

# **Theoretical and Computational Studies on the Dynamics and Regulation of Cell Phenotypic Transitions**

by

Hang Zhang

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Genetics, Bioinformatics, and Computational Biology

Jianhua Xing, Chair

John J. Tyson, Co-Chair

Mark A. Stremler, Co-Chair

Hehuang Xie

February 17, 2016

Blacksburg, VA

Keywords: Mathematical Modeling, Epigenetics, Cell Differentiation,  
Mono-allelic expression, Epithelial-to-Mesenchymal-Transition

Copyright 2016, Hang Zhang

# Theoretical and Computational Studies on the Dynamics and Regulation of Cell Phenotypic Transitions

Hang Zhang

## Abstract

Cell phenotypic transitions, or cell fate decision making processes, are regulated by complex regulatory networks composed of genes, RNAs, proteins and metabolites. The regulation can take place at the epigenetic, transcriptional, translational, and post-translational levels to name a few.

Epigenetic histone modification plays an important role in cell phenotype maintenance and transitions. However, the underlying mechanism relating dynamical histone modifications to stable epigenetic cell memory remains elusive. Incorporating key pieces of molecular level experimental information, we built a statistical mechanics model for the inheritance of epigenetic histone modifications. The model reveals that enzyme selectivity of different histone substrates and cooperativity between neighboring nucleosomes are essential to generate bistability of the epigenetic memory. We then applied the epigenetic modeling framework to the differentiation process of olfactory sensory neurons (OSNs), where the observed ‘one-neuron-one-allele’ phenomenon has remained as a long-standing puzzle. Our model successfully explains this singular behavior in terms of epigenetic competition and enhancer cooperativity during the differentiation process. Epigenetic level events and transcriptional level events cooperate synergistically in the OSN differentiation process. The model also makes a list of testable experimental predictions. In general, the epigenetic modeling framework can be used to study phenotypic transitions when histone modification is a major regulatory element in the system.

Post-transcriptional level regulation plays important roles in cell phenotype maintenance. Our integrated experimental and computational studies revealed such a motif regulating the differentiation of definitive endoderm. We identified two RNA binding proteins, hnRNPA1 and KSRP, which repress each other through microRNAs miR-375 and miR-135a. The motif can generate switch behavior and serve as a

noise filter in the stem cell differentiation process. Manipulating the motif could enhance the differentiation efficiency toward a specific lineage one desires.

Last we performed mathematical modeling on an epithelial-to-mesenchymal transition (EMT) process, which could be used by tumor cells for their migration. Our model predicts that the IL-6 induced EMT is a stepwise process with multiple intermediate states.

In summary, our theoretical and computational analyses about cell phenotypic transitions provide novel insights on the underlying mechanism of cell fate decision. The modeling studies revealed general physical principles underlying complex regulatory networks.

# **Theoretical and Computational Studies on the Dynamics and Regulation of Cell Phenotypic Transitions**

Hang Zhang

## **General Audience Abstract**

The human body is composed of more than 200 different types of cells, characteristic of muscles, skin, brain, etc. Although different types of cells exhibit many differences in morphology and other properties, they all originate from a single fertilized egg. When a cell divides, the new cell has the potential either to maintain its own cell type or become another type of cell. For example, an unspecialized stem cell can produce another stem cell or develop into other types of cell with specialized functions. The transition between different cell types is called cell phenotypic transition. Cell phenotypic transitions are commonly observed in living organisms. For example, olfactory stem cells can transform to olfactory sensory neuron (OSN) cells. As a result, each olfactory sensory neuron (OSN) is specialized to express one and only one type of olfactory receptor (OR) protein on the cell surface, and by this mechanism, the human nose can sense trillions of different odors. As a second example, epithelial cells can convert to mesenchymal cells in human development, cancer development and wound healing. Epithelial cells are tightly connected to each other and to the basement, thus hardly have the ability to move. On the other hand, mesenchymal cells are less connected and more easily to migrate. Epithelial cancers are the most prevalent type of cancer. During cancer metastasis, epithelial cells transform into mesenchymal cells and gain the ability to move to a distant positions. Epithelial to mesenchymal transition (EMT) is not only utilized by cancer cells; normal cell will also utilize this process to move. In human development and wound healing, epithelial cells will transform to mesenchymal cells to migrate to the desired locations. Despite intensive studies, the underlying regulatory mechanisms of cell phenotypic transitions remain largely elusive.

In this dissertation, I present a theoretical framework to elucidate some underlying mechanisms of cell-type maintenance and transitions. Based on experimentally observed physical interactions, we built a

statistical mechanics model which successfully explained the mechanism of cell type maintenance. As an application, I used this model to study the differentiation process of OSNs. Each OSN cell randomly expresses one and only one OR gene from thousands of candidates with equal probability, however, the mechanism remains one of the biggest puzzles in neuroscience. Our model not only recapitulated singular OR expression, but also elucidated how the olfactory system maximizes and maintains the diversity of OR expression.

In addition to the differentiation of OSNs, I also studied the differentiation of embryonic stem cell in human development. Previous studies mainly focused on protein signaling networks regulating this process, our studies revealed a novel regulatory mechanism with microRNA (an important type of regulatory RNA) playing a crucial role. Manipulating the network could enhance the efficiency of differentiation toward a specific lineage that one might desire. Lastly, I examined a mathematical model of the EMT process. The model revealed potential relationships between inflammation and tumor metastasis regarding molecule IL-6 induced EMT. Here, our model predicted that IL-6 induced EMT is a stepwise process with multiple intermediate states and helped us to fully characterize the tumor progression process.

Overall, the broad goal of this research is to advance our understanding of the underlying mechanisms of cell phenotypic transitions. It also sheds light on the processes of tissue regeneration and cancer development.

## Acknowledgements

I would like to give my sincere thanks to all the people who have helped me during my doctoral study.

I would like to thank my advisor Dr. Jianhua Xing, who has always been patient with me, providing tremendous mentoring, continuous support, and exemplary guidance throughout my doctoral training, who encouraged me to explore new fields, and looked into problems both beyond the immediate obstacle, and at the big picture. His endless dedication and love of science have been a great example and inspiration for me to think critically and work hard. To me, he is not only a great advisor, but also a role model who has been fostered my growth as a scientist in the past years.

I would like to thank my co-advisor Dr. John Tyson. The lesson I learned from his systems biology class gave me a first taste of systems biology. The TA experience in his ISC classes further made me understand the history of systems biology and integrate different courses I learned before. His work ethic in science and great responsibilities to students will always remind me to work in discipline and diligent. I would like to thank my co-advisor Dr. Mark Stremler, who is always helpful and willing to discuss with me about my research problems, and who warm-hearted teaches me dynamics, and would like to share his own growth experience with me and encourage me like a friend and advisor. I would like to thank Dr. Hehuang Xie, who is always there to help me whenever I go to ask any questions. I greatly appreciate his time, encouragement and invaluable career advice.

I would like to thank my friends and colleagues in Dr. Xing's and Dr. Tyson's lab. I would like to give specially thanks to Dr. Xiaojun Tian and Ms. Jingyu Zhang on the work of epigenetics and EMT. We studied together and worked together with good time, which makes my PhD a happy and wonderful process. I also would like to thank my friend Li Fei, Xiaojin Sun, my colleague students from GBCB program, and Dr. David Bevan and Ms. Dennie Munson, who have given tremendous help during my study.

Finally, I would like to thank my mother Ruiying Zhang, my father Shuxu Zhang, my sister Zhe Zhang and my girlfriend Lu Chen, for their love, encouragement and support during my doctoral training.

## **Attribution**

Chapter 2 of the dissertation is based on a published paper. Zhang, Hang, et al. "Statistical mechanics model for the dynamics of collective epigenetic histone modification." *Physical Review Letters* 112.6 (2014): 068101. Dr. Jianhua Xing conceived the study, Hang Zhang, Xiao-Jun Tian, Abhishek Mukhopadhyay, and Jianhua Xing analyzed the data. All authors contributed to the writing.

Chapter 3 of the dissertation is based on a submitted manuscript. Dr. Xing constructed the model and conceived the study. Xiao-Jun Tian performed most simulations. Hang Zhang collected biological background information and participated in the simulations. He also worked on generalizing the model of Dodd et al. Xiao-Jun Tian, Hang Zhang and Dr. Xing analyzed the data. Dr. Xing wrote the draft with input from Xiao-Jun Tian and Hang Zhang.

Chapter 4 of the dissertation is from a submitted manuscript. Dr. Xing, Hang Zhang, and Dr Yue Teng conceived the study. Dr. Yue Teng performed the experimental study, Hang Zhang performed bioinformatics and modeling study. Hang Zhang, Yue Teng and Dr. Xing analyzed the data and contributed to the writing.

# Table of Contents

Abstract .....	ii
General Audience Abstract .....	iv
Acknowledgements.....	vi
Attribution.....	vii
Table of Contents .....	viii
List of Figures.....	xii
List of Tables .....	xiv
Chapter 1. Introduction .....	1
References .....	4
Chapter 2. Statistical Mechanics Model for the Dynamics of Collective Epigenetic Histone Modification	5
2.1. Abstract.....	5
2.2. Introduction .....	5
2.3. Model and Method .....	8
2.4. Parameter estimation .....	13
2.4.1. Nonspecific background free energy of binding of enzymes .....	13
2.4.2. Free energy of binding of enzymes within the nucleation region .....	15
2.4.3. Enzyme lateral interactions: .....	15
2.4.4. Enzyme rate constants: .....	15
2.4.5. Histone exchange: .....	16
2.5. Results .....	17
2.6. Robustness of the proposed physical mechanism over parameters .....	21
2.6.1. Binding affinities and lateral interactions.....	21
2.6.2. Nucleation region length .....	24
2.6.3. Histone exchange rate .....	25
2.6.4. Cell cycle time.....	27
2.7. Ideas for future experimental measurements.....	28
2.8. Relation between the statistical mechanics model of Zhang et al. and the rule-based model of Dodd et al. ....	29
2.9. Conclusion and Discussion.....	32

References .....	34
Chapter 3. Three-layer Regulation Leads to Monoallelic and Diverse Olfactory Receptor Expression ....	39
3.1. Abstract.....	39
3.2. Introduction .....	39
3.3. Results .....	43
3.3.1. Mathematical formulation of OR activation.....	43
3.3.2. Low Noise and lack of demethylases kinetically freeze allele epigenetic state before and after differentiation .....	45
3.3.3. Elevation of bifunctional demethylase level leads to a barrier-crossing like dynamics.....	47
3.3.4. Elevation of unifunctional demethylase level leads to ratchet-like dynamics.....	52
3.3.5. Mathematically controlled comparison explains why bifunctional LSD1 improves the ratio of single allele activation. ....	55
3.3.6. Epigenetic competition model predicts zebrafish but not mouse experiments on inhibiting methyltransferases/demethylases .....	57
3.3.7. Competition of cooperatively bound enhancers further reduces coexpression of multi-allele ORs. ....	59
3.3.8. Model studies identify multiple schemes of OR transcription switching.....	63
3.4. Discussion.....	64
3.4.1. A sequential three-layer regulation mechanism controls single allele activation.....	65
3.4.2. The OR selection process is optimized to satisfy hierarchical multi-task and even opposing requirements.....	66
3.4.3. Counter-intuitive bifunctionality of the LSD1 maximizes single allele epigenetic activation. ....	67
3.4.4. Seemingly subtle differences on the regulation schemes lead to qualitatively different mechanisms.....	68
3.4.5. The model makes multiple testable predictions and many have been confirmed. ....	70
3.4.6. Future studies may reveal fine-tuned regulation on OR and other allele-specific activation processes. ....	73
3.5. Method:.....	73
3.5.1. Mathematical analysis of simple models.....	76
3.6. Author contributions.....	78
3.7. Acknowledgements .....	78
References .....	79

Chapter 4. Post-transcriptional motif regulating differentiation of human definitive endoderm.....	86
4.1. Abstract.....	87
4.2. Introduction .....	87
4.3. Results .....	89
4.3.1. Differentially expressed proteins hnRNP A1 and KSRP were identified from inversely altered proteins in DE derived from hESCs.....	89
4.3.2. Bioinformatics analysis predicted potential hnRNPA1 and KSRP target microRNAs.....	92
4.3.3. Mathematical modeling suggests a switch-like motif formed by hnRNPA1-miR375-KSRP-miR135a.....	94
4.3.4. Priming hESCs with AW and hnRNP A1 or miRNA-375 induces rapid and efficient differentiation into the DE .....	97
4.4. Discussion.....	100
4.5. Model and methods .....	103
4.6. Parameter robustness and sensitivity analysis .....	106
4.6.1. Local Sensitivity Analysis:.....	106
4.6.2. Global Sensitivity Analysis: .....	107
4.7. The model could also leads to sigmoidal response under different parameter selections .....	111
References .....	112
Chapter 5. Mathematical modeling of the epithelial-to-mesenchymal transition .....	116
5.1. Abstract.....	116
5.2. Introduction .....	116
5.3. Model and methods .....	120
5.3.1. IL-6R/pSTAT3/miR34a feedback loop:.....	122
5.3.2. Core mutual inhibitory regulatory motifs of SNAIL/miR-34a and ZEB/miR-200 .....	122
5.3.3. IL-6 autocrine secretion module (IL-6/pSTAT3/Lin28/Let-7): .....	122
5.4. Results .....	123
5.4.1. IL-6 induced EMT in a sequential manner.....	123
5.4.2. IL-6-induced EMT is composed of three bistable switches regulated by the corresponding mutual inhibition pairs .....	124
5.4.3. Autocrine production of IL-6 adds another irreversible switch .....	128
5.5. Discussion.....	130
Supplemental Information: .....	131
References: .....	136

Chapter 6. Conclusions .....	139
References .....	141

## List of Figures

Figure 2.1 Illustration of nucleosome structure and histone covalent modification .....	6
Figure 2.2: Illustration of histone epigenetic regulation.. .....	9
Figure 2.3: Simulation results using model parameters corresponding to Oct4 .....	17
Figure 2.4: Typical trajectories of the fraction of nucleosomes with repressive marks (left) and the corresponding probability distribution of observing given number of nucleosomes with repressive marks (right). .....	18
Figure 2.5: Mechanism of bistability .....	21
Figure 2.6: Bistable parameter sets distribution heatmaps of 4096 sets of parameters.....	23
Figure 2.7 Effect of lateral interactions between removal enzymes on generating bimodal nucleosome mark distribution .....	24
Figure 2.8 Average state dwelling time as a function of the nucleation region size.....	25
Figure 2.9 Dependence of average dwelling time on histone exchange rate .....	27
Figure 2.10 Average state dwelling time as a function of the cell cycle.....	28
Figure 2.11 Relaxation dynamics of the total number of nucleosomes with repressive marks .....	29
Figure 3.1: Mathematical model of the experimentally revealed regulatory system of olfactory receptor activation.....	44
Figure 3.2: Low Noise and demethylation enzyme concentration kinetically freeze allele epigenetic state. ....	46
Figure 3.3: Bifunctional LSD1 leads to barrier-crossing-like dynamics and ensures mono-allelic epigenetic activation .....	49
Figure 3.4: Unifunctional LSD1 leads to ratchet-like dynamics and cannot ensure mono-allelic epigenetic activation.....	54
Figure 3.5: Competition of cooperatively bound enhancers further reduces co-expression of multi-allele ORs .....	58

Figure 3.6: Predicted OR expression switching schemes .....	63
Figure 3.7: The three-layer mechanism ensures mono-allele activation of OR genes.....	66
Figure 4.1: Validation of differentially expressed proteins by immunofluorescence and WB .....	91
Figure 4.2 Identification of potential microRNAs regulates and regulated by hnRNPA1 and KSRP during hESCs differentiation to DE process .....	93
Figure 4.3 Post-transcriptional regulatory network regulates the phenotypic transition .....	96
Figure 4.4 High hnRNP A1 and microRNA-375 level drive definitive endoderm differentiation from hESCs .....	99
Figure 4.5 A Post-transcription level toggle switch may act as a noise filter during the differentiation process .....	102
Figure 4.6 Identified post-transcriptional regulatory network .....	103
Figure 4.7 Local parameter sensitivity of the two bifurcation points .....	107
Figure 4.8 Global sensitivity Analysis.....	110
Figure 4.9 Sigmoidal transition of the system. ....	111
Figure 5.1: Signaling pathways involved in EMT. ....	118
Figure 5.2: Schematic illustration of TGF- $\beta$ -induced EMT.....	119
Figure 5.3: Influence wiring diagram of the gene regulatory network of IL-6 induced EMT based on experimental studies.....	121
Figure 5.4: Activation of the EMT switches by IL-6.....	124
Figure 5.5: IL-6 induced EMT are composed of three bistable switches. ....	126
Figure 5.6: The effects of strength of inhibition from pSTAT3 to miR-34a on IL6-induced four state EMT system .....	127
Figure 5.7: Autocrine production of IL-6 makes the last switch irreversible. ....	129

## List of Tables

Table 2.1: Model parameters .....	12
Table 2.2: Summarizes our estimations based on available experimental data. ....	14
Table 2.3 Comparison between the present model and that of Dodd et al.....	31
Table 3.1 Model predictions and corresponding experimental confirmations and suggestions. ....	71
Table 3.2 Values of model parameters used in this work. ....	74
Table 4.1 Parameter description and corresponding control value and uncertainty range.....	108
Table 5.1 Parameter values .....	135

## Chapter 1. Introduction

A living cell is a complicated and delicately regulated biochemical machine. Complicated interaction networks composed of genes, RNAs and proteins known as gene regulatory networks orchestrate all cellular processes in a biological system. Cellular differentiation is a complex cellular process through which generic stem cells become specialized mature cells. The process of cellular differentiation is a direct result of differential gene expression. During differentiation, cells change drastically in their phenotypes. A fertilized egg can develop into over 200 different cell types in the human body, which present drastic variability in the phenotypes while all these cells share the same genetic information. Moreover, under certain circumstances, a differentiated cell may undergo a phenotypic transition in response to an environmental signal.

In 1942, Waddington coined the term “epigenetics” to denote the relationship between genotype and phenotype and events that could not be explained by genetic principles<sup>1,2</sup>. Epigenetics, in a narrow sense, denotes any potential heritable change in gene expression without changes in the underlying DNA sequences, which could be reflected in histone epigenetic memory. Histone epigenetic memory denotes the fact that the same genes in cells of a specific phenotype tend to have roughly the same histone epigenetic pattern. Therefore, histone epigenetic memory plays a key role in maintaining cell phenotype. However, one lasting question in the field is how to correlate the molecular level histone dynamics with the collective ensemble histone epigenetic memory. In this dissertation, we present a mathematical model to bridge the gap between molecular and cellular level experimental observations. By using quasi-equilibrium approximation and implementing the Boltzmann distribution, we constructed a Potts-type model based on experimentally observed lateral interactions of nearest-neighbor enzymes and enzyme recruitment biased by nucleosome covalent modification state. The model leads to effective nonlocal interactions among nucleosomes. Moreover, the simulation demonstrates that epigenetic memory is robustly inheritable against stochastic cellular processes. The model and simulation results are presented in Chapter 2.

Epigenetic histone dynamics greatly affects the differentiation of olfactory sensory neuron (OSNs). OSNs are highly unique and interesting in that they carry hundreds of genes encoding olfactory receptors (ORs) but each cell expresses one and only one gene. A fundamental issue in neuron biology is how the biological system design guarantees mono-allelic activation, which is important for correct function of the olfaction system. One interesting phenomena of the differentiation process is that before differentiation, all OR genes are occupied by repressive histone marks, and after differentiation, one and only one OR gene gains active histone mark. By applying the framework we developed previously, we built a model that captures key biological features of the differentiation process. Our model successfully explains the singular OR expression behavior in terms of epigenetic competition and enhancer cooperativity during the differentiation process and further makes testable predictions. The corresponding analysis and results are described in Chapter 3.

Furthermore, in Chapter 4, we studied the human embryonic stem cell differentiation process and cell reprogramming and transdifferentiation process. Numerous master regulators control human embryonic stem cell differentiation process. In this study, we are interested in definitive endoderm differentiation process. The aim of this work is to reveal a novel regulatory motif acting at the post-transcriptional level and regulating the cell fate decision process. Using high-throughput proteomics studies, we identified two differentially expressed RNA binding proteins, hnRNPA1 and KSRP. Furthermore microRNA array data determines microRNA miR-15 and miR-27 as their targets. Our analysis reveals a classical double negative feedback motif that can generate bistable switch behavior. Further experimental studies confirmed the regulating roles of the motif.

Besides cell phenotypic transition processes in development and cell reprogramming, some cell phenotypic transitions take place during pathological conditions. The epithelial-to-mesenchymal transition (EMT) is one example. EMT occurs in normal developmental and physiological processes, but it also contributes to cancer progression and metastasis. In order to understand the underlying control mechanism of EMT, we built a kinetic rate equation-based mathematical model of TGF- $\beta$  induced EMT. The model demonstrates that EMT is a stepwise activation process of a cascade of two bistable switches regulated by

two double-negative feedback loops: one between the transcription factor SNAIL1 and the miR-34 family and another between the transcription factor ZEB1 and the miR-200 family. The model is further been experimentally verified by us from single cell and population level dynamics of key mRNAs, microRNAs and proteins <sup>3</sup>. Besides TGF-beta, EMT can be induced by many other signals and IL-6 is one example. Tumor progression is modulated by the tumor's microenvironment, which (as is well known) is typically enriched with inflammatory cytokines. In cytokine IL-6 induced EMT process, two additional double negative feedback loops of miR34a/IL-6R and let-7/IL6 are observed. Together with the core mutual inhibition of Snail/miR34a and Zeb1/miR200, they regulate the IL-6 induced EMT process. We construct a mathematical model of these multiple positive feedback loops, and the model suggests that the signaling network proceeds through two intermediate states between the initial epithelial state and the final mesenchymal state. Furthermore, IL-6 and TGF-beta signaling work synergistically during the EMT process. Therefore our model analysis demonstrates that for developing more efficacious cancer therapies, it could be of great clinical importance to fully characterize the dynamic regulation of the composite TGF- $\beta$ /IL6 axis.

## References

1. Goldberg, A.D., Allis, C.D. & Bernstein, E. Epigenetics: a landscape takes shape. *Cell* **128**, 635-8 (2007).
2. Waddington, C.H. The epigenotype. *Endeavour* **1**, 18-20 (1942).
3. Zhang, J. *et al.* TGF- $\beta$ -induced epithelial-to-mesenchymal transition proceeds through stepwise activation of multiple feedback loops. *Science signaling* **7**, ra91-ra91 (2014).

# Chapter 2. Statistical Mechanics Model for the Dynamics of Collective Epigenetic Histone Modification

This Chapter is based on a published paper: Physical review letters 112.6 (2014): 068101.

Hang Zhang<sup>1</sup>, Xiao-Jun Tian<sup>1</sup>, Abhishek Mukhopadhyay<sup>2</sup>, K. S. Kim<sup>3</sup> and Jianhua Xing<sup>1,2,4,\*</sup>

<sup>1</sup>*Department of Biological Science, Virginia Tech, Blacksburg, Virginia 24061-0406, USA;* <sup>2</sup>*Department of Physics, Virginia Tech, Blacksburg, Virginia 24061-0406, USA;* <sup>3</sup>*Lawrence Livermore National Laboratory and University of California, Livermore, California 94550, USA;* <sup>4</sup>*Beijing Computational Science Center, Beijing 100084, China*

## 2.1. Abstract

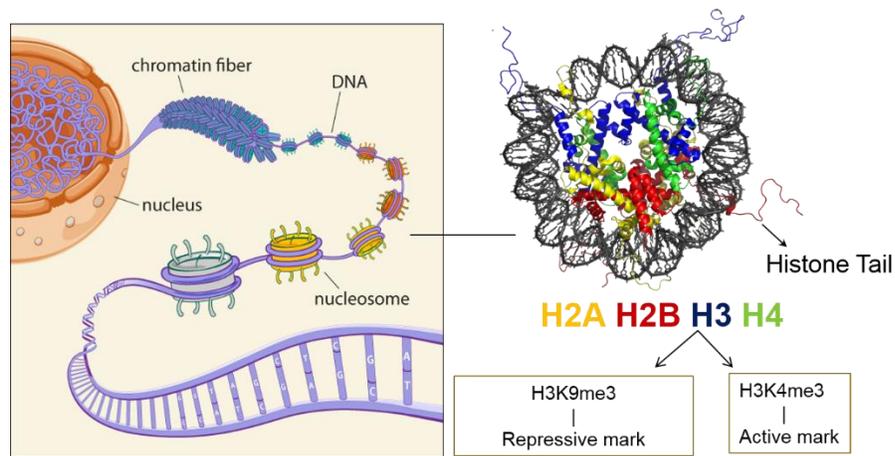
Epigenetic histone modifications play an important role in the maintenance of different cell phenotypes. The exact molecular mechanism for inheritance of the modification patterns over cell generations remains elusive. We construct a Potts-type model based on experimentally observed nearest-neighbor enzyme lateral interactions and nucleosome covalent modification state biased enzyme recruitment. The model can lead to effective nonlocal interactions among nucleosomes suggested in previous theoretical studies, and epigenetic memory is robustly inheritable against stochastic cellular processes.

## 2.2. Introduction

Epigenetics, in a narrow sense, denotes any potential heritable change in gene expression without changes in the underlying DNA sequences; for example, DNA methylation, histone variants incorporation and histone post-translational modifications are different ways of epigenetics modifications. Clearly, epigenetic dynamics affects gene expression patterns. Epigenetics is important due to different layers of regulation it could participate and regulate, also due to its special characteristics and crosstalk with other regulators. Histone modification could cross talk with other layer of regulation mechanisms, e.g.

transcription level (transcription factor), post-transcription level (e.g. microRNAs), DNA methylation level etc. In this chapter, I will mainly focus on the histone modification memory and dynamics.

Nucleosome is the basic unit of chromatin. It is composed of ~200bp DNA wrapping around eight proteins -- which is called histone octamer as shown in Figure 2.1. Each histone octamer consists of two copies of the four core histone proteins, H2A, H2B, H3 and H4. As time goes by, people discovered various covalent marks modified in the histone tail in the nucleosome. Methylation, acetylation, phosphorylation and sumoylation can take place on a number of side residues of the histone proteins<sup>1</sup>. Moreover, researchers found that different histone modification pattern may correspond to different gene expression state, e.g. acetylation and H3K4me3 would always imply an active state while methylation on histone lysine 9 (H3K9me3) or 27 (H3K27me3) would always indicate a repressive state<sup>2</sup>. The so-called ‘histone code’ proposal, analogy to the DNA codon, although still under debate, has drawn extensive attention from the field<sup>3</sup>. Furthermore, in the past few years, several groups discovered multiple enzymes regulating the histone modification dynamics<sup>4,5</sup>.



**Figure 2.1** Illustration of nucleosome structure and histone covalent modification. Chromatin in nucleus is zoomed into chromatin fiber and further zoomed into DNA sequence and nucleosome (left panel). Right panel shows the crystal structure of a nucleosome. Different covalent modification can take place on the histone tail. Two well-studied covalent modifications (H3K9me3 and H3k4me3) which correlate with active or repressive gene activity are listed. H3k9me3 denotes tri-methylation

*on histone H3 lysine 9. H3K4me3 denotes tri-methylation on histone H3 lysine 4. Picture modified from Broad Communications.*

Epigenetics plays a crucial role in development and cell fate decision. In the 2012 Noble prize winning study of cell reprogramming from Yamanaka's group<sup>6</sup>, fibroblast cells crossed the barrier between different cell types and reprogrammed into stem cells. In this reprogramming process, researchers find that cells undergo global epigenetic remodeling<sup>7</sup>. Also researchers suggest that epigenetic regulation is a key step and main barrier of the reprogramming process. Due to the fact that all genes or DNA sequences are subjected to the same set of epigenetic regulating enzymes, epigenetics perturbation (e.g. the enzyme concentration perturbation) could lead to widespread gene expression changes, thus globally regulate the behavior of all the genes and facilitate the fibroblast cell to stem cell phenotypic transition process. Indeed, numerous studies state the importance of these epigenetic regulating enzymes in the phenotypic transition process<sup>8-10</sup>. Moreover, people discovered that many phenomena could be explained by histone modification dynamics, e.g., 'vernalization', the phenomena that flowering in certain plants need to remember the previous winter (cold days) in order to safely give birth to the next generation<sup>11</sup>. The prolonged cold exposure process enables the plants undergo an epigenetic change that ensures the plants could flower in the spring, which comes after winter. In the olfactory sensory neuron cell differentiation, the olfactory epithelium cells need to correctly select only one gene among thousands of candidates, which is known as 'one neuron-one allele' of olfactory receptors. This brought another advantage of epigenetic regulation -- it could affect the gene expression profile in a highly wide-range. Recent findings prove that epigenetics play crucial roles in establishing and maintaining the singularity<sup>12-14</sup>. There are more phenomena recently found that could fall into the epigenetic regulation area, e.g. genetic imprinting, random mono-allelic expression, etc.<sup>15-17</sup>

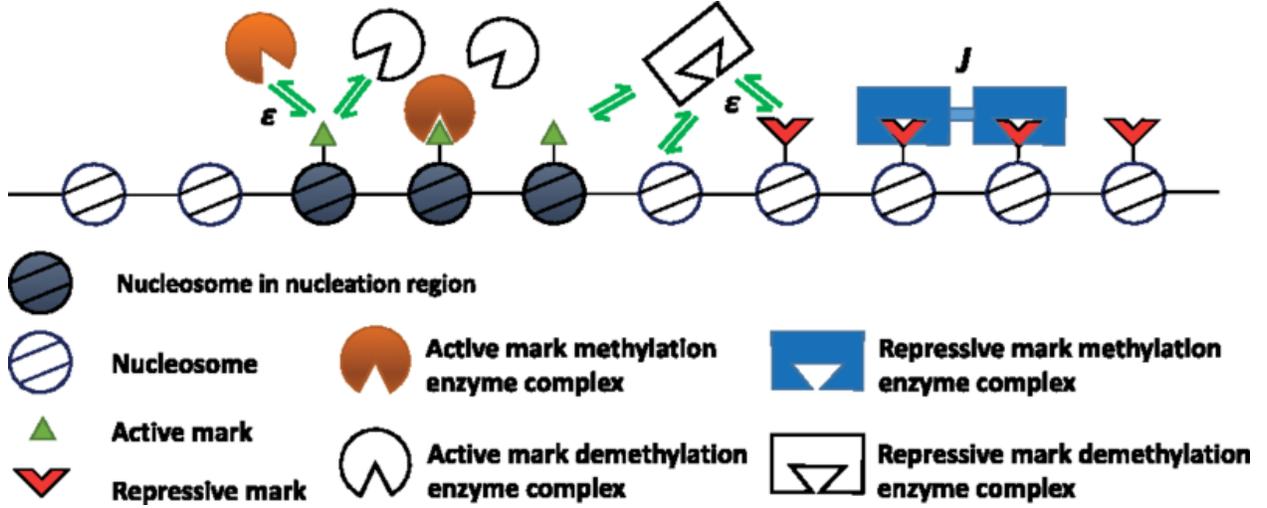
All these experimental studies demonstrated the importance of histone modification patterns on the cell memory and development process. Experiments suggest that at least some of the histone covalent patterns can be transmitted over a number of generations<sup>18</sup>. Although the actual mechanism for this epigenetic

memory is unclear, a simple rule-based model by Dodd et al.<sup>19</sup> shows that robust bistability requires cooperative effects beyond neighboring nucleosomes, which they suggest is due to compact chromatin structures. Subsequent theoretical studies on yeast chromatin silencing<sup>20</sup>, mouse stem cell differentiation<sup>21</sup>, and plant flowering regulation<sup>22</sup> also conclude that this nonlocal cooperativity is necessary for generating stable epigenetic memory.

In recent years molecular details on nucleosome covalent modification dynamics have been extensively studied. Numerous evidences have changed biologists' view of epigenetics memory. Instead of static memory, epigenetics regulation is in highly dynamic fashion<sup>23</sup>. Measurements show that the typical residence time of a modification on chromatin is within sub-seconds to seconds<sup>23</sup>. Experimental observations also suggest that a modified nucleosome may have higher binding affinity for the corresponding enzymes<sup>2,24</sup>. Another interesting observation is that a nucleosome bound modification enzyme complex laterally interacts with another bound to neighboring nucleosomes<sup>24,25</sup>. Although the functional consequences of these interactions on epigenetic dynamics are unclear, recent work suggests that increased enzyme lateral interactions lead to sustained repression or activation of genes, and cancer cells show mutations linked with such lateral interactions<sup>26,27</sup>.

### **2.3. Model and Method**

We construct a theoretical model aiming to bridge the gap between detailed molecular events occurring at the subsecond time scale, and the long-time scale epigenetic change dynamics that is typically in days or longer. To be specific, we focus on lysine 4 (active) and lysine 9 (repressive) methylation on histone H3, although we expect the mechanism discussed here can be general.



*Figure 2.2: Illustration of histone epigenetic regulation.  $\epsilon$  denotes the enzyme binding energy and  $J$  denotes the enzyme lateral interaction energy.*

Consider a gene with  $N$  nucleosomes, as shown in Figure 2.2. Each nucleosome can be in one of the three covalent states: repressively modified ( $s = -1$ ), unmodified (0), and actively modified (1). Here, for simplicity, we only consider one covalent modification site on each histone octamer, and do not distinguish multiple modification (e.g., mono-, di-, and trimethylation) states. Our model is flexible enough to admit straightforward extensions with increasing complexity. Each nucleosome can be in one of the enzyme binding states with corresponding binding energies: empty ( $\sigma = 1$ ,  $\epsilon = \epsilon_{1s} = 0$ ), repressive modification enzyme bound ( $\sigma = 2$ ,  $\epsilon = \epsilon_{2s}$ ), repressive modification removal enzyme bound ( $\sigma = 3$ ,  $\epsilon = \epsilon_{3s}$ ), active modification enzyme bound ( $\sigma = 4$ ,  $\epsilon = \epsilon_{4s}$ ), active modification removal enzyme bound ( $\sigma = 5$ ,  $\epsilon = \epsilon_{5s}$ ). To account for the  $s$  dependence of binding affinity, we assume that the binding energies for the adding or removing enzymes to a nucleosome bearing the corresponding (or antagonizing) mark are  $\Delta\epsilon$  lower (or higher) than those binding to an unmodified nucleosome. Furthermore, if two neighboring [ $i$ -th and  $(i+1)$ -th] nucleosomes are both bound, the binding enzymes interact laterally with energy  $J_{\sigma_i\sigma_{i+1}}$ , otherwise  $J_{\sigma_i=1,\sigma_{i+1}} = J_{\sigma_i,\sigma_{i+1}=1} = 0$ . The above background enzyme-nucleosome binding has no DNA sequence specificity, and the corresponding binding energies estimated from experimental data are weak.

It is suggested that transcription factors and other molecules recruit the enzymes to bind on specific nucleosomes with significantly stronger binding affinity<sup>28</sup>. DNA sequence elements, e.g., CpG islands, have also been shown to have higher but less sequence-specific enzyme binding affinity<sup>29-33</sup>. Therefore, we denote a special nucleation region of nucleosomes [for H3K4me3 and H3K9me3 centered around the transcription start site (TSS)] with lower binding energies. We will index the middle nucleosome within this region as 0, those on its left negative, and those on its right positive. Specifically, there is a nucleosome-free region near the TSS<sup>34</sup>, and some DNA distortion may be needed.

As shown in Figure 2.2, considering a gene with  $N$  nucleosome, each could reside in one of the three covalent state:  $s = 0$  (unmodified),  $s = 1$  (activated), and  $s = -1$  (repressed). There are 5 enzyme binding state of each nucleosome, which means there will be 15 ( $3 \times 5$ ) states for each nucleosome. This will lead to high number of calculations:  $15^N$ .

The system appears very complicated in the first look, but there is one clear time-scale separation, which allows us to greatly simplify the problem. The enzyme binding time scale is in sub-seconds to seconds level, where the epigenetic memory is in weeks. We borrow the quasi-equilibrium approximation assumption from the michaelis-menten equation we learned before,  $E + S \xrightleftharpoons[k_{-1}]{k_1} ES \xrightarrow{v} P + E$ .

Here the enzyme  $E$  binds to substrate  $S$  with much faster speed compared to the enzymatic reactions. Following the similar approximation, we could use the Boltzman distribution to calculate the possibility of histone modification enzyme binding.

Here is the algorithm to calculate the possibility: The interactions between covalent modification enzymes and a collection of nucleosomes at given  $s$  states can be described by the following Hamiltonian:

$$H_s = \sum_{i=1}^N \mathcal{E}_{\sigma_i, s} - \sum_{i=1}^{N-1} J_{\sigma_i \sigma_{i+1}} \quad (1)$$

Throughout this work, energy is given in units of  $k_B T$ , where  $k_B$  is the Boltzmann constant, and  $T$  is the temperature. The partition function is  $Z_s = \sum_{\{\sigma\}} \exp(-H_s) = Tr[T_1 \cdots T_N]$ , where the transfer

matrix  $T_i$  has elements  $(T_i)_{\alpha\beta} = \exp(-\frac{1}{2}(\varepsilon_{\alpha s_i} + \varepsilon_{\beta s_{i+1}}) + J_{\alpha\beta})$  for  $i=1, \dots, N-1$ , and

$(T_N)_{\alpha\beta} = \exp(-\frac{1}{2}(\varepsilon_{\alpha s_N} + \varepsilon_{\beta s_1}))$ . For notational simplicity, we omit the  $s$  dependence of the transfer

matrices. Then, the probability of finding site  $i$  in state  $\sigma_i$  is

$$P_{\sigma_i}(\{s\}) = \text{Tr}[T_1 \cdots T_{i-2} G_{\sigma_i} T_{i+1} \cdots T_N] / Z_s \quad (2)$$

where  $(G_{\sigma_i})_{\alpha\beta} = (T_{i-1})_{\alpha, \sigma_i} (T_i)_{\sigma_i, \beta}$  except  $(G_{\sigma_1})_{\alpha\beta} = (T_N)_{\alpha, \sigma_1} (T_1)_{\sigma_1, \beta}$ .

After we get the possibility of histone modification enzyme binding, we treat the enzyme binding/unbinding processes by the above quasi-equilibrium approximation, the following events can take place:

- 1) An enzymatic reaction at site  $i$  with rate  $k_i = \delta_{s_i, 0}(v'_{0 \rightarrow -1} \bar{p}_2 + v'_{0 \rightarrow 1} \bar{p}_4) + \delta_{s_i, -1} v'_{-1 \rightarrow 0} \bar{p}_3 + \delta_{s_i, 1} v'_{1 \rightarrow 0} \bar{p}_5$ . Here  $\delta_{ij}$  is the Kronecker delta function, which assumes a value 1 when  $i = j$ , and 0 when  $i \neq j$ . Notice here we take into account the fact that for an enzymatic reaction to take place, the corresponding enzyme has to bind to the nucleosome.
- 2) Histone replacement due to stochastic turnover ( $s_i \rightarrow 0$ ) with rate  $d$ .
- 3) Every time when cell divides, each histone has 50% probability to be partitioned into one of the daughter cells.

Therefore the overall simulation procedure is as follows:

For each step with covalent state,

- 1) Calculate  $\bar{p}_{\sigma_i}$  and  $\{k_i\}$ .
- 2) Define the transition rate array  $k = [k_1, k_2, \dots, k_N, d, \dots, d]$ . Then at a given simulation step, the total

reaction rate is given by  $\alpha_0 = Nd + \sum_{i=1}^N k_i$ , and an accumulative reaction rate array is given by

$$\alpha = [k_1, k_1 + k_2, \dots, \alpha_0].$$

- 3) Generate two random numbers  $r_1$  and  $r_2$  from a uniform distribution within  $[0,1]$ . The next time that an event will take place is given by  $dt = \frac{1}{\alpha_0} \ln\left(\frac{1}{r_1}\right)$ , so  $t \rightarrow t + dt$ , and the reaction channel taking place is the smallest integer  $m$  satisfying  $\alpha_m \geq r_2$ . In our case if  $m \leq N$ , and  $s_m = 0$ , generate another random number  $r_3$  from uniform distribution within  $[0,1]$ , repressive modification on site  $m$  if  $r_3 \leq (v'_{0 \rightarrow 1} \bar{p}_1) / (v'_{0 \rightarrow 1} \bar{p}_1 + v'_{0 \rightarrow -1} \bar{p}_3)$ , otherwise active modification. Update histone state  $s_i$  ( $i = \text{mod}(m, N)$ ) according to the selected reaction channel.
- 4) Repeat.
- 5) When time  $t$  reaches the cell division time, for each nucleosome  $i$  generate a random number  $r_4$  from a uniform distribution within  $[0,1]$ . If  $r_4 \leq 0.5$  then  $s_i = 0$ , meaning that the histone is replaced by a nascent unmarked one; otherwise keep the original value of  $s_i$ , meaning the original histone is partitioned to this daughter cell being monitored. Here for simplicity we assume that the cell cycle time is fixed, which can be easily modified if variation of cell cycle time needs to be considered.

Detailed description of the parameter and the parameter value are listed in Table 2.1. Parameter estimation procedure will be discussed in next section.

*Table 2.1: Model parameters*

Methylation enzyme free energy of binding within nucleation region $\varepsilon_{\sigma 0}$ , $\sigma = 2, 4$	-1
Demethylation enzyme free energy of binding within nucleation region $\varepsilon_{\sigma 0}$ , $\sigma = 3, 5$	1
Methylation enzyme free energy of binding outside nucleation region $\varepsilon_{\sigma 0}$ , $\sigma = 2, 4$	3
Demethylation enzyme free energy of binding outside nucleation region $\varepsilon_{\sigma 0}$ , $\sigma = 3, 5$	3
$s$ state related free energy of binding $\Delta\varepsilon$	2
Lateral interaction between two identical enzyme molecules $J_{\alpha\alpha}$	3
Lateral interaction between two identical enzyme molecules $J_{\alpha\beta}$ , $\alpha \neq \beta$	-2
Enzymatic reaction rate constant $v_{\alpha \rightarrow \beta}$	1.5/hour
Histone exchange rate $d$	0.6/hour
Cell cycle time	20 hours

## 2.4. Parameter estimation

Some parameter can be determined easily. If one assumes that some insulating elements constrain the histone modification patterns<sup>34</sup>, one can estimate the number  $N$  from the DNA length within the constraints. For a gene length  $\sim 10$ k bp including the promoter regions, the nucleosome length  $N = 40$ . Without insulating elements, the model studies of Hodges and Crabtree show that an inherently bound histone pattern domain can be formed when the mark addition and removal enzymes have comparable catalytic activities<sup>35</sup>. In that case the length of the domain is determined by the relative ratio between the addition and removal enzyme activities.

### 2.4.1. Nonspecific background free energy of binding of enzymes

Several experimental techniques, such as fluorescence recovery after photobleaching (FRAP) and fluorescence correlation spectroscopy (FCS), can provide quantitative information about protein-chromatin binding<sup>23</sup>. In the literature what is usually reported is the fraction of enzymes bound to the histones. Below we discuss how to roughly estimate the free energy of binding from the data. Since these measurements are genome wide, therefore they reflect nonspecific protein-chromatin bindings instead of specific bindings facilitated by DNA-sequence specific elements.

Experimental data reveals that nonspecific protein-chromatin bindings are weak. Therefore we assume that the probability of having two neighboring nucleosomes occupied (from nonspecific background binding) at the same time is negligible. That is, for parameter estimation purpose we can neglect possible effects of the lateral interaction  $J$ , and treat each nucleosome as independent. Each histone can have two states: empty or occupied. Then respect to an arbitrary reference state with binding energy  $\varepsilon_0$  and free enzyme concentration  $c_0$ , the binding energy with free enzyme concentration  $c_{free}$  is:

$$\varepsilon = \varepsilon_0 - k_B T \ln(c_{free} / c_0) \quad (3)$$

From the Boltzmann distribution, the probability to observe a histone in the bound state is

$$p_H = \frac{\exp(-\varepsilon/k_B T)}{1 + \exp(-\varepsilon/k_B T)} \quad (4)$$

then

$$\varepsilon = -k_B T \ln \frac{p_H}{1 - p_H} \quad (5)$$

$$\varepsilon_0 = -k_B T \ln \frac{p_H c_0}{(1 - p_H) c_{free}}$$

We could roughly estimate probability of histone in a bound state to be the percentage of nucleosome observed in a bound state. From data collected in literatures, we estimate the enzyme binding energy in the  $\sim 4 k_B T$  scale.

**Table 2.2:** Summarizes our estimations based on available experimental data.

	H3K9me3	H3K27me3	Refs
Enzyme	<i>HP1<math>\alpha</math></i>	Polycomb group (PcG) proteins	23,36,37
Cell source	Mouse L cells	Drosophila Neuroblasts / Embryo	23,36,37
Nuclear volume ( $\mu\text{M}^3$ )	435	200	23
Estimated nucleosome number	21,120,000 (L cells)	960,000 (Embryo (cycle 14))	23
Nucleosome concentration	80.6 mM (L cells)	7.97 $\mu\text{M}$ (Embryo (cycle 14))	23
Measured enzyme bound fraction	65% (Mouse NIH 3T3/iMEFs)	18.93% (Drosophila Neuroblasts cells)	36,37
Total enzyme concentration	1 $\mu\text{M}$	380 nM (Drosophila Neuroblasts cells)	36,37
Number of bound enzymes	149477	10350	Derived
$P_H$	0.004	0.0045	Derived based on eq. (2)
$c_{free}$	0.35 $\mu\text{M}$	0.308 $\mu\text{M}$	Derived
$\varepsilon$	4.5 $k_B T$	4.2 $k_B T$	Derived based on eq. (3)

The above table of values is based on *very rough* estimations. For example, we do not consider competition of binding from different types of enzymes. The total number of nucleosomes is expected to

be overestimated since there are nucleosome-free regions. Including these corrections reduce the number of free nucleosomes, and leads to a lower free energy of binding.

The key information we can infer from experimental data is that nonspecific binding of enzymes on DNA is very weak. Mechanistically this weak binding is reasonable. From the above table, the total number of nucleosomes is far more than that of the enzymes. Strong nonspecific binding would not allow a binding enzyme to move and interact with other nucleosomes, and seriously deplete the pool of free enzymes.

#### **2.4.2. Free energy of binding of enzymes within the nucleation region**

There is no quantitative information on the enzyme free energy of binding. Below we examine model behaviors within a range of binding affinities.

#### **2.4.3. Enzyme lateral interactions:**

Currently there is no quantitative information regarding the lateral interaction strength between two identical methylation enzymes available. Since the lateral interaction can be experimentally observed, it is reasonable to assume that it is larger than  $1 k_B T$ . The values of  $J_{aa}$  is chosen to reproduce the bell-like shaped histone methylation pattern centered around the nucleation in Figure 2.3.

There is no evidence of interactions between demethylation enzymes and between enzymes of different types. Below we will show results with different choices of these interactions.

#### **2.4.4. Enzyme rate constants:**

There is no information on the enzyme rate constants. Therefore for simplicity we use the same rate constants for all four enzymes, and choose the value which reproduces the experimental observation that it takes about 5 cell cycles to switch *Oct4* on or off<sup>38</sup>.

We want to point out that the enzyme rate constant and the free energy of binding can compensate each other. As described in the main text, an enzyme reaction rate depends on the product of the enzyme rate constant and the enzyme binding probability.

### 2.4.5. Histone exchange:

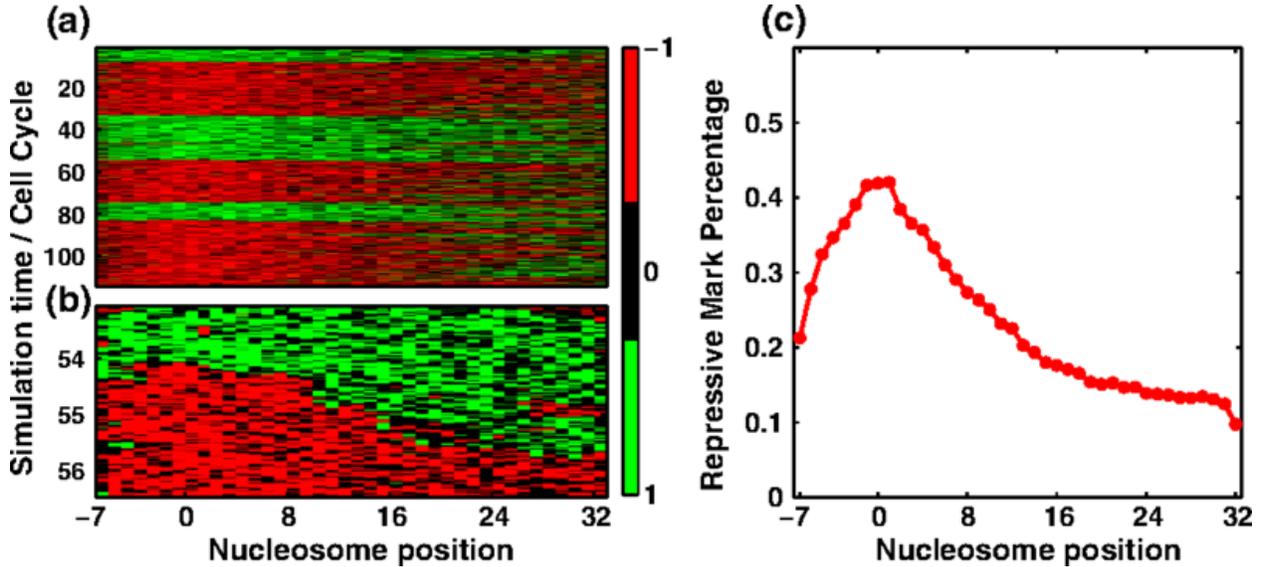
The reported value of the histone exchange rate varies over a broad range and show cell-type dependence. In reality one may also expect dependence of histone exchange rate on the covalent marks. Active transcriptions can lead to higher histone exchange rate<sup>31,39</sup>, and thus different histone exchange rates may exist for euchromatins and heterochromatins. For simplicity though, Angel *et al.* uses a single value estimated from measurements on *Drosophila* cells<sup>22</sup>. In this work, we will adopt this value as well. The rate may be different for cells from other organisms and may vary among cell types. Below we will show how different values of the histone exchange rate may affect the model behavior.

Another assumption we made is that after a histone is detached from the DNA, the rebinding histone is an unmodified one. While this assumption is typically assumed in the current literature<sup>22,40</sup>, validation of this assumption depends on how fast the rebinding process takes place after a histone is detached, that is, if there is sufficient time for this histone to diffuse away (or be demarked) and others to diffuse in. In other words, the exchange rate used in this work should be treated as an effective exchange rate, which excludes contributions from events with the same histone detaching and reattaching to the same nucleosome. We want to point out that in our model a detached histone replaced by an unmarked one is a perturbation to the epigenetic state. It is reasonable to assume that the local concentration of free histones bearing the same mark as those on the nucleosomes might be higher. This would lead to a more stable epigenetic state than what we examine in this work.

The effective histone exchange rate also contains contribution from another possible effect we do not consider explicitly here: active transcriptions can lead to higher histone exchange rate<sup>31,39</sup>, since histones need to be released to let the RNA polymerase pass then rebind.

In this work we do not consider the dependence of histone exchange rate on the covalent marks due to lack of experimental information.

## 2.5. Results

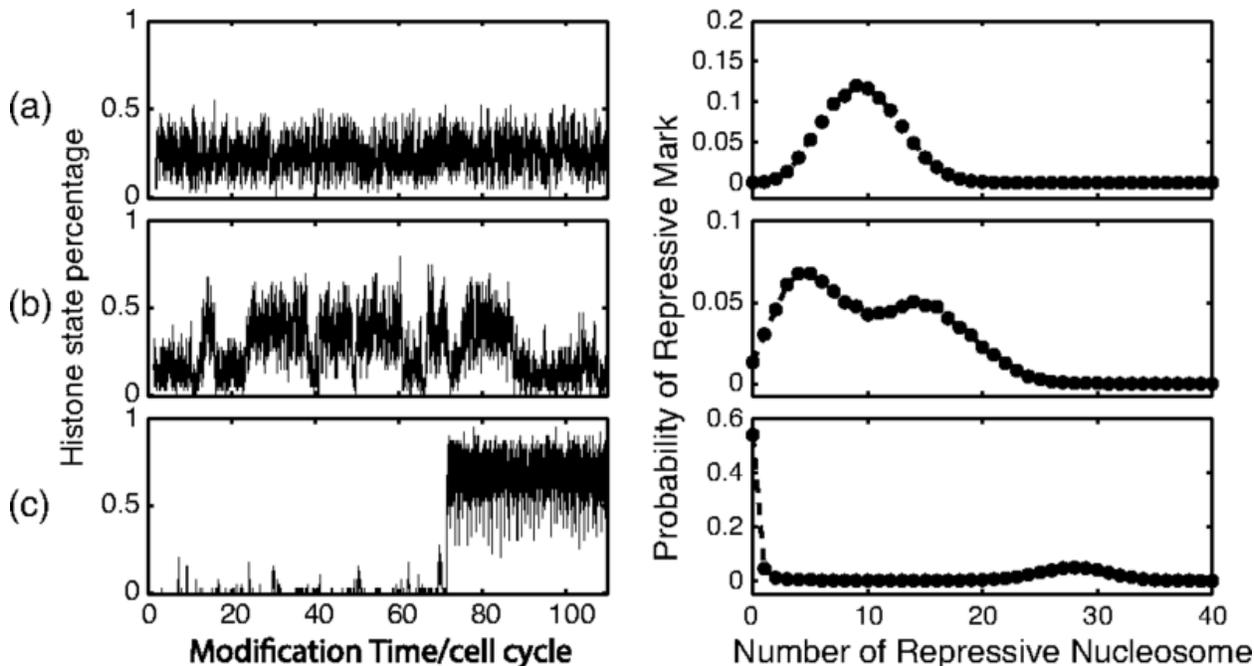


*Figure 2.3: Simulation results using model parameters corresponding to Oct4. (a): Heat map representation of a typical trajectory. (b): Zoom-in of the heat map in (a) showing epigenetic state transition. (c): Probability of observing repressive marks at different nucleosome sites. The nucleation region is at nucleosomes -1, 0, and 1.*

Figure 2.3(a) shows a typical simulated trajectory using parameters roughly representing the gene *Oct4*. Clearly the  $s$  state of each nucleosome changes randomly and frequently. However, the system can maintain one collective epigenetic state, dominated by either repressive or active marks, for many cell cycles before stochastically switch to another state. A zoom-in of the trajectory (Figure 2.3(b)) shows that a transition usually starts at one place, often within the nucleation region, and then propagates outwards. Statistically the system still spends most of the time around either the repressive or active mark dominated states. That is, if one plots the fraction of time the system have  $n$  nucleosomes bearing repressive marks out of the  $N$  nucleosomes, one obtains a histogram with a bimodal distribution. In other words, the system exists as a bistable system as shown in Figure 2.4.

Figure 2.4 shows that with enzyme lateral interaction  $J = 0$ , the percentage of nucleosomes with repressive marks fluctuates but shows a uni-modal distribution. A cell with this dynamical property cannot maintain a memory of its epigenetic state over generations. With  $J = 2.5$ , however, we could see a clear

two-state separation from the trajectory shows that the gene of the cell begin to have memory; the right panel also shows bi-modal distribution indicate the cell is in the bistable region. As  $J$  further increased to 3.5, the memory further increased and the cell could stay in one state for many generations.



*Figure 2.4: Typical trajectories of the fraction of nucleosomes with repressive marks (left) and the corresponding probability distribution of observing given number of nucleosomes with repressive marks (right). All simulations are performed with  $J = 2$ , but different  $J$  values, (a):  $J = 0$ , (b):  $J = 2.5$ , (c):  $J = 3.5$ . Other parameters values are from Table 2.1. The dwelling time distribution is obtained by averaging over 100 trajectories, each started with a randomly selected initial histone modification configuration, simulated for  $10^3$  Gillespie steps, then followed by another  $2 \times 10^3$  Gillespie steps for sampling.*

Close examination of the trajectory in Figure 2.4(c) reveals that a major contribution to nucleosome mark fluctuations is due to random replacements during every cell cycle. After each cell division, the fraction of repressive marks relaxes quickly to a steady state value before the next cell division. Figure 2.11 shows that the relaxation time is about 6 hours, which is also consistent with experimental measurement on HeLa cells<sup>41</sup>. It is natural to conjecture that this fast relaxation (less than one cell cycle) is necessary for maintaining a stable epigenetic state against cell division perturbation. We also define an average dwelling

time at an epigenetic state as the average time the system stays in the epigenetic state with one mark dominating before it switches to the state with another mark dominating; this is calculated using the algorithm adapted from Ref.<sup>42</sup>. Figure 2.10 shows that it increases with the cell cycle time. That is, a shorter cell cycle makes the epigenetic state less stable. This is consistent with experimental findings that increasing cell division rate accelerates the epigenetic reprogramming from differentiated cells to induced pluripotent stem cells<sup>43</sup>.

To further analyze the memory (bistability) dependence on the parameter, we explored the bistable region in the  $\Delta\varepsilon$ - $J$  plane. Figure 2.5(a) shows that a finite value of  $J$  is necessary for generating bimodal distributions of the fraction of histones with repressive marks. Below a critical value  $\sim 2$ , the system only shows unimodal distribution even with very large  $\Delta\varepsilon$  values. The required value of  $J$  also increases sharply upon decreasing  $\Delta\varepsilon$ . With  $\Delta\varepsilon = 0$ , the system can not generate a bimodal distribution with an arbitrarily large value of  $J$ . While one should be cautious of results with large (possibly unphysical) values of  $J$  and  $\Delta\varepsilon$  since the time-scale separation argument then becomes questionable, the results in Figure 2.5(a) suggest that both  $J$  and  $\Delta\varepsilon$  are necessary to generate bimodal distributions.

Previous results shows that the local neighbor interaction  $J$  is enough to generate the inheritable memory that could resist the histone exchange and cell division perturbation. Furthermore, the trajectory heatmap in Figure 2.2 also suggests that the histone mark could propagate and occupy the whole gene. However, Dodd *et al.*'s study states that the non-local correlation is required to achieve the bistability. To gain mechanistic understand and test our current model is contradicted with the results of Dodd *et al.*, we present a statistical analysis on possible correlations between different nucleosomes.

We define the correlations for the  $\sigma$  states of two nucleosomes  $i$  and  $j$  with a given set of  $s$  configurations, i.e., the correlation between nucleosome  $i$  in state  $\sigma_i = \alpha$  and nucleosome  $j$  in  $\sigma_j = \beta$  as

$$\begin{aligned}
 C_{\alpha,\beta}(\sigma_i, \sigma_j; \{s\}) &= \frac{\langle \delta_{\sigma_i,\alpha} \delta_{\sigma_j,\beta} \rangle_s - \langle \delta_{\sigma_i,\alpha} \rangle_s \langle \delta_{\sigma_j,\beta} \rangle_s}{\langle \delta_{\sigma_i,\alpha} \delta_{\sigma_i,\alpha} \rangle_s - \langle \delta_{\sigma_i,\alpha} \rangle_s \langle \delta_{\sigma_i,\alpha} \rangle_s} \\
 &= \frac{P_{\sigma_i=\alpha, \sigma_j=\beta}(\{s\}) - P_{\sigma_i=\alpha}(\{s\})P_{\sigma_j=\beta}(\{s\})}{P_{\sigma_i=\alpha}(\{s\})(1 - P_{\sigma_i=\alpha}(\{s\}))}.
 \end{aligned} \tag{6}$$

Averaging over  $N_s$  consecutive samples of Gillespie simulations with the waiting time at each step  $\tau_l$ , and the total simulation time  $t$ , we obtain the correlation functions averaged over the  $s$  states

$$\bar{C}_{\alpha\beta}(\sigma_i, \sigma_j) = \sum_{l=1}^{N_s} C_{\alpha\beta}(\sigma_i, \sigma_j; \{s\})\tau_l/t \quad (7)$$

And  $s$ -state correlation:

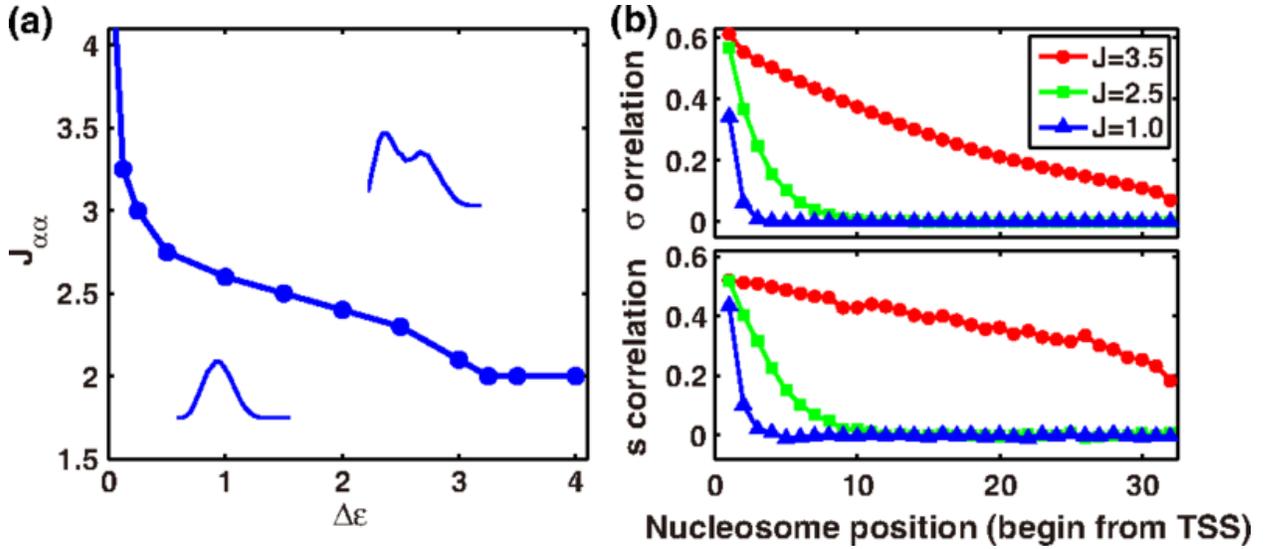
$$C_{a,b}(s_i, s_j) = \frac{\langle \delta_{s_i,a} \delta_{s_j,b} \rangle - \langle \delta_{s_i,a} \rangle \langle \delta_{s_j,b} \rangle}{\langle \delta_{s_i,a} \delta_{s_i,a} \rangle - \langle \delta_{s_i,a} \rangle \langle \delta_{s_i,a} \rangle} \quad (8)$$

Here

$$\langle \delta_{s_i,a} \rangle = \sum_{l=1}^{N_s} \delta_{s_i(l),a} \tau_l / t \quad (9)$$

As shown in Figure 2.5, the nucleosome enzyme binding states show correlations from the smallest length scale, nearest neighbors for small  $J$  values, to the larger length scales spanning the whole region for sufficiently large  $J$  values.

The nucleosome enzyme binding states show correlations from the smallest length scale, nearest neighbors for small  $J$  values, to the larger length scales spanning the whole region for sufficiently large  $J$  values [Figure 2.5 (b)]. It is not surprising for a Potts-type model with nearest-neighbor interactions to give rise to beyond-nearest-neighbor correlations of  $\sigma$  states. Because the  $\sigma$  and  $s$  state dynamics are coupled, the  $s$  states of nucleosomes also show similar correlations. These nonlocal nucleosome-nucleosome  $s$  state correlations are mediated through enzyme binding.



**Figure 2.5: Mechanism of bistability.** (a): Phase diagram on the  $\Delta\epsilon$ - $J$  plane. All other parameters take values in Table 2.1. (b): Correlation functions of enzyme bound state  $\sigma$  and nucleosome states.

Overall, the above results demonstrate that the inheritable memory could be achieved with the energy difference of enzyme  $\Delta\epsilon$  on different marks and enzyme lateral interactions  $J$ . This corresponds to a simple story, the  $\Delta\epsilon$  term represents the enzyme preference, which would facilitate the enzyme recognize its corresponding histone mark, and the  $J$  term could help the enzyme correctly be recruited to the desired nucleosome position and make corresponding modification. In short,  $\Delta\epsilon$  will help reading and  $J$  will help writing, form a simple ‘reader and writer’ scheme to establish the inheritable histone epigenetic memory.

## 2.6. Robustness of the proposed physical mechanism over parameters

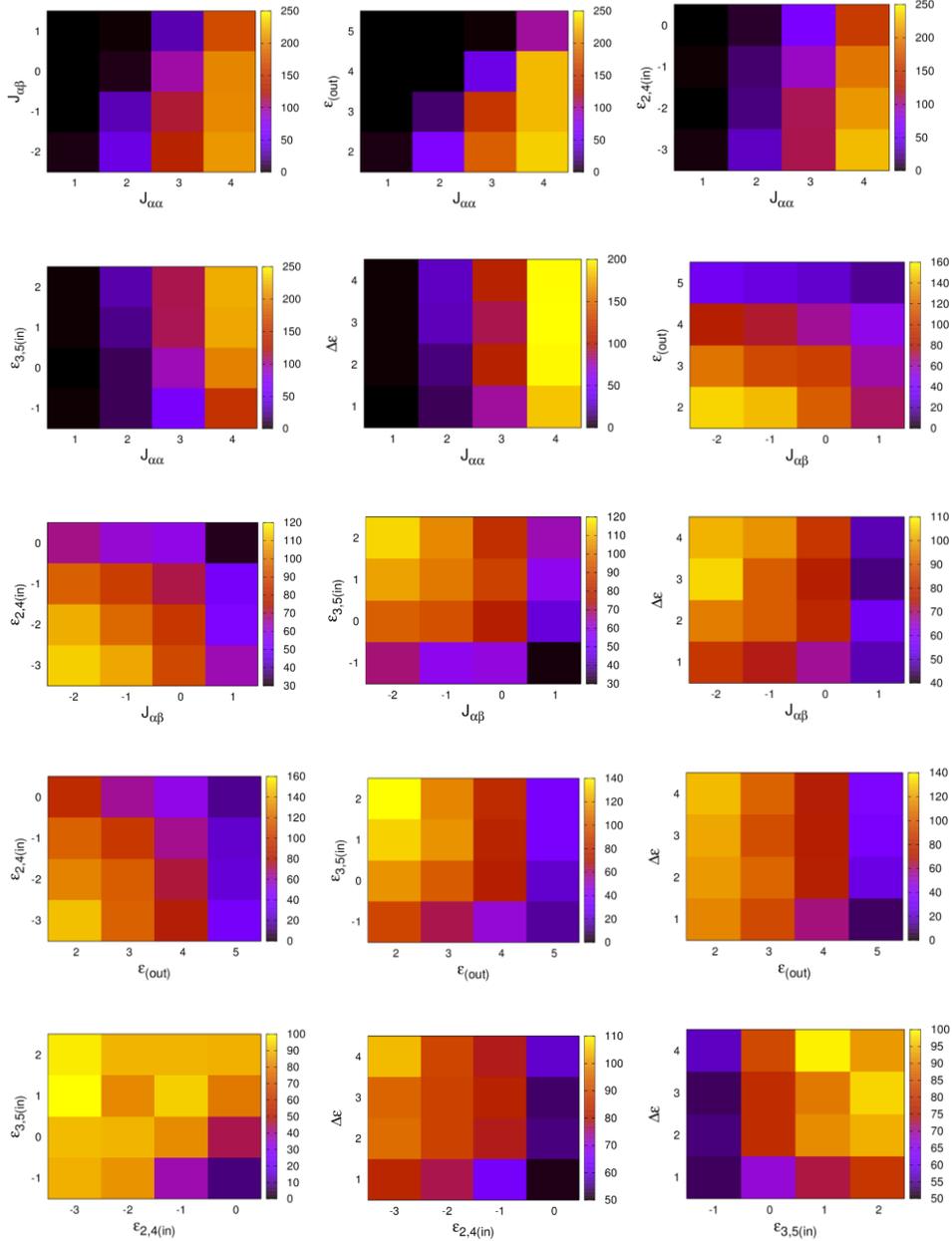
### 2.6.1. Binding affinities and lateral interactions

As described above, most of the model parameters are crudely estimated from experimental data. To show that the physical mechanism proposed in this work is not a result of fine-tuning the model parameters, we performed model studies within a range of parameters. In previous section, we have shown the phase diagram between  $\Delta\epsilon$  and  $J_{\alpha\alpha}$ . To further explore how the proposed physical mechanism is robust against parameters, we performed simulations using 4096 sets of parameters in the 6-dimensional parameter space, with each dimension divided into 4 equally distributed grid points within a physically reasonable range.

The 6 parameters are the binding affinities and lateral interactions. We also used a more stringent criterion for the bistable region compared to what was used to generate Figure 2.5: clear separation between the epigenetic states with high and low average number of nucleosomes with repressive marks ( $>4.5$ ), significant epigenetic memory with the average dwelling time on each epigenetic state  $> 2$  cell cycles. We obtain 1238 (30%) parameter sets satisfying the above requirement. This clearly indicates that the mechanism we propose here is robust against parameter choices.

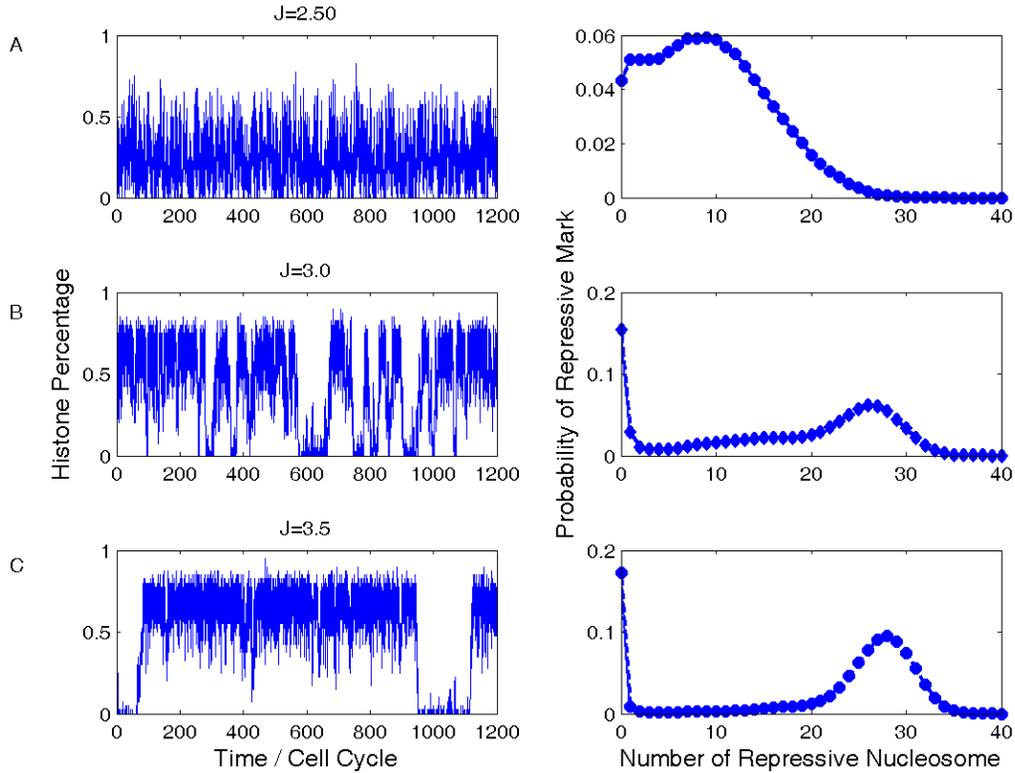
The heatmaps in Figure 2.6 reveal the correlations between 15 pairs of parameters. From the results one can see several trends,

- 1)  $J_{aa}$ , the lateral interactions between two enzymes of the same type, have major contributions for stabilizing bound enzymes and thus epigenetic states.  $J_{ab}$ , the lateral interactions between two enzymes of different types, can be positive (attraction) or negative (repulsion), but should be smaller than  $J_{aa}$ . Physically it is easy to understand since otherwise an enzyme has no preference to recruit an enzyme of the same type, a key ingredient in our model. A smaller  $J_{ab}$  leads to more stable epigenetic states.
- 2) A higher free energy of binding (high positive value) of methylation enzymes would require compensation from stronger lateral interactions  $J_{aa}$  so an enzyme can recruit another one of its kind for subsequent covalent modification. On the other hand, a higher free energy of binding of demethylation enzymes is preferred for a stable epigenetic state.
- 3) Higher  $\Delta\epsilon$  means better distinction between the covalent states of a nucleosome and thus more stable epigenetic states.



**Figure 2.6: Bistable parameter sets distribution heatmaps of 4096 sets of parameters.** *X label and y label of each heatmap are varying parameters. All the energy units are in  $k_B T$ . In each heatmap, there are total 4096 simulated parameter sets. Each grid has  $256 (= 4096/16)$  parameter sets. The value in each grid describes the number of parameter sets which meet the bistable requirement. The total number of parameter sets which satisfying the requirement is 1238.*

Since no evidence on the lateral interactions between removal enzymes have been reported, we also examined the case with  $J_{33} = J_{55} = 0$ . Figure 2.7 shows that these two interaction terms are not essential for generating memory, but existence of these interactions do help stabilize the epigenetic states.

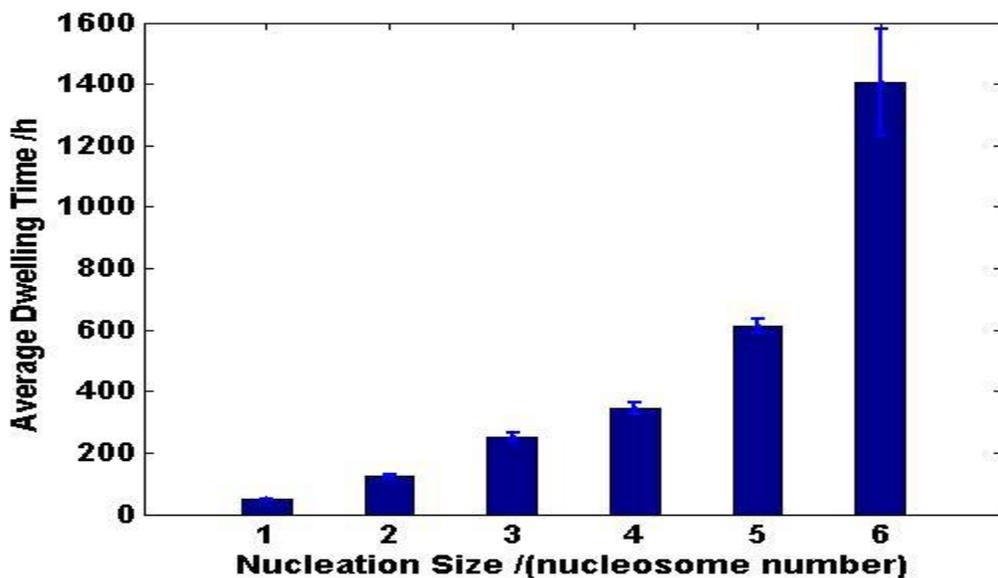


**Figure 2.7** *Effect of lateral interactions between removal enzymes on generating bimodal nucleosome mark distribution. (A-C) Typical trajectories of the fraction of nucleosomes with repressive marks (left) and the corresponding probability distribution of observing given number of nucleosomes with repressive marks (right). All simulations are performed with  $\Delta\varepsilon = 2$ , but different  $J_{aa}$  values, A:  $J_{aa} = 2.5$ , B:  $J_{aa} = 3.0$ , C:  $J_{aa} = 3.5$ . Here, the simulations are performed without these removal enzyme lateral interactions ( $J_{33} = J_{55} = 0$ ), while other parameters values are taken in Table 2.1.*

### 2.6.2. Nucleation region length

Since the exact length of the nucleation region is unknown, we examined how the average dwelling time of an epigenetic state depends on the nucleation region size. Not surprisingly, Figure 2.8 shows that

the dwelling time increases with the nucleation region size. This is because nucleosomes within the nucleation region have lower free energy of binding for the modification enzymes, and thus a faster rate to restore the epigenetic state after perturbations from cell division and histone exchange.



*Figure 2.8 Average state dwelling time as a function of the nucleation region size. All simulations are performed with  $\Delta\epsilon = 2$ , but different nucleation size while other parameters values are taken in Table 2.1. The number of nucleosomes within nucleation region is chosen as 1, 2, 3, 4, 5, 6 centered around nucleosome 0.*

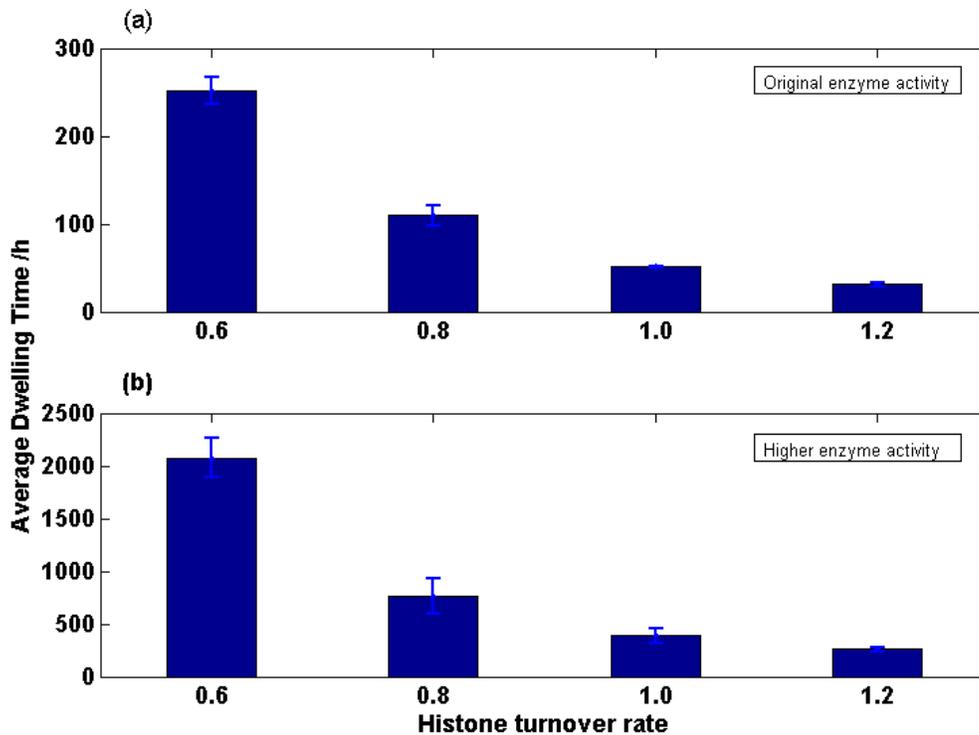
### 2.6.3. Histone exchange rate

In this work, we used a histone exchange rate measured for *Drosophila* cells. The rate may be different for cells of other organisms and may vary among cell types. We performed simulations with different exchange rates. The results in Figure 2.9(a) show that the epigenetic state stability depends strongly on the histone exchange rate, indicating that histone exchange is one of the main perturbations to the epigenetic memory. At an exchange rate  $d = 1.2 \text{ h}^{-1}$ , the epigenetic state average dwelling time has dropped to about 1 cell cycle.

There is no quantitative information on another parameter, the enzymatic reaction rate constant  $\nu$ , so we used a rough estimate. We hypothesize that if the “true” histone exchange rate is larger than that

estimated from *Drosophila* cells, the epigenetic state can still be stabilized with an increasing value of  $\nu$ . Indeed, even with a  $\nu$  value changed from  $1.5 \text{ h}^{-1}$  to  $3 \text{ h}^{-1}$ , and a 2-fold increase of the exchange rate from  $0.6 \text{ h}^{-1}$  used in the main text, the average dwelling time on an epigenetic state is still about 10 cell cycles (Figure 2.9(b)). That is, an increase of  $d$  can be compensated by an increase of  $\nu$  about the same fold. Therefore we suggest that accurate measurements of both the histone exchange rate and the enzyme rate constants are needed to evaluate epigenetic state stability.

The simulation results showing increasing histone exchange rate decreases epigenetic state stability are consistent with existing experimental observations. For example, embryonic stem cells tend to have higher histone exchange rate<sup>44</sup>, and spontaneous transitions between different epigenetic states are observed<sup>45</sup>. We hypothesize that different cell types may have different requirements on maintaining epigenetic histone memory. It may not be necessary to maintain long epigenetic histone memory during the early embryo development stage since cells differentiate into different cell types quickly.



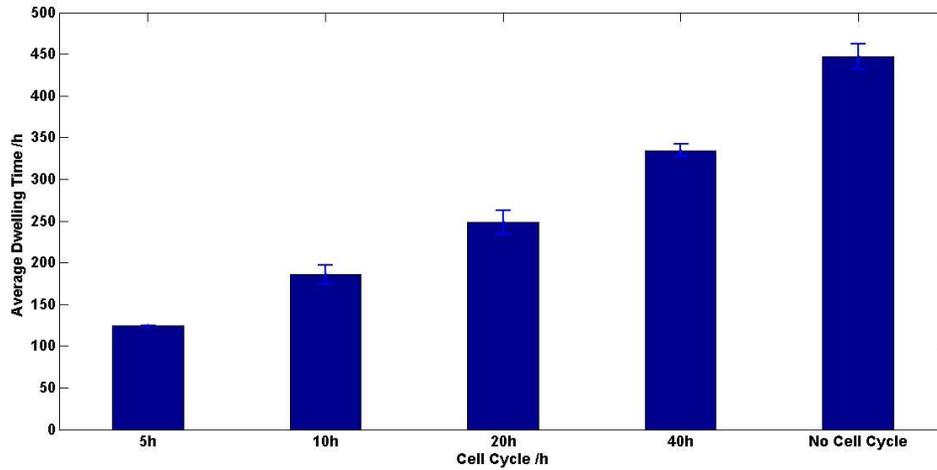
**Figure 2.9 Dependence of average dwelling time on histone exchange rate.** (a) All other parameters are from Table 2.1. (b) Parameters similar to the upper panel are used, except with a higher enzyme rate constant  $v = 3$ .

Our model predicts that a faster histone exchange rate can be counter-balanced by more active enzymes or higher enzyme concentrations. It remains to be tested whether, for example acetylation enzymes have higher activity or concentration than methylation enzymes, since in HeLa cells experiments show histones with acetylation marks have faster exchange rate than those with methylation marks. For embryonic stem cells, where histone exchange rates are high<sup>40,44</sup>, indeed the covalent modification enzyme concentrations are higher than those in differentiated cells<sup>46</sup>.

Interestingly, Yang *et al.* reported auto-acetylation of the active site on a histone acetyltransferase (HAT) result in only slight modification (less than 1 fold) of the enzyme activity<sup>47</sup>. However, Yuan *et al.* showed this slight change of enzyme activity was fatal for yeast survival<sup>48</sup>. Our result in Figure 2.9 suggests that a small change of the histone modification enzyme activity may lead to dramatic change of epigenetic memory. Consequently, expression activities of many genes may be affected. This may explain the above mentioned sensitive dependence of yeast survival on HAT activity.

#### **2.6.4. Cell cycle time**

Different cell types may have different cell cycle time. Human embryonic stem cells have a cell cycle time of 10 hours or less<sup>49</sup>. As shown in Figure 2.10, the average dwelling time of an epigenetic state do increase with increasing cell cycle time. This is consistent with the observation that histone repartitioning during cell divisions is a strong perturbation to the epigenetic states. We also notice that even with a hypothetical cell cycle time of 5 hours, the average dwelling time is  $\sim 120$  hours, indicating that the epigenetic memory can still be maintained over a number of cell generations.



**Figure 2.10** Average state dwelling time as a function of the cell cycle. In these calculations  $J_{aa} = 3.0$ , and other parameters are the same as in Table 2.1 of the main text. The dwelling time on the repressive state is estimated as the time from first reaching the average nucleosome fraction (0.5) of the repressive state to the first reaching the average fraction (0.1) of the active state. Each average is performed over three trajectories of 10000 cell cycles in length.

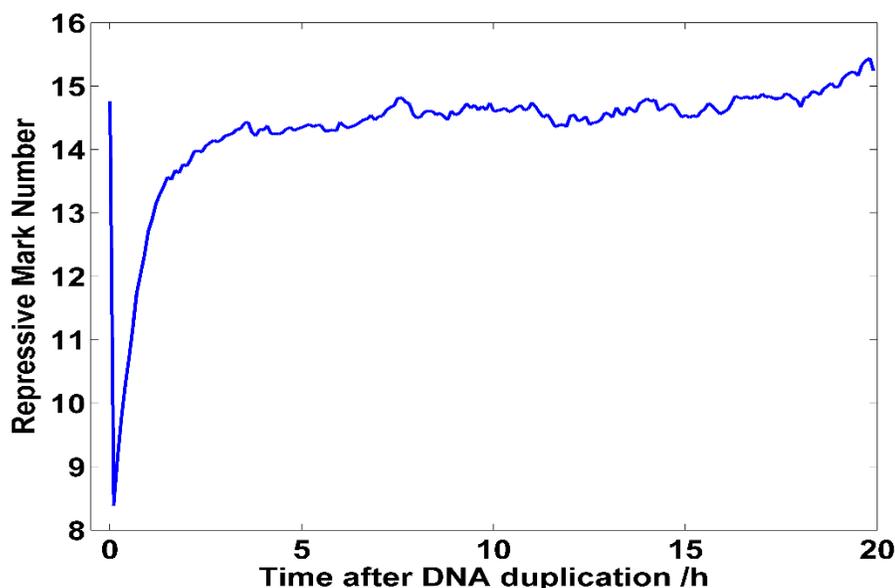
## 2.7. Ideas for future experimental measurements

In this work we only use parameters roughly estimated from the available experimental measurements, and do not thoroughly explore the model behaviors with different choices of model parameters. Our analysis suggests that enzyme lateral interactions of modest strengths are necessary for maintaining epigenetic memory. There are structural evidences for the interactions between H3K4me3 enzymes<sup>50</sup>, and between H3k9 enzymes<sup>25</sup>, either directly or possibly through some scaffold proteins. To the best of our knowledge, there are no existing report on the lateral interactions between demethylation enzymes. We performed a set of simulations without these removal enzyme lateral interactions  $J_{33} = J_{55} = 0$ , and find they are not necessary for generating the bistable behavior (see Figure 2.7). The exact size of the nucleation region is also currently not known. Figure 2.8 shows that a longer nucleation region requires a smaller  $J$  for generating bistability.

We suggest that more accurate measurements of the parameters in the model<sup>51</sup>, binding affinities and enzymatic reaction rates, are necessary to help determine the relevance of the discussed mechanism here.

Specifically, existing measurements on the enzyme binding fractions and residence times are all genome wide. However, other experimental evidences suggest strong genome position dependence<sup>23</sup>.

In this model a key ingredient is that a bound enzyme can recruit another enzyme of its type through lateral interactions. We suggest that one may repeat the experiments of Canzio *et al.*<sup>25</sup> but with different linker lengths for the DNA and/or for different domains of the enzyme, and examine how the probability of observing oligomers bound to nucleosomes would be affected. Indeed Yuan *et al.* shows that H3K27 methylation requires chromatin to form compact structure<sup>52</sup>.



*Figure 2.11 Relaxation dynamics of the total number of nucleosomes with repressive marks. At time 0, DNA is duplicated and nucleosomes bearing repressive marks are randomly divided into two sister DNA sequences, and the previous number of nucleosomes bearing repressive marks roughly decreased into half of the number before DNA duplication. In these calculations  $J_{\alpha\alpha} = 3.0$ , and other parameters are the same as in Table 2.1 of the main text.*

## **2.8. Relation between the statistical mechanics model of Zhang *et al.* and the rule-based model of Dodd *et al.***

We summarize the two approaches for modeling the mechanism for epigenetic memory, our statistical mechanics model and the rule-based model of Dodd *et al.*<sup>19</sup>. Our model Hamiltonian describes local

interactions between covalent modification enzymes bound to neighboring nucleosomes and enzyme-nucleosome binding. After integrating out the enzyme binding/unbinding processes as performed in the main text, each enzymatic reaction rate is given by  $k_i = \nu p_{s_i}(\{s\})$ . The term  $p_{s_i}(\{s\})$  involves enzyme mediated effective nonlocal interactions among nucleosomes. In the rule-based model of Dodd *et al*<sup>19</sup>, the corresponding effective nonlocal interactions are given *a priori*. To further see relations between the two models, let's examine a simplified model.

Consider a collection of three nucleosomes bound by insulating elements. Here for simplicity and to compare with that of Dodd *et al*<sup>19</sup>, we neglect the nucleation region. We apply the same rule-based algorithm to the model used in Dodd *et al*<sup>19</sup>:

Step 1—A random nucleosome  $n_1$  to be modified is selected among the  $N (= 3)$  nucleosomes. With probability  $\alpha$ , a positive feedback (recruited) conversion of  $n_1$  is attempted (Step 2A), OR (with probability  $1 - \alpha$ ), a noisy change of  $n_1$  is attempted (Step 2B).

Step 2A—Recruited conversion: A second random nucleosome  $n_2$  is selected from anywhere within the region, and if  $n_2$  is in either the R (repressed,  $s = -1$ , denoted M in Dodd *et al*<sup>19</sup>.) or the A (activated,  $s = 1$ , also denoted A in Dodd *et al*<sup>19</sup>) state,  $n_1$  is changed one step toward the state of  $n_2$ . That is, if  $n_2$  is R, then  $n_1$  is changed from A/U to U/M; if  $n_2$  is A, then  $n_1$  is changed from R/U to U/A, where U represents a unmarked state ( $s = 0$ ). If  $n_1$  and  $n_2$  are in the same state, or if  $n_2$  is a U state, then no change is performed.

Step 2B—Noisy conversion: Nucleosome  $n_1$  is changed one step toward either of the other types (i.e., no direct A and R inter-conversions) with a probability of one-third.

Notice that in this model the time step is arbitrary and discrete. Following Dodd *et al.*, we use the parameter values  $\alpha = 1.4/2.4$  (corresponds to  $F = \alpha/(1 - \alpha) = 1.4$ ), and we will specifically focus on the transition probabilities of nucleosome 3 from  $s = 0$  to  $s = -1$ .

For comparison, we also calculate the probabilities of  $s_3 = 0 \rightarrow s_3 = -1$  using the present model. We use the parameters given in Table 2.1 of the main text. Following the procedure discussed in the main text, we can calculate  $p_3$ , the probability of having repressive mark adding enzyme bound to nucleosome 3 that is at

covalent state  $s = 0$ . The corresponding enzymatic reaction rate for adding the repressive mark is  $k_3 = p_3 v$ . Suppose that in a Gillespie step nucleosome 3 has been chosen to change, we calculate the probability of adding a repressive mark (with rate  $k_3$ ) versus other two possibilities, adding an active mark (with rate  $k_3' = p_1 v$ , where  $p_1$  is the probability of having active mark adding enzyme bound), and remaining unmarked (due to histone exchange with rate  $d$ ), which is given by  $k_3 / (k_3 + k_3' + d)$ .

Table 2.3 summarizes the transition probabilities from the two models. Note that although it is not meaningful to compare the two quantitatively, one can compare the qualitative trend of the two models. Our statistical mechanics model and Dodd's full rule-based model show that the transition dynamics of nucleosome 3 can be influenced by both of the other two nucleosomes, which differs from the next neighbor-limited model also analyzed by Dodd *et al*<sup>19</sup>. The latter is shown not to be a robust mechanism for epigenetic memory. In other words, if one integrates out the enzyme binding/unbinding process, one qualitatively recovers the rule-based model of Dodd *et al*<sup>19</sup>. The main difference is that in our statistical mechanics model, the beyond-nearest-neighbor influence (or effective nonlocal interaction) is not described by a set of rules set *a priori*, but derived from a fundamental Hamiltonian.

Table 2.3 Comparison between the present model and that of Dodd *et al*<sup>19</sup>.

$s_1, s_2, s_3$	Probability of $s_3 = 0 \rightarrow s_3 = -1$ from statistical mechanics model	Probability of $s_3 = 0 \rightarrow s_3 = -1$ from Dodd's full model		Probability of $s_3 = 0 \rightarrow s_3 = -1$ from Dodd's neighbor-limited model	
-1,-1,0	0.35	$(1-\alpha)/3+2\alpha/2$	0.72	$(1-\alpha)/3+ \alpha/2$	0.43
0,-1,0	0.30	$(1-\alpha)/3+\alpha/2$	0.43	$(1-\alpha)/3+ \alpha/2$	0.43
1,-1,0	0.23	$(1-\alpha)/3+\alpha/2$	0.43	$(1-\alpha)/3+ \alpha/2$	0.43
-1,0,0	0.22	$(1-\alpha)/3+ \alpha/2$	0.43	$(1-\alpha)/3$	0.14
0,0,0	0.13	$(1-\alpha)/3$	0.14	$(1-\alpha)/3$	0.14
1,0,0	0.07	$(1-\alpha)/3$	0.14	$(1-\alpha)/3$	0.14
-1,1,0	0.06	$(1-\alpha)/3+ \alpha/2$	0.43	$(1-\alpha)/3$	0.14
0,1,0	0.03	$(1-\alpha)/3$	0.14	$(1-\alpha)/3$	0.14
1,1,0	0.01	$(1-\alpha)/3$	0.14	$(1-\alpha)/3$	0.14

Since Dodd's full model is based on a set of rules, while the statistical mechanics model explicitly take into account physical interactions, there are some major differences as well. The statistical mechanics model has dependence on the linear position of the nucleosomes, but Dodd's full model has no spatial dependence,

although it can be generalized to have such a dependence, as discussed in the original paper of Dodd *et al*<sup>19</sup>. For the statistical mechanics model, even if neither nucleosome 1 nor 2 is in state  $s = -1$ , nucleosome 3 still has a finite probability to recruit a repressive mark adding enzyme to change  $s_3$  to  $-1$ . However for Dodd's model,  $s_3$  can switch to  $-1$  only through noisy conversion. In other words, one cannot directly relate the recruited conversion and noisy conversion in Dodd's model to the enzyme catalyzed mark adding/removal and histone exchange processes discussed here.

## 2.9. Conclusion and Discussion

In summary, our model analysis shows that the experimentally observed nearest-neighbor interaction and modification state biased enzyme recruitment of individual nucleosomes work synergistically and sufficiently to result in collective active and repressive epigenetic states. Unlike a simple 1D model with nearest neighbor interactions that shows no phase transition, the coupled two-layer model here gives rise to bistability due to positive feedback of the nucleosome mark state to enzyme recruitment. The model supports the proposal of Dodd *et al*<sup>19</sup>. that nonlocal "effective interactions" among nucleosomes affect the covalent modification rates (as evidenced by the dependence of  $P_{\sigma_i}$  on  $s$  states of all nucleosomes) and are necessary for generating robust bistable epigenetic states. In section 2.8, we compare the two models. Our analysis demonstrates a possible molecular mechanism of generating these effective interactions, and epigenetic memory, mediated through nearest-neighbor enzyme lateral interactions. Let us focus on a specific unmarked nucleosome. Without interactions from other nucleosomes, with a set of symmetrically chosen parameters, the nucleosome has equal probability of being actively or repressively modified. The term  $\Delta\epsilon$  determines what types of enzymes are likely to bind on other nucleosomes within the correlation region. The enzyme lateral interactions ( $J$ ) result in the stabilization of enzyme binding on this tagged site by the binding events at other nucleosomes within the correlation region. This allows the nucleosome to "read" the majority epigenetic mark type of these nucleosomes, bias its recruitment of the corresponding enzyme, and "write" on itself accordingly. As shown in the previous section, this mechanism is robust with different choices of model parameter values, with the essential requirement that the time scale for mark

restoration must be faster than that of perturbations, mainly from mark removal reactions, cell division, and histone exchange, all of which may vary significantly among different cell types.

Our analysis does not rule out other possible mechanisms for epigenetic memory, such as direct interactions among distant nucleosomes due to compact histone structure<sup>50</sup>. Inclusion of these interactions extends the present one dimensional two-layer Potts model into higher dimensions and one expects even richer physics. Furthermore, epigenetic memory is maintained by a closed network coupling regulations at different levels including gene expression, epigenetic modification, chromatin remodeling, etc.<sup>53,54</sup>, and requires an integrated treatment in the future.

## References

1. Kouzarides, T. Chromatin Modifications and Their Function. *Cell* **128**, 693-705 (2007).
2. Bannister, A.J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res* **21**, 381-95 (2011).
3. Henikoff, S. & Shilatifard, A. Histone modification: cause or cog? *Trends Genet* **27**, 389-96 (2011).
4. Black, J.C., Van Rechem, C. & Whetstone, J.R. Histone lysine methylation dynamics: establishment, regulation, and biological impact. *Mol Cell* **48**, 491-507 (2012).
5. Greer, E.L. & Shi, Y. Histone methylation: a dynamic mark in health, disease and inheritance. *Nat Rev Genet* **13**, 343-57 (2012).
6. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-76 (2006).
7. Maherali, N. *et al.* Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* **1**, 55-70 (2007).
8. Mansour, A.A. *et al.* The H3K27 demethylase Utx regulates somatic and germ cell epigenetic reprogramming. *Nature* **488**, 409-13 (2012).
9. Nottke, A., Colaiacovo, M.P. & Shi, Y. Developmental roles of the histone lysine demethylases. *Development* **136**, 879-89 (2009).
10. Schuettengruber, B. & Cavalli, G. Recruitment of polycomb group complexes and their role in the dynamic regulation of cell fate choice. *Development* **136**, 3531-42 (2009).
11. Kim, D.H., Doyle, M.R., Sung, S. & Amasino, R.M. Vernalization: winter and the timing of flowering in plants. *Annu Rev Cell Dev Biol* **25**, 277-99 (2009).
12. Magklara, A. *et al.* An epigenetic signature for monoallelic olfactory receptor expression. *Cell* **145**, 555-70 (2011).
13. Clowney, E.J. *et al.* Nuclear Aggregation of Olfactory Receptor Genes Governs Their Monogenic Expression. *Cell* **151**, 724-737 (2012).

14. Lyons, D.B. *et al.* An epigenetic trap stabilizes singular olfactory receptor expression. *Cell* **154**, 325-36 (2013).
15. Goldberg, A.D., Allis, C.D. & Bernstein, E. Epigenetics: a landscape takes shape. *Cell* **128**, 635-8 (2007).
16. D'Urso, A. & Brickner, J.H. Mechanisms of epigenetic memory. *Trends Genet* **30**, 230-6 (2014).
17. Tuesta, L.M. & Zhang, Y. Mechanisms of epigenetic memory and addiction. *EMBO J* **33**, 1091-103 (2014).
18. Alberts, B. *et al.* *Molecular Biology of the Cell*, 1725-1725 (Garland Press, 2008).
19. Dodd, I.B., Micheelsen, M.A., Sneppen, K. & Thon, G. Theoretical analysis of epigenetic cell memory by nucleosome modification. *Cell* **129**, 813-22 (2007).
20. Sedighi, M. & Sengupta, A.M. Epigenetic chromatin silencing: bistability and front propagation. *Phys Biol* **4**, 246-55 (2007).
21. Binder, H. *et al.* Transcriptional regulation by histone modifications: towards a theory of chromatin re-organization during stem cell differentiation. *Phys Biol* **10**, 026006 (2013).
22. Angel, A., Song, J., Dean, C. & Howard, M. A Polycomb-based switch underlying quantitative epigenetic memory. *Nature* **476**, 105-8 (2011).
23. Steffen, P.A., Fonseca, J.P. & Ringrose, L. Epigenetics meets mathematics: towards a quantitative understanding of chromatin biology. *Bioessays* **34**, 901-13 (2012).
24. Schmitges, F.W. *et al.* Histone Methylation by PRC2 Is Inhibited by Active Chromatin Marks. *Molecular Cell* **42**, 330-341 (2011).
25. Canzio, D. *et al.* Chromodomain-mediated oligomerization of HP1 suggests a nucleosome-bridging mechanism for heterochromatin assembly. *Molecular Cell* **41**, 67-81 (2011).
26. So, C.W., Lin, M., Ayton, P.M., Chen, E.H. & Cleary, M.L. Dimerization contributes to oncogenic activation of MLL chimeras in acute leukemias. *Cancer Cell* **4**, 99-110 (2003).
27. Norwood, L.E. *et al.* A requirement for dimerization of HP1Hsa in suppression of breast cancer invasion. *Journal of Biological Chemistry* **281**, 18668-18676 (2006).

28. Ku, M. *et al.* Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet* **4**, e1000242 (2008).
29. Boyer, L.A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *cell* **122**, 947-956 (2005).
30. Probst, A.V., Dunleavy, E. & Almouzni, G. Epigenetic inheritance during the cell cycle. *Nature reviews Molecular cell biology* **10**, 192-206 (2009).
31. Deal, R.B., Henikoff, J.G. & Henikoff, S. Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. *Science* **328**, 1161-4 (2010).
32. Figueiredo, M.L., Philip, P., Stenberg, P. & Larsson, J. HP1a recruitment to promoters is independent of H3K9 methylation in *Drosophila melanogaster*. (2012).
33. Buscaino, A. *et al.* Distinct roles for Sir2 and RNAi in centromeric heterochromatin nucleation, spreading and maintenance. *EMBO J* (2013).
34. Bushey, A.M., Dorman, E.R. & Corces, V.G. Chromatin Insulators: Regulatory Mechanisms and Epigenetic Inheritance. *Molecular Cell* **32**, 1-9 (2008).
35. Hodges, C. & Crabtree, G.R. Dynamics of inherently bounded histone modification domains. *Proceedings of the National Academy of Sciences* (2012).
36. Muller, K.P. *et al.* Multiscale analysis of dynamics and interactions of heterochromatin protein 1 by fluorescence fluctuation microscopy. *Biophys J* **97**, 2876-85 (2009).
37. Fonseca, J.P. *et al.* In vivo Polycomb kinetics and mitotic chromatin binding distinguish stem cells from differentiated cells. *Genes Dev* **26**, 857-71 (2012).
38. Hathaway, N.A. *et al.* Dynamics and memory of heterochromatin in living cells. *Cell* **149**, 1447-60 (2012).
39. Gaffney, D.J. *et al.* Controls of Nucleosome Positioning in the Human Genome. *PLoS Genet* **8**, e1003036 (2012).
40. Zee, B.M., Levin, R.S., Dimaggio, P.A. & Garcia, B.A. Global turnover of histone post-translational modifications and variants in human cells. *Epigenetics Chromatin* **3**, 22 (2010).

41. Xu, M., Wang, W., Chen, S. & Zhu, B. A model for mitotic inheritance of histone lysine methylation. *EMBO Rep* **13**, 60-7 (2012).
42. Canny, J. A computational approach to edge detection. *IEEE Trans Pattern Anal Mach Intell* **8**, 679-98 (1986).
43. Hanna, J. *et al.* Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature* **462**, 595-U63 (2009).
44. Bhattacharya, D., Talwar, S., Mazumder, A. & Shivashankar, G.V. Spatio-temporal plasticity in chromatin organization in mouse cell differentiation and during *Drosophila* embryogenesis. *Biophys J* **96**, 3832-9 (2009).
45. Hayashi, K., Lopes, S.M.C.d.S., Tang, F. & Surani, M.A. Dynamic Equilibrium and Heterogeneity of Mouse Pluripotent Stem Cells with Distinct Functional and Epigenetic States. *Cell stem cell* **3**, 391-401 (2008).
46. Sparmann, A. & van Lohuizen, M. Polycomb silencers control cell fate, development and cancer. *Nat Rev Cancer* **6**, 846-56 (2006).
47. Yang, C., Wu, J. & Zheng, Y.G. Function of the Active Site Lysine Autoacetylation in Tip60 Catalysis. *PLoS ONE* **7**, e32886 (2012).
48. Yuan, H. *et al.* MYST protein acetyltransferase activity requires active site lysine autoacetylation. *EMBO J* **31**, 58-70 (2012).
49. White, J. & Dalton, S. Cell cycle control of embryonic stem cells. *Stem Cell Rev* **1**, 131-8 (2005).
50. Ruthenburg, A.J., Allis, C.D. & Wysocka, J. Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Molecular Cell* **25**, 15-30 (2007).
51. Zeisberg, M. & Neilson, E.G. Biomarkers for epithelial-mesenchymal transitions. *The Journal of clinical investigation* **119**, 1429 (2009).
52. Yuan, W. *et al.* Dense chromatin activates Polycomb repressive complex 2 to regulate H3 lysine 27 methylation. *Science* **337**, 971-5 (2012).

53. Padinhateeri, R. & Marko, J.F. Nucleosome positioning in a model of active chromatin remodeling enzymes. *Proc Natl Acad Sci U S A* **108**, 7799-803 (2011).
54. Ptashne, M. Epigenetics: Core misconception. *Proceedings of the National Academy of Sciences* **110**, 7101-7103 (2013).

# Chapter 3. Three-layer Regulation Leads to Monoallelic and Diverse Olfactory Receptor Expression

This Chapter is based on a manuscript submitted to PNAS on 02/02/2016

Xiao-Jun Tian<sup>1,†</sup>, Hang Zhang<sup>2,†</sup>, and Jianhua Xing<sup>1,\*</sup>

<sup>1</sup>Department of Computational and Systems Biology, School of Medicine, University of Pittsburgh, Pittsburgh, PA, 15260, USA; <sup>2</sup>Genetics, Bioinformatics, and Computational Biology Program, Virginia Polytechnic Institute and State University, Blacksburg, VA, 24060, USA; <sup>†</sup>The two contributed equally

\*To whom correspondence should be sent: [xing1@pitt.edu](mailto:xing1@pitt.edu)

## 3.1. Abstract

Each mammalian olfactory sensory neuron stochastically expresses only one out of thousands of olfactory receptor alleles and the molecular mechanism for this selection remains as one of the biggest puzzles in neurobiology. Through constructing and analyzing a mathematical model based on extensive experimental observations, we identified an evolutionarily optimized three-layer regulation mechanism that robustly generates single-allele expression. Zonal separation reduces the number of competing alleles. Bifunctional LSD1 and cooperative histone modification dynamics minimize multiple allele epigenetic activation and alleles trapped in incomplete epigenetic activation states. Subsequent allele competition for a limited number of enhancers through cooperative binding serves as final safeguard for single allele expression. The identified design principles demonstrate the importance of molecular cooperativity in selecting and maintaining monoallelic olfactory receptor expression.

## 3.2. Introduction

Olfaction, or the sense of smell, can be essential for the survival and reproduction of an organism. Thus, most species have evolved a highly sensitive olfactory system. A major functional unit of the olfactory system is the main olfactory epithelium where up to millions of olfactory sensory neurons (OSNs) reside. These OSNs sense odor molecules through transmembrane olfactory receptors (ORs), and transmit electric

signals to the brain. OR genes are the largest gene superfamily in vertebrates. There are ~60 OR genes in *Drosophila*, 100-200 in fish, ~ 1,300 (including ~20% pseudogenes, *i.e.*, dysfunctional genes that have lost protein-coding ability) in mouse and ~ 900 (including ~63% pseudogenes) in humans<sup>2-6</sup>.

In their classical studies<sup>5,7-9</sup>, Axel, Buck and coworkers showed that in mammals an individual OSN only stochastically expresses one type of functional ORs, or more precisely one allele of the gene. This monoallelic expression of OR proteins with rare violations has also been shown in other organisms such as zebrafish, and is essential for specificity and sensitivity of olfactory sensing. Expression of more than one type of OR would lead to improper stimulation and wiring of the olfactory system and thus misinterpretation of chemical signals<sup>10</sup>.

The above observations raise one of the most intriguing puzzles in neurobiology that remains elusive after several decades of intensive investigations: how is a single allele selected for activation from a large number of possible OR genes and maintained throughout the lifespan of the neuron? While accumulating evidences and several theoretical studies reveal that a selection is maintained through a feedback loop elicited by expression of the chosen functional OR gene to inhibit further activation of other OR genes<sup>11-18</sup>, proposals on the selection mechanism can be divided into two categories: individual-allele centered selection, and enhancer-regulated selection.

The individual-allele centered proposal emphasizes that properties and dynamics within a single allele lead to the single-allele expression. Indeed, the epigenetic signature of an active OR allele in mice converts from H3K9me3, a covalent histone mark typically repressing gene transcription, to H3K4me3, a mark typically activating gene transcription, and this change is likely conserved in mammals<sup>19</sup>. Similar epigenetic regulation was reported in zebrafish and *Drosophila*<sup>20,21</sup>. Furthermore, disruption of either histone methyltransferases or demethylases leads to violations of the rule of one-allele-activation<sup>15,21,22</sup>. Together with the observation that during OSN differentiation a histone demethylase LSD1 is transiently expressed, an individual-allele centered model suggests a competition among OR alleles for the H3K9me3-to-H3K4me3 transition (see Figure 3.1A)<sup>15</sup>.

In comparison, the enhancer-regulated selection proposal is based on the observations that multiple regulatory genome sequences, *i.e.*, enhancers, can associate with OR promoters<sup>23-25</sup>. Specifically, multiple enhancers bind to the active OR alleles, but not the silenced ones, and form a dense interaction network, possibly mediated by DNA and histone binding proteins such as transcription factor Bptf<sup>26</sup>. Therefore the enhancer-regulated selection model propose that these enhancers act as *cis* or *trans* elements during the OR selection process.

Each of the two proposed mechanisms has experimental supports and complications. The individual-allele epigenetic competition model reveals a natural feedback mechanism that expression of the winning allele causes endoplasmic reticulum stress and expression of Adcy3 enzyme, which then down-regulates LSD1, leading to an epigenetic trap that stabilizes the OR choice.<sup>15</sup> Theoretical analyses demonstrate the feasibility of this model (an epigenetic race followed by a negative feedback) to generate single-allele expression<sup>16,17</sup>. The epigenetic competition model, however, also leaves unanswered questions. First, LSD1 functions counter-intuitively as a bifunctional demethylase. Because it removes both the activating methylation mark from H3K4 and the repressive methylation mark from H3K9, it seems not an efficient way to activate ORs. This bifunctional LSD1 is not taken into account in previous model studies. Furthermore, the mechanism may cause accumulation of "semi-converted" OR genes. A typical OR gene including the regulatory region has ~ 40 nucleosomes as inferred from the transgene experiments<sup>27</sup>, then the stepwise propagation of conversion in epigenetic marks through an OR gene likely takes longer time than other related processes including gene transcription and translation<sup>28</sup>. Consequently by the time LSD1-induced epigenetic conversion is frozen, one would expect a large number of OR genes in a hybrid state, *i.e.*, with some nucleosomes bearing H3K9me3 and others bearing H3K4me3. Such hybrid state is not normally present in stable cell phenotypes, and extended period of existence in this hybrid state is likely detrimental for a cell since histone marks can affect higher-order chromatin structures and gene activities<sup>29</sup>. Thus the hybrid state must relax back to the H3K9me3 dominated state. This relaxation, however, places OSNs in a dilemma, since a sufficiently high level of LSD1 is necessary for removing the H3K4 methylation, but would destabilize the epigenetic state of the activated OR gene as well. On the other hand, for an

enhancer-based model the molecular mechanism for negative feedback is unclear. Furthermore, all OR genes share similar promoter sequence and regulatory elements, which suggests that specific binding to regulatory DNA sequence alone may not be sufficient to regulate OR selection<sup>30</sup>.

The present work aims to reconcile the above two proposals and provides a mechanistic explanation on single-allele OR expression. Our analysis starts with the hypothesis that OSNs have evolved an optimal strategy for olfactory receptor activation as a multi-task design problem: before differentiation, all OR genes should remain transcriptionally silent; one allele is stochastically selected to become transcriptionally active within a biologically relevant period of time (5-10 days for mice) and the error rate of multiple-allele activation should be minimized; the diversity of activated OR genes should be maximized so each gene has approximately equal probability of being activated; if a pseudo gene is selected, it should be recognized and reselected until a functional allele is chosen; after differentiation the selected allele should be kept transcriptionally active while others remains inactive for the life time of an OSN, which is about 100 days for mice.

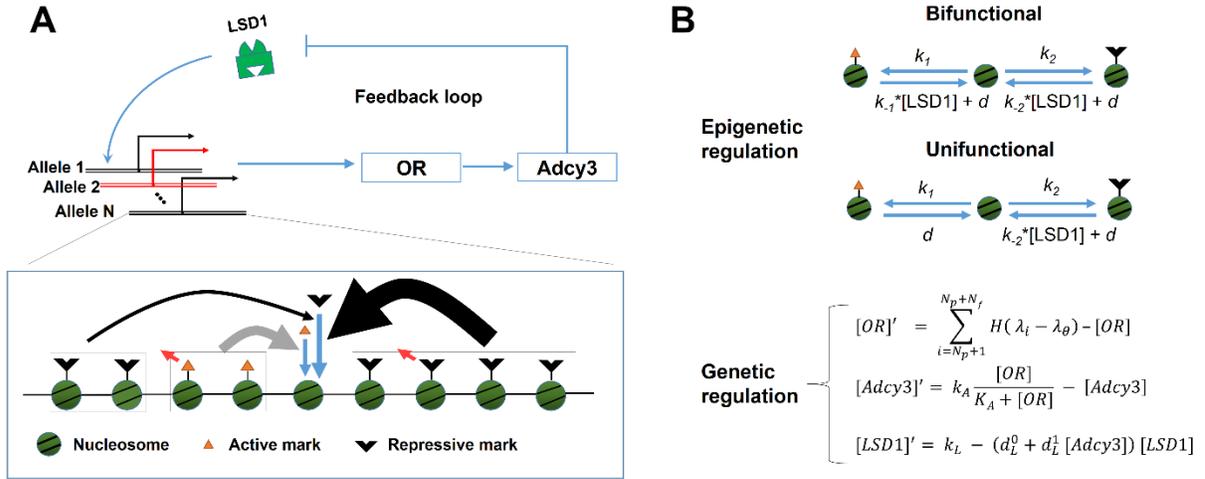
Through mathematical and computational analysis, we demonstrate that OSNs achieve the above multi-task problem through synergistic and sequential selection processes governed by the epigenetic competition and enhancer-regulated transcription, respectively. First, multiple alleles compete for the repressive-to-active epigenetic state transition. Transient expression of the bifunctional LSD1 and cooperativity among nucleosomes lead to an effective two-state transition dynamics, which results in mostly one, occasionally two and rarely more alleles performing such transition, and others remaining predominantly in the repressive epigenetic state throughout the time of competition. Next, while this epigenetic activation is sufficient for OR selection in organisms such as zebrafish, for vertebrates with more sophisticated olfactory sensing, epigenetically active alleles still have to compete for a limited number of enhancers to be transcriptionally active. Cooperative interactions among the enhancers lead to nearly binary OR promoter transcription activity, with negligible probability of having two or more actively expressed alleles simultaneously. Finally the actively transcribed OR induces a negative feedback loop to lower LSD1 level and traps the epigenetic thus transcriptional states of OR alleles.

### 3.3. Results

#### 3.3.1. Mathematical formulation of OR activation

OSNs expressing different subsets of ORs topologically segregate into circumscribed zones. For example, Zone 1 of the mouse main olfactory epithelium contains OSNs that express a subset of 150 OR alleles<sup>3</sup>. Within each zone, the OR alleles in the corresponding subset are expressed with nearly equal probability<sup>8,32,33</sup>. Similar segregated distribution has been found in zebrafish<sup>34</sup>. Zonal segregation reduces the number of OR alleles competing for single allele expression from thousands to hundreds within a zone. We therefore modeled a cell with 100 alleles to recapitulate the selection process from within a single zone of olfactory epithelium.

First we formulated the following mathematical model for the OR activation problem as illustrated in Figure 3.1. Throughout this paper for simplicity of presentation we treat the OR genes within a cell as a number of individual alleles. Each OR allele consists of a linear array of  $N = 41$  nucleosomes, and each nucleosome can bear repressive H3K9 (R), no (E), or active H3K4 (A) methylations. Transition between these states is governed by enzyme concentration dependent rates. Specifically, demethylation  $R \rightarrow E$  and  $A \rightarrow E$  can take place either through stochastic exchange between nucleosome histones and the reservoir of unmarked histones with a turnover rate constant  $d$ , or through demethylation reactions with rates proportional to concentration of the catalyzing enzyme LSD1. We analyzed the real system that bifunctional LSD1 catalyzes both H3K4 and H3K9 demethylation, and a hypothetical system in which unifunctional LSD1 only catalyzes H3K9 demethylation for comparison. To maintain stable collective epigenetic state of an allele, previous studies reveal that the methylation state change on a nucleosome needs to be influenced by the methylation states of other nucleosomes beyond immediate neighbors<sup>31,35</sup>. Therefore we set the methylation rate constants  $k_1$  and  $k_2$  as functions of methylation states of other nucleosomes:  $k_1$  ( $k_2$ ) is promoted by H3K4 (H3K9) methylation in other nucleosomes, and the influence decreases with the nucleosome spatial separation. Details of the model are given in the Method section.



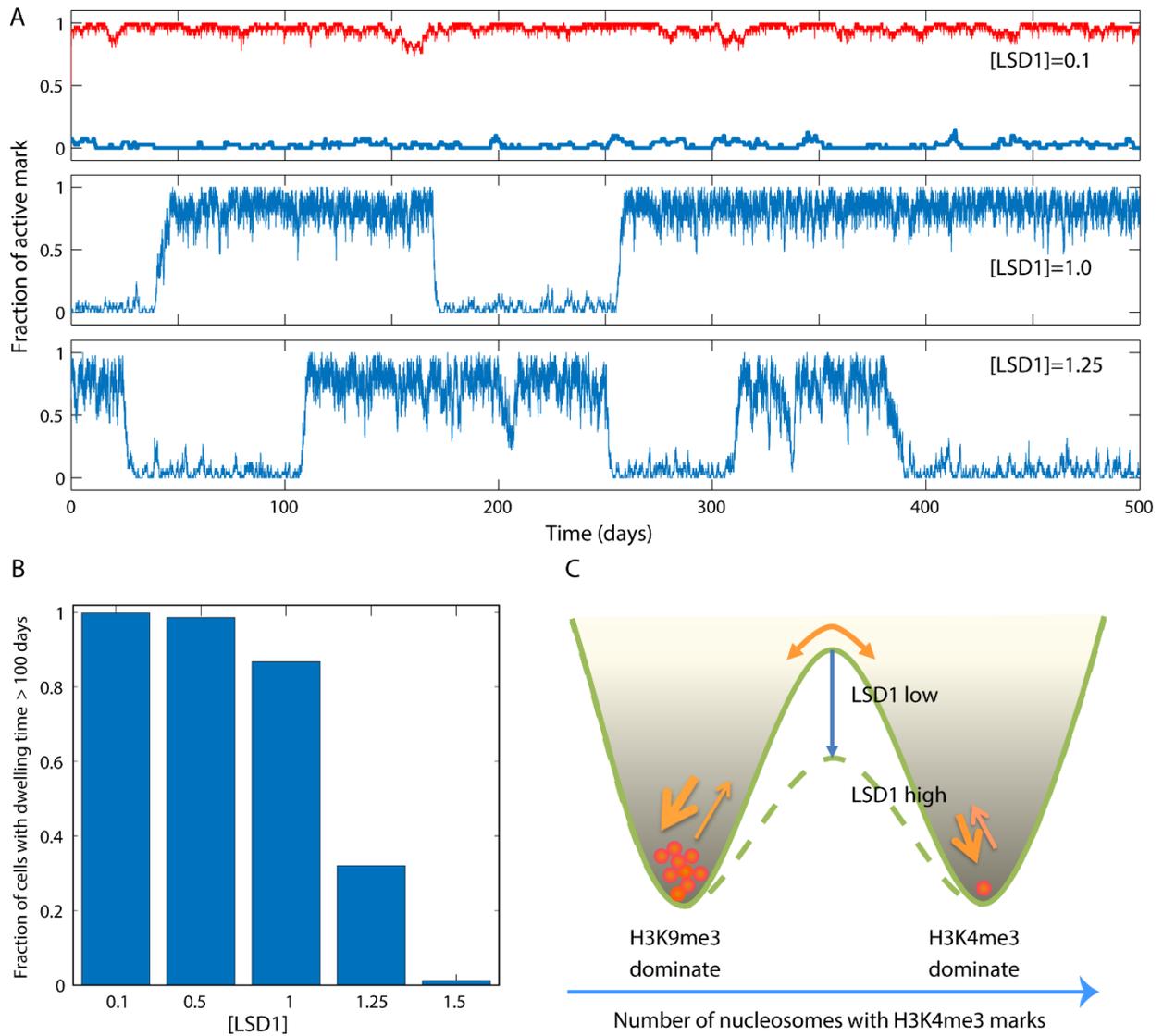
**Figure 3.1: Mathematical model of the experimentally revealed regulatory system of olfactory receptor activation.** (A) Feedback regulated OR allele epigenetic activation. Each OSN contains  $N_p$  ( $= 30$ ) pseudo OR alleles and  $N_f$  ( $= 70$ ) functional OR alleles. Each allele is composed of a linear array of 41 nucleosomes. Each nucleosome bears active, no, or repressive mark, and a mark-bearing nucleosome facilitates an empty nucleosome to add the same mark in a distance dependent manner. Expression of an OR protein elicits a feedback to induce expression of enzyme Adcy3, which removes the demethylase LSD1. (B) The corresponding mathematical formulation. A nucleosome changes its covalent modification state stochastically with the indicated rate constants. The methylation rate constants  $k_1$  and  $k_2$  are influenced by nearby nucleosomes. Protein level changes are simulated by ordinary differentia equations.  $H(x)$  is a Heaviside function which assumes value 0 for  $x < 0$ , and 1 otherwise.  $\lambda_i$  is the fraction of active mark in allele  $i$ , while  $\lambda_\theta$  is the cutoff fraction of nucleosomes with active marks so an allele is regarded as epigenetically activated.

We propagated the nucleosome methylation states using stochastic Gillespie simulations, and simultaneously updated the levels of the expressed OR protein, Adcy3, and LSD1 by solving deterministic

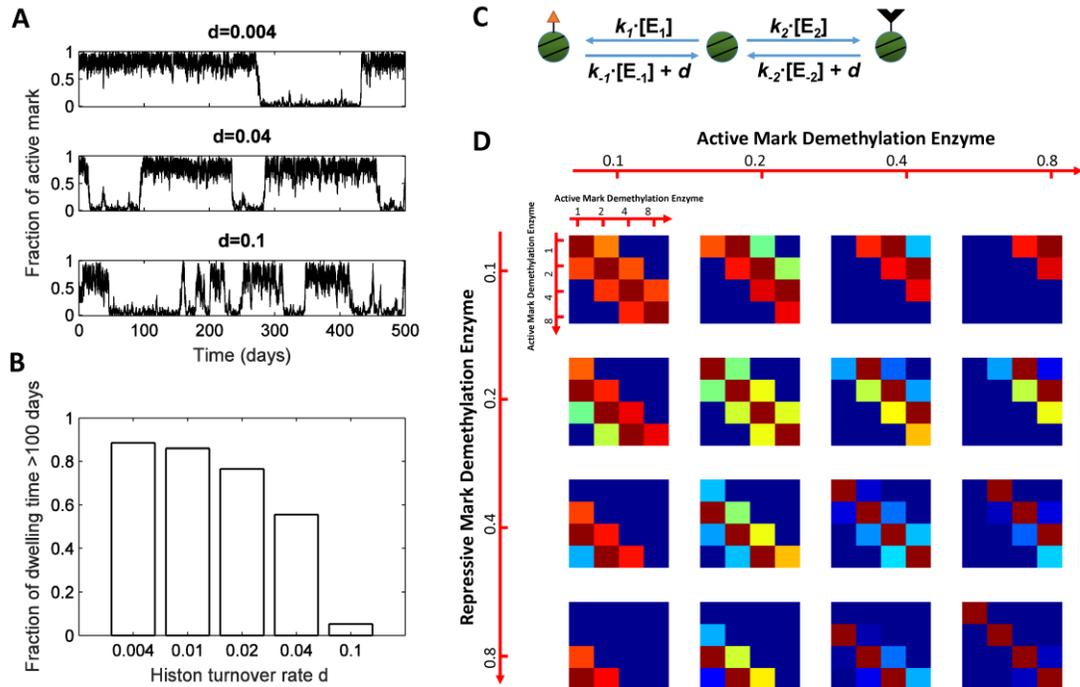
rate equations shown in Figure 3.1B. We assumed that the gene is only transcriptionally active when the fraction of nucleosomes bearing active marks,  $\lambda$ , is larger than a threshold value  $\lambda_\theta$ .

### **3.3.2. Low Noise and lack of demethylases kinetically freeze allele epigenetic state before and after differentiation**

We first examined the model under conditions prior to and after OSN differentiation. As illustrated by the simulated trajectories in Figure 3.2A, cooperation among nucleosomes biases them to have the same histone marks. This cooperation leads to collective epigenetic state dominated by either repressive or active marks, which is consistent with previous studies<sup>31,35</sup>. Increasing LSD1 concentration (Figure 3.2A & B) or the level of system noise due to stochastic histone turnover (Figure 3.2 supplement 1 A-B) facilitates removal of existing methylation marks on a nucleosome and thus destabilizes the collective epigenetic states. Therefore, prior to and after differentiation, maintaining high levels of methyltransferases and low level of demethylases forces an allele to be kinetically trapped at one of the two possible epigenetic states throughout the life time of an OSN, analogous to a system trapped (or frozen) in a double-well shaped potential with a very high barrier (see Figure 3.2C). The above mechanism is confirmed with additional simulations through scanning 256 sets of parameters (Figure 3.2 supplement 1 C-D). In general maintaining stable epigenetic states require that the methylation rates much faster than the demethylation rates, and  $k_1/k_2 \sim k_1/k_2$ , but the latter can be relaxed when either of the demethylation rates is reduced.



**Figure 3.2: Low Noise and demethylation enzyme concentration kinetically freeze allele epigenetic state.** (A) Typical single allele trajectories of the fraction of nucleosomes with active marks under various constant concentrations of LSD1. (B) The fraction of alleles that maintain epigenetic state longer than 100 days under various constant concentrations of LSD1. The result was sampled over 1000 alleles initially in the collective repressive mark dominated state. (C) Analogous double-well potential system with the barrier height inversely related to LSD1 concentration.

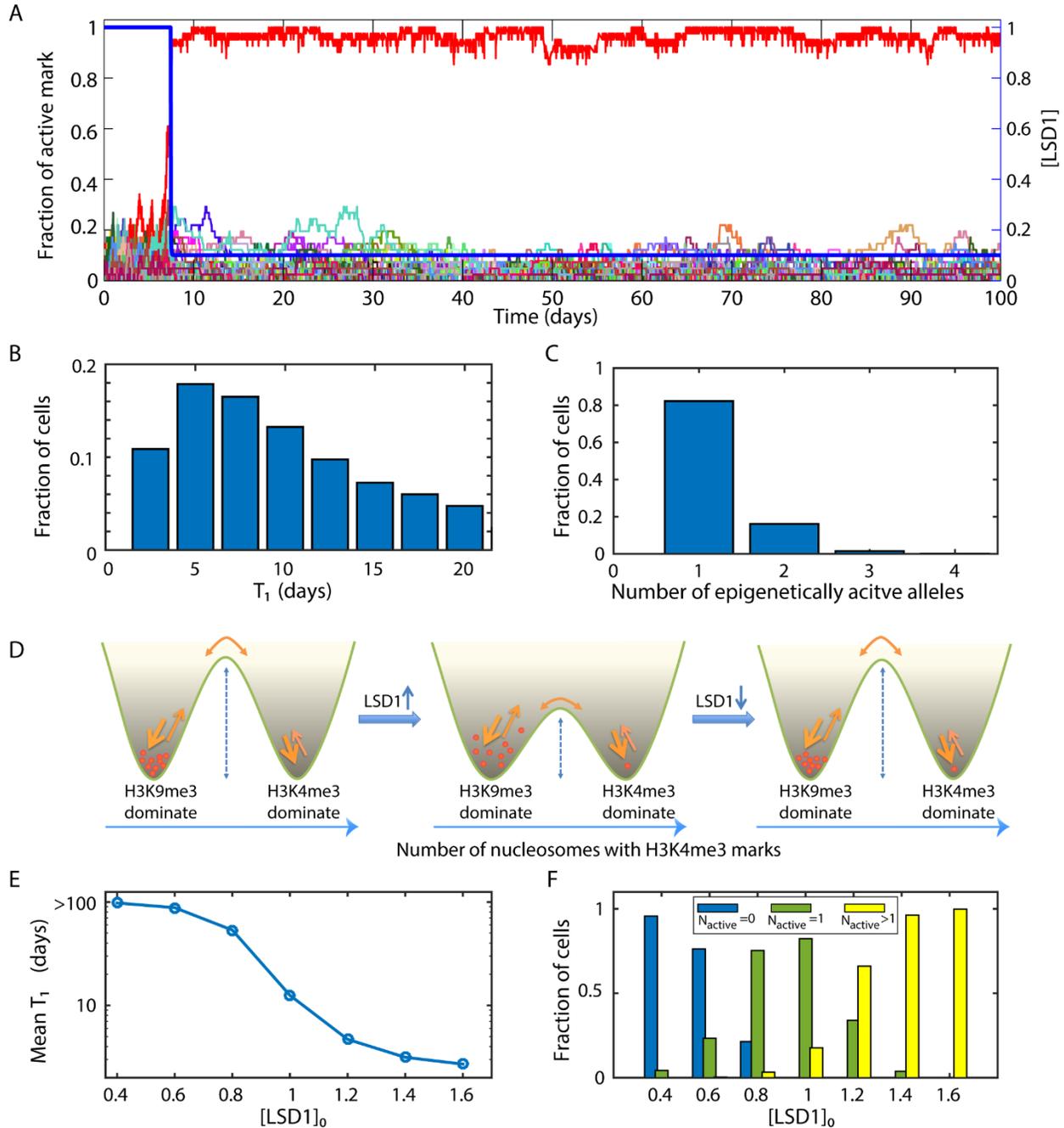


**Figure 3.2. supplemental I: Low histone turnover rate stabilizes the epigenetic state of an allele.** (A) Typical trajectories of the epigenetic state of an allele with different values of the histone turnover rate. (B) The fraction of alleles that maintain epigenetic state within 100 days as a function of the histone turnover rate  $d$ . (C) Expanded model of histone modification reactions on a nucleosome to include methyltransferases and demethylases explicitly. (D) Fraction of alleles that maintain its epigenetic state within 100 days simulated with 256 different sets of active/repressive methylation/demethylation rates. For each parameter set 1000 independent alleles were simulated.

### 3.3.3. Elevation of bifunctional demethylase level leads to a barrier-crossing like dynamics

Next we analyzed OSN differentiation with bifunctional LSD1. As shown in Figure 3.3A, after elevation of the LSD1 concentration at time 0, the OR alleles remain as repressive mark dominated, until one allele becomes active mark dominated, which leads to the corresponding OR expression and subsequent *Adcy3* expression. *Adcy3* down regulates LSD1, then the system maintains at a steady state with one OR allele active and the remaining ones inactive. Notice that the inactive alleles remain H3K9me3 dominated throughout the time. Due to stochasticity of the histone modification process, sampling over 1000 cells

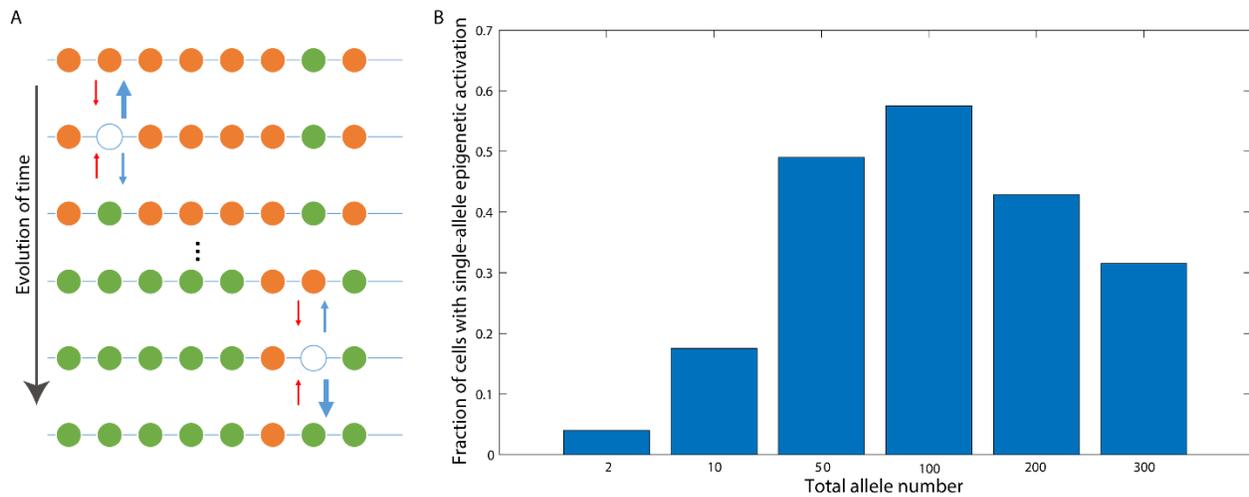
gives a broad distribution of  $T_1$ , the time of having the first allele epigenetically active, ranging from a few to 20 or more days and roughly centered around day 8 (Figure 3.3B). Throughout their lifespan most of the OSNs only have one allele epigenetically activated, while a small fraction has two and rarely 3 alleles epigenetically activated (Figure 3.3C), consistent with the functional requirements and experimental observations.



**Figure 3.3: Bifunctional LSD1 leads to barrier-crossing-like dynamics and ensures mono-allelic epigenetic activation.** (A) Typical trajectories of the fraction of nucleosomes with active marks on one allele for 100 alleles (represented by different colors) within a cell. The temporal change of LSD1 level (blue curve, in relative unit) is also indicated. (B) Distribution of  $T_1$ , the time observing the first epigenetically active allele (75% nucleosomes bearing active marks). Sampled over 1000 cells. (C) Fraction of cells with various numbers of epigenetically active alleles at day 100. (D) The analogous potential system during activation. (E) Dependence of the average of  $T_1$  on the elevated LSD1 level ( $[LSD1]_0$ ) during differentiation. (F) Dependence of the fraction of cells with various numbers of epigenetically active alleles at day 100 on  $[LSD1]_0$ . In all simulations a cell has 100 OR alleles, and at time 0 the LSD1 level is elevated 10 folds from its basal value to simulate the onset of differentiation.

Close examination of the simulated trajectories reveals a simple mechanistic explanation illustrated in Figure 3.3 supplement 1A. Starting with the repressive mark dominated state, transient increase of LSD1 after initiation of OSN differentiation demethylates nucleosomes, and allows resetting of methylation states in the nucleosomes. As a consequence, small patches of H3K4me3 nucleosomes may form, but are flanked by extended regions of H3K9me3 nucleosomes. Such H3K4me3 patches are unlikely to expand because of the cooperativity of methylation between nucleosomes and the dominance of H3K9me3 marks at the current stage. Nevertheless, when an H3K4me3 patch reaches a critical size -- as a rare event, it is able to propagate spontaneously and generate an epigenetic conversion of the OR gene into the H3K4me3 dominated state. That is, LSD1 increase resembles lowering the transition barrier between the double-well shaped potential shown in the previous section, and allows rare transition to happen (Figure 3.3D). Once one allele converts to the H3K4me3 dominated state, and triggers the negative feedback loop to remove LSD1, the system is kinetically trapped again with high transition barrier. The converted allele is kept active with H3K4me3 marks, while the remaining alleles bear repressive H3K9me3 marks. A prominent feature of this barrier-crossing-like dynamics is that throughout the process the probability of having an allele with hybrid pattern of epigenetic marks is low, and most alleles only fluctuate around the H3K9me3 dominated state.

Based on the above analogy to a double-well potential, we reasoned that increasing the LSD1 concentration facilitates epigenetic state transitions. Indeed simulation results show that upon increasing the LSD1 concentration,  $\langle T_I \rangle$ , the average of  $T_I$ , decreases (Figure 3.3E), but the fraction of cells with multi-allele activation increases (Figure 3.3F). Lyons et al. also observed fewer mature OSNs in mice with reduced LSD1<sup>15</sup>, as what predicted in Figure 3.3E. Therefore a given organism may have evolved an optimal LSD1 concentration to compromise the requirements of single-allele activation and efficient OSN differentiation.

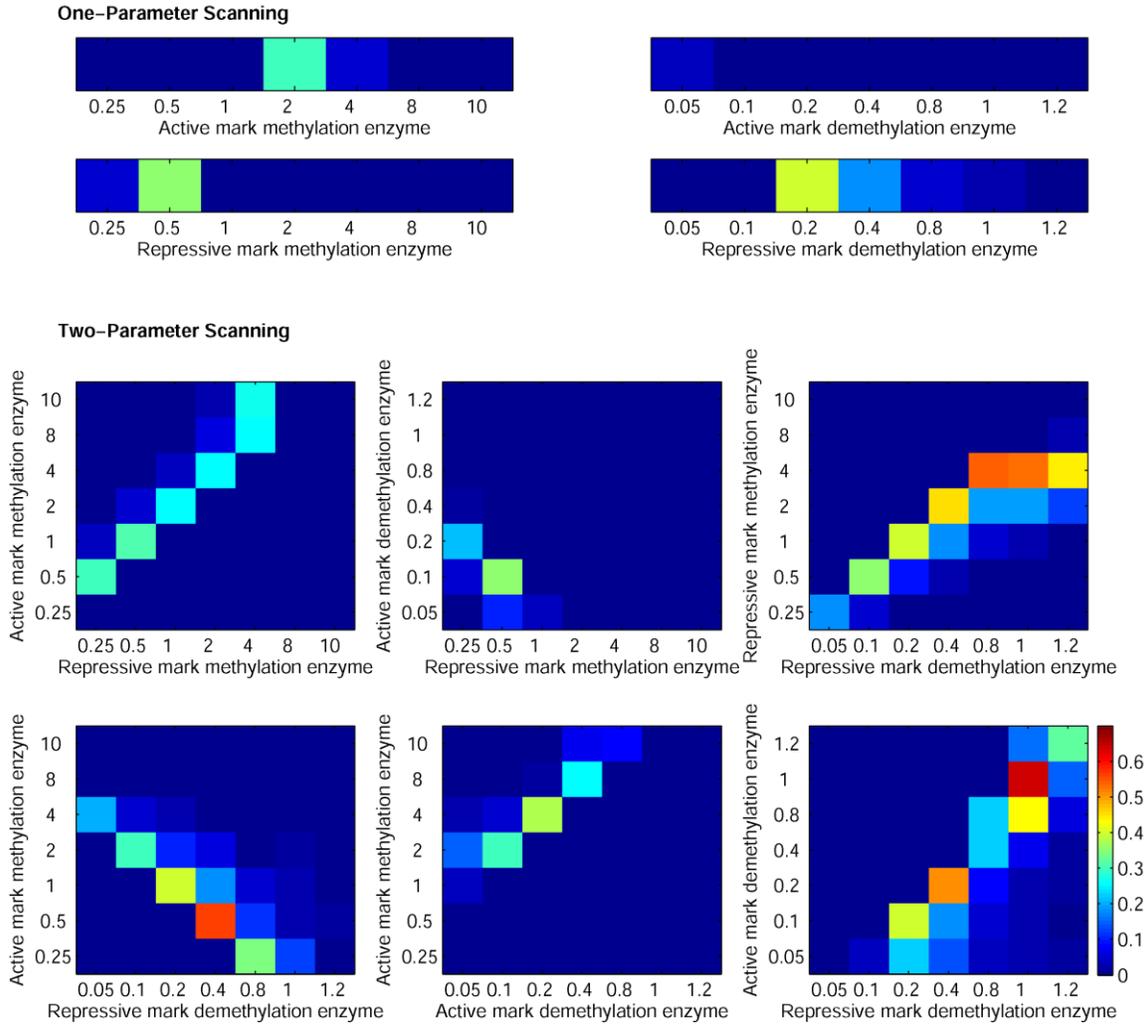


**Figure 3.3 supplement 1. Additional results of bifunctional LSD1 elicited epigenetic conversion.** (A) Schematic illustration of a conversion process elicited by bifunctional LSD1. Orange, green, and white balls represent nucleosomes bearing repressive, active, and no marks, respectively. The widths of arrows represent the relative values of corresponding rates. (B) The monoallelic ratio changes non-monotonically over the number of the allele. Except for the total number of alleles, the simulations are performed in the same way as those in Figure 3.3D.

Next we asked how the number of permitted alleles affects the ratio of single allele activation (Figure 3.3 supplement 1B). The ratio first increases since a cell with more alleles has higher probability to have at least one allele epigenetically activated during the differentiation period. Then it decreases after a peak value since the probability of having more than one allele activated also increases with the number of alleles

per cell. While the exact position of the peak depends on model parameters, the model results predict that the number of OR genes within a zone is under selective pressure.

The feedback regulates a bifunctional demethylase, LSD1. It seems both counter-intuitive and inefficient, since the enzyme moves both the repressive and active methylation marks, with the latter being what added to an active allele. Theoretically the feedback could act on any one or any combination of the four groups of enzymatic reactions (Figure 3.1B). Therefore we simulated all the 10 cases that the feedback regulates one or two of the reaction rates. All these cases have the same set of parameters for cells after activating the feedback, and they differ only on value(s) of one or two rates prior to feedback taking effect. By scanning each combination of parameter pairs over a  $7 \times 7$  grid and performing 500 independent simulations for each parameter set, indeed the case of negative feedback on both of the two demethylation rate constants, *i.e.*, a bifunctional LSD1, leads to the highest monoallelic activation ratio (Figure Supplemental 2). A less robust scheme requires regulating the H3K9 demethylase and H3K4 methyltransferase oppositely at the same time. Both of these schemes modulate the effective transition “barrier” without necessarily changing the relative stability of the two collective epigenetic states.

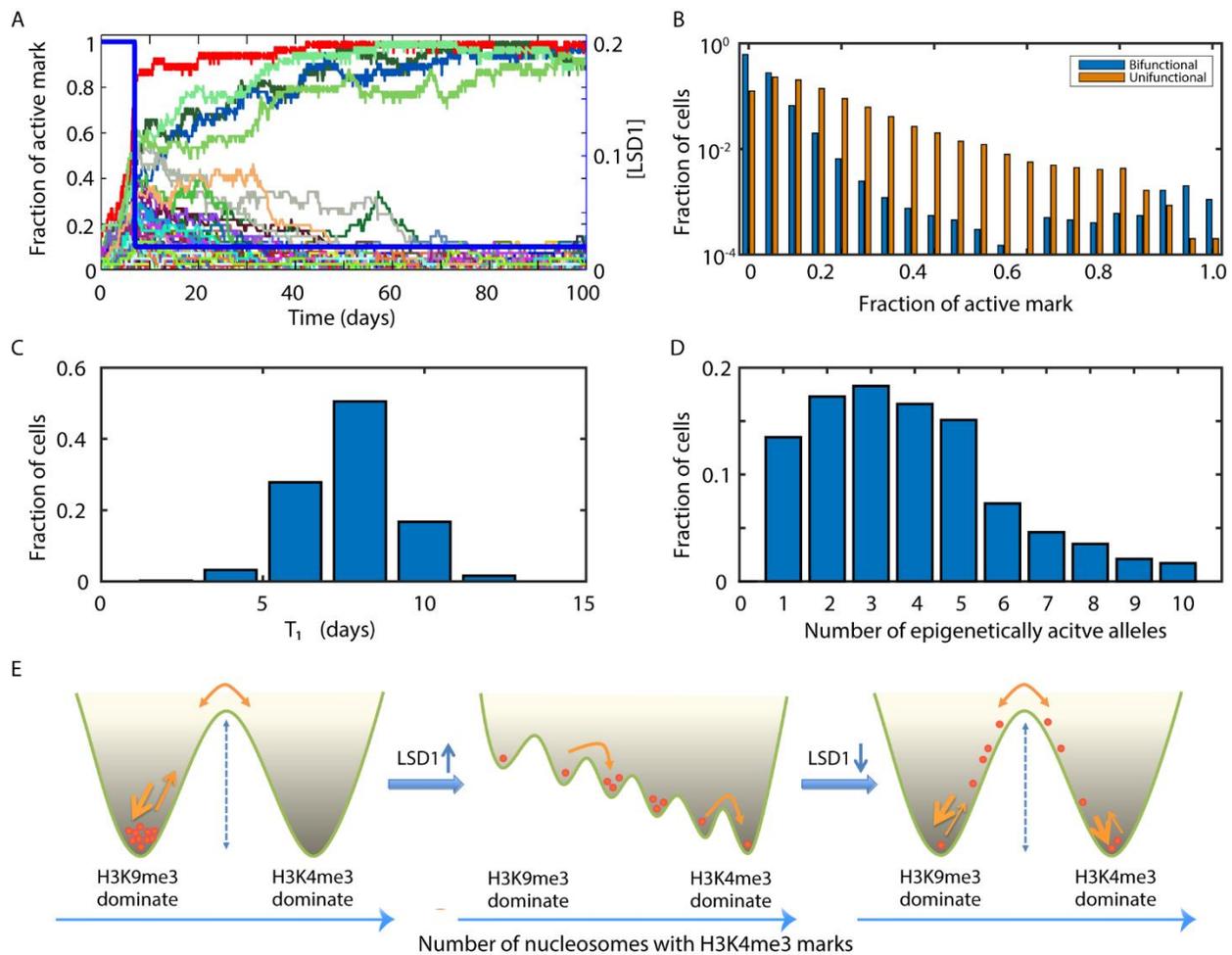


**Figure 3.3 supplement 2.** *Comparative studies on all possible one-rate and two-rate feedback regulation schemes demonstrate that it is optimal to regulate both two demethylation reactions. For each data point, the fraction of cells with one epigenetically active allele at day 20 is calculated from 500 independent simulations.*

### 3.3.4. Elevation of unfunctional demethylase level leads to ratchet-like dynamics

To further understand the critical role of the bifunctional LSD1, we examined one of the above hypothetical cases that unfunctional LSD1 only catalyzes H3K9 demethylation. The dynamics is completely different. As revealed by a typical simulated trajectory (Figure 3.4A), starting at time 0 all the OR alleles begin to remove repressive marks and gain active marks, while some are faster than the others. After one epigenetically active OR allele activates the feedback loop and lowers the LSD1 concentration,

some OR alleles return to the epigenetically repressive state, and a number of others propagate further to the epigenetically active state. A significant fraction of alleles are present in some hybrid epigenetic state during the activation process, and subsequently retard in the hybrid state due to slow demethylation (Figure 3.4A & B). In contrast, when LSD1 bifunctions, the probability of catching an allele in an intermediate state is very low (Figure 3.4B) by the time LSD1 depletes, leading to efficient relaxation. Therefore, bifunctional LSD1 avoids hybrid state by maintaining one single large barrier between the repressive mark dominated and active mark dominated states. While making this comparison, we have chosen model parameters so the distribution of  $T_1$  with a unifunctional LSD1 also centers roughly around day 8 (Figure 3.4C). Then sampling over 1000 cells shows that ~ 87% of the cells have two or more alleles activated at the end of differentiation (Figure 3.4D). These simulation results demonstrate that the hypothetical unifunctional LSD1 scheme is not as robust as the bifunctional enzyme scheme on achieving single allele epigenetic activation.



**Figure 3.4: Unifunctional LSD1 leads to ratchet-like dynamics and cannot ensure mono-allelic epigenetic activation.** (A) Typical trajectories of the fraction of nucleosomes with active marks on one allele for 100 alleles within a cell. The temporal change of LSD1 level is also indicated. (B) Distribution of the fraction of nucleosomes with active marks on day 8 with bifunctional and unfunctional LSD1. Sampled over 1000 cells. (C) Distribution of  $T_1$ . Sampled over 1000 cells. (D) Fraction of cells with various numbers of epigenetically active alleles at day 100. (E) The analogous potential system during activation.

Again the above dynamics of allele selection has a simple mechanistic explanation (Figure 3.4 supplement 1A). Let us start with an allele originally in the H3K9me3 dominated state. Increased LSD1 level promotes empty nucleosomes. Following the same argument in the previous section, an empty nucleosome still has higher probability to add H3K9me3 compared to H3K4me3. However, once an

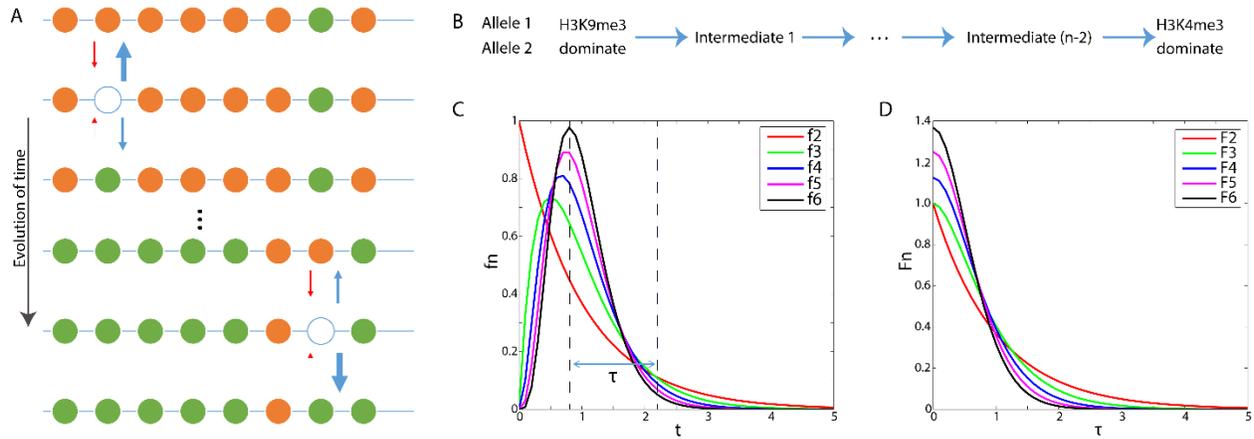
H3K4me3 is added, the removal rate is very low. Consequently, the dynamics is ratcheted<sup>36</sup> towards increasing the number of nucleosomes with H3K4me3, as illustrated by Figure 3.4E. A prominent feature of this ratchet-like dynamics is that every allele has its total number of H3K4me3 nucleosomes increasing with time, with a leading allele followed by other alleles on the way of converting their nucleosome marks, thus intuitively it is difficult to ensure single allele epigenetic activation.

### **3.3.5. Mathematically controlled comparison explains why bifunctional LSD1 improves the ratio of single allele activation.**

To understand why the effective two-state barrier-crossing dynamics is advantageous over the multi-state ratchet-like dynamics on generating single allele activation, we performed further mathematical analysis based on the following reasoning. In the OR system a number of alleles convert their epigenetic state independently and stochastically under an elevated LSD1 concentration. Let us denote the activation time separation between the first two converted alleles as  $\tau$ . Then from an engineering perspective, a better design to achieve single-allele activation is the one with a larger  $\tau$ , which means that the two activation events are better separated temporally, and thus more time for the first allele to elicit the feedback loop and prevent activation of another allele.

Therefore we performed mathematically controlled comparison among a set of simple models shown in Figure 3.4 supplement 1B. Consider two alleles transiting independently from the repressive mark dominated state to the active mark dominated state through various numbers of intermediate states, but with the same mean first arrival time. Figure 3.4 supplement 1C shows that the two-state model has an exponentially shaped first-arrival-time distribution  $f_2$ , while those with  $(n - 2)$  intermediate states have peaked ones that at large  $t$  decrease faster with increasing  $n$ . One can randomly draw two points  $p_{11}$  and  $p_{12}$  from a distribution, corresponding to the stochastic activation events of the two independent alleles. Clearly the temporal separation of the two points,  $\tau$ , is likely to be larger if they are drawn from a broader  $f$

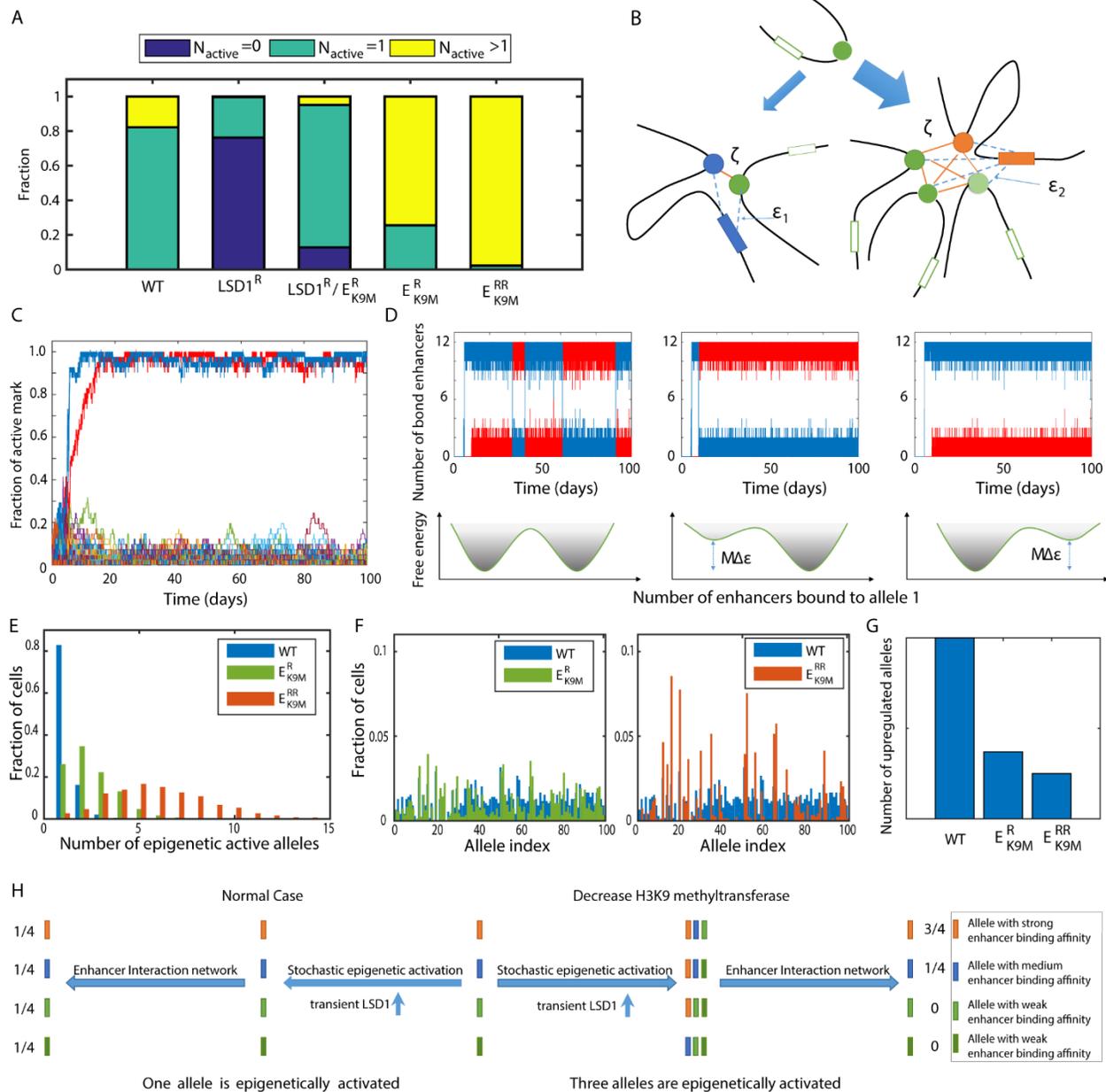
corresponding to smaller  $n$ . Indeed Figure 3.4 supplement 1D shows that the distribution of  $\tau$  has longer tail for smaller  $n$ . That is, a design with the two-state dynamics is better than that with the multi-state dynamics.



**Figure 3.4 supplement 1. Simple mathematical analysis reveals the mechanistic advantage of bidirectional over unidirectional demethylation enzyme.** (A) Schematics of a conversion process elicited by unifunctional LSD1. Same notations as in Figure 3.3 supplement 1A. (B) Minimal effective Markovian transition models for an OR allele changing from H3K9me3 dominate state to H3K4me3 dominate state with no ( $n = 2$ , corresponding to the barrier-crossing dynamics with the bifunctional LSD1), and various number ( $n > 2$ , corresponding to the ratchet-like dynamics with the unifunctional LSD1) of intermediate states. Here we compare two independent alleles that undergo the transition. (C) The first-arrival time ( $t$ ) distribution  $f_n$  of a single allele transiting from H3K9me3 dominate state to H3K4me3 dominate state as a function of the overall state number  $n$ . (D) The distribution ( $F_n$ ) of first-arrival time separation ( $\tau$ ) between two kinetically independent alleles as a function of the overall state number  $n$ . From engineering design perspective a larger  $\tau$  is desirable since it gives the system more response time to elicit the feedback after the first allele becomes epigenetically active and prevents the second allele from making the transition.

### **3.3.6. Epigenetic competition model predicts zebrafish but not mouse experiments on inhibiting methyltransferases/demethylases**

In the illustrative double-well potential shown in Figure 3.3D, lifting the left well allows easier transition to the right well thus higher probability of multiple allele activation, while elevating the barrier height leads to an opposite effect. Experimentally, shallowing of the left well can be realized by reducing the enzymatic activity of H3K9 methyltransferases, G9a and GLP. Similarly decreasing LSD1 concentration corresponds to increasing the barrier height. The simulation results in Figure 3.5A show that reducing the LSD1 concentration ( $LSD1^R$ ) impedes OR activation, which can be partially restored by decreasing the enzymatic activities of H3K9 methyltransferases ( $LSD1^R/E_{K9M}^R$ ), consistent with that observed in mice<sup>22</sup>. Furthermore, partially inhibiting the enzymatic activities of G9a/GLP ( $E_{K9M}^R$ ) leads to increased number of cells coexpressing multiple ORs, which is confirmed in zebrafish<sup>21</sup>.



**Figure 3.5: Competition of cooperatively bound enhancers further reduces co-expression of multi-allele ORs.** (A) Predicted fractions of cells with various numbers of epigenetic active alleles under different conditions. WT: wide type.  $LSD1^R$ :  $LSD1$  level reduced.  $E_{K9M}^R/E_{K9M}^{RR}$ :  $H3K9$  methyltransferase level reduced and further reduced. (B) Model of alleles competing for  $M$  enhancers. (C) Simulated allele trajectories of one cell with two epigenetically active alleles. (D) Simulated dynamics of enhancers binding to two epigenetic active alleles corresponding to the cell in panel C with the same (left) or different (by  $\Delta\epsilon = \pm 0.5 k_B T$ , middle and right) binding affinity.

Also shown are schematic free energy profiles. (E) Simulated distribution of 1000 cells with various numbers of epigenetically active alleles under  $E_{K9M}^R, E_{K9M}^{RR}$  and WT on day 100. (F) Fractions of overall protein expression of each allele simulated with a population of 1000 cells under  $E_{K9M}^R$  and  $E_{K9M}^{RR}$  comparing to those with WT. (G) The number of transcriptionally upregulated alleles under  $E_{K9M}^R, E_{K9M}^{RR}$  and in WT. (H) Schematic illustration on the mechanism of reduced OR expression diversity with  $E_{K9M}^R$  and  $E_{K9M}^{RR}$  compared to that in WT.

When the left well in Figure 3.3D lifts further the system resembles more and more that in Figure 3.4E, and the epigenetic activation process evolves from a barrier-crossing-like dynamics to a ratchet-like dynamics. The simulated results in Figure 3.5 supplement 1A & B indeed predict that in contrast to that in wild type (WT), with further reduced level of H3K9 methyltransferases ( $E_{K9M}^{RR}$ ) a majority of the OR alleles assume a hybrid methylation pattern during the differentiation process, and thus significant increase of multi-OR activation (Figure 3.5A). However, experimental studies using G9a/GLP double knockout (dKO) mice only observed elevated but still rare multi-OR coexpression compared to the WT control in mice<sup>22</sup>. Therefore the epigenetic conversion mechanism is insufficient to explain the experimental results.

### **3.3.7. Competition of cooperatively bound enhancers further reduces coexpression of multi-allele ORs.**

To explain the G9a/GLP dKO mouse experiment, we generalized the model based on recent studies<sup>22,26</sup>. We hypothesize that for terrestrial vertebrates such as humans and mice, active expression of an OR allele requires both the gene bearing active epigenetic marks (H3K4me3) and co-localization of a sufficient number of enhancers to the allele. In the following, we elaborate on how we model this process.

Suppose  $M$  enhancers are available for an OR genomic cluster with  $L$  OR alleles (see Figure 3.5B). Each enhancer can bind to the epigenetically active  $l$ -th OR allele with a free energy of binding  $\varepsilon_l$ , and can interact with any other enhancer bound to the same allele with energy  $\zeta$ . Enhancer binding to alleles with repressive marks is weak and can be neglected<sup>26</sup>. Notice that enhancer binding to active alleles is

cooperative: when two or more epigenetically active alleles compete for the enhancers, an enhancer prefers to binding to the one that already has more enhancers bound since it can form more enhancer-enhancer interactions. Consequently, enhancers collectively bind to and transcriptionally activate one allele at a given time, and switching from binding to one allele to another one is rare since it requires breaking many interactions. While the exact value of  $M$  is currently not known, the experimental observation that ectopically introduced multiple copies of a specific H enhancer increase the probability of multi-OR coexpression (Lomvardas et al., 2006) supports our assumption: alleles have to compete for an unsaturating number of enhancers in WT cells; due to cooperative binding the enhancers first saturate one allele and then extra enhancers bind to others.

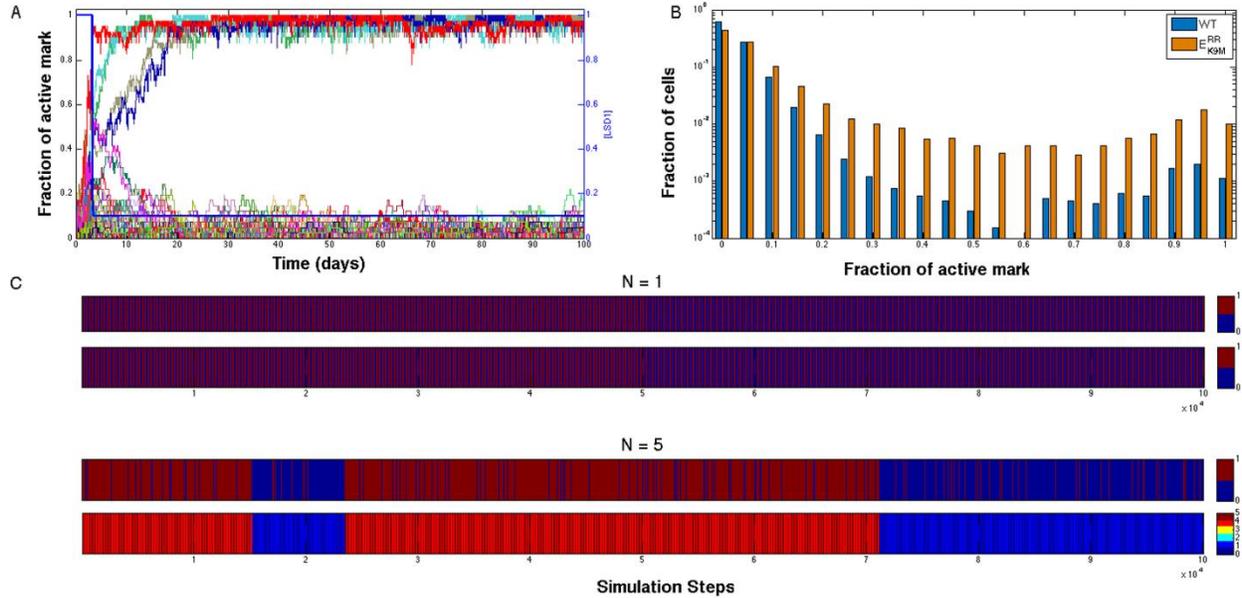
We then performed Gillespie simulations for the enhancer-allele binding/unbinding dynamics. Figure 3.5C&D give an example of a cell with two alleles becoming epigenetically active (Figure 3.5C), but only one of them is transcriptionally active at a given time (Figure 3.5D). If the enhancers bind to the two alleles with equal strength, i.e.,  $\zeta$  and  $\varepsilon_l$  assume the same values for different alleles and enhancers, the enhancers jump stochastically and collectively between the two alleles, showing a two-state dynamics alike a particle moving in a symmetric double-well potential (left panel of Figure 3.5D). The frequency of transitions depends on the actual binding strength and the number of enhancers. However, it is likely that the values of  $\zeta$  and  $\varepsilon_l$  are slightly allele-dependent. Then cooperative enhancer binding can amplify this difference by many folds. For example, suppose that there exists a free energy difference of enhancer-allele binding  $\Delta\varepsilon = \varepsilon_1 - \varepsilon_2$  between allele 1 and allele 2. Then the free energy difference between allele 1 bound with  $M$  enhancers and allele 2 bound with  $M$  enhancers is  $M\Delta\varepsilon$ , which can be significant due to the factor  $M$ . So the allele with stronger enhancer binding dominates transcriptionally alike a particle moving in an asymmetric double-well potential (middle and right panels of Figure 3.5D).

The above model reaches a surprising prediction on the OR expression pattern when the level of H3K9 methyltransferases is reduced. Compared to WT cells, the cells with  $E_{K9M}^R$  tend to have more OR alleles being epigenetic active (Figure 3.5E), as expected. However, except for a small group of OR genes becoming transcriptionally upregulated, most of them instead show decreased expression compared to those

in the WT (Figure 3.5F). Further reduction of the enzyme level ( $E_{K9M}^{RR}$ ) causes fewer OR alleles to be expressed, but each with higher expression level (Figure 3.5F & G). Seemingly counterintuitive, this prediction is what observed experimentally<sup>22</sup>.

The reduced diversity in Figure 3.5F has a simple mechanistic explanation. For illustration purpose let us consider a toy system in which  $L$  ( $= 4$ ) OR alleles exist in a zone, and these alleles have strong (allele 1), medium (allele 2), and weak (alleles 3 and 4) binding strength to the enhancers, respectively (Figure 3.5H). Existing experimental evidences suggest that the epigenetic activation step is stochastic and each allele has roughly equal probability  $1/L$  to be chosen. For WT OSNs, most cells have only one epigenetically active allele, and the allele becomes transcriptionally active as well. Therefore the overall transcriptional probability of each allele in the zone is  $\sim 1/4$ . On the other hand, with the H3K9 methyltransferase level reduced ( $E_{K9M}^R$  and  $E_{K9M}^{RR}$ , or G9a KO and G9a/GLP dKO experimentally), an OSN may have multiple epigenetically active OR alleles. For simplicity of argument let us assume that in a cell three alleles compete for enhancers. Since each allele has the same probability becoming epigenetically active, there are 4 possible combinations with equal probability, (123), (124), (134) and (234). As an allele with stronger enhancer binding dominates transcription, one expects that the first 3 combinations mainly express allele 1, and the last one expresses allele 2. That is, the expression of allele 1 is upregulated while that of alleles 3/4 are down regulated. Similarly, with more epigenetically active alleles coexisting in individual OSNs, more OSNs are likely to have the strongest alleles active and express them; fewer OSNs have the chance to express the weaker alleles. Consequently, the OR diversity in the OSN population becomes further reduced. In the above simulations we assumed that only the number of enhancers bound to an allele affects its transcription. It is possible that enhancers have certain OR gene specificity<sup>12,23,25</sup>. Therefore we considered the alternative possibility that only one of the binding enhancers, say enhancer 1, is necessary for activating a given OR gene. Compared to the case with only enhancer 1 (Figure 3.5 supplement 1C upper panel), with other enhancers being present enhancer 1 shows increased dwelling time of binding to allele 1 and this binding correlates with the overall collective binding state of enhancers (Figure 3.5 supplement 1C lower

panel). That is, the presence of other enhancers stabilizes the binding of the enhancer who actually affects the allele transcription, and the above results discussed in this section still hold in this case.



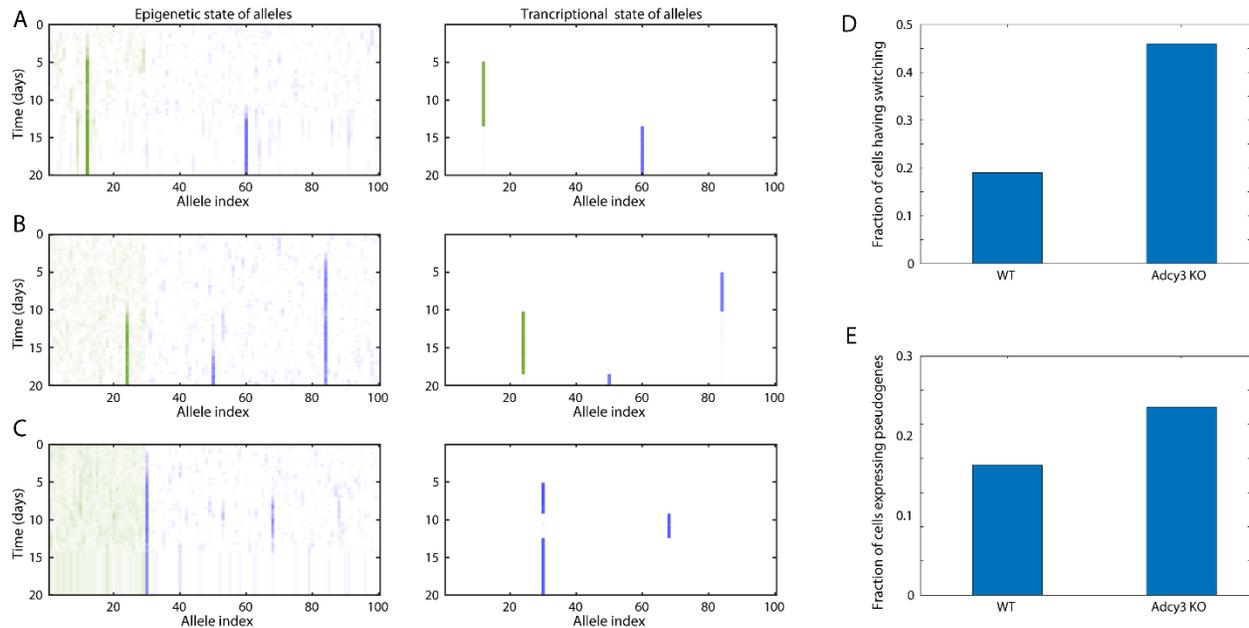
**Figure 3.5 supplement 1. Enhancer competition assures transcriptional activation of single allele.**

(A) Typical single-allele trajectories of the fraction of nucleosomes with active marks for 100 allele within an  $E_{K9M}^R$  cell. (B) The distribution of the fraction of nucleosomes with active marks on day 8 averaged over 1000 cells. (C) Auxiliary enhancers stabilize binding of a specific enhancer to an allele. For each result with  $M$  enhancers, the upper one shows the trajectory of enhancer 1, and the lower one shows the corresponding number of enhancers bound to allele 1. The time is given by the number of Gillespie simulation steps. In these simulations,  $\varepsilon_1 = \varepsilon_2 = -1 \text{ k}_B T$ ,  $\zeta = -3 \text{ k}_B T$ . Similar results were obtained with broad range of parameter values (e.g.,  $\varepsilon_1$  and  $\varepsilon_2$  assuming values from  $-2$  to  $-0.5 \text{ k}_B T$  and  $\zeta$  assuming values from  $-3$  to  $-0.5 \text{ k}_B T$ ), and more enhancers involved.

### 3.3.8. Model studies identify multiple schemes of OR transcription switching

The trajectories in Figure 3.5D reveal that an OSN cell occasionally switches off an active OR allele and chooses another one. This switching phenomenon has been widely reported in the literature<sup>33</sup>, thus we examined these trajectories in detail.

For WT OSNs, most switching takes place as a pseudo gene allele gets epigenetically and transcriptionally activated first, then a functional allele transcriptionally switches on after certain time (Figure 3.6A). This is because the products of a pseudo gene fail to elicit the *Adcy3* mediated feedback loop to reduce the LSD1 level, and permit another allele to be epigenetically activated. Figure 3.6B shows another type of switch also found in simulations with WT OSNs. A functional allele is activated first and then switches to a pseudogene, the cell reenters to the selection process until an alternative functional is activated. In both two schemes an allele remains epigenetically active even after switching off transcriptionally.



**Figure 3.6: Predicted OR expression switching schemes.** Typical switching examples: active pseudo gene switches to intact gene (A), active intact gene switches to pseudo gene and then switches to intact gene (B), and intact active gene switch off itself (C). (D) Simulated switching

*frequency under WT and Adcy3 KO conditions. (E) Simulated fraction of cells expressing pseudo genes under WT and Adcy3 KO conditions.*

In Adcy3 KO mice the feedback loop is disrupted and the simulation results in Figure 3.6C reveal a new switching pattern. First a transcriptional switch takes place between two alleles. Then the newly activated allele switches off epigenetically and thus transcriptional, and the original allele switches back to be transcriptionally active. Mechanistically the sustained high level of LSD1 in Adcy3 KO cells leads to collective removal of H3K4 methylation from the activated allele.

Not surprisingly, the switching frequency increases in Adcy3 KO OSNs compared to that in WT OSNs (Figure 3.6D) since more cells have multiple epigenetically active alleles. Furthermore, the fraction of cells expressing pseudo ORs increases while that expressing functional ORs decreases in Adcy3 KO simulations (Figure 3.6E), consistent with the experimental results<sup>15</sup>. Mechanistically in Adcy3 KO system transcription of a functional allele does not inhibit further epigenetic activation of pseudo gene alleles, and the latter then competes with the former for transcription.

Therefore the simulations suggest two possible mechanisms for OR allele transcriptional switching. One is that an allele converts from the active H3K4me3 epigenetic state back to the repressive H3K9me3 state. Experimental testing of this mechanism requires monitoring the histone modification state of one allele over time. Another is that the enhancers cooperatively change their binding from one allele to another one, with both being epigenetically active. The present model predicts that the genes showing upregulated expression in the G9a/GLP dKO mice, such as Olfr231, have slighter stronger interactions with the enhancers than the remaining genes do. Then an experimentally testable prediction is that in normal mice, OSNs that express one of these genes should have lower frequency of switching than those cells express other genes in the same zone do.

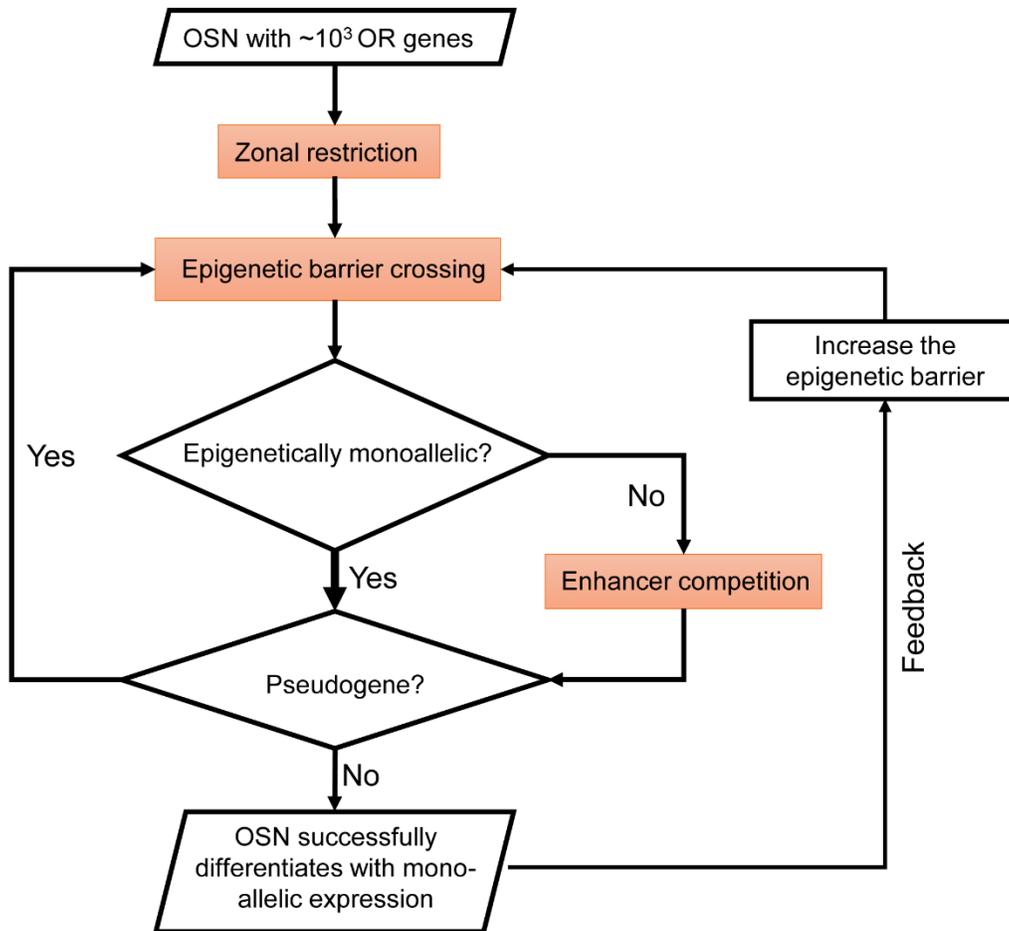
### **3.4. Discussion**

Single allele activation in olfactory sensory neurons is a multi-decade-long puzzle in neurobiology. Recently several mathematical models have been formulated to examine various proposed mechanisms for

explaining the phenomenon<sup>16-18</sup>. Compared to these existing modeling studies, the present model is based on some key experimental observations available only recently. The model, while coarse-grained, has every of its components corresponding directly to an experimentally measurable quantity, which makes comparison to experimental results and prediction test transparent. More importantly, there are some essential differences between the present model and the existing ones both mechanistically and conceptually.

### **3.4.1. A sequential three-layer regulation mechanism controls single allele activation.**

Through integrating a large body of existing experimental studies, our theoretical studies suggest that single allele activation is achieved through a series of selection processes functioning synergistically (Figure 3.7). A subset of the alleles is selected by the zonal segregation. Then they are randomly chosen to be epigenetically activated through transient elevation of bifunctional LSD1. Most of the cells only have one epigenetically active and thus transcriptional active allele. If more than one allele are epigenetically activated, they compete for the enhancers to be transcriptionally active, resulting in only one epigenetically and transcriptional active allele. If the activated allele is not a pseudo gene, it triggers the feedback to prevent further epigenetic state change. Therefore, this coordinated three-layer regulation mechanism faithfully assures that only one OR allele is expressed in one OSN. All other existing models consider only epigenetic switching or competition for certain regulatory elements. For example, the epigenetic switching model of Alsing et al.<sup>17</sup> states that “singular gene selection does not require transient mechanisms, enhancer elements or transcription factors to separate choice from maintenance”.



*Figure 3.7: The three-layer mechanism ensures mono-allele activation of OR genes.*

### **3.4.2. The OR selection process is optimized to satisfy hierarchical multi-task and even opposing requirements.**

All existing models recognize that OSNs need to achieve monoallelic OR expression. Based on extensive experimental results, the present model further argues that the regulation system is optimized to achieve multi-task functional requirements. Some of these requirements, such as maximum OR diversity and minimal hybrid state, while not recognized in earlier studies, naturally reconcile many seemingly counter-intuitive and contradictory observations and ideas in the field.

It might seem redundant to have both epigenetic activation and enhancer competition to achieve monoallelic expression. Epigenetic activation selects OR alleles with approximately equal probability, but leads to a small percentage of cells having multiple epigenetically active alleles. On the other hand,

enhancer competition is more effective on ensuring single allele activation, but it also introduces strong bias towards allele selection. Therefore, to achieve single allele activation as the top priority and maximize the diversity of expressed ORs at the same time, the OR selection system has evolved a combined procedure. The epigenetic activation step is optimized with a bifunctional LSD1 to achieve maximal single allele activation. When multiple allele epigenetic activation does happen but with low probability, the enhancer competition in allele selection serves as the last “safeguard” without severely distorting the overall diversity of OR expression. Similarly our analysis reveals that other variables are also subject to the multi-task optimization. For example, the LSD1 concentration may be optimized as a result of compromise between maximum single-allele activation and fast allele activation.

### **3.4.3. Counter-intuitive bifunctionality of the LSD1 maximizes single allele epigenetic activation.**

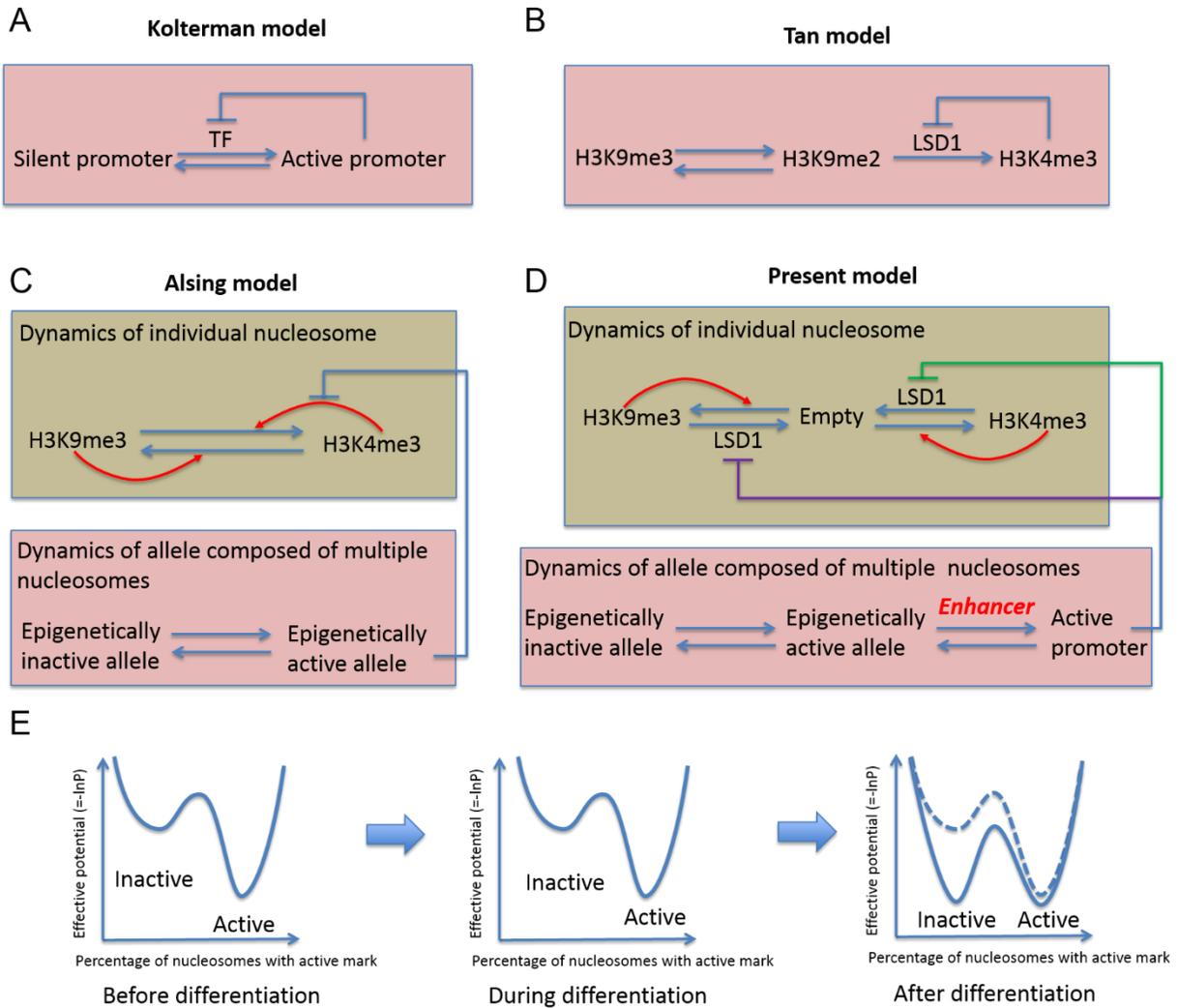
An intriguing feature of the OR selection system is that the selection is initialized then maintained through regulating the level of the bifunctional LSD1. While not discussed in previous studies, our analysis reveals that it is important to remove both repressive and active marks during the activation process. This bifunctionality leads to a barrier-crossing-like dynamics with high single allele epigenetic activation ratio and minimization of alleles trapped in hybrid epigenetic states. Theoretically, one might conceive many possible designs of modulating the methyltransferases to activate OR in the cell. For example, LSD1 might first remove H3K9 methylation to activate the alleles, subsequently remove H3K4 methylation on those “unsuccessful” alleles including pseudogenes. However, besides problems of the ratchet-like dynamics discussed above, it is also practically difficult to prevent the enzyme from destabilizing the activated allele by removing its H3K4 methylation marks as well.

Therefore a central prediction of our model is that throughout the selection process a “tug-of-war” exists for adding and removing H3K9 and H3K4 methylations. This “tug-of-war” is analogous to that of ultrasensitive phosphorylation-dephosphorylation cycle observed in signal transduction networks <sup>37</sup>, and work together with nucleosome crosstalks to generate the kinetic cooperativity during the epigenetic

activation process. Furthermore the enzymatic activities of methyltransferases are in excess over that of demethylase, i.e., LSD1. Lyons *et al.* indeed observed that G9a/GLP at excessive concentration coexist with LSD1 during OSN differentiation<sup>22</sup>.

#### **3.4.4. Seemingly subtle differences on the regulation schemes lead to qualitatively different mechanisms.**

Existence of a feedback loop has long been recognized to be necessary for maintaining the choice of OR selection. Not surprisingly all mathematical studies including ours contain a feedback loop. These models, however, differ on how the feedback loop is implemented. Figure 3.7 Supplement 1A-D summarizes the basic structures of the four existing mathematical models. Some of these models differ only in some subtle details, but suggest qualitatively different mechanisms. We have already discussed models with bifunctional and unifunctional LSD1. The two models differ only on whether H3K4 methylation is regulated by the negative feedback loop, but lead to barrier-crossing versus ratchet-like dynamics. Similarly, the model of Alsing *et al.* (Alsing and Sneppen 2013) emphasizes that “the only requirement is that the coupling feedback must favour the silent nucleosome state”. Figure 3.7 Supplement 1E gives a corresponding potential analogy, where activation of the feedback loop stabilizes the repressive mark dominated well, while slightly destabilizes the active mark dominated well. On the contrary, the feedback in the present model modulates the “barrier” without necessarily favoring any of the nucleosome states (Figure 3.3D). The two models respond differently to enzyme concentration fluctuations. In that of Alsing *et al.*, fluctuations of any of the methyltransferases or demethylases affect the relative stability of the two states, and these fluctuations are coupled. On the other hand, in the present model fluctuations of the two methyltransferases are largely decoupled provided that the LSD1 concentration is kept low compared to that of the methyltransferases.



**Figure 3.7 supplement 1: Comparison of existing mathematical models.** (A) In the model of Kolterman et al.<sup>18</sup>, OR promoters compete for cooperative binding of a limited number of transcription factors, and OR expression elicits a feedback to degrade the transcription factors. While the mechanism is not consistent with known experimental information, the model suggests the importance of cooperative binding of trans-elements. (B) The model of Tan et al.<sup>16</sup> assumes that the transcriptional activity of an OR allele is controlled by the epigenetic state of one nucleosome in the promoter region. A slow LSD1-dependent  $\text{H3K9me2} \rightarrow \text{H3K4me3}$  transition is controlled by a fast feedback that removes LSD1. Notice that the transition is irreversible and LSD1 is considered unifunctional. (C) Alsing et al.<sup>17</sup> considers that each allele is composed of a number of nucleosomes. Each nucleosome exists as  $\text{H3K9me3}$  or  $\text{H3K4me3}$ , and the transition is affected by other

nucleosomes as positive feedbacks. A negative feedback elicited by the expressed OR decreases the  $H3K9me3 \rightarrow H3K4me3$  transition rate. (D) The present model also considers that each allele is composed of a number of nucleosomes. Treating each nucleosome dynamics as transitions among three states, the model is able to examine the role of bifunctional LSD1. The feedback affects both H3K9 and H3K4 demethylation. Furthermore an epigenetically active allele needs to bind enhancers to become transcriptionally active. A hypothetical unifunctional LSD1 model differs only in absence of the negative feedback on H3K4 demethylation (the green line and dependence of demethylation on LSD1), but generates qualitatively different dynamics. (E) Analogous potential system corresponding to the Alsing model. Initially the inactive well is shallower than the active well, reflecting the requirement that for mice OSN differentiation takes place in 5-10 days, but the choice needs to be maintained for ~3 months. The feedback reduces the rate of  $H3K9me3 \rightarrow H3K3me3$  transition, and thus stabilizes the inactive well while slightly destabilizes the active well.

#### **3.4.5. The model makes multiple testable predictions and many have been confirmed.**

A prominent feature of the present modeling study is that it is not only based on extensive experimental information, but also makes a number of predictions and experimental suggestions. As summarized in Table 3.1 and discussed above, our model predicts many seemingly “strange” dynamic features of the OR selection system, which have been experimentally observed but remained unexplained. Here we discuss a few of suggested new experiments in detail.

To reach the diversity change prediction in Figure 3.5F & G, a key ingredient in the model is that the values of  $\zeta$  and/or  $\epsilon_l$  are allele-dependent. The difference may come from DNA sequence, and it may be even less than the thermal energy  $k_B T$ , the product of Boltzmann’s constant and temperature. However, this free energy difference can be significantly amplified by enhancer cooperative binding (Figure 3.5D). This amplification explains why strong OR expression bias occurs in G9a/GLP dKO mice while Lyon *et al.* could not identify any significant differences between the promoters of the most upregulated ORs and the remaining ones in predicting the transcription-factor-binding-motifs<sup>22</sup>. Another possible source of different OR-enhancer binding strength lies in the different distances between enhancers and alleles. Different allele-enhancer distances may require slight different DNA distortion to form the OR-enhancer binding complex,

as implied by the observation that moving the H enhancer closer to MOR28 dramatically up-regulates its expression while down regulates other neighboring ORs<sup>12</sup>. To further test this mechanism, one can replace an upregulated OR gene and its promoter by a down-regulated one, and test whether the latter becomes upregulated in a G9a/GLP dKO main olfactory epithelium. Another suggested experiment is to introduce enhancers ectopically to G9a KO mice (Lomvardas et al., 2006), which should at least partially rescue the reduction of OR diversity if the model holds.

**Table 3.1** Model predictions and corresponding experimental confirmations and suggestions.

<b>Model predictions</b>	<b>Experimental confirmation or suggestions</b>
OSNs need to maintain saturating levels of methyltransferases, but low levels of demethylases and stochastic histone exchange rate before and after differentiation (Figure 3.2 and 3.2 Supplement 1).	Compare the enzyme levels and histone exchange rate (e.g., using isotope labeled histones) within OSNs and other types of cells.
The number of OR alleles in a zone affects the single-allele ratio nonmonotonically (Figure 3.3B).	
Decreasing LSD1 concentration impedes OR activation (less OSN differentiation), which can be partially restored by inhibiting G9a/GLP.	Confirmed in mice <sup>22</sup> .
Epigenetic switching assumes a barrier-crossing-like dynamics for WT (Figure 3.3, Figure 3.4B), but a ratchet-like dynamics with G9a/GLP dKO (Figure 3.5 Supplement 1 A & B).	Following Magklara et al. <sup>19</sup> , sort GFTP+ cells from OMP-IRES-GFP mice, and perform CHIP-qPCR for selected silent OR genes. Perform similar studies with OMP-IRES-GFP and G9a/GLP dKO mice. We predict that silent OR alleles from the former are dominated by H3K9me3, but those from the latter have mixed nucleosomes with H3K4 and H3K9 methylations. One can further measure the epigenetic pattern at different time points before and after differentiation to test the prediction that it takes long time for the alleles with mixed methylations to relax to one of the epigenetic states with one mark dominated.
A cell may have more than one epigenetically active alleles (Figure 3.3F).	Following Shykind et al. <sup>33</sup> , cross mice bearing MOR28-IRES-Cre allele with strains bearing the reporter Rosa-loxP-stop-loxP-CFP, sort CFP <sup>+</sup> Cre <sup>-</sup> cells and perform epigenetic histone modification analysis as in Magklara et al. <sup>19</sup> .
Inhibition of H3K9 methyltransferases G9a/GLP leads to multiple allele activation (Figure 3.5A).	Confirmed in Zebrafish <sup>21</sup> and mice <sup>22</sup> .

With enhancer competition, Inhibition of H3K9 methyltransferases G9a/GLP leads to transcriptional downregulation of most OR genes upregulation of a small number of genes, and thus lead to decrease of diversity of expressed OR genes (Figure 3.5E-G).	Confirmed in mice <sup>22</sup> .
Multiple epigenetically active alleles compete for a finite number of enhancers, which contributes to the diversity reduction in G9a/GLP KO mice (Figure 3.5).	Introduce enhancers ectopically to G9a KO mice (Lomvardas et al., 2006), which should at least partially rescue the reduction of OR diversity if the model holds.
The proximity difference of enhancers to a gene leads to different OR-enhancer binding strength.	Replace an upregulated OR gene together with the promoter by a down-regulated one, and test whether the latter becomes upregulated in a G9a/GLP dKO MOE.
The binding strength differences between an OR promoter and individual enhancers can be small thus experimentally hard to detect, but are amplified by cooperative enhancer binding (Figure 3.5D).	Lyon <i>et al.</i> could not identify any significant differences between the promoters of the most upregulated ORs and the remaining ones in predicting the transcription-factor-binding-motifs <sup>22</sup>
The switching frequency increases in Adcy3 KO OSNs compared to that in WT OSNs (Figure 3.6D). Furthermore, the fraction of cells expressing pseudo ORs increases while that expressing functional ORs decreases in Adcy3 KO mice (Figure 3.6E).	Confirmed <sup>15</sup>
The genes showing upregulated expression in the G9a/GLP dKO mice, such as Olfr231, have slighter stronger interactions with the enhancers than the remaining genes do. Then in normal mice, OSNs that express one of these genes should have lower frequency of switching than those cells express other genes in the same zone do.	Follow the study procedure of Shykind et al. <sup>33</sup> . It can reveal more information if one use techniques such as the CRISPR-Cas9 gene editing approach to fluorescently label genes like Olfr231, and perform time-lapse studies.

To test the prediction given in Figure 3.5 supplement 1A&B, one may sort GFP<sup>+</sup> cells from OMP-IRES-GFP control mice and G9a/GLP dKO OMP-IRES-GFP mice<sup>19</sup>, respectively, then perform CHIP-qPCR for selected silent OR genes. We expect that H3K9me3 dominates on silent OR alleles from the control mice, but H3K4 and H3K9 methylations mixed at various extent on silent OR alleles from the dKO mice (Figure 3.5 supplement 1B). One can further measure the epigenetic pattern at different time points before and after differentiation to test the prediction that it takes long time for the alleles with mixed methylations to relax to a steady state distribution.

### **3.4.6. Future studies may reveal fine-tuned regulation on OR and other allele-specific activation processes.**

This work aims at revealing the essential elements that regulate the OR selection process. Therefore we adopted a coarse-grained model without specifying many details. For example, the model requires kinetic cooperativity of epigenetic mark change among nucleosomes, and energetic cooperative binding of a limited number of enhancers. Molecular details of these cooperativities require clarification. For simplicity, we did not distinguish possible differences among enhancers and their OR specificity, and leave the effective enhancer number in a cell as an unspecific parameter. The molecular mechanism for zonal segregation is unclear. All these unresolved questions require further studies.

Besides the mechanisms discussed in this work, living organisms have likely evolved additional mechanisms for fine-tuning OR expression. For example, chromatin structures in OSNs are highly dynamic to expose or sequester specific OR genes. Specific patterns of DNA methylation and other histone covalent modifications have been observed for OR promoters and enhancers. OR genes are not expressed with exact equal probability, and coordinated expression might exist<sup>38</sup>. Further studies are needed to reveal these subtle regulation mechanisms.

Our model studies reveal some design principles to achieve robust single allele activation, which may apply to other single allele systems as well<sup>10</sup>. Stochastic monoallelic expression has been widely observed in diploid organisms, with an estimate of >20% of the autosomal genes, and OR expression is one prominent example<sup>39</sup>. Further studies can examine whether the regulatory mechanisms discussed here are adopted in other cell types.

### **3.5. Method:**

Each OSN is modeled to have  $N_p = 30$  pseudo gene alleles and  $N_f = 70$  functional OR alleles, with the only difference being that the product of the former does not elicit Adcy3 mediated feedback.

**Epigenetic dynamics:** For simplicity we treated step-wise methylations/demethylations on a nucleosome as single steps, and treated participating enzymes other than LSD1 implicitly. Denote

methylation state of a nucleosome R, E, and A as  $s = -1, 0, 1$ , respectively. We set the methylation rate constants for an empty nucleosome  $i$  as

$$k_1^i = k_1^0 \exp\left(\sum_{j \neq i} \frac{\mu}{|j-i|} \delta_{s_j,1}\right), k_2^i = k_2^0 \exp\left(\sum_{j \neq i} \frac{\mu}{|i-j|} \delta_{s_j,-1}\right),$$

where the sum is over all other nucleosomes, and  $\delta$  is a Kroneck-delta function. That is, each of the other nucleosomes influences the nucleosome to add the same mark of the latter, and the influence decreases with the nucleosome spatial separation. An insulating boundary is assumed, and three nucleosomes in the middle form a nucleation region with higher enzymatic rate constants than other nucleosomes have. We modeled  $E_{K9M}^R$  and  $E_{K9M}^{RR}$  by reducing the value of  $k_2^0$  for the WT to 90% and 80%, respectively. A theoretical justification of the model from a more detailed physical model <sup>31</sup> is given below, and values of the model parameters can be found in Table 3.2.

**Table 3.2** Values of model parameters used in this work.

Parameter	Value
Active mark methylation rate constant $k_1$	within nucleation region $0.125 \text{ h}^{-1}$ , outside nucleation region $0.025 \text{ h}^{-1}$ .
Repressive mark methylation rate constant $k_2$	within nucleation region, $0.125 \text{ h}^{-1}$ (WT), outside nucleation region, $0.025 \text{ h}^{-1}$ (WT)
Active mark demethylation rate constant $k_{-1}$	$0.125 \text{ h}^{-1}$
Repressive mark demethylation rate constant $k_{-2}$	$0.125 \text{ h}^{-1}$
Nucleosome correlation length $\mu$	0.64
Histone turnover rate $d$	$0.002 \text{ h}^{-1}$
Cutoff fraction of nucleosomes with active marks so an allele is regarded as epigenetically activated $\lambda_\theta$	0.75
Adcy3 synthesis rate $k_A$	$1 \text{ h}^{-1}$
Michaelis-Menten constant of OR induced Adcy3 expression $K_A$	0.8
LSD1 basal degradation rate constant $d_L^0$	$0.5 \text{ h}^{-1}$
Adcy3 facilitated LSD1 degradation rate constant $d_L^1$	$8 \text{ h}^{-1}$

Prefactor for the enhancer switching rate constant $\nu^*$	$1 \text{ h}^{-1}$
Free energy of enhancer-enhancer interaction $\zeta$	$-0.5 k_B T$
Free energy of enhancer-allele interaction $\varepsilon$	$\sim -1 k_B T$
Total number of enhancers $M$	12

\* This parameter is only used in generating Figure 3.5D for illustration purpose, and its actual value can be better estimated if time-course of OR switching becomes available.

Three nucleosomes located at the center of the nucleosome array form the nucleation region. Existence of this nucleation region reflects the observation that some DNA sequence specific molecular species, such as transcription factors and noncoding RNAs, help on recruiting histone modification enzymes. We also performed simulations with the nucleation region and the found no qualitative change of the mechanisms discussed in the present paper.

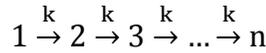
**Enhancer binding dynamics:** For simplicity we assumed that there is no free enhancer. This assumption is not essential for the present discussions and can be easily removed at the expense of a few additional parameters. Also we treated the enhancers equally, although generalization is straightforward when additional experimental information becomes available. An enhancer can jump from allele  $i$  to  $j$  with rate,  $k_{i \rightarrow j} = \nu \exp[0.5(\varepsilon_i - \varepsilon_j + (M_i - 1 - M_j)\zeta)]$  to satisfy the detailed balance requirement, where  $M_i$  and  $M_j$  are the number of enhancers bound to allele  $i$  and  $j$  before the jump, respectively, and  $\sum_i M_i = M$ . We chose the factor 0.5 to satisfy the detailed balance requirement, *i.e.*,  $k_{i \rightarrow j}/k_{j \rightarrow i}$  equals to the Boltzmann factor corresponding to the system free energy after the transition divided by that prior to the transition. At each Gillespie simulation step, one of all possible enhancer binding changes is randomly selected. Since an allele with higher enhancer binding affinity dominates enhancer competition, for computational efficiency we only simulated enhancer dynamics explicitly for the results in Figure 3.5D and Figure 3.5 supplement 1. For other simulations in Figure 3.5 and 6 we adopted a simplified procedure as schematically illustrated in Figure 3.5H. That is, we obtained the probabilities of having a cell with various numbers of epigenetically active alleles (Figure 3.5E), and stochastically ranked the enhancer binding affinities of the 100 alleles, Assuming that each allele has the same probability of being epigenetically active, we consider all the

possible combinations and the associated weights of having two or more alleles being epigenetically active, and for each case the allele with strongest enhancer binding affinity is chosen to be transcriptionally active.

**Gene expression dynamics:** All gene expression is modeled by solving ordinary differential equations (Figure 3.1B). For simulations with enhancer binding dynamics in Figure 3.5, Figure 3.6, we multiplied to the first synthesis term of OR expression a Kroneck-delta function, which assumes 1 if the allele is epigenetically active and other alleles are epigenetically silent, or if its enhancer binding affinity is stronger than that of other epigenetically active alleles, and 0 otherwise. For Adcy3 KO simulations,  $k_A$  is set to be 0. All concentrations are dimensionless values.

### 3.5.1. Mathematical analysis of simple models

Here we show how to calculate the results in Figure 3.4 supplement. Consider the process (also shown in Figure 3.4 supplement B)



Denote  $p_i$  the probability of an allele in state  $i$ , which is given by

$$\frac{d}{dt} \begin{pmatrix} p_1 \\ p_2 \\ \dots \\ p_n \end{pmatrix} = \begin{pmatrix} -k & 0 & 0 & 0 & 0 \\ k & -k & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & k & 0 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ \dots \\ p_n \end{pmatrix} \text{ with } \begin{pmatrix} p_1 \\ p_2 \\ \dots \\ p_n \end{pmatrix}_0 = \begin{pmatrix} 1 \\ 0 \\ \dots \\ 0 \end{pmatrix}$$

The solution of the system is,

$$\begin{aligned} p_1(t) &= e^{-kt}, \\ p_2(t) &= e^{-kt} kt, \\ p_3[t] &\rightarrow \frac{1}{2} e^{-kt} k^2 t^2, \\ &\dots \\ p_n[t] &\rightarrow \frac{1}{(n-1)!} e^{-kt} k^{n-1} t^{n-1} \end{aligned}$$

The first-arrival time distribution is  $f_n(t) = \frac{d}{dt} p_n(t)$ , and  $f_n$  is normalized ( $\int_0^\infty f_n * dt = 1$ ). Then

$$f_2[t] \rightarrow e^{-kt} k^1,$$

$$f_3[t] \rightarrow e^{-kt} k^2 t,$$

...

$$f_n[t] \rightarrow \frac{1}{(n-2)!} e^{-kt} k^{n-1} t^{n-2}$$

The mean first arrival time  $T$  is given by  $T = \int_0^\infty f_n * t * dt$ . Require that the mean first arrival time  $T$  is the same for different  $n$ , one has  $k = (n-1)/T$ .

The formula below gives the distribution that the arrival time difference between two alleles is  $\tau$

$$F_n = 2 \int_0^\infty f_n(t) * f_n(t + \tau) * dt$$

Thus,

$$\begin{aligned} F_n &= 2 \int_0^\infty \frac{1}{(n-2)!} e^{-kt} k^{n-1} t^{n-2} * \frac{1}{(n-2)!} e^{-k(t+\tau)} k^{n-1} (t+\tau)^{n-2} * dt \\ &= 2 \frac{1}{(n-2)! (n-2)!} e^{-k\tau} k^{2n-2} \int_0^\infty e^{-2kt} t^{n-2} (t+\tau)^{n-2} * dt \end{aligned}$$

Choose the time unit so that  $T = 1$ , one has

$$F_2 = e^{-\tau}$$

$$F_3 = e^{-2\tau} (1 + 2\tau)$$

$$F_4 = \frac{9}{8} e^{-3\tau} (1 + 3\tau(1 + \tau))$$

$$F_5 = \frac{1}{12} e^{-4\tau} (15 + 4\tau(15 + 8\tau(3 + 2\tau)))$$

$$F_6 = \frac{25}{384} e^{-5\tau} (21 + 5\tau(21 + 5\tau(9 + 5\tau(2 + \tau))))$$

### **3.6. Author contributions**

JX conceived the project, constructed the model and wrote the paper with input from XT and HZ. XT performed most simulations. HZ collected biological background information and participated simulations. XT, HZ and JX analyzed the data.

### **3.7. Acknowledgements**

We thank Drs David Lyons and Andrew Chess for many helpful discussions. The research was supported by the U.S. National Science Foundation (DMS-0969417).

## References

1. Rinaldi, A. The scent of life. The exquisite complexity of the sense of smell in animals and humans. *EMBO Rep* **8**, 629-33 (2007).
2. Glusman, G., Yanai, I., Rubin, I. & Lancet, D. The complete human olfactory subgenome. *Genome Res* **11**, 685-702 (2001).
3. Zhang, X. & Firestein, S. The olfactory receptor gene superfamily of the mouse. *Nat Neurosci* **5**, 124-33 (2002).
4. Godfrey, P.A., Malnic, B. & Buck, L.B. The mouse olfactory receptor gene family. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 2156-2161 (2004).
5. Vosshall, L.B., Amrein, H., Morozov, P.S., Rzhetsky, A. & Axel, R. A spatial map of olfactory receptor expression in the *Drosophila* antenna. *Cell* **96**, 725-36 (1999).
6. Nei, M., Niimura, Y. & Nozawa, M. The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. *Nat Rev Genet* **9**, 951-63 (2008).
7. Buck, L. & Axel, R. A Novel Multigene Family May Encode Odorant Receptors - a Molecular-Basis for Odor Recognition. *Cell* **65**, 175-187 (1991).
8. Ressler, K.J., Sullivan, S.L. & Buck, L.B. A zonal organization of odorant receptor gene expression in the olfactory epithelium. *Cell* **73**, 597-609 (1993).
9. Chess, A., Simon, I., Cedar, H. & Axel, R. Allelic inactivation regulates olfactory receptor gene expression. *Cell* **78**, 823-834 (1994).
10. Magklara, A. & Lomvardas, S. Stochastic gene expression in mammals: lessons from olfaction. *Trends Cell Biol* **23**, 449-56 (2013).
11. Lewcock, J.W. & Reed, R.R. A feedback mechanism regulates monoallelic odorant receptor expression. *Proc Natl Acad Sci U S A* **101**, 1069-74 (2004).
12. Serizawa, S. *et al.* Negative feedback regulation ensures the one receptor-one olfactory neuron rule in mouse. *Science* **302**, 2088-94 (2003).

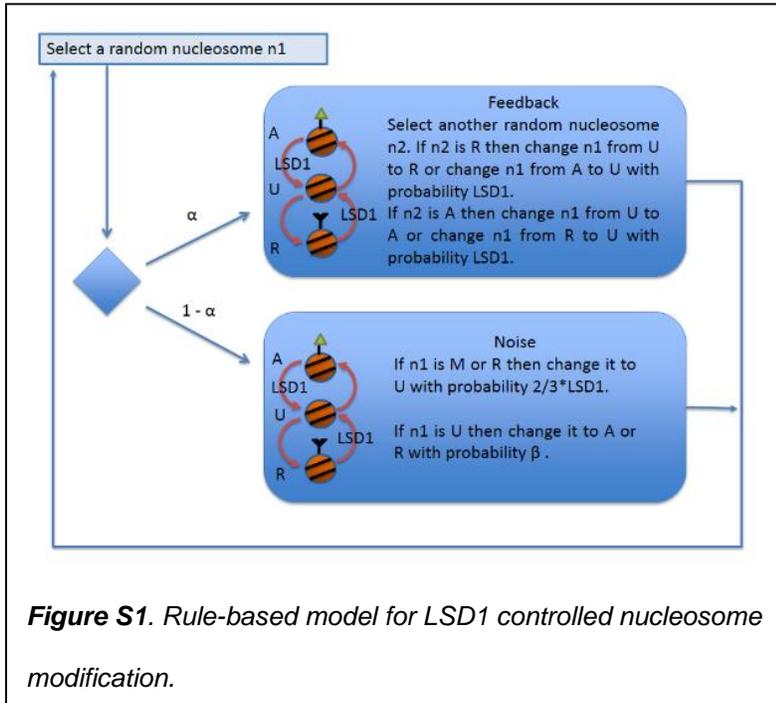
13. Nguyen, M.Q., Zhou, Z., Marks, C.A., Ryba, N.J. & Belluscio, L. Prominent roles for odorant receptor coding sequences in allelic exclusion. *Cell* **131**, 1009-17 (2007).
14. Dalton, R.P., Lyons, D.B. & Lomvardas, S. Co-opting the unfolded protein response to elicit olfactory receptor feedback. *Cell* **155**, 321-32 (2013).
15. Lyons, D.B. *et al.* An epigenetic trap stabilizes singular olfactory receptor expression. *Cell* **154**, 325-36 (2013).
16. Tan, L., Zong, C. & Xie, X.S. Rare event of histone demethylation can initiate singular gene expression of olfactory receptors. *Proceedings of the National Academy of Sciences* (2013).
17. Alsing, A.K. & Sneppen, K. Differentiation of developing olfactory neurons analysed in terms of coupled epigenetic landscapes. *Nucleic Acids Res* **41**, 4755-64 (2013).
18. Kolterman, B.E., Iossifov, I. & Koulakov, A.A. A race model for singular olfactory receptor expression. *arXiv preprint arXiv:1201.2933* (2012).
19. Magklara, A. *et al.* An epigenetic signature for monoallelic olfactory receptor expression. *Cell* **145**, 555-70 (2011).
20. Sim, C.K., Perry, S., Tharadra, S.K., Lipsick, J.S. & Ray, A. Epigenetic regulation of olfactory receptor gene expression by the Myb-MuvB/dREAM complex. *Genes Dev* **26**, 2483-98 (2012).
21. Ferreira, T. *et al.* Silencing of odorant receptor genes by G protein betagamma signaling ensures the expression of one odorant receptor per olfactory sensory neuron. *Neuron* **81**, 847-59 (2014).
22. Lyons, D.B. *et al.* Heterochromatin-mediated gene silencing facilitates the diversification of olfactory neurons. *Cell Rep* **9**, 884-92 (2014).
23. Lomvardas, S. *et al.* Interchromosomal interactions and olfactory receptor choice. *Cell* **126**, 403-13 (2006).
24. Khan, M., Vaes, E. & Mombaerts, P. Regulation of the probability of mouse odorant receptor gene choice. *Cell* **147**, 907-21 (2011).
25. Fuss, S.H., Omura, M. & Mombaerts, P. Local and cis effects of the H element on expression of odorant receptor genes in mouse. *Cell* **130**, 373-84 (2007).

26. Markenscoff-Papadimitriou, E. *et al.* Enhancer Interaction Networks as a Means for Singular Olfactory Receptor Expression. *Cell* **159**, 543-557 (2014).
27. Plessy, C. *et al.* Promoter architecture of mouse olfactory receptor genes. *Genome Res* **22**, 486-97 (2012).
28. Hathaway, N.A. *et al.* Dynamics and memory of heterochromatin in living cells. *Cell* **149**, 1447-60 (2012).
29. Bannister, A.J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res* **21**, 381-95 (2011).
30. Alexander, J.M. & Lomvardas, S. Nuclear architecture as an epigenetic regulator of neural development and function. *Neuroscience* **264**, 39-50 (2014).
31. Zhang, H., Tian, X.J., Mukhopadhyay, A., Kim, K.S. & Xing, J. Statistical mechanics model for the dynamics of collective epigenetic histone modification. *Phys Rev Lett* **112**, 068101 (2014).
32. Vassar, R., Ngai, J. & Axel, R. Spatial segregation of odorant receptor expression in the mammalian olfactory epithelium. *Cell* **74**, 309-18 (1993).
33. Shykind, B.M. *et al.* Gene switching and the stability of odorant receptor gene choice. *Cell* **117**, 801-15 (2004).
34. Weth, F., Nadler, W. & Korsching, S. Nested expression domains for odorant receptors in zebrafish olfactory epithelium. *Proc Natl Acad Sci U S A* **93**, 13321-6 (1996).
35. Dodd, I.B., Micheelsen, M.A., Sneppen, K. & Thon, G. Theoretical analysis of epigenetic cell memory by nucleosome modification. *Cell* **129**, 813-22 (2007).
36. Peskin, C.S., Odell, G.M. & Oster, G.F. Cellular motions and thermal fluctuations: the Brownian ratchet. *Biophys. J.* **65**, 316-324 (1993).
37. Goldbeter, A. & Koshland, D.E., Jr. An amplified sensitivity arising from covalent modification in biological systems. *Proc Natl Acad Sci U S A* **78**, 6840-4 (1981).
38. Rodriguez-Gil, D.J. *et al.* Chromosomal location-dependent nonstochastic onset of odor receptor expression. *J Neurosci* **30**, 10067-75 (2010).

39. Burgess, D.J. Gene Regulation: Characterizing monoallelic expression. *Nat Rev Genet* **15**, 143 (2014).
40. Grubb, M.S. Olfactory Nerve. in *Encyclopedia of Neuroscience* 2998-3002 (Springer, 2009).
41. Olender, T., Lancet, D. & Nebert, D.W. Update on the olfactory receptor (OR) gene superfamily. *Hum Genomics* **3**, 87-97 (2008).
42. Serizawa, S., Miyamichi, K. & Sakano, H. One neuron–one receptor rule in the mouse olfactory system. *Trends in Genetics* **20**, 648-653 (2004).
43. Malnic, B., Gonzalez-Kristeller, D.C. & Gutiyama, L.M. Odorant Receptors. in *The Neurobiology of Olfaction* (ed. Menini, A.) (Boca Raton (FL), 2010).
44. Chess, A. Random and non-random monoallelic expression. *Neuropsychopharmacology* **38**, 55-61 (2013).
45. Eckersley-Maslin, M.A. *et al.* Random Monoallelic Gene Expression Increases upon Embryonic Stem Cell Differentiation. *Dev Cell* **28**, 351-65 (2014).
46. Deng, Q., Ramskold, D., Reinius, B. & Sandberg, R. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **343**, 193-6 (2014).
47. Chess, A., Simon, I., Cedar, H. & Axel, R. Allelic inactivation regulates olfactory receptor gene expression. *Cell* **78**, 823-34 (1994).
48. Rodriguez, I. Singular expression of olfactory receptor genes. *Cell* **155**, 274-7 (2013).
49. Li, J.S., Ishii, T., Feinstein, P. & Mombaerts, P. Odorant receptor gene choice is reset by nuclear transfer from mouse olfactory sensory neurons. *Nature* **428**, 393-399 (2004).
50. Clowney, E.J. *et al.* High-throughput mapping of the promoters of the mouse olfactory receptor genes reveals a new type of mammalian promoter and provides insight into olfactory receptor gene regulation. *Genome Research* **21**, 1249-1259 (2011).
51. Clowney, E.J. *et al.* Nuclear Aggregation of Olfactory Receptor Genes Governs Their Monogenic Expression. *Cell* **151**, 724-737 (2012).

52. Johnston, R.J. & Desplan, C. Interchromosomal Communication Coordinates Intrinsically Stochastic Expression Between Alleles. *Science* **343**, 661-665 (2014).
53. Lyons, D.B., Magklara, A., Schotta, G. & Lomvardas, S. Heterochromatin-Mediated Gene Silencing Facilitates the Diversification of Olfactory Neurons. *Cell Rep* (2014).

### Supplemental rule-based model:



### Rule-based modeling:

The working model is based on very general physics of a dynamic system: barrier-crossing dynamics versus ratchet-like dynamics. Therefore we anticipate that many model details are irrelevant. To test this irrelevance, instead of our physical interaction-based model we will repeat the above comparison with a much-simplified rule-based model

generalized from the work of Dodd *et al.*.

Here is the detailed description of how we set up the rule-based model.

Since the LSD1 enzyme demethylates both repressive and active histone marks, we want to add the LSD1 influence to both the active and repressive remove rule.

During the OR differentiation process, LSD1 concentration will be up-regulated first and then down-regulated after one OR turns on and elicits the feedback. As illustrated in Figure S1, here is the detailed mathematical rule for this process:

Step 1: A random nucleosome  $n1$  is selected to be modified among  $N$  nucleosomes. The selected nucleosome  $n1$  has a probability of  $\alpha$  to go to the feedback module, or with probability  $1-\alpha$  to go to the noise module. As Dodd *et al.* used,  $F = \alpha/(1-\alpha)$  denotes the signal to noise ratio.

Step 2A: Recruited conversion. Another random nucleosome  $n2$  is selected from anywhere within the region. If  $n2$  is R, if  $n1$  is A, then  $n1$  will change to U with probability LSD1; if  $n1$  is U, then  $n1$  will change

to R. If n2 is A, if n1 is R, then n1 will change to U with probability LSD1; if n1 is U, then n1 will change to A.

Step 2B: Noise conversion. If n1 is A or R, change it to U with probability  $\beta$ . Else if n1 is A or R, change it to U with probability  $2/3 * LSD1$ .

From the following figure S2, we can see that rule-based model also could achieve the one allele activation scenario.

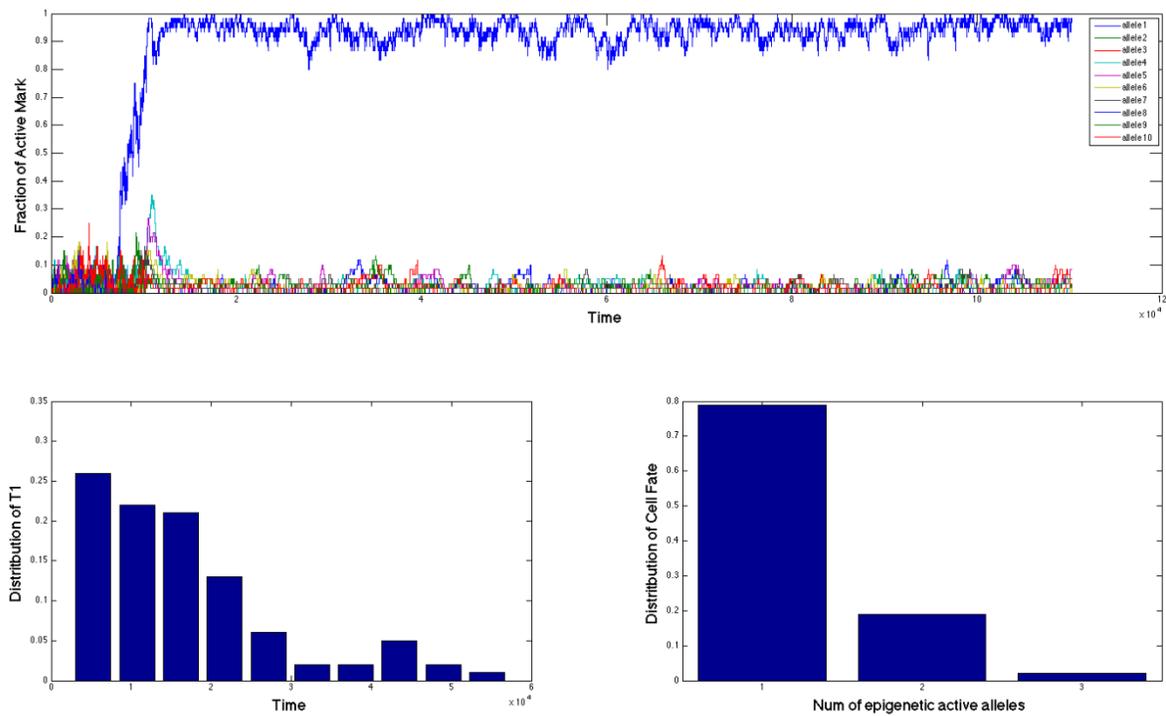


Figure S2. Simulation result with rule-based model of 10 alleles.

## Chapter 4. Post-transcriptional motif regulating differentiation of human definitive endoderm

This chapter is based on a manuscript which is under review in Scientific Report.

Yue Teng<sup>\*1,9</sup>, Hang Zhang<sup>\*5</sup>, Wen Yue<sup>\*1</sup>, Jingxue Wang<sup>1</sup>, Jichao Du<sup>1</sup>, Jingli Li<sup>2</sup>, Jun Zhou<sup>6</sup>, Xiaojun Tian<sup>4</sup>, Bowen Zhang<sup>1</sup>, Liangtao Zheng<sup>7</sup>, Quan Zeng<sup>1</sup>, Lijuan He<sup>1</sup>, Xue Nan<sup>1</sup>, Ying Jiang<sup>+2</sup>, Jianhua Xing<sup>+4,8</sup>, Fuchu He<sup>+2,3</sup>, Xuetao Pei<sup>+1</sup>

\*These authors contributed equally to this work.

<sup>1</sup> Stem Cell and Regenerative Medicine Laboratory, Beijing Institute of Transfusion Medicine, 27 Taiping Road, Beijing 100850, P.R. China; <sup>2</sup> State Key Laboratory of Proteomics, Beijing Proteome Research Center, Beijing Institute of Radiation Medicine, Beijing 102206, P.R. China; <sup>3</sup> Institutes of Biomedical Sciences and Department of Chemistry, Fudan University, Shanghai 200032, P.R. China; <sup>4</sup>Department of Computational and Systems Biology, School of Medicine, University of Pittsburgh, Pittsburgh, PA, 15260, USA, <sup>5</sup> Genetics, Bioinformatics, and Computational Biology Program, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA; <sup>6</sup> College of computer and information science, Northeastern University, Boston, MA, US; <sup>7</sup> The Key Laboratory of Molecular Biology in Medical Sciences of Zhejiang Province, Cancer Institute, Hangzhou, Zhejiang, China; <sup>8</sup>Beijing Computational Science Research Center, Beijing 100084, China; <sup>9</sup>The present address is The State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, 20 Dong-Da Street, Fengtai District, Beijing 100071, China.

<sup>+</sup> To whom correspondence should be addressed: Xuetao Pei, Tel: 8610-68164807; Fax: 8610-68167357; E-mail: [peixt@nic.bmi.ac.cn](mailto:peixt@nic.bmi.ac.cn);

Correspondence may also be addressed to: Fuchu He, Tel: 8610-80705299; Fax: 8610-80705002;

E-mail: [hffc@nic.bmi.ac.cn](mailto:hffc@nic.bmi.ac.cn);

Correspondence may also be addressed to: Jianhua Xing, Tel/Fax: 001-4126488679; E-mail: [xing1@pitt.edu](mailto:xing1@pitt.edu).

## 4.1. Abstract

Systematic investigation of the developmental stage from human embryonic stem cells (hESCs) to definitive endoderm (DE), through which liver organogenesis occurs, may shed light on the underlying mechanisms of human liver development. Furthermore, these investigations may lead to more efficient and effective production of hepatocytes for cell therapies. Here, using two-dimensional difference gel electrophoresis in conjunction with mass spectrometry, we identified two significantly inversely altered splicing-related gene products during the differentiation process, heteronuclear ribonucleoprotein A1 (hnRNP A1) and KH-type splicing regulatory protein (KSRP). Combined bioinformatics and microRNA-Array data analysis suggests hnRNPA1 and KSRP antagonizing each other through miR-375 and miR-135a respectively. Further mathematical modeling analysis demonstrated that this motif could generate switch-like responses to the differentiation signal, which can serve as a noise filter to control hESCs self-renewal and differentiation. Simulations predicted that elevated hnRNP A1 or miRNA-375 expression lead to rapid and efficient differentiation of hESCs into DE was further experimentally validated. Taken together, we revealed a novel mechanism which functions in post-transcriptional level to regulate stem cell differentiation.

## 4.2. Introduction

Human embryonic stem cells (hESCs) can self-renew and differentiate into any cell type found in the three embryonic germ layers<sup>1,2</sup>, making them an attractive source of cells for use in regenerative medicine. For example, hESCs are capable to generate definitive endoderm cells (DECs), the precursor cell type to the liver, pancreas, lungs, thyroid, and intestines<sup>3,4</sup>. Within these organs, chronic shortage of liver donors limits the transplantation of whole liver organs and isolated hepatocytes, also makes deriving endoderm with hepatic potential for therapeutic and pharmaceutical applications of great clinical importance<sup>5</sup>. A substantial effort has been made to develop protocols that promote hepatic differentiation<sup>6</sup>. However, *in vitro* differentiation of hESCs toward the hepatic lineage has been challenging. More efficient ways of

controlling the hESC differentiation to DECs are required. Thus, understanding the underlying mechanism of hESC differentiate into DECs is not only of fundamental importance, but also of great clinical importance

Numerous studies have been performed about the stem cell fate decision process. It is well known that transcription factors Oct4, Sox2 and Nanog are the core regulators of pluripotency and could be regarded as the hallmark for ESCs<sup>7-9</sup>. These trio of transcription factors, known as master regulators, cooperate and antagonize each other during ESC self-renew and lineage specification process. Similar to ESCs, different lineage specifiers, are identified in Pu.1 and Gata1 will determine the choice of erythroid and myelomonocytic fates in haematopoietic progenitor cells<sup>10,11</sup>; naïve CD4<sup>+</sup> T cell use Foxp3 and ROR $\gamma$ t as master regulators to determine the cell fate during differentiation<sup>12</sup>. For definitive endoderm differentiation, transcription factors Sox17 and FoxA2 have been reported to facilitate differentiation of hESCs into DECs and are considered as the hallmark for DECs<sup>13,14</sup>. Directing embryonic stem cell differentiation towards definitive endoderm has been achieved by manipulating the Nodal and Wnt signaling pathways. Activin A, which activates the Nodal pathway, directs DE formation from mesendoderm precursors in mouse and human ESCs<sup>15</sup>. In human ESCs, synergistic activation of Nodal and Wnt- $\beta$ -catenin signaling promotes more efficient DE generation<sup>16</sup>. Besides the transcription factor master regulators, in recent year, microRNAs also emerge as important factors controlling the differentiation process and serve as hallmarks of different cell phenotype, e.g. miR-290 and miR-302 are hallmarks for ESCs<sup>17</sup>. Similar to transcription factors, microRNAs usually have plenty of downstream targets. Together with other regulators, they could form multiple negative feedback loops which will generate rich dynamics of different molecules involving in the cell fate decision process. While most previous studies in this area have been focused on identifying gene expression and signalling pathways, we want to investigate the key proteins and microRNAs associated with the differentiation process.

Here, we report results of a comparative proteomic analysis on DE derived from hESCs in feeder layer-free conditions, using two-dimensional difference gel electrophoresis (2D-DIGE) and mass spectrometry (MS). We identified two significantly inversely altered splicing-related gene products during the differentiation process, heteronuclear ribonucleoprotein A1 (hnRNP A1) and KH-type splicing regulatory

protein (KSRP). Combined bioinformatics and microRNA-Array data analysis suggests hnRNPA1 and KSRP antagonizing each other through miR-375 and miR-135a respectively. Based on these analyses, we revealed a potential regulatory motif in the post-transcriptional level to regulate the differentiation process. Furthermore, simulation results and experiments demonstrated that perturbing this motif could increase the differentiation efficiency of hESCs towards DE.

### **4.3. Results**

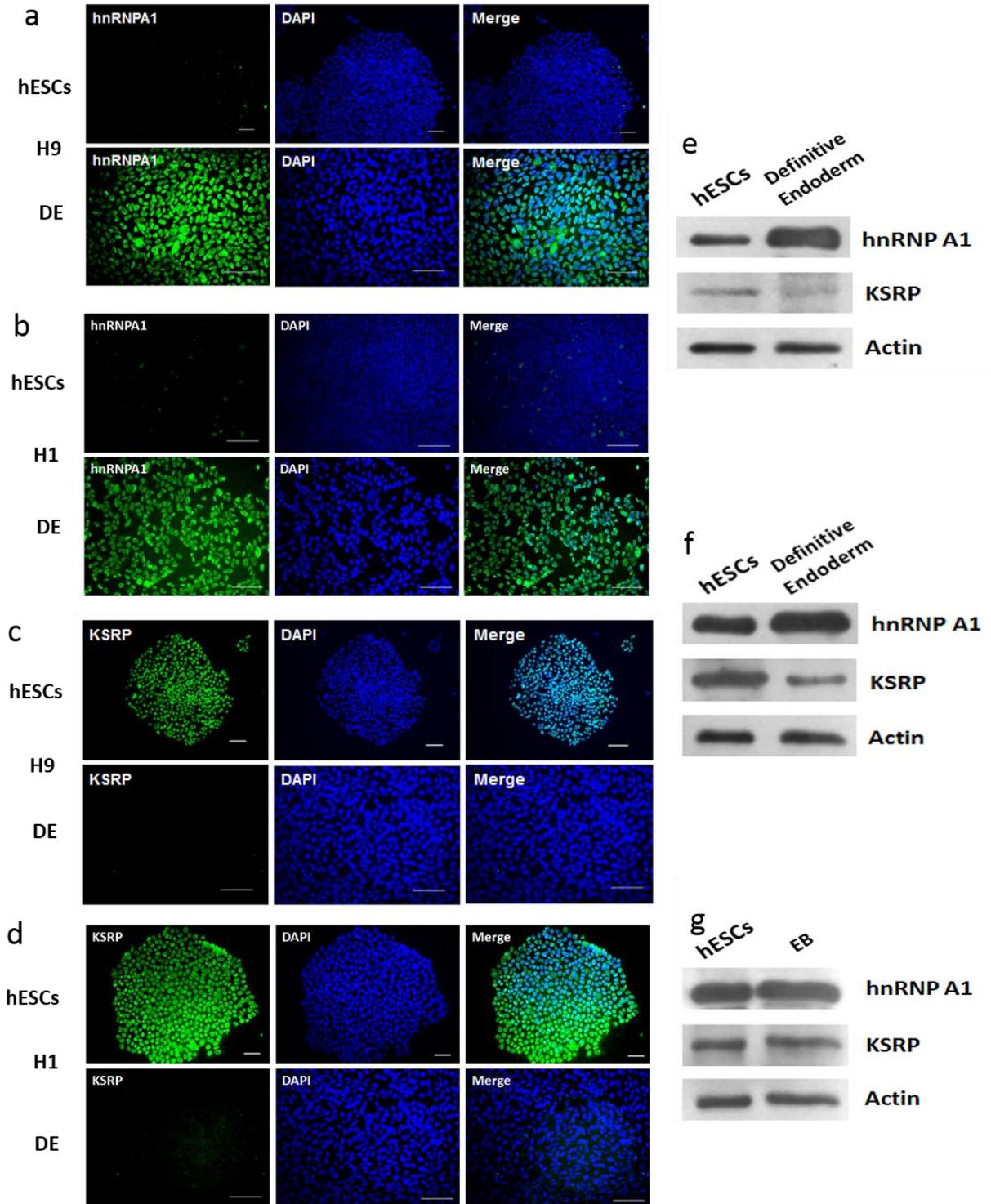
#### **4.3.1. Differentially expressed proteins hnRNP A1 and KSRP were identified from inversely altered proteins in DE derived from hESCs**

My collaborators identified two differentially expressed RNA binding proteins hnRNPA1 and KSRP. These two proteins act as splicing-associated factors and are responsible for the alternative splicing programs of genes. They have been found to interact with Drosha and Dicer to regulate the biogenesis of a subset of miRNAs, and may play a role in the gene regulatory network that controls the level of microRNA expression in response to intracellular or extracellular stimuli<sup>18</sup>. We report here, for the first time, that differential expression of hnRNP A1 and KSRP was identified by 2D-DIGE and mass spectrometry in the differentiation of hESCs into the DECs. In order to further verify DIGE results and investigate protein expression levels in the H1 and H9 cell lines, these protein spots were undergone western blotting and immunocytochemistry. Consistent with 2D-DIGE results, the up-regulated protein hnRNP A1 was found to be immunolocalized to nuclei (Figure 4.1A and B) and expressed in DECs (Figure 4.1E and F), but not in undifferentiated H1 and H9 hESCs. In contrast, KSRP was found to be expressed but down regulated in DECs, compared to that in undifferentiated hESCs. The decreased amount of KSRP in DECs was confirmed by immunocytochemistry (Figure 4.1C and D), and the results were consistent with findings from western blotting (Figure 4.1E and F).

Interestingly, expression of hnRNP A1 and KSRP proteins was not changed in the embryonic body from hESCs, H9 (Figure 4.1G). Similar to hESCs cultured in monolayer, hESCs within embryonic bodies undergo spontaneous differentiation and cell specification along the three germ lineages (endoderm,

ectoderm, and mesoderm), which comprise all somatic cell types. The expression of hnRNP A1 and KSRP in definitive endoderm also implies that they may be involved in germ layer specification and initiation of endoderm formation.

The hnRNP A1 and KSRP are known to regulate microRNA expression<sup>19,20</sup>. In the case of hnRNP A1, promotion of microRNA expression is achieved by antagonizing Wnt signaling, whereas KSRP regulates cyclin-dependent kinases (CDK) to promote microRNA expression<sup>21,22</sup>. Therefore, we hypothesized that these proteins may also regulate the microRNAs of hESCs and DEC.



**Figure 4.1: Validation of differentially expressed proteins by immunofluorescence and WB. (A,B)** The expression of hnRNP A1 in the H9 (A) and H1 (B) hESCs as well as the definitive endoderm cells derived from them was showed by immunofluorescence staining. Nuclei were stained by DAPI

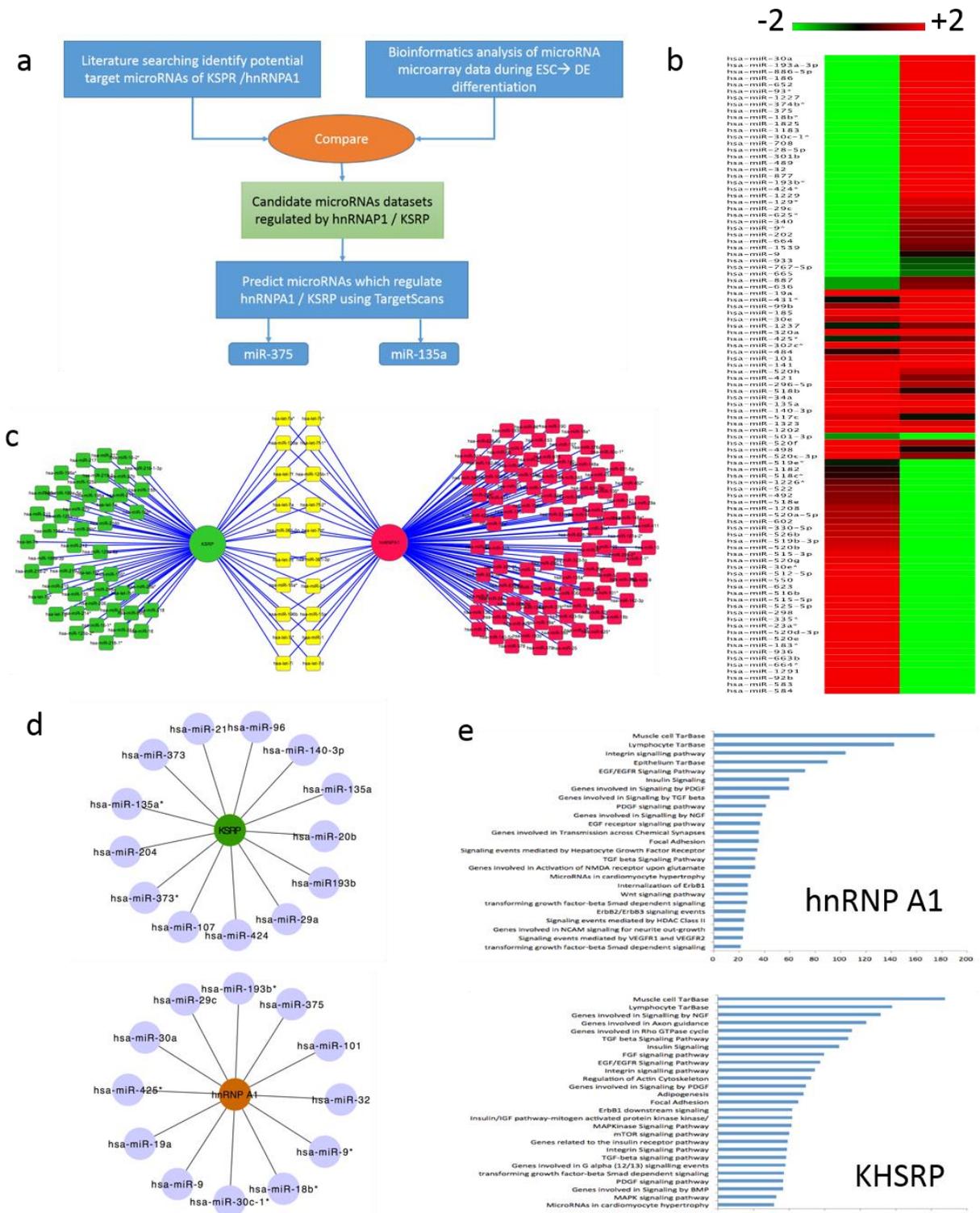
(Bars: 100 $\mu$ m). (C, D) The expression of KSRP in the H9 (C) and H1 (D) hESCs as well as the definitive endoderm cells derived from them was showed by immunofluorescence staining. Nuclei were stained by DAPI (Bars: 100 $\mu$ m). (E, F) Western blotting was performed to analyze the protein expression of hnRNP A1 and KSRP in the H9 (E) and H1 (F) hESCs as well as the definitive endoderm cells derived from them, using  $\beta$ -actin as an internal standard. (G) Western blotting was performed to analyze the protein expression of hnRNP A1 and KSRP in hESCs and embryonic body cells, using  $\beta$ -actin as an internal standard.

#### **4.3.2. Bioinformatics analysis predicted potential hnRNPA1 and KSRP target microRNAs**

To find out the potential microRNAs regulated by hnRNPA1 and KSRP during DE differentiation process, we adopted a strategy schematically illustrated in Figure 4.2A. In order to investigate the microRNA behaviors during the DE differentiation process, we carried out microRNA microarray experiments, and identified that miR-30a, miR-186 and miR-375 were significantly up-regulated while miR-584, miR-583 and miR-936 were down-regulated during the DE differentiation process. Figure 4.2B summarizes the microRNA expression profile and correlation with hnRNPA1 and KSRP.

Secondly, we collected potential microRNA targets of hnRNPA1 and KSRP<sup>18,20,23</sup> and constructed a predicted microRNA targets set of hnRNPA1/KSRP by literature review and TargetScan analysis (Figure 4.2C). Comparing the measured correlation microRNA dataset and predicted microRNA dataset, we identified the microRNAs, such as miR373, miR135a, miR29c, miR30a and miR37, potentially regulated by the two proteins as shown in Figure 4.2D.

Further pathway enrichment analysis of candidate target microRNAs showed that the target microRNAs were enriched in multiple signal transduction pathways, such as TGF-Beta, Wnt, FGF, MAPK and SMAD, which are important for hESCs self-renewal and DE differentiation (Figure 4.2E).



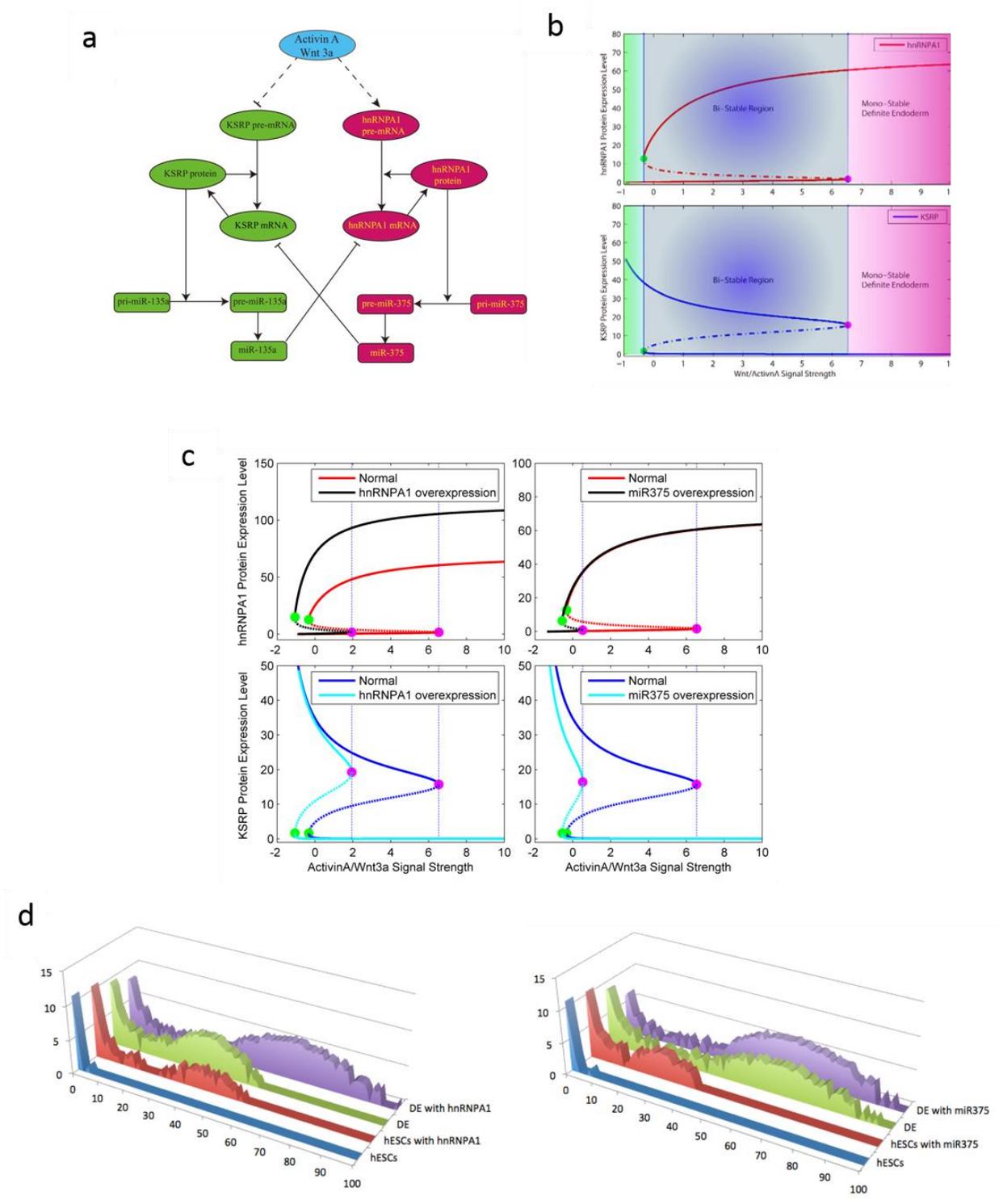
identification process. From literature search, we get the potential microRNA targets of hnRNPA1 and KSRP. Prediction of potential hnRNPA1 and KSRP target microRNAs based on microRNA microarray data expression data during ESC to DE. Intersection of literature predicted microRNA datasets and potential targets based on microRNA microarray analysis. (B) Heatmap depicting microRNA expression profile during hESC to DE differentiation process. (C) Potential microRNA targets of hnRNPA1 and KSRP collected from literatures. Green ones are potential targets of hnRNPA1, red ones are potential targets of KSRP, and yellow ones are potential targets of both proteins. (D) Intersections of predicted targeted microRNAs and significantly expressed microRNAs during hESC to DE differentiation process. (E) Pathway enrichment analysis of candidate target microRNAs dataset from Figure 4.2(A) Top enriched pathways are listed ( $p$ -value  $< 0.01$ ). (KEGG pathway enrichment analysis is performed using the DAVID bioinformatics tool)

### **4.3.3. Mathematical modeling suggests a switch-like motif formed by hnRNPA1-miR375-KSRP-miR135a**

To study the potential mechanism underlying the anti-correlated hnRNPA1 / KSRP pairs, we used *TargetScans* and predicted that miR-375, the candidate targets of hnRNPA1 protein, has target sites in 3'UTR of KSRP protein, and miR-135a, the candidate target of KSRP protein, has target sites in 3'UTR of hnRNPA1 protein.

Integrating the above information and reported functional roles of the two proteins, we constructed a regulatory network shown in Figure 4.3A. KSRP catalyzes maturation of its own mRNA and miR135a. The latter inhibits hnRNPA1 expression through binding its mRNA for degradation. Similarly, hnRNPA1 catalyzes maturation of its own mRNA and miR375. The latter inhibits KSRP expression through binding its mRNA for degradation. Through the Activin A/Wnt pathways, the stimulation signal Activin A/Wnt3a (AW) activates hnRNPA1 and/or inhibits KSRP, with the exact molecular mechanism to be determined. A high hnRNPA1 and low KSRP expression pattern may induce ESC-to-DE differentiation. The overall network has the structure of a toggle switch, with auto-activation of hnrnPA1 and KSRP, and mutual

inhibition between them. This network structure can generate bistable behavior (Figure 4.4B): the system can be in two types of steady states, high hnrnPA1/low KSRP (state 1), or low hnrnPA1/high KSRP (state 2); the system remains at state 1 until the stimulation signal reaches a threshold, where it jumps to state 2; within the bistable region, the two states coexist. The supporting text gives more details of the model. We also performed parameter sensitivity analysis to show that the model can generate bistable behavior over a range of parameter values. Alternatively, the network can generate a sigmoidal shaped switch-like stimulation-response curve, as exemplified in Figure 4.9, which again predicts that cells differentiate only under signal concentrations above a threshold value. Indeed it has been well documented that a threshold concentration of Activin A exists on inducing ESC-to-DE differentiation<sup>24-27</sup>.



**Figure 4.3 Post-transcriptional regulatory network regulates the phenotypic transition.** (A) Schematic depiction of the identified regulatory network. Activin A/Wnt3a signals up-regulate hnRNPA1 module and down-regulate KSRP module. hnRNPA1/KSRP protein has self-activation as to promote the alternative splicing of their pre-mRNAs.

*hnRNPA1/KSRP protein could promote the biogenesis of miR-375/miR-135a, on the other hand, miR-375/miR-135a could repress KSRP/hnRNPA1 mRNAs. (B) Bifurcation diagram of hnRNPA1 (Upper panel) and KSRP (lower panel) on the level of exogenous Activin A/Wnt3a signal level. Bifurcation analysis demonstrates that the system undergoes a two state transition process during hESC to DE differentiation. (C) Influence of hnRNPA1/miR-375 overexpression on the system dynamics. The bifurcation point shift left upon hnRNPA1 / miR-375 synthesis rate increment. (D) Population level stochastic simulation demonstrates the influence of exogenous WA signal and hnRNPA1/miR-375 overexpression on the system. We use hnRNPA1 low, KSRP high denotes ESC state, and hnRNPA1 high, KSRP low denotes DE state. The two hnRNPA1 level distributions reflect the cell phenotype distribution under different conditions.*

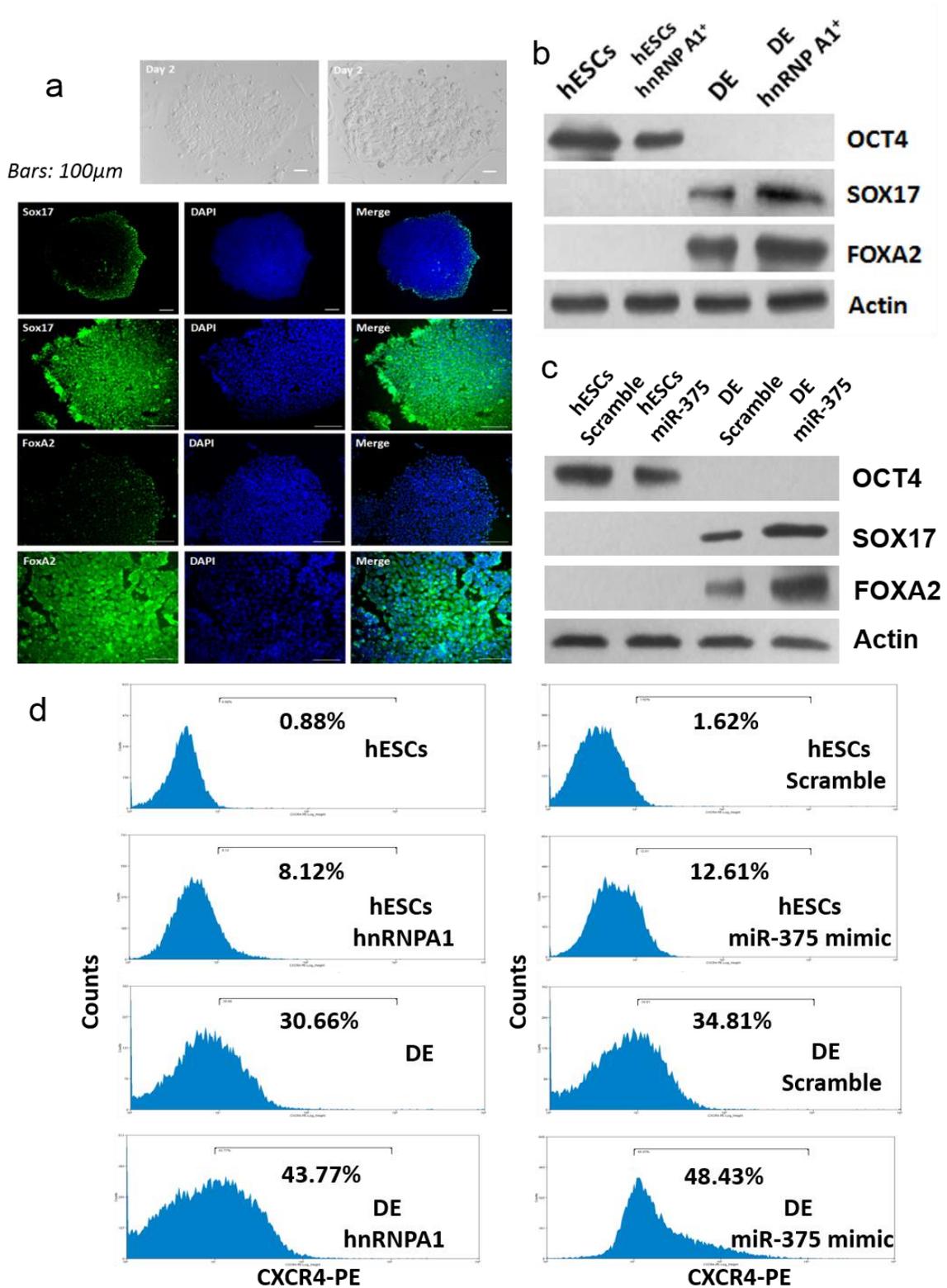
#### **4.3.4. Priming hESCs with AW and hnRNP A1 or miRNA-375 induces rapid and efficient differentiation into the DE**

If the above network regulates ESC differentiation to DE, varying the level of one or more components of the network should affect the differentiation process. Figure 4.3C shows a model prediction that increasing hnRNPA1/miR-375 synthesis rate leads to transition from the ESC branch to DE branch at a lower AW concentration. Considering cell-to-cell variation, we further simulated the model with each model parameter randomly and uniformly chosen with Latin Hypercube sampling centered around the current parameter selection<sup>28</sup>. Figure 4.3D shows simulation results with 2000 cells. The first two lanes show that the ESC differentiation efficiency increases with higher hnRNPA1 (left panel) / miR-375 (right panel) synthesis rate. The backward two lanes demonstrate that at a fixed AW concentration, the colony with increased hnRNPA1 level has more cells differentiated to DE. Similar results were obtained with increased miR375 level.

To test this prediction experimentally, H9 hESCs were treated with AW alone or AW plus hnRNP A1, and were transferred to media that is known to specify differentiation into the endoderm. The number of DECs in culture was assessed morphologically, by their polygonal shape and prominent nucleoli, and confirmed by Sox17 and FoxA2 staining and western blotting. Cells were maintained in culture for 2 days and then underwent western blotting to assess protein expression. Cells primed with AW plus hnRNP A1

exhibited endoderm morphology by day 2 (Figure 4.4A Upper), while cells primed with AW alone exhibited endoderm morphology after 3 days in culture. Furthermore, on day 2, hESCs primed with AW and hnRNP A1 yielded more Sox17-positive and FoxA2-positive endoderm (Figure 4.4A Middle and Lower) than hESCs primed with AW alone, which was confirmed by western blotting (Figure 4.4B). In this study, we also used a cell surface marker, CXCR4, to identify the definitive endoderm population. The CXCR4 can be used to distinguish early definitive endoderm in both human and mouse ESCs. To investigate the expression of CXCR4, we used FACS analysis to sort fractions using PE-conjugated sorting antibodies. FACS analysis showed that on day 2, CXCR4 expression in hESCs primed with AW plus hnRNP A1 was greater than CXCR4 expression in hESCs primed with AW alone (Figure 4.4D Left). Our results show that priming hESCs with AW plus hnRNP A1 indeed promotes rapid and efficient differentiation into DECs, consistent with the model predictions in Figure 4.3D.

We further performed experiments with miRNA-375. H9 hESCs were treated with AW plus scramble mimics or AW plus miR-375 mimic, and were transferred to media that is known to specify differentiation into the endoderm. Cells were maintained in culture for 2 days and then underwent western blotting to assess protein expression. On day 2, hESCs primed with AW and miR-375 yielded more expression of Sox17 and FoxA2 than hESCs primed with AW and Scramble mimics, which was confirmed by western blotting (Figure 4.4C). We also detect the expression of CXCR4 to identify the definitive endoderm population and we used FACS analysis to sort fractions using PE-conjugated sorting antibodies. FACS analysis showed that on day 2, CXCR4 expression in hESCs primed with AW plus miR-375 mimic was greater than CXCR4 expression in hESCs primed with AW plus Scramble mimics (Figure 4.4D Right). Our results show that priming hESCs with AW plus miR-375 indeed promotes rapid and efficient differentiation into DECs, consistent with the model predictions in Figure 4.3D.



**Figure 4.4** High hnRNP A1 and microRNA-375 level drive definitive endoderm differentiation from hESCs. (A) Morphology of hESCs-derived definitive endoderm differentiation by 2 days with AW and hnRNP A1 or AW alone

(upper, Bars: 100 $\mu$ m). The expression of Sox17 and FoxA2 in H9 and H1 hESCs primed with AW alone or AW plus hnRNP A1 was showed by immunofluorescence staining. Nuclei were stained by DAPI (middle and lower, Bars: 100 $\mu$ m). (B) Western blotting was performed to analyze the protein expression of Oct4, Sox17 and FoxA2 in the hESCs primed with AW alone or AW plus hnRNP A1, using  $\beta$ -actin as an internal standard. (C) Western blotting was performed to analyze the protein expression of Oct4, Sox17 and FoxA2 in the hESCs primed with AW plus scramble mimics or AW plus miR-375 mimic, using  $\beta$ -actin as an internal standard. (D) FACS analysis showed CXCR4 expression in hESCs primed with AW alone or AW plus hnRNP A1 (left), and in hESCs primed with AW plus scramble mimics or AW plus miR-375 mimic (right).

#### 4.4. Discussion

For this study, we focused on proteins that are involved in or related to differentiation. Two proteins, hnRNP A1 and KSRP, showed significantly different expression levels depending on the state of cell differentiation (hESC or DE). The presence of hnRNP A1 and KSRP was confirmed by immunocytochemistry and immunoblotting (Figure 4.1 and Figure 4.2). To our knowledge, this is the first study to identify differentially expressed proteins using proteomic techniques, and to suggest a role for hnRNP A1 and KSRP during differentiation of hESCs into the DE. Notably, microRNAs (miRNAs) are endogenous single-stranded RNAs that base-pair with target mRNAs to negatively regulate their expression. Production of miRNA is a tightly regulated process that can be modulated at different steps during the biogenesis pathway<sup>29</sup>. One step related to the discussion in this work is that pri-miRNA precursors need to be processed by RNase enzymes such as Drosha. Failure of this step correlates with cancer progression<sup>30</sup>.

The hnRNP A1 protein, which has well-known functions in nucleocytoplasmic shuttling and mRNA metabolism, has been found to play a role in the production of miRNAs. Guil and Cáceres recently reported that hnRNP A1 specifically binds to a miRNA cluster containing miR-18a and facilitates the Drosha-mediated processing of miR-18a, but does not affect any other member of the cluster<sup>31</sup>. They showed that hnRNP A1 binds to the conserved terminal loop of pri-miR-18a and induces a relaxation at the stem, creating a more favorable cleavage site for Drosha. Furthermore, they found that 14% of all human pri-miRNAs have terminal loops that are well conserved throughout evolution. This group also reported that

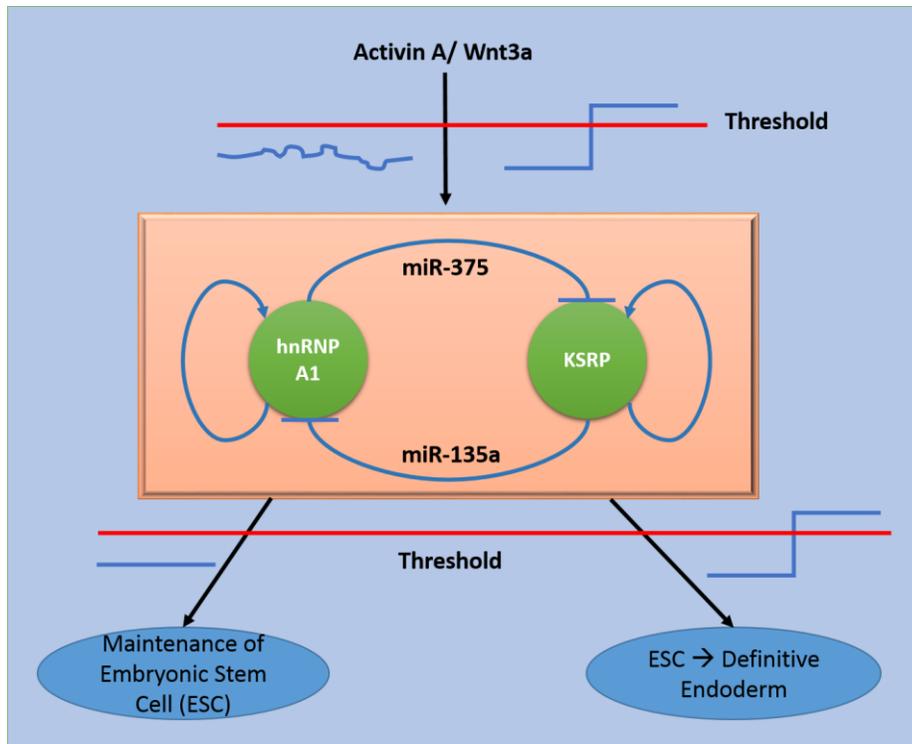
hnRNP A1 binds the terminal loop of pri-let-7a miRNA<sup>18</sup>, which harbors a perfect hnRNP A1 consensus binding site (UAGGGA/U). Thus, they suggest that hnRNP A1 could act as a landing pad for *trans*-acting factors that regulate biogenesis of miRNAs. It is also of interest that hnRNP A1 has been shown to have roles in the maturation of certain miRNAs, including miR-18a and let7a<sup>32</sup>.

The splicing factor KH-type splicing regulatory protein (KSRP) contains four K homology RNA-binding domains and promotes exon inclusion of the c-src alternative exon through an intronic splicing enhancer<sup>33</sup>. KSRP has recently been shown to be a component of both Drosha and Dicer complexes, whereby it positively regulates the biogenesis of a subset of miRNAs, including miR-155 and let-7<sup>34,35</sup>. Altogether, there is ample evidence to suggest the existence of *trans*-acting factors that bind conserved terminal loops can influence the processing of specific miRNAs. The remaining challenges in this realm will involve elucidating the regulatory networks involved in controlling inducible splicing events and exploring the involvement of candidate miRNAs that have recently been implicated in regulating hESC differentiation toward the endoderm. Powerful tools are now available to further explore functional links between splicing factors and miRNA.

The two microRNA, miR-135a and miR-375, has also been widely studied for the embryonic stem cell differentiation. miR-135a has been identified to target FoxO1<sup>36</sup>, which is essential for maintenance of human stem cell and known to participate in suppressing the mesoderm and endoderm lineage commitment<sup>37</sup>. While miR-375, is known to be the hallmark of definitive endoderm. It is highly up-regulated in endoderm and its expression level declines later in the differentiation process<sup>38-40</sup>. How miR-375 participated in the endoderm differentiation process is not very clear.

Our combined bioinformatics analysis and mathematical modeling suggest that the two enzymes antagonize each other. This miRNA mediated antagonization, which leads to either bistable or sigmoidal switch-like behavior in response to differentiation signaling molecules, could function in different ways, as summarized in Figure 4.5. When the extrinsic differentiation signal (AW) is above the threshold, the cell differentiates into DE, otherwise the stem cell remains self-renewal. The ESC state needs to be delicately maintained. Accumulating evidences suggest that an ESC cell stochastically transit among subphenotypes

carrying different competing lineage properties. It is intriguing how an ESC can maintain its stemness despite the large fluctuations of competing transcription factor expressions. Our work suggest that antagonizing enzyme pairs can serve as a noise filter to set a threshold of signal strength on differentiation initiation, and prevent inadvertent differentiation due to transient fluctuations from the Activin A/Wnt3a signaling pathway.

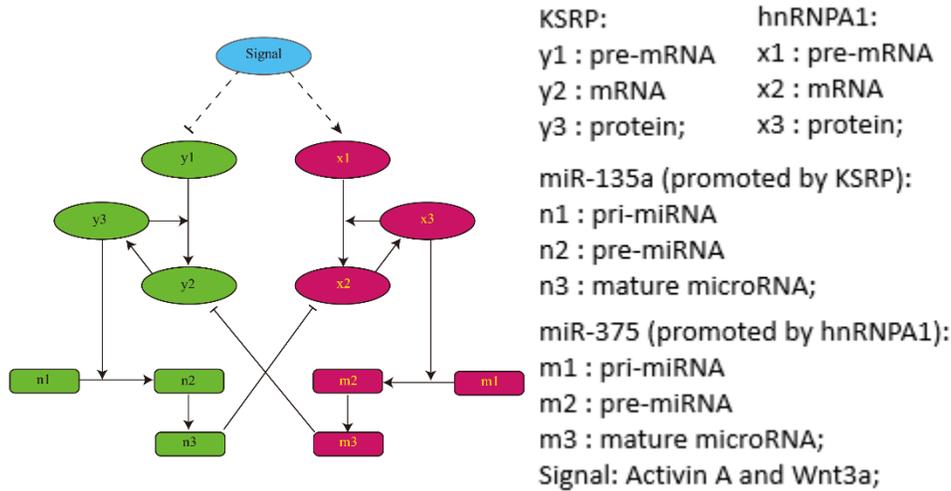


**Figure 4.5** A post-transcription level toggle switch may act as a noise filter during the differentiation process. Due to its property of switch-like behavior, the module could filter out extrinsic noise to help maintain the ESC self-renewal process (shown on the left side of the figure). The module may further integrate the exogenous Activin A/Wnt3a signals into the post-transcriptional level regulatory network and facilitate the hESC to DE differentiation (shown on the right side of the figure).

The present work is the first direct proteomic comparison of hESCs and their differentiation into the definitive endoderm by 2D-DIGE proteomics. The 2D-DIGE proteomic approach coupled with mass spectrometry proved to be an effective tool for studying differentiation into the definitive endoderm by allowing us to identify specific proteins implicated in this process. In particular, hnRNP A1 and KSRP were

confirmed to have significant differential expression, suggesting their importance in the differentiation process and thus highlighting their potential for use as biomarkers during definitive endoderm derivation from hESCs. Bioinformatics and mathematical modeling further suggested these two proteins as part of a network motif buffering noises leading to ESC differentiation.

#### 4.5. Model and methods

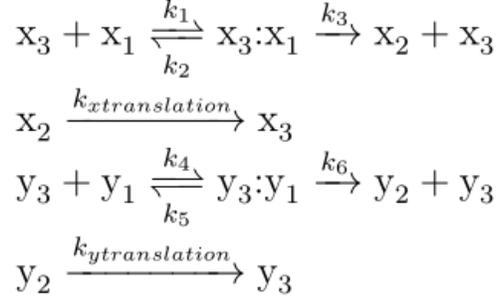


**Figure 4.6 Identified post-transcriptional regulatory network.** For simplicity, different molecules are represented by  $x$ ,  $y$ ,  $m$ ,  $n$  as illustrated on the right panel. This network is the same as Figure 4.3A.

The post-transcriptional regulatory network we identified is shown in Figure 4.6. RNA binding protein KSRP and hnRNPA1 could bind to primary microRNA of miR-135a and miR-375 respectively and facilitate the maturation of the microRNA. Meanwhile, it is assumed that the two RNA binding protein could promote the alternative splicing of their own mRNA.

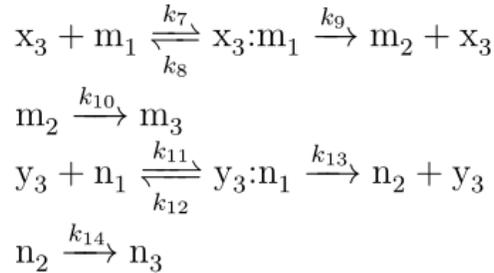
The network motif shown in Figure 4.6 can be decomposed into 3 modules:

(1) X, Y protein self-activation: Proteins hnRNPA1 and KSRP are known to facilitate mRNA alternative splicing events, and hnRNPA1 is known to have self-splicing<sup>3</sup>. Here we assume that KSRP also binds to its own pre-mRNA and facilitate the alternative splicing events. Therefore, these two proteins bind to their own pre-mRNAs and promote mRNAs maturation, and have self-activation behavior.



Here  $x_1, x_2, x_3$  represent pre-mRNA hnRNA1, mRNA hnRNPA1, and protein hnRNPA1;  $y_1, y_2, y_3$  represent pre-mRNA KSRP, mRNA KSRP, and protein KSRP;  $x_3:x_1 / y_3:y_1$  represent the enzyme binding intermediate states;  $k_1, k_2, k_4, k_5$  represent enzyme binding on/off rates;  $k_3, k_6$  represent enzymatic reaction rate;  $k_{xtranslation}$  and  $k_{ytranslation}$  represent the mRNA translation rates.

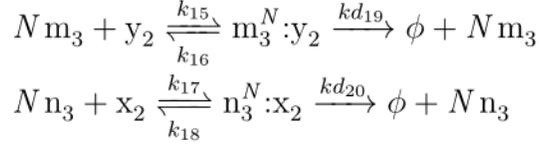
(2) X, Y proteins promote microRNA biogenesis: hnRNPA1 and KSRP are RNA binding proteins. They are also well known to bind to pri-microRNAs to facilitate Dicer / Drosha cutting events and promote the microRNA maturation <sup>41</sup>. From *in silico* predication and experiment report, we get ~150 microRNAs as potential microRNA targets of the two proteins <sup>42,43</sup>. Combined with our microRNA microarray experiment results, we assume that hnRNPA1 binds to pri-miR375 and promotes miR-375 biogenesis, and KSRP binds to pri-miR135a and promotes miR-135a biogenesis.



Here  $m_1 / n_1$  represent pri-miR375 / pri-miR135a,  $x_3:m_1 / y_3:n_1$  represent the enzyme binding complex,  $m_2 / n_2$  represent pre-miR375 / pre-miR135a,  $m_3 / n_3$  represent miR375 / miR135a.  $k_7, k_8, k_{11}, k_{12}$  represent enzyme binding on/off rates;  $k_9, k_{13}$  represent enzymatic reaction rates and  $k_{10}$  and  $k_{14}$  are pre-microRNA processing rates to mature microRNA.

(3) Mutual Inhibition: miR-375 and miR135a silence KSRP and hnRNPA1 at post-transcription level.

After the microRNA binds to the mRNA and form microRNA:mRNA complex, the mRNA will be degraded and the microRNAs will be fully recycled.



Here  $N$  is the number of microRNA binding sites number on the mRNA for protein KSRP and hnRNPA1.  $k_{15}$ ,  $k_{16}$ ,  $k_{17}$ ,  $k_{18}$  are enzyme binding on/off rates,  $Kd_{10}$  and  $Kd_{20}$  are microRNA:mRNA complex turnover rates. Here we assume  $N$  microRNAs will cooperatively bind to the mRNA target and lead to the degradation of the mRNAs, all microRNAs are recycled. We do not discuss the intermediate states of the intermediate state of mRNA bound by  $1 \dots N-1$  microRNA(s) and the degradation resulting from the intermediate states.

The corresponding rate equations are listed below.

$$\frac{d[x_1]}{dt} = k_{0x_1} + k_{x_1} * S / (Kx + S) - k_1 * [x_3] * [x_1] + k_2 * [x_1x_3] - k_{dx_1} * [x_1]$$

$$\frac{d[x_1x_3]}{dt} = k_1 * [x_1] * [x_3] - (k_2 + k_3) * [x_1x_3]$$

$$\frac{d[x_2]}{dt} = k_{0x_2} * [x_1] + k_3 * [x_1x_3] - k_{17} * [n_3]^N * [x_2] + k_{18} * [n_3][x_2] - k_{dx_2} * [x_2]$$

$$\frac{d[x_3]}{dt} = k_{xtrs} * [x_2] - k_1 * [x_1] * [x_3] + (k_2 + k_3) * [x_1x_3] - k_7 * [x_3] * [m_1] + (k_8 + k_9) * [x_3m_1] - k_{dx_3} * [x_3]$$

$$\frac{d[n_3x_2]}{dt} = k_{17} * [n_3]^N * [x_2] - k_{18} * [n_3x_2] - k_{d20} * [n_3x_2]$$

$$\frac{d[y_1]}{dt} = k_{0y_1} + k_{y_1} * Ky / (Ky + S) - k_4 * [y_3] * [y_1] + k_5 * [y_1y_3] - k_{dy_1} * [y_1]$$

$$\frac{d[y_1y_3]}{dt} = k_4 * [y_3] * [y_1] - (k_5 + k_6) * [y_1y_3]$$

$$\frac{d[y_2]}{dt} = k_{0y_2} * [y_1] + k_6 * [y_1y_3] - k_{15} * [m_3]^N * [y_2] + k_{16} * [m_3y_2] - k_{dy_2} * [y_2]$$

$$\frac{d[y_3]}{dt} = k_{ytrs} * [y_2] - k_4 * [y_3] * [y_1] + (k_5 + k_6) * [y_1y_3] - k_{11} * [y_3] * [n_1] + (k_{12} + k_{13}) * [y_3n_1] - k_{dy_3} * [y_3]$$

$$\frac{d[m_3y_2]}{dt} = k_{15} * [m_3]^N * [y_2] - k_{16} * [m_3y_2] - k_{d19} * [m_3y_2]$$

$$\begin{aligned}
\frac{d[m_1]}{dt} &= k_{0m_1} - k_7 * [x_3] * [m_1] + k_8 * [x_3m_1] - k_{dm_1} * [m_1] \\
\frac{d[x_3m_1]}{dt} &= k_7 * [x_3] * [m_1] - (k_8 + k_9) * [x_3m_1] \\
\frac{d[m_2]}{dt} &= k_{0m_2} * [m_1] + k_9 * [x_3m_1] - k_{10} * [m_2] - k_{dm_2} * [m_2] \\
\frac{d[m_3]}{dt} &= k_{10} * [m_2] - N * k_{15} * [m_3]^N * [y_2] + N * k_{16} * [m_3y_2] - k_{dm_3} * [m_3] + N * k_{d19} * [m_3y_2] \\
\frac{d[n_1]}{dt} &= k_{0n_1} - k_{11} * y_3 * [n_1] + k_{12} * [y_3n_1] - k_{dn_1} * [n_1] \\
\frac{d[y_3n_1]}{dt} &= k_{11} * [y_3] * [n_1] - (k_{12} + k_{13}) * [y_3n_1] \\
\frac{d[n_2]}{dt} &= k_{0n_2} * [n_1] + k_{13} * [y_3n_1] - k_{14} * [n_2] - k_{dn_2} * [n_2] \\
\frac{d[n_3]}{dt} &= k_{14} * [n_2] - N * k_{17} * [n_3]^N * [x_2] + N * k_{18} * [n_3x_2] - k_{dn_3} * [n_3] + N * k_{d20} * [n_3x_2]
\end{aligned}$$

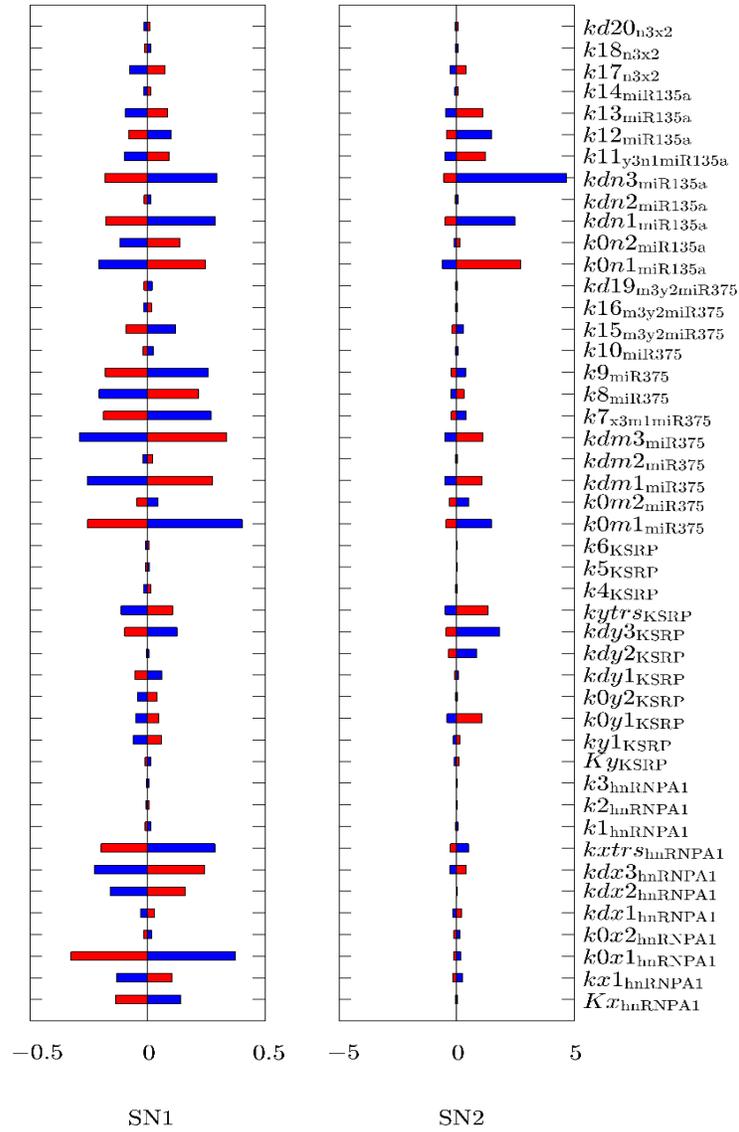
Here [X] denotes the concentration of molecule X, other parameters' meaning are listed in Table 4.1.

## 4.6. Parameter robustness and sensitivity analysis

To demonstrate that the model can generate robust bistable behavior, not due to fine-tuning of parameters, we performed sensitivity analysis against different parameter selections.

### 4.6.1. Local Sensitivity Analysis:

We first performed local parameter sensitivity analysis by perturbing each parameter by 15% around the control parameter set to examine the corresponding impact on the bifurcation threshold. From Figure 4.7, we see that the system still have two bifurcation points. That is, the switch behavior is robust to parameter variation. Specifically, the bifurcation points are not sensitive to the enzyme binding rates, but sensitive to the mRNA and miRNA synthesis/degradation rates.



**Figure 4.7** Local parameter sensitivity of the two bifurcation points. SN1 and SN2 denote saddle node bifurcation points on the left and right respectively of Figure 4.3B. Blue bar and red bar show the percentage change of exogenous AW signal in the presence of a 15% decrease and increase of the parameter value respectively.

#### 4.6.2. Global Sensitivity Analysis:

We used Latin-Hypercubic sampling to draw 1,000,000 parameter sets from the parameter space as shown in Table 4.1. We run the simulation with 2 different initial conditions,  $X_{on}Y_{off}$  and  $X_{off}Y_{on}$ , for each parameter set. After the system reaches a steady state, we consider the system as bi-stable and the

corresponding parameter set as good parameter set, if  $[X]_{ss(XonYoff)} - [X]_{ss(XoffYon)} > \lambda$ . Here we use  $\lambda = 5$ . We finally get around 5,000 good parameter sets.

Since the half-life of mRNA, protein, miRNA are in 2h-20h scale<sup>44 45</sup>, the parameters regarding the turnover rate ( $k_{dx1}, k_{dx2}, k_{dx3}, k_{dy1}, k_{dy2}, k_{dy3}, k_{dm1}, k_{dm2}, k_{dm3}, k_{dn1}, k_{dn2}, k_{dn3}$ ) are assigned with a range of [0.035 0.35], with the unit of  $\text{hour}^{-1}$  ( $\text{hr}^{-1}$ ), other parameter ranges are based on these turnover rates. The concentration unit is “ $\mu\text{M}$ ” and time unit is “hour”. Table 4.1 shows the parameter ranges we used in the sensitivity analysis.

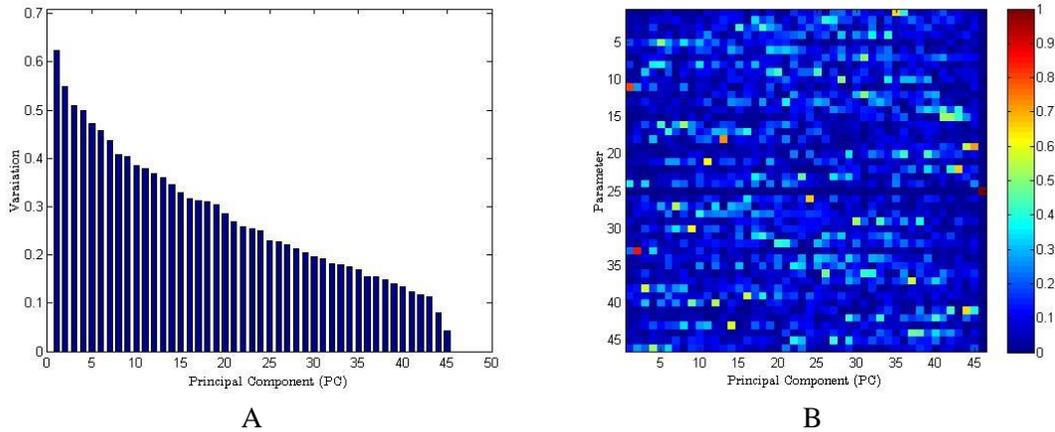
**Table 4.1** Parameter description and corresponding control value and uncertainty range

Parameters	Parameter Description	Control Value	Parameter uncertainty range
$k_{dx1}, k_{dx2}, k_{dx3}$	mRNA/miRNA/protein turnover rate	0.09,0.09,0.17	[0.035, 0.35]
$k_{dy1}, k_{dy2}, k_{dy3}$		0.09,0.09,0.17	
$k_{dm1}, k_{dm2}, k_{dm3}$		0.035, 0.035, 0.035	
$k_{dn1}, k_{dn2}, k_{dn3}$		0.035, 0.035, 0.035	
$K_{d19}, k_{d20}$	miRNA:mRNA complex turnover rate	1, 1	[0.035, 0.35]
$k_{0x1}, k_{0y1}$	Basal mRNA transcription rate	0.3, 0.2	[0.1, 1]
$k_{0x2}, k_{0y2}$	Basal pre-mRNA processing rate	0.05, 0.05	[0.01 0.1]
$k_{xtrs}, k_{ytrs}$	mRNA translation rate	1.5, 1.5	[0.5-5]
$k_{0m1}, k_{0n1}$	Basal miRNA transcription rate	0.6, 0.6	[0.1, 1]
$k_{0m2}, k_{0n2}$	Basal Transition rate from pri-miRNA to pre-miRNA	0.001, 0.001	[0, 0.01]
$k_{10}, k_{14}$	Transition rate from pre-miRNA processed to miRNA	0.5, 0.5	[0.1, 1]
$k_1, k_2, k_3$		1, 1, 1	[0.01, 10]

$k_4, k_5, k_6$	Enzyme binding on/off rate (proten x,y binds to pre- mRNAs; proten x,y binds to pri-miRNAs and miRNA binding rates to mRNAs)	1, 1, 1	
$k_7, k_8, k_9$		0.01, 0.2, 0.01	
$k_{11}, k_{12}, k_{13}$		0.01, 0.2, 0.01	
$k_{15}, k_{16}, k_{17}, k_{18}$		1, 2, 1, 2	
$S$	Signal strength	0-10	[0, 6]
$N$	Number of miRNA binding sites to mRNAs	3	[1, 5]
$K_x, K_y$	Coefficient of signal strength	2,2	[0.5, 5]
$K_{x1}, K_{y1}$	Coefficient of signal strength	0.5, 0.2	[0.1, 1]

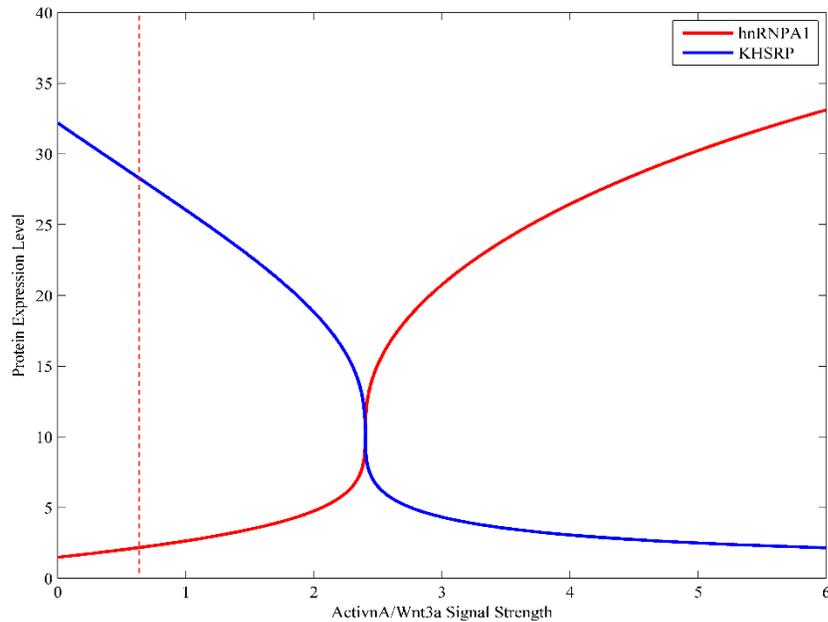
After we get good parameter sets, we nondimensionalized the data with auto-scaling, removed the units of the data and made them comparable. We then performed principal component analysis (PCA).

PCA summarizes variation in the system with uncorrelated principal components (PCs). From the first PC up to the last PC, the variance of the parameter space in that direction decreases as shown in Figure 4.8A. The first several PCs specify the direction with the largest variance or “Sloppiest”/most robust direction. The last several PCs specify the direction with the smallest variance or “stiffest” /most sensitive direction. For sensitivity analysis, we mainly look for the stiffest direction. However, the PCA analysis results shown in Figure 4.8A didn’t show a significant separation among different directions. We evaluated the contribution of each parameter to the system based on PCA loading analysis as shown in Figure 4.8B. From the last several columns of Figure 4.8B, we get similar conclusion to the local sensitivity analysis. The parameters representing mRNAs/miRNAs synthesis rates are the most sensitive parameters.



**Figure 4.8 Global sensitivity Analysis. Panel A demonstrates variation of all the principal components. Panel B shows the PCA loading factor heatmap for all the parameters. The x axis, from left to right, denotes PC1, PC2, PC3 up to PC46. The y axis, from top to bottom denotes  $K_x$ ,  $K_y$ ,  $k_{0x1}$ ,  $k_{x1}$ ,  $k_1$ ,  $k_2$ ,  $k_{dx1}$ ,  $k_3$ ,  $k_{0x2}$ ,  $k_{17}$ ,  $k_{18}$ ,  $k_{dx2}$ ,  $k_{xrs}$ ,  $k_{dx3}$ ,  $k_{d20}$ ,  $k_{11}$ ,  $k_{12}$ ,  $k_{13}$ ,  $k_{0n1}$ ,  $k_{dn1}$ ,  $k_{0n2}$ ,  $k_{14}$ ,  $k_{dn2}$ ,  $k_{dn3}$ ,  $k_{0y1}$ ,  $k_{y1}$ ,  $k_4$ ,  $k_5$ ,  $k_{dy1}$ ,  $k_6$ ,  $k_{0y2}$ ,  $k_{15}$ ,  $k_{16}$ ,  $k_{dy2}$ ,  $k_{yrs}$ ,  $k_{dy3}$ ,  $k_{d19}$ ,  $k_7$ ,  $k_8$ ,  $k_9$ ,  $k_{0m1}$ ,  $k_{dm1}$ ,  $k_{0m2}$ ,  $k_{10}$ ,  $k_{dm2}$ ,  $k_{dm3}$ . Each grid represents the parameter coefficient of the corresponding parameter in that principal component. We use different colors to denote the magnitude of the principal loading coefficient of each parameter in that direction. The deeper the color, the more that parameter contributes to that principal component.**

#### 4.7. The model could also leads to sigmoidal response under different parameter selections



**Figure 4.9 Sigmoidal transition of the system.** Under different parameter choices, the system can also demonstrate sigmoidal behavior with the increase of AW signal strength. As the signal strength reaches certain threshold value, the system jumps into DE State.

Additional research can uncover how these two proteins act in the signaling pathways and developmental processes that regulate differentiation into the definitive endoderm. In the future, this information may help shape understanding of the underlying differentiation mechanisms and encourage the development of new derivation strategies for hepatocytes.

## References

1. Thomson, J.A. *et al.* Embryonic stem cell lines derived from human blastocysts. *science* **282**, 1145-1147 (1998).
2. Reubinoff, B.E., Pera, M.F., Fong, C.-Y., Trounson, A. & Bongso, A. Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro. *Nature biotechnology* **18**, 399-404 (2000).
3. Blanchette, M. & Chabot, B. Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization. *Embo Journal* **18**, 1939-1952 (1999).
4. Grapin-Botton, A. & Melton, D.A. Endoderm development: from patterning to organogenesis. *Trends in genetics* **16**, 124-130 (2000).
5. Fox, I.J. *et al.* Treatment of the Crigler–Najjar syndrome type I with hepatocyte transplantation. *New England Journal of Medicine* **338**, 1422-1427 (1998).
6. Behbahan, I.S. *et al.* New approaches in the differentiation of human embryonic stem cells and induced pluripotent stem cells toward hepatocytes. *Stem Cell Reviews and Reports* **7**, 748-759 (2011).
7. Boyer, L.A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947-56 (2005).
8. Shu, J. *et al.* Induction of pluripotency in mouse somatic cells with lineage specifiers. *Cell* **153**, 963-75 (2013).
9. Chickarmane, V., Troein, C., Nuber, U.A., Sauro, H.M. & Peterson, C. Transcriptional dynamics of the embryonic stem cell switch. *PLoS Comput Biol* **2**, e123 (2006).
10. Huang, S., Guo, Y.P., May, G. & Enver, T. Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Dev Biol* **305**, 695-713 (2007).
11. Chang, H.H., Hemberg, M., Barahona, M., Ingber, D.E. & Huang, S. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* **453**, 544-7 (2008).
12. Hong, T., Xing, J., Li, L. & Tyson, J.J. A mathematical model for the reciprocal differentiation of T helper 17 cells and induced regulatory T cells. *PLoS Comput Biol* **7**, e1002122 (2011).

13. Séguin, C.A., Draper, J.S., Nagy, A. & Rossant, J. Establishment of endoderm progenitors by SOX transcription factor expression in human embryonic stem cells. *Cell stem cell* **3**, 182-195 (2008).
14. Kanda, S. *et al.* In vitro differentiation of hepatocyte-like cells from embryonic stem cells promoted by gene transfer of hepatocyte nuclear factor 3  $\beta$ . *Hepatology research* **26**, 225-231 (2003).
15. Takenaga, M., Fukumoto, M. & Hori, Y. Regulated Nodal signaling promotes differentiation of the definitive endoderm and mesoderm from ES cells. *Journal of cell science* **120**, 2078-2090 (2007).
16. Hay, D.C. *et al.* Highly efficient differentiation of hESCs to functional hepatic endoderm requires ActivinA and Wnt3a signaling. *Proceedings of the National Academy of Sciences* **105**, 12301-12306 (2008).
17. Martinez, N.J. & Gregory, R.I. MicroRNA gene regulatory pathways in the establishment and maintenance of ESC identity. *Cell Stem Cell* **7**, 31-5 (2010).
18. Michlewski, G., Guil, S., Semple, C.A. & Cáceres, J.F. Posttranscriptional regulation of miRNAs harboring conserved terminal loops. *Molecular cell* **32**, 383-393 (2008).
19. Nielsen, A.F., Leuschner, P.J. & Martinez, J. Not miR-ly a splicing factor: hnRNP A1 succumbs to microRNA temptation. *Nature structural & molecular biology* **14**, 572-573 (2007).
20. Trabucchi, M. *et al.* The RNA-binding protein KSRP promotes the biogenesis of a subset of microRNAs. *Nature* **459**, 1010-1014 (2009).
21. Markovtsov, V. *et al.* Cooperative assembly of an hnRNP complex induced by a tissue-specific homolog of polypyrimidine tract binding protein. *Molecular and Cellular Biology* **20**, 7463-7479 (2000).
22. Bikkavilli, R.K. & Malbon, C.C. Dishevelled-KSRP complex regulates Wnt signaling through post-transcriptional stabilization of  $\beta$ -catenin mRNA. *Journal of cell science* **123**, 1352-1362 (2010).
23. Towbin, H. *et al.* Systematic screens of proteins binding to synthetic microRNA precursors. *Nucleic acids research* **41**, e47-e47 (2013).

24. Gadue, P., Huber, T.L., Paddison, P.J. & Keller, G.M. Wnt and TGF- $\beta$  signaling are required for the induction of an in vitro model of primitive streak formation using embryonic stem cells. *Proceedings of the National Academy of Sciences* **103**, 16806-16811 (2006).
25. Hansson, M. *et al.* A late requirement for Wnt and FGF signaling during activin-induced formation of foregut endoderm from mouse embryonic stem cells. *Developmental biology* **330**, 286-304 (2009).
26. Tahamtani, Y. *et al.* Small molecule induction of human embryonic stem cells towards definitive endoderm. (2013).
27. Xiao, L., Yuan, X. & Sharkis, S.J. Activin A maintains self - renewal and regulates fibroblast growth factor, Wnt, and bone morphogenic protein pathways in human embryonic stem cells. *Stem cells* **24**, 1476-1486 (2006).
28. Tian, X.-J., Zhang, H. & Xing, J. Coupled reversible and irreversible bistable switches underlying TGF $\beta$ -induced epithelial to mesenchymal transition. *Biophysical journal* **105**, 1079-1089 (2013).
29. Treiber, T., Treiber, N. & Meister, G. Regulation of microRNA biogenesis and function. *Thrombosis and haemostasis* **107**, 605 (2012).
30. Li, Q. & Gregory, R.I. MicroRNA regulation of stem cell fate. *Cell Stem Cell* **2**, 195-196 (2008).
31. Guil, S. & Cáceres, J.F. The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. *Nature structural & molecular biology* **14**, 591-596 (2007).
32. Michlewski, G. & Cáceres, J.F. Antagonistic role of hnRNP A1 and KSRP in the regulation of let-7a biogenesis. *Nature structural & molecular biology* **17**, 1011-1018 (2010).
33. García-Mayoral, M.F. *et al.* The structure of the C-terminal KH domains of KSRP reveals a noncanonical motif important for mRNA degradation. *Structure* **15**, 485-498 (2007).
34. Davis, B.N. & Hata, A. Regulation of MicroRNA Biogenesis: A miRiad of mechanisms. *Cell Commun Signal* **7**, 10.1186 (2009).

35. Krol, J., Loedige, I. & Filipowicz, W. The widespread regulation of microRNA biogenesis, function and decay. *Nature Reviews Genetics* **11**, 597 (2010).
36. Moritoki, Y. *et al.* Expression profiling of microRNA in cryptorchid testes: miR-135a contributes to the maintenance of spermatogonial stem cells by regulating FoxO1. *J Urol* **191**, 1174-80 (2014).
37. Zhang, X. *et al.* FOXO1 is an essential regulator of pluripotency in human embryonic stem cells. *Nature Cell Biology* **13**, 1092-U118 (2011).
38. Kim, N. *et al.* Expression profiles of miRNAs in human embryonic stem cells during hepatocyte differentiation. *Hepatol Res* **41**, 170-83 (2011).
39. Wei, R. *et al.* Dynamic expression of microRNAs during the differentiation of human embryonic stem cells into insulin-producing cells. *Gene* **518**, 246-255 (2013).
40. Li, X. MiR-375, a microRNA related to diabetes. *Gene* **533**, 1-4 (2014).
41. Michlewski, G., Guil, S., Semple, C.A. & Caceres, J.F. Posttranscriptional regulation of miRNAs harboring conserved terminal loops. *Mol Cell* **32**, 383-93 (2008).
42. Trabucchi, M. *et al.* The RNA-binding protein KSRP promotes the biogenesis of a subset of microRNAs. *Nature* **459**, 1010-4 (2009).
43. Towbin, H. *et al.* Systematic screens of proteins binding to synthetic microRNA precursors. *Nucleic Acids Res* **41**, e47 (2013).
44. Schwanhausser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337-42 (2011).
45. Gantier, M.P. *et al.* Analysis of microRNA turnover in mammalian cells following Dicer1 ablation. *Nucleic Acids Research* (2011).

# Chapter 5. Mathematical modeling of the epithelial-to-mesenchymal transition

Manuscript in preparation

## 5.1. Abstract

The epithelial-to-mesenchymal transition (EMT) is an important biological process during development, wound healing and cancer metastasis. Tumor progression is modulated by the tumor's microenvironment, which (as is well known) is typically enriched with inflammatory cytokines. Recently, several studies have revealed the crucial roles of inflammatory cytokines in inducing EMT. Cytokine IL-6, when bound to IL-6 receptor (IL-6R) activates the transcription factor STAT3, which down-regulates microRNA-34a (miR34a) and upregulates the expression of Lin28 protein, an inhibitor of the maturation of microRNA Let-7. miR34a and Let-7 directly inhibit the expression of IL-6R and IL-6 respectively. These two double-negative feedback loops, together with a two "core" mutual inhibitions between transcription factor Snail and miR34a and transcription factor ZEB1 and microRNA-200 (miR-200), regulate the IL-6 induced EMT process. We have constructed a mathematical model of these multiple positive feedback loops, and the model suggests that the signaling network proceeds through two intermediate states between the initial epithelial state and the final mesenchymal state. This model prediction suggests that the IL-6 pathway contributes to a spectrum of partial EMT states, which possess both epithelial and mesenchymal features. Furthermore, IL-6 and TGF $\beta$  signaling may work synergistically during EMT. Therefore our model analysis demonstrates that for developing more efficacious cancer therapies, it could be of great clinical importance to fully characterize the dynamic regulation of the composite TGF $\beta$  /IL-6 axis.

## 5.2. Introduction

Epithelial to mesenchymal transition (EMT) is a central biological process required for fetal development and wound healing. Moreover, dysregulated EMT has been linked to fibrosis, tumor progression and metastasis. During this process, cells lose many epithelial characteristics, e.g. close

junctions to neighbor cells, and gain mesenchymal characteristics, e.g. spindle cell shape, loose connections with neighbor cells, which could be reflected and measured by several well-known surface markers, such as E-cadherin, N-cadherin and Vimentin. During the EMT process, E-cadherin is downregulated while N-cadherin and Vimentin are up-regulated.

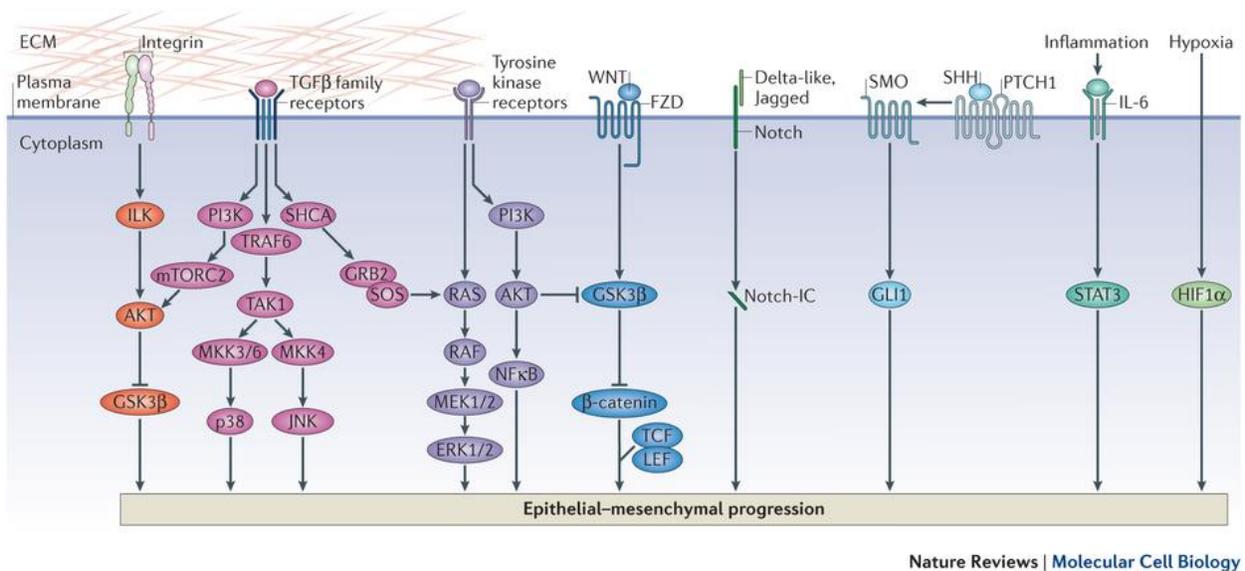
During the EMT process, at least three cell states could be observed: an epithelial state (E), a partial EMT (pEMT) state characterizing the intermediate state, and a mesenchymal (M) state. The pEMT state is believed to be connected to stemness and tumor circulation in the bloodstream, and has attracted intensive attention from the field recently. Actually accumulating evidences suggest that the intermediate state is not a single pEMT state but is composed of multiple intermediate states<sup>1-3</sup>. Huang et al. characterized ovarian cancer cell lines into four different tumor categories (E, Intermediate E, Intermediate M, M) according to EMT genes and cell surface markers (E-cadherin, PCK and Vimentin)<sup>1</sup>. In a recent study, Hong et al. theoretically identified four states in EMT when they add the newly discovered *Ovol2* and *ZEB1* mutual inhibition into earlier models, which they also verified experimentally.<sup>3</sup> These intermediate states may result from crosstalks among different signaling pathways and existence of multiple positive feedback loops in the regulatory network. Moreover, a new study has shown that cells in different states of the EMT spectrum respond differently to therapeutic regimes, which may provide new drug targets in the future<sup>2</sup>.

There are multiple stimuli that can induce EMT<sup>4</sup>, and cytokine interleukin-6 (IL-6) is one of them<sup>5-9</sup>. IL-6 is known to play important roles in immune response<sup>10</sup>. Activated STAT3, which is downstream transcription factor of IL-6 signaling pathway, is found in tumor cells and inflammatory cells, particularly on the invasive edge of tumors<sup>11</sup>. This observation indicates that IL-6 may provide an autocrine and paracrine stimulus for the migration and infiltration of inflammatory cells<sup>11,12</sup>.

Moreover, elevated serum IL-6 concentrations in patients have been associated with tumor progression and prognosis, and advanced stages of various cancers such as multiple myeloma, esophageal squamous cell carcinoma, non-small cell lung carcinoma, renal cell carcinoma, colorectal cancer, prostate cancer, breast cancer and ovarian cancer<sup>13-16</sup>. IL-6 is also found to be closely related to cancer patient survival rate: Higher level of circulating IL-6 indicates worse survival rate in patients with metastatic breast cancer<sup>17</sup>.

Level of secreted IL-6 from monocyte in patients also predicts the survival rate for colorectal carcinoma and squamous cell carcinoma<sup>18,19</sup>.

Over the past decade, cancer immunotherapy has been intensively studied and appears to be a promising tool to cure tumor. IL-6 signaling pathway, as a pathway involved in the immune response, may also be a therapeutic target for different cancers. Therefore, blocking IL-6 signaling (i.e., anti-IL-6 therapy) is a potential therapeutic strategy for cancer characterized by pathological IL-6 overproduction. Preliminary clinical evidence has shown that antibody targeted IL-6 therapy is well tolerated by cancer patients.<sup>15</sup>

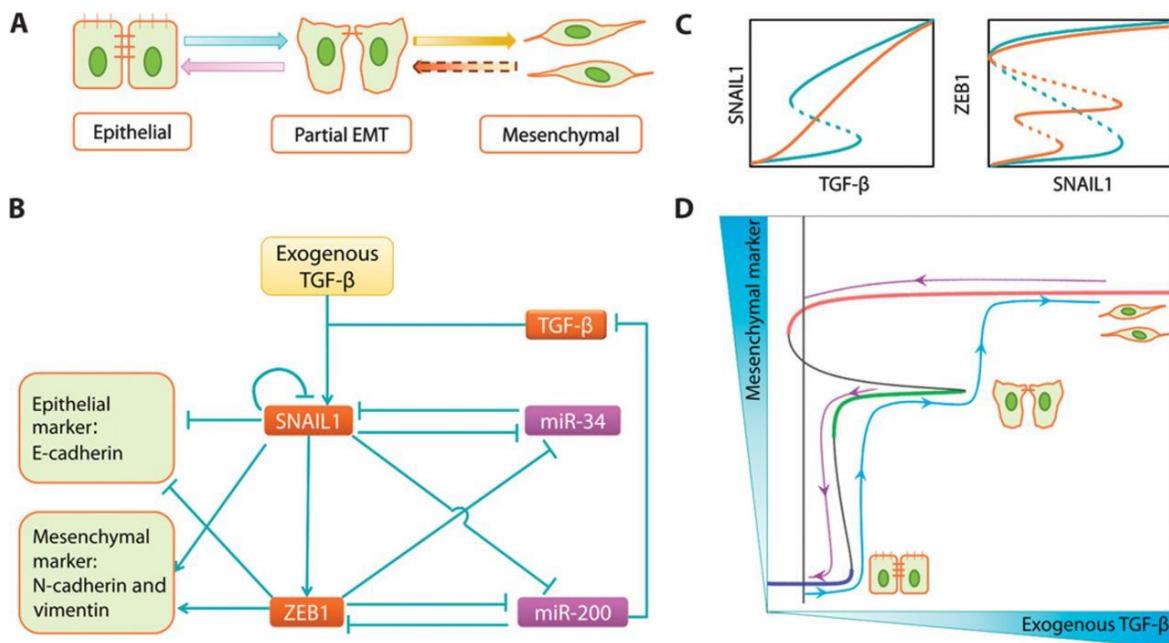


**Figure 5.1: Signaling pathways involved in EMT (from <sup>4</sup>).**

It is well known that the TGFβ/IL-6 axis is crucial for tumor therapy<sup>20</sup>. Due to the critical nature of the cytokines, clinical trials have been performed to test biological therapies targeted toward these signaling molecules. Interestingly, anti-TGF-β therapies alone are not successful for treating tumor progression <sup>21</sup>. Hence, mechanistic study of IL-6 induced EMT and its possible cross-talk with TGFβ may help on improving the treatment efficacy.

Previously we have studied the EMT dynamical process induced by TGFβ. We have predicted and verified the existence of an intermediate pEMT state during the EMT process. We uncovered that a cascade of two positive feedback loops (SNAIL/miR-34a and ZEB1/miR-200) govern the two sequential bistable

switches that regulate the E to pEMT and pEMT to M transition shown in Figure 5.2<sup>22,23</sup>. Recently, Rokavec et al. <sup>8</sup> revealed that one positive feedback loop composed of IL-6R/STAT3/miR-34a could promote EMT in colorectal cancer (CRC) cell lines, and suggested SNAIL may be one downstream effector of STAT3. Also, Liu et al. demonstrated that IL-6 signaling pathway could enhance TGF $\beta$  induced EMT process by activating Smad3 and Smad4, the down-stream effectors of TGF $\beta$  signaling pathway<sup>24</sup>. Here, we begin by studying the IL-6 induced EMT process, aiming to understanding the linkage between inflammation and EMT from a systems perspective.

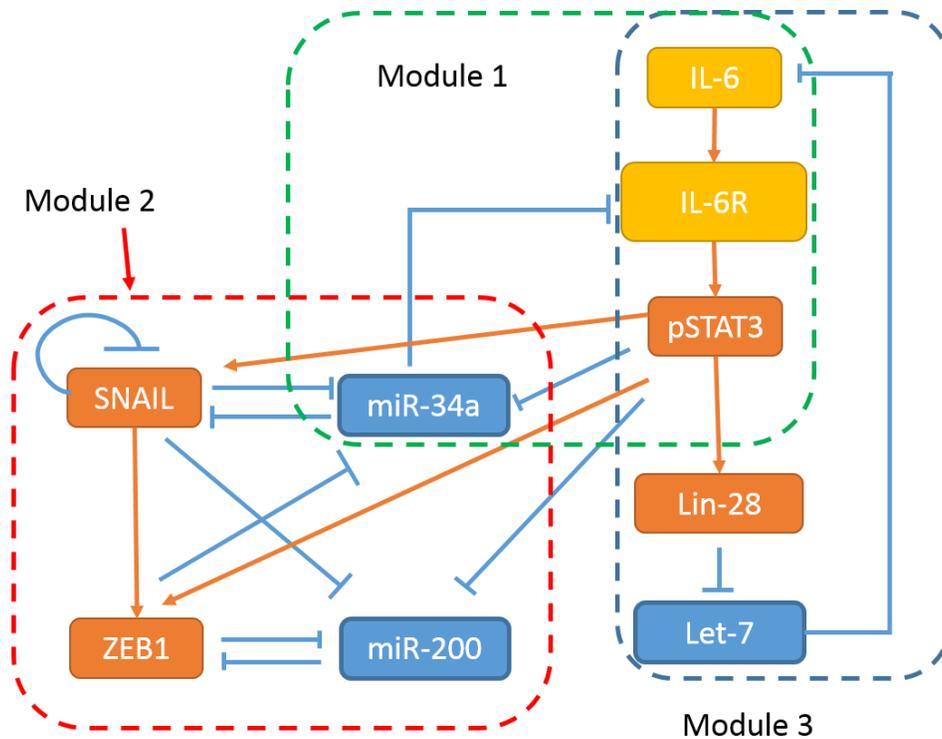


**Figure 5.2: Schematic illustration of TGF- $\beta$ -induced EMT.** (A) Three states of EMT. (B) The core regulatory network of TGF- $\beta$ -induced EMT. The input of the system is exogenous TGF- $\beta$ , which induces SNAIL1 expression. SNAIL1 and miR-34 participate in a double-negative feedback loop (17, 18). SNAIL1 also inhibits its own expression (36). SNAIL1 stimulates expression of ZEB1 and inhibits expression of miR-200. Another similar double-negative feedback loop is formed between ZEB1 and miR-200 (19–21). Furthermore, miR-200 inhibits the autocrine production of TGF- $\beta$  (13), forming another feedback loop. E-cadherin, an epithelial cell marker, is inhibited by SNAIL1 and ZEB1, whereas N-cadherin and vimentin, markers of mesenchymal cells, are promoted by SNAIL1 and ZEB1. (C) Different functional roles of SNAIL1 and ZEB1 modules are proposed in the

*“Cascading Bistable Switches” (CBS) model<sup>23</sup> (blue lines) and “Ternary Chimera Switch” (TCS) model<sup>25</sup> (orange lines). SNAIL1 and miR-34 form a bistable switch (blue line) in the CBS model but a monostable noise filter (orange line) in the TCS model. ZEB1/miR-200 is bistable (blue line) in the CBS model but tristable (orange line) in the TCS model. (D) The predicted bifurcation diagram of CBS model. The thick blue, green, and red lines correspond to E, pEMT, and M states, respectively. The blue and purple arrow lines show how cell fates flip or maintain when the exogenous TGF- $\beta$  dose increases or decreases. Reproduced from <sup>23</sup>.*

### **5.3. Model and methods**

Figure 5.3 gives a schematic diagram of the core regulatory network of IL-6 induced EMT. Cytokine IL-6 binds to the cell membrane receptor IL-6R, which then activates STAT3 to its phosphorylated form. Dimerized phosphorylated STAT3 (pSTAT3) promotes transcription of SNAIL, ZEB1 and Lin28, and represses transcription of miR-34a and miR-200. miR34a can silence IL-6R by binding to the IL-6R mRNA 3'UTR. Thus, IL-6R, pSTAT3 and miR-34a form a double negative feedback loop, shown as module 1 in Figure 5.3. In our previous studies, we analyzed another two pairs of mutual inhibition loops on the core regulatory network of EMT. One mutual inhibition is between SNAIL and miR-34a, and the other is between ZEB1 and miR-200, shown as module 2 in Figure 5.3. A fourth double negative feedback, which could lead to the autocrine production of IL-6, is composed of IL-6/pSTAT3/Lin-28 and Let-7. The RNA binding protein Lin-28 inhibits the maturation of Let-7, while Let-7 silences translation of IL-6 mRNA, shown as module 3 in Figure 5.3. We expect that crosstalks among these feedback loops lead to multistability of the network.



**Figure 5.3: Influence wiring diagram of the gene regulatory network of IL-6 induced EMT based on experimental studies.** IL-6 binds to the IL-6 receptor (IL-6R) on the cell membrane, which then triggers activation of the downstream STAT3. Activated STAT3 serves as a transcriptional factor, which binds to the promoter regions of SNAIL, ZEB, Lin-28 and miR-34a, activate or repress their transcription as shown in the diagram. (Orange arrow indicates activation and blue line with a bar indicates repression). The four proteins in orange promote the EMT process while the three microRNAs in blue antagonize the process. Yellow icons are extracellular input cytokine IL-6 and its receptor IL-6R on the cell membrane.

As shown in the previous paragraph, the network can be decomposed into three different modules. The first one is the IL-6R/STAT3/miR34a module to down-regulate miR34a and activate SNAIL in turn. The second one is the core cascade switches regulated by SNAIL/miR-34a and ZEB1/miR-200 as we studied in the previous papers. The third one is the autocrine module for IL-6, which stabilize the final cell phenotype. We will discuss these three modules in detail below.

### **5.3.1. IL-6R/pSTAT3/miR34a feedback loop:**

Rokavec et al. find that direct repression of SNAIL in DLD-1 (colorectal cancer cell line) cells does not affect the repression of miR-34a by IL-6, but does reduce the expression level of the cell surface marker E-cadherin<sup>8</sup>. They postulate that SNAIL may be one downstream effector of IL-6/STAT3 pathway. They also suggest that transcription factors STAT3 and SNAIL work as an “OR” logic gate for repressing miR34a. Thus, we propose that an immune-related EMT undergoes a STAT3-mediated switch, followed by the cascading switches we demonstrated in the previous papers. Also, in contrast to functioning as a proportional integrator, the switch could filter out noise and link the immune responses with the EMT process.

### **5.3.2. Core mutual inhibitory regulatory motifs of SNAIL/miR-34a and ZEB/miR-200**

The mutual inhibition between SNAIL and miR-34, and between ZEB and miR-200 has been verified in many publications, and we already study this part in our previous model.

### **5.3.3. IL-6 autocrine secretion module (IL-6/pSTAT3/Lin28/Let-7):**

Autocrine of IL-6 can cause multidrug resistance in breast cancer cells<sup>16,26</sup>, and malignant progression in liver cancer<sup>27,28</sup>. It is well documented that IL-6 activates NFkB, which in turn activate IL-6 itself<sup>29,30</sup>. IL-6 can also activate Lin-28, which downregulates Let-7<sup>31</sup>. Lin-28 can bind to let-7 miRNA precursors, both primary microRNA pri-let-7 and precursor microRNA pre-let-7 to inhibit the maturation of Let-7. Furthermore, Lin-28 can increase the decay of the pre-miRNA<sup>32</sup>. Let-7 can degrade the mRNAs of IL-6<sup>27</sup>. Put all together, IL-6 promotes its own synthesis through the IL-6-Lin-28-Let-7 double negative feedback loop. In this work, we only consider the self-activation of IL-6 through STAT3 activate Lin-28 pathway as shown in Figure 5.3.

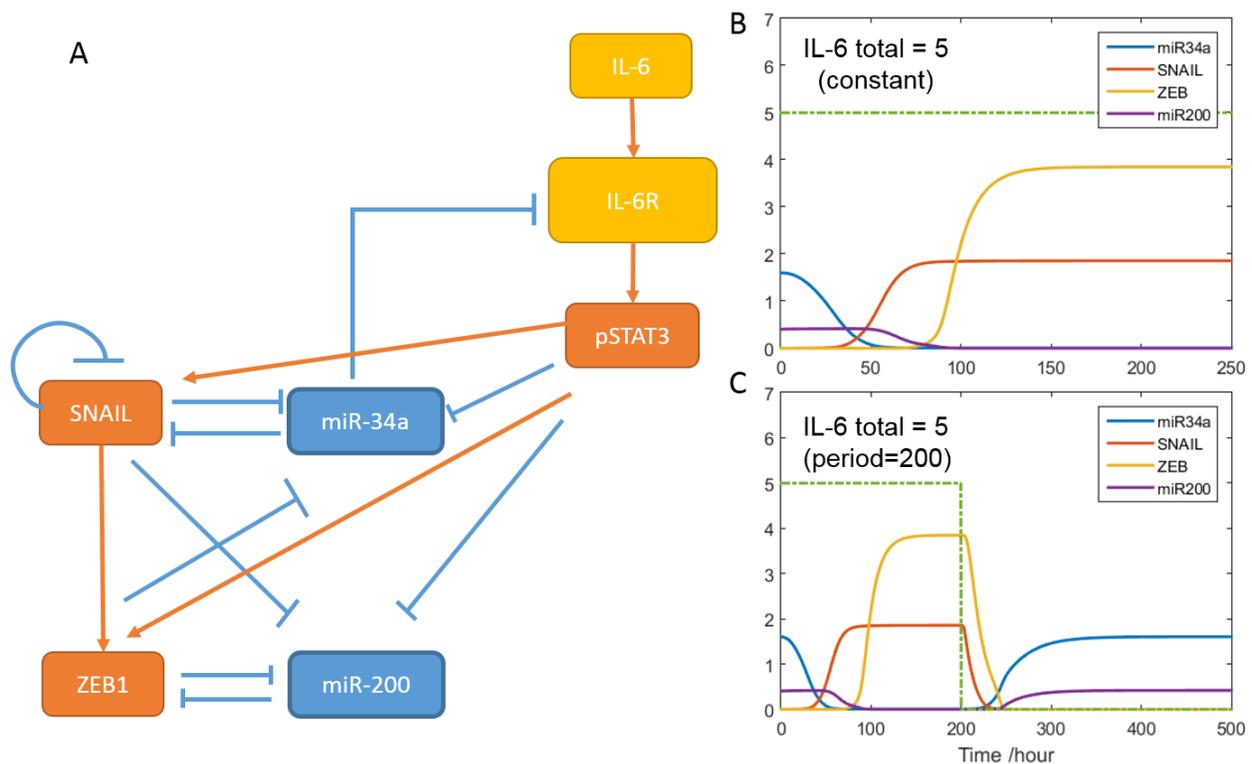
We use ordinary differential equations (ODEs) to model the regulatory network of each module. Similar to our model framework in<sup>23</sup>, we employ Hill equations and mass-action equilibrium equations to model the system. Specifically, microRNA binding sites on their targets are explicitly considered in the model.

Bifurcation analysis was performed with Oscill8 (<http://sourceforge.net/projects/oscill8>). Lists of equations, parameters and assumptions can be found in supplementary materials.

## 5.4. Results

### 5.4.1. IL-6 induced EMT in a sequential manner

We begin the model without the IL-6 autocrine loop as shown in Figure 5.4A. Figure 5.4B clearly demonstrates that in IL-6 induced EMT, hallmark factors of the E state (here, miR-34a and miR-200) are down-regulated, while hallmark factors of the M state (here, SNAIL and ZEB) are upregulated. Similar to the TGF $\beta$  induced EMT, IL-6 induced EMT also has stepwise activation process. From Figure 5.4B, we can see that upon IL-6 treatment, miR-34a decreases over the course of 100 hours, then SNAIL begins to increase. With the increment of SNAIL, miR-200 begins to decrease, which lead to the final upregulation of ZEB. When we only model module 1 and module 2, IL-6 induced EMT is a reversible process (Figure 5.4C). After the removal of the stimulus IL-6, the system reverts back to its initial state. Next, we want to find out whether there exists different thresholds in the sequential activation process.



**Figure 5.4: Activation of the EMT switches by IL-6.** A) Influence wiring diagram of the gene regulatory network of IL-6 induced EMT without module 3. B) When applied a constant stimulus of IL-6 ( $IL-6=5$ ), miR-34a is downregulated and then SNAIL is upregulated. miR-200 turns off later while ZEB turns on lastly. C) When applied the stimulus for a limited period of time ( $t=200$  hours), the system switch from E state to M state, but after the signal is removed, the system reverts to the E state. Initial conditions of different variables was chosen to be E state (miR34/miR-200 high, SNAIL/ZEB low) when  $IL-6=0$ . Detailed description of initial condition can be found in supplemental information. X axis is the time with unit of hour,

#### **5.4.2. IL-6-induced EMT is composed of three bistable switches regulated by the corresponding mutual inhibition pairs**

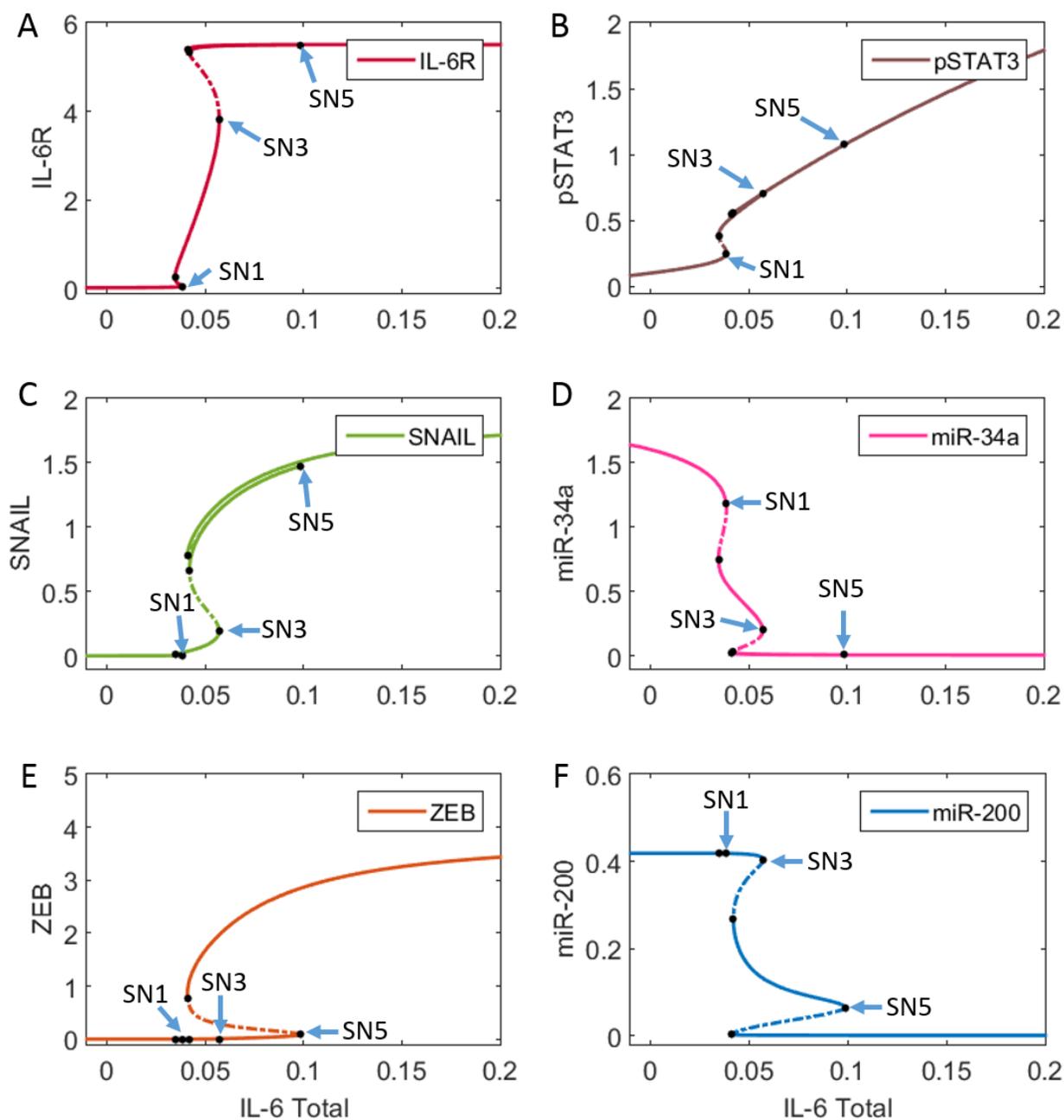
IL-6-induced EMT may exhibit four different steady states in the process as shown in Figure 5.5. With the increment of stimulus IL-6, the system undergoes three bistable switches across four different states, E state, Intermediate E (IE) state, Intermediate M (IM) state and M state. The three bistable switches are represented by saddle node SN1, SN3 and SN5 respectively as labeled in Figure 5.5.

The transitions between the four states are governed by three reversible bistable switches. Suppose that one cell originally starts in the lower branch corresponding to the E state. When IL-6 level is less than the threshold of first switch SN1, the expression level of IL-6R/SNAIL/ZEB is low, and that of miR-34a/miR-200 is high, and miR-34a/miR-200 control the cell. Upon gradually increasing the stimulus IL-6 level, the cell jumps across the first saddle-node bifurcation point SN1 to IE state, IL6R/pSTAT3/SNAIL are increased and miR-34a level is greatly decreased. At the second switch SN3, SNAIL level is increased dramatically and miR-34a level continue to decrease to almost 0, and they switch their dominance at SN3, and the cell transitions to the IM state from the IE state. Finally, when the level of stimulus IL-6 is larger than SN5, miR-200, repressed by ZEB, SNAIL and pSTAT3 (as shown in Figure 5.4A), turns off and ZEB finally turns on. The cell transitions from IM state to M state. In these three switching process, all the three switching behaviors are reversible process, which means as the stimulus IL-6 decreased, the cell could reverts back to the previous steady state, similar to Figure 5.4C.

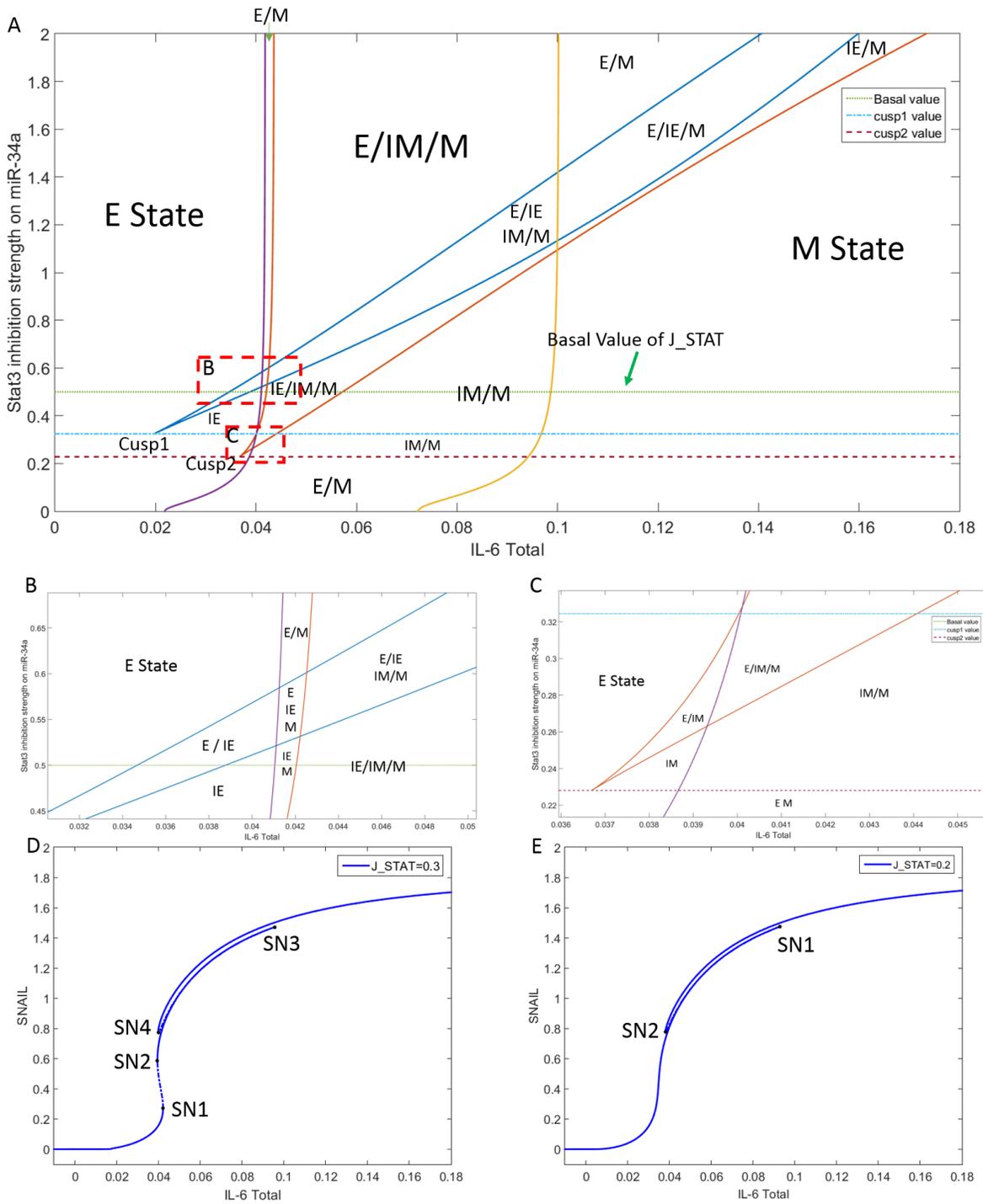
Next, we want to see what mechanism leads to the four state behavior. We plan to perform parameter sensitivity analysis and change the strength of each mutual inhibition loop to study the effects of each mutual inhibition on the system.

Figure 5.6 shows the impact of the first double negative feedback loop IL-6R/pSTAT3/miR34a on the IL-6-induced EMT system. Parameter  $J_{\text{STAT}}$  represents the inhibition strength of pSTAT3 on miR-34a, which can partially account for the strength of the first double negative feedback loop. From the two parameter bifurcation diagram of  $J_{\text{STAT}}$  and IL-6T (Figure 5.6A,B,C), we can see the two signals can produce various combinations of cell phenotypes in different parameter regions. For example, as shown in Figure 5.6B, under the selection of basal value of  $J_{\text{STAT}}$ , when increasing IL-6 level, the cell could have E state, coexistence of E and IE state, IE state, co-existence of IE and M state, co-existence of IE, IM and M state, co-existence of IM and M state and only M state. Also, from Figure 5.6A, we can see that robustness of intermediate state IE and IM (can be reflected by the area of the regions which have this cell phenotype) require minimum  $J_{\text{STAT}}$  and IL-6 signal.

From figure 5.6C, D and E, we can see that when we reduce the strength of pStat3 inhibition on miR-34a, or block this inhibition, the robustness of intermediate states is significantly reduced. From Figure 5.6A and D, we can see when  $J_{\text{STAT}}$  is smaller than  $\text{cusp1}$  and larger than  $\text{cusp2}$ , we can see that intermediate E (IE) state vanished, only E, IM and M states can exist under this condition. When  $J_{\text{STAT}}$  is smaller than  $\text{cusp2}$ , the intermediate M (IM) state also disappeared, only left E and M state. Figure 5.6A and E illustrate this.



**Figure 5.5: IL-6 induced EMT are composed of three bistable switches.** We plot signal-response curves for the steady state levels of IL-6R(A), pSTAT3(B), SNAIL (C), miR-34a (D), ZEB (E), and miR-200 (F) as a function to total IL-6 respectively. Black dots are saddle node bifurcation points, the dashed line connects a pair of saddle node bifurcation point to form a saddle node bifurcation. Solid lines, stable steady states; dashed lines, unstable steady states. Blue arrow tells the name of the saddle node bifurcation point.

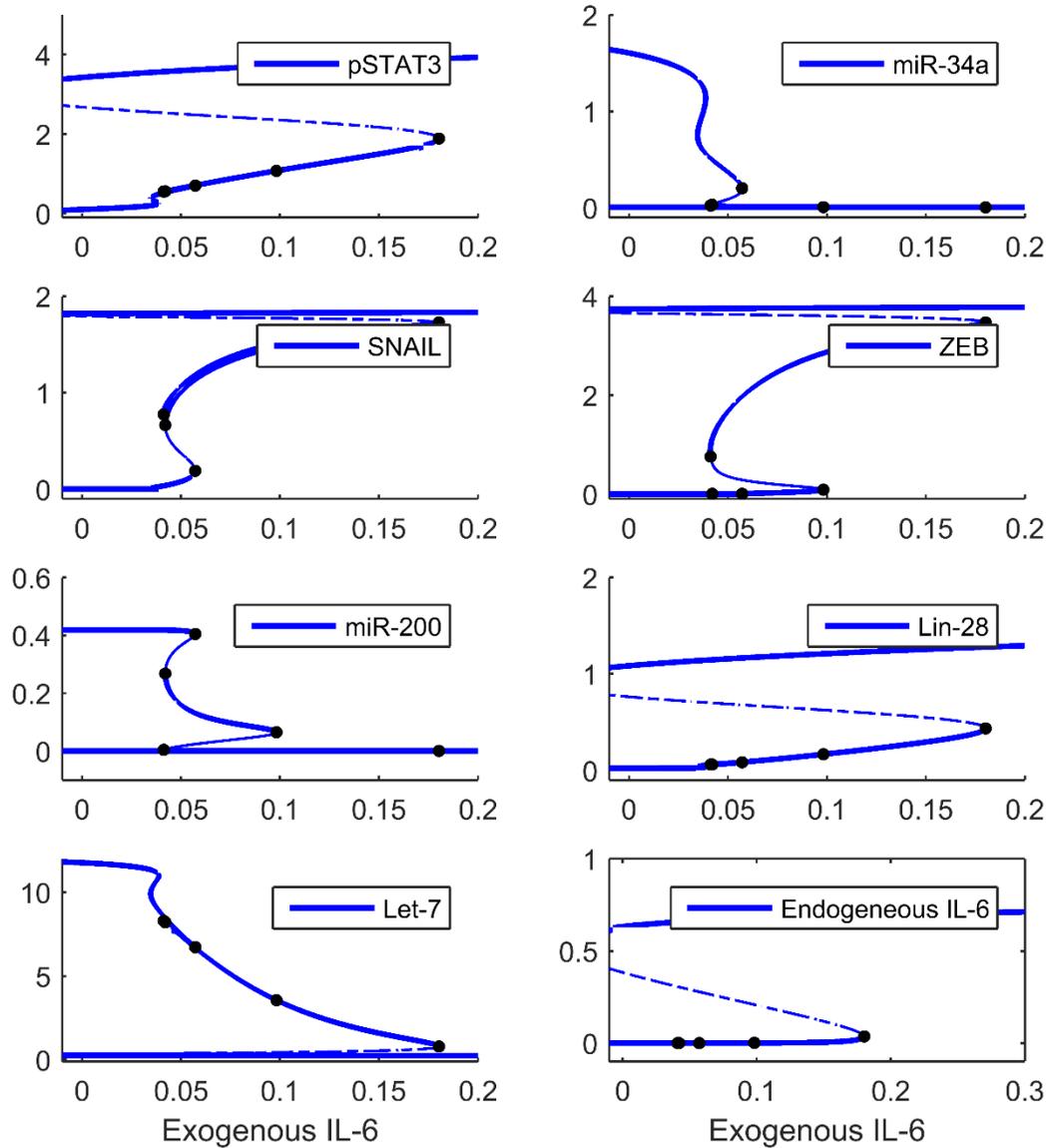


**Figure 5.6:** The effects of strength of inhibition from pSTAT3 to miR-34a on IL6-induced four state EMT system. A) Two parameter bifurcation diagram with respect to IL-6 total and pStat3 inhibition strength on miR-34a ( $J_{STAT}$ ). The blue, red, yellow and purple curves are computed by

extending the saddle-node bifurcation points obtained in one parameter bifurcation analysis and they define different parameter regions that can be mono-stable, bi-stable, tri-stable or tetra-stable as labeled in the figure. Each multi-stable region labeled with more than one state can be viewed as an area where multiple phenotypes co-exist. Region B and C are zoomed into panel B) and panel C) of the figure. Green dotted line: the basal parameter value for  $J_{STAT}$ , the *stat3* inhibition strength on *miR-34a*, blue dotted line:  $J_{STAT}=Cusp1$ , purple dotted line:  $J_{STAT}=Cusp2$ . *Cusp1*: the two saddle node bifurcation curve of the first switch meet at *cusp1*. *Cusp2*: the two saddle node bifurcation curve of the second switch meet at *cusp2*. Basal value of  $J_{STAT}=0.5$ . Different  $J_{STAT}$  value will affect the robustness of different cell phenotypes. We can see that when the when  $J_{STAT}$  is between *cusp2* and *cusp1* (D), the first bistable switch and intermediate E (IE) state disappear. When  $J_{STAT}$  is smaller than *cusp2*, the second switch and intermediate M state (IM) also disappear.

### **5.4.3. Autocrine production of IL-6 adds another irreversible switch**

In the previous sections, we discussed IL-6-induced EMT without participating of module 3: the IL-6/pStat3/Lin-28/Let-7 double negative feedback. Activation of this double negative feedback could induce the cell to synthesize its own IL-6, thus making the EMT process irreversible. In our TGF $\beta$ -induced EMT system, autocrine production of TGF $\beta$  (achieved by TGF $\beta$  and miR-200 double negative feedback loop) is responsible for irreversibility of the second switch (the switch from green to red steady state in Figure 5.2). However, in IL-6-induced EMT system, the IL-6:Let-7 double negative feedback loop adds another irreversible switch, which provides the IL-6 induced EMT process with three intermediate states. In the following work, we will discuss the control mechanism of different switches and which factor has the most impact on the system. Also, I will compare EMT induced by TGF $\beta$  alone with EMT induced by the activations of TGF $\beta$  and IL-6 together, in order to understand the mechanism of cooperativity between TGF $\beta$  and IL-6 pathway regarding EMT induction.



*Figure 5.7: Autocrine production of IL-6 makes the last switch irreversible. We plot signal-response curves for the steady state levels of pSTAT3, miR-34a, SNAIL, ZEB, miR-200, Lin-28, Let-7 and Endogenous IL-6 as a function to exogenous IL-6 respectively. We can see that the expression level of these regulators has undergoes four switches and the last switch is irreversible. Solid lines, stable steady states; dashed lines, unstable steady states.*

## 5.5. Discussion

EMT is a stepwise activation process composed of multiple intermediate states. Our work of IL-6-induced EMT system further illustrates this point. In this chapter, we study the underlying design principle to generate these intermediate states. These intermediate states arise from multiple positive feedback loops involved in the regulatory network. It is highly possible that when more positive feedback loops are included, more intermediate states will be discovered and a spectrum of EMT will be established. However, the physiological significance and evolutionary advantage of these intermediate states still needs to be explored.

In this model, we didn't consider the possible negative feedback towards pSTAT3 activation which could lead to a pulse or oscillatory behavior of pSTAT3<sup>8</sup>. Future work discussing priming effect may take this into account. Also, we integrated the autocrine production of IL-6 in the regulatory network. Autocrine production of the stimulus signal is one common way to generate irreversible switches in the system. These autocrine signaling pathways are also known as “vicious cycle of autocrine”, due to the fact that they can perpetuate the system by the signal itself, and they are always correlated with advanced stages of tumor development. In our study, the IL-6 autocrine loop adds another irreversible bistable switches to the system.

In summary, we built a mathematical model of IL-6-induced EMT and identified different intermediate states during this process. We further studied the effect of strength of mutual inhibition on the robustness of different states in the process. In the future, we plan to compare EMT induced by TGF- $\beta$  and IL-6 alone and study the mechanism of cooperative effect between TGF- $\beta$  and IL-6 pathway regarding EMT induction.

## Supplemental Information:

### Equations:

$$IL6 = IL6\_0 + IL6_a - IL6\_IL6R \quad (1)$$

$$\frac{d[mIL6R]}{dt} = k_{s\_mIL6R} - 2k_{d\_LR1}[LR1] - k_{d\_LR2}[LR2] - k_{d\_mIL6R}[mIL6R] \quad (2)$$

$$[mIL6R] = [mIL6RT] - 2[LR1] - LR2 \quad (3)$$

$$\frac{d[IL6R]}{dt} = k_{r\_mIL6R}[mIL6R] - k_1[IL6][IL6R] + k_2[LR6\_IL6R] - k_{d\_IL6R}[IL6R] \quad (4)$$

$$\frac{d[IL6\_IL6R]}{dt} = k_1[IL6][IL6R] - k_2[IL6\_IL6R] - k_{d\_IL6\_IL6R}[LR6\_IL6R] \quad (5)$$

$$\frac{d[pSTAT3]}{dt} = k_{s\_stat} + V_{sstat} \frac{[IL6\_IL6R]}{J\_IL6S + [IL6\_IL6R]} - k_{d\_STAT}[pSTAT3] \quad (6)$$

$$\frac{d[miR34aT]}{dt} = k_{s\_r34} + \frac{k_{miR34a}}{1 + \left(\frac{[pSTAT3]}{J_{stat}}\right)^2 + \left(\frac{[SNAIL]}{J_{1\_34}}\right)^2 + \left(\frac{[ZEB]}{J_{2\_34}}\right)^2} - k_{d\_34}[miR34a] - 2k_{d\_LR1}(1-\lambda)[LR2] - k_{d\_SR1}(1-\lambda)[SR1] \quad (7)$$

$$[miR34a] = [miR34aT] - 2[LR1] - 2[LR2] - [SR1] \quad (8)$$

$$\frac{d[LR1]}{dt} = k_3[miR34a][mIL6R] - k_4[LR1] \quad (9)$$

$$\frac{d[LR2]}{dt} = k_3[miR34a][LR1] - k_4[LR2] \quad (10)$$

$$[snail] = [snailT] - [SR1] \quad (11)$$

$$\frac{d[snailT]}{dt} = k_{0\_snail} + k_{snail} \frac{\left(\frac{[pSTAT3]}{J_{snail0}}\right)^2}{\left(1 + \left(\frac{[pSTAT3]}{J_{snail0}}\right)^2\right)\left(1 + \frac{[SNAIL]}{J_{snail1}}\right)} - k_{d\_msnail}[snail] - k_{d\_SR1}[SR1] \quad (12)$$

$$\frac{d[SR1]}{dt} = k_s[snail][miR34a] - [SR1] \quad (13)$$

$$\frac{d[SNAIL]}{dt} = k_{r\_SNAIL}[snail] - k_{d\_SNAIL}[SNAIL] \quad (14)$$

$$\begin{aligned} \frac{d[zebt]}{dt} = & k_{0\_zeb} \frac{\left(\frac{[SNAIL]}{J_{zeb}}\right)^2}{1 + \left(\frac{[SNAIL]}{J_{zeb}}\right)^2} - k_{d\_zeb}[zeb] - 5dk\_ZR1[ZR1] - 10dk\_ZR2[ZR2] \\ & - 10dk\_ZR3[ZR3] - 5dk\_ZR4[ZR4] - dk\_ZR5[ZR5] \end{aligned} \quad (15)$$

$$\frac{d[ZEB]}{dt} = k_{ZEB}[zeb] - k_{d\_ZEB}[ZEB] \quad (16)$$

$$\begin{aligned} \frac{d[miR200T]}{dt} = & k_{0\_200} + \frac{k_{200}}{1 + \left(\frac{[SNAIL]}{J_{1\_200}}\right)^3 + \left(\frac{[ZEB]}{J_{2\_200}}\right)^2} - k_{d\_200}[miR200] \\ & - 5dk\_ZR1[ZR1] - 20dk\_ZR2[ZR2] - 30dk\_ZR3[ZR3] - 20dk\_ZR4[ZR4] - 5dk\_ZR5[ZR5] \\ & + 5\lambda_1 dk\_ZR1[ZR1] + 20\lambda_2 dk\_ZR2[ZR2] + 30\lambda_3 dk\_ZR3[ZR3] + 20\lambda_4 dk\_ZR4[ZR4] + 5\lambda_5 dk\_ZR5[ZR5] \end{aligned} \quad (17)$$

$$\frac{d[ZR1]}{dt} = K_1[miR200][zeb] - [ZR1] \quad (18)$$

$$\frac{d[ZR2]}{dt} = K_2[miR200][ZR1] - [ZR2] \quad (19)$$

$$\frac{d[ZR3]}{dt} = K_3[miR200][ZR2] - [ZR3] \quad (20)$$

$$\frac{d[ZR4]}{dt} = K_4[miR200][ZR3] - [ZR4] \quad (21)$$

$$\frac{d[ZR5]}{dt} = K_5[miR200][ZR4] - [ZR5] \quad (22)$$

$$[zeb] = [zebt] - (5[ZR1] + 10[ZR2] + 10[ZR3] + 5[ZR4] + [ZR5]) \quad (23)$$

$$[miR200] = [miR200T] - (5[ZR1] + 20[ZR2] + 30[ZR3] + 20[ZR4] + 5[ZR5]) \quad (24)$$

$$\frac{d[Lin28]}{dt} = k_{s\_LIN28} + V_{s\_lin28} \frac{\left(\frac{[pSTAT3]}{J_{lin28}}\right)^2}{1 + \left(\frac{[pSTAT3]}{J_{lin28}}\right)^2} - k_{d\_lin28}[Lin28] \quad (25)$$

$$\frac{d[let7T]}{dt} = k_{s\_let7} + V_{s\_let7} \frac{1}{1 + \left(\frac{[LIN28]}{J_{i\_let7}}\right)^{en}} - k_{d\_let7}[let7] - k_{d\_mIL6let7}(1 - \lambda)[mIL6let7] \quad (26)$$

$$[let7] = [let7T] - 3[mIL6let7] \quad (27)$$

$$\frac{d[mIL6T]}{dt} = k_{s\_mil6} - k_{d\_mil6}[mIL6] - k_{d\_mIL6let7}[mIL6let7] \quad (28)$$

$$\frac{d[mIL6let7]}{dt} = k_{on}[mIL6][let7]^3 - k_{off}[mIL6let7] \quad (29)$$

$$[mIL6] = [mIL6T] - [mIL6let7] \quad (30)$$

$$\frac{d[IL6a]}{dt} = k_{r\_il6}[mIL6] - k_{d\_il6}[IL6a] \quad (31)$$

Here, [x] means the concentration of molecule x.

**Meaning of each variable:**

IL-6: free IL-6 protein which can bind to IL-6R

IL6\_0: IL-6 protein in the culture media, control parameter

IL6<sub>a</sub>: autocrine secreted IL-6 protein

mIL6R: messenger RNA of IL-6R

IL6R: IL-6R protein

IL6\_IL6R: The complex formed by IL-6 binds to IL-6R in the membrane

pSTAT3: Activated STAT3

LR1: The microRNA:mRNA complex of miR-34a and IL-6R mRNA, with one miR-34a bounded

LR2: The microRNA:mRNA complex of miR-34a and IL-6R mRNA, with two miR-34a bounded

SR1: The microRNA:mRNA complex of miR-34a and SNAIL mRNA, with two miR-34a bounded

miR34a: Free miR-34a

miR34aT: Total miR-34a, including free form of miR-34a, and miR-34a bound by its targets (miR-34a bound to IL-6R mRNA, and SNAIL mRNA, or LR1, LR2 and SR1)

snail: Free mRNA of SNAIL protein

snailT: Total mRNA of SNAIL protein, including free snail mRNA and snail mRNA bounded by miR-34a (SR1)

SNAIL: SNAIL protein

miR200: Free form of miR-200

ZR1: The microRNA:mRNA complex of miR-200 and ZEB mRNA, with one miR-200 bounded

ZR2: The microRNA:mRNA complex of miR-200 and ZEB mRNA, with two miR-200 bounded

ZR3: The microRNA:mRNA complex of miR-200 and ZEB mRNA, with three miR-200 bounded

ZR4: The microRNA:mRNA complex of miR-200 and ZEB mRNA, with four miR-200 bounded

ZR5: The microRNA:mRNA complex of miR-200 and ZEB mRNA, with five miR-200 bounded

zeb: Free form of ZEB mRNA

zebt: Total mRNA of ZEB, including free form of zeb mRNA, and ZEB mRNAs bounded by miR-200

(including ZR1, ZR2, ZR3, ZR4, ZR5)

miR200T: Total microRNA miR-200, including free form of miR-200 and all miR-200 bounded by its targets (including ZR1, ZR2, ZR3, ZR4, ZR5)

Lin28: microRNA-binding protein Lin-28, which inhibit the maturation of microRNA Let-7

let-7: Free form of microRNA Let-7

let7T: Total microRNA Let-7, including free form of let-7 and let-7 bounded by its target IL-6 mRNA

mIL6let7: The microRNA:mRNA complex of Let-7 and IL-6 mRNA

mIL6: messenger RNA of IL-6 protein

### **Initial Conditions:**

Initially cells are in the E state. The initial values are selected to be their steady state value when  $[IL-6T] = 0$ .

$[mIL6RT] = 0.28$ ,  $[IL-6R] = .001$ ,  $[STAT3] = 0.1$ ,  $[SNAIL] = 0.001$ ,  $[ZEB] = 0.001$

$[miR34aT] = 1.6$ ,  $[miR200T] = 0.4$ .

*Table 5.1 Parameter values*

Parameter	Value	Parameter	Value
Module 1			
ks_mIL6R	0.05 $\mu\text{M/hr}$	ks_stat	0.0006 $\mu\text{M/hr}$
kd_mIL6R	0.05 $\text{hr}^{-1}$	Vsstat	3 $\mu\text{M/hr}$
ktr_mIL6R	1 $\text{hr}^{-1}$	J_IL6s	0.5 $\mu\text{M}$
k1	200 $\mu\text{M}^{-1}\text{hr}^{-1}$	kd_STAT	0.5 $\text{hr}^{-1}$
K2	20 $\text{hr}^{-1}$		
kd_IL6R	0.1 $\text{hr}^{-1}$	lambda	0.5
kd_LR1	0.1 $\text{hr}^{-1}$	kd_LR2	0.2 $\text{hr}^{-1}$
ks_R34	0.001	kd_34	0.035 $\text{hr}^{-1}$
k_miR34a	0.09	J_stat	0.5 $\mu\text{M}$
Module 2			
J2_34	0.5 $\mu\text{M}$	J1_34	0.15 $\mu\text{M}$
k0_snail	0.001	k_snail	0.05
J_snail0	0.62 $\mu\text{M}$	J_snail1	0.67 $\mu\text{M}$
kd_msnail	0.09 $\text{hr}^{-1}$	ktr_SNAIL	17
kd_SNAIL	1.2 $\text{hr}^{-1}$	kd_SR1	0.9 $\text{hr}^{-1}$
K1	10 $\mu\text{M}^{-1}\text{hr}^{-1}$	lamda1	0.5
dk_ZR1	0.9 $\text{hr}^{-1}$		
k0_zeb	0.003 $\mu\text{M/hr}$	k_zeb	0.1 $\mu\text{M/hr}$
J_zeb	3.5 $\mu\text{M}$	Kd_zeb	0.09 $\text{hr}^{-1}$
k_ZEB	17 $\mu\text{M/hr}$	J_ZEB	0.06 $\mu\text{M}$
kd_ZEB	1.2 $\text{hr}^{-1}$		
k0_200	0.0002 $\mu\text{M/hr}$	k_200	0.02
J1_200	3 $\mu\text{M}$	J2_200	0.2 $\mu\text{M}$
Kd_200	0.035 $\text{hr}^{-1}$		
Module 3			
ks_LIN28	0.006 $\mu\text{M/hr}$	Vs_lin28	1 $\mu\text{M/hr}$
J_lin28	5 $\mu\text{M}$	kd_lin28	0.3
ks_let7	0.01 $\mu\text{M/hr}$	Vs_let7	0.5
Ji_let7	0.5 $\mu\text{M}$	en	5
kd_let7	0.035 $\text{hr}^{-1}$		
ks_mil6	0.01 $\mu\text{M/hr}$	kd_mil6	0.1 $\text{hr}^{-1}$
ktr_il6	1.2 $\text{hr}^{-1}$	kd_il6	0.1 $\text{hr}^{-1}$
kd_mIL6let7	0.9 $\text{hr}^{-1}$		
kon	1000 $\mu\text{M}^{-1}\text{hr}^{-1}$	koff	100 $\text{hr}^{-1}$

## References:

1. Huang, R.Y.J. *et al.* An EMT spectrum defines an anoikis-resistant and spheroidogenic intermediate mesenchymal state that is sensitive to e-cadherin restoration by a src-kinase inhibitor, saracatinib (AZD0530). *Cell Death & Disease* **4**(2013).
2. Tan, T.Z. *et al.* Epithelial-mesenchymal transition spectrum quantification and its efficacy in deciphering survival and drug responses of cancer patients. *EMBO Mol Med* **6**, 1279-93 (2014).
3. Hong, T. *et al.* An Ovol2-Zeb1 Mutual Inhibitory Circuit Governs Bidirectional and Multi-step Transition between Epithelial and Mesenchymal States. *PLoS Comput Biol* **11**, e1004569 (2015).
4. Lamouille, S., Xu, J. & Derynck, R. Molecular mechanisms of epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol* **15**, 178-96 (2014).
5. Rokavec, M., Wu, W. & Luo, J.L. IL6-mediated suppression of miR-200c directs constitutive activation of inflammatory signaling circuit driving transformation and tumorigenesis. *Mol Cell* **45**, 777-89 (2012).
6. Yadav, A., Kumar, B., Datta, J., Teknos, T.N. & Kumar, P. IL-6 promotes head and neck tumor metastasis by inducing epithelial-mesenchymal transition via the JAK-STAT3-SNAIL signaling pathway. *Mol Cancer Res* **9**, 1658-67 (2011).
7. Rojas, A. *et al.* IL-6 promotes prostate tumorigenesis and progression through autocrine cross-activation of IGF-IR. *Oncogene* **30**, 2345-2355 (2011).
8. Rokavec, M. *et al.* IL-6R/STAT3/miR-34a feedback loop promotes EMT-mediated colorectal cancer invasion and metastasis. *J Clin Invest* **124**, 1853-67 (2014).
9. Sullivan, N.J. *et al.* Interleukin-6 induces an epithelial-mesenchymal transition phenotype in human breast cancer cells. *Oncogene* **28**, 2940-7 (2009).
10. Scheller, J., Chalaris, A., Schmidt-Arras, D. & Rose-John, S. The pro- and anti-inflammatory properties of the cytokine interleukin-6. *Biochim Biophys Acta* **1813**, 878-88 (2011).
11. Bromberg, J. & Wang, T.C. Inflammation and cancer: IL-6 and STAT3 complete the link. *Cancer Cell* **15**, 79-80 (2009).

12. Grivennikov, S.I., Greten, F.R. & Karin, M. Immunity, inflammation, and cancer. *Cell* **140**, 883-99 (2010).
13. O'Toole, A. *et al.* Tumour microenvironment of both early- and late-stage colorectal cancer is equally immunosuppressive. *Br J Cancer* **111**, 927-32 (2014).
14. Yoneda, M. *et al.* Prognostic impact of tumor IL-6 expression after preoperative chemoradiotherapy in patients with advanced esophageal squamous cell carcinoma. *Anticancer Res* **33**, 2699-705 (2013).
15. Guo, Y., Xu, F., Lu, T., Duan, Z. & Zhang, Z. Interleukin-6 signaling pathway in targeted therapy for cancer. *Cancer Treat Rev* **38**, 904-10 (2012).
16. Sansone, P. *et al.* IL-6 triggers malignant features in mammospheres from human ductal breast carcinoma and normal mammary gland. *J Clin Invest* **117**, 3988-4002 (2007).
17. Salgado, R. *et al.* Circulating interleukin-6 predicts survival in patients with metastatic breast cancer. *Int J Cancer* **103**, 642-6 (2003).
18. Clinchy, B. *et al.* Preoperative interleukin-6 production by mononuclear blood cells predicts survival after radical surgery for colorectal carcinoma. *Cancer* **109**, 1742-9 (2007).
19. Heimdal, J.H., Kross, K., Klementsén, B., Olofsson, J. & Aarstad, H.J. Stimulated monocyte IL-6 secretion predicts survival of patients with head and neck squamous cell carcinoma. *BMC Cancer* **8**, 34 (2008).
20. Yao, Z. *et al.* TGF- $\beta$  IL-6 axis mediates selective and adaptive mechanisms of resistance to molecular targeted therapy in lung cancer. *Proc Natl Acad Sci U S A* **107**, 15535-40 (2010).
21. Connolly, E.C., Freimuth, J. & Akhurst, R.J. Complexities of TGF- $\beta$  targeted cancer therapy. *Int J Biol Sci* **8**, 964-78 (2012).
22. Tian, X.-J., Zhang, H. & Xing, J. Coupled reversible and irreversible bistable switches underlying TGF $\beta$ -induced epithelial to mesenchymal transition. *Biophysical journal* **105**, 1079-1089 (2013).
23. Zhang, J. *et al.* TGF- $\beta$ -induced epithelial-to-mesenchymal transition proceeds through stepwise activation of multiple feedback loops. *Science signaling* **7**, ra91-ra91 (2014).

24. Liu, R.-Y. *et al.* JAK/STAT3 signaling is required for TGF- $\beta$ -induced epithelial-mesenchymal transition in lung cancer cells. *International journal of oncology* **44**, 1643-1651 (2014).
25. Lu, M., Jolly, M.K., Levine, H., Onuchic, J.N. & Ben-Jacob, E. MicroRNA-based regulation of epithelial-hybrid-mesenchymal fate determination. *Proc Natl Acad Sci U S A* **110**, 18144-9 (2013).
26. Conze, D. *et al.* Autocrine production of interleukin 6 causes multidrug resistance in breast cancer cells. *Cancer Res* **61**, 8851-8 (2001).
27. Taniguchi, K. & Karin, M. IL-6 and related cytokines as the critical lynchpins between inflammation and cancer. *Seminars in Immunology* **26**, 54-74 (2014).
28. He, G. *et al.* Identification of liver cancer progenitors whose malignant progression depends on autocrine IL-6 signaling. *Cell* **155**, 384-96 (2013).
29. Iliopoulos, D., Hirsch, H.A. & Struhl, K. An epigenetic switch involving NF-kappaB, Lin28, Let-7 MicroRNA, and IL6 links inflammation to cell transformation. *Cell* **139**, 693-706 (2009).
30. Grivennikov, S. & Karin, M. Autocrine IL-6 signaling: a key event in tumorigenesis? *Cancer Cell* **13**, 7-9 (2008).
31. Guo, L. *et al.* Stat3-coordinated Lin-28-let-7-HMGA2 and miR-200-ZEB1 circuits initiate and maintain oncostatin M-driven epithelial-mesenchymal transition. *Oncogene* **32**, 5272-82 (2013).
32. Nam, Y., Chen, C., Gregory, R.I., Chou, J.J. & Sliz, P. Molecular basis for interaction of let-7 microRNAs with Lin28. *Cell* **147**, 1080-91 (2011).

## Chapter 6. Conclusions

How can a cell maintains its phenotype and how can a cell transitions to another phenotype? We begin to answer these questions by considering the mechanisms that maintain a cell's phenotype.

We first looked into the mechanism of epigenetic memory maintenance by dynamic histone modification. In Chapter 2 of this dissertation, we presented a mathematical model of epigenetic memory of collective histone modifications. The model suggests that the enzyme selectivity of different histone substrates and the cooperativity between nearest neighbor are essential to generate bistability and epigenetic memory. This corresponds to a simple 'read and write scheme' of the enzymes: the selectivity will help the enzyme to read its correct substrate histone epigenetic marks and the cooperativity will help the enzyme to write the same histone epigenetic marks to the neighbor. The model bridges the gap between molecular interactions (time scale  $\leq 1$  second) and epigenetic memory (time scale  $\geq 1$  month).

During differentiation, histone epigenetic memory is lost and the cell transitions to a new state. In Chapter 3, we studied the differentiation process of olfactory sensory neurons (OSNs). Before differentiation, all olfactory receptor genes are occupied by repressive marks, upon differentiation, enzyme LSD1, which could remove both the repressive mark and active mark is upregulated. OR genes compete for LSD1 to jump from the silencing state to active state. After a slow activation process, one gene will be activated and elicit one fast negative feedback toward LSD1, thus trapping other genes in the epigenetically silenced state. Moreover, with the newly discovered enhancer interaction network among OR genes, our model could further eliminate the co-expression of several genes, this will also lead to biased gene expression, which agrees with the results of G9a/GLP dKO experiment<sup>1</sup>. From a modeling perspective, epigenetic switching assumes a barrier-crossing-like dynamics for wild type OSNs, but a ratchet-like dynamics with G9a/GLP dKO OSNs.

In the modeling work of OSN differentiation, we related slow epigenetic activation to fast transcriptional regulation (LSD1 silencing) and further to the cis-regulatory elements (enhancers) in the DNA sequences. Histone-level epigenetic regulation usually cross-talks to other levels of regulation with

different time scales. These cross-talk mechanisms, which may involve large difference in time scale, could be very interesting regulatory motifs in cell fate decision process.

Gene regulatory networks composed of regulatory motifs as building blocks can generate a variety of dynamics and control the cell fate decision making process. Numerous studies in systems biology have revealed the dynamical behaviors and functional roles of different motifs in diverse biological systems<sup>2,3</sup>. With powerful high-throughput experimental techniques, the community of systems biologists will discover many new regulatory motifs. Integrating high-throughput approaches with mathematical modeling, the community will begin to understand the functional significance of these motifs. For example, in Chapter 4, we used high-throughput methods to identify differentially expressed proteins and microRNAs, and then studied the underlying relationship among them by mathematical modeling.

In Chapter 5, we looked into an important phenotypic transition, the epithelial-to-mesenchymal transition (EMT), which is dysregulated in many diseases, including cancer. EMT could be triggered by different cytokines, and there usually are many intermediate cell states in the EMT process created by numerous mutual inhibitions. These intermediate cell states may be of great clinical importance. However, from a design point of view (or an evolutionary point of view), the reason for so many intermediate states is still unclear. More experimental and computational studies are needed to quantify different intermediate states and reveal their underlying control mechanisms.

## References

1. Lyons, D.B. *et al.* Heterochromatin-mediated gene silencing facilitates the diversification of olfactory neurons. *Cell reports* **9**, 884-892 (2014).
2. Tyson, J.J. & Novak, B. Functional motifs in biochemical reaction networks. *Annu Rev Phys Chem* **61**, 219-40 (2010).
3. Tyson, J.J. & Novak, B. Models in biology: lessons from modeling regulation of the eukaryotic cell cycle. *BMC Biol* **13**, 46 (2015).