2
- Eichenberger, P., Fujita, M., Jensen, S. T., Conlon, E. M., Rudner, D. Z., Wang, S. T., et al. (2004). The program of gene transcription for a single differentiating cell type during sporulation in *Bacillus subtilis*. *PLoS Biol.* 2:e328. doi: 10.1371/journal.pbio.0020328
- Feng, H., and Wang, J. (2012). A new mechanism of stem cell differentiation through slow binding/unbinding of regulators to genes. *Sci. Rep.* 2:550. doi: 10.1038/srep00550
- Forger, D. B., and Peskin, C. S. (2003). A detailed predictive model of the mammalian circadian clock. *Proc. Natl. Acad. Sci. U.S.A.* 100, 14806–14811. doi: 10.1073/pnas.2036281100
- François, P., and Siggia, E. D. (2008). A case study of evolutionary computation of biochemical adaptation. *Phys. Biol.* 5:026009. doi: 10.1088/1478-3975/5/2/026009
- Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* 81, 2340–2361. doi: 10.1021/j100540a008
- Goentoro, L., Shoval, O., Kirschner, M. W., and Alon, U. (2009). The incoherent feedforward loop can provide fold-change detection in gene regulation. *Mol. Cell* 36, 894–899. doi: 10.1016/j.molcel.2009.11.018
- Guo, D., and Li, C. (2009). Stochastic and coherence resonance in feed-forward-loop neuronal network motifs. *Phys. Rev. E* 79:051921. doi: 10.1103/PhysRevE.79.051921
- Gupta, P. B., Fillmore, C. M., Jiang, G., Shapira, S. D., Tao, K., Kuperwasser, C., et al. (2011). Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell* 146, 633–644. doi: 10.1016/j.cell.2011.07.026
- Hager, G. L., McNally, J. G., and Misteli, T. (2009). Transcription dynamics. *Mol. Cell* 35, 741–753. doi: 10.1016/j.molcel.2009.09.005

- Hasegawa, Y., and Struhl, K. (2019). Promoter-specific dynamics of tata-binding protein association with the human genome. *Genome Res.* 29, 1939–1950. doi: 10.1101/gr.254466.119
- Huang, S., Ernberg, I., and Kauffman, S. (2009). Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective. *Semin. Cell Dev. Biol.* 20, 869–876. doi: 10.1016/j.semcdb.2009.07.003
- Hurles, M. (2004). Gene duplication: the genomic trade in spare parts. *PLoS Biol.* 2:e206. doi: 10.1371/journal.pbio.0020206
- Ivey, K. N., and Srivastava, D. (2010). MicroRNAs as regulators of differentiation and cell fate decisions. *Cell Stem Cell* 7, 36–41. doi: 10.1016/j.stem.2010.06.012
- Jones, R. D., Qian, Y., Siciliano, V., DiAndreth, B., Huh, J., Weiss, R., et al. (2020). An endoribonuclease-based feedforward controller for decoupling resource-limited genetic modules in mammalian cells. *Nat. Commun.* 11:5690. doi: 10.1038/s41467-020-19126-9
- Kondrashov, F. A. (2012). Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc. R. Soc. B* 279, 5048–5057. doi: 10.1098/rspb.2012.1108
- Kurtz, T. G. (1971). Limit theorems for sequences of jump Markov processes approximating ordinary differential processes. *J. Appl. Probabil.* 8, 344–356. doi: 10.2307/3211904
- Kurtz, T. G. (1972). The relationship between stochastic and deterministic models for chemical reactions. *J. Chem. Phys.* 57, 2976–2978. doi: 10.1063/1.1678692
- Kuwahara, H., and Mura, I. (2008). An efficient and exact stochastic simulation method to analyze rare events in biochemical systems. *J. Chem. Phys.* 129:165101. doi: 10.1063/1.2987701
- Lee, T. I., Rinaldi, N. J., Robert, F., Odom, D. T., Bar-Joseph, Z., Gerber, G. K., et al. (2002). Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 298, 799–804. doi: 10.1126/science.1075090
- Liang, J., and Qian, H. (2010). Computational cellular dynamics based on the chemical master equation: A Challenge for understanding complexity. *J. Comput. Sci. Technol.* 25, 154–168. doi: 10.1007/s11390-010-9312-6
- Lickwar, C. R., Mueller, F., Hanlon, S. E., McNally, J. G., and Lieb, J. D. (2012). Genome-wide protein-dna binding dynamics suggest a molecular clutch for transcription factor function. *Nature* 484, 251–255. doi: 10.1038/nature10985
- Lin, Y. T., Hufton, P. G., Lee, E. J., and Potoyan, D. A. (2018). A stochastic and dynamical view of pluripotency in mouse embryonic stem cells. *PLoS Comput. Biol.* 14:e1006000. doi: 10.1371/journal.pcbi.1006000
- Lu, T., Shen, T., Bennett, M. R., Wolynes, P. G., and Hasty, J. (2007). Phenotypic variability of growing cellular populations. *Proc. Natl. Acad. Sci. U.S.A.* 104, 18982–18987. doi: 10.1073/pnas.0706115104
- Lynch, M., and Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science* 290, 1151–1155. doi: 10.1126/science.290.5494.1151
- Ma, W., Trusina, A., El-Samad, H., Lim, W. A., and Tang, C. (2009). Defining network topologies that can achieve biochemical adaptation. *Cell* 138, 760–773. doi: 10.1016/j.cell.2009.06.013
- Mangan, S., and Alon, U. (2003). Structure and function of the feed-forward loop network motif. *Proc. Natl. Acad. Sci. U.S.A.* 100, 11980–11985. doi: 10.1073/pnas.2133841100
- Mangan, S., Itzkovitz, S., Zaslaver, A., and Alon, U. (2006). The incoherent feed-forward loop accelerates the response-time of the gal system of *Escherichia coli*. *J. Mol. Biol.* 356, 1073–1081. doi: 10.1016/j.jmb.2005.12.003
- Mangan, S., Zaslaver, A., and Alon, U. (2003). The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks. *J. Mol. Biol.* 334, 197–204. doi: 10.1016/j.jmb.2003.09.049
- Manuchehrfar, F., Li, H., Tian, W., Ma, A., and Liang, J. (2021). Exact topology of dynamic probability surface of an activated process by persistent homology. *J. Phys. Chem. B* 125, 4667–4680. doi: 10.1021/acs.jpcc.1c00904
- Mayshar, Y., Ben-David, U., Lavon, N., Biancotti, J.-C., Yakir, B., Clark, A. T., et al. (2010). Identification and classification of chromosomal aberrations in human induced pluripotent stem cells. *Cell Stem Cell* 7, 521–531. doi: 10.1016/j.stem.2010.07.017
- Momin, M. S. A., and Biswas, A. (2020). Extrinsic noise of the target gene governs abundance pattern of feed-forward loop motifs. *Phys. Rev. E* 101:052411. doi: 10.1103/PhysRevE.101.052411
- Okawa, S., Nicklas, S., Zickenrott, S., Schwamborn, J., and del Sol, A. (2016). A generalized gene-regulatory network model of stem cell differentiation for predicting lineage specifiers. *Stem Cell Rep.* 7, 307–315. doi: 10.1016/j.stemcr.2016.07.014
- Onichtchouk, D., Geier, F., Polok, B., Messerschmidt, D. M., Massner, R., Wendik, B., et al. (2010). Zebrafish pou5f1-dependent transcriptional networks in temporal control of early development. *Mol. Syst. Biol.* 6:354. doi: 10.1038/msb.2010.9
- Papatsenko, D., Darr, H., Kulakovskiy, I. V., Waghray, A., Makeev, V. J., MacArthur, B. D., et al. (2015). Single-cell analyses of escs reveal alternative pluripotent cell states and molecular mechanisms that control self-renewal. *Stem Cell Rep.* 5, 207–220. doi: 10.1016/j.stemcr.2015.07.004
- Ptashne, M. (2004). *A Genetic Switch: Phage Lambda Revisited, 3rd Edn.* New York, NY: Cold Spring Harbor Laboratory Press.
- Re, A., Corá, D., Taverna, D., and Caselle, M. (2009). Genome-wide survey of microRNA-transcription factor feed-forward regulatory circuits in human. *Mol. Biosci.* 5, 854–867. doi: 10.1039/b900177h
- Schultz, D., Jacob, E. B., Onuchic, J. N., and Wolynes, P. G. (2007). Molecular level stochastic model for competence cycles in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U.S.A.* 104, 17582–17587. doi: 10.1073/pnas.0707965104
- Shalek, A. K., Satija, R., Adiconis, X., Gertner, R. S., Gaublomme, J. T., Raychowdhury, R., et al. (2013). Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* 498, 236–240. doi: 10.1038/nature12172
- Shen-Orr, S. S., Milo, R., Mangan, S., and Alon, U. (2002). Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat. Genet.* 31, 64–68. doi: 10.1038/ng881
- Shiraishi, T., Matsuyama, S., and Kitano, H. (2010). Large-scale analysis of network bistability for human cancers. *PLoS Comput. Biol.* 6:e1000851. doi: 10.1371/journal.pcbi.1000851
- Siegal-Gaskins, D., Grotewold, E., and Smith, G. D. (2009). The capacity for multistability in small gene regulatory networks. *BMC Syst. Biol.* 3(1):1. doi: 10.1186/1752-0509-3-96
- Sorrells, T. R., and Johnson, A. D. (2015). Making sense of transcription networks. *Cell* 161, 714–723. doi: 10.1016/j.cell.2015.04.014
- Swiers, G., Patient, R., and Loose, M. (2006). Genetic regulatory networks programming hematopoietic stem cells and erythroid lineage specification. *Dev. Biol.* 294, 525–540. doi: 10.1016/j.ydbio.2006.02.051
- Tej, S., Gaurav, K., and Mukherji, S. (2019). Small RNA driven feed-forward loop: critical role of sRNA in noise filtering. *Phys. Biol.* 16:046008. doi: 10.1088/1478-3975/ab1563
- Terebus, A., Cao, Y., and Liang, J. (2019). “Sensitivities of regulation intensities in feed-forward loops with multistability,” in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Berlin: IEEE), 1969–1972. doi: 10.1109/EMBC.2019.8856532
- Thomas, P., Popović, N., and Grima, R. (2014). Phenotypic switching in gene regulatory networks. *Proc. Natl. Acad. Sci. U.S.A.* 111, 6994–6999. doi: 10.1073/pnas.1400049111
- Tsang, J., Zhu, J., and van Oudenaarden, A. (2007). MicroRNA-mediated feedback and feedforward loops are recurrent network motifs in mammals. *Mol. Cell* 26, 753–767. doi: 10.1016/j.molcel.2007.05.018
- Tuğrul, M., Paixão, T., Barton, N. H., and Tkačik, G. (2015). Dynamics of transcription factor binding site evolution. *PLoS Genet.* 11:e1005639. doi: 10.1371/journal.pgen.1005639
- Vellela, M., and Qian, H. (2007). A quasistationary analysis of a stochastic chemical reaction: Keizer’s paradox. *Bull. Math. Biol.* 69, 1727–1746. doi: 10.1007/s11538-006-9188-3
- Vellela, M., and Qian, H. (2009). Stochastic dynamics and non-equilibrium thermodynamics of a bistable chemical system: the schl#xf6:gl model revisited. *J. R. Soc. Interface* 6, 925–940. doi: 10.1098/rsif.2008.0476
- Xia, K., Feng, X., Tong, Y., and Wei, G. W. (2015). Persistent homology for the quantitative prediction of fullerene stability. *J. Comput. Chem.* 36, 408–422. doi: 10.1002/jcc.23816
- Xia, K., and Wei, G.-W. (2014). Persistent homology analysis of protein structure, flexibility, and folding. *Int. J. Num. Methods Biomed. Eng.* 30, 814–844. doi: 10.1002/cnm.2655
- Xia, K., and Wei, G.-W. (2015). Multidimensional persistence in biomolecular data. *J. Comput. Chem.* 36, 1502–1520. doi: 10.1002/jcc.23953

- Ye, Y., Kang, X., Bailey, J., Li, C., and Hong, T. (2019). An enriched network motif family regulates multistep cell fate transitions with restricted reversibility. *PLoS Comput. Biol.* 15:e1006855. doi: 10.1371/journal.pcbi.1006855
- Zhang, Y., Liu, N., Lin, W., and Li, C. (2019). Quantifying the interplay between genetic and epigenetic regulations in stem cell development. *N. J. Phys.* 21:103042. doi: 10.1088/1367-2630/ab4c82
- Zhu, X.-M., Yin, L., Hood, L., and Ao, P. (2004a). Calculating biological behaviors of epigenetic states in the phage λ life cycle. *Funct. Integr. Genomics* 4, 188–195. doi: 10.1007/s10142-003-0095-5
- Zhu, X.-M., Yin, L., Hood, L., and Ao, P. (2004b). Robustness, stability and efficiency of phage lambda genetic switch: dynamical structure analysis. *J. Bioinform. Comput. Biol.* 2, 785–817. doi: 10.1142/S0219720004000946

Conflict of Interest: YC was employed by company Merck & Co., Inc., Kenilworth, NJ, United States.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Terebus, Manuchehrfar, Cao and Liang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



A Modeling and Machine Learning Pipeline to Rationally Design Treatments to Restore Neuroendocrine Disorders in Heterogeneous Individuals

Tongli Zhang*

Department of Pharmacology and Systems Physiology, College of Medicine, University of Cincinnati, Cincinnati, OH, United States

OPEN ACCESS

Edited by:

Pavel Kraikivski,
Virginia Tech,
United States

Reviewed by:

Amogh Jalihal,
Harvard Medical School,
United States
Guang Yao,
University of Arizona,
United States

*Correspondence:

Tongli Zhang
zhangtl@ucmail.uc.edu

Specialty section:

This article was submitted to
Systems Biology Archive,
a section of the journal
Frontiers in Genetics

Received: 20 January 2021

Accepted: 11 August 2021

Published: 09 September 2021

Citation:

Zhang T (2021) A Modeling and Machine Learning Pipeline to Rationally Design Treatments to Restore Neuroendocrine Disorders in Heterogeneous Individuals. *Front. Genet.* 12:656508. doi: 10.3389/fgene.2021.656508

Heterogeneity among individual patients presents a fundamental challenge to effective treatment, since a treatment protocol working for a portion of the population often fails in others. We hypothesize that a computational pipeline integrating mathematical modeling and machine learning could be used to address this fundamental challenge and facilitate the optimization of individualized treatment protocols. We tested our hypothesis with the neuroendocrine systems controlled by the hypothalamic–pituitary–adrenal (HPA) axis. With a synergistic combination of mathematical modeling and machine learning (ML), this integrated computational pipeline could indeed efficiently reveal optimal treatment targets that significantly contribute to the effective treatment of heterogeneous individuals. What is more, the integrated pipeline also suggested quantitative information on how these key targets should be perturbed. Based on such ML revealed hints, mathematical modeling could be used to rationally design novel protocols and test their performances. We believe that this integrated computational pipeline, properly applied in combination with other computational, experimental and clinical research tools, can be used to design novel and improved treatment against a broad range of complex diseases.

Keywords: neuroendocrine dysfunction, stress, depression, machine learning, computational psychiatry, computational modeling, post-traumatic stress disorder

INTRODUCTION

Proper response to stress signal is essential to maintain the physiological and psychological health. Upon the stimulation by stress signals, the corticotropin-releasing hormone (CRH) is released from the hypothalamus and results in the release of adrenocorticotrophic hormone (ACTH). Through the circulation system, ACTH then travels to the adrenal glands, binds to ACTH receptors, and stimulates the secretion of corticosteroids such as cortisol. Cortisol then stimulates the increases of glucose concentration in the blood to provide energy to cope with the stresses. The proper functioning of this hypothalamic–pituitary–adrenal (HPA) axis is important for physiological response to stress (Tsigos and Chrousos, 2002; Dunlop and Wong, 2019); while the dysregulation of the HPA axis is closely associated with stress order, such as post trauma stress disorder (PTSD) and depression (Bisson et al., 2015; Yehuda et al., 2015; Shalev et al., 2017). If the dysregulated dynamics of the HPA axis is reversed and the normal dynamics and

function of the HPA axis were restored, it might help with treating stress disorders (Ronaldson et al., 2018; Menke, 2019).

However, the effective restoration of HPA function is challenged by the heterogeneous dynamics of the dysregulated axis in patients with stress disorders. For example, in patients with PTSD, both lower cortisol levels and higher cortisol have been reported. In patients with other stress disorders, the cortisol levels are also reported to be bimodal (Yehuda et al., 1995; Gold and Chrousos, 2002; Bremner et al., 2007; Meewisse et al., 2007).

To cope with this challenge of heterogeneity and facilitate the optimization of treatment protocols that can effectively restore HPA axis dynamics, we explored the potential of an integrated computational pipeline that combines mathematical modeling and machine learning. The computational model incorporates several feedbacks controlling the HPA axis, with which we can computationally scanned the effects of potential targets. Machine learning analysis of the random scanning results then revealed the effective targets and how these targets should be perturbed. These ML derived insights aided us to design novel, optimized treatment protocols, which could be further tested with mathematical models.

Our analysis demonstrated a “*proof of concept*” that an integration of mathematical modeling and machine learning can be used to efficiently explore a heterogeneous patient population and facilitate the design of optimized treatment protocols. In the discussion, we have also commented on the strength and limitation of this computational pipeline and envisioned how it could be used together with other tools to improve clinical treatments of complex diseases.

MATERIALS AND METHODS

Time Series Simulation

Simulations were carried out using the ordinary differential equations built following the standard formula. All parameter values were randomly selected from uniform distributions of broad ranges. Time series simulations were performed using XPPaut,¹ the simulated data were then plotted using Matlab.² The detailed simulation protocols for each figure was described along with the figure.

Classification Tree Analysis

Tree models were run using the model parameters in addition to the steady state values for model components. Trees were computed in R³ using the rpart2 algorithm.

Random Forest Analysis

Random Forest analysis was carried out with the value change of parameters as input features and the outcome (effective or non-effective) as prediction targets. The analysis was performed using the standard package in R (see footnote 3). Permutation feature importance were scaled to the maximum (100%) and plotted.

Implementation of Treatments

All treatments were implemented as transient changes of the model parameters. In the random against random targets, a random target parameter is chosen and then either increased or decreased by a value between zero and 10. In the targeted treatments, the parameters with the top rank were decreased.

Selection of Parameter Ranges

If the ranges of parameter changes were too small (i.e., 5 or 10%), the small changes of parameters result in mostly homogeneous behaviors and the sampling of heterogeneous response was computationally inefficient. With trials and errors in preliminary exploration, we chose all parameters randomly from uniform distributions that ranged 10-fold up and down their basal values (10–1,000%) to sample heterogeneous responses efficiently. Since the patient behavior of interest were already covered by the current ranges, the ranges of the parameter changes were not further expanded.

RESULTS

A Mathematical Model Integrated One Negative Feedback and Two Positive Feedbacks Controlling the HPA Axis

The HPA axis is characterized by a negative feedback: after the increase of the stress signal results in the sequential release of corticotropin-releasing hormone (CRH), adrenocorticotrophic hormone (ACTH), and cortisol, the activated glucocorticoid receptors then represses both CRH and ACTH. This negative feedback has been implemented in previous mathematical models (Gudmand-Hoeyer et al., 2014; Bangsgaard and Ottesen, 2017; Stanojević et al., 2018a).

In addition, the glucocorticoid receptor is characterized by a positive feedback that potentially results in bistable switching (Sriram et al., 2012). Meanwhile, Kim et al. (2016) proposed the positive feedback regulating corticotropin-releasing hormone (CRH) could also result in switch like behavior.

Our current mathematical model has integrated the above mentioned negative and positive feedbacks to generate complex and heterogeneous dynamics, which serve as ideal tests to examine whether our analysis pipeline works. Since both the machine learning tools and modeling tools in our pipeline are applicable to systems with additional components, we expect that our analysis pipeline will be able to continue to provide useful and realistic insights even if the model of the HPA axis is expanded to incorporate more regulatory details than these three feedbacks.

Facilitating the Exploration of Heterogeneity With a Standard Model Formula

The heterogeneity of different individuals could be implemented with different model structures, different model parameter values, or both. However, it would be computationally expensive to explore the heterogeneity by composing a different set of mathematical equations for every single individual. In order to facilitate the exploration of heterogeneity and reduce the

¹<http://www.math.pitt.edu/~bard/xpp/xpp.html>

²<https://www.mathworks.com>

³<https://www.r-project.org/>

computational expense of the computational pipeline, we have adopted a standard formula to describe the model structure. In this way, we can simply change the parameter values to explore the heterogeneity between individuals.

In this standard formula, the dynamics of each model component (x) is described as: $\frac{dX_i}{dt} = \tau_i (F_i - X_i)$, with $F_i = \frac{1}{1 + e^{-\sigma W_i}}$ and $W_i = R_i^0 + \sum_j R_i^j \cdot X_j$.

In which, τ_i describes the time scale of the component change, F_i describes the steady state level of the component, and W_i describes the net regulation received by the component.

In this formula, a positive R_i^j specifies an activating effect, while a negative R_i^j specifies an inhibitory one. R_i^0 sums effects that origin from all other components not explicitly

incorporated in the model, and this parameter can be replaced with additional regulatory terms when the model is expanded. Additional elaboration of this approach and its dynamical properties are available in the literature (Mjolsness et al., 1991; Tyson and Novak, 2010) and our previous publication (Ballweg et al., 2018).

By changing the values of the regulatory parameters (R s), we could conveniently explore the nature (activation or inhibition) as well as the strength of the interactions between the model components. The model simulations then allowed us to explore the dynamical properties of the HPA axis that resulted from the different interactions. The parameters explored in this work and the interaction they regulated have been illustrated in **Figure 1**, the regulatory roles of the parameters are described in **Table 1**. The ordinary differential equations and the initial values of the model parameters are recorded in the **Supplementary Table 1**.

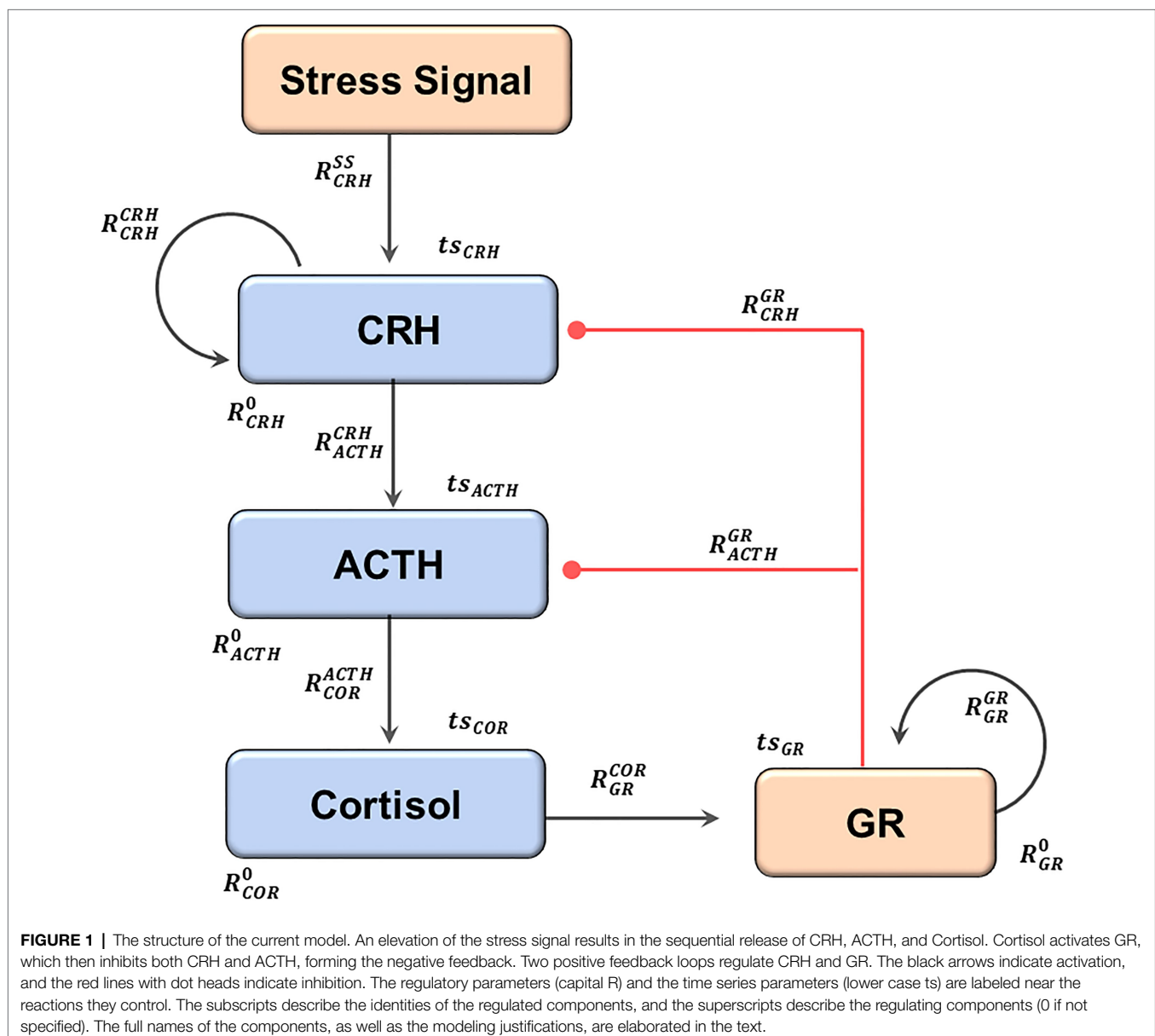


TABLE 1 | The matrix of model parameters.

Regulator	8 Regulatory parameters			
Stress signal	R_{CRH}^{SS}			
CRH	R_{CRH}^{CRH}	R_{ACTH}^{CRH}		
ACTH			R_{COR}^{ACTH}	
Cortisol				R_{GR}^{COR}
GR	R_{CRH}^{GR}	R_{ACTH}^{GR}		R_{GR}^{GR}
Component	CRH	ACTH	Cortisol	GR
8 Component specific parameters				
Background	R_{CRH}^0	R_{ACTH}^0	R_{COR}^0	R_{GR}^0
Time scale	ts_{CRH}	ts_{ACTH}	ts_{COR}	ts_{GR}

Fractional Development of Stress Disorders Within Heterogeneous Individuals

With the standard model formula, we then mimicked a population of heterogeneous individuals by assigning random values to the control parameters. The dynamical simulations of the models represented the dynamical responses of these individuals to stress signals.

A transient elevation of the stress signal (**Figure 2A**) was applied to every individual within the heterogeneous population. In response to this increase of the stress signal, the cortisol levels were transiently elevated in almost all individuals. For the control population, the cortisol levels returned to base line after the stress signal decreased (**Figure 2B**), representing the return to the physiological homeostasis. On the other hand, the cortisol levels sustained at either lower or higher levels represented patients with stress disorders (**Figures 2C,D**).

An analysis with the classification and regression tree (CART) provided an overview of features (including both model component and parameters) characterizing these three different populations (**Figure 2E**). The high cortisol patients are characterized by high levels of CRH and ACTH, which makes mechanistic sense since CRH and ACTH promote cortisol release in the HPA axis. The lower cortisol patients, on the other hand, are characterized by low ACTH and high level of GR. The high level of GR in these patients repress their cortisol release.

The simulated dynamics indicates that the systems changes might be sustained even after the decrease of stress signal, which is a hallmark of post trauma stress disorder (PTSD). Upon the exposure to transient traumatic events, the symptoms in PTSD patients could sustain for a long time. Many military personnel suffer from stress disorder years after departing from the battle field (O'Toole et al., 2009; Marmar et al., 2015; Armenta et al., 2018).

Our simulations used a strong stress signal, which was necessary to trigger cortisol disorders. If the stress signal were reduced, less simulated individuals developed cortisol disorder. Meanwhile, even for the strong stress signal used in our simulation, not all the affected individuals developed cortisol disorder. Rather, a large portion of the stimulated individuals (>50%) could recover their physiological cortisol levels after the stress disappears. The low percentage of stress disorder development was also reported in the literature. The World Health Organization World Mental Health Surveys reported a cross-national lifetime prevalence rate of 3.9% (Koenen et al., 2017); while a National Epidemiologic Survey reported a lifetime prevalence rate of 6.1% in the United States (Goldstein et al., 2016).

Our simulations indicated that the heterogeneous levels of cortisol, either lower or higher, might naturally emerge after heterogeneous individuals were stimulated with stress signal. The heterogeneous cortisol levels in the patient populations were consistent to the literature reports (Yehuda et al., 1995; Gold and Chrousos, 2002; Bremner et al., 2007; Meewisse et al., 2007). The heterogeneous levels of cortisol made sense when we examined the structure of the mathematical model. The level of cortisol was regulated by a combination of two positive feedbacks and one negative feedbacks. The positive feedbacks allowed the model to have the potential to generate different attracting stable steady states, which potentially could explain the distinction between the higher cortisol population, the lower cortisol population, and the control.

Meanwhile, the negative feedback had the potential to generate sustained oscillations (one example illustrated in **Figure 2D**). The regulation and dysregulation of the oscillatory dynamics of cortisol might also play a role in the response to stress stimulation (Kalafatakis et al., 2018; Stanojević et al., 2018b; Lightman et al., 2020).

With the intervened positive and negative feedbacks, the current model has been able to mimic the complex, heterogeneous dynamics of different individuals who develop stress disorder, hence we proceed to use the current model to test our integrated computational pipeline.

Random Scanning of Treatment Targets

We first selected around 30,000 patients with lower pathological cortisol and subjected them to for potential treatments that were applied on the control parameters. Since we lacked both qualitative information (what targets should be targeted) as well as quantitative information (to what levels should the targets be changed), we randomly selected an individual parameter target and changed its level randomly and transiently (details in Materials and Methods). Five representative trajectories of effective treatments were illustrated in **Figure 3A**. These five individuals started with normal, physiological levels of cortisol. After transient stimulation by the stress signal (applied around time 20), their cortisol levels were decreased to lower, pathological levels. After the effective treatments, their cortisol levels were restored to normal, physiological ones.

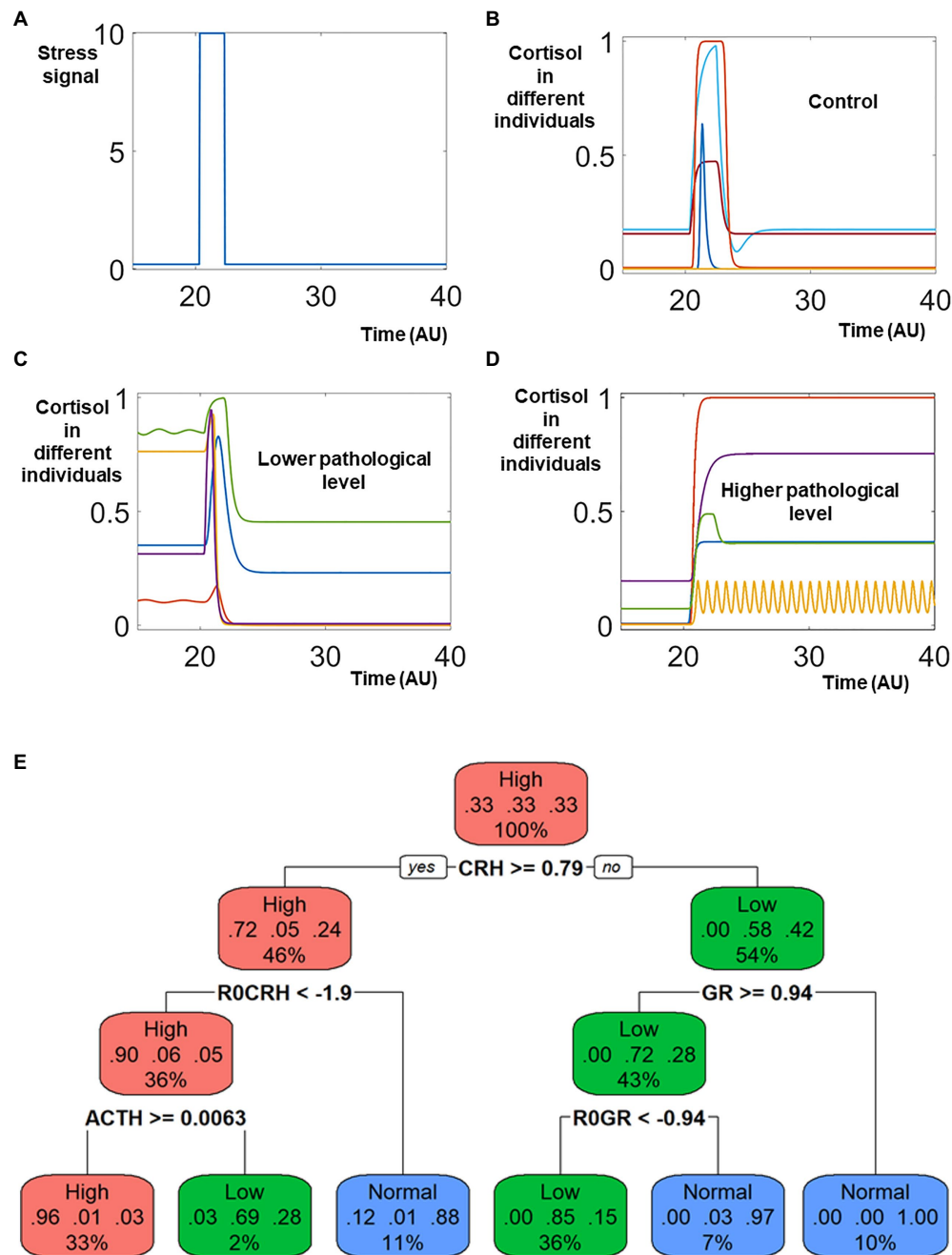


FIGURE 2 | Sampling the heterogeneous response triggered by stress signal. Identical model structure and different values of all parameters are used to simulate the heterogeneous response to stress signal. In response to a transient elevation of the stress signal (around time 20, **A**), the cortisol levels in the control population (**B**) returned to normal levels after transient elevation, representing the healthy population who do not develop neuron endocrine disorder after stress. On the contrary, the levels of cortisol decreased and were sustained lower in patient populations who were characterized by lower pathological levels of cortisol (**C**). The cortisol levels were elevated and remained higher in patients whose pathological cortisol levels were higher after stress (**D**). (**E**) Classification Tree. The different colors of different nodes indicate the types of dominant populations: red nodes indicates that most individual in the node had higher cortisol after stress; green nodes indicates that most individuals were characterized by lower levels of cortisol; blue nodes indicate the ones with most individuals from the control population. In the top node, individuals from these three populations were of identical number, and the node is labeled red.

On the contrary, the cortisol levels were not restored in five individuals representing patients receiving ineffective treatments (Figure 3B).

Such bimodal responses were also observed in around 30,000 individuals whose pathological cortisol levels were higher than their physiological ones. In five representing individuals whose

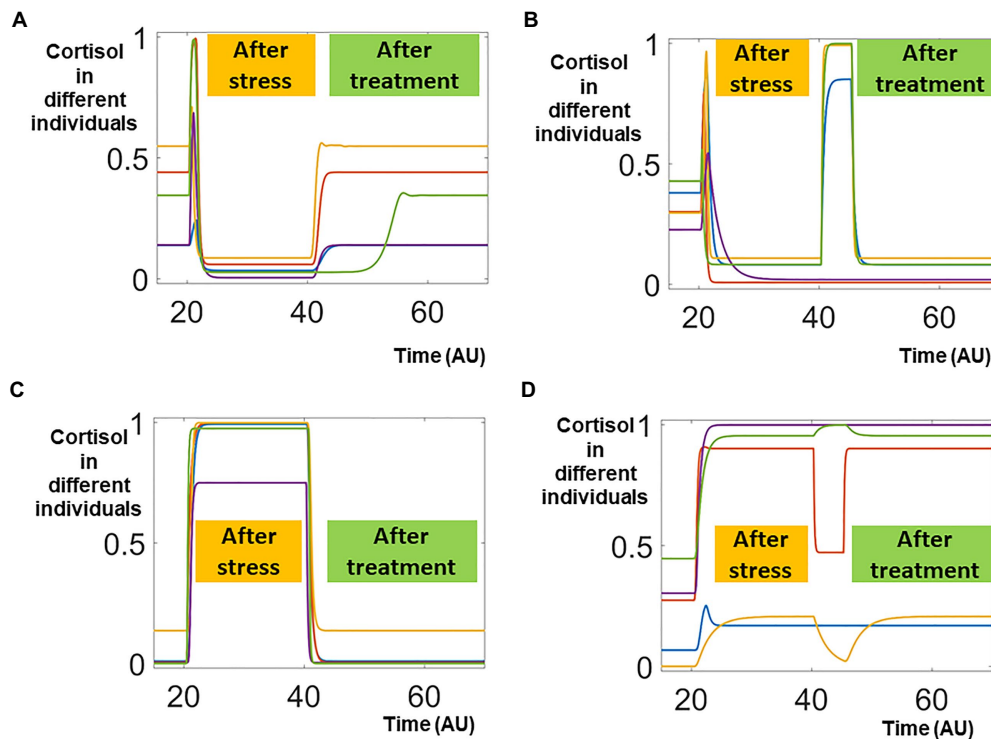


FIGURE 3 | Sampling effective and non-effective treatments. In this random scanning, one parameter is randomly perturbed around time 40. The effective treatments, which restored the cortisol distorted by the stress signal, were shown in panels (A,C). On the other hand, the non-effective treatments, which failed to restore the cortisol levels, were shown in panels (B,D).

cortisol levels were sustained higher after the transient stress signal, effective treatment restored their cortisol levels to physiological ones (Figure 3C). On the other hand, the cortisol levels remained higher after ineffective treatments (Figure 3D).

Treatments against random targets were effective in small portions of the patients but ineffective in the majorities of the treated individuals. For our pipeline, even the ineffective treatments provided important information on how the system responded to perturbations. Hence, proceeded to apply machine learning analysis on the simulation data that included both effective and ineffective treatments. In the future, we envision that our pipeline could potentially be applied to the clinical data that combined effective and ineffective treatments, to improve the design of clinical trials and treatment protocols.

Improving the Treatment Protocols

The individual responses to random treatments were binary (Effective vs. Non-Effective), and such binary data were fed into the Random Forest (RF) analysis. RF identified the most influential factors that distinguished the effective treatments from the non-effective ones. By computing the consequential error resulted from permuting features, the RF analysis also ranked the relative importance of all potential the targets.

The RF analysis revealed that for individuals whose pathological cortisol levels were lower, the changes of R_{GR}^0 and R_{GR}^{GR} were most important for effective treatments

(Figure 4A). Furthermore, the analysis with decision tree provided quantitative information on how the targets should be changed (Figure 4B). In the treatments that sufficiently decreased the level of R_{GR}^{GR} , most of them would be effective (node on the bottom left in Figure 4B); similarly, most of the treatments that decreased the level of R_{GR}^0 were also effective (2nd node from left on the bottom of Figure 4B).

These RF identified targets, R_{GR}^0 and R_{GR}^{GR} , make mechanistic sense in the context of the model structure. A positive feedback controlling GR might result in pathological steady state with lower cortisol levels. Hence, it is reasonable to expect that the pathological states could be reverted by targeting the key parameters that control the positive feedback. On the basis of these qualitative and quantitative information, we designed targets treatments against the top targets, R_{GR}^0 or R_{GR}^{GR} . Simulation showed that transient decrease of these two targets were able to effectively treat patients whose pathological cortisol levels were lower. These patients were characterized with lower levels of cortisol after their stimulation by transient stress signal (applied around time 20). Then, their cortisol levels were restored by the transient decrease of R_{GR}^0 or R_{GR}^{GR} applied around time 60 (Figures 4C,D).

Similarly, RF analysis indicated that R_{CRH}^0 and R_{CRH}^{CRH} were the top targets in patients whose pathological cortisol levels were higher (Figure 5A). The decision tree analysis suggested that these two targets should be decreased in effective treatments (Figure 5B), and rationally designed treatments based on such

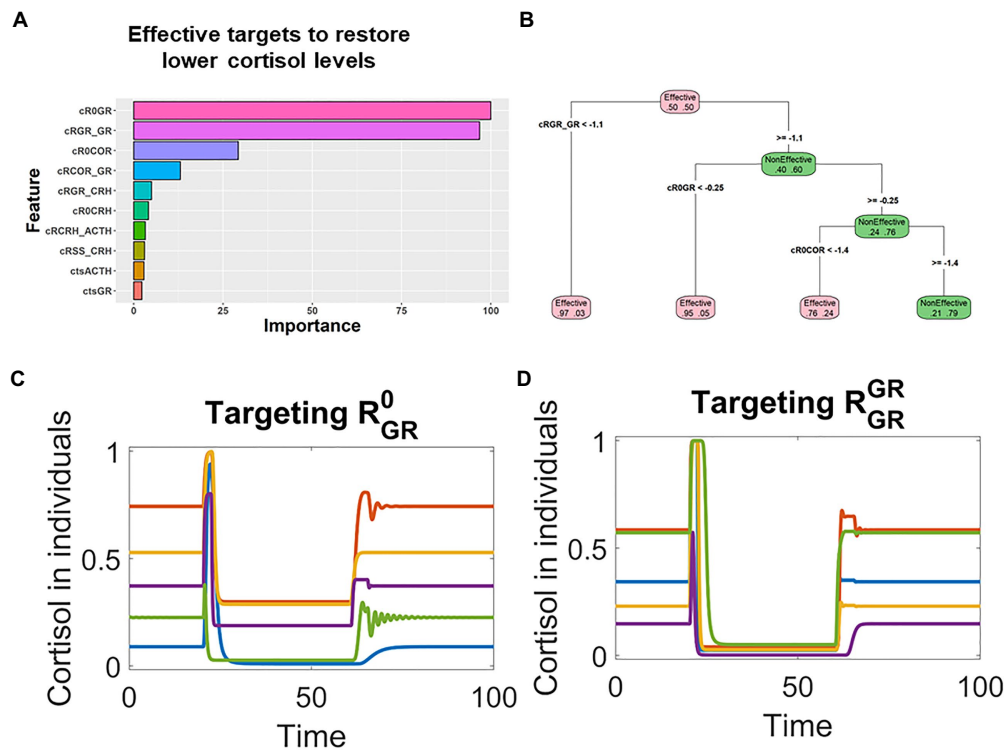


FIGURE 4 | Machine learning analysis revealed that the key parameters controlling the GR positive feedback were associated with effective treatments for patients with lower cortisol levels. **(A)**, Random forest analysis ranked the targets based on their association with the effective treatments in patients whose pathological cortisol levels were lower. **(B)**, Decision tree analysis indicated that these parameters were decreased in effective treatments. **(C,D)** Time series simulations illustrated how the decrease of R_{GR}^0 or R_{GR}^{GR} restored the cortisol levels in patients with lower pathological cortisol levels.

information could effectively restore the levels in these patients. After stimulation by transient stress signal (applied around time 20), cortisol levels were sustained higher in these individual patients. Then, R_{CRH}^0 and R_{CRH}^{CRH} were transiently decreased around time 60, the cortisol levels were decreased and remained low (Figures 5C,D).

It was encouraging that the treatments effects were sustained even though the treatments against these highly ranked targets were only transient. In clinical terms, this means that it would be possible to fully cure the patients suffering stress disorder if the correct targets were identified and perturbed.

Hence, starting with scanning treatments against random targets, machine learning analysis with the scanning results could lead us to rationally design novel and improved treatments. What is more, since the machine learning analysis revealed more than one target that could contribute to the effective treatments, it would be plausible to design multiple effective treatments and select the practical ones based on clinical constraints.

DISCUSSION

Heterogeneity within individual patients underlies partial responses to treatment and calls for the design of personalized

and optimized treatment protocols. In this work, we have demonstrated the performance of a computational pipeline that integrated mathematical modeling and machine learning. The pipeline was able to address this fundamental challenge of heterogeneity: starting with little qualitative clue (target identification) and quantitative clue (perturbation strength), the pipeline was able to deliver rational designs of effective treatment plans that clearly answered “*what to targets?*” and “*how much to change?*”

With such “*proof of principle*,” we hope that the computational pipeline could be integrated into clinical practice to design novel and more effective treatments for complex diseases. We envision that after clinical data were fed into this analysis pipeline, it would lead to insights that are clinically applicable.

Two theoretical approaches, one driven by data and the other based on mechanism, have been widely applied in the field of systems biology and quantitative systems pharmacology. In this work, we have illustrated that these two approaches could be integrated together to achieve synergistic effects (Figure 6): the machine learning methods could be used to efficiently extract insights from heterogeneous behaviors, while the mechanistic models could be used to design mechanistic and dynamical protocols that are directly translatable on the basis of the machine learning revealed insight.

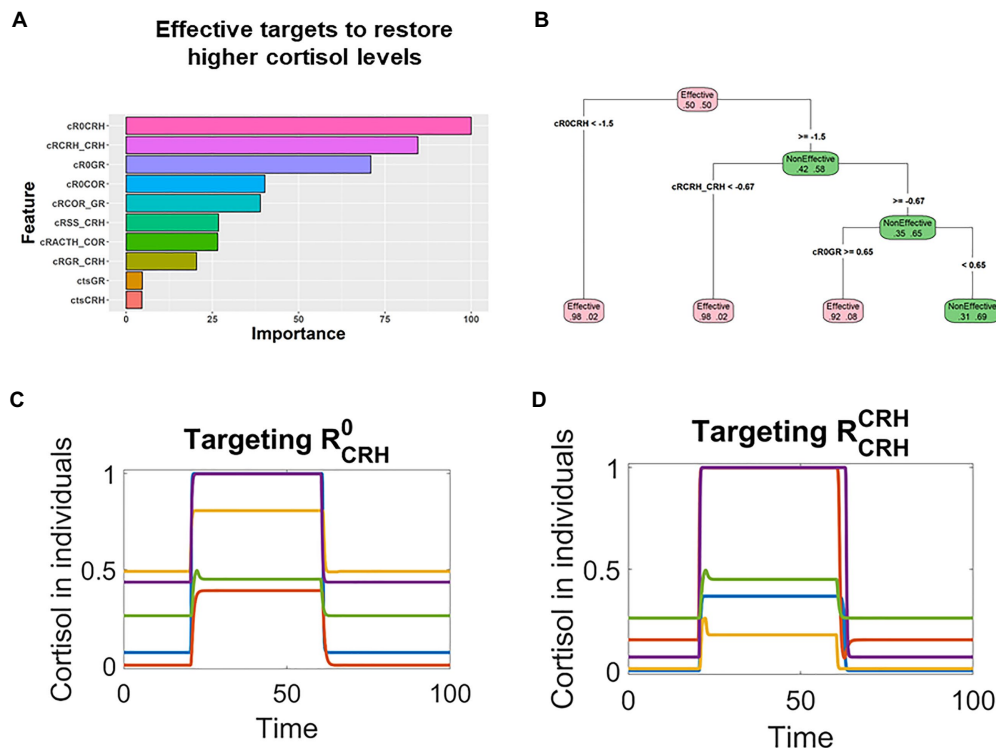
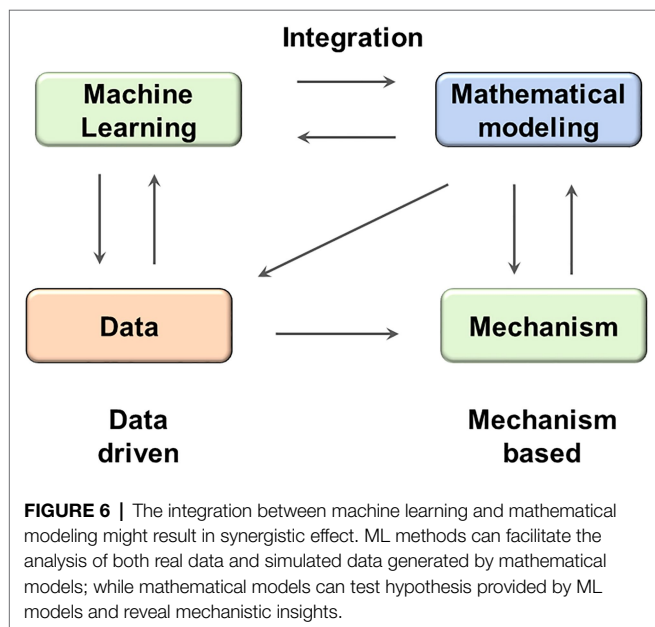


FIGURE 5 | Machine learning analysis revealed that parameters controlling the CRH positive feedback were associated with effective treatments for patients with higher cortisol levels. **(A)** Random forest analysis ranked the targets based on their association with the effective treatments in patients whose pathological cortisol levels were higher. **(B)** Decision tree analysis indicated that these key parameters should be decreased for the treatments to be effective. **(C,D)** Time series simulations illustrated how the decrease of R_{CRH}^0 or R_{CRH}^{CRH} restored the cortisol levels in patients who had higher levels of cortisol after stress stimulation.



The machine learning methods could help modeling to be more efficient. It would be computationally expensive to carry out a comprehensive scanning with all possible targets

and all possible perturbation values, in comparison, a random scanning followed by machine learning analysis in our work was able to efficiently yield hints on some top ranked targets and how they should be perturbed. Meanwhile, modeling puts the insights extracted with machine learning back into the mechanistic context and could help designing novel protocols directly translatable to clinical practice.

Consistent with its purpose of showing a “proof of principle,” the current model has represented the heterogeneous patient populations only with the qualitative changes (either lower or higher) of their cortisol levels. As the next step, it would be beneficial to fit the model parameters with cortisol levels observed in individual patients (Bangsgaard and Ottesen, 2017), so that the heterogeneous models could represent individual patients in quantitative details.

With the continuous contribution from the communities of systems biology and quantitative systems pharmacology, we expect the simplified model would be expanded to more realistic, multi-scale ones that include more biochemical, genetic, epigenetic, molecular, cellular, and neurological details. This process would be time and effort consuming; however, the overall process could be facilitated by taking advantage of the existing models that have been developed to describe HPA axis and its role in stress orders. For example, our current work has been benefiting from the modeling works

or Sriram et al. (2012) and Kim et al. (2016). We envision the further expansion of the current model will also be able to utilize many other modeling works in the field of computational psychiatry, such as the PKA-PP2A model of fear conditioning (Yang et al., 2010), the model for protein kinase M feedback (Ogasawara and Kawato, 2010), and the modeling work on the positive feedback loop controlling brain-derived neurotrophic factor (BDNF; Bambah-Mukku et al., 2014; Zhang et al., 2016). The incorporation of additional regulatory pathways might result in the co-existence of even more attractors, which would further increase the heterogeneous subtypes of stress disorder patients. Combination of different machine learning algorithms, including those used here, promises to facilitate the analysis of these additional subtypes.

In addition, we expect that the expanded models of stress disorders will also integrate multi-scale neural circuits within the corresponding regions of the brain (Smith, 2005), the pharmacokinetics of various drugs such as the selective serotonin reuptake inhibitors sertraline (Zoloft) and paroxetine (Paxil; Alhadab and Brundage, 2020; Heydorn, 1999), as well as the pharmacodynamic effect of these drugs such as the serotonin production and regulation (Best et al., 2010).

Though it is going to be time consuming and effort consuming to develop such realistic models with elaborated biological and pharmaceutical control details, we expect that the effort will eventually pay off and the realistic models will make contributions of clinical significance. For example, the realistic models may be able to guide us to further understand the genetic and biochemical basis of different patients whose cortisol levels are either lower or higher when developing depression; these models may point out to optimized targets for patients who are not responding to the currently available treatments; also, comprehensive models will have the potential to aid us to examine whether novel treatments would result in undesired side effects or toxicities in healthy, control populations.

Computational psychiatry promises to address some of the hard challenges faced by psychobiological researchers, and encouraging results have been accumulating along this direction (Ferrante and Gordon, 2021; Huys et al., 2021). From a methodological perspective, we have tested an integrated computational pipeline (ICP) that combines computational modeling and machine learning and shown “*proof of principle*” that this pipeline could be used to aid with the design of

novel treatment protocols which can effectively restore neuroendocrine dysregulation in a population of heterogeneous individuals. We expect that the further expansion of the model as well as this pipeline would be able to deliver more clinically useful insights for psychological disorders.

What is more, this computational framework of integrated modeling and machine learning can be readily applied to other research areas beyond neuroendocrine and psychological disorders. The field of computational medicine and quantitative systems pharmacology have already started to integrate complimentary tools to achieve greater benefits (Hutchinson et al., 2019; Zhang et al., 2019; Benzekry, 2020), and we believe that the broader application of our pipeline will contribute to the design of novel and effective treatments for a board range of complex diseases.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

The work was supported by grant 1016183 ARMY W911NF-20-1-0192 to TZ.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.656508/full#supplementary-material>

Supplementary Table 1 | The ordinary differential equations and basal values of the model parameters.

REFERENCES

- Alhadab, A. A., and Brundage, R. C. (2020). Physiologically-based pharmacokinetic model of sertraline in human to predict clinical relevance of concentrations at target tissues. *Clin. Pharmacol. Ther.* 108, 136–144. doi: 10.1002/cpt.1824
- Armenta, R. F., Rush, T., LeardMann, C. A., Millegan, J., Cooper, A., Hoge, C. W., et al. (2018). Factors associated with persistent posttraumatic stress disorder among U.S. military service members and veterans. *BMC Psychiatry* 18:48. doi: 10.1186/s12888-018-1590-5
- Ballweg, R., Lee, S., Han, X., Maini, P. K., Byrne, H., Hong, C. I., et al. (2018). Unraveling the control of cell cycle periods during intestinal stem cell differentiation. *Biophys. J.* 115, 2250–2258. doi: 10.1016/j.bpj.2018.10.025
- Bambah-Mukku, D., Travaglia, A., Chen, D. Y., Pollonini, G., and Alberini, C. M. (2014). A positive autoregulatory BDNF feedback loop via C/EBPβ mediates hippocampal memory consolidation. *J. Neurosci.* 34, 12547–12559. doi: 10.1523/JNEUROSCI.0324-14.2014
- Bangsgaard, E. O., and Ottesen, J. T. (2017). Patient specific modeling of the HPA axis related to clinical diagnosis of depression. *Math. Biosci.* 287, 24–35. doi: 10.1016/j.mbs.2016.10.007
- Benzekry, S. (2020). Artificial intelligence and mechanistic modeling for clinical decision making in oncology. *Clin. Pharmacol. Ther.* 108, 471–486. doi: 10.1002/cpt.1951
- Best, J., Nijhout, H. F., and Reed, M. (2010). Serotonin synthesis, release and reuptake in terminals: a mathematical model. *Theor. Biol. Med. Model.* 7:34. doi: 10.1186/1742-4682-7-34

- Bisson, J. I., Cosgrove, S., Lewis, C., and Robert, N. P. (2015). Post-traumatic stress disorder. *BMJ* 351:h6161. doi: 10.1136/bmj.h6161
- Bremner, D., Vermetten, E., and Kelley, M. E. (2007). Cortisol, dehydroepiandrosterone, and estradiol measured over 24 hours in women with childhood sexual abuse-related posttraumatic stress disorder. *J. Nerv. Ment. Dis.* 195, 919–927. doi: 10.1097/NMD.0b013e3181594ca0
- Dunlop, B. W., and Wong, A. (2019). The hypothalamic-pituitary-adrenal axis in PTSD: pathophysiology and treatment interventions. *Prog. Neuro-Psychopharmacol. Biol. Psychiatry* 89, 361–379. doi: 10.1016/j.pnpbp.2018.10.010
- Ferrante, M., and Gordon, J. A. (2021). Computational phenotyping and longitudinal dynamics to inform clinical decision-making in psychiatry. *Neuropsychopharmacology* 46, 243–244. doi: 10.1038/s41386-020-00852-z
- Gold, P. W., and Chrousos, G. P. (2002). Organization of the stress system and its dysregulation in melancholic and atypical depression: high vs low CRH/NE states. *Mol. Psychiatry* 7, 254–275. doi: 10.1038/sj.mp.4001032
- Goldstein, R. B., Smith, S. M., Chou, S. P., Saha, T. D., Jung, J., Zhang, H., et al. (2016). The epidemiology of DSM-5 posttraumatic stress disorder in the United States: results from the National Epidemiologic Survey on Alcohol and Related Conditions-III. *Soc. Psychiatry Psychiatr. Epidemiol.* 51, 1137–1148. doi: 10.1007/s00127-016-1208-5
- Gudmand-Hoyer, J., Timmermann, S., and Ottesen, J. T. (2014). Patient-specific modeling of the neuroendocrine HPA-axis and its relation to depression: Ultradian and circadian oscillations. *Math. Biosci.* 257, 23–32. doi: 10.1016/j.mbs.2014.07.013
- Heydorn, W. E. (1999). Paroxetine: a review of its pharmacology, pharmacokinetics and utility in the treatment of a variety of psychiatric disorders. *Expert Opin. Investig. Drugs* 8, 417–441. doi: 10.1517/13543784.8.4.417
- Hutchinson, L., Steiert, B., Soubret, A., Wagg, J., Phipps, A., Peck, R., et al. (2019). Models and machines: how deep learning will take clinical pharmacology to the next level. *CPT Pharmacometrics Syst. Pharmacol.* 8, 131–134. doi: 10.1002/psp4.12377
- Huys, Q. J. M., Browning, M., Paulus, M. P., and Frank, M. J. (2021). Advances in the computational understanding of mental illness. *Neuropsychopharmacology* 46, 3–19. doi: 10.1038/s41386-020-0746-4
- Kalafatakis, K., Russell, G. M., Harmer, C. J., Munafo, M. R., Marchant, N., Wilson, A., et al. (2018). Ultradian rhythmicity of plasma cortisol is necessary for normal emotional and cognitive responses in man. *Proc. Natl. Acad. Sci. U. S. A.* 115, E4091–E4100. doi: 10.1073/pnas.1714239115
- Kim, L. U., D'Orsogna, M. R., and Chou, T. (2016). Onset, timing, and exposure therapy of stress disorders: mechanistic insight from a mathematical model of oscillating neuroendocrine dynamics. *Biol. Direct* 11:13. doi: 10.1186/s13062-016-0117-6
- Koenen, K. C., Ratanatharathorn, A., Ng, L., McLaughlin, K. A., Bromet, E. J., Stein, D. J., et al. (2017). Posttraumatic stress disorder in the world mental health surveys. *Psychol. Med.* 47, 2260–2274. doi: 10.1017/S0033291717000708
- Lightman, S. L., Birnie, M. T., and Conway-Campbell, B. L. (2020). Dynamics of ACTH and cortisol secretion and implications for disease. *Endocr. Rev.* 41, 470–490. doi: 10.1210/edrv/bnaa002
- Marmar, C. R., Schlenger, W., Henn-Haase, C., Qian, M., Purchia, E., Li, M., et al. (2015). Course of posttraumatic stress disorder 40 years after the Vietnam war: findings from the National Vietnam Veterans Longitudinal Study. *JAMA Psychiatr.* 72, 875–881. doi: 10.1001/jamapsychiatry.2015.0803
- Meewisse, M. L., Reitsma, J. B., de Vries, G. J., Gersons, B. P., and Olff, M. (2007). Cortisol and post-traumatic stress disorder in adults: systematic review and meta-analysis. *Br. J. Psychiatry* 191, 387–392. doi: 10.1192/bjp.bp.106.024877
- Menke, A. (2019). Is the HPA axis as target for depression outdated, or is there a new hope? *Front. Psych.* 10:101. doi: 10.3389/fpsyg.2019.00101
- Mjolsness, E., Sharp, D. H., and Reinitz, J. (1991). A connectionist model of development. *J. Theor. Biol.* 152, 429–453. doi: 10.1016/S0022-5193(05)80391-1
- Ogasawara, H., and Kawato, M. (2010). The protein kinase Mzeta network as a bistable switch to store neuronal memory. *BMC Syst. Biol.* 4:181. doi: 10.1186/1752-0509-4-181
- O'Toole, B. I., Catts, S. V., Outram, S., Pierse, K. R., and Cockburn, J. (2009). The physical and mental health of Australian Vietnam veterans 3 decades after the war and its relation to military service, combat, and post-traumatic stress disorder. *Am. J. Epidemiol.* 170, 318–330. doi: 10.1093/aje/kwp146
- Ronaldson, A., Carvalho, L. A., Kostich, K., Lazzarino, A. I., Urbanova, L., and Steptoe, A. (2018). The effects of six-day SSRI administration on diurnal cortisol secretion in healthy volunteers. *Psychopharmacology* 235, 3415–3422. doi: 10.1007/s00213-018-5050-1
- Shalev, A., Liberzon, I., and Marmar, C. (2017). Post-traumatic stress disorder. *N. Engl. J. Med.* 376, 2459–2469. doi: 10.1056/NEJMra1612499
- Smith, M. E. (2005). Bilateral hippocampal volume reduction in adults with post-traumatic stress disorder: a meta-analysis of structural MRI studies. *Hippocampus* 15, 798–807. doi: 10.1002/hipo.20102
- Sriram, K., Rodriguez-Fernandez, M., and Doyle, F. J. 3rd (2012). Modeling cortisol dynamics in the neuro-endocrine axis distinguishes normal, depression, and post-traumatic stress disorder (PTSD) in humans. *PLoS Comput. Biol.* 8:e1002379. doi: 10.1371/journal.pcbi.1002379
- Stanojević, A., Marković, V. M., Čupić, Ž., Kolar-Anić, L., and Vukojević, V. (2018a). Advances in mathematical modelling of the hypothalamic-pituitary-adrenal (HPA) axis dynamics and the neuroendocrine response to stress. *Curr. Opin. Chem. Eng.* 21, 84–95. doi: 10.1016/j.coche.2018.04.003
- Stanojević, A., Marković, V. M., Mačević, S., Kolar-Anić, L., and Vukojević, V. (2018b). Kinetic modelling of testosterone-related differences in the hypothalamic-pituitary-adrenal axis response to stress. *React. Kinet. Mech. Catal.* 123, 17–30. doi: 10.1007/s11444-017-1315-7
- Tsigos, C., and Chrousos, G. P. (2002). Hypothalamic-pituitary-adrenal axis, neuroendocrine factors and stress. *J. Psychosom. Res.* 53, 865–871. doi: 10.1016/S0022-3999(02)00429-4
- Tyson, J. J., and Novak, B. (2010). Functional motifs in biochemical reaction networks. *Annu. Rev. Phys. Chem.* 61, 219–240. doi: 10.1146/annurev.physchem.012809
- Yang, R., Sriram, K., and Doyle, F. J. (2010). “Control circuitry for fear conditioning associated with post-traumatic stress disorder (PTSD).” in *49th IEEE Conference on Decision and Control (CDC)*; December 15–17, 2010; Atlanta, GA, USA (IEEE), 2541–2546.
- Yehuda, R., Hoge, C. W., McFarlane, A. C., Vermetten, E., Lanius, R. A., Nievergelt, C. M., et al. (2015). Post-traumatic stress disorder. *Nat. Rev. Dis. Primers.* 1:15057. doi: 10.1038/nrdp.2015.57
- Yehuda, R., Kahana, B., Binder-Brynes, K., Southwick, S. M., Mason, J. W., and Giller, E. L. (1995). Low urinary cortisol excretion in holocaust survivors with posttraumatic stress disorder. *Am. J. Psychiatry* 152, 982–986. doi: 10.1176/ajp.152.7.982
- Zhang, Y., Huynh, J. M., Liu, G. S., Ballweg, R., Aryeh, K. S., Paek, A. L., et al. (2019). Designing combination therapies with modeling chaperoned machine learning. *PLoS Comput. Biol.* 15:e1007158. doi: 10.1371/journal.pcbi.1007158
- Zhang, Y., Smolen, P., Alberini, C. M., Baxter, D. A., and Byrne, J. H. (2016). Computational model of a positive BDNF feedback loop in hippocampal neurons following inhibitory avoidance training. *Learn. Mem.* 23, 714–722. doi: 10.1101/lm.042044.116

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read
for greatest visibility
and readership



FAST PUBLICATION

Around 90 days
from submission
to decision



HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,
and constructive
peer-review



TRANSPARENT PEER-REVIEW

Editors and reviewers
acknowledged by name
on published articles

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org

Contact us: frontiersin.org/about/contact



REPRODUCIBILITY OF RESEARCH

Support open data
and methods to enhance
research reproducibility



DIGITAL PUBLISHING

Articles designed
for optimal readership
across devices



FOLLOW US

@frontiersin



IMPACT METRICS

Advanced article metrics
track visibility across
digital media



EXTENSIVE PROMOTION

Marketing
and promotion
of impactful research



LOOP RESEARCH NETWORK

Our network
increases your
article's readership