

# A logic-incorporated gene regulatory network deciphers principles in cell fate decisions

Gang Xue<sup>1</sup>, Xiaoyi Zhang<sup>2</sup>, Wanqi Li<sup>1</sup>, Lu Zhang<sup>2</sup>, Zongxu Zhang<sup>2</sup>, Xiaolin Zhou<sup>1</sup>, Di Zhang<sup>2</sup>, Lei Zhang<sup>2,3\*</sup>, Zhiyuan Li<sup>1,2\*</sup>

<sup>1</sup>Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China; <sup>2</sup>Center for Quantitative Biology, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China; <sup>3</sup>Beijing International Center for Mathematical Research, Center for Machine Learning Research, Peking University, Beijing, China

**Abstract** Organisms utilize gene regulatory networks (GRN) to make fate decisions, but the regulatory mechanisms of transcription factors (TF) in GRNs are exceedingly intricate. A longstanding question in this field is how these tangled interactions synergistically contribute to decision-making procedures. To comprehensively understand the role of regulatory logic in cell fate decisions, we constructed a logic-incorporated GRN model and examined its behavior under two distinct driving forces (noise-driven and signal-driven). Under the noise-driven mode, we distilled the relationship among fate bias, regulatory logic, and noise profile. Under the signal-driven mode, we bridged regulatory logic and progression-accuracy trade-off, and uncovered distinctive trajectories of reprogramming influenced by logic motifs. In differentiation, we characterized a special logic-dependent priming stage by the solution landscape. Finally, we applied our findings to decipher three biological instances: hematopoiesis, embryogenesis, and trans-differentiation. Orthogonal to the classical analysis of expression profile, we harnessed noise patterns to construct the GRN corresponding to fate transition. Our work presents a generalizable framework for top-down fate-decision studies and a practical approach to the taxonomy of cell fate decisions.

## eLife assessment

The study presented in this manuscript makes **important** contributions to our understanding of cell fate decisions and the role of noise in gene regulatory networks. Through computational and theoretical analysis, the authors provide **solid** support for distinguishing distinct driving forces behind fate decisions based on noise profiles and reprogramming trajectories. While acknowledging the potential limitations of small gene regulatory networks in capturing the richness of whole-transcriptome sequencing datasets, this study offers a creative approach for formulating hypotheses about gene regulation during stem cell differentiation using single-cell sequencing data.

## Introduction

Waddington's epigenetic landscape is a fundamental and profound conceptualization of cell fate decisions (*Waddington, 1957*). Over decades, this insightful metaphor has facilitated researchers to distill a myriad of models regarding cell fate decisions (*MacArthur, 2022; Schiebinger et al., 2019; Olsson et al., 2016; Bhattacharya et al., 2011; Fisher and Merckenschlager, 2010; Hota et al., 2022; Huang, 2012; Li et al., 2021*). While introducing various quantitative models and dissecting diverse

### \*For correspondence:

zhangl@math.pku.edu.cn (LZ);  
zhiyuanli@pku.edu.cn (ZL)

**Competing interest:** The authors declare that no competing interests exist.

**Funding:** See page 28

**Preprint posted**

21 April 2023

**Sent for Review**

05 May 2023

**Reviewed preprint posted**

27 September 2023

**Reviewed preprint revised**

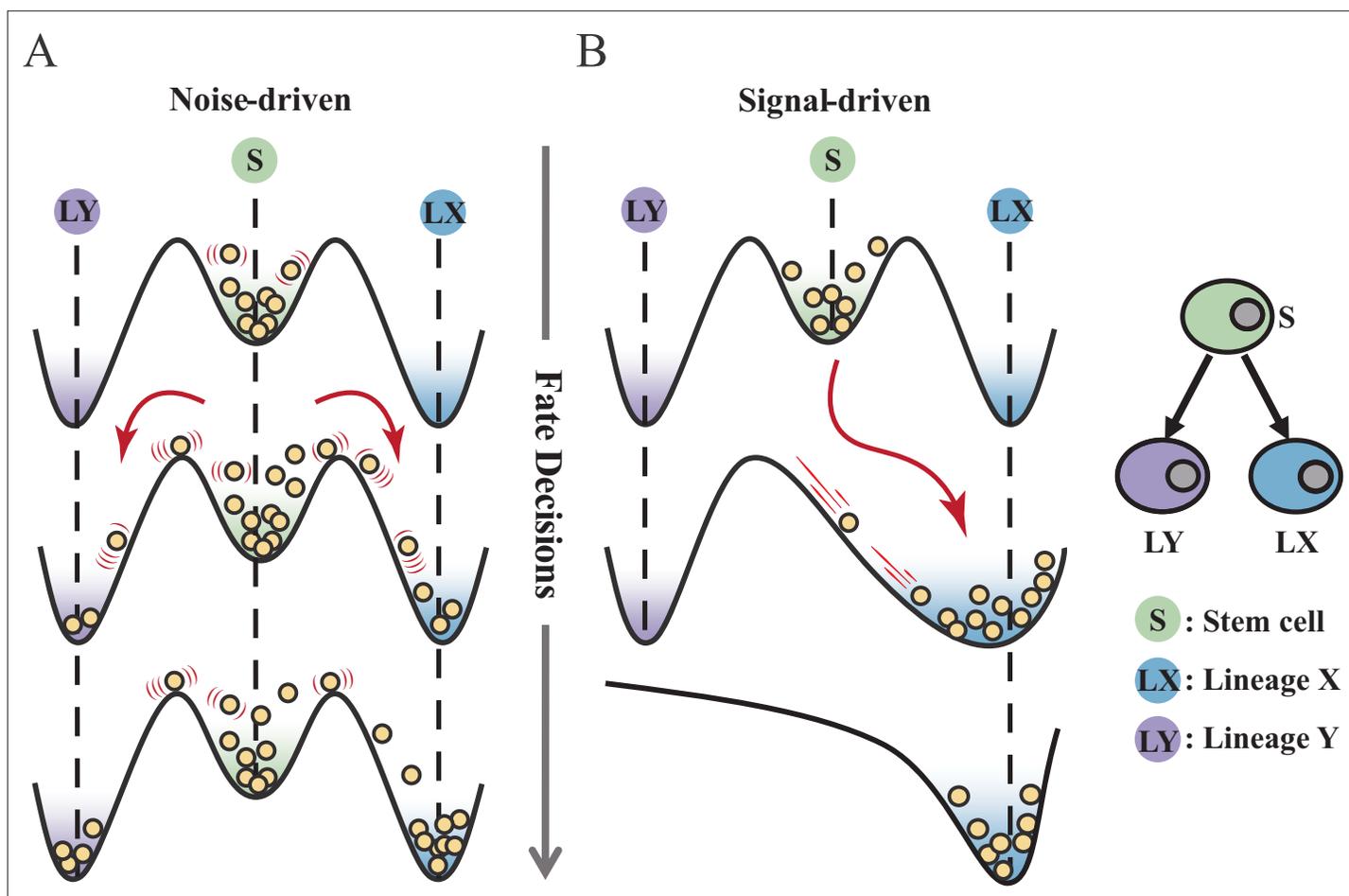
26 March 2024

**Version of Record published**

23 April 2024

**Reviewing Editor:** Mariana Gómez-Schiavon, Universidad Nacional Autónoma de México, Mexico

© Copyright Xue et al. This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.



**Figure 1.** Schematic representation of cell fate decisions driven by noise (A) and signal (B) from a view of epigenetic landscape. (A–B) Valleys represent stable attractors. Cells (yellow balls) in stem cell fate (denoted as ‘S’, green well in landscape) differentiate into downstream fates, lineage X (denoted as ‘LX’, blue well), and lineage Y (denoted as ‘LY’, purple well). These abbreviations were used for following **Figures 2–7**.

fate-decision processes, researchers have further elaborated the Waddington landscape (*Shakiba et al., 2022; Loeffler and Schroeder, 2019; Coomer et al., 2022; Shi et al., 2022; Stanoev and Koseska, 2022; Glauche and Marr, 2021*). An outstanding question is whether the landscape is static or not, i.e., whether cell fate decisions are driven by noise or signal (*Stanoev and Koseska, 2022; Simon et al., 2018; Zernicka-Goetz and Huang, 2010*). On one hand, some perspectives hold that cells reside in a stationary landscape, where decisions are made by switching through discrete valleys (*Lord et al., 2019; Desai et al., 2021*), as a result of gene expression noise (*Chang et al., 2008; Guillemain and Stumpf, 2021*) (termed as ‘noise-driven’; **Figure 1A**). Meanwhile, some researchers argued that the epigenetic landscape is dynamic during fate decisions. That is, the distortion of the landscape orchestrates fate transitions (*Hota et al., 2022; Zhu et al., 2022; Li et al., 2020*) and is driven by extrinsic signals (termed as ‘signal-driven’; **Figure 1B**).

Under the noise-driven mode, the bias of cell fate decisions largely depends on the spontaneous heterogeneity of gene expressions in the cell population (*Wheat et al., 2020; Kovary et al., 2018*). Consequently, the initial cellular state predominantly impacts the direction of the fate decision. *Chang et al., 2008* uncovered that hematopoietic stem cell (HSC) population possesses intrinsic and robust heterogeneity of *Scal-1* expression (also known as *Ly6a* [*van de Rijn et al., 1989*]). Notably, populations with discrete expression levels of *Scal-1* confer different propensities for downstream lineage commitment. Considering the signal-driven mode, cell fates are tightly steered by extrinsic signals (e.g., cytokines, chemical molecules, mechanical strength, and genetic operations) that reshape the landscape (**Figure 1B**). In this circumstance, the impact of the initial state on fate decisions is relevantly inconsequential. Additionally, due to the accessibility of signal manipulation, the signal-driven

mode has been widely utilized for cell fate engineering (Xu et al., 2015; Del Vecchio et al., 2017), leading to in-vitro induction systems centered on induced pluripotent stem cells (iPSC) for obtaining desired cell types (Ng et al., 2021). Recently, researchers reported a 'fate-decision abduction' of erythroid-to-myeloid trans-differentiation induced by various types of cancer, which facilitates tumor escape from the individual's immune system (Long et al., 2022). Collectively, driving forces couple the foundational and crucial features of fate decisions, serving as an essential basis for further decoding fate decisions and interpreting the development of organisms (Simon et al., 2018). By examining the two driving modes, we can gain a better understanding and characterization of cell fate decisions, including in-vivo cell differentiation, oncogenesis, and in-vitro reprogramming systems.

Nevertheless, the driving forces that underlie fate decisions remain largely elusive. The intricate nature of GRNs presents a challenge in deciphering driving modes. It has been generally acknowledged that corresponding core GRNs orchestrate cell fate decisions (Qiu et al., 2022; Graf and Enver, 2009), where the lineage-specifying TFs interact to implement fate-decision procedures. Furthermore, researchers transferred specific TFs into donor cells to reconfigure the intracellular GRNs for acquiring cell types of interest (Hammelman et al., 2022; Mazid et al., 2022). Although some studies suggested that perturbation of a single TF is sufficient to transform certain cell fates (Ng et al., 2021), a large number of TFs are inevitably involved in most differentiation/reprogramming processes (Hersbach et al., 2022). In particular to orchestrate decisions among multiple cell fates, it is necessary for TFs to regulate target genes cooperatively (Trojanowski and Rippe, 2022). As crucial determinants of cell fates, TFs function via binding to cis-regulatory elements (CRE, e.g., promoter and enhancer). CREs of a single gene in metazoans can simultaneously accommodate numerous TFs (Lambert et al., 2018; Reiter et al., 2017). While experimental protocols have been developed to assess TF binding and one-to-one up- or down-regulatory relationships, it is more challenging to quantify these combinatorial regulations. For instance, given two factors activate and inhibit the same target gene, respectively, does the target gene turn on or off when both factors are present in its CREs?

Computational approaches in systems biology can be utilized to tackle complex networks (MacArthur et al., 2009; Krumsiek et al., 2011; Collombet et al., 2017; Moignard et al., 2015; Kirouac et al., 2009). A concise GRN model typically entails the following two elements. The element 1 is the topology. Much research efforts have been devoted to investigating network topologies on cellular behaviors, e.g., toggle switch (Graf and Enver, 2009; Davies et al., 2021), and feed-forward loop (Kittisopikul and Süel, 2010; Oliver Metzger et al., 2020; Sciammas et al., 2011). In particular, the Cross-Inhibition with Self-activation (CIS) network is one of the most studied two-node GRNs in cell fate decisions (Orkin and Zon, 2008; Kato and Igarashi, 2019), with examples found in *Gata1-PU.1* and *FLI1-KLF1* in hematopoiesis (Hoppe et al., 2016; Pali et al., 2019), *Nanog-Gata6* and *Oct4-Cdx2* in gastrulation (Simon et al., 2018; Zhou and Huang, 2011), and *Sir2* and *HAP* in yeast aging (Li et al., 2020). In this topology, two lineage-specifying factors inhibit each other while active themselves. For example, in the well-known *Gata1-PU.1* circuit, *Gata1* directs fate of megakaryocyte-erythroid progenitor (MEP), and *PU.1* (also known as *SPI1* in humans) specifies the fate of granulocyte-monocyte progenitor (GMP) (Orkin and Zon, 2008). Namely, the antagonism of two TFs implicates two cell fates in competition with each other.

The another element is the logic for regulatory functions (Ocone et al., 2015; Matsuura et al., 2018). Exemplified by the CIS network, each node (e.g., X and Y) receives the activation by itself and inhibition by the counterpart. Hence there is naturally the logic function between these two inputs. Given the logic function is AND, in the context of biological mechanism of regulation by TFs, X gene expresses only when X itself is present in X's CREs but Y is not. Researchers observed in the *E. coli* lac operon system that changes in one single base can shift the regulatory logic significantly, suggesting that logic functions of GRNs can be adapted on the demand of specific functions in organisms (Collombet et al., 2017; Mayo et al., 2006; Buchler et al., 2003). Additionally, considering that the combination and cooperativity of TFs are of great significance in development (Zaret, 2020; Spitz and Furlong, 2012; Balsalobre and Drouin, 2022), theoretical investigation of the logic underlying GRNs should be concerned in cell fate decisions. However, despite the existence of large number of mathematical models on fate decisions, the role played by the regulatory logic in cell fate decisions is still obscure. Some theoretical studies put emphasis on specific biological instances, adopting logic functions that best fit the observations derived from experiments (Hota et al., 2022; Krumsiek et al., 2011; Huang et al., 2007). As a result, the models incorporated different regulatory logic

received limited attention. Other research assigned logic to large-scale multi-node GRNs, confining the interpretation of the role of logic (Moignard et al., 2015; Wu et al., 2008). Collectively, the bridge between the logic of nodes in GRNs and cell fate decisions has not yet been elucidated systematically and adequately. Current research already encompassed a wealth of cell fate decisions: embryogenesis (Xu et al., 2022; Yu et al., 2022; Simunovic et al., 2022), lineage commitment (Palii et al., 2019; Pei et al., 2020; Psaila et al., 2020), oncogenesis (Aguadé-Gorgorió et al., 2021; Yang et al., 2022; Uthamacumaran, 2021), in-vitro reprogramming (Tang et al., 2022; Liang et al., 2019; Yu et al., 2021), and large-scale perturbations (Ng et al., 2021; Dixit et al., 2016; Joung et al., 2023; Replogle et al., 2022). Analogous to the effort on taxonomy of cell types and tumors (Domcke and Shendure, 2023), how cell fate decisions can be classified and distilled to the common properties is a challenge for further exploring systematically and application on fate engineering (MacArthur, 2023; Casey et al., 2020).

In this work, we integrated the fate-decision modes (noise-driven/signal-driven) and the classical logic operations (AND/OR) underlying GRNs in a continuous model. Based on our model, we investigated the impact of distinct logic operations on the nature of fate decisions with driving modes in consideration. Additionally, we extracted the difference in properties between the two driving modes. We unearthed that in the noise-driven stem cells, regulatory logic results in the opposite bias of fate decisions. We further distilled the relationship among noise profiles, logic motifs, and fate-decision bias, showing that knowledge of two of these allows inference of the third heuristically. Under the signal-driven mode, we identified two basic patterns of cell fate decisions: progression and accuracy. Moreover, based on our findings in silico, we characterized cell fate decisions in hematopoiesis and embryogenesis and unveiled their decision modes and logic motifs underlying GRNs. Ultimately, we applied our framework to a reprogramming system. We deciphered the driving force of this trans-differentiation, and utilized noise patterns for nominating key regulators. We underscored that clustering of gene noise patterns is an informative approach to investigate high-throughput datasets. Together, we underlined regulatory logic is of the significance in cell fate decisions. Our work presents a generalizable framework for classifying cell fate decisions and a blueprint for circuit design in synthetic biology.

## Results

### Section 1: Mathematical model of the CIS network with logic motifs

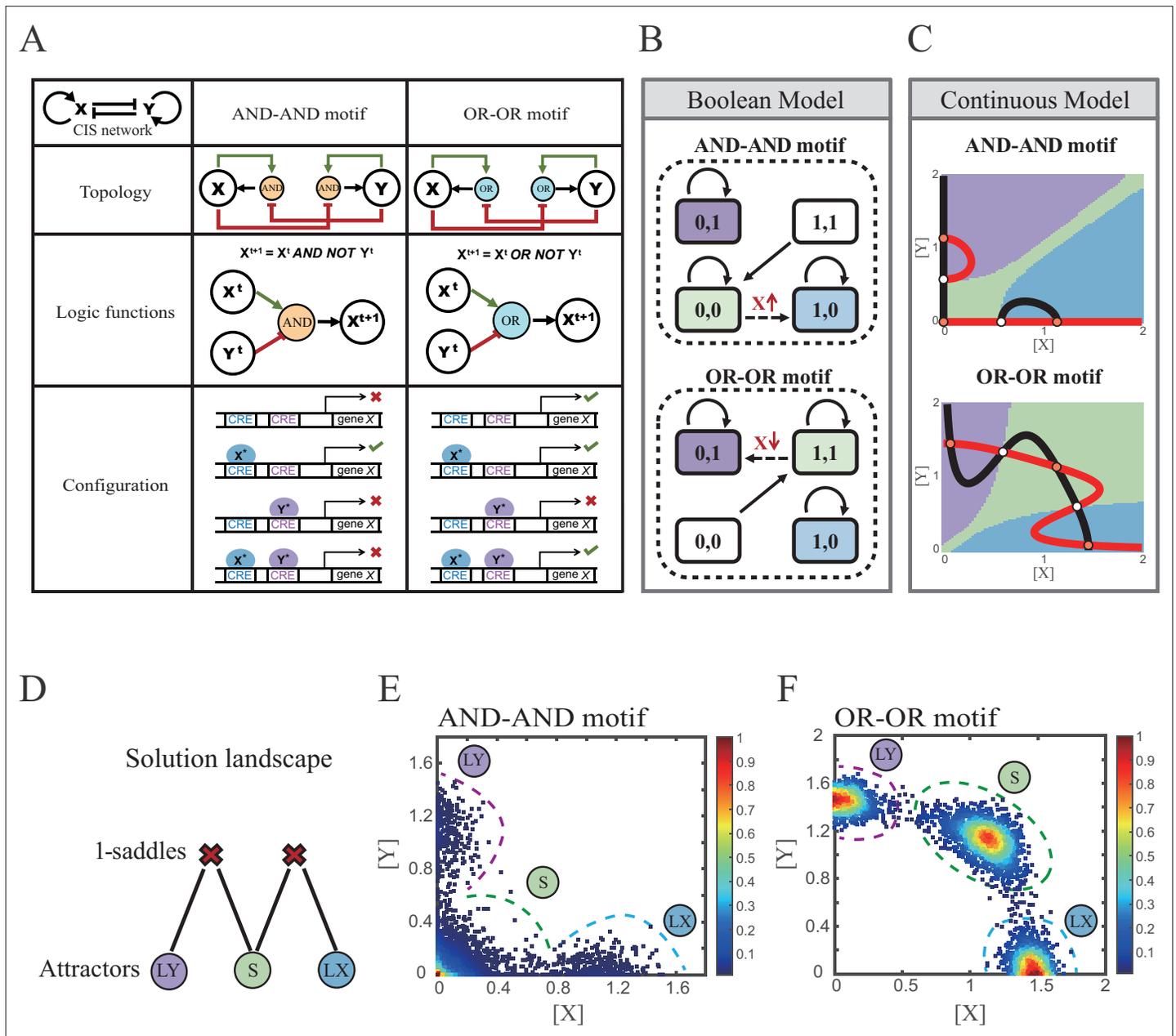
Binary tree-like cell fate decisions are prevalent in biological systems (Zhou and Huang, 2011; Domcke and Shendure, 2023; Stadler et al., 2021; Macnair et al., 2019), orchestrated by a series of the CIS networks. Accordingly, we developed our ordinary differential equations (ODE) model based on this paradigmatic and representative topology (Equations 1 and 2; see 'Materials and methods' for details).

$$\frac{d[X]}{dt} = \frac{r_0^x + r_1 k_1 [X]^{n_1} + r_2 k_2 [Y]^{n_2} + r_3 k_3 [X]^{n_1} [Y]^{n_2}}{1 + k_1 [X]^{n_1} + k_2 [Y]^{n_2} + k_3 [X]^{n_1} [Y]^{n_2}} - d_1 [X] \quad (1)$$

$$\frac{d[Y]}{dt} = \frac{r_0^y + r_4 k_4 [Y]^{n_2} + r_5 k_5 [X]^{n_1} + r_6 k_6 [X]^{n_1} [Y]^{n_2}}{1 + k_4 [Y]^{n_2} + k_5 [X]^{n_1} + k_6 [X]^{n_1} [Y]^{n_2}} - d_2 [Y] \quad (2)$$

X and Y are TFs in the CIS network.  $n_1$  and  $n_2$  are the coefficients of molecular cooperation.  $k_1$ - $k_3$  in Equation 1 and  $k_4$ - $k_6$  in Equation 2 represent the relative probabilities for possible configurations of binding of TFs and CREs. (Figure 2A).  $d_1$  and  $d_2$  are degradation rates of X and Y, respectively. Here, we considered a total of four CRE's configurations as shown in Figure 2A (i.e., TFs bind to the corresponding CREs or not,  $2^2=4$ ). Accordingly, depending on the transcription rates (i.e.,  $r_0^x$ ,  $r_1$ ,  $r_2$ ,  $r_3$  in Equation 1, similarly in Equation 2) of each configuration, we can model the dynamics of TFs in the Shea-Ackers formalism (Shea and Ackers, 1985; Olariu and Peterson, 2019).

Thus, the distinct logic operations (AND/OR) of two inputs (e.g., activation by X itself and inhibition by Y) can be further implemented by assigning the corresponding profile of transcription rates in four configurations (Figure 2A). From the perspective of molecular biology, the regulatory logics embody the complicated nature of TF regulation that TFs function in a context-dependent manner. Considering the CIS network, when X and Y bind respective CREs concurrently, whether the expression of



**Figure 2.** Models of the cross-inhibition with self-activation (CIS) network incorporated logic motifs. **(A)** A table listing the topologies with logic nodes, logic functions and Cis-regulatory elements (CRE) configurations in the CIS network incorporated AND-AND and OR-OR logic (denoted as AND-AND motif and OR-OR motif). X and Y are lineage-specifying transcription factors (TF).  $X^{t+1}$  indicates the value of X at the next time step.  $X^t$ ,  $Y^t$  represent activated forms of X and Y, respectively. The true or false signs denote whether gene X can be transcribed, respectively. These annotations were used for the following **Figures 3–7**. **(B)** State spaces of the AND-AND (top panel) and OR-OR (bottom panel) motifs in Boolean models. Updated rules of Boolean models are stated in **Figure 2A**. Rectangles indicate cell states. Green, blue, and purple represent S, LX, and LY, respectively. Solid arrows indicate transitions between states under corresponding Boolean models. Dotted arrows indicate forced transition imposed by external perturbations. **(C)** State spaces of the AND-AND (top panel) and OR-OR (bottom panel) motifs in ODE models. Dark and red lines represent nullclines of  $\frac{d[X]}{dt} = 0$ ,  $\frac{d[Y]}{dt} = 0$ , respectively. Stable steady states (SSS) are denoted as orange dots. Unstable steady states (USS) are denoted as white dots. Each axis represents the concentration of each transcription factor, which units are arbitrary. Blue, green, and purple areas in state spaces indicate attractor basins representing LX, S, and LY, respectively. Color of each point in state space was assigned by the attractors they finally enter according to the deterministic models (**Equation 1**, **Equation 2**). These annotations were used for the following **Figures 3–7**. **(D)** The solution landscape for the AND-AND and OR-OR motifs. The crimson X-cross sign denotes the first-order saddle node. Blue, green, and purple circles indicate attractors. These annotations were used for the following **Figures 3–7**. **(E–F)** Simulation result of stochastic differential equation models of the AND-AND (**E**) and OR-OR (**F**) motifs. Other than adding a white noise, parameters were identical with those in **(C)**. Initial values were set to the attractor representing S fate in **Figure 2C** top

Figure 2 continued on next page

Figure 2 continued

panel (E) and **Figure 2C** bottom panel (F). Noise levels of  $X$  ( $\sigma_x$ ) and  $Y$  ( $\sigma_y$ ) are both set to 0.14 in the AND-AND motif (E), and 0.1 in the OR-OR motif (F). Stochastic simulation was preformed 3500 times, with each final state recorded as a dot on the plot. Color of heatmap corresponds to the density of points. Unit of concentration is arbitrary.

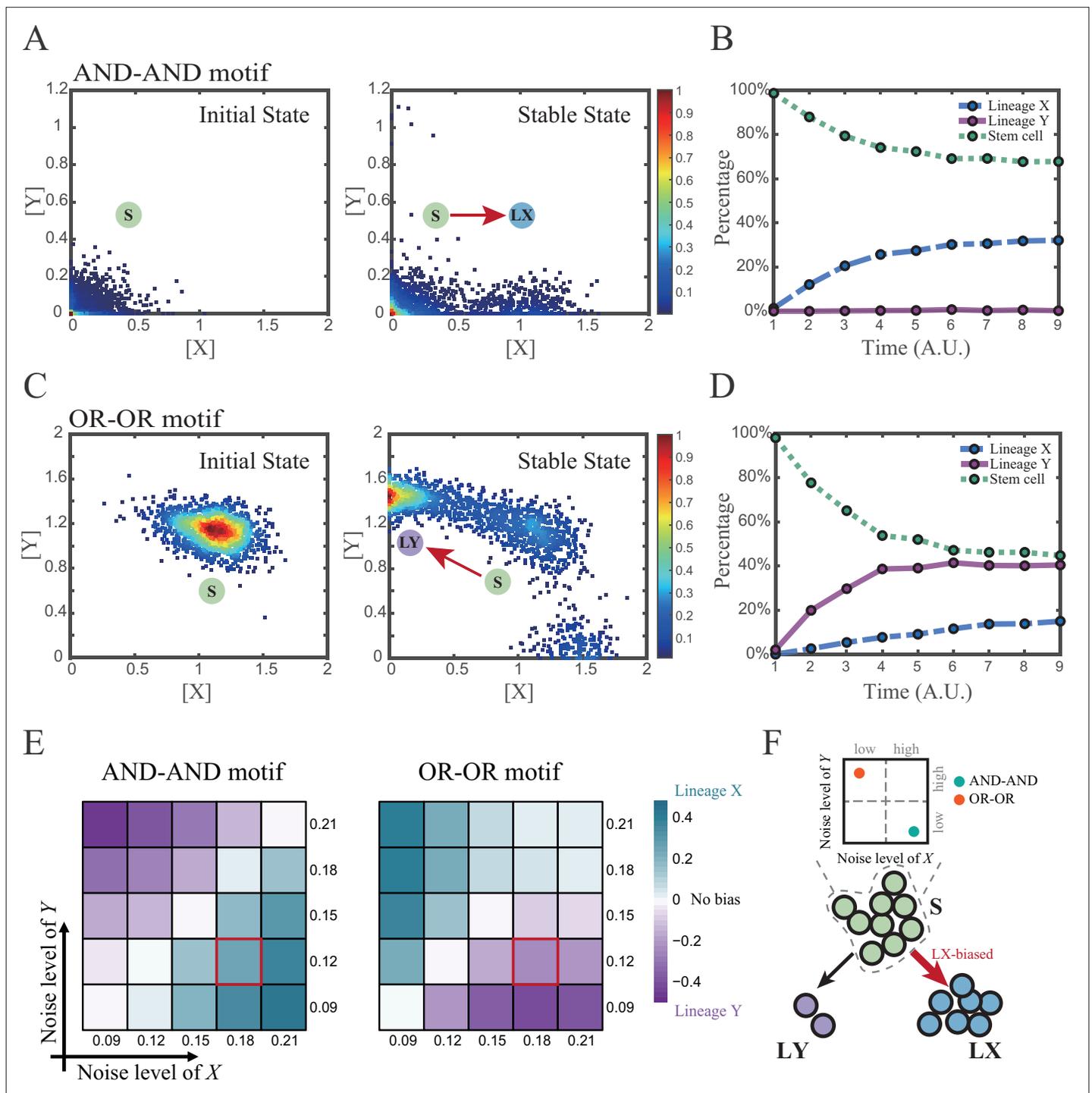
target gene is turned on or off depends on the different regulatory logics (specifically, off in the AND logic and on in the OR logic; **Figure 2A**). Notably, instead of exploring the different logics of one certain gene (*Kittisopikul and Süel, 2010*), we focus on different combinations of regulatory logics due to dynamics in cell fate decisions is generally orchestrated by GRNs with multiple TFs.

Benchmarking the Boolean models with different logic motifs (**Figure 2B**; see 'Materials and methods'), we reproduced the geometry of the attractor basin in the continuous models resembling those represented by corresponding Boolean models (**Figure 2C**; see 'Materials and methods'). Under double AND and double OR motifs (termed as AND-AND motif and OR-OR motif, respectively), there are typically three stable steady states (SSS) in the state spaces (**Figure 2C**): two attractors near the axes representing the fate of lineage X (denoted as LX,  $X^{\text{high}}Y^{\text{low}}$ ) and the fate of lineage Y (denoted as LY,  $X^{\text{low}}Y^{\text{high}}$ ), and the attractor in the center of the state space representing stem cell fate (denoted as S).

Evidently, the stem cell states exhibit different expression patterns between the two logic motifs. Stem cells in the AND-AND motif do not express X nor Y (**Figure 2B** top panel; expressed in low level in **Figure 2C** top panel), while in the OR-OR motif, stem cells express both lineage-specifying TFs (**Figure 2B** bottom panel; expressed in high level in **Figure 2C** bottom panel). The difference in the status of S attractors relates to the co-expression level of lineage-specifying TFs in stem cells in real biological systems (*Loeffler and Schroeder, 2019; Pali et al., 2019*). Intuitively, from the view of the Boolean model, stem cell state in the AND-AND motif ([0,0] state) needs to switch on lineage-specifying TFs to transit to downstream fates (**Figure 2B** top panel). Whereas in the OR-OR motif, fate transitions are subject to the switch-off of TF expression (**Figure 2B** bottom panel). Furthermore, we introduced the solution landscape method. Solution landscape is a pathway map consisting of all stationary points and their connections, which can describe different cell states and transfer paths of them (*Yin et al., 2020; Yin et al., 2021*). From the perspective of the solution landscape, two logic motifs possess akin geometric topologies in their steady-state adjacencies (**Figure 2D**): when there are three fates coexisting in the state space, S attractor resides in the middle of LX and LY as the possible pivot for fate transitions (**Figure 2D**). To investigate noise, we developed models with stochastic forms (see 'Materials and methods'). Simulations display the primary distribution of cell populations, corresponding to SSSs in deterministic models (**Figure 2E and F**).

## Section 2: Two logic motifs exhibit opposite bias of fate decisions under the noise-driven mode

We first investigated the difference between the AND-AND and OR-OR motifs under the noise-driven mode. Here, we assigned the stem cell state as the starting point in simulation. In biological systems, it is unlikely that the noise level of different genes is kept perfectly the same. Asymmetry of the noise levels was thus introduced. First, we set the noise level of TF X higher than that of Y ( $\sigma_x=0.18, \sigma_y=0.12$ ). Under this asymmetric noise, we observed that stem cells shifted toward LX in the AND-AND motif (**Figure 3A and B**), but toward LY in the OR-OR motif (**Figure 3C and D**). From the perspective of the state space, such properties intuitively originate from the distinctive status of stem cell attractors in two logic motifs (**Figure 2B and C**). In the AND-AND motif, the stem cell state resides at the origin of coordinates. Thus, with increasing X's noise level, the stem cell population crossed the boundary between S and LX basins with a rising probability (**Figure 2C** top panel). Consequently, the fate decision of the stem cell population manifests a bias toward LX. Likewise, in the OR-OR motif, the stem cell population has a higher probability of entering LY basin following an increase in X's noise (**Figure 2C** bottom panel). Next, we simulated multiple sets of noise levels for X and Y. We quantified the distribution of cell types, which was determined by the basin in which the final state of each round of stochastic simulation ended up. We observed that stem cell population displays almost opposite differentiation preference under identical noise levels but distinct logic motifs (**Figure 3E**). Conversely, when two distinct logic motifs exhibiting the same fate-decision bias, cell populations need to employ



**Figure 3.** Two logic motifs exhibit opposite bias of fate decisions under the noise-driven mode. **(A and C)** Stochastic simulation in both the AND-AND and OR-OR motifs.  $\sigma_x$  is set to 0.18, and  $\sigma_y$  is 0.12. In both **(A)** and **(C)**, initial values were identical with attractors of stem cell fate in **Figure 2C** (SSSs in green attractor basins). Simulation was preformed 1500 times, with each initial (A left and C left) and final (A right and C right) states recorded as a dot on the plot. **(B and D)** Time courses of the percentage of cells in different fates in stochastic simulation, under the AND-AND motif **(B)** and OR-OR motif **(D)**. Fates of cells were assigned by their final states according to the basins of the deterministic models in **Figure 2C**. Unit of time is arbitrary. **(E)** Heatmaps showing the bias of cell fate decisions under different noise levels of X and Y. Color of heatmap indicates the extent of bias. Here,  $bias = \frac{n_{LX} - n_{LY}}{n_{total}}$ .  $n_{LX}$ ,  $n_{LY}$  represent number of LX, LY, respectively.  $n_{total}$  represents the total number of cells ( $n_{total} = 1500$ ). The method of assigning fate to cells is identical with **Figure 3B and D**. The red-marked cells correspond to the noise conditions simulated in **(A)** and **(C)**. **(F)** Schematic illustration in **Figure 3** continued on next page

Figure 3 continued

that stem cell populations possessing the same bias of fate decisions need to have opposite noise patterns, according to whether they are in the AND-AND or OR-OR motif. The red and bold arrow indicates the bias of fate decisions.

The online version of this article includes the following figure supplement(s) for figure 3:

**Figure supplement 1.** Noise as a driving force for reprogramming of LY to S in two logic motifs.

opposite noise patterns (**Figure 3E and F**). Collectively, if two of the three (noise profile, logic motif, fate-decision bias) are accessible, the last is inferential.

Next, we wondered whether noise could act as a driving force for reprogramming (e.g., from LY to S). We assigned LY state as the starting cell type in simulation. Apparently, in the AND-AND motif, transition of LY to S can be realized by increasing the noise level of TF Y (**Figure 3—figure supplement 1A**). Meanwhile, in the OR-OR motif, it is the increased noise level of X that can drive the transition from LY to S (**Figure 3—figure supplement 1B**), which is also intuitive by viewing the basin geometry of the state space (**Figure 2C**). These observations suggested that under the noise-driven mode, experimental reprogramming strategy need to take consideration of the regulatory logic (e.g., in reprogramming of LY to S, perturb the high expression TF of LY in the AND-AND motif, while in the OR-OR motif, perturb the low expression TF).

### Section 3: Two logic motifs decide oppositely between differentiation and maintenance under the signal-driven mode

In addition to noise, cell fate decisions can also be driven by signals, e.g., GM-CSF in hematopoiesis (**Mojtahedi et al., 2016**), CHIR99021 in chemically induced reprogramming (**Zhao, 2019**). The change conducted by signals corresponds to the distortions of the cell fate landscape. To simulate the signal-driven mode, we focused on the effect of parameters in the mathematical models on the system's dynamical properties. To simulate models feasibly and orthogonally, we added parameters  $u$  ( $u_x$  in **Equation 3**,  $u_y$  in **Equation 4**) to **Equations 1; 2**:

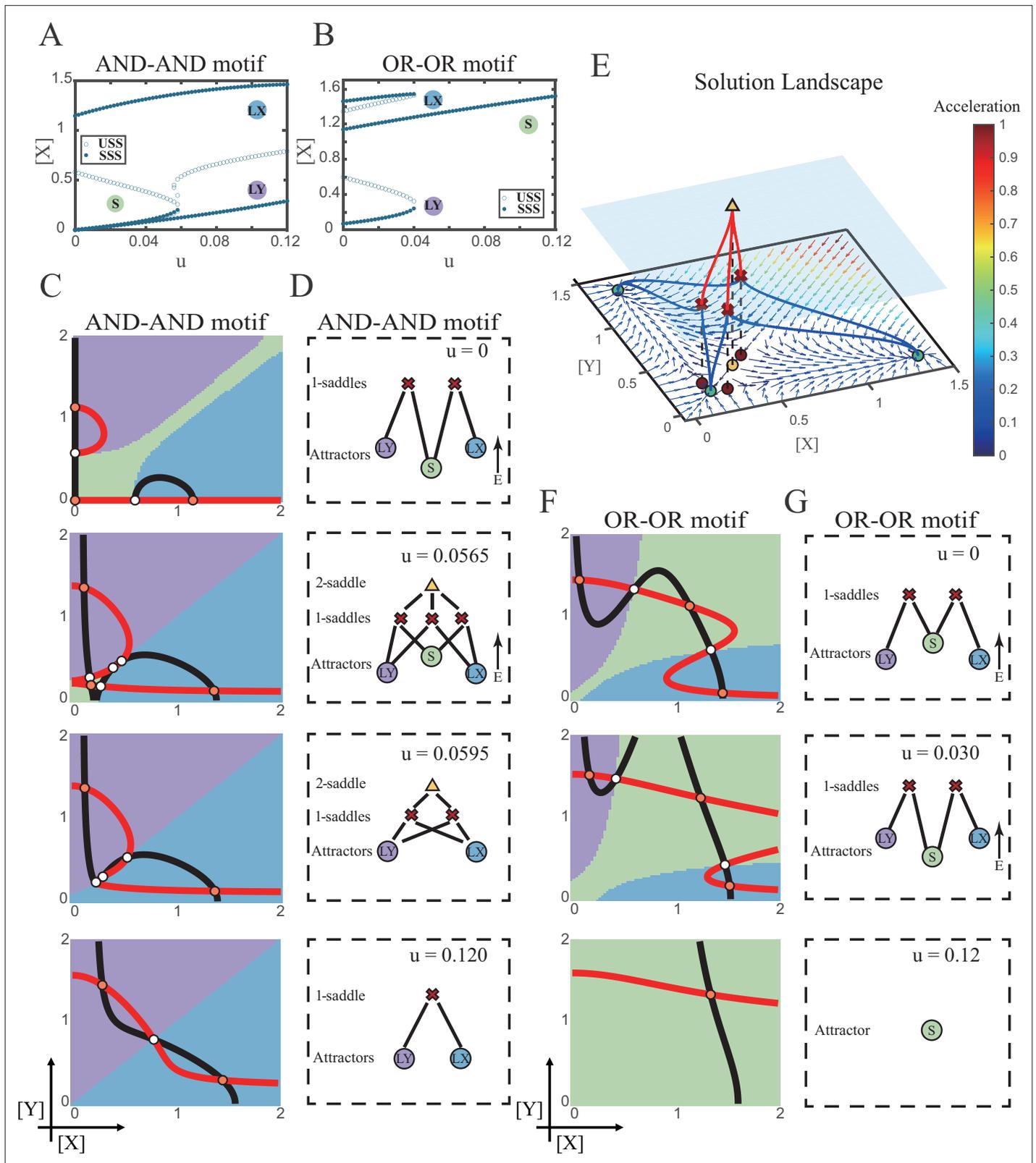
$$\frac{d[X]}{dt} = \frac{r_0^x + r_1 k_1 [X]^{n_1} + r_2 k_2 [Y]^{n_2} + r_3 k_3 [X]^{n_1} [Y]^{n_2}}{1 + k_1 [X]^{n_1} + k_2 [Y]^{n_2} + k_3 [X]^{n_1} [Y]^{n_2}} - d_1 [X] + u_x \quad (3)$$

$$\frac{d[Y]}{dt} = \frac{r_0^y + r_4 k_4 [Y]^{n_2} + r_5 k_5 [X]^{n_1} + r_6 k_6 [X]^{n_1} [Y]^{n_2}}{1 + k_4 [Y]^{n_2} + k_5 [X]^{n_1} + k_6 [X]^{n_1} [Y]^{n_2}} - d_2 [Y] + u_y \quad (4)$$

The increase of  $u$  represents an elevation in the basal expression level of lineage-specifying TFs, reflecting an induction signal from the extracellular environment. From an experimental standpoint, this signal can be the induction of small molecules or overexpression by gene manipulations, such as the transfection of cells with expression vectors containing specific genes.

We first explored the impact on the system when the two induction parameters are changed symmetrically ( $u = u_x = u_y$ ). As the increase of  $u$ , the number of SSS in the AND-AND system decreases from three to two, where S attractor evaporates after a subcritical pitchfork bifurcation (**Figure 4A** and **Figure 4—figure supplement 1A**). Whereas in the OR-OR motif, after the increase of  $u$ , LX and LY attractors disappear with saddle-node bifurcations, respectively. Only the SSS representing stem cell fate is retained in the state space (**Figure 4B**). We then portrayed all the topology of the steady-state adjacency that accompanied the increase of  $u$ , from the perspective of the solution landscape. In the AND-AND motif, the attractor basin of LX and LY started to adjoin and occupied the vanishing S attractor basin together (**Figure 4C and D**). Accordingly, the stem cells cannot maintain themselves and decided to differentiate into either one of the lineages. Moreover, if the cell population possesses the same noise levels in both X and Y, then the fate decisions are unbiased (**Figure 4—figure supplement 1B**).

Notably, in the AND-AND motif, we observed a brief intermediated stage before S attractor disappears, where all three fates are directly interconnected (**Figure 4C** second panel and **Figure 4D** second panel and **Figure 4E**). To manifest the generality, we globally screened 6231 groups of parameter sets under the AND-AND motif, and this logic-dependent intermediated stage can be observed for 82.7% of them (see 'Materials and methods'; **Supplementary file 1**), indicating little dependence on particular parameter setting (1.8% in the OR-OR motif). Unlike the indirect attractor adjacency



**Figure 4.** Two logic motifs decide oppositely between differentiation and maintenance under the signal-driven mode. **(A–B)** Bifurcation diagrams for the AND-AND motif **(A)** and OR-OR motif **(B)** driven by parameter  $u$  ( $u = u_x = u_y$ ) in the cross-inhibition with self activation (CIS) model. Stable steady states (SSS) and unstable steady states (USS) are denoted as solid dots and hollow dots, respectively. **(C and F)** Changes in the state spaces for the AND-AND motif **(C)** and OR-OR motif **(F)** with increasing parameter  $u$ , from top to down. **(D and G)** Changes in the solution landscape with increasing of  $u$ , in *Figure 4 continued on next page*

Figure 4 continued

company with these in (C and F). The crimson X-cross sign and yellow triangle denote first-order and second-order saddle nodes, respectively. Relative energy is quantified by the geometric minimum action method (Vanden-Eijnden and Heymann, 2008), see 'Materials and methods'. (E) The solution landscape with parameter  $u=0.0565$  for the AND-AND motif from a view of three dimensions. It describes a hierarchical structure of the steady states. From top to bottom, it represents 2-saddle (yellow triangle), 1-saddles (crimson X-cross sign), and the attractors (green dot). The layer of 1-saddles is represented by a blue translucent plane, and the bottom layer is the flow field diagram. The connections from 2-saddle to 1-saddles are represented by red lines, and the connection from 1-saddles to the attractors are represented by blue lines. In the flow field diagram, the direction and color of the arrows correspond to the direction and size of the flow at that location. The corresponding positions of 2-saddle and 1-saddles in the flow field are marked with yellow and red dots, respectively, with black dashed lines indicating the corresponding relationship.

The online version of this article includes the following figure supplement(s) for figure 4:

**Figure supplement 1.** Two logic motifs are associated with opposite fate-decision choices under bidirectional induction.

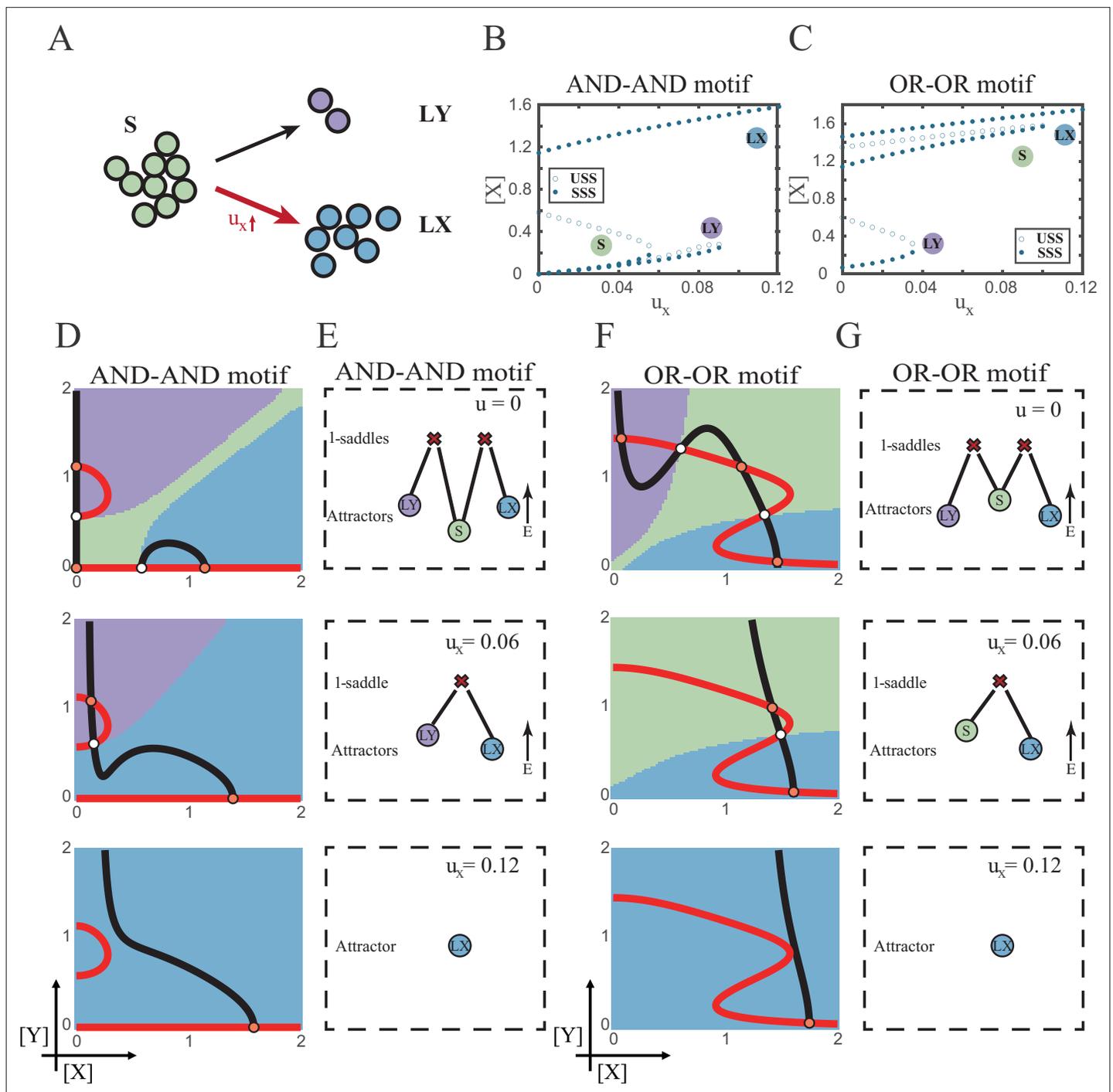
structure mediated by S attractor (Figure 2D), the solution landscape with fully-connected structure facilitates transitions between any two pairs of fates. Furthermore, this transitory fully-connected stage is located between the fate-undetermined stage (Figure 4C top panel) and fate-determined stage (Figure 4C third panel), comparable to the initiation (or activation) stage before the lineage commitment in experimental observations (Brand and Morrissey, 2020; Zhang et al., 2018; Arinobu et al., 2007). Therefore, we suspected that the robust fully-connected stage in the AND-AND motif may correspond to a specific period in cell fate decisions.

From the standpoint of reprogramming of differentiated cells back into progenitors, in the AND-AND motif, differentiated cells are more capable of maintaining their own fates during the symmetrical increase of the induction signals on both lineages (Figure 4A and C). Whereas in the OR-OR motif, the attractor basin of LX or LY is progressively occupied by the stem cell fate as  $u_x$  and  $u_y$  increase together (Figure 4F and G). In this scenario, the downstream fates are eventually reversed back to the undifferentiated state (Figure 4B and F). Namely, reprogramming engaged in the OR-OR motif can be accomplished by bi-directional induction of downstream antagonistic fates. In sum, we found that under symmetrical signal induction, the behavior of stem cells is subject to core GRN's logic motifs. In the AND-AND motif, stem cells prefer to differentiate, while under the OR-OR motif, the stem cell population inclines to maintain its undifferentiated state.

#### Section 4: The trade-off between progression and accuracy of cell fate decisions under the signal-driven mode

According to experimental observations, the majority of fate decisions exhibit lineage preference, also known as 'symmetry breaking of fate decisions' (Stanoev and Koseska, 2022; Pei et al., 2020; Psaila et al., 2020; Chen et al., 2018). Take the lineage choices in hematopoiesis as an example, Some HSCs prefer myeloid over lymphoid (Pei et al., 2020; de Haan and Lazare, 2018). This fate-decision bias also further shifts along with aging and infection (Zhang et al., 2018; Reya et al., 2001). In studying this preference in fate decisions, we broke the symmetry in the signal-driven models, by solely increasing  $u_x$  while keeping  $u_y=0$  (Figure 5A). First, it is apparent that the fate decision will significantly steer toward LX along with the increase of  $u_x$ , regardless of the logic motifs. Ultimately the state spaces contain only LX attractor when  $u_x$  is sufficiently high (Figure 5B and C and Figure 5—figure supplement 1A). However, the changes in the state space and the solution landscape follow different routes for two logic motifs. In the AND-AND motif, S attractor basin disappears at first, leaving a state space with two differentiated fates (Figure 5D and E). Then the basin of LY attractor shrinks and finally disappears (Figure 5D and E). Whereas in the OR-OR motif, LY attractor disappears first. Then S attractor, with an enlarged basin, shares the state space with LX attractor. Finally, S attractor basin abruptly disappears by a saddle-node bifurcation (Figure 5F and G).

The distinct sequences of attractor basin disappearance as  $u_x$  increasing can be viewed as a trade-off between progression and accuracy. In the AND-AND motif, the attractor basin of LX and LY adjoins (Figure 5D middle panel) when S attractor disappears due to the first saddle-node bifurcation (Figure 5B). Notwithstanding the bias of differentiation toward LX, the initial population still possesses the possibility of transiting into LY (Figure 5—figure supplement 1B). That is, in the AND-AND motif, as the increase of induction signal  $u_x$ , the 'gate' for the stem cell renewal is closed first. Stem cells are immediately compelled to make fate decisions toward either LX or LY, with a bias toward LX but a nonignorable probability of entering LY. Albeit the accuracy of differentiation is,



**Figure 5.** The progression-accuracy trade-off in cell fate decisions. **(A)** Schematic illustration of S-to-LX cell fate decisions with X-inducing signals. The red and bold arrow indicates the direction of fate decisions. **(B–C)** Bifurcation diagrams for the AND-AND motif **(B)** and OR-OR motif **(C)** driven by parameter  $u_x$ . **(D and F)** Changes in the state spaces for the AND-AND motif **(D)** and OR-OR motif **(F)** with increasing values of  $u_x$  from top to down. **(E and G)** Changes in the solution landscape with increasing of  $u_x$  in company with these in **(D and F)**.

The online version of this article includes the following figure supplement(s) for figure 5:

**Figure supplement 1.** The PA (progression and accuracy) trade-off of cell fate decisions under the signal-driven mode.

**Figure supplement 2.** Distinct trajectories of reprogramming under the signal-driven mode.

therefore, compromised, the overall progression of differentiation is ensured (i.e., all stem cells have to make the fate decisions downward. This causes the pool of stem cells to be exhausted rapidly). Whereas in the OR-OR motif, the antagonistic fate, LY, disappears first (**Figure 5C**). The attractor basin of S and LX are adjacent in the state space (**Figure 5F**). In this case, the orientation of the fate decisions is generally unambiguous since the stem cell population can only shift to LX, ensuring the accuracy of differentiation. Next, to check if the observed sequences of basin disappearance are artifacts of specific parameter choice, we randomly sampled parameter sets to check the sequence of attractor changes in their state spaces (6207 groups of the AND-AND motifs and 6634 groups of the OR-OR motifs; **Supplementary file 1**). We found that 96% AND-AND motifs and 70% OR-OR motifs exhibit the same sequence of attractor vanishment mentioned above (**Figure 5—figure supplement 1C and D** and **D**; see ‘Materials and methods’). These results of the global screen demonstrated that the sequence of attractor vanishment is robust to parameter settings. In sum, we proposed that logic motifs couple the trade-off between progression and accuracy as a general phenomenon in the signal-driven asymmetrical fate decisions (**Figure 5—figure supplement 1E**).

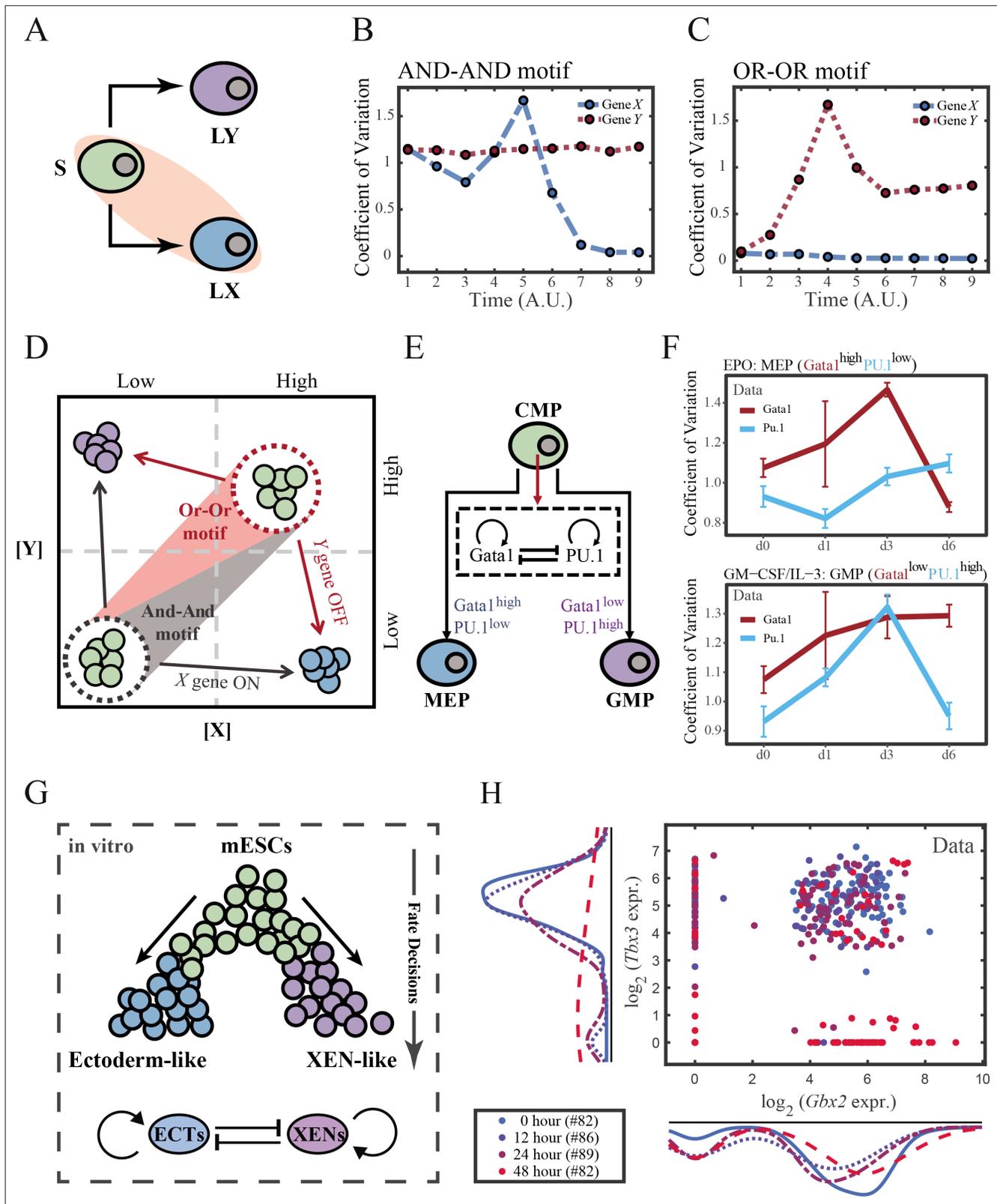
Next, we examined the trans-differentiation from LY into LX by increasing  $u_x$ . In the AND-AND motif, with the induction of X, LY directly transitioned into LX as the stem cell state disappears before LY (**Figure 5D** and **Figure 5—figure supplement 2A**). Intriguingly, for the OR-OR motif under the same induction, LY population first returned to the S state and then flows into LX (**Figure 5F** and **Figure 5—figure supplement 2B**). Namely, different logic motifs conduct distinct trajectories in response to identical induction in reprogramming. The AND-AND motif renders a one-step transition between downstream fates (**Figure 5—figure supplement 2C**). While in the OR-OR motif, it is a two-step transition mediated by the stem cell state (**Figure 5—figure supplement 2D**). This phenomenon suggests the observation that cells may be reprogrammed to distinct cell types depending on the induction dose (**Zhao, 2019**) is more realizable in the OR-OR motif. Integrated with the foregoing symmetrical induction, we recapitulated that in the OR-OR motif, the bi-directional induction or a unilateral induction from a counterpart (e.g., solely induced Y to realize reprogramming of LX to S) confer downstream cell fates to return to the undifferentiated state (**Figure 4F** and **Figure 5F**). Whereas in the AND-AND motif, it is substantially more difficult to achieve de-differentiation. This observation may explain why some cell types are not feasible to reprogram (**Shi et al., 2017**).

## Section 5: The CIS network performs differently during hematopoiesis and embryogenesis

In prior sections, we systematically investigated two logic motifs under the noise- and signal-driven modes in silico. With various combinations of logic motifs and driving forces, features about fate-decision behaviors were characterized by computational models. Next, we questioned whether observations in computation can be mapped into real biological systems. And how to discern different logic motifs and driving modes is a prerequisite for answering this question.

To end this, we first evaluated the performance of different models, specifically in simulating the process of stem cells differentiating towards LX (**Figure 6A**). Under four models with different combinations of driving modes and logic motifs (**Figure 6—figure supplement 1A and B**), we assessed the expression level and expression variance (defined as the coefficient of variation) of TFs X and Y among the cell population over time in stochastic simulation. We observed that, under the same logic motifs, different driving modes change in the patterns of expression variance rather than expression levels (**Figure 6B and C** and **Figure 6—figure supplement 1C and D**). Overall, under the noise-driven differentiation from S to LX, the variance of expression exhibits a continuous and monotonic trend (**Figure 6—figure supplement 1D**) for both logic motifs. For different logic motifs, in the AND-AND motif, the expression variance of X (highly expressed in LX) declines (**Figure 6—figure supplement 1D** top panel). Whereas in the OR-OR motif, it is the expression variance of Y (low expressed in LX) displays a rising trend (**Figure 6—figure supplement 1D** bottom panel). Nevertheless, under the signal-driven mode, the expression variance increases and then decreases, exhibiting a non-monotonic transition due to signal-induced bifurcation. During S to LX differentiation, comparable to the noise-driven mode, it is the expression variance of TF X in the AND-AND motif and TF Y in the OR-OR motif display a nonmonotonic pattern.

Such disparities between logic motifs originate from the location of S attractor (**Figure 6D** and **Figure 2C**). Although the target cell types are the same (LX), the AND-AND motif requests the



**Figure 6.** The cross-inhibition with self activation (CIS) network performs differently during hematopoiesis and embryogenesis. **(A)** Schematic illustration of S differentiating into LX. We took fate transition labeled in light pink shade as an example in the following simulation. **(B)** Time courses on the coefficient of variation in expression levels of X and Y genes in silico during differentiation towards LX ( $u_x$  switches from 0 to 0.08 from time point 1–9) in the AND-AND motif. Initial values were set to the attractors of stem cell fate in **Figure 2C** top panel (SSS in green attractor basin).  $\sigma_x$  and  $\sigma_y$  are both

Figure 6 continued on next page

Figure 6 continued

set to 0.07. Stochastic simulation was performed 1000 times for each pseudo-time point. Unit of time is arbitrary. (C) Time courses on the coefficient of variation in expression levels of X and Y genes in silico during differentiation towards LX ( $u_x$  switches from 0 to 0.24 from time point 1–9) in the OR-OR motif. Initial values were set to the attractors of stem cell fate in **Figure 2C** bottom panel (stable steady state, SSS in green attractor basin).  $\sigma_x$  and  $\sigma_y$  are both set to 0.05. Stochastic simulation was performed 1000 times for each pseudo-time point. Unit of time is arbitrary. (D) Schematic illustration of distinctive cell fate decision patterns under the AND-AND and OR-OR motifs in the state space. Dark and red gradients represent the extent of 'AND-AND' and 'OR-OR' in the actual regulatory network, respectively. Each axis represents expression levels of the lineage-specifying transcription factors. Blue, green, and purple circles indicate the cell fates of LX, S, and LY, respectively. (E) Schematic illustration of *Gata1-PU.1* circuit that dominates the primary fate decisions in hematopoiesis (CMP: Common myeloid progenitor; MEP: megakaryocyte-erythroid progenitor; GMP: Granulocyte-monocyte progenitor). (F) Measured coefficient of variation of expression levels of *Gata1* and *PU.1* changing over time during differentiation from CMPs to MEPs and GMPs. Expression levels were quantified via single-cell RT-qPCR (**Mojtahedi et al., 2016**). Error bars on points represent standard deviation (SD). For details of data processing, see 'Materials and methods'. (G) Schematic illustration of the differentiation from mESCs in an induction system (**Semrau et al., 2017**). (H) Measured expression levels of *Gbx2* and *Tbx3* among cells in embryogenesis quantified via single-cell SMART-seq2 (**Semrau et al., 2017**). For details of data processing, see 'Materials and methods'.

The online version of this article includes the following figure supplement(s) for figure 6:

**Figure supplement 1.** Simulation of fate transitions of S to LX in silico in two logic motifs under two driving forces.

**Figure supplement 2.** Gene regulatory networks (GRNs) perform differently in differentiations during hematopoiesis and embryogenesis.

expression of the TF X to be turned on, while the OR-OR motif requests the TF Y to be turned off. These key fate-transition genes, namely TF X in the AND-AND motif and TF Y in the OR-OR motif, both exhibit a sharp increase of variation in response to saddle-node bifurcation driven by  $u_x$  induction (first saddle node in **Figure 5B**; second saddle node in **Figure 5C**). Overall, these computational results suggest that we may be able to distinguish the two driving modes according to the expression variance over time series, then logic motifs can be correspondingly assigned by the expression level of the genes in the target cell types. For instance, if the expression variance of the X gene exhibits a nonmonotonic pattern and X is highly expressed in target cell types, then this cell fate decision can be assigned as the signal-driven fate decision in an AND-AND-like motif.

To support our findings with real-world correspondence, we first focused on the differentiation of CMPs in hematopoiesis (**Figure 6E**). It is acknowledged that the transcriptional regulation of *Gata1-PU.1* circuit dominates this cell fate decisions, which conforms to the CIS topology (**Figure 6E**). **Mojtahedi et al., 2016** stimulated murine multipotent hematopoietic precursor cell line EML with erythropoietin (EPO) or granulocyte-macrophage colony-stimulating factor/interleukin 3 (GM-CSF/IL-3) to examine the commitment into an erythroid or a myeloid fate, respectively. Based their curated dataset, we found that the expression level of *Gata1* (highly expressed in MEPs) gradually increased during EPO induction (**Figure 6—figure supplement 2A**), while the expression variance exhibits a nonmonotonic trend (**Figure 6F** top panel). Symmetrically, during GM-CSF/IL-3 induced differentiation toward GMPs, the expression level of *PU.1* (highly expressed in GMPs) gradually increased (**Figure 6—figure supplement 2B**), while the expression variance also presents a nonmonotonic pattern (**Figure 6F** bottom panel). The trends shown in the dataset resemble the signal-driven mode with the AND-AND motif. In addition, we quantified the expression of *Gata1* and *PU.1* via the single molecule FISH dataset (**Figure 6—figure supplement 2C**; **Wheat et al., 2020**). They are at low levels in CMPs, corresponding to the expression patterns in the AND-AND motifs (**Figure 2E** and **Figure 6D**). Together, we suggested that the *Gata1-PU.1* circuit performs in an AND-AND-like manner, and this differentiation system (**Mojtahedi et al., 2016**) is under the signal-driven mode.

Another paradigmatic model of fate decision is the differentiation of embryonic stem cells (ESC). **Semrau et al., 2017** found that under the retinoic acid (RA) exposure system in vitro, mouse embryonic stem cells (mESC) differentiated into two lineages: extraembryonic endoderm (XEN)-like and ectoderm-like. The investigators recapitulated that two clusters of TFs with the CIS topology determined this lineage specification (**Figure 6G**). We observed that the expression variance in most of these fate-decision TFs (16/22 73%) are gradually increasing during time, and 14% (3/22) of them exhibit nonmonotonic behavior (**Figure 6—figure supplement 2D**), suggesting the process is more likely driven by noise (**Figure 6—figure supplement 2E**). Furthermore, we focused on potential key regulators: *Gbx2* and *Tbx3*, the two likely targets of RA that are crucial for this fate decision (**Semrau et al., 2017**). The expression variances over time of these two TFs are consistently increasing (**Figure 6—figure supplement 2D**). In addition, their initial expressions are at a high level, in agreement with that

of the OR-OR motif (**Figure 6D and H**). In short, we proposed that the mESCs differentiation system under RA exposure performs in an OR-OR-like manner, and its differentiation is under the noise-driven mode in this experimental setting.

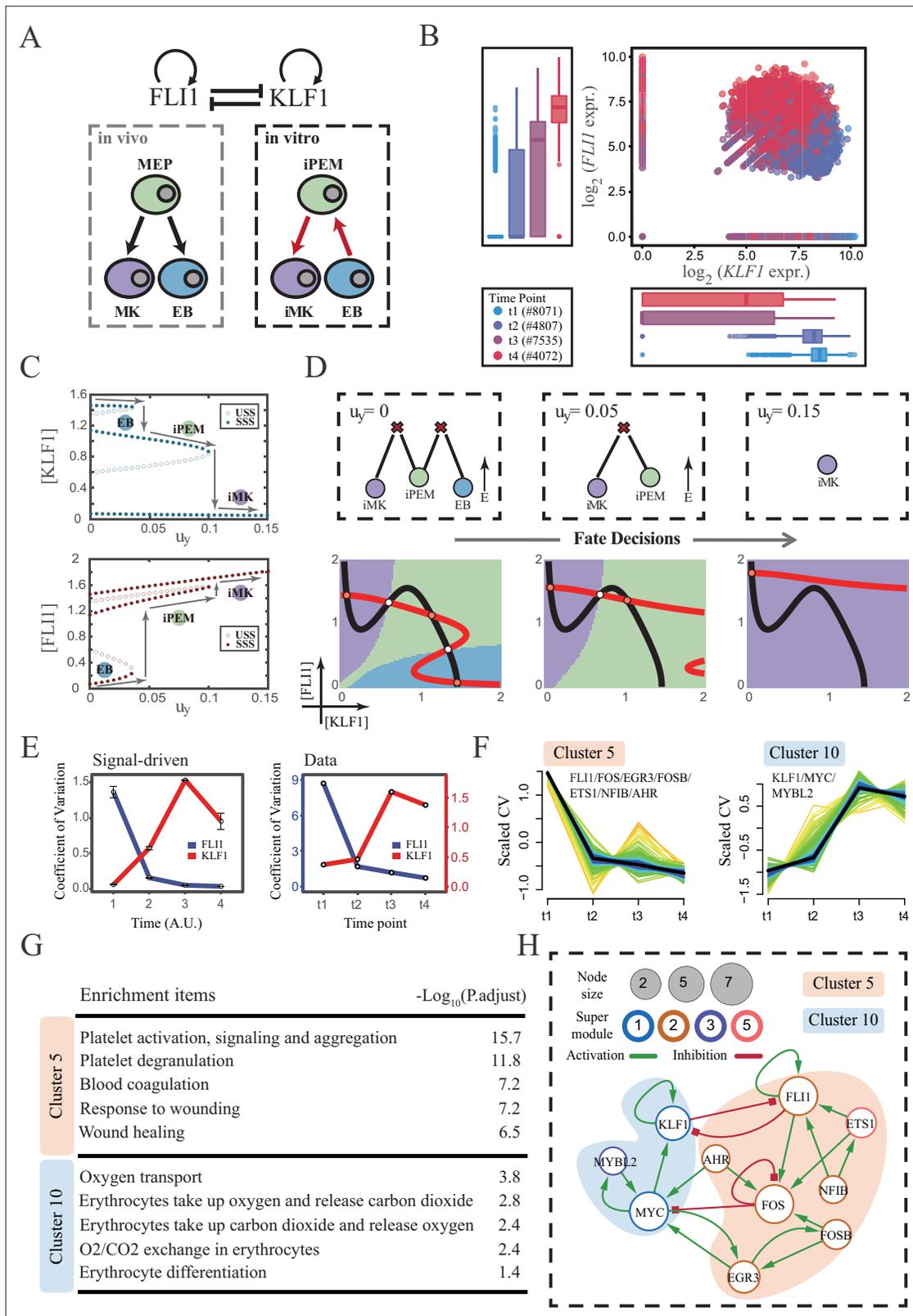
## Section 6: The chemical-induced reprogramming of human erythroblasts (EB) to induced megakaryocytes (iMK) is the signal-driven fate decisions with an OR-OR-like motif

The foregoing cell fate decisions initiate from pluripotent cells in mice (mESC, CMP) corresponding to a typical 'downhill' in Waddington's metaphor. In 2006, Yamanaka et al. accomplished the reprogramming from mouse fibroblasts into iPSC state via the noted 'OSKM' factors, representing 'uphill' in Waddington's metaphor (**Shi et al., 2017; Hamazaki et al., 2017; Abdallah and Del Vecchio, 2019**). Likewise, trans-differentiations from one lineage to another have been realized by overexpression or chemical inductions (**Xu et al., 2015**), whether they correspond to direct 'trespassing' of the ridge or an 'up-and-down' through the peak in Waddington landscape are still elusive. We then applied our models to reprogramming systems, with a primary focus on hematopoiesis. **Qin et al., 2022** recently achieved the direct chemical reprogramming of EBs to iMKs using a four-small-molecule cocktail (**Figure 7A**). Investigators presented that EBs underwent an induced bipotent precursor for erythrocytes and MKs (iPEM) to finally desired iMKs. It is acknowledged that the *FLI1*-*KLF1* circuit with the CIS topology dominates this fate-decision process (**Orkin and Zon, 2008; Palii et al., 2019**). To deduce the logic motif of the *FLI1*-*KLF1* circuit, we quantified the expression patterns of *FLI1* and *KLF1* based on published single-cell RNA-seq data (**Qin et al., 2022**). We can observe the fate transition from the EB population ( $FLI1^{low}, KLF1^{high}$ ) to the iMK population ( $FLI1^{high}, KLF1^{low}$ ) (**Figure 7B**). According to their expression level, the cell populations can be primarily classified into three clusters. In addition, both *FLI1* and *KLF1* are highly expressed in the intermediate cell population suspected to be the progenitors of iMKs and EBs (**Qin et al., 2022**). Namely, the pattern of expression level is concordant with the OR-OR motif in our framework (**Figure 6D**).

Next, to investigate the driving force of this reprogramming system, we simulated the fate transition from EB (corresponding to LX, blue) to iMK (LY, purple) under both driving modes. Under the noise-driven mode, we assumed that the reprogramming system facilitated the noise levels of both TFs. For simplicity, the starting cell population (EBs) was assigned symmetrical high noise levels. While under the signal-driven mode, we assumed that the four-small-molecule cocktail upregulated the expression of *FLI1* (highly expressed in iMKs). Then, we simulated the transition from EBs to iMKs by lifting the basal expression level of *FLI1*, corresponding to parameter  $u$ , in model (**Figure 7C**). The bifurcation diagrams indicate that the signal-driven fate transition is mediated by the iPEM state (**Figure 7C**). In particular, overexpression of *FLI1* renders sequential saddle-node bifurcations. Thus, EBs are converted to iPEMs before steering toward the terminal iMK state (**Figure 7C and D**), which is consistent with the experiment's findings.

Furthermore, we next assessed the dynamic behaviors in models of this system under different driven modes with the OR-OR logic. As discussed before, there is no discernible difference in the expression levels between the two driving modes. Overall, the common tendency is an up-regulation in *KLF1* and a down-regulation in *FLI1* (**Figure 7—figure supplement 1A**). We then quantified the expression variances during this fate transition by the model. Under the noise-driven mode, expression variances of *FLI1* and *KLF1* would gradually decrease and increase, respectively, until stabilizing (**Figure 7—figure supplement 1B**). Under the signal-driven mode, the expression variance of *FLI1* would first decline and then remain nearly constant, while *KLF1* would exhibit a nonmonotonic pattern (**Figure 7E** left panel). From the view of modeling, the nonmonotonic pattern presented by *KLF1* originates from the rapid shut-off during the transition from iPEM state to iMK state (second saddle node in **Figure 7C**).

Accordingly, we next quantified the expression variances in the real dataset over time. Impressively, the pattern emerging from the data accommodates the hypothesis of the signal-driven mode (**Figure 7E**). Altogether, we proposed that this reprogramming system (**Qin et al., 2022**) is the signal-driven process underlying the OR-OR-like motif. Moreover, the high expression level of *FLI1* induced by small molecules is the key driving force of the fate transition, as suggested by the properties of the OR-OR motif. We underlined that it is a classical two-step fate-decision process mediated by the



**Figure 7.** The chemical-induced reprogramming of human erythroblast (EB) to induced megakaryocyte (iMK) is the signal-driven fate decisions with an OR-OR-like motif. (A) Schematic illustration of the differentiation from megakaryocyte-erythroid progenitors (MEPs) in vivo and in vitro. Red arrows represent the route of reprogramming (Qin et al., 2022) (B) Measured expression levels of *KLF1* and *FLI1* in reprogramming quantified via single-cell 10 X. For details of data processing, see ‘Materials and methods’. (C) Bifurcation diagrams for the OR-OR motif driven by parameter  $u_y$  in the CIS model.

Figure 7 continued on next page

Figure 7 continued

(D) Fate transition representing reprogramming of EB to iMK in silico. Top panel: changes in the solution landscape with increasing of parameter  $u_y$ , from left to right; Bottom panel: changes in the state spaces for the OR-OR motif with increasing values of  $u_y$ , in company with these in top panel. Unit of concentration is arbitrary. (E) Left panel: coefficient of variation of expression levels of *KLF1* and *FLI1* changes in silico over time under given parameter ( $u_y=0.11$ ) in the OR-OR motif. Noise level of *KLF1* ( $\sigma_x$ ) and *FLI1* ( $\sigma_y$ ) are set to 0.087. Initial values were identical with LX attractor in **Figure 2C** bottom panel (stable steady state, SSS in blue attractor basin). Stochastic simulation was preformed 1000 times per round for each time point. We totally preformed three round simulations. Error bars on points represent standard deviation (SD); Right panel: measured coefficient of variation of expression levels of *KLF1* and *FLI1* changing over time in the processes from EBs to iMKs. Unit of time is arbitrary. (F) Identification of distinct temporal patterns of expression variance by fuzzy c-means clustering. The x-axis represents four time points, while the y-axis represents scaled CV (coefficient of variation) in each time point. Dark trend lines in the middle indicate the average of scaled CV over genes in cluster. (G) Enriched major Gene Ontology terms for clusters 5 and 10. (H) Regulatory network of transcription factors (TFs) in cluster 5 and 10. Circle size indicates the sum of in-degree and out-degree. Node colors indicate different Supermodules (adapted from **Qin et al., 2022**). Green and red edges indicate activation and inhibition, respectively. The light blue and light pink shades denote genes in clusters 5 and 10, respectively.

The online version of this article includes the following figure supplement(s) for figure 7:

**Figure supplement 1.** The chemical-induced reprogramming of human erythroblast (EB) to induced megakaryocyte (iMK) is the signal-driven fate decisions.

upstream progenitor state, which is in agreement with the phenomenon articulated by **Qin et al., 2022**.

We then searched for genes with similar patterns of expression variance to those of *KLF1* and *FLI1*, with a hypothesis that genes possessing comparable expression variance patterns, especially TFs, may synergistically perform fate-decision related functions. Thus, we applied the fuzzy c-means algorithm (**Kumar and Futschik, 2007**) to cluster genes based on their expression variances, rather than expression levels. In total, we observed 12 distinct clusters of temporal patterns (**Figure 7—figure supplement 1C; Supplementary file 2**). We focused primarily on clusters 5 and 10, where *KLF1* and *FLI1* are found respectively (**Figure 7F**). To testify our hypothesis, we conducted an enrichment analysis using the gene set in clusters 5 and 10. Functions associated with specific cell types are significantly enriched (**Figure 7G**). In particular, cluster 5 with *FLI1* is largely related to platelet-related functions, whereas cluster 10 with *KLF1* is largely related to the energy metabolism of blood cells. Furthermore, we filtered out 15 TFs in clusters 5 and 10 (11 in cluster 5; 4 in cluster 10). Next, by harnessing the TF interaction database (see ‘Materials and methods’; **Supplementary file 3**), we collected 21 regulons associated with 10 TFs to construct the TF regulatory Network (**Figure 7H**). Intriguingly, in the original article, genes were classified into six ‘Supermodules’ according to the patterns of expression levels. Genes in cluster 5 are located in Supermodules 1 and 3, representing a decreasing tendency for expression level. Meanwhile, genes in cluster 10 are distributed in Supermodules 2 and 5, and their expression levels raise from low to high (Figure 6D from **Qin et al., 2022**). Of note, most of the TFs filtered by expression variance patterns appear in the GRN constructed in the original article (8/10, 80%; Figure 6G from **Qin et al., 2022**). As a result, we underscored that *EGR3* and *ETS1*, which did not appear in the GRN of the original article, have been suggested to play important roles in the trans-differentiation. In addition, we observed that *MYC* and *FOS* possess the largest connectivity in our 10-node GRN, suggesting that these two TFs as hubs are essential regulators of this reprogramming system.

Together, in mapping our framework to the real reprogramming system, we assigned the cell fate decisions of EBs to iMKs to the signal-driven mode incorporated the OR-OR-like motif, by comparing the expression and expression variance patterns measured from the real dataset with pseudo-data produced by our models in a ‘top-down’ fashion. According to the model, the reprogramming is primarily driven by induced up-regulation of *FLI1*. Additionally, from the view of expression variance, we recapitulated a concise 10-node GRN and identified some TF nodes not previously recognized as major regulators, like *EGR3* and *ETS1*.

## Discussion

Comprehending the driving forces behind cell fate decisions is crucial for both fundamental scientific research and biomedical engineering. Recent advances in data collection and statistical methods have greatly enhanced our understanding of the mechanisms that regulate cell fate. However, despite the widespread use of the Waddington landscape as a metaphor in experiments, there has been

little examination into whether the fate decision observed in a particular experiment corresponds to a stochastic shift from one attractor to another on the landscape (i.e., noise-driven) or an overall distortion of the landscape (i.e., signal-driven). The application of appropriate computational models in systems biology can aid in uncovering the underlying mechanisms (Cahan et al., 2021; del Sol and Jung, 2021). One of the most representative work is that Huang et al., 2007 modeled the bifurcation in hematopoiesis to reveal the lineage commitment quantitatively. Compared to simply modularizing activation or inhibition effect by employing Hill function in previous work, our models reconsidered the multiple regulations from the level of TF-CRE binding.

Our computational investigations have emphasized the importance of the combinatorial logic in connecting gene expression patterns with the driving forces underlying cell fate decisions. Utilizing a representative network topology known as the CIS, our analysis demonstrated how both driving forces and regulatory logic jointly shape expression patterns during fate transitions. In turn, mean and variance in gene expression patterns can reveal logic motifs and driving forces. Our analytical framework promotes the interpretability of fate decisions and can be employed to speculate on the driving factors of fate decisions using a 'top-down' approach, thereby providing a reference for investigating the causality of fate decisions and experimental validation.

The roles of noise as a possible driver of fate transitions are intriguing. By our models, the relationship among noise configuration, logic motifs, and fate-decision bias has been unveiled, and we noticed the opposite fate bias for the AND-AND and OR-OR motifs. Conversely, on the demand of the same bias, the progenitors with different logic motifs tend to employ a different profile of noise level (Figure 3E and F). Therefore, we suggested that under the noise-driven mode, the logic motif works like a 'broker' to shape the fate preferences. Based on the assumption that the preference of fate decisions is the result of evolution and adaptation, we posited that if an organism has a functional demand for a specific bias, the noise profile will be iterated via logic motifs. In turn, changes in noise levels will be mediated by logic motifs to shape the differentiation bias. One intuitive example is the significant shift in differentiation preferences of HSCs over aging (López-Otín et al., 2013; Beerman and Rossi, 2015; Dorshkind et al., 2020). It has been reported that aging HSCs show different levels of DNA methylation and epigenetic histone modifications from young HSCs (de Haan and Lazare, 2018; Kovtonyuk et al., 2016; Pang et al., 2017). How changes in epigenetic level shape the noise profile of the cell population and further affect the shift of fate-decision preference is a fascinating question. We underscored that the dissection of logic motifs underlying associated GRNs is a prerequisite for answering that question.

With the ever-expanded of single-cell sequencing data, characterizing genes by their mean expressions does not make the most of high-throughput datasets. As an intrinsic characteristic in central dogma, expression variance has been utilized to locate the 'critical transitions' in complex networks (Wang and Chen, 2020; Scheffer et al., 2012), e.g., identify the critical transitions in diseases like lymphoma (Liu et al., 2012). Instead of differentially expressed genes, Rosales-Alvarez et al., 2022 harnessed differentially noisy genes to characterize the functional heterogeneity in HSC aging. Our work presents that the patterns of expression variance can also be used to indicate driving forces and key regulators during the fate-decision processes. Nevertheless, compared to traditional gene expression, the interpretation of expression variance patterns is generally not intuitively accessible (Grün et al., 2014). Additionally, extra a priori knowledge is needed to filter out the cluster of interest. To this end, our framework enables researchers to locate functional clusters via mathematic model based on appropriate hypothesis. Notably, if the genes that constituting the CIS network are not specified, we can conversely leverage the patterns of temporal expression variance to nominate key regulators in a model-guided manner. Collectively, our framework provides a mechanistic explanation for expression variance patterns and qualitatively characterize key expression variance patterns to locate core regulators of fate decisions without reliance on a priori knowledge.

Comparing to tuning noise, altering signals is a more accessible approach for experimentally manipulating cell fates. When the basal expressions of two lineage-specifying genes grow symmetrically, we have shown that opposite trends of fate transitions occur under two logic motifs: In the AND-AND motif, it promotes differentiation, whereas the OR-OR motif stabilizes stem cell fates. This is reminiscent of the 'seesaw' model where maintenance of stemness can be achieved by overexpression of antagonistic lineage-specifying genes (Shu et al., 2013). Our model suggests that restoring stemness by inducing two antagonistic lineage-specifying genes is more likely under the OR-OR-like

motif (**Figure 4—figure supplement 1C**). This is in concert with our analysis that mESC differentiation system performs in an OR-OR-like manner. In addition, **Mojtahedi et al., 2016** found that, under simultaneous induction of two antagonistic fates (EPO and GM-CSF/IL-3), although differentiation was delayed, it eventually occurred. This observation is consistent with models in the AND-AND motif, and we suggested that the core regulatory circuits in hematopoiesis performs in an AND-AND-like manner. More experimental validations would be needed to validate this hypothesis that ‘the seesaw model prefers the OR-OR motif.’ Conversely, insight from the ‘seesaw’ model also provides a candidate approach to further testify logic motifs underlying GRNs in experiments. For instance, stem cells with an AND-AND-like circuit are expected to display differentiation rather than maintenance in response to bidirectional induction.

In quantifying the signal-driven landscape changes, the solution landscape enables intuitive interpretation even in high-dimension GRNs (**Yin et al., 2020; Sáez et al., 2022; Corson and Siggia, 2012**). In this work, we used both the state space and the solution landscape, in order to relate them for further investigations involving more than two TFs. Interestingly, from the perspective of the solution landscape, we found a robust fully-connected stage in the AND-AND motif. We envisioned that this period corresponds to the priming stage of differentiation. Notably, this fully-connected stage was not found in the OR-OR motif, suggesting that the necessity for priming during differentiation may be subject to the logic motifs of core GRNs.

Actual cell fate decisions are seldom purely unbiased. Under the asymmetrical signal-driven mode, we summarized the progression-accuracy trade-off in cell fate decisions: If a large number of cells are ensured to differentiate, then concessions have to be made in the accuracy of differentiation, and vice versa (**Figure 5—figure supplement 1E**). An intuitive example is the large-scale apoptosis occurs daily in hematopoiesis. Hence, maintaining homeostasis *in vivo* inevitably requests cells to respond rapidly to differentiation. A recent study reported that in response to systemic inflammation by polymicrobial sepsis, pool of CMPs is rapidly depleted to accelerate the production of downstream cell fates (**Fanti et al., 2023**). This is concordant with our result that *Gata1-PU.1* circuit in hematopoiesis performs in an AND-AND-like manner. On the other hand, in embryonic development like *C. elegans*, the accuracy of cell fate decisions is considerably emphasized. How this nature of differentiation has been adapted in evolution is an interesting question. Our work highlights that this property is associated with the logic motifs of GRNs, suggesting the emphasis on progression or accuracy may be embedded in the logic motifs of core GRNs.

We classified three examples of cell fate decisions based on patterns of expression and expression variance. In hematopoiesis, we took fate choice between erythroid and myeloid as a paradigm, and assigned it an AND-AND-like motif under the signal-driven mode. In embryogenesis, we suggested the fate decision in RA exposure system is an OR-OR-like motif under the noise-driven mode. In reprogramming, the chemical-induced trans-differentiation is the signal-driven fate decisions incorporated an OR-OR-like motif. For simplicity and intuitiveness, we devised our model with two symmetrical combinations of regulatory logic (AND-AND/OR-OR). Albeit there are merely four types of cell fate decisions in consideration, our framework enables to be generalized and expanded to accommodate multi-node GRNs and complex logic combinations. Plenty of studies zoomed in one particular fate-decision events. However, from the standpoint of systems biology, we underlined that classification of fate decisions is a vital step for further investigation, as is the case for the typing of cells and tumors. Theoretically, appropriate classification of fate-decision systems enables the enrichment of common properties. So, accumulated knowledge can be inherited to new fate-decision cases. Taking reprogramming as an example, **Zhao, 2019** recapitulated five kinds of trajectories in chemical-induced reprogramming. We suggested that the reprogramming trajectory is coupled with the logic motifs. On one hand, it is possible to answer why a certain reprogramming system exhibit a particular trajectory. On the other hand, it is possible to postulate achievable reprogramming according to the logic motifs of core GRNs (e.g., the AND-AND motif is more likely to enable direct conversion; model 4 mentioned in **Zhao, 2019**). Recently, synthetic biology has realized the insertion of the CIS network in mammalian cells (**Zhu et al., 2022**). One of the prerequisites for recapitulating the complex dynamics of fate transitions in synthetic biology is systematical understanding of the role of GRNs and driving forces in differentiation. And the logic motifs are the essential and indispensable elements in GRNs. Our work also provides a blueprint for designing logic motifs with particular functions. We are also interested in validating the conclusions drawn from our models in a synthetic biology system.

## Limitation of this study

Although our framework enables the investigation of more logic motifs, we chose two classical and symmetrical logic combinations for our analysis. Future work should involve more logic gates like XOR and explore asymmetrical logic motifs like AND-OR. The gene expression datasets analyzed here are only available for a limited number of time points. Though they meet the need for discerning trends, it is evident that the application to the datasets with more time points will yield clearer and less ambiguous changing trends to support the conclusions of this paper more generally. Notwithstanding the fact that the CIS network is prevalent in fate-decision programs, there are other topologies of networks that serve important roles in the cell-state transitions, like feed-forward loop, etc. The framework should further incorporate diverse network motifs in the future. In addition, for simplicity and intuition, we here considered signals as uncoupled and additive effects in ODE models, due to feasible mapping in real biological systems, such as ectopic overexpression.

## Materials and methods

### Derivation of the CIS network

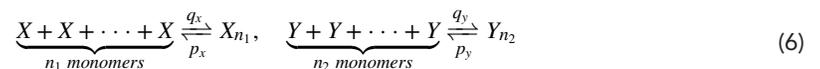
In GRN, each TFs is represented by a node and the edge between nodes represents the regulatory relationship. In our work, we considered a GRN comprised of 2 TFs (i.e., X, Y), formulated with 2 ODEs describing the change of each TF. Define  $[X]$ ,  $[Y]$  to be the concentration of TF X and Y, the ODE model is described as follows:

$$\begin{cases} \frac{d[X]}{dt} = H_X([X], [Y]) - d_X[X] \\ \frac{d[Y]}{dt} = H_Y([X], [Y]) - d_Y[Y] \end{cases} \quad (5)$$

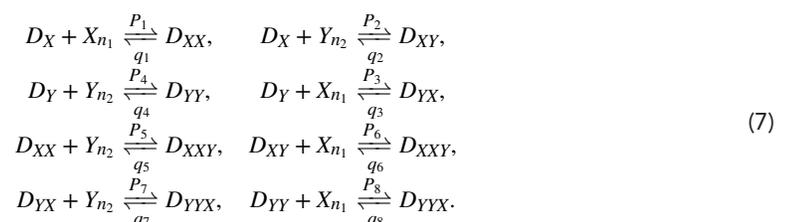
where  $H_X([X], [Y])$  is the production rate of TF X that combines the effects from both activators and suppressors of the X and  $d_X$  is the decay rate from the X and Y, which integrate degradation and dilution.

The exact form of  $H_X([X], [Y])$ ,  $H_Y([X], [Y])$  are derived based on the simple GRN with nodes X and Y via a set of molecular interactions between these TFs themselves, genes that encode for them, and the mRNAs (Abdallah and Del Vecchio, 2019).

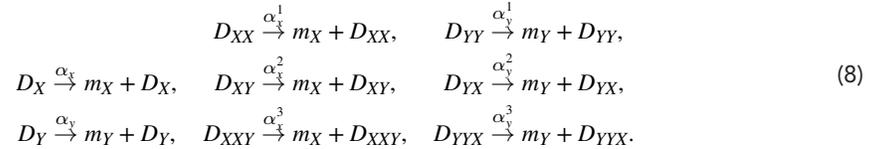
TFs in a GRN act as multimers to implement regulatory interaction. In our model, we treated TFs X and Y as acting in their homo-multimer forms,  $X_{n_1}$  and  $Y_{n_2}$ , respectively (i.e.,  $n_1$  TF X monomers reversibly form an activated homo-multimerized form  $X_{n_1}$  and  $n_2$  monomers of TF Y to reversibly form activated homo-multimer  $Y_{n_2}$ ). Of note,  $n_1$  and  $n_2$  here are able to be further generalized beyond the number of binding elements (Nam et al., 2022; Santillán, 2008). The multimerization biochemical reaction of TFs X and Y can be represented as follows,



There are many diverse mechanisms of regulation (Lambert et al., 2018; Balsalobre and Drouin, 2022). For simplicity, our model considered transcriptional regulation as a major mechanism since it is feasibly characterized in experiments and can well represent the interaction between genes in Cross-Inhibition with Self-activation (CIS) network. To activate downstream transcription, TFs bind with their Cis-Regulatory Elements (CREs). We denoted  $D_X$  for the no-bound CREs of TF X and  $D_Y$  for the TF Y in an independent manner. Here, we posited different TFs bind to exclusive, non-overlapping CREs to regulate target genes. Hence, there are totally eight binding patterns described as follows,



Next, we modeled the biochemical reactions of transcription and translation. Transcription is an elaborate process involving many steps, including initiation, elongation, and termination of an mRNA transcript. Likewise, translation includes peptide formation and elongation, followed by protein folding before function. However, to evaluate the behavior of GRN concisely, we treated transcription and translation as single-step reactions and then use lumped rates to encompass the time it takes for all steps in the elaborate machinery to complete (**Abdallah and Del Vecchio, 2019**). Hence, the transcription process can be represented as follows,



Here, we distinguished the basal rate of transcription of a gene without activation or repression (termed as constitutive transcription). The basal transcription rate is shown in the left panel in **Equation 8**.

Though we considered eight configurations, these ultimately led to two mRNA transcripts,  $m_X$  and  $m_Y$ . Translation can also be represented by a one-step process from these transcripts to their protein products. The one-step translation biochemical reactions are described as follows,



Next, we considered the decay of these proteins as well as their respective mRNA species. In general, these proteins and mRNA species will undergo decay to some extent which is a combination of both degradation and dilution.



here, the left and middle panels represent the degradation of TFs X and Y as well as mRNA transcripts  $m_X$  and  $m_Y$ . The right panel represents the dilution of TFs X and Y due to cell division. In general, the degradation rate  $\gamma$  can be determined using the following relations,

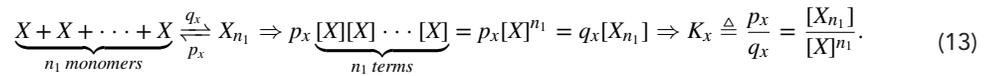
$$\delta = \frac{\ln 2}{t_{1/2}}, \quad \eta = \frac{\ln 2}{t_1}, \quad \beta \approx \frac{\ln 2}{t_{doubling}},
 \tag{11}$$

where  $t_{1/2}$  represents the half-lives of each protein or mRNA transcript,  $t_{doubling}$  represents the cell division rate. For protein products X and Y, we used the notation  $d$  to illustrate the total rate of decay rather than two separate parameters in our model, i.e.,  $d_1 = \delta_x + \beta$ ,  $d_2 = \delta_y + \beta$ , respectively. In general,  $t_{1/2}$  and  $t_{doubling}$  are robust and stationary, thus we treated these parameters as constant in our model, i.e.,  $d_i = \text{Const}$ , for  $i = 1, 2$ .

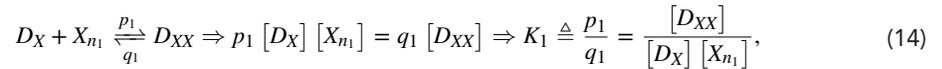
We have described the reactions that comprise the endogenous components of our GRN. In our work, fate decisions are classified into two modes, i.e., driven by the noise of intracellular gene expression or driven by extracellular signals. Under the noise-driven mode, a Gaussian white noise is added to the concentration of TFs to illustrate the random intracellular noise. Whereas if driven by signals, the expression of a gene will be affected by extracellular molecular cocktails, physical stimulation, etc. Therefore, we added the production rate of the TF's mRNA from the ectopic DNA to generalize our model.



here  $u_x$ ,  $u_y$  represent the additional mRNA species  $m_X$  and  $m_Y$  via ectopic overexpression at rates  $u_x$  and  $u_y$ , respectively. The biochemical reactions take place on a faster timescale. This enables us to derive the change in concentration of species from a biochemical reaction based on the law of mass action. For example, for multimerization reaction of X in **Equation 6**, once equilibrium, we can get



Also, for TF X bind with its CRE  $D_x$ , we can get



here, we define  $K_1 \triangleq \frac{p_1}{q_1}$  as a constant, i.e., chemical equilibrium constant. Furthermore, we define  $K_i \triangleq \frac{p_i}{q_i}, i = x, y, 1, 2, \dots, 8$  for biochemical reactions in **Equation 6** and **Equation 7**, respectively.

Then, we constructed an ODE model for the GRN. For a given species  $S$ , the principle to get a single equation is as follows:

$$\text{Change} \in \text{concentration of } S = \frac{d[S]}{dt} = [\dot{S}] = \sum \{ \text{all biochemical reaction rates involving } S \} \quad (15)$$

To explain how this principle assists us to build the ODE model, we took species  $m_x$  as a running example. Summarizing all of these biochemical reaction rates involving  $m_x$  we can get the change in concentration  $[m_x]$ ,

$$[\dot{m}_x] = \alpha_x [D_x] + \alpha_x^1 [D_{XX}] + \alpha_x^2 [D_{XY}] + \alpha_x^3 [D_{XXY}] - \eta_x [m_x] + u_x. \quad (16)$$

Doing so for our biochemical reaction network yielded the equations of our ODE model,

$$\left\{ \begin{array}{l} [\dot{X}] = \kappa_x [m_x] - d_1 [X] \\ [\dot{Y}] = \kappa_y [m_y] - d_2 [Y] \\ [\dot{m}_x] = \alpha_x [D_x] + \alpha_x^1 [D_{XX}] + \alpha_x^2 [D_{XY}] + \alpha_x^3 [D_{XXY}] - \eta_x [m_x] + u_x \\ [\dot{m}_y] = \alpha_y [D_y] + \alpha_y^1 [D_{YY}] + \alpha_y^2 [D_{YX}] + \alpha_y^3 [D_{YYX}] - \eta_y [m_y] + u_y \\ [\dot{D}_x] = -p_1 [X_{n_1}] [D_x] + q_1 [D_{XX}] - p_2 [Y_{n_2}] [D_x] + q_2 [D_{XY}] \\ [\dot{D}_y] = -p_4 [Y_{n_2}] [D_y] + q_4 [D_{YY}] - p_3 [X_{n_1}] [D_y] + q_3 [D_{YX}] \\ [\dot{D}_{XX}] = p_1 [D_x] [X_{n_1}] - q_1 [D_{XX}] - p_5 [D_{XX}] [Y_{n_2}] + q_5 [D_{XXY}] \\ [\dot{D}_{XY}] = p_2 [D_x] [Y_{n_2}] - q_2 [D_{XY}] - p_6 [D_{XY}] [X_{n_1}] + q_6 [D_{XXY}] \\ [\dot{D}_{YX}] = p_3 [D_y] [X_{n_1}] - q_3 [D_{YX}] + p_7 [D_{YX}] [Y_{n_2}] - q_7 [D_{YYX}] \\ [\dot{D}_{YY}] = p_4 [D_y] [Y_{n_2}] - q_4 [D_{YY}] + p_8 [D_{YY}] [X_{n_1}] - q_8 [D_{YYX}] \\ [\dot{D}_{XXY}] = p_5 [D_{XX}] [Y_{n_2}] - q_5 [D_{XXY}] + p_6 [D_{XY}] [X_{n_1}] - q_6 [D_{XXY}] \\ [\dot{D}_{YYX}] = p_7 [D_{YX}] [Y_{n_2}] - q_7 [D_{YYX}] + p_8 [D_{YY}] [X_{n_1}] - q_8 [D_{YYX}] \\ [\dot{X}_{n_1}] = p_x [X]^{n_1} - q_x [X_{n_1}] - p_1 [D_x] [X_{n_1}] + q_1 [D_{XX}] - p_3 [D_y] [X_{n_1}] + \\ q_3 [D_{YX}] - p_6 [D_{XY}] [X_{n_1}] + q_6 [D_{XXY}] - p_8 [D_{YY}] [X_{n_1}] + q_8 [D_{YYX}] \\ [\dot{Y}_{n_2}] = p_y [Y]^{n_2} - q_y [Y_{n_2}] - p_2 [D_x] [Y_{n_2}] + q_2 [D_{XY}] - p_4 [D_y] [Y_{n_2}] + \\ q_4 [D_{YY}] - p_5 [D_{XX}] [Y_{n_2}] + q_5 [D_{XXY}] - p_7 [D_{YX}] [Y_{n_2}] + q_7 [D_{YYX}] \end{array} \right. \quad (17)$$

The 14-dimension ODE model of the GRN contains overwhelming parameters that make it impractical to analyze for cell fate. To reduce the dimensionality of the model, we assumed the multimerization, DNA binding/unbinding and mRNA dynamic occurs sufficiently faster than protein production and decay, the temporal derivatives of the respective species can be set to 0, indicating that the species concentration reaches its quasi-steady state (**Abdallah and Del Vecchio, 2019**). Thus, we can get

$$\begin{cases} [\dot{X}] = \kappa_x [m_X] - d_1 [X] \\ [\dot{Y}] = \kappa_y [m_Y] - d_2 [Y] \\ [m_X] = 0 = \alpha_x [D_X] + \alpha_x^1 [D_{XX}] + \alpha_x^2 [D_{XY}] + \alpha_x^3 [D_{XXY}] - \eta_x [m_X] + u_x \\ [m_Y] = 0 = \alpha_y [D_Y] + \alpha_y^1 [D_{YY}] + \alpha_y^2 [D_{YX}] + \alpha_y^3 [D_{YYX}] - \eta_y [m_Y] + u_y. \end{cases} \quad (18)$$

To better demonstrate the deviation of the model, we take TF X as a running example. Since GRN is symmetric, we can deviate TF Y's equation in the same way. Once the biochemical reaction reaches its equilibrium, using the law of mass action we can get

$$\begin{aligned} [X_{n_1}] &= K_x [X]^{n_1} \\ [D_{XX}] &= K_1 [D_X] [X_{n_1}] = K_1 [D_X] [X_{n_1}] = K_x K_1 [D_X] [X]^{n_1} \\ [D_{XY}] &= K_1 [D_X] [Y_{n_2}] = K_1 [D_X] [Y_{n_2}] = K_y K_2 [D_X] [Y]^{n_2} \\ [D_{XXY}] &= K_5 [D_{XX}] [Y_{n_2}] = K_5 [D_{XX}] [Y_{n_2}] = K_x K_y K_1 K_5 [D_X] [X]^{n_1} [Y]^{n_2} \end{aligned} \quad (19)$$

Noticing that  $[m_X] = 0$  in **Equation 18**, the species  $[m_X]$  can be represented by CRE terms,

$$\begin{aligned} [m_X] &= \frac{1}{\eta_x} \left( \alpha_x [D_X] + \alpha_x^1 [D_{XX}] + \alpha_x^2 [D_{XY}] + \alpha_x^3 [D_{XXY}] \right) + \frac{u_x}{\eta_x} \\ &= \frac{1}{\eta_x} \left( \alpha_x [D_X] + \alpha_x^1 K_1 [D_X] [X_{n_1}] + \alpha_x^2 K_2 [D_X] [Y_{n_2}] + \alpha_x^3 K_1 K_5 [D_X] [X_{n_1}] [Y_{n_2}] \right) + \frac{u_x}{\eta_x} \\ &= \frac{1}{\eta_x} \left( [D_X] + \alpha_x^1 K_x K_1 [D_X] [X]^{n_1} + \alpha_x^2 K_y K_2 [D_X] [Y]^{n_2} + \alpha_x^3 K_x K_y K_1 K_5 [D_X] [X]^{n_1} [Y]^{n_2} \right) + \frac{u_x}{\eta_x} \end{aligned} \quad (20)$$

In our well-stirred system, for TF X's CREs, the conservation law holds, i.e., the total CRE concentration of X, as a constant, equals the sum of CRE concentration that is bound by X, by Y, and by X and Y. By employing so and combining with **Equation 19** we can get a representation of  $[D_X]$ ,

$$\frac{[D_{TX}] = [D_X] + [D_{XY}] + [D_{XX}] + [D_{XXY}] \Rightarrow [D_X] = [D_{TX}]}{(1 + K_x K_1 [X]^{n_1} + K_y K_2 [Y]^{n_2} + K_x K_y K_1 K_5 [X]^{n_1} [Y]^{n_2})}. \quad (21)$$

Substitute  $[D_X]$  in **Equation 20** with **Equation 21**, we can get

$$\begin{aligned} [m_X] &= \frac{1}{\eta_x} [D_X] (\alpha_x + \alpha_x^1 K_x K_1 [X]^{n_1} + \alpha_x^2 K_y K_2 [Y]^{n_2} + \alpha_x^3 K_x K_y K_1 K_5 [X]^{n_1} [Y]^{n_2}) + \frac{u_x}{\eta_x} \\ &= \frac{D_{TX} \kappa_x (\alpha_x + \alpha_x^1 K_x K_1 [X]^{n_1} + \alpha_x^2 K_y K_2 [Y]^{n_2} + \alpha_x^3 K_x K_y K_1 K_5 [X]^{n_1} [Y]^{n_2})}{\eta_x (1 + K_x K_1 [X]^{n_1} + K_y K_2 [Y]^{n_2} + K_x K_y K_1 K_5 [X]^{n_1} [Y]^{n_2})} + \frac{u_x}{\eta_x}. \end{aligned} \quad (22)$$

Then finally we reach to the point that the deviation of  $[X]$ . In **Equation 18** we substituted  $[m_X]$  with **Equation 22**, and got

$$\begin{aligned} [\dot{X}] &= \kappa_x [m_X] - d_1 [X] \\ &= \frac{D_{TX} \kappa_x (\alpha_x + \alpha_x^1 K_x K_1 [X]^{n_1} + \alpha_x^2 K_y K_2 [Y]^{n_2} + \alpha_x^3 K_x K_y K_1 K_5 [X]^{n_1} [Y]^{n_2})}{\eta_x (1 + K_x K_1 [X]^{n_1} + K_y K_2 [Y]^{n_2} + K_x K_y K_1 K_5 [X]^{n_1} [Y]^{n_2})} + \frac{\kappa_x u_x}{\eta_x} - d_1 [X] \end{aligned} \quad (23)$$

here, we define  $\gamma_0^x = \alpha_x \frac{D_{TX} \kappa_x}{\eta_x}$ ,  $\gamma_1 = \alpha_x^1 \frac{D_{TX} \kappa_x}{\eta_x}$ ,  $\gamma_2 = \alpha_x^2 \frac{D_{TX} \kappa_x}{\eta_x}$ ,  $\gamma_3 = \alpha_x^3 \frac{D_{TX} \kappa_x}{\eta_x}$  as relative protein generation rates,  $k_1 = K_1 K_x$ ,  $k_2 = K_2 K_y$ ,  $k_3 = K_x K_y K_1 K_5$  as the weight of each TF binding patterns, as ectopic DNA (to simplify the notation, we may drop the tilde of  $\tilde{u}_i, i = x, y$ ), then we can get

$$[\dot{X}] = \frac{r_0^x + r_1 k_1 [X]^{n_1} + r_2 k_2 [Y]^{n_2} + r_3 k_3 [X]^{n_1} [Y]^{n_2}}{1 + k_1 [X]^{n_1} + k_2 [Y]^{n_2} + k_3 [X]^{n_1} [Y]^{n_2}} - d_1 [X] + u_x, \quad (24)$$

Similarly, for TF Y, we can get

$$\begin{aligned}
 [\dot{Y}] &= \frac{D_{TY}\kappa_y (\alpha_y + \alpha_y^1 K_y K_3 [Y]^{n_2} + \alpha_y^2 K_y K_4 [X]^{n_1} + \alpha_y^3 K_x K_y K_3 K_7 [X]^{n_1} [Y]^{n_2})}{\eta_y (1 + K_y K_3 [Y]^{n_2} + K_y K_4 [X]^{n_1} + K_x K_y K_3 K_7 [X]^{n_1} [Y]^{n_2})} + \frac{\kappa_y u_y}{\eta_y} - d_2 [Y] \\
 &= \frac{r_0^y + r_4 k_4 [Y]^{n_2} + r_5 k_5 [X]^{n_1} + r_6 k_6 [X]^{n_1} [Y]^{n_2}}{1 + k_4 [Y]^{n_2} + k_5 [X]^{n_1} + k_6 [X]^{n_1} [Y]^{n_2}} - d_2 [Y] + u_y.
 \end{aligned}
 \tag{25}$$

Ultimately, by combining **Equation 24** and **Equation 25**, we reduced our 14-dimension model to a 2-dimension dynamic system,

$$\begin{cases} \frac{d[X]}{dt} = \frac{r_0^x + r_1 k_1 [X]^{n_1} + r_2 k_2 [Y]^{n_2} + r_3 k_3 [X]^{n_1} [Y]^{n_2}}{1 + k_1 [X]^{n_1} + k_2 [Y]^{n_2} + k_3 [X]^{n_1} [Y]^{n_2}} - d_1 [X] + u_x \\ \frac{d[Y]}{dt} = \frac{r_0^y + r_4 k_4 [Y]^{n_2} + r_5 k_5 [X]^{n_1} + r_6 k_6 [X]^{n_1} [Y]^{n_2}}{1 + k_4 [Y]^{n_2} + k_5 [X]^{n_1} + k_6 [X]^{n_1} [Y]^{n_2}} - d_2 [Y] + u_y, \end{cases}
 \tag{26}$$

### Stochastic simulations

In **Equation 22**, parameters of the AND-AND motif (**Figure 2C** top panel) are:  $(r_0^x, r_0^y, r_1, r_2, r_3, r_4, r_5, r_6) = (0, 0, 0.95, 0, 0, 0.95, 0, 0)$ ,  $(k_1, k_2, k_3, k_4, k_5, k_6) = (1.5, 0.8, 1, 1.5, 0.8, 1)$  and  $(d_1, d_2) = (0.55, 0.55)$ . Parameters of the OR-OR motif (**Figure 2C** bottom panel) are:  $(r_0^x, r_0^y, r_1, r_2, r_3, r_4, r_5, r_6) = (0.15, 0.15, 0.95, 0, 0.95, 0, 0.95, 0)$ ,  $(k_1, k_2, k_3, k_4, k_5, k_6) = (2.1, 1.7, 1.8, 2.1, 1.7, 1.8)$  and  $(d_1, d_2) = (0.55, 0.55)$ . In both the AND-AND and OR-OR motifs,  $(n_1, n_2) = (2, 2)$ . We simulated noise in the model using a Langevin equation (**Foster et al., 2009**):

$$dx_i = F_i(x) dt + dW_i,
 \tag{27}$$

Where  $F_i(x)$  is the deterministic function of **Equation 22**.  $W_i$  is a Wiener process which introduces additive noise with no dependence on the state  $x$ . We integrated the Wiener process over the interval which gives us  $\sqrt{\Delta t} \xi_i$  and  $\xi_i$  is a Gaussian white noise with zero mean and given variance  $(\sigma^2)$ . Thus we can compute the new state of the dynamical systems using the Runge–Kutta method.

### Parameter screening

We conducted global parameters screening of the fully connected stage and Progression-Accuracy trade-off in our models. First, to collect parameter sets with 3 SSSs, we used Latin hypercube sampling (LHS) to screen k-series parameters symmetrically (i.e.,  $k_1 = k_4, k_2 = k_5, k_3 = k_6$ ) ranging from 0.001 to 5 both in the AND-AND and OR-OR motifs. We ultimately collected 6,231 sets for the AND-AND motif and 6682 sets for the OR-OR motifs (**Supplementary file 1**).

To analyze the sequence of vanishing SSSs. We further filtered parameter sets with 2 SSSs remained as increasing  $u_x$  (6207 sets for the AND-AND motif; 6634 sets for the OR-OR motif). For each SSS, we quantified the Minus of  $[X]$  and  $[Y]$  (**Figure 5—figure supplement 1C, D**).

### Saddle points and saddle dynamics

Set an autonomous dynamical system (**Yin et al., 2021**)

$$\dot{x} = F(x), x \in R^n
 \tag{28}$$

where  $F: R^n \rightarrow R^n$  is a  $C^r$  ( $r \geq 2$ ) function, and a point  $x^* \in R^n$  is called a stationary point (or equilibrium solution) of **Equation 28** if  $F(x^*) = 0$ . Let  $J(x) = \nabla F(x)$  denote the Jacobian of  $F(x)$ . For a stationary point  $x^*$ , taking  $x = x^* + y$  in **Equation 28**, we have

$$\dot{y} = J(x^*) y + O(\|y\|^2),
 \tag{29}$$

where  $\| \cdot \|$  denotes the norm induced by the inner product. The associated linear system

$$\dot{y} = J(x^*) y,
 \tag{30}$$

is used to determine the stability of  $x^*$ . Depending on the eigenvalues of  $J(x)$  with positive, negative, and zero real parts, we can define unstable, stable, and center manifolds of the Jacobian  $J(x)$  spanned by the corresponding eigenvectors, as  $W^u(x^*)$ ,  $W^s(x^*)$  and  $W^c(x^*)$ .

From the primary decomposition theorem,  $R^n$  can be decomposed as a direct sum:

$$\mathbb{R}^n = W^u(x) \oplus W^s(x) \oplus W^c(x), \tag{31}$$

A hyperbolic stationary point is called a saddle if  $W^u(x^*)$  and  $W^s(x^*)$  are nontrivial. The hyperbolic stationary point  $x^*$  is called a sink (source) if all the eigenvalues of  $J(x^*)$  have negative (positive) real parts. The index of a stationary point  $x^*$  is defined as the dimension of the unstable subspace  $W^u(x^*)$ .

The HiOSD method (Yin et al., 2019) is designed for finding index- $k$  saddles of an energy function  $E(x)$ . For gradient systems,  $F(x) = -\nabla E(x)$ , and the Jacobian  $J(x) = -\nabla^2 E(x) = -G(x)$ , where  $G(x)$  denotes the Hessian of  $E(x)$ . The  $k$ -saddle  $x^*$  is a local maximum on the linear manifold  $x^* + W^u(x^*)$  and a local minimum on  $x^* + W^s(x^*)$ . So, the high-index saddle dynamics (HiSD) for a index  $k$  saddle ( $k$ -saddle) is a transformed gradient flow

$$\dot{x} = -P_{W^u(x)}F(x) + (F(x) - P_{W^u(x)}F(x)) = (I - 2P_{W^u(x)})F(x), \tag{32}$$

where  $P_V$  denotes the orthogonal projection operator on a finite-dimensional subspace  $V$ . Here,  $-P_{W^u(x)}F(x)$  is taken as an ascent direction on the subspace  $W^u(x)$  and  $F(x) - P_{W^u(x)}F(x)$  is a descent direction on the subspace  $W^s(x)$ . The subspace  $W^u(x) = span\{v_1, \dots, v_k\}$  where  $v_i$  is the unit eigenvector corresponding to the smallest  $i$ -th eigenvalues, which can be obtained by many methods such as minimize the Rayleigh quotients. According to the above the HiSD for a  $k$ -saddle ( $k$ -HiSD) is:

$$\begin{cases} \beta^{-1}\dot{x} = \left(I - \sum_{i=1}^k 2v_i v_i^T\right) F(x), \\ \gamma^{-1}\dot{v}_i = -\left(I - v_i v_i^T - \sum_{j=1}^{i-1} 2v_j v_j^T\right) G(x) v_i, i = 1, 2, \dots, k, \end{cases} \tag{33}$$

which coupled with the initial condition

$$x(0) = x^{(0)} \in \mathbb{R}^n, v_i(0) = v_i^{(0)} \in \mathbb{R}^n, \text{ s.t. } \langle v_j^{(0)}, v_i^{(0)} \rangle = \delta_{ij}, i, j = 1, \dots, k \tag{34}$$

Where  $I$  is the identity operator and  $\beta, \gamma > 0$  are relaxation parameters.

Similarly, the GHiSD (Yin et al., 2021) for a  $k$ -saddle ( $k$ -GHiSD) of the dynamical system (28) has the following form:

$$\begin{cases} \dot{x} = \left(I - 2 \sum_{j=1}^k v_j v_j^T\right) F(x) \\ \dot{v}_i = \left(I - v_i v_i^T\right) J(x) v_i - \sum_{j=1}^{i-1} v_j v_j^T \left(J(x) + J^T(x)\right) v_i, i = 1, \dots, k \end{cases} \tag{35}$$

which coupled with the initial condition(34). The  $k$ -GHiSD can be accelerated by the Heavy Ball method.

### Constructing the solution landscape

For a given set of parameters, we can use  $k$ -GHiSD to find each order saddle point of the dynamical system, and then construct the solution landscape of the system. The solution landscape is a pathway map consisting of all stationary points and their connections (Yin et al., 2020; Yin et al., 2022). Solution landscape is a new tool used to describe the dynamic behavior of stationary points in a dynamic system which can show the connection and transfer path between stationary points. In general, we can use the downward search algorithm and the upward search algorithm to construct the solution landscape, which starting from the  $k$ -saddle points and the stable points respectively. Because of the symmetry and other prior knowledge of the dynamic system used in our paper, The saddle points of the system tend to occur in the range of  $[0, x_{max}]^2$ , where the  $x_{max}$  means the maximum coordinate of the stable point. We grid the range, select each grid point as the initial state, and find saddle points of corresponding order ( $k = 1, 2$ ) using  $k$ -GHiSD. Taking the saddle points found before as the initial

states, we disturb its unstable directions to search other stationary points of lower index, establish the connection relationship between stationary points and finally construct the solution landscape. In addition, we use geometric minimum action method (gMAM) ([Vanden-Eijnden and Heymann, 2008](#)) to find the minimum action path between the stable points, use the obtained action to represent the stability of the stable points, and represent them at different heights in the solution landscape.

## Screening the fully-connected stage

As shown in the case of  $u = 0.0565$  in [Figure 4D](#), the fully connected stage (FCS) has three stable points and three 1-index saddle points to connect the stable points with each other. Therefore, we posited that finding a fully connected stage can be equivalent to finding three different 1-index saddle points in this system.

First, build a set of parameters with three stable points. In two different logics, 10,000 sets of dynamic parameters are randomly generated respectively, and all stable points of the system corresponding to each set of parameters are found. If there are three stable points, the corresponding dynamic parameters are recorded as a new set of parameters for the subsequent search of 1-index saddle points.

Second, finding the 1-index saddle points in a system with three stable points. According to the above results, we can find that saddle points tend to appear in the range of  $[0, x_{max}]^2$ , where the  $x_{max}$  means the maximum coordinate of the stable point. So we can grid the range in the same process as before, and search 1-index saddle points with each grid point as the initial structure. In order to find the relevant saddle points as detailed as possible, denser grid points were used as the initial structure of the system without three 1-index saddle points to find the 1-index saddle points.

Due to the calculation error of numerical calculation itself, the saddle points found by saddle point dynamics may have the following problems:

1. The same saddle points found from different initial structures were mistaken for two.
2. Find a negative or infinite saddle point.
3. The saddle-point judgment is inaccurate due to the influence of zero eigenvalue.

We can set corresponding judgment conditions to solve the above problems.

In addition, since there is also a parameter  $u$  representing the foreign signal in the system, we can let  $u$  vary to find if there are three 1-index saddle points. At the beginning,  $u = 0$  is set, and then  $u$  is changed according to the number of saddle points found and the corresponding judgment conditions, so as to find the system with three 1-index saddle points. We use the idea of adaptive step size to speed up the search process in the algorithm of finding three 1-index saddle points in the system. The specific process is as follows:

- Set the initial  $u = 0$ ,  $du = 1$ .
- Saddle Dynamics is used to find the 1-index saddle points in the parameters.
- When the number of 1-index saddle points is equal to 0: there is only one stable point in the system ( $u = 0.12$  in [Figure 4F](#)), it indicates that  $u$  is too large and needs to be reduced to make  $u = u - du$ .
- When the number of 1-index saddle points is equal to 1: the system has two stable points ( $u = 0.12$  in [Figure 4C](#)), it indicates that  $u$  is too large and needs to be reduced so that  $u = u - du$ .
- When the number of 1-index saddle points is equal to 2: the system has two kinds of cases, one of which needs to make  $u$  increase ( $u = 0$  in [Figure 4C](#)). In this case, the coordinate of stable point near  $y = x$  is smaller than the midpoint of two 1-index saddle points, so we let  $u = u + du$ , and let  $du = 0.1 * du$ . In the other case, there are two possibilities. One is like  $u = 0$  in [Figure 4F](#), and the other is like  $u = 0.0595$  in [Figure 4F](#). In this case, there is no stable point near  $y = x$  or the coordinate of the stable point is larger than the midpoint of two 1-index saddle points.
- When the system finds that the number of 1-index saddle points is equal to 3: we need to judge whether there are three non-degenerate saddle points that are different from each other. If the condition is sufficient, it indicates that we have found the fully connected solution landscape; otherwise, let  $u = u + du$ ,  $du = 0.1 * du$ , and continue searching.
- When the number of 1-index saddle points is greater than 3, it indicates that  $u$  is too large and needs to be reduced to make  $u = u - du$ .
- If three 1-index saddle points are not found after 100 changes of  $u$ , it is considered that there is no fully connected solution landscape in the system.

We, respectively, used the above algorithm to search for the fully connected solution landscape under two different logics and obtained the following results:

Logic	# of 3 SSS	# of FCS	Ratio
AND-AND	6231	5151	82.7%
OR-OR	6682	120	1.8%

It can be seen from the simulation results that the fully connected solution landscape can appear in most AND-AND motifs, but almost not in the OR-OR motifs. Due to the error of numerical calculation itself, the obtained results may have a certain deviation. In addition, in our search process, due to the limitation of computing resources and the complex nature of some systems, in order to take into account the majority of cases, some systems with full connectivity have not been found. Subsequent manual search found that the fully connected solution landscape is existed in some systems which not found under the AND-AND motifs previously (17.3%). Therefore, the obtained proportion can be further improved.

### Data processing of single-cell data

There are totally four public datasets used in our work (related to *Figure 6F and H* and *Figure 7* and *Figure 6—figure supplement 2C*). In mice's hematopoiesis (*Mojtahedi et al., 2016*), expression of genes is quantified as  $2^{\text{LOD}-\text{Cq}}-1$  (LOD: limit of detection; Cq-value is the cutoff of amplification signal in single-cell qPCR data). Thus, we can compute the coefficient of variation (CV) over time. In embryogenesis (*Semrau et al., 2017*), public dataset used in our work was generated by single-cell SMART-seq2 (GEO: GSE79578). We utilized preprocessed data to compute the CV over time. In reprogramming, public dataset was generated by 10 x Genomics (GEO: GSE207654). Counts Per Million (CPM) matrix are used to compute the CV. We assigned time points by the Leiden clustering algorithm with parameters consistent with the original paper (see methods in *Qin et al., 2022*).

### Fuzzy C-means clustering

CV of 2000 high variable genes given by original paper (*Qin et al., 2022*) over four time points were computed. We filtered out 1677 genes by removing NA due to missing values. Noise of 1677 genes were then grouped into 12 clusters using the Mfuzz package in R with fuzzy c-means algorithm (*Kumar and Futschik, 2007; Supplementary file 2*).

### GO enrichment analysis

GO and pathway enrichment analyses were performed using Metascape (*Zhou et al., 2019*).

### Integration of human TF-target interactions

A collection of 456,698 TF-target interactions for 1541 human TFs and their mode of regulation (MoR, activation or repression) were integrated by merging a built-in collection of human regulons from the DoRothEA R package (version 1.8.0) (*Garcia-Alonso et al., 2019*) and human TRRUST database (*Han et al., 2018*). To be specific, resource from DoRothEA includes 454,504 TF-target interactions for 1541 human TFs, retrieved from (1) literature-curated resources, (2) ChIP-seq binding data, (3) TFBS (TF binding sites) predictions, and (4) transcriptional regulatory interactions inferred from published gene expression profiles. Human TRRUST database is a manually curated database of transcriptional regulatory networks, whose current version contains 8427 TF-target regulatory relationships of 795 human TFs with annotation of MoR (activation, repression, or unknown). For our purpose, we excluded TF-target interactions with unknown MoR, yet 207 TF-target pairs still remained to show conflict MoRs in records from literature, according to TRRUST database. To integrate data from both resources, when a conflict occurred in TRRUST database, we retained the MoR recorded in DoRothEA resource for the same pair of TF-target, but omitted the conflict records if the TF-target pair was not included in DoRothEA resource. As for TF-target pairs without conflict MoRs in TRRUST database but shows conflicts across the two resources, records in TRRUST database were taken as more credible (*Supplementary file 3*).

## Integration of human TF list

In addition to the 1541 human TFs we obtained from the collection of TF-target interactions mentioned earlier, 1639 human TFs reported by *Lambert et al., 2018* and 1564 by *Ng et al., 2021* were used as complements, ending up with a total of 2051 human TFs (*Supplementary file 3*).

## Acknowledgements

We thank Y Zhao for advice on model application on real biological processes; D Grün for insightful and generous feedback on noise interpretation; J Zhang, Y Li, and X Pei for providing original raw-data processing R scripts; Z Zhou for illustrations and useful discussion; and the entire Zhiyuan laboratory for support and advice. This work was supported by the National Key Research and Development Program of China (No. 2021YFF1200500, 2021YFA0910700), and the National Natural Science Foundation of China (No.12225102, 12050002, and 12226316). It is also supported by grants from the Peking-Tsinghua Center for Life Sciences.

---

## Additional information

### Funding

Funder	Grant reference number	Author
National Key Research and Development Program of China	2021YFF1200500	Zhiyuan Li
National Natural Science Foundation of China	12225102	Zhiyuan Li
National Key Research and Development Program of China	2021YFA0910700	Zhiyuan Li
National Natural Science Foundation of China	12050002	Zhiyuan Li
National Natural Science Foundation of China	12226316	Zhiyuan Li

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

### Author contributions

Gang Xue, Conceptualization, Resources, Data curation, Software, Formal analysis, Investigation, Visualization, Methodology, Writing - original draft, Writing - review and editing; Xiaoyi Zhang, Software, Visualization, Methodology, Writing - review and editing; Wanqi Li, Data curation, Writing - review and editing; Lu Zhang, Xiaolin Zhou, Writing - review and editing; Zongxu Zhang, Visualization, Methodology, Writing - review and editing; Di Zhang, Resources; Lei Zhang, Supervision, Funding acquisition, Project administration, Writing - review and editing; Zhiyuan Li, Conceptualization, Supervision, Funding acquisition, Project administration, Writing - review and editing

### Author ORCIDs

Gang Xue  <http://orcid.org/0000-0002-4116-5819>  
Xiaoyi Zhang  <https://orcid.org/0009-0001-7015-4158>  
Zongxu Zhang  <https://orcid.org/0009-0006-2909-5842>  
Lei Zhang  <https://orcid.org/0000-0001-9972-2051>  
Zhiyuan Li  <http://orcid.org/0000-0001-6662-2636>

### Peer review material

Joint public review: <https://doi.org/10.7554/eLife.88742.3.sa1>  
Author response <https://doi.org/10.7554/eLife.88742.3.sa2>

## Additional files

### Supplementary files

- Supplementary file 1. Parameter sets for screening fully-connected stage in two logic motifs. The spreadsheet titled 'MAP' gives the relative energy in **Figure 4D and G**, **Figure 5E and G**, **Figure 7D**.
- Supplementary file 2. 1677 TFs divided into 12 clusters based on time-course CV. The spreadsheet lists all 12 clusters of TFs, as described in **Figure 7F**, **Figure 7—figure supplement 1C**.
- Supplementary file 3. Integrated human TF list and regulon database. The spreadsheet records the human TF list and regulon used in **Figure 7H**.
- MDAR checklist

### Data availability

All data analysed during this study are included in the manuscript and supporting files.

The following previously published datasets were used:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Semrau S, van Oudenaarden A	2017	Dynamics of lineage commitment revealed by single-cell transcriptomics of differentiating embryonic stem cells	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE79578">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE79578</a>	NCBI Gene Expression Omnibus, GSE79578
Qin J, Zhang J, Jiang J, Li Y, Pei X	2022	Direct chemical reprogramming of human cord blood erythroblasts to induced megakaryocytes that produce platelets	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE207654">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE207654</a>	NCBI Gene Expression Omnibus, GSE207654

## References

- Abdallah HM**, Del Vecchio D. 2019. Computational analysis of altering cell fate. *Methods in Molecular Biology* **1975**:363–405. DOI: [https://doi.org/10.1007/978-1-4939-9224-9\\_17](https://doi.org/10.1007/978-1-4939-9224-9_17), PMID: 31062319
- Aguadé-Gorgorió G**, Kauffman S, Solé R. 2021. Transition therapy: Tackling the ecology of tumor phenotypic plasticity. *Bulletin of Mathematical Biology* **84**:24. DOI: <https://doi.org/10.1007/s11538-021-00970-9>, PMID: 34958403
- Arinobu Y**, Mizuno S, Chong Y, Shigematsu H, Iino T, Iwasaki H, Graf T, Mayfield R, Chan S, Kastner P, Akashi K. 2007. Reciprocal activation of GATA-1 and PU.1 marks initial specification of hematopoietic stem cells into myeloerythroid and myelolymphoid lineages. *Cell Stem Cell* **1**:416–427. DOI: <https://doi.org/10.1016/j.stem.2007.07.004>, PMID: 18371378
- Balsalobre A**, Drouin J. 2022. Pioneer factors as master regulators of the epigenome and cell fate. *Nature Reviews. Molecular Cell Biology* **23**:449–464. DOI: <https://doi.org/10.1038/s41580-022-00464-z>, PMID: 35264768
- Beerman I**, Rossi DJ. 2015. Epigenetic control of stem cell potential during homeostasis, aging, and disease. *Cell Stem Cell* **16**:613–625. DOI: <https://doi.org/10.1016/j.stem.2015.05.009>, PMID: 26046761
- Bhattacharya S**, Zhang Q, Andersen ME. 2011. A deterministic map of Waddington's epigenetic landscape for cell fate specification. *BMC Systems Biology* **5**:85. DOI: <https://doi.org/10.1186/1752-0509-5-85>, PMID: 21619617
- Brand M**, Morrissey E. 2020. Single-cell fate decisions of bipotential hematopoietic progenitors. *Current Opinion in Hematology* **27**:232–240. DOI: <https://doi.org/10.1097/MOH.0000000000000592>, PMID: 32427596
- Buchler NE**, Gerland U, Hwa T. 2003. On schemes of combinatorial transcription logic. *PNAS* **100**:5136–5141. DOI: <https://doi.org/10.1073/pnas.0930314100>
- Cahan P**, Cacchiarelli D, Dunn S-J, Hemberg M, de Sousa Lopes SMC, Morris SA, Rackham OJL, Del Sol A, Wells CA. 2021. Computational stem cell biology: Open questions and guiding principles. *Cell Stem Cell* **28**:20–32. DOI: <https://doi.org/10.1016/j.stem.2020.12.012>, PMID: 33417869
- Casey MJ**, Stumpf PS, MacArthur BD. 2020. Theory of cell fate. *Wiley Interdisciplinary Reviews. Systems Biology and Medicine* **12**:e1471. DOI: <https://doi.org/10.1002/wsbm.1471>, PMID: 31828979
- Chang HH**, Hemberg M, Barahona M, Ingber DE, Huang S. 2008. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* **453**:544–547. DOI: <https://doi.org/10.1038/nature06965>, PMID: 18497826
- Chen Q**, Shi J, Tao Y, Zernicka-Goetz M. 2018. Tracing the origin of heterogeneity and symmetry breaking in the early mammalian embryo. *Nature Communications* **9**:1819. DOI: <https://doi.org/10.1038/s41467-018-04155-2>

- Collombet S**, van Oevelen C, Sardina Ortega JL, Abou-Jaoudé W, Di Stefano B, Thomas-Chollier M, Graf T, Thieffry D. 2017. Logical modeling of lymphoid and myeloid cell specification and transdifferentiation. *PNAS* **114**:5792–5799. DOI: <https://doi.org/10.1073/pnas.1610622114>
- Coomer MA**, Ham L, Stumpf MPH. 2022. Noise distorts the epigenetic landscape and shapes cell-fate decisions. *Cell Systems* **13**:83–102. DOI: <https://doi.org/10.1016/j.cels.2021.09.002>, PMID: 34626539
- Corson F**, Siggia ED. 2012. Geometry, epistasis, and developmental patterning. *PNAS* **109**:5568–5575. DOI: <https://doi.org/10.1073/pnas.1201505109>
- Davies J**, Vallejo AF, Sirvent S, Porter G, Clayton K, Qumbelo Y, Stumpf P, West J, Gray CM, Chigorimbo-Murefu NTL, MacArthur B, Polak ME. 2021. An IRF1-IRF4 toggle-switch controls tolerogenic and immunogenic transcriptional programming in human langerhans cells. *Frontiers in Immunology* **12**:665312. DOI: <https://doi.org/10.3389/fimmu.2021.665312>, PMID: 34211464
- de Haan G**, Lazare SS. 2018. Aging of hematopoietic stem cells. *Blood* **131**:479–487. DOI: <https://doi.org/10.1182/blood-2017-06-746412>
- del Sol A**, Jung S. 2021. The importance of computational modeling in stem cell research. *Trends in Biotechnology* **39**:126–136. DOI: <https://doi.org/10.1016/j.tibtech.2020.07.006>
- Del Vecchio D**, Abdallah H, Qian Y, Collins JJ. 2017. A blueprint for a synthetic genetic feedback controller to reprogram cell fate. *Cell Systems* **4**:109–120. DOI: <https://doi.org/10.1016/j.cels.2016.12.001>, PMID: 28065574
- Desai RV**, Chen X, Martin B, Chaturvedi S, Hwang DW, Li W, Yu C, Ding S, Thomson M, Singer RH, Coleman RA, Hansen MMK, Weinberger LS. 2021. A DNA repair pathway can regulate transcriptional noise to promote cell fate transitions. *Science* **373**:eabc6506. DOI: <https://doi.org/10.1126/science.abc6506>, PMID: 34301855
- Dixit A**, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, Marjanovic ND, Dionne D, Burks T, Raychowdhury R, Adamson B, Norman TM, Lander ES, Weissman JS, Friedman N, Regev A. 2016. Perturb-Seq: Dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* **167**:1853–1866. DOI: <https://doi.org/10.1016/j.cell.2016.11.038>, PMID: 27984732
- Domcke S**, Shendure J. 2023. A reference cell tree will serve science better than A reference cell atlas. *Cell* **186**:1103–1114. DOI: <https://doi.org/10.1016/j.cell.2023.02.016>, PMID: 36931241
- Dorshkind K**, Höfer T, Montecino-Rodriguez E, Pioli PD, Rodewald H-R. 2020. Do haematopoietic stem cells age? *Nature Reviews. Immunology* **20**:196–202. DOI: <https://doi.org/10.1038/s41577-019-0236-2>, PMID: 31740804
- Fanti A-K**, Busch K, Greco A, Wang X, Cirovic B, Shang F, Nizharadze T, Frank L, Barile M, Feyerabend TB, Höfer T, Rodewald H-R. 2023. Flt3- and Tie2-Cre tracing identifies regeneration in sepsis from multipotent progenitors but not hematopoietic stem cells. *Cell Stem Cell* **30**:207–218. DOI: <https://doi.org/10.1016/j.stem.2022.12.014>, PMID: 36652946
- Fisher AG**, Merckenschlager M. 2010. Fresh powder on Waddington's slopes. *EMBO Reports* **11**:490–492. DOI: <https://doi.org/10.1038/embor.2010.95>, PMID: 20559322
- Foster DV**, Foster JG, Huang S, Kauffman SA. 2009. A model of sequential branching in hierarchical cell fate determination. *Journal of Theoretical Biology* **260**:589–597. DOI: <https://doi.org/10.1016/j.jtbi.2009.07.005>, PMID: 19615382
- Garcia-Alonso L**, Holland CH, Ibrahim MM, Turei D, Saez-Rodriguez J. 2019. Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Research* **29**:1363–1375. DOI: <https://doi.org/10.1101/gr.240663.118>, PMID: 31340985
- Glauche I**, Marr C. 2021. Mechanistic models of blood cell fate decisions in the era of single-cell data. *Current Opinion in Systems Biology* **28**:e. DOI: <https://doi.org/10.1016/j.coisb.2021.100355>, PMID: 34950807
- Goldberg AD**, Allis CD, Bernstein E. 2007. Epigenetics: a landscape takes shape. *Cell* **128**:635–638. DOI: <https://doi.org/10.1016/j.cell.2007.02.006>, PMID: 17320500
- Graf T**, Enver T. 2009. Forcing cells to change lineages. *Nature* **462**:587–594. DOI: <https://doi.org/10.1038/nature08533>, PMID: 19956253
- Grün D**, Kester L, van Oudenaarden A. 2014. Validation of noise models for single-cell transcriptomics. *Nature Methods* **11**:637–640. DOI: <https://doi.org/10.1038/nmeth.2930>, PMID: 24747814
- Guillemin A**, Stumpf MPH. 2021. Noise and the molecular processes underlying cell fate decision-making. *Physical Biology* **18**:011002. DOI: <https://doi.org/10.1088/1478-3975/abc9d1>, PMID: 33181489
- Hamazaki T**, El Rouby N, Fredette NC, Santostefano KE, Terada N. 2017. Concise review: Induced pluripotent stem cell research in the era of precision medicine. *Stem Cells* **35**:545–550. DOI: <https://doi.org/10.1002/stem.2570>, PMID: 28100040
- Hammelman J**, Patel T, Closser M, Wichterle H, Gifford D. 2022. Ranking reprogramming factors for cell differentiation. *Nature Methods* **19**:812–822. DOI: <https://doi.org/10.1038/s41592-022-01522-2>, PMID: 35710610
- Han H**, Cho J-W, Lee S, Yun A, Kim H, Bae D, Yang S, Kim CY, Lee M, Kim E, Lee S, Kang B, Jeong D, Kim Y, Jeon H-N, Jung H, Nam S, Chung M, Kim J-H, Lee I. 2018. TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Research* **46**:D380–D386. DOI: <https://doi.org/10.1093/nar/gkx1013>, PMID: 29087512
- Hersbach BA**, Fischer DS, Masserdotti G, Mojžišová K, Waltzhöni T, Rodriguez-Terrones D, Heinig M, Theis FJ, Götz M, Stricker SH. 2022. Probing cell identity hierarchies by fate titration and collision during direct reprogramming. *Molecular Systems Biology* **18**:e11129. DOI: <https://doi.org/10.15252/msb.202211129>, PMID: 36106915

- Hoppe PS**, Schwarzfischer M, Loeffler D, Kokkalis KD, Hilsenbeck O, Moritz N, Ende M, Filipczyk A, Gambardella A, Ahmed N, Etzrodt M, Coutu DL, Rieger MA, Marr C, Strasser MK, Schaubberger B, Burtscher I, Ermakova O, Bürger A, Lickert H, et al. 2016. Early myeloid lineage choice is not initiated by random PU.1 to GATA1 protein ratios. *Nature* **535**:299–302. DOI: <https://doi.org/10.1038/nature18320>, PMID: 27411635
- Hota SK**, Rao KS, Blair AP, Khalilimeybodi A, Hu KM, Thomas R, So K, Kameswaran V, Xu J, Polacco BJ, Desai RV, Chatterjee N, Hsu A, Muncie JM, Blotnick AM, Winchester SAB, Weinberger LS, Hüttenhain R, Kathirya IS, Krogan NJ, et al. 2022. Brahma safeguards canalization of cardiac mesoderm differentiation. *Nature* **602**:129–134. DOI: <https://doi.org/10.1038/s41586-021-04336-y>, PMID: 35082446
- Huang S**, Guo Y-P, May G, Enver T. 2007. Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Developmental Biology* **305**:695–713. DOI: <https://doi.org/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: <https://doi.org/10.1002/bies.201100031>, PMID: 22102361
- Jung J**, Ma S, Tay T, Geiger-Schuller KR, Kirchgatterer PC, Verdine VK, Guo B, Arias-Garcia MA, Allen WE, Singh A, Kuksenko O, Abudayyeh OO, Gootenberg JS, Fu Z, Macrae RK, Buenrostro JD, Regev A, Zhang F. 2023. A transcription factor atlas of directed differentiation. *Cell* **186**:209–229. DOI: <https://doi.org/10.1016/j.cell.2022.11.026>, PMID: 36608654
- Kato H**, Igarashi K. 2019. To be red or white: lineage commitment and maintenance of the hematopoietic system by the “inner myeloid.” *Haematologica* **104**:1919–1927. DOI: <https://doi.org/10.3324/haematol.2019.216861>, PMID: 31515352
- Kirouac DC**, Madlambayan GJ, Yu M, Sykes EA, Ito C, Zandstra PW. 2009. Cell-cell interaction networks regulate blood stem and progenitor cell fate. *Molecular Systems Biology* **5**:293. DOI: <https://doi.org/10.1038/msb.2009.49>, PMID: 19638974
- Kittisopikul M**, Süel GM. 2010. Biological role of noise encoded in a genetic network motif. *PNAS* **107**:13300–13305. DOI: <https://doi.org/10.1073/pnas.1003975107>
- Kovary KM**, Taylor B, Zhao ML, Teruel MN. 2018. Expression variation and covariation impair analog and enable binary signaling control. *Molecular Systems Biology* **14**:e7997. DOI: <https://doi.org/10.15252/msb.20177997>, PMID: 29759982
- Kovtonyuk LV**, Fritsch K, Feng X, Manz MG, Takizawa H. 2016. Inflamm-aging of hematopoiesis, hematopoietic stem cells, and the bone marrow microenvironment. *Frontiers in Immunology* **7**:502. DOI: <https://doi.org/10.3389/fimmu.2016.00502>, PMID: 27895645
- Krumsiek J**, Marr C, Schroeder T, Theis FJ. 2011. Hierarchical differentiation of myeloid progenitors is encoded in the transcription factor network. *PLOS ONE* **6**:e22649. DOI: <https://doi.org/10.1371/journal.pone.0022649>, PMID: 21853041
- Kumar L**, Futschik ME. 2007. Mfuzz: A software package for soft clustering of microarray data. *Bioinformatics* **23**:5–7. DOI: <https://doi.org/10.1093/bioinformatics/btl582>
- Lambert SA**, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, Chen X, Taipale J, Hughes TR, Weirauch MT. 2018. The human transcription factors. *Cell* **172**:650–665. DOI: <https://doi.org/10.1016/j.cell.2018.01.029>, PMID: 29425488
- Li Y**, Jiang Y, Paxman J, O'Laughlin R, Klepin S, Zhu Y, Pillus L, Tsimring LS, Hasty J, Hao N. 2020. A programmable fate decision landscape underlies single-cell aging in yeast. *Science* **369**:325–329. DOI: <https://doi.org/10.1126/science.aax9552>, PMID: 32675375
- Li C-J**, Liao ES, Lee Y-H, Huang Y-Z, Liu Z, Willems A, Garside V, McGlenn E, Chen J-A, Hong T. 2021. MicroRNA governs bistable cell differentiation and lineage segregation via a noncanonical feedback. *Molecular Systems Biology* **17**:e9945. DOI: <https://doi.org/10.15252/msb.20209945>, PMID: 33890404
- Liang J**, Wang N, He J, Du J, Guo Y, Li L, Wu W, Yao C, Li Z, Kee K. 2019. Induction of Sertoli-like cells from human fibroblasts by NR5A1 and GATA4. *eLife* **8**:e48767. DOI: <https://doi.org/10.7554/eLife.48767>, PMID: 31710289
- Liu R**, Li M, Liu Z-P, Wu J, Chen L, Aihara K. 2012. Identifying critical transitions and their leading biomolecular networks in complex diseases. *Scientific Reports* **2**:813. DOI: <https://doi.org/10.1038/srep00813>, PMID: 23230504
- Loeffler D**, Schroeder T. 2019. Understanding cell fate control by continuous single-cell quantification. *Blood* **133**:1406–1414. DOI: <https://doi.org/10.1182/blood-2018-09-835397>, PMID: 30728141
- Long H**, Jia Q, Wang L, Fang W, Wang Z, Jiang T, Zhou F, Jin Z, Huang J, Zhou L, Hu C, Wang X, Zhang J, Ba Y, Gong Y, Zeng X, Zeng D, Su X, Alexander PB, Wang L, et al. 2022. Tumor-induced erythroid precursor-differentiated myeloid cells mediate immunosuppression and curtail anti-PD-1/PD-L1 treatment efficacy. *Cancer Cell* **40**:674–693. DOI: <https://doi.org/10.1016/j.ccell.2022.04.018>, PMID: 35594863
- López-Otín C**, Blasco MA, Partridge L, Serrano M, Kroemer G. 2013. The Hallmarks of Aging. *Cell* **153**:1194–1217. DOI: <https://doi.org/10.1016/j.cell.2013.05.039>
- Lord ND**, Norman TM, Yuan R, Bakshi S, Losick R, Paulsson J. 2019. Stochastic antagonism between two proteins governs a bacterial cell fate switch. *Science* **366**:116–120. DOI: <https://doi.org/10.1126/science.aaw4506>, PMID: 31604312
- Macarthur BD**, Ma'ayan A, Lemischka IR. 2009. Systems biology of stem cell fate and cellular reprogramming. *Nature Reviews. Molecular Cell Biology* **10**:672–681. DOI: <https://doi.org/10.1038/nrm2766>, PMID: 19738627
- MacArthur BD**. 2022. The geometry of cell fate. *Cell Systems* **13**:1–3. DOI: <https://doi.org/10.1016/j.cels.2021.12.001>, PMID: 35051370

- MacArthur BD.** 2023. Stem cell biology needs a theory. *Stem Cell Reports* **18**:3–5. DOI: <https://doi.org/10.1016/j.stemcr.2022.11.005>, PMID: 36630903
- Macnair W,** De Vargas Roditi L, Ganscha S, Claassen M. 2019. Tree-ensemble analysis assesses presence of multifurcations in single cell data. *Molecular Systems Biology* **15**:e8552. DOI: <https://doi.org/10.15252/msb.20188552>, PMID: 30918107
- Matsuura S,** Ono H, Kawasaki S, Kuang Y, Fujita Y, Saito H. 2018. Synthetic RNA-based logic computation in mammalian cells. *Nature Communications* **9**:4847. DOI: <https://doi.org/10.1038/s41467-018-07181-2>, PMID: 30451868
- Mayo AE,** Setty Y, Shavit S, Zaslaver A, Alon U. 2006. Plasticity of the cis-regulatory input function of a gene. *PLOS Biology* **4**:e45. DOI: <https://doi.org/10.1371/journal.pbio.0040045>, PMID: 16602820
- Mazid MA,** Ward C, Luo Z, Liu C, Li Y, Lai Y, Wu L, Li J, Jia W, Jiang Y, Liu H, Fu L, Yang Y, Ibañez DP, Lai J, Wei X, An J, Guo P, Yuan Y, Deng Q, et al. 2022. Rolling back human pluripotent stem cells to an eight-cell embryo-like stage. *Nature* **605**:315–324. DOI: <https://doi.org/10.1038/s41586-022-04625-0>, PMID: 35314832
- Moignard V,** Woodhouse S, Haghverdi L, Lilly AJ, Tanaka Y, Wilkinson AC, Buettner F, Macaulay IC, Jawaid W, Diamanti E, Nishikawa S-I, Piterman N, Kouskoff V, Theis FJ, Fisher J, Göttgens B. 2015. Decoding the regulatory network of early blood development from single-cell gene expression measurements. *Nature Biotechnology* **33**:269–276. DOI: <https://doi.org/10.1038/nbt.3154>, PMID: 25664528
- Mojtahedi M,** Skupin A, Zhou J, Castaño IG, Leong-Quong RYY, Chang H, Trachana K, Giuliani A, Huang S. 2016. Cell fate decision as high-dimensional critical state transition. *PLOS Biology* **14**:e2000640. DOI: <https://doi.org/10.1371/journal.pbio.2000640>, PMID: 28027308
- Nam KM,** Martinez-Corral R, Gunawardena J. 2022. The linear framework: using graph theory to reveal the algebra and thermodynamics of biomolecular systems. *Interface Focus* **12**:20220013. DOI: <https://doi.org/10.1098/rsfs.2022.0013>, PMID: 35860006
- Ng AHM,** Khoshakhlagh P, Rojo Arias JE, Pasquini G, Wang K, Swiersy A, Shipman SL, Appleton E, Kiaee K, Kohman RE, Vernet A, Dysart M, Leeper K, Saylor W, Huang JY, Graveline A, Taipale J, Hill DE, Vidal M, Melero-Martin JM, et al. 2021. A comprehensive library of human transcription factors for cell fate engineering. *Nature Biotechnology* **39**:510–519. DOI: <https://doi.org/10.1038/s41587-020-0742-6>, PMID: 33257861
- Ocone A,** Haghverdi L, Mueller NS, Theis FJ. 2015. Reconstructing gene regulatory dynamics from high-dimensional single-cell snapshot data. *Bioinformatics* **31**:i89–i96. DOI: <https://doi.org/10.1093/bioinformatics/btv257>, PMID: 26072513
- Olariu V,** Peterson C. 2019. Kinetic models of hematopoietic differentiation. *Wiley Interdisciplinary Reviews. Systems Biology and Medicine* **11**:e1424. DOI: <https://doi.org/10.1002/wsbm.1424>, PMID: 29660842
- Oliver Metzger M,** Tang Y, Mitchell S, Taylor B, Foreman R, Wollman R, Hoffmann A. 2020. An incoherent feedforward loop interprets NFκB/RelA dynamics to determine TNF-induced necroptosis decisions. *Molecular Systems Biology* **16**:e9677. DOI: <https://doi.org/10.15252/msb.20209677>, PMID: 33314666
- Olsson A,** Venkatasubramanian M, Chaudhri VK, Aronow BJ, Salomonis N, Singh H, Grimes HL. 2016. Single-cell analysis of mixed-lineage states leading to a binary cell fate choice. *Nature* **537**:698–702. DOI: <https://doi.org/10.1038/nature19348>, PMID: 27580035
- Orkin SH,** Zon LI. 2008. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* **132**:631–644. DOI: <https://doi.org/10.1016/j.cell.2008.01.025>, PMID: 18295580
- Palii CG,** Cheng Q, Gillespie MA, Shannon P, Mazurczyk M, Napolitani G, Price ND, Ranish JA, Morrissey E, Higgs DR, Brand M. 2019. Single-cell proteomics reveal that quantitative changes in co-expressed lineage-specific transcription factors determine cell fate. *Cell Stem Cell* **24**:812–820. DOI: <https://doi.org/10.1016/j.stem.2019.02.006>, PMID: 30880026
- Pang WW,** Schrier SL, Weissman IL. 2017. Age-associated changes in human hematopoietic stem cells. *Seminars in Hematology* **54**:39–42. DOI: <https://doi.org/10.1053/j.seminhematol.2016.10.004>, PMID: 28088986
- Pei W,** Shang F, Wang X, Fanti A-K, Greco A, Busch K, Klapproth K, Zhang Q, Quedenau C, Sauer S, Feyerabend TB, Höfer T, Rodewald H-R. 2020. Resolving fates and single-cell transcriptomes of hematopoietic stem cell clones by polyloxexpress barcoding. *Cell Stem Cell* **27**:383–395. DOI: <https://doi.org/10.1016/j.stem.2020.07.018>, PMID: 32783885
- Psaila B,** Wang G, Rodriguez-Meira A, Li R, Heuston EF, Murphy L, Yee D, Hitchcock IS, Sousos N, O'Sullivan J, Anderson S, Senis YA, Weinberg OK, Calicchio ML, NIH Intramural Sequencing Center, Iskander D, Royston D, Milojkovic D, Roberts I, Bodine DM, et al. 2020. Single-cell analyses reveal megakaryocyte-biased hematopoiesis in myelofibrosis and identify mutant clone-specific targets. *Molecular Cell* **78**:477–492. DOI: <https://doi.org/10.1016/j.molcel.2020.04.008>, PMID: 32386542
- Qin J,** Zhang J, Jiang J, Zhang B, Li J, Lin X, Wang S, Zhu M, Fan Z, Lv Y, He L, Chen L, Yue W, Li Y, Pei X. 2022. Direct chemical reprogramming of human cord blood erythroblasts to induced megakaryocytes that produce platelets. *Cell Stem Cell* **29**:1229–1245. DOI: <https://doi.org/10.1016/j.stem.2022.07.004>
- Qiu X,** Zhang Y, Martin-Rufino JD, Weng C, Hosseinzadeh S, Yang D, Pogson AN, Hein MY, Hoi Joseph Min K, Wang L, Grody EI, Shurtleff MJ, Yuan R, Xu S, Ma Y, Replogle JM, Lander ES, Darmanis S, Bahar I, Sankaran VG, et al. 2022. Mapping transcriptomic vector fields of single cells. *Cell* **185**:690–711. DOI: <https://doi.org/10.1016/j.cell.2021.12.045>, PMID: 35108499
- Reiter F,** Wienerroither S, Stark A. 2017. Combinatorial function of transcription factors and cofactors. *Current Opinion in Genetics & Development* **43**:73–81. DOI: <https://doi.org/10.1016/j.gde.2016.12.007>
- Replogle JM,** Saunders RA, Pogson AN, Hussmann JA, Lenail A, Guna A, Mascibroda L, Wagner EJ, Adelman K, Lithwick-Yanai G, Iremadze N, Oberstrass F, Lipson D, Bonnar JL, Jost M, Norman TM, Weissman JS. 2022.

- Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell* **185**:2559–2575. DOI: <https://doi.org/10.1016/j.cell.2022.05.013>, PMID: 35688146
- Reya T, Morrison SJ, Clarke MF, Weissman IL. 2001. Stem cells, cancer, and cancer stem cells. *Nature* **414**:105–111. DOI: <https://doi.org/10.1038/35102167>, PMID: 11689955
- Rosales-Alvarez RE, Rettkowski J, Herman JS, Dumbović G, Cabezas-Wallscheid N, Grün D. 2022. Gene expression noise dynamics unveil functional heterogeneity of ageing hematopoietic stem cells. *SSRN*. DOI: <https://doi.org/10.2139/ssrn.4197395>
- Sáez M, Blassberg R, Camacho-Aguilar E, Siggia ED, Rand DA, Briscoe J. 2022. Statistically derived geometrical landscapes capture principles of decision-making dynamics during cell fate transitions. *Cell Systems* **13**:12–28. DOI: <https://doi.org/10.1016/j.cels.2021.08.013>, PMID: 34536382
- Santillán M. 2008. On the use of the hill functions in mathematical models of gene regulatory networks. *Mathematical Modelling of Natural Phenomena* **3**:85–97. DOI: <https://doi.org/10.1051/mmnp:2008056>
- Scheffer M, Carpenter SR, Lenton TM, Bascompte J, Brock W, Dakos V, van de Koppel J, van de Leemput IA, Levin SA, van Nes EH, Pascual M, Vandermeer J. 2012. Anticipating critical transitions. *Science* **338**:344–348. DOI: <https://doi.org/10.1126/science.1225244>, PMID: 23087241
- Schiebinger G, Shu J, Tabaka M, Cleary B, Subramanian V, Solomon A, Gould J, Liu S, Lin S, Berube P, Lee L, Chen J, Brumbaugh J, Rigollet P, Hochedlinger K, Jaenisch R, Regev A, Lander ES. 2019. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell* **176**:928–943. DOI: <https://doi.org/10.1016/j.cell.2019.01.006>, PMID: 30712874
- Sciammas R, Li Y, Warmflash A, Song Y, Dinner AR, Singh H. 2011. An incoherent regulatory network architecture that orchestrates B cell diversification in response to antigen signaling. *Molecular Systems Biology* **7**:495. DOI: <https://doi.org/10.1038/msb.2011.25>, PMID: 21613984
- Semrau S, Goldmann JE, Soumillon M, Mikkelsen TS, Jaenisch R, van Oudenaarden A. 2017. Dynamics of lineage commitment revealed by single-cell transcriptomics of differentiating embryonic stem cells. *Nature Communications* **8**:1096. DOI: <https://doi.org/10.1038/s41467-017-01076-4>, PMID: 29061959
- Shakiba N, Li C, Garcia-Ojalvo J, Cho K-H, Patil K, Walczak A, Liu Y-Y, Kuehn S, Nie Q, Klein A, Deco G, Kringelbach M, Iyer-Biswas S. 2022. How can Waddington-like landscapes facilitate insights beyond developmental biology? *Cell Systems* **13**:4–9. DOI: <https://doi.org/10.1016/j.cels.2021.12.003>, PMID: 35051373
- Shea MA, Ackers GK. 1985. The OR control system of bacteriophage lambda. A physical-chemical model for gene regulation. *Journal of Molecular Biology* **181**:211–230. DOI: [https://doi.org/10.1016/0022-2836\(85\)90086-5](https://doi.org/10.1016/0022-2836(85)90086-5), PMID: 3157005
- Shi Y, Inoue H, Wu JC, Yamanaka S. 2017. Induced pluripotent stem cell technology: a decade of progress. *Nature Reviews. Drug Discovery* **16**:115–130. DOI: <https://doi.org/10.1038/nrd.2016.245>, PMID: 27980341
- Shi J, Aihara K, Li T, Chen L. 2022. Energy landscape decomposition for cell differentiation with proliferation effect. *National Science Review* **9**:wac116. DOI: <https://doi.org/10.1093/nsr/nwac116>, PMID: 35992240
- Shu J, Wu C, Wu Y, Li Z, Shao S, Zhao W, Tang X, Yang H, Shen L, Zuo X, Yang W, Shi Y, Chi X, Zhang H, Gao G, Shu Y, Yuan K, He W, Tang C, Zhao Y, et al. 2013. Induction of pluripotency in mouse somatic cells with lineage specifiers. *Cell* **153**:963–975. DOI: <https://doi.org/10.1016/j.cell.2013.05.001>
- Simon CS, Hadjantonakis AK, Schröter C. 2018. Making lineage decisions with biological noise: Lessons from the early mouse embryo. *Wiley Interdisciplinary Reviews. Developmental Biology* **7**:e319. DOI: <https://doi.org/10.1002/wdev.319>, PMID: 29709110
- Simunovic M, Siggia ED, Brivanlou AH. 2022. In vitro attachment and symmetry breaking of a human embryo model assembled from primed embryonic stem cells. *Cell Stem Cell* **29**:962–972. DOI: <https://doi.org/10.1016/j.stem.2022.05.001>, PMID: 35659878
- Spitz F, Furlong EEM. 2012. Transcription factors: from enhancer binding to developmental control. *Nature Reviews. Genetics* **13**:613–626. DOI: <https://doi.org/10.1038/nrg3207>, PMID: 22868264
- Stadler T, Pybus OG, Stumpf MPH. 2021. Phylodynamics for cell biologists. *Science* **371**:eaah6266. DOI: <https://doi.org/10.1126/science.aah6266>, PMID: 33446527
- Stanoev A, Koseska A. 2022. Robust cell identity specifications through transitions in the collective state of growing developmental systems. *Current Opinion in Systems Biology* **31**:100437. DOI: <https://doi.org/10.1016/j.coisb.2022.100437>
- Tang Y, Aryal S, Geng X, Zhou X, Fast VG, Zhang J, Lu R, Zhou Y. 2022. TBX20 improves contractility and mitochondrial function during direct human cardiac reprogramming. *Circulation* **146**:1518–1536. DOI: <https://doi.org/10.1161/CIRCULATIONAHA.122.059713>
- Trojanowski J, Rippe K. 2022. Transcription factor binding and activity on chromatin. *Current Opinion in Systems Biology* **31**:100438. DOI: <https://doi.org/10.1016/j.coisb.2022.100438>
- Uthamacumaran A. 2021. A review of dynamical systems approaches for the detection of chaotic attractors in cancer networks. *Patterns* **2**:100226. DOI: <https://doi.org/10.1016/j.patter.2021.100226>, PMID: 33982021
- Vanden-Eijnden E, Heymann M. 2008. The geometric minimum action method for computing minimum energy paths. *The Journal of Chemical Physics* **128**:061103. DOI: <https://doi.org/10.1063/1.2833040>, PMID: 18282021
- van de Rijn M, Heimfeld S, Spangrude GJ, Weissman IL. 1989. Mouse hematopoietic stem-cell antigen Sca-1 is a member of the Ly-6 antigen family. *PNAS* **86**:4634–4638. DOI: <https://doi.org/10.1073/pnas.86.12.4634>
- Waddington CH. 1957. *The Strategy of the Genes: A Discussion of Some Aspects of Theoretical Biology*. London: Allen & Unwin.
- Wang P, Chen L. 2020. Critical transitions and tipping points in EMT. *Quantitative Biology* **8**:195–202. DOI: <https://doi.org/10.1007/s40484-020-0219-0>

- Wheat JC**, Sella Y, Willcockson M, Skoultchi AI, Bergman A, Singer RH, Steidl U. 2020. Single-molecule imaging of transcription dynamics in somatic stem cells. *Nature* **583**:431–436. DOI: <https://doi.org/10.1038/s41586-020-2432-4>, PMID: 32581360
- Wu X**, Sun Z, Jiang R. 2008. Logic motif of combinatorial control in transcriptional networks. *Nature Precedings*:2227. DOI: <https://doi.org/10.1038/npre.2008.2227.1>
- Xu J**, Du Y, Deng H. 2015. Direct lineage reprogramming: strategies, mechanisms, and applications. *Cell Stem Cell* **16**:119–134. DOI: <https://doi.org/10.1016/j.stem.2015.01.013>, PMID: 25658369
- Xu R**, Li S, Wu Q, Li C, Jiang M, Guo L, Chen M, Yang L, Dong X, Wang H, Wang C, Liu X, Ou X, Gao S. 2022. Stage-specific H3K9me3 occupancy ensures retrotransposon silencing in human pre-implantation embryos. *Cell Stem Cell* **29**:1051–1066. DOI: <https://doi.org/10.1016/j.stem.2022.06.001>, PMID: 35803226
- Yang D**, Jones MG, Naranjo S, Rideout WM, Min KHJ, Ho R, Wu W, Replogle JM, Page JL, Quinn JJ, Horns F, Qiu X, Chen MZ, Freed-Pastor WA, McGinnis CS, Patterson DM, Gartner ZJ, Chow ED, Bivona TG, Chan MM, et al. 2022. Lineage tracing reveals the phylodynamics, plasticity, and paths of tumor evolution. *Cell* **185**:1905–1923. DOI: <https://doi.org/10.1016/j.cell.2022.04.015>, PMID: 35523183
- Yin J**, Zhang L, Zhang P. 2019. High-index optimization-based shrinking dimer method for finding high-index saddle points. *SIAM Journal on Scientific Computing* **41**:A3576–A3595. DOI: <https://doi.org/10.1137/19M1253356>
- Yin J**, Wang Y, Chen JZY, Zhang P, Zhang L. 2020. Construction of a pathway map on a complicated energy landscape. *Physical Review Letters* **124**:090601. DOI: <https://doi.org/10.1103/PhysRevLett.124.090601>, PMID: 32202879
- Yin J**, Yu B, Zhang L. 2021. Searching the solution landscape by generalized high-index saddle dynamics. *Science China Mathematics* **64**:1801–1816. DOI: <https://doi.org/10.1007/s11425-020-1737-1>
- Yin J**, Zhang L, Zhang P. 2022. Solution landscape of the Onsager model identifies non-axisymmetric critical points. *Physica D* **430**:133081. DOI: <https://doi.org/10.1016/j.physd.2021.133081>
- Yu L**, Wei Y, Duan J, Schmitz DA, Sakurai M, Wang L, Wang K, Zhao S, Hon GC, Wu J. 2021. Blastocyst-like structures generated from human pluripotent stem cells. *Nature* **591**:620–626. DOI: <https://doi.org/10.1038/s41586-021-03356-y>, PMID: 33731924
- Yu H**, Chen M, Hu Y, Ou S, Yu X, Liang S, Li N, Yang M, Kong X, Sun C, Jia S, Zhang Q, Liu L, Hurst LD, Li R, Wang W, Wang J. 2022. Dynamic reprogramming of H3K9me3 at hominoid-specific retrotransposons during human preimplantation development. *Cell Stem Cell* **29**:1031–1050. DOI: <https://doi.org/10.1016/j.stem.2022.06.006>, PMID: 35803225
- Zaret KS**. 2020. Pioneer transcription factors initiating gene network changes. *Annual Review of Genetics* **54**:367–385. DOI: <https://doi.org/10.1146/annurev-genet-030220-015007>, PMID: 32886547
- Zernicka-Goetz M**, Huang S. 2010. Stochasticity versus determinism in development: a false dichotomy? *Nature Reviews. Genetics* **11**:743–744. DOI: <https://doi.org/10.1038/nrg2886>, PMID: 20877326
- Zhang Y**, Gao S, Xia J, Liu F. 2018. Hematopoietic hierarchy - an updated roadmap. *Trends in Cell Biology* **28**:976–986. DOI: <https://doi.org/10.1016/j.tcb.2018.06.001>, PMID: 29935893
- Zhao Y**. 2019. Chemically induced cell fate reprogramming and the acquisition of plasticity in somatic cells. *Current Opinion in Chemical Biology* **51**:146–153. DOI: <https://doi.org/10.1016/j.cbpa.2019.04.025>, PMID: 31153758
- Zhou JX**, Huang S. 2011. Understanding gene circuits at cell-fate branch points for rational cell reprogramming. *Trends in Genetics* **27**:55–62. DOI: <https://doi.org/10.1016/j.tig.2010.11.002>, PMID: 21642993
- Zhou Y**, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. 2019. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications* **10**:1523. DOI: <https://doi.org/10.1038/s41467-019-09234-6>
- Zhu R**, Del Rio-Salgado JM, Garcia-Ojalvo J, Elowitz MB. 2022. Synthetic multistability in mammalian cells. *Science* **375**:eabg9765. DOI: <https://doi.org/10.1126/science.abg9765>, PMID: 35050677