Single-molecule techniques to probe the dynamic gene regulatory network formed by core pluripotency circuit in embryonic stem cells

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th>Citation</th>
<th>Lin, Ya. 2014. Single-molecule techniques to probe the dynamic gene regulatory network formed by core pluripotency circuit in embryonic stem cells. Doctoral dissertation, Harvard University.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:13070085">http://nrs.harvard.edu/urn-3:HUL.InstRepos:13070085</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
Single-molecule techniques to probe the dynamic gene regulatory network formed by core pluripotency circuit in embryonic stem cells

A dissertation presented
by
Ya Lin
to
The Committee on Higher Degrees in Biophysics
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Biophysics

Harvard University
Cambridge, Massachusetts
July 2014
Abstract

This work investigates the dynamics of gene regulatory network formed by Oct4, Sox2 and Nanog in embryonic stem cells (ESCs). Despite a large number of existing studies on stem cells, current technologies used often force a compromise between quantification of gene expression via bulk measurements and qualitative imaging of cell heterogeneity. There are few options that allow for accurate and quantitative single-cell analysis that is robust yet not associated with a high degree of technical difficulty or obscured by amplification. Here, we adapted a high resolution, single-molecule RNA fluorescent in situ hybridization technique (smFISH) to study gene expression of the core pluripotency circuit upon various types of perturbations such as differentiation, induction or knockdown of one of the three pluripotent factors. We used previously-published smFISH procedures as our initial template for investigating gene regulatory dynamics of the core pluripotency circuit during those perturbation assays. To obtain a more comprehensive picture of the regulatory circuit, we developed a modified smFISH strategy to measure mRNA and protein expression simultaneously in single ESCs. By incorporating a novel modification into the smFISH technique which allows accurate quantification of transcripts that differ by short sequences, we managed to identify a few interesting features of the core pluripotency circuit. Taken together, we demonstrated our ability to perform single-cell, single-molecule assays that reveal highly quantitative information in unprecedented detail.
Table of Contents

Abstract...........................................................................................................................................iii

Table of Contents ..............................................................................................................................iv

List of Figures..................................................................................................................................viii

List of Tables .....................................................................................................................................xi

Acknowledgements..........................................................................................................................xii

Chapter 1 Introduction.....................................................................................................................1
  1.1 Thesis Contributions..................................................................................................................1
  1.2 Thesis Outline...........................................................................................................................2

Chapter 2 Effects of all-trans retinoic acid on the differentiation of mouse embryonic stem cells....5
  2.1 Introduction...............................................................................................................................5
  2.2 Results.......................................................................................................................................8
    2.2.1 RA-induced differentiation assay without removing RA afterwards.................................8
        2.2.1.1 Oct4, Sox2 and Nanog mRNA expression profiles all showed a monotonic decrease.....8
        2.2.1.2 Sox2 was most gradually down-regulated among the three factors.............................11
        2.2.1.3 Oct4, Sox2 and Nanog mRNA expression did not completely shut down after 7-day RA treatment.................................................................14
        2.2.1.4 NOS transcriptional dynamics upon RA treatment..........................................................16
        2.2.1.5 Single-cell scatterplots suggested the potential effect of RA on the differentiation of ECSs........................................................................................................................17
    2.2.2 RA-induced differentiation assay with RA removal afterwards........................................19
        2.2.2.1 Oct4, Sox2 and Nanog mRNA expressions were all down-regulated at the population level..........................................................................................................................20
        2.2.2.2 Oct4 was down-regulated more gradually when treated with RA for 13hrs compared with 24hr or 2day RA treatment..................................................................................22
        2.2.2.3 Sox2 and Nanog responded to RA treatment differently from Oct4.............................23
  2.3 Discussion.................................................................................................................................28
  2.4 Potential pitfalls and caveats.....................................................................................................29
  2.5 Methods....................................................................................................................................30
    2.5.1 ESC culture medium............................................................................................................30
2.5.2 ESC adherent monoculture for RA-induced differentiation assays.........................30
2.5.3 SmFISH probe design..........................................................................................31
2.5.4 Sample preparations for smFISH.........................................................................32
2.5.5 Image acquisition...............................................................................................32
2.5.6 Image analysis......................................................................................................33
2.6 References................................................................................................................34

Chapter 3 Dynamics on a transcript pathway formed by core pluripotency circuit in ESCs.......37
3.1 Introduction................................................................................................................37
3.2 Results.......................................................................................................................38
  3.2.1 Positive correlation between every pair of the genes in the unperturbed NOS network.......38
  3.2.2 NOS circuit’s transcript space may have upper bounds...........................................44
  3.2.3 The population average transcript levels were constrained to the expression space spanned
      by the unperturbed cells.........................................................................................50
  3.2.4 All perturbations eventually led to the contraction of explored areas to a nearly common
      region within the transcript plane.........................................................................51
3.3 Discussion..................................................................................................................53
3.4 Potential pitfalls and caveats.....................................................................................54
3.5 Methods.....................................................................................................................55
  3.5.1 ESC culture medium............................................................................................55
  3.5.2 ESC adherent monoculture for doxycycline-induced perturbation assays..................55
  3.5.3 Sample preparations for smFISH...........................................................................55
  3.5.4 Image acquisition and analysis.............................................................................55
  3.5.5 Calculations and Analyses....................................................................................56
3.6 References..................................................................................................................58

Chapter 4 Gene regulation of the core pluripotency circuit in embryonic stem cells.........60
4.1 Introduction...............................................................................................................60
4.2 Results.......................................................................................................................61
  4.2.1 mRNA expression levels of Sox2, Oct4 and Nanog were doxycycline dose-dependent.....61
  4.2.2 The regulatory patterns of Sox2 and Oct4 were symmetric......................................61
  4.2.3 NOS circuit seemed to have an upper bound to keep the gene expression below certain
      level.........................................................................................................................68
  4.2.4 Nanog seemed to minimally regulate Sox2 and Oct4............................................69
4.2.5 Protein and mRNA expression profiles of the NOS circuit correlated well with each other...77
4.2.6 Upon perturbation Sox2 and Oct4 expression profiles became anti-correlated at both mRNA and protein levels.................................................................78
4.2.7 Overexpressing Sox2 or Oct4 increased transcription heterogeneity of the other two factors in the NOS circuit...........................................................................95
4.2.8 Overexpressing Nanog transiently decreased the heterogeneity of Nanog transcripts while having minimal influence on Sox2 and Oct4..................................................95
4.3 Discussion.................................................................................................96
4.4 Methods..................................................................................................97
4.4.1 ESC culture medium...............................................................................97
4.4.2 ESC adherent monoculture for doxycycline-induced perturbation assays........97
4.4.3 Sample preparations for smFISH/IF.......................................................98
4.4.4 Immunoblot..........................................................................................99
4.4.5 Image acquisition and analysis..............................................................99
4.5 References...............................................................................................100

Chapter 5 Autorepression helps maintain the expression level of Oct4, Sox2 and Nanog in the ESCs..............................................................................................................102
5.1 Introduction..............................................................................................102
5.2 Results......................................................................................................103
5.2.1 smFISH probes were designed to accurately measure the exo- and endo-genous expression of Oct4, Sox2 and Nanog.................................................................103
5.2.2 Autorepression of Oct4, Sox2 and Nanog in both serum containing and 2i media........114
5.2.3 Autorepression of Sox2, Oct4 and Nanog could be reversed upon doxycycline removal....122
5.2.4 Reversal of autorepression of Sox2, Oct4 and Nanog occurred at different rates........123
5.2.5 Anti-correlation between Oct4 and Sox2 disappeared upon doxycycline removal........124
5.3 Discussion..............................................................................................128
5.4 Methods..................................................................................................129
5.4.1 ESC culture medium...............................................................................129
5.4.2 ESC adherent monoculture for doxycycline-induced perturbation assays...............129
5.4.3 Sample preparations for smFISH............................................................129
5.4.4 Image analysis for quantifying exogenous and endogenous transcripts...............129
5.5 References...............................................................................................131
Chapter 6 Potential network connectivity of the core pluripotency circuit upon perturbation......133
  6.1 Introduction ..................................................................................................................133
  6.2 Results ..........................................................................................................................134
    6.2.1 NOS network upon perturbation over time .........................................................134
    6.2.2 MRA revealed network connectivity of the NOS circuit upon perturbation ..........135
  6.3 Discussion ....................................................................................................................136
  6.4 Calculation and Analysis .............................................................................................142
  6.5 References ..................................................................................................................145

Chapter 7 Final Conclusions .................................................................................................147
  7.1 Future Directions .........................................................................................................147
  7.2 References ..................................................................................................................150
List of Figures

Figure 2.1: Plots of population average for Oct4, Sox2 and Nanog transcripts in the ESCs that were treated with 0.5 µM of RA for various durations……………………………………………………………………..10
Figure 2.2: Plots of population median for Oct4, Sox2 and Nanog transcripts in the ESCs that were treated with 0.5 µM of RA for various durations……………………………………………………………………..11
Figure 2.3: Single-cell histograms of Oct4, Sox2 and Nanog transcripts in the ESCs for the RA treatment experiment……………………………………………………………………………12
Figure 2.4: Three-dimensional scatterplot of Oct4, Sox2 and Nanog transcripts for the 7-day RA-treated sample…………………………………………………………………………..15
Figure 2.5: Three-dimensional scatterplot of Oct4, Sox2 and Nanog transcripts for the RA treatment time course experiment……………………………………………………………………………16
Figure 2.6: Single-cell pairwise scatterplots of Oct4, Sox2 and Nanog transcripts for the RA treatment time course experiment………………………………………………………………………….18
Figure 2.7: Plots of population average of Oct4, Sox2 and Nanog transcripts in the ESCs that were first treated with 0.5 µM of RA for various durations followed by RA removal………………………….22
Figure 2.8: Single-cell histograms of Oct4, Sox2 and Nanog transcripts in the ESCs for the RA removal experiment………………………………………………………………………………24
Figure 2.9: Three-dimensional scatterplot of Oct4, Sox2 and Nanog transcripts in the ESCs that were first treated with 0.5 µM of RA for various durations followed by RA removal……………………………27

Figure 3.1: Schematics of smFISH and using irradiated MEFs as a negative control to test binding specificity of Nanog, Oct4, and Sox2 RNA FISH probes……………………………………………………………………..40
Figure 3.2: Single-cell histograms of Nanog, Sox2, and Oct4 transcripts in unperturbed population of cells……………………………………………………………………………………41
Figure 3.3: Transcript space explored by the unperturbed core pluripotency circuit…………………………..42
Figure 3.4: Schematic of the three cell lines used for ectopically perturbing transcript levels of Nanog, Sox2, and Oct4 individually…………………………………………………………………………….43
Figure 3.5: Inducibility of $P_{TET}$ inferred from GFP mRNA counts as a function of time with addition of 2 µg/ml of doxycycline to the KH2-GFP cell line…………………………………………………………………43
Figure 3.6: Plots of population average for Oct4, Sox2 and Nanog in ESCs that were treated with various concentrations of doxycycline for different durations……………………………………………………………………45
Figure 3.7: Perturbing the core circuit produces highly complex transcriptional dynamics. Dynamics of population average transcript levels show complex non-linear behaviors……………………46
Figure 3.8: Collapse of scaled Oct4-Sox2 transcript counts as a function of Nanog transcript counts. Two-dimensional transcriptional trajectory………………………………………………..47
Figure 3.9: Collapse of scaled Oct4-Sox2 transcript counts as a function of Nanog transcript counts. Two- dimensional transcriptional trajectory with the ‘average transcript velocity’ along the pathway………48
Figure 3.10: Restricted transcript space explored by the core pluripotency circuit…………………………….49
Figure 4.1: Doxycycline dose-dependent mRNA expression profiles of Sox2, Oct4 and Nanog in ESCs grown in serum containing medium. .................................................................63
Figure 4.2: Two-dimensional transcriptional trajectory of the NOS circuit ...........................................64
Figure 4.3: smFISH measurement of population average of Sox2, Oc4 and Nanog for the wild type KH2 cells ..................................................................................................................65
Figure 4.4: Single-cell histograms of Sox2, Oct4 and Nanog transcripts in wild type KH2 cells at t = 12hr ...............................................................................................................................66
Figure 4.5: smFISH measurement of population average of Sox2, Oc4 and Nanog for the three inducible cell lines grown in serum containing medium ..................................................................................70
Figure 4.6: Single-cell histograms of Sox2, Oct4 and Nanog transcripts (after normalization by cell volume) in three inducible cell lines grown in serum containing medium ..................................................71
Figure 4.7: smFISH measurement of population average of Sox2, Oc4 and Nanog for the three inducible cell lines grown in 2i medium ..................................................................................................73
Figure 4.8: Single-cell histograms of Sox2, Oct4 and Nanog transcripts (after normalization by cell volume) in three inducible cell lines grown in 2i medium ..................................................................................74
Figure 4.9: Single-cell histograms of Nanog transcripts in unperturbed population of cells ..........................................................76
Figure 4.10: smFISH/IF images of individual unperturbed ESCs ........................................................................79
Figure 4.11: Sox2 antibody optimization and quantification ...........................................................................80
Figure 4.12: Oct4 antibody optimization and quantification ...........................................................................81
Figure 4.13: Nanog antibody optimization and quantification ...........................................................................82
Figure 4.14: Scatter plots of IF intensity and mRNA counts normalized by cell volume for Sox2, Oct4 and Nanog in unperturbed ESCs ..................................................................................................83
Figure 4.15: IF measurement of population average of Sox2, Oc4 and Nanog for the three inducible cell lines grown in serum containing medium ..................................................................................84
Figure 4.16: Single-cell histograms of Sox2, Oct4 and Nanog IF intensity (after normalization by cell volume) in three inducible cell lines ..................................................................................85
Figure 4.17: IF measurement of population average of Sox2, Oc4 and Nanog for the three inducible cell lines grown in 2i medium ..................................................................................................87
Figure 4.18: Western blot analysis of Oct4, Sox2 and Nanog expression in the three inducible cell lines grown in serum containing medium for various durations with the indicated concentrations of doxycycline ............................................................................................................88
Figure 4.19: smFISH and Immunofluorescence (IF) population average data for Oct4 knockdown cells grown in serum containing medium ..............................................................................................91
Figure 4.20: Plots of coefficients of variation (CV) of Sox2, Oc4 and Nanog for the three inducible cell lines grown in 2i medium ..........................................................................................................92

Figure 5.1: Schematic representation of the construct used to generate the three inducible cell lines in this study and two potential approaches to measure exo- and endo-genous gene expressions of Sox2, Oct4 and Nanog .........................................................................................................................104
Figure 5.2: Schematic of smFISH probe design to detect exo- and endo-genous gene expressions of Oct4, Sox2 and Nanog ..........................................................107
Figure 5.3: Using unperturbed cells as a negative control to test binding specificity of the “CMV promoter” smFISH probe ..................................................................108
Figure 5.4: Distribution of mRNA transcripts using identification and endogenous probes in the three inducible cell lines grown in serum containing medium with no doxycycline added ......................... 109
Figure 5.5: Portion of detected dots .....................................................................111
Figure 5.6: Scatter plots of the quantified relative intensities of Alexa 594 and Cy5 signals for Sox2 (a), Oct4 (b) and Nanog (c) transcripts respectively in cells that express both the exogenous (green) and endogenous (magenta) form of a particular factor .......................................................112
Figure 5.7: The average correct assignment rates for Oct4 (4 probes), Sox2 (14 probes) and Nanog (9 probes) quantified with (orange) and without (blue) information from the “identification” channel .............113
Figure 5.8: smFISH population average data for the three inducible cell lines ..........................................................115
Figure 5.9: Single-cell histograms of Sox2, Oct4 and Nanog transcripts (total, exogenous, endogenous) in three inducible cell lines grown in serum containing medium ..............................................116
Figure 5.10: Single-cell scatterplots of Sox2, Oct4 and Nanog mRNA transcripts (exogenous versus endogenous) in three inducible cell lines grown in serum containing medium ...................................... 118
Figure 5.11: Single-cell histograms of Sox2, Oct4 and Nanog transcripts (total, exogenous, endogenous) in three inducible cell lines grown in 2i medium ................................................................119
Figure 5.12: Single-cell scatterplots of Sox2, Oct4 and Nanog mRNA transcripts (exogenous versus endogenous) in three inducible cell lines grown in 2i medium .......................................................121
Figure 5.13: smFISH population average data for the doxycycline removal assay in serum containing medium ..................................................................................................................125
Figure 5.14: smFISH population average data of total, exo- and endo-genous transcripts of Sox2, Oct4 and Nanog for the doxycycline removal assay in serum containing medium ..........................126
Figure 5.15: Schematic of the core pluripotency circuit formed by Nanog, Oct4 and Sox2 ..................127
Figure 5.16: Schematic representation of the potential binding sites of Sox2, Oct4 and Nanog to their own promoters ........................................................................................................127

Figure 6.1: Stem cell regulatory networks (adapted from MacArthur et al., 2009) ............................138
Figure 6.2: Core pluripotency circuit formed by Nanog, Oct4 and Sox2 ..................................................138
Figure 6.3: Examples of proposed architecture of the NOS circuit ..........................................................139
Figure 6.4: Matrix to generate NOS network diagram .............................................................................140
Figure 6.5: Inferring the regulatory network within the NOS network using the MRA algorithm ........141
List of Tables

Table 2.1: Population average of mRNA transcripts of Oct4, Sox2 and Nanog upon RA treatment........9
Table 2.2: Population median of mRNA transcripts of Oct4, Sox2 and Nanog upon RA treatment........10
Table 2.3: Percentage of down-regulation of Oct4, Sox2 and Nanog transcripts compared with those in the control samples.................................................................12
Table 2.4: Proportion of cells in the 7-day RA-treated sample expressing certain range for Oct4, Sox2 and Nanog transcripts.................................................................................................................15
Table 2.5: Population average of Oct4, Sox2 and Nanog transcripts upon RA treatment followed by RA removal..............................................................................................................................21
Table 4.1: Correlation coefficient of transcript profiles for all pairs of genes in the NOS circuit for the three inducible cell lines..........................................................................................................................89
Table 4.2: Correlation coefficient of protein profiles for all pairs of genes in the NOS circuit for the three inducible cell lines..........................................................................................................................89
Table 4.3: Correlation coefficient between smFISH and IF population average data for the three inducible cell lines grown in serum containing medium.........................................................................................90
Table 5.1 Sequences used to design smFISH probes to measure exo- and endo-genous gene expressions of Sox2, Oct4 and Nanog.........................................................................................................................105
Table 5.2: smFISH probe sequences used in this study to measure exo- and endo-genous gene expressions of Sox2, Oct4 and Nanog.........................................................................................................................126
Table 5.3: Percentage of endogenous transcripts of Sox2, Oct4 and Nanog compared with those in the unperturbed cells across various time points in the doxycycline removal assay........................................106
Acknowledgments

First and foremost, I would like to thank Alexander van Oudenaarden, my thesis advisor and the world’s leading figure in the field of systems biology, for his guidance and support, and the freedom to pursue my research interests. I have always been impressed by his scientific instincts and vision in research directions. I feel very lucky to be able to learn from him and the talented people in his lab. Although having many professional and personal commitments, Alexander always finds time to discuss my research projects, providing guidance and encouragements. I realize this experience is rare, and I will forever be grateful for this opportunity and for the experiences I had while working in the van Oudenaarden Laboratory.

I would like to thank the members of my committee, Jeremy Gunawardena, Laurie A. Boyer, Leonid Mirny and James M. Hogle. During the dynamic latter stages of my PhD, my committee offered additional scientific guidance and support needed to reach the finish line. I am grateful to Dr. Jeremy Gunawardena, for his guidance, insights and criticism. I was very moved when Jeremy was still willing to spend time giving me advices when his mom was under critical medical condition in May this year. I would also like to thank Dr. Laurie A. Boyer, for whom taught me how to be a better scientist which is a priceless gift that I was lucky to receive. I learned much during our interactions and will use these tools in my future career. I am also grateful towards Dr. Leonid Mirny and Dr. James M. Hogle for serving on my thesis committee and providing me with invaluable comments.

I would like to express my heartfelt gratitude to Hyun Youk, my collaborator in the past three years. I am amazed and motivated by his passion for science, sharp intuition and persistence to success. It has been a really enjoyable and rewarding experience working with Hyun on my main thesis project. He is always willing to offer help and useful advice despite a super busy transition period from postdoc to junior young faculty.
Science is rarely done in a vacuum. I want to thank the members of the van Oudenaarden Laboratory, both past and present. Thank you so much to Philipp J. Junker for his training on smFISH when I first joined the lab, and to Magda Bienko and Nicola Crosetto for sharing their expertise on various molecular biology techniques. I especially want to thank Yannan Zheng, Jialing Li, Miaoqing Fang, Annalisa Pawlosky, Sandy Klemm, Clinton Hansen, Apratim Sahay and Dylan Mooijman not only for their scientific and technical aid, but also for their friendship. Thank you to Monica Wolf for your support and making sure things always worked out.

Thank you so much to my friends, both at Harvard and MIT, within Boston and around the world. A special thanks to my best friend Kristen Bartoli, her fiancé Jason Thomas and their lovely dog Dazzle. They very generously gave me lots of encouragement, listened to my distress, and shared my happiness – I always enjoy talking to them and thank them for their kindness. I will definitely miss the daily coffee trip to Dunkin' Donuts (we called it “DD” for short) with Kristen!

Most importantly, thank you to my family. Thank you to my father. Dad, you are one who continues to dedicate his life’s work, both professionally and personally, to helping others. Your strength and intelligence continue to amaze me. Thank you to my mom, for whom is there is not enough thanks. Thank you for pushing me to try new things, to never give up, to reach higher and to continue to explore in life. Thank you for your love, support, and disciplinary actions that have made me a better person. I will always love you, to whom this thesis is dedicated.
Chapter 1 Introduction

Although the importance of Oct4, Sox2 and Nanog for the regulation of pluripotency and differentiation in ESCs is well established, their mutual interaction and the resulting regulatory dynamics are only partially understood and controversially discussed. Relying on the bulk assays to study the core pluripotency circuit could be misleading because most of those techniques are not sensitive enough to detect subtle changes in gene expressions and also do not capture the vast transcript space that individual cells within the population span. In this thesis, we adapted a high resolution smFISH technique to directly measure the enormous heterogeneity in the levels of the core pluripotency circuit’s transcripts in individual cells upon various types of perturbations such as differentiation, induction or knockdown of one of the three pluripotent factors. A simplified model of the pluripotency circuit is thought to consist of the three factors with nine regulatory links among them. When the network is perturbed, the strengths of these interaction links will change over time and the whole circuit will be rewired in complicated ways that are not yet well understood. With feedback loops, it will be even more challenging to unravel an already simplified three-factor network. To obtain a more comprehensive picture and unprecedented details of the core pluripotency regulatory circuit, we developed modified smFISH strategies to 1) measure mRNA and protein simultaneously in single ESCs; 2) accurately quantify exo- and endo-genous transcripts that varied by only short sequences. On top of the accurate and quantitative single-cell measurements, we built robust network connectivity matrices that helped us identify a few interesting novel features of the core pluripotency circuit.

1.1 Thesis Contributions

The overall objective of this research was to apply a novel single molecule technique to gain a deeper understanding of the dynamics of gene regulatory network formed by Oct4, Sox2 and Nanog in
ESCs. The study of how the three factors respond to various regulatory signals will shed light on the architecture of the core pluripotency network. To achieve this objective, we utilized novel tools from molecular biology in combination with smFISH technique that allows for the accurate quantification of individual molecules within cells. The aims of this research that are described in the chapters of this thesis are as follows:

1. Explore the precise effects of retinoic acid signaling on the differentiation and cell lineage decisions of ESCs (Chapter 2).

2. Experimentally reveal a restricted transcript space formed by the core pluripotency circuit (Chapter 3).

3. Modify high resolution smFISH approach to measure mRNA and protein simultaneously in single ESCs to reveal novel features of the core pluripotency circuit (Chapter 4).

4. Investigate autoregulation of the three factors by incorporating a novel modification into smFISH technique which accurately quantifies exo- and endo-genous transcripts that varied by only short sequences (Chapter 5).

5. Retrieve the topology and strengths of network connections of the core pluripotency circuit from experimentally measured network responses (Chapter 6).

1.2 Thesis Outline

The thesis is structured into seven chapters, which includes one brief introduction chapter, five chapters that focus on the development and implementation of smFISH for studies of gene expression of the core pluripotency circuit upon various types of perturbations, followed by one conclusion chapter.
Chapter 2 focuses on the realization of using smFISH to study the effects of RA-induced differentiation in ESCs. It begins with a brief literature review of the important roles of RA in various biological processes, especially neural differentiation of ESCs. To quantify the effect of RA on the differentiation of ESCs and with the hope of finding a “threshold” for the responses, two types of RA-induced differentiation assays are performed: 1) the cells are exposed to RA for a range of time durations; 2) the cells are treated with RA for various durations followed by RA removal. The results of this work suggest that upon the addition of a single chemical cue (RA in this case) the stem cells could develop from a homogenous starting population to a mixture of different cell types.

Chapter 3 focuses on revealing dynamics on transcript pathway formed by Oct4, Sox2 and Nanog. This chapter briefly reviews the current challenges in building quantitative models to understand complex pluripotency network in ESCs, followed by descriptions of how to perturb the core pluripotency circuit by using three stable cell lines each of which has an extra copy of one of the three pluripotency genes. The quantitative data from smFISH measurement is then used to construct a two-dimensional transcriptional trajectory. The results suggest that perturbing the core circuit over a wide temporal range and strengths tunes the range of heterogeneity in transcript levels. The unperturbed core circuit spans near maximal volume in transcript space and could detect overexpression of the three factors and respond by turning down the expression levels to a lower but non-zero levels.

Chapter 4 is the continuation of Chapter 3 by improving the experimental setups based on the discussion about the potential pitfalls at the end of Chapter 3. In this part of study, the ESCs are grown on feeders in serum containing medium or in 2i medium without feeders for shorter period of time (up to 24hrs). A modified smFISH approach is adapted to measure mRNA and protein simultaneously in single ESCs. The results reveal a few striking features of the core pluripotency circuit: 1) The circuit’s transcript space might have a threshold that restricts how high the transcription levels could overshoot beyond those in the unperturbed cells; 2) Sox2 and Oct4 behave in a symmetric way in response to perturbations. On the other hand, their expression profiles become anti-correlated at both transcript and protein levels. 3)
Perturbation of Nanog has minimal influence on the expression level and transcription heterogeneity of Oct4 and Sox2.

**Chapter 5** experimentally tests the hypothesis mentioned in Chapter 4 that the core pluripotency circuit exerts a negative feedback regulation loop to keep the expression of each of its components at a steady level. A novel modification is incorporated into the existing smFISH technique to accurately quantify exo- and endo-genous transcripts that varied by only short sequences. The experimental results convincingly validate our hypothesis. When the expression level of one of the factors in the core circuit rises above a certain level, it represses its endogenous promoter to restrict how high the transcription level can overshoot beyond that in the unperturbed cells.

**Chapter 6** focuses on building simple yet robust models based on accurate and quantitative experimental measurements (from Chapter 4 and 5) to illustrate the interconnection structure of the core pluripotency network. Our results suggest that most of the regulatory interactions are inhibitory among the three factors. The network connectivity generated by the Modular Response Analysis (MRA) algorithm helps us retrieve much more detailed information about the architecture of the core circuit.

**Chapter 7** describes final conclusion and future directions for these projects.
Chapter 2 Effects of all-trans retinoic acid on the differentiation of mouse embryonic stem cells

2.1 Introduction

Embryonic stem cells (ESCs) are pluripotent cells derived from the inner cell mass (ICM) that are able to self-renew or undergo differentiation into cellular derivatives of all three primary germ layers depending on a complex interplay of extracellular signals and intracellular factors. They are considered to be a unique biological system for disclosing the mechanisms of pluripotency and lineage commitment, and a useful resource for disease modeling, pharmacological screening and cell therapy.

All-trans retinoic acid (ATRA), a biologically active form of vitamin A, plays an important role in many diverse biological processes, such as embryogenesis and cellular differentiation (Ross et al., 2000; Clagett-Dame and DeLuca, 2002; Pan and Baker, 2007). RA functions as a specific ligand for retinoic acid receptor signaling pathway, which is required in embryonic development, bone formation, and in the maintenance of normal epithelial structures (Mark et al., 2006; Duester, 2007; Niederreither and Dolle, 2008). It also plays a critical role in neuronal patterning, neural differentiation and axon outgrowth (Lee et al., 2009; Siegenthaler et al., 2009). It is of great importance to thoroughly understand the mechanism of action of RA on the development of the nervous system. RA regulates differentiation and metabolism through interaction with two of nuclear receptors families, the RAR receptors that bind and activate by both ATRA and 9-cis retinoic acid (9CRA), and the retinoid X receptors (RXR) that exclusively bind 9-cis-RA. Both receptors families have α, β, γ subtypes with varying expression patterns in different tissue and cells (Wan 1993; Smith et al. 1998).

ATRA is one of the strongest and most thoroughly studied differentiation inducers. Endogenous RA could be detected transiently during the stage of blastocyst before implantation (Rossant et al., 1991; Parrow et al., 1998; Clagett-Dame and DeLuca, 2002). However, the role of such transient RA expression in the stage of blastocyst is still unclear. In mESC in vitro differentiation model, RA induces embryonic
carcinoma (EC) and ESC commitment into a number of different cell types, such as primitive endoderm-like cells (Maden and Holder, 1992), cardiomyocytes (Wobus et al., 1997) and pancreatic β cells (Shi et al., 2005) in a time- and concentration-dependent manner. It has been reported that 0.5 µM RA induced differentiation of ESCs into neural cells in vitro (Herget et al., 1998; Wang et al., 2008). A brief exposure to RA was shown to facilitate neuronal maturation of ESCs (Kim et al., 2009). One study suggested that ATRA pre-treatment could promote neuronal differentiation of mesenchymal stem cells (MSCs) in vivo and improve its survival (Bi et al., 2010). All together the existing studies have demonstrated that RA plays a central role during embryo development and ESC differentiation. However, the effect and mechanism through RA signaling regulates ESC differentiation are still poorly understood.

In our study, single-cell measurement techniques were used to explore the precise effects of RA on neural differentiation of ESCs. In particular, we examined the gene expression profile of the three core pluripotency factors (Oct4, Sox2 and Nanog) upon RA treatment. People have noticed that ESCs, selected from the same culture dish are highly variable in gene expression profile, and global gene expression can measure only the average level of all the genes of a population. Our results demonstrated the strength of the single-cell technique to complement conventional techniques and to provide additional information about the expression patterns of Oct4/Sox2/Nanog during RA-induced differentiation.

We used a monolayer culture system of neural induction and treated the cultures with 0.5 µM of RA for various durations. Some studies have reported that low concentration of ATRA ranging from 0.01 to 1 µM improved neuronal differentiation efficiency (Bi et al., 2010; Woodbury et al., 2000). However, higher concentration of ATRA (e.g. 10 and 100 µM) resulted in significantly higher cell death rate than that of the controls (Diaz et al., 2006). One study suggested that RA treatment at 0.05 µM up to 24hr appeared to be an optimum concentration for maintaining undifferentiated state of ESCs (Wang et al., 2008). We chose to treat the ESCs with 0.5 µM of RA because it was supposed to be potent enough to induce neuronal differentiation while not causing massive cell death. The reason we did not carry out the experiments in embryonic bodies (EBs) was that it is difficult to dissect and manipulate differentiation
within EBs because they are multicellular agglomerations of extra-embryonic endoderm and definitive ectodermal, mesodermal and endodermal derivatives (Gordon et al., 1995).

To quantify the effect of ATRA on the differentiation of ESCs and with the hope of finding a “threshold”, we carried out two types of RA-induced differentiation assays. In the first part of our study, the ESCs were treated with 0.5µM of RA for a range of time durations (1, 3, 4, 5 and 7 days), and then at the end of each time point the gene expression profiles of Oct4, Sox2 and Nanog were examined. In the second part, we aimed to determine whether or not the RA-induced differentiation could be reversed upon removal of RA. The ESCs were first treated with 0.5 µM of RA for 13, 24, or 48 hrs, and at the end of each time point RA was removed and LIF was added back to the system. Then, the cells were collected 0hr, 12hrs, 24hrs, 2days, 4 days and 6 days post RA-removal. The combination of this differentiation protocol and techniques to read out the states of single cells allow us to study differentiation in unprecedented detail. Insights from our studies will also likely improve the production of clinically relevant cell types from ESCs.
2.2 Results

2.2.1 RA-induced differentiation assay without removing RA afterwards

The ESCs were treated with 0.5µM of RA in serum containing medium without LIF for a range of time durations. To be consistent with the standard protocol for the differentiation assay, the stem cells were cultured on gelatin coated petri dish without the presence of the feeder cells. The cells were collected after being treated with the RA for 1, 3, 4, 5 and 7 days. For the control, we cultured the ESCs exactly the same way as described above except that the same amount of 1xPBS instead of RA was added to the medium. The control sample was collected at the same time when RA was first added to the treatment samples, and it was denoted as “Day0” sample. This control might not be ideal, a better approach would be to collect cells that were not treated with RA at exactly the same time points as those treatment samples. Thus, besides “Day0” we should also include control samples at day 1, 3, 4, 5 and 7. Once all the samples had been collected, the mRNA expression profiles of Oct4, Sox2 and Nanog were then examined using 3-color smFISH imaging. For this particular set of experiment, raw counts of mRNA transcripts were measured using ImageM in MatLab, and the effect of cell size was not taken into account.

2.2.1.1 Oct4, Sox2 and Nanog mRNA expression profiles all showed a monotonic decrease

Table 2.1 and Fig. 2.1 provided a summary of the population average of the mRNA counts for Oct4, Sox2 and Nanog. The control sample (“Day0”) on average expressed 245 Oct4 transcripts, 124 Sox2 transcripts and 58 Nanog transcripts. Upon RA treatment, all three factors showed a monotonic decrease in the mRNA expression. By day 7, on average individual ESCs only expressed 6 Oct4 transcripts, 8 Sox2 transcripts and 4 Nanog transcripts. These results are consistent with what have been reported in the existing literature that stem cells tend to lose pluripotency factors (e.g. Oct4, Sox2 and Nanog) and to differentiate after being treated with RA (Lu et al., 2009). Besides the mean, another way to examine the mRNA expression profiles is to look at the median. Similarly, Table 2.2 and Fig. 2.2
suggested that the population median of the three factors also decreased monotonically upon RA treatment. We noticed that 24hr RA treatment was enough to cause a more than 50% down-regulation of the mRNA expression for Oct4, Sox2 and Nanog. Our results disagree with one exiting study that has suggested that short-term treatment with RA in the early stage of differentiation is able to prevent spontaneous differentiation of ESCs and mains self-renewal (Wang et al., 2008). More specifically, in their study most of the ESCs maintained compact colonies after the cells were exposed to 0.05 – 1 µM of RA for the first 24hr of spontaneous differentiation process. However, our data are consistent with their claim that treatment of RA at 0.05 -1 µM for 48hr or longer had no differentiation-inhibitory effect. One interesting follow-up experiment would be to treat the ESCs with 0.5 µM of RA for less than 24 hours (e.g. 2, 4, 8 and 12 hours) to quantify the extent of gene down-regulation and hopefully to find a “threshold” of the RA-induced differentiation effect on the stem cells. We could also add one more dimension to the differentiation assay by treating the ESCs with a range of concentrations of RA (e.g. 0.05 – 1 µM).

Table 2.1: Population average of mRNA transcripts of Oct4, Sox2 and Nanog upon RA treatment. “Day 0” denotes the control sample that was not treated with RA. “Day 1” denotes the sample that was treated with 0.5 µM of RA for 1 day. The denotation is similarly applied to “Day 3”, “Day 4”, “Day 5”, and “Day 7”.

<table>
<thead>
<tr>
<th></th>
<th>Oct4</th>
<th>Sox2</th>
<th>Nanog</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day 0</td>
<td>245</td>
<td>124</td>
<td>58</td>
</tr>
<tr>
<td>Day 1</td>
<td>55</td>
<td>46</td>
<td>29</td>
</tr>
<tr>
<td>Day 3</td>
<td>4</td>
<td>27</td>
<td>3</td>
</tr>
<tr>
<td>Day 4</td>
<td>2</td>
<td>21</td>
<td>2</td>
</tr>
<tr>
<td>Day 5</td>
<td>3</td>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>Day 7</td>
<td>6</td>
<td>8</td>
<td>4</td>
</tr>
</tbody>
</table>
Table 2.2: Population median of mRNA transcripts of Oct4, Sox2 and Nanog upon RA treatment. “Day 0” denotes the control sample that was not treated with RA. “Day 1” denotes the sample that was treated with 0.5 µM of RA for 1 day. The denotation is similarly applied to “Day 3”, “Day 4”, “Day 5”, and “Day 7”.

<table>
<thead>
<tr>
<th></th>
<th>Oct4</th>
<th>Sox2</th>
<th>Nanog</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day 0</td>
<td>236</td>
<td>123</td>
<td>34</td>
</tr>
<tr>
<td>Day 1</td>
<td>48</td>
<td>39</td>
<td>14</td>
</tr>
<tr>
<td>Day 3</td>
<td>0</td>
<td>18</td>
<td>1</td>
</tr>
<tr>
<td>Day 4</td>
<td>0</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>Day 5</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Day 7</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 2.1: Plots of population average of Oct4, Sox2 and Nanog transcripts in the ESCs that were treated with 0.5 µM of RA for various durations. Each data point represents the population average (total mRNA transcript counts) at a particular time point.
2.2.1.2 Sox2 was most gradually down-regulated among the three factors

As discussed in the previous section, all the three factors were down-regulated upon RA treatment. However, Table 2.3 suggested that Oct4, Sox2 and Nanog responded to RA treatment at different rates. More specifically, Sox2 was down-regulated most gradually compared with Oct4 and Nanog. Histograms in Fig. 2.3 further illustrated the differential down-regulation rates. Sox2 mRNA distribution gradually shifted towards the left from day 0 to day 7. On the other hand, both Oct4 and Nanog mRNA showed more than 95% down-regulation compared with the control sample by day 3. This interesting observation might be explained by the possibilities that the three factors are regulated by RA in a sequential manner or just simply that their mRNA transcripts have different half-lives.
Table 2.3: Percentage of down-regulation of Oct4, Sox2 and Nanog transcripts compared with those in the control samples. “Day 1” denotes the sample that was treated with RA for 1 day. The denotation is similarly applied to “Day 3”, “Day 4”, “Day 5”, and “Day 7”. The percentage is calculated based on the formula 
\[
\frac{(\text{number of transcripts in the treatment sample} - \text{number of transcripts in the control sample})}{\text{number of transcripts in the control sample}} \times 100
\]

<table>
<thead>
<tr>
<th></th>
<th>Oct4</th>
<th>Sox2</th>
<th>Nanog</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day 1</td>
<td>- 77.5%</td>
<td>- 62.8%</td>
<td>- 49.6%</td>
</tr>
<tr>
<td>Day 3</td>
<td>- 98.5%</td>
<td>- 77.9%</td>
<td>- 95%</td>
</tr>
<tr>
<td>Day 4</td>
<td>- 99.2%</td>
<td>- 83%</td>
<td>- 96.5%</td>
</tr>
<tr>
<td>Day 5</td>
<td>- 98.8%</td>
<td>- 90.6%</td>
<td>- 95.1%</td>
</tr>
<tr>
<td>Day 7</td>
<td>- 97.7%</td>
<td>- 93.9%</td>
<td>- 93.7%</td>
</tr>
</tbody>
</table>

Figure 2.3: Single-cell histograms of Oct4, Sox2 and Nanog transcripts in the ESCs for the RA treatment experiment. Transcript counts of Oct4 (blue), Sox2 (red) and Nanog (green) in individual ESCs that make up the unperturbed population of cells ([RA] = 0 µM) or perturbed populations ([RA] = 0.5 µM). “Day 0” denotes the control sample that was not treated with RA. “Day 1” denotes the sample that was treated with 0.5 µM of RA for 1 day. The denotation is similarly applied to “Day 3”, “Day 4”, “Day 5”, and “Day 7”.
Figure 2.3 (Continued): Single-cell histograms of Oct4, Sox2 and Nanog transcripts in the ESCs for the RA treatment experiment. Transcript counts of Oct4 (blue), Sox2 (red) and Nanog (green) in individual ESCs that make up the unperturbed population of cells ( [RA] = 0 µM ) or perturbed populations ( [RA] = 0.5 µM ). “Day 0” denotes the control sample that was not treated with RA. “Day 1” denotes the sample that was treated with 0.5 µM of RA for 1 day. The denotation is similarly applied to “Day 3”, “Day 4”, “Day 5”, and “Day 7”. 
2.2.1.3 Oct4, Sox2 and Nanog mRNA expression did not completely shut down after 7-day RA treatment

Table 2.3 showed that the mRNA expression level of the three factors decreased ~93-98% compared with the control sample by day7. It suggested that even after treating the ESCs with RA for such a long period of time, a small population of the cells still expressed some or all of the three pluripotency factors. For example, 4% of the population expressed more than 50 Oct4 mRNA transcripts, 3% of the population had more than 50 Sox2 transcripts, and 2.5% of the total 268 cells still had more than 50 Nanog mRNA transcripts. These results could be further supported by the long-tailed distribution of Oct4, Sox2 and Nanog at day7 shown in Fig. 2.3. Moreover, the three-dimensional scatterplot for day7 RA-treated sample in Fig. 2.4 again illustrated the fact that the NOS mRNA expressions in a small subpopulation of the ESCs did not completely disappear even after 7-day RA treatment. In contrast, some existing studies that used bulk measurements (e.g. qRT-PCR, Northern blot) reported that upon stimulation with RA, Oct4 expression was down-regulated more rapidly in RA-treated cultures and could not be detected by the third day of RA treatment (Lu et al., 2009; Wang et al., 2008). Once again the smFISH technique was proved to be much more sensitive and quantitative compared with conventional techniques.

Our results suggested that after adding a single chemical cue the stem cells could develop from a homogeneous starting population to a mixture of different cell types. Using single-cell techniques to measure the state (i.e. gene expression) of single cells we were able to show that two mutually exclusive subpopulations coexist upon prolonged RA treatment. One potential interesting experiment is to isolate that “persistent” subpopulation and to conduct a genome-wide assay to compare its gene expression profiles with the subpopulation that has undetectable NOS expression as well as with non-RA treated control sample. One possible observation would be that cells in the “persistent” subpopulation also express lineage-specific markers despite the relatively high NOS gene expression. All together these
experiments could potentially show that the decision to develop into a particular cell type, the actual change of the cell state, and commitment to the chosen cell type are separate events.

Table 2.4: Proportion of cells in the 7-day RA-treated sample expressing certain range of Oct4, Sox2 and Nanog transcripts. A total of 268 cells were analyzed for the 7-day RA-treated sample.

<table>
<thead>
<tr>
<th>mRNA counts</th>
<th>Oct4</th>
<th>Sox2</th>
<th>Nanog</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>196 (73%)</td>
<td>102 (38%)</td>
<td>136 (50%)</td>
</tr>
<tr>
<td>1-20</td>
<td>60 (22%)</td>
<td>147 (55%)</td>
<td>122 (46%)</td>
</tr>
<tr>
<td>21-50</td>
<td>2 (1%)</td>
<td>11 (4%)</td>
<td>4 (1.5%)</td>
</tr>
<tr>
<td>51-100</td>
<td>3 (1%)</td>
<td>2 (1%)</td>
<td>4 (1.5%)</td>
</tr>
<tr>
<td>&gt; 100</td>
<td>7 (3%)</td>
<td>6 (2%)</td>
<td>2 (1%)</td>
</tr>
</tbody>
</table>

Figure 2.4: Three-dimensional scatterplot of Oct4, Sox2 and Nanog transcripts for the 7-day RA-treated sample. A total of 268 cells were analyzed. x-axis represents number of Nanog transcripts, y-axis represents number of Oct4 transcripts, z-axis represents number of Sox2 transcripts.
2.2.1.4 NOS transcriptional dynamics upon RA treatment

In our study, directly measuring the enormous heterogeneity in the levels of the core pluripotency circuit’s transcripts among single cells led us to reveal a restricted transcript space formed by the core circuit in the ESCs during RA-induced differentiation process. Fig. 2.5 demonstrated how the heterogeneous transcript levels among individual cells within a population changed upon RA treatment. During the time course experiment, the single-cell “cloud” in the transcript space shifted towards the zero-coordinates corner. This was consistent with the fact that ESCs tend to gradually down-regulate the NOS network upon treatment with RA. With the three-dimensional scatterplot, we again demonstrated that relying solely on the bulk measurements to study the core pluripotency circuit could be misleading because the average gene expression levels do not capture the vast transcript space that individual cells within the population span.

![Figure 2.5: Three-dimensional scatterplot of Oct4, Sox2 and Nanog transcripts for the RA treatment time course experiment. x-axis represents number of Nanog transcripts. y-axis represents number of Oct4 transcripts. z-axis represents number of Sox2 transcripts. “Day 0” denotes the control sample that was not treated with RA. “Day 1” denotes the sample that was treated with 0.5 µM of RA for 1 day. The denotation is similarly applied to “Day 3”, “Day 4”, “Day 5”, and “Day 7”.

]
2.2.1.5 Single-cell scatterplots suggested the potential effect of RA on the differentiation of ECSs

Our results are consistent with what have been reported in the existing literature that RA induces ESCs to neural differentiation with reduced Oct4 expression (Bi et al., 2010; Kim et al., 2009; Herget et al., 1998; Wang et al., 2008). The correlated expression pattern of the pluripotency factors in the non-RA-treated ESCs (first column in Fig. 2.6) contrasted sharply with those in the RA-treated cells (column 2-6 in Fig. 2.6). In particular, Oct4 and Sox2 showed a strong positive correlation in the control sample. After the addition of RA, their mRNA levels diverged in time, the positive correlation gradually disappeared and by day3 majority of the cells within the population still expressed some Sox2 but very little Oct4. Our observation of Sox2 being expressed in a complementary pattern to Oct4 is again consistent with what other groups have suggested that differential modulation of Oct4 and Sox2 precedes cell fate selection (Thomson et al., 2011; Ema et al., 2008; Han et al., 2010; Jiang et al., 2008; Pasini et al., 2010; Peng et al., 2009; Schuettengruber and Cavalli, 2009; Silva and Smith, 2008). More specifically, when these two factors are differentially regulated leading to high Oct4 and low Sox2 levels, or low Oct4 and high Sox2 levels, either the mesendodermal (ME) fate or the neural ectodermal (NE) fate becomes available to the cell. There are two possible mechanisms that how various differentiation signals induce the ESCs towards a specific cell fate. One possibility is that ECSs initially express markers from all the three germ layers (i.e. endoderm, ectoderm and mesoderm); upon induced differentiation, expression of some of the germ layer(s) markers are down-regulated whereas the other germ layer(s) marks are significantly up-regulated. The alternative mechanism is that ESCs initially do not express any of the germ layer markers; after the addition of some differentiation signal to the cells, makers of particular germ layer(s) are turned on. To get a fuller understanding of the underlying mechanism, for the future experiments it would be interesting to examine the expression pattern of various differentiation markers in the ESCs treated with RA and then compare the results with those non-RA-treated cells. For example, we might expect to observe co-expression of Sox2 and NE lineage specific markers such as Sox1.
Figure 2.6: Single-cell pairwise scatterplots of Oct4, Sox2 and Nanog transcripts for the RA treatment time course experiment. “Day 0” denotes the control sample that was not treated with RA. “Day 1” denotes the sample that was treated with 0.5 μM of RA for 1 day. The denotation is similarly applied to “Day 3”, “Day 4”, “Day 5”, and “Day 7”. 1st row: scatterplots of Sox2 transcripts (y coordinate) versus Oct4 transcripts (x coordinate). 2nd row: scatterplots of Nanog transcripts (y coordinate) versus Oct4 transcripts (x coordinate). 3rd row: scatterplots of Nanog transcripts (y coordinate) versus Sox2 transcripts (x coordinate).
2.2.2 RA-induced differentiation assay with RA removal afterwards

In section 2.2.1, we have demonstrated that even a short-term RA treatment (24 hrs) was enough to cause a more than 50% down-regulation of the mRNA expression of Oct4, Sox2 and Nanog in the ESCs. By the third day of RA treatment, both Oct4 and Nanog mRNA showed more than 95% down-regulation, 78% decrease for Sox2 compared with the control sample. These results all together suggested that majority of the cells in the RA-treated ESCs population started to shut down the NOS circuit expression, prepared for or already committed to a specific cell fate. The next question we would like to explore was whether the RA-induced differentiation effect observed in section 2.2.1 could be reversed. The answers could potentially help us understand the following three important questions: 1) Is the differentiation induced by RA terminal? 2) Are there several transition states in between the starting pluripotent ESCs population and the final differentiated cells? 3) Is there a threshold that determines whether or not the RA treatment induces terminal differentiation?

In section 2.2.2, we conducted RA treatment followed by RA removal to determine whether the differentiation could be reversed upon removing RA. The ESCs were first treated with 0.5 µM of RA in serum containing medium without LIF for 13, 24, or 48 hrs. Similar to the culture conditions described in section 2.2.1, the stem cells were cultured on gelatin coated petri dish without the presence of the feeder cells. At the end of each RA treatment time point, RA was removed and LIF was added back to the culture medium. Then, the cells were collected 0 hr, 12 hrs, 24 hrs, 2 days, 4 days and 6 days post RA-removal. For the control, we cultured the ESCs exactly the same way as described above except that the same amount of 1xPBS instead of RA was added to the medium. The control sample was collected at the same time when RA was first removed from the treatment samples, thus it was denoted as “Day 0” sample. Again, this type of control might not be ideal. A better approach would be to collect cells that were not treated with RA at exactly the same time points as those treatment samples. Thus, besides “Day 0” we should also include control samples at 12 hrs, 24 hrs, 2 days, 4 days and 6 days post RA-removal. Once all the samples had been collected, the mRNA expression profiles of Oct4, Sox2 and Nanog were then...
examined using 3-color smFISH imaging. For this particular set of experiment, number of mRNA transcripts were measured using ImageM in MatLab, and the effect of cell size was not taken into account.

2.2.2.1 Oct4, Sox2 and Nanog mRNA expressions were all down-regulated at the population level

Table 2.4 and Fig. 2.7 gave us a summary of the population average gene expression of Oct4, Sox2 and Nanog for the RA removal experiment. Similar to what we observed in section 2.2.1, the three factors were down-regulated for all the experimental conditions tested here. A 13hrs short exposure to RA was already enough to cause a big decrease in Oct4, Sox2 and Nanog gene expression. More specifically, even when we removed RA and added back LIF after inducing the cells with RA for 13hrs, the transcript levels of the three pluripotency factors still kept going down and eventually approached almost zero 6days post RA removal. Our results suggested that 13hrs of RA treatment was sufficient to induce terminal differentiation in majority of the cells within the population. Moreover, the down-regulation pattern obtained from the 13hrs RA treatment experiment was quite similar to those from the 24hrs or 2day RA treatment experiments. Again, it suggested that 13hrs of RA exposure was able to induce differentiation in the ESCs to almost the same extent as the longer RA treatment conditions (e.g. 24hrs, 2days). Thus, to find the “threshold” of RA-induced differentiation effect, in the future experiments we could try shorter RA exposures (e.g. 2hrs, 4hrs, 8hrs, etc). We could also add one more dimension to further fine tune the search for “threshold” by treating the ESCs with a range of concentrations of RA (e.g. 0.05 – 1 µM).
Table 2.5: Population average of Oct4, Sox2 and Nanog transcripts upon RA treatment followed by RA removal. 1st column: duration of RA treatment on the ESCs. 2nd column: duration of the cells cultured in the LIF containing medium post RA removal.

<table>
<thead>
<tr>
<th>RA treatment</th>
<th>after RA removal</th>
<th>Oct4</th>
<th>Sox2</th>
<th>Nanog</th>
</tr>
</thead>
<tbody>
<tr>
<td>0hr</td>
<td>0hr</td>
<td>224</td>
<td>114</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>0hr</td>
<td>131</td>
<td>70</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>12hr</td>
<td>137</td>
<td>40</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>24hr</td>
<td>134</td>
<td>24</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>2days</td>
<td>146</td>
<td>16</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>4days</td>
<td>90</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>6days</td>
<td>6</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>13hr</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0hr</td>
<td>159</td>
<td>63</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>12hr</td>
<td>130</td>
<td>40</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>24hr</td>
<td>100</td>
<td>27</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>2days</td>
<td>60</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>4days</td>
<td>9</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>6days</td>
<td>3</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>24hr</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0hr</td>
<td>132</td>
<td>37</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>12hr</td>
<td>110</td>
<td>23</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>24hr</td>
<td>70</td>
<td>19</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>2days</td>
<td>15</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>4days</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>2days</td>
<td>6days</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>
Figure 2.7: Plots of population average of Oct4, Sox2 and Nanog transcripts in the ESCs that were first treated with 0.5 µM of RA for various durations followed by RA removal. Each data point represents the population average (total mRNA transcript counts) at a particular time point post RA removal. Black dotted lines correspond to data obtained from the control sample (non-RA treated). Blue lines correspond to data obtained from the cells treated with RA for 13hrs. Red lines correspond to data obtained from the cells treated with RA for 24hrs. Magenta lines correspond to data obtained from the cells treated with RA for 2days.

2.2.2.2 Oct4 was down-regulated more gradually when treated with RA for 13hrs compared with 24hr or 2day RA treatment

The blue line in the left plot of Fig. 2.7 showed that Oct4 transcripts first showed a decrease after 13hr RA exposure and stayed at almost the same expression level for 2 days post RA removal followed by a monotonic decrease all the way till day 6. The “2day delay” was not observed for samples treated with RA for 24hrs (red line) or 2days (magenta line). The blue histograms in Fig. 2.8 further demonstrated the “2day delay” of Oct4 down-regulation for the cells treated with RA for only 13hrs and that was not observed in the 24hr or 2day RA treated samples. These results suggested that there might be a “temporary threshold” for the RA-induced differentiation effect on Oct4 gene expression. Moreover, the three colored lines in the left plot of Fig. 2.7 demonstrated that Oct4 transcript level was down-regulated
at differential rates when the ESCs were treated with RA for various durations. More specifically, Oct4
gene expression was more rapidly down-regulated upon longer RA exposure. To further fine tune the RA-
exposure-dependent down-regulation of Oct4 transcripts, one interesting experiment we could do is to
include shorter RA exposure conditions (e.g. 2hrs, 4hrs, 8hrs, etc) as well as longer exposures (e.g. 3days,
4days, etc).

2.2.2.3 Sox2 and Nanog responded to RA treatment differently from Oct4

Unlike what we described in section 2.2.2.2 for Oct4, both Sox2 and Nanog transcripts did not
show the temporal delay in gene down-regulation upon 13hr treatment with RA. Instead, Sox2 and Nanog
gene expression demonstrated a monotonic decrease for all the three RA treatments (middle and right
plots in Fig. 2.7). The red (Sox2) and green (Nanog) histograms in Fig. 2.8 further demonstrated the
“2day delay” observed for Oct4 down-regulation for the cells treated with RA for only 13hrs was not
detected for Sox2 and Nanog expression. Moreover, the duration of RA treatment seemed to have a non-
significant effect on the rate of down-regulation of both Sox2 and Nanog transcripts. The three colored
lines in the middle or right plot of Fig. 2.7 showed that Sox2 or Nanog expression level was down-
regulated at similar rates when the ESCs were treated with RA for various durations.
Figure 2.8: Single-cell histograms of Oct4, Sox2 and Nanog transcripts in the ESCs for the RA removal experiment. Transcript counts of Oct4 (blue), Sox2 (red) and Nanog (green) in individual ESCs that make up the unperturbed population of cells ([RA] = 0 µM) or perturbed populations ([RA] = 0.5 µM). The far left histograms correspond to data obtained from the control sample (“wild type”). Histograms in the 2nd column correspond to data obtained from samples collected right after the RA treatment (“@ 0hr”). Histograms in the 3rd column correspond to data obtained from samples collected 2 days after RA removal (“@ 2day”). For example, “13hr RA @ 0hr” denotes samples treated with RA for 13hrs and then collected right at the end of the treatment.
Figure 2.8 (Continued): Single-cell histograms of Oct4, Sox2 and Nanog transcripts in the ESCs for the RA removal experiment. Transcript counts of Oct4 (blue), Sox2 (red) and Nanog (green) in individual ESCs that make up the unperturbed population of cells ( [RA] = 0 µM ) or perturbed populations ( [RA] = 0.5 µM ). The far left histograms correspond to data obtained from the control sample (“wild type”). Histograms in the 2nd column correspond to data obtained from samples collected right after the RA treatment (“@ 0hr”). Histograms in the 3rd column correspond to data obtained from samples collected 2 days after RA removal (“@ 2day”). For example, “13hr RA @ 0hr” denotes samples treated with RA for 13hrs and then collected right at the end of the treatment.
Figure 2.8 (Continued): Single-cell histograms of Oct4, Sox2 and Nanog transcripts in the ESCs for the RA removal experiment. Transcript counts of Oct4 (blue), Sox2 (red) and Nanog (green) in individual ESCs that make up the unperturbed population of cells ( [RA] = 0 µM ) or perturbed populations ( [RA] = 0.5 µM ). The far left histograms correspond to data obtained from the control sample (“wild type”). Histograms in the 2nd column correspond to data obtained from samples collected right after the RA treatment (“@ 0hr”). Histograms in the 3rd column correspond to data obtained from samples collected 2 days after RA removal (“@ 2day”). For example, “13hr RA @ 0hr” denotes samples treated with RA for 13hrs and then collected right at the end of the treatment.
Figure 2.9: Three-dimensional scatterplot of Oct4, Sox2 and Nanog transcripts in the ESCs that were first treated with 0.5 µM of RA for various durations followed by RA removal. x-axis represents number of Nanog transcripts. y-axis represents number of Oct4 transcripts. z-axis represents number of Sox2 transcripts. “13hr RA @ 0hr” denotes samples treated with RA for 13hrs and then collected right at the end of the treatment. “13hr RA @ 2day” denotes samples treated with RA for 13hrs and then collected 2days after RA was removed.
2.3 Discussion

The pluripotent ESC is a valuable *in vitro* model for studying the effects of various factors on cell lineage decisions in very early embryonic stages of mammalian development. In particular, the effects of RA signaling on the differentiation of ESCs and neural induction have been studied extensively. In addition to previous studies that have shown that RA promotes neural differentiation of ESCs, directly measuring the enormous heterogeneity in the levels of the core pluripotency circuit’s transcripts among single cells led us to reveal a restricted transcript space formed by the core circuit in the ESCs during RA-induced differentiation process. Our results demonstrated that relying solely on the bulk measurements to study the core pluripotency circuit could be misleading because the average gene expression levels do not capture the vast transcript space that individual cells within the population span. For example, in contrast with some existing literature that used bulk measurements, our study showed that the NOS mRNA expressions in a small subpopulation of the ESCs did not completely disappear even after 7-day RA treatment. Our results suggested that after adding a single chemical cue (RA in this case) the stem cells could develop from a homogeneous starting population to a mixture of different cell types. The highly sensitive and quantitative smFISH technique helped us to identify two mutually exclusive subpopulations coexist upon prolonged RA treatment.

Similar to normal stem cells, which can self-renew and give rise to differentiated cells, cancer stem cells can also self-renew and produce more differentiated tumor cell progeny in tumors (Gupta et al., 2009). According to this model, therapeutic approaches must focus on the selective elimination of cancer stem cells and/or induction of irreversible differentiation of these cells. Differentiation therapy is a treatment that induces differentiation of cancer cells, and potentially cancer stem cells, thereby preventing further proliferation (Kashyap et al., 2009). RA has been used in animals and humans to induce differentiation cancer cells in various types of cancers (Reynolds et al., 2000; Matthey et al., 2009; Guibal et al, 2009; Nasr et al., 2008; Nasir et al., 2009; Ross and Spengler, 2007). Therefore, the data obtained
from our study may shed light on general mechanisms underlying RA induced differentiation in stem cells and perhaps lead to new strategies for cancer treatment.

2.4 Potential pitfalls and caveats

First, the effect of cell size was not taken into account when we analyzed the smFISH results in both section 2.2.1 and 2.2.2. All the data presented were raw counts of mRNA transcripts of Oct4, Sox2 and Nanog. We noticed that on average the cells varied in size across different time points, therefore in the future studies a more quantitative way of presenting the gene expression data would be normalizing the mRNA counts over cell volume.

Second, the control samples chosen in both section 2.2.1 and 2.2.2 might not be ideal. As discussed earlier, a better approach would be to collect non-RA-treated cells at exactly the same time points as those RA-treated samples. The rationale is that even for the non-RA-treated cells they exhibit fluctuations in gene expressions at various time points. To achieve a more valid comparison, it is important to control for those fluctuations in the control samples.

Third, the effect of serum was not taken into consideration in our study. It has been shown that serum contains a large number of factors that might influence the differentiation process. For example, one study has provided evidence for the presence of RA-signaling activity in undifferentiated ESCs that come from the ATRA content of the serum (Simandi et al., 2010). Thus, further comparative studies are required to evaluate how low RA content of the serum might influence the ESC pluripotency and differentiation in comparison with serum-free growth medium.

Fourth, the effect of cell density was not evaluated thoroughly. We chose a different plating density for each time point with the hope that at the end of each time point all the culture dishes would have similar confluency of cells. However, some studies have suggested that RA affects neural
specification of ESCs in a cell-density dependent manner (Lu et al., 2009; Ying et al., 2003). Neural
induction of ESCs during monolayer differentiation has been found to be strongly inhibited by even
modest increase in cell density (Lu et al., 2009). Therefore, it is important to account for the effect of cell
density in our study. A way to get around the problem would be to include proper controls that have been
mentioned earlier in this section.

Finally, the effect of off feeders was not taken into account for the experiments in section 2.2.2. The transcript levels of Oct4, Sox2 and Nanog kept decreasing even after RA was removed and LIF was added back to the culture medium. Besides the RA-induced differentiation effect, the down-regulation of the three pluripotency factors might also be caused by the absence of the feeders. It has been shown that even in the LIF + serum containing medium the ESCs tend to differentiate if they are not grown on feeders. One way to get around this problem would be to culture the cells in 2i medium that does not require feeders to support the pluripotent state of the ESCs.

2.5 Methods

2.5.1 ESC culture medium

ESCs V6.5 (129SvJae and C57BL/6; male) were cultured in a standard ESC growth medium on
gelatinized tissue culture plates. In brief, knockout DMEM (Gibco) was supplemented with 15% FBS
(HyClone), 100 μM MEM non-essential amino acids (Gibco), 0.1 mM beta-Mercaptoethanol, 2 mM L-
glutamine (Gibco), 100 U/ml penicillin, 100μg/ml streptomycin (Gibco), and leukemia inhibitory factor.

2.5.2 ESC adherent monoculture for RA-induced differentiation assays

ESCs V6.5 were first cultured with irradiated mouse embryonic fibroblast (MEFs) on
gelatinized culture plates for 2-3days till reached ~ 70-80% confluency. The cells were then trypsinized,
and the mixture of both the ESCs and MEFs were resuspended in culture medium. The cell suspension
was incubated at 37 °C on non-gelatinized culture plates for 20mins. The supernatant, that contained mostly ESCs but very few MEFs, was collected and then spin down. The supernatant was removed and the cell pellet was resuspended in 10ml of culture medium. The cell number was counted using hemocytometer. The cells were then seeded onto gelatinized culture plates in the absence of MEFs and cultured in LIF containing medium described in 2.5.1 for 12hrs. By the end of 12hr, the medium was removed and the cells were gently washed with 1xPBS to get rid of any residual LIF. The cells were cultured in medium without LIF throughout the entire RA-induction assay. All-trans RA (Sigma-Aldrich) was added to the medium to reach a final concentration of 0.5 μM. The cells were then collected by the end of each time point for smFISH imaging. For RA-induced differentiation assay with RA removal part, the medium was removed by the end of RA treatment and the cells were gently washed with 1xPBS to get rid of any residual RA and then cultured in LIF containing medium. To prepare the cells for smFISH imaging, cells were trypsinized for 5 min, fixed for 10 min in 4% formaldehyde in 1× PBS, washed twice with 1× PBS and stored in 70% ethanol. We did not detect any mycoplasma by DAPI staining during imaging.

2.5.3 SmFISH probe design

SmFISH probes were designed using a custom algorithm [now publicly available at https://www.biosearchtech.com/stellarisdesigner/] to locate twenty oligonucleotide regions with 45 – 65% GC content in the cDNA of the open reading frame of a gene of interest. The restrictions on the GC content are to ensure that all probes have a relatively uniform binding energy. The 5’ and 3’ regions of the gene were included if the open reading frame search did not result in sufficient probe candidates. For optimum performance, probes were separated by 2 nucleotides to ensure approximate spacing for the fluorophore molecule. Selected probes were then subjected to BLAST (Basic Local Alignment Search Tool) analysis of the entire genome, and those with significant alternative targets were removed from the selection process. Forty-eight to 96 oligonucleotide probes with 3’ amine end modification for coupling to fluorophores were synthesized by Biosearch Technologies (Novato, CA). For coupling of probes to
fluorophores, 300 ng of Alexa594 (Invitrogen), TMR (Invitrogen) or Cy5 was combined with 48-96 nmol probes in 0.1 M sodium bicarbonate buffer and incubated at 37°C overnight. To each sample, 0.1X volume of 3M sodium acetate and 2.5X volume of 100% ethanol was added. The solution was vortexed thoroughly and incubated at -80°C overnight. The solution was centrifuged for 14000xg at 4°C for 20 minutes, and the supernatant was removed and discarded. Probes were then HPLC purified using triethylamine acetate (0.1 M) and acetonitrile as buffers. The final probe concentration was measured using Nanodrop.

2.5.4 Sample preparations for smFISH

We performed smFISH and counted the mRNAs in individual cells as described previously (Raj et al., 2008). Briefly, harvested cells were fixed with 3.7% formaldehyde for 10 minutes, washed twice with PBS, and permeabilized in 70% ethanol. For hybridization, the samples were resuspended in 100 μl of hybridization solution containing labeled DNA probes in 2xSSC, 1 mg/ml BSA, 10mM VRC, 0.5 mg/ml Escherichia coli tRNA and 0.1 g/ml dextran sulfate, with 10 to 25% formamide, which varies for different probes, and incubated overnight at 30°C. All hybridizations were performed in solution using probes coupled to either tetramethylrhodamine (TMR) (Invitrogen), Alexa 594 (Invitrogen) or Cy5 (GE Amersham). We used TMR for the probes against Nanog mRNA, Alexa 594 for Sox2 mRNA, and Cy5 for Oct4 mRNA. Optimal probe concentrations during hybridization were determined empirically. The next day, the samples were washed twice by incubating in 1 ml of wash solution consisting of 10 to 25% formamide and 2xSSC for 30 minutes.

2.5.5 Image acquisition

The samples were resuspended in glucose oxidase anti-fade solution, which contains 10 mM Tris (pH 7.5), 2xSSC, 0.4% glucose, supplemented with glucose oxidase and catalase. Then 8 μl of cell suspension were sandwiched between two coverglasses, and mounted on a glass slides using a silicone gasket. Images were taken with a Nikon TE2000 inverted fluorescence microscope equipped with a 100x
oil-immersion objective and a Princeton Instruments camera using MetaMorph software. Stacks of images were taken automatically with 0.4 microns between the z-slices.

2.5.6 Image analysis

To segment the cells, a marker-guided watershed algorithm was used. Briefly, cell boundaries were obtained by running an edge detection algorithm on the bright-field image of the cells. To generate markers, the centroid of the region enclosed by individual cell boundaries is computed. A marker-guided watershed algorithm was then run on the distance transformation of the cell boundaries, using the markers located within the cell boundaries. The resultant cell segmentation image was then manually curated for occasional mis-segmentations. To quantify the number of RNA molecules in each cell, a log filter was run over each optical slice of an image stack to enhance signals. A threshold was taken on the resultant image stack to pick up mRNA spots. The locations of mRNA spots were then taken to be the regional maximum pixel value of each connected region. The number of mRNA spots located within the cell boundaries of an individual cell was thus quantified.
2.6 References


Chapter 3 Dynamics on a transcript pathway formed by core pluripotency circuit in ESCs

3.1 Introduction

A large-scale network built from hundreds of interconnected genes helps control pluripotency in ESCs (Huang et al., 2005; Wang et al., 2006; Chen et al., 2008; Jaenisch and Young, 2008; Kim et al., 2008; Liu et al., 2008; Orkin et al., 2008; Chambers and Tomlinson, 2009). As a quantitative framework for understanding how such complex networks respond to perturbations, researchers have proposed mathematical models in which perturbing the network yields transcriptional dynamics that are limited to just a few temporal profiles among vast possibilities (Thomas, 1998; Chickarmane et al., 2006; Chickarmane et al., 2008; Ma’ayan 2008; Jothi et al., 2009). A key idea underlying such models, based on a theoretical concept proposed by Conrad Waddington (1957), is that a stem cell could be thought of as moving through a multi-dimensional space whose “spatial” coordinate is specified by the expression level of each gene. In this space, perturbations to the network would push the cells along a limited set of “canals” over time and cells would not be allowed to occupy positions beyond specified regions in this multi-dimensional space. Experimentally verifying that such restricted “transcript space” exists for ESCs requires measuring the key transcript levels in single cells because it is precisely the variations in transcript levels among single cells within an isogenic population that provides the information on the shape and “ruggedness” of such a restricted space. This has so far been a challenging task to accomplish for transcripts in single stem cells. Little is known about the shape of the restricted transcriptional space. We sought to experimentally reveal the existence, shape, and dynamics of restricted transcript space that is formed by very general perturbations to the core pluripotency circuit in ESCs.

The core pluripotency circuit consists of the three master transcription factors, Nanog, Oct4, and Sox2 with nine regulatory links among them (Boyer et al., 2005; Loh et al., 2006; Chambers et al., 2007; Loh et al., 2011) (Fig. 3.3a). Each regulatory link (arrows in Fig. 3.3a) represents the combination of direct and indirect interactions between the two genes it connects. For example, an arrow from Nanog to
Oct4 represents Nanog directly affecting the transcription of Oct4 (by binding to the promoter of Oct4), as well as Nanog interacting with hundreds of factors outside the circuit whose downstream actions affect the transcription of Oct4 (indirect interaction). When this network is perturbed, it is thought that the strengths of these interaction links change over time and rewiring of the circuit occurs in complicated ways that are not yet well understood. The enormous number of interactions that each regulatory link actually represents has hindered understanding the transcriptional dynamics of the core circuit that arise from very general perturbations (Fig. 3.3a). In our study, we experimentally revealed a restricted transcript space formed by the core circuit. Our work may inform future studies of mapping complex transcriptional networks in stem cells and other organisms to an intuitive landscape picture that captures the essence underlying complexity of the networks.

3.2 Results

3.2.1 Positive correlation between every pair of the genes in the unperturbed NOS network

As our starting point, we applied smFISH to measure the integer counts of Nanog, Oct4 and Sox2 mRNAs in single ESCs whose core circuit was unperturbed. We attached fluorophores of three different colors to label each of the single-stranded DNA probes that bind Nanog, Sox2, and Oct4 mRNA (Fig. 3.1). As a result the three different transcripts simultaneously became visible as diffraction limited spots within single cells (Fig. 3.3b-e and Fig. 3.2). Hence each cell within an isogenic population is now assigned a position in a three-dimensional “transcript space” whose coordinates are defined by the integer counts of Nanog, Oct4, and Sox2 mRNA within that cell (Fig. 3.3f). In this transcript space, we found a positive correlation between every pair of the genes in the core circuit (Fig. 3.3g-i). In particular, Oct4 and Sox2 transcript levels had the strongest positive correlation (Fig. 3.3g), Nanog and Sox2 transcript levels exhibited a weaker positive correlation (Fig. 3.3h). Nanog and Oct4 transcript levels had the weakest positive correlation among all three pairs (Fig. 3.3i). The most common interpretation of these
positive correlations is that if any one of the three genes is over-expressed, then the transcript levels of the other two genes would simply increase. This would mean that the single-cell “cloud” in the transcript space would shift up along the grey positive-sloped linear regression lines shown in Figs. 3.3g-i. If the cells can shift up along the regression lines of Fig. 3.3g-i without an upper bound upon overexpression of each of the three factors, then models of stem cell landscapes would describe the core circuit as having no restriction on transcript space. We also noticed that some regions of the transcript space were more densely populated than others. In particular about 65% of the cells populated regions marked by low levels of Nanog (i.e., less than the population average of 57 counts of Nanog mRNAs).
Figure 3.1: Schematics of smFISH and using irradiated MEFs as a negative control to test binding specificity of Nanog, Oct4, and Sox2 RNA FISH probes. 

a, Schematic of RNA FISH probes. Each probe is a single-stranded DNA that is about 20 bases long and complimentarily binds to a portion of a single mRNA molecule. 48 such probes, each labeled with a single fluorophore, bind in tandem along a single mRNA molecule. Three different colors of fluorophores – TMR, Alexa594, Cy5 – that do not spectrally overlap, are used to detect Nanog, Sox2, and Oct4 within the same cell. 

b-d, Example images of smFISH in individual mouse embryonic stem cells. Each bright diffraction-limited spot is a single Nanog mRNA (b), Sox2 mRNA (c), and Oct4 mRNA (d) molecule. Integer counts of all three transcripts are simultaneously measured in the same single cell. 

e-h, Negative control for probe specificity. We used the smFISH probes for Nanog, Sox2 and Oct4 in irradiated MEFs which do not express any of the three pluripotency genes. In MEFs, no spots were detected (e), as quantified by single-cell counts (Mean±S.E.M; 106 cells) for transcripts of Nanog (0.32±0.07) (f), Sox2 (0.27±0.07) (g), and Oct4 (0.25±0.06) (h).
Figure 3.2: Single-cell histograms of Nanog, Sox2, and Oct4 transcripts in unperturbed population of cells. Transcript counts of Nanog (a), Sox2 (b), and Oct4 (c) in individual mouse embryonic cells that make up the unperturbed population of cells (also see Fig. 3.1). The single-cell averages of the three transcripts are (Mean±S.E.M; 775 cells) 56.6±2.2 for Nanog, 114.4±2.3 for Sox2, and 224.5±3.0 for Oct4. Correlations among the three transcripts are shown in Fig.3.1.
Figure 3.3: Transcript space explored by the unperturbed core pluripotency circuit. 

**a.** Schematic of the core pluripotency circuit formed by Nanog, Oct4 and Sox2 with regulatory links among them. 

**b.** Example image of DAPI staining of nucleus in individual ESCs. 

**c-e.** Example images of smFISH in individual ESCs. 

**f.** Three-dimensional transcript space whose coordinates are defined by the integer counts of Nanog, Oct4 and Sox2. 

**g-i.** Scatterplots between every pair of the genes in the NOS circuit.
Figure 3.4: Schematic of the three cell lines used for ectopically perturbing transcript levels of Nanog, Sox2, and Oct4 individually. In each cell line, the endogenous copies of the gene is kept intact, thereby leaving all endogenous regulations that exist within the core circuit intact. An extra copy of the gene, under the control of the doxycycline-inducible promoter $P_{TET}$, is stably integrated in the Col1a locus in each cell line. rtTA is used for doxycycline-induction.

Figure 3.5: Inducibility of $P_{TET}$ inferred from GFP mRNA counts as a function of time with addition of 2 µg/ml of doxycycline to the KH2-GFP cell line. smFISH was used to measure single-cell average GFP mRNA counts in KH2-GFP cells as a function of time.
3.2.2 NOS circuit’s transcript space may have upper bounds

To measure how the heterogeneous transcript space spanned by the cells (Fig. 3.3f) morphs when the core circuit is perturbed, we used three stable cell lines, each of which had an extra copy of one of the three pluripotency genes controlled by the inducible promoter $P_{TET01}$, and stably integrated at the ColA1 locus (Fig. 3.4 and 3.5). These cell lines all contained two endogenous copies of *Nanog*, *Oct4* and *Sox2* genes each as well. By growing these three cell lines in a wide range of doxycycline concentrations and time durations, we measured temporal changes in the total counts of the three transcripts in single cells (Fig. 3.6, Fig. 3.7a-c). Under these perturbations, the population average of transcript levels followed highly complex temporal dynamics under our perturbations (Fig. 3.6, Fig. 3.7a-c; 0 is the population average counts for unperturbed cells). But despite their complexity, we found a common trend in the dynamics. Namely, *Oct4* and *Sox2* transcript levels increased only transiently, and was eventually pushed down to much lower levels compared to the unperturbed population’s expression level (Fig. 3.6, last column in Figs. 3.7a and 3.7b) for any perturbation. *Nanog* transcript level also increased only transiently, and eventually was pushed down to an average expression level in the vicinity of the mean transcript level of the unperturbed population, This suggested to us that the core circuit’s transcript space may have upper bounds that restrict how high the transcript levels can overshoot beyond those of the unperturbed cells.
Figure 3.6: Plots of population average for Oct4, Sox2 and Nanog in ESCs that were treated with various concentrations of doxycycline for different durations. Each data point represents the population mean (total mRNA transcript counts) at a particular condition (doxycycline concentration, time point). Plots in the left panel correspond to data from the Sox2-inducible cell line. Plots in the middle panel correspond to data from the Oct4-inducible cell line. Plots in the right panel correspond to data from the Nanog-inducible cell line.
Figure 3.7: Perturbing the core circuit produces highly complex transcriptional dynamics. Dynamics of population average transcript levels show complex non-linear behaviors.
Figure 3.8: Collapse of scaled Oct4-Sox2 transcript counts as a function of Nanog transcript counts. Two-dimensional transcriptional trajectory. Each data point represents the population mean (total mRNA transcript counts measured by smFISH) at a particular condition compared to the population mean in unperturbed state. a, All the population average data plotted together: allowed positions in Nanog-Oct4 plane (colored by cell-line). b, All the population average data plotted together: allowed positions in Nanog-Sox2 plane (colored by cell-line). c, All the population average data plotted together: allowed positions in Sox2-Oct4 plane. The red dash line is the best-fitted line. d, Two-dimensional transcriptional trajectory of Oct4-Sox2 as a function of Nanog.
Figure 3.9: Collapse of scaled Oct4-Sox2 transcript counts as a function of Nanog transcript counts. Two-dimensional transcriptional trajectory with the ‘average transcript velocity’ along the pathway. Each data point represents the population mean (total mRNA transcript counts measured by smFISH) at a particular condition compared to the population mean in unperturbed state. Each arrow represents the full velocity vector along each point on the trajectory. The pink line represents the best-fitted trajectory based on the given velocity vectors.
Figure 3.10: Restricted transcript space explored by the core pluripotency circuit – Both the population average expression levels and single-cell expression levels are confined within the 95-percentile boundaries. The unperturbed cells span a near maximal area (bounded by 95-percentile boundaries for the cells) in the transcript plane. Any perturbation eventually leads to a contraction of the 95-percentile boundaries and thus the area explored by the single cells within the population. Subsampling equal number of cells from each population and computing statistical quantities for 1000 iterations. Reported values are averages with standard errors. 95-percentile boundaries that show enclosure of cells in each population were also computed by subsampling equal number of cells from each population.
3.2.3 The population average transcript levels were constrained to the expression space spanned by the unperturbed cells

To see where in the transcript space the “upper bounds” (discussed in section 3.2.2) might be imposed, we first plotted together the population average transcript counts of Oct4 and Sox2 from all the perturbations shown in Fig. 3.6 and Fig. 3.7 (all cell lines, doxycycline level, temporal duration of perturbation) onto one graph (Fig. 3.10a). Here the coordinate (0,0) corresponds to the unperturbed population’s average counts of Sox2 and Oct4. Positive (negative) transcript counts in Fig. 3.10a represent perturbations that increase (decrease) the transcript counts above (below) the unperturbed state. This analysis established a strong positive correlation between Oct4 and Sox2 transcript levels (Fig. 3.10a), just as we had observed a positive correlation in the single-cell positions within the Oct4-Sox2 plane in the unperturbed state (Fig. 3.3g). Importantly, this analysis shows that despite the positive correlation, all Oct4 and Sox2 population averages were constrained well within the boundaries of the Oct4-Sox2 transcript plane that the unperturbed cells occupy, below the 75-percentile boundaries of Oct4 and Sox2 (grey lines in Fig. 3.10a).

Given the positive correlation between Oct4 and Sox2, we can analyze the relationship between Oct4 and Nanog levels, and infer the relationship between Sox2 and Nanog (or vice versa). Equivalently, we used the linear regression fit line in the Oct4-Sox2 transcript plane (red dashed line in Fig. 3.10a) as a new axis to project onto it all the data points from our perturbations (Fig. 3.8) onto this new axis. This way we obtained the population averages of Nanog transcript counts from all perturbations as a function of this new coordinate axis (See supplementary information for details of calculations). We call the coordinates on this new axis” Scaled Sox2-Oct4 transcript counts” (Fig. 3.8, 3.9 and 3.10). This reduces the three-dimensional transcript space into a two-dimensional “transcript plane”. We also plotted the positions of all the cells within the unperturbed population shown in the original three-dimensional transcript space (Fig. 3.3f) onto this transcript plane (grey data points in Fig. 3.10b). Here the
unperturbed population average is at the coordinate (0,0). Surprisingly, this transformation collapsed the complex transcriptional dynamics of the population averages obtained from all the perturbations in Fig. 3.6 and 3.7 onto one single master curve (Fig. 3.8d, 3.9 and 3.10b; see supplementary information for details of analysis). This curve shows that the population average moves along a very restricted path that lies within the heterogeneous expression space that the unperturbed cells span. Importantly, despite the positive correlation between the transcript levels of Nanog and of scaled Oct4-Sox2 among the single cells in the unperturbed population (shown as dashed grey line in Fig. 3.10b), we do not observe such a positive correlation in the population averages. Namely, the population average of scaled Oct4-Sox2 transcript level eventually decreases upon overexpression of any of the three factors in the core circuit. These analyses demonstrate that the population average transcript levels alone cannot capture the core circuit’s transcriptional dynamics.

3.2.4 All perturbations eventually led to the contraction of explored areas to a nearly common region within the transcript plane

A fuller understanding of the core circuit’s transcriptional dynamics may arise from following how the heterogeneous transcript levels among individual cells within a population changes due to perturbations. Given our revelation of the constrained population average transcript levels, the positions that individual cells within a population can occupy within the transcript plane may be restricted as well. To see this, we plotted the position of every single cell within the unperturbed (day 0) population (yellow points in Fig. 3.10c) as well as the position of every single cell from all perturbed populations (blue points in Fig. 3.10c) onto one graph (Fig. 3.10c). This showed that the transcript space explored by the perturbed cells do not differ much from the space explored by the unperturbed cells. In fact, in the reduced transcript-plane, the 95-percentile (top 5-percentile) transcript levels of Nanog and reduced Oct4-Sox2 of the unperturbed cells (brown dashed borders in Fig. 3.10c) and (all) perturbed cells (blue dashed borders in Fig. 3.10c) are very close to each other.
To quantify how these boundaries dynamically change due to perturbations, we analyzed the expression level in every single cell for each of the perturbations. This enabled us to quantify the variation in heterogeneity in the expression levels of the core circuit due to each perturbation. In particular, for each perturbed population, we computed the 95-percentile (top 5-percentile) transcript levels of Nanog and the scaled Oct4-Sox2 that the cells in the population had. This is a measure of the area spanned by the single cells within the population. To compute these 95-percentile boundaries for each perturbed and unperturbed population, we sub-sampled an equal number of cells from each perturbation (to normalize the different number of cells analyzed per perturbation) and repeatedly computed the area by bootstrapping this procedure (See details of calculation supplementary information). Doing so, we found that the unperturbed population of cells spanned nearly maximal volume (proportional to the area bounded by the 95-percentile boundary lines in the reduce transcript plane) in the three-dimensional transcript space (Fig. 3.10d). Perturbations could cause the 95-percentile boundaries (and thus the area explored by single-cells within a population) to transiently increase in the transcript plane but they all retracted to a region smaller than the area that is spanned by the unperturbed population. This analysis shows that the heterogeneity in expression levels of the core circuit, but not their mean levels, show the most drastic changes under our general perturbations. The volume-contraction we found, along with the constraint on the mean-trajectory, suggest the existence of restricted transcript space, and that the unperturbed population maximally spans this space. The constrained trajectory of the population average expression levels, along with the contraction of area spanned by the single cells within the population upon overexpression of any of the three factors, support mathematical models with a restricted transcription space for the core pluripotency circuit.
3.3 Discussion

Directly measuring the enormous heterogeneity in the levels of the core pluripotency circuit’s transcripts among single cells led us to reveal a restricted transcript space formed by the core circuit in embryonic stem cells. We have also shown that relying on only the mean gene expression levels to study the core pluripotency circuit is misleading because the average expression levels do not capture the vast transcript space that individual cells within the population span. Conceptual and mathematical models of stem cells in which gene expressions are given a geometric interpretation of abstract landscapes, particularly of transcript landscapes, have been difficult to validate and put to a practical use due to challenges in simultaneously measuring several key transcripts in single cells. In this study, we managed to experimentally reveal a restricted transcript space formed by the core circuit and our work may shed light on the underlying complexity of the NOS network.

Perturbing the core circuit over a wide temporal range and strengths tuned the range of heterogeneity in transcript levels. This manifested as contractions in volume occupied by a population of cells in a measurable transcript space. Our results suggested that unperturbed core circuit spanned near maximal volume in transcript space. The “exploration area” spanned by the unperturbed cells was such that it detected overexpression perturbation to each node of the circuit. The circuit initially responded by transiently increasing the exploration area of the cells. The circuit then negatively regulated the transcript levels of NOS and brought down the exploration area to a much lower than that of the unperturbed area. This smaller, but non-zero area (Nanog, Sox, Oct had not completely shut off) could represent a “primed” state for differentiation for perturbation of all three transcripts but not necessarily the same primed state. The unperturbed cells therefore occupied a state that could detect overexpression of the NOS genes and responded by turning down the expression levels to a lower, but non-zero levels. This may lead to a faster shutting down of the factors when ESCs receive a bona-fide signal for differentiation (e.g., retinoic acid) but not fully commit to differentiation.
3.4 Potential pitfalls and caveats

First, the effect of off-feeders was not taken into consideration in this study. Besides the doxycycline-induced perturbation effect, the down-regulation of the pluripotency factors might also be caused by the absence of MEFs. As discussed in Chapter 2, it has been shown that even in the LIF containing medium the ESCs tend to differentiate if they are not supported by the feeders. And this effect will become more pronounced if the ESCs are grown off-feeders for a long period of time for example in our case 6 days. A way to get around this problem would be to culture the ESCs on MEFs in serum containing medium or simply in 2i medium that does not require the presence of the feeders.

Second, the control sample in our study might not be ideal. Similar to what has been discussed in Chapter 2, a better approach would be to collect unperturbed cells at exactly the same time point as those perturbed samples. Again, the rationale is that even for the unperturbed cells their gene expression profiles tend to fluctuate across various time points. To achieve a valid comparison between the perturbed and unperturbed cells, it is important to monitor those fluctuations in the control samples.

Third, our study only provided information about the transcript pathway. It would be more informative if we could examine the expression of both mRNA and protein in single cells. That would help us gain a deeper understanding of the architecture of the NOS circuit in ESCs.

Finally, we were unable to distinguish between the exo- and endo-genous gene expressions for Oct4, Sox2 and Nanog. All the smFISH data presented were total mRNA expression. It is critical to tease apart the transgene and endogenous gene expressions to fully comprehend the underlying mechanisms of NOS circuit regulation upon perturbations.

Based on the existing experimental system described in this chapter, we made some modifications and improvements to address the four issues mentioned above. In Chapter 4, the ESCs were grown on feeders in serum containing medium or in 2i medium for a shorter period of time (up to 24hr). Moreover, we managed to perform smFISH and immunofluorescence (IF) together in single cells to study
the expression profiles of mRNA and protein simultaneously. Western blot was also conducted to complement the IF results. In Chapter 5, we developed a highly sensitive and quantitative way to measure exo- and endo-genous gene expressions separately.

3.5 Methods

3.5.1 ESC culture medium (same as in 2.5.1)

3.5.2 ESC adherent monoculture for doxycycline-induced perturbation assays

V6.5 (129SvJae and C57BL/6; male) and KH2 ESCs (Beard et al., 2006) were used to establish wildtype smFISH measurements. All the ESC lines with inducible Nanog, Oct4 and Sox2 expression levels were a kind gift from the R. Jaenisch lab and were constructed as previously described (Hochedlinger et al., 2005; Beard et al., 2006). The cells were first cultured with irradiated MEFs on gelatinized culture plates for 2-3days till reached ~ 70-80% confluency. The cells were then trypsinized, and the mixture of both the ESCs and MEFs were resuspended in culture medium. The cell suspension was incubated at 37 °C on non-gelatinized culture plates for 20mins. The supernatant, that contained mostly ESCs but very few MEFs, was collected and then spin down. The supernatant was removed and the cell pellet was resuspended in 10ml of culture medium. The cell number was counted using hemocytometer. The cells were then seeded onto gelatinized culture plates in the absence of MEFs and cultured in LIF containing medium described in 2.2.1 for 12hrs. By the end of 12hr, the medium was removed and fresh medium with appropriate level of doxycycline (Invitrogen) was added to the plates. The cells were then collected by the end of each time point for smFISH imaging.

3.5.3 Sample preparations for smFISH (same as in 2.5.4)

3.5.4 Image acquisition and analysis (same as in 2.5.5 and 2.5.6)
3.5.5 Calculations and Analyses

Analysis for revealing unified transcript trajectory

To obtain the transcript trajectory (path) in Fig. 3.10b, we first computed the average single-cell Nanog, Sox2, and Oct4 transcript counts in all the perturbations shown in Fig. 3.7. As a result, each perturbation condition (i.e., cell line, doxycycline-induction duration, [doxycycline]) is now endowed with a position-coordinate in the three-dimensional transcript space:

\[(x, y, z) = \text{(single-cell average Sox2, single-cell average Oct4, single-cell average Nanog)}.\]

First, we plot just \((x,y)\) for all the perturbation conditions together onto one plot (Fig. 3.10a). From this plot, we see a strong positive correlation between the population averages of Sox2 (i.e., \(x\)) and Oct4 (i.e., \(y\)). The linear regression fit of this plot (Fig. 3.10a) yields the following equation for straight-line (red line in Fig. 3.10a):

\[\hat{y} = ax + b\]  \[1\]

where \(\hat{x}\) and \(\hat{y}\) are the Sox2 and Oct4 coordinates of the linear regression fit line. Here, \(a = 1.15\pm0.24\) and \(b = -45.51\pm11.56\) are the slope and the \(\hat{y}\)-intercept of the linear regression line. Next, we projected each perturbation condition’s \((x,y)\)-coordinate (population-level average of Nanog transcript counts) onto this fit line by computing the line that is perpendicular to the fit line \([1]\) and joins the perturbation’s \((x,y)\)-point. This line joins the point \((\tilde{x}, \tilde{y})\) on the fit line to the data point \((x,y)\), where:

\[(\tilde{x}, \tilde{y}) = (a(y - b) + x, a\hat{x} + b)\]  \[2\]

Next, we compute the distance of the projected data point \((\tilde{x}, \tilde{y})\) from the unperturbed population’s projected point onto the axis. The displacement \(d\) relative to the unperturbed population’s projected point is...
\[ d = \text{sign}(\tilde{x}) \sqrt{\tilde{x}^2 + (\tilde{y} - b)^2} - c \]  

where \( c = 34.34 \) is the unperturbed population’s coordinate. Figure 3.10b is obtained by plotting the displacement \( d \) against the average nanog transcript count \( z \). Savitzky-Golay smoothing filter with a sliding window along the nanog-axis (window size = 9 transcripts) yielded the smoothing fit line (red line) that describing the trajectory in fig. 3.10b.

**Calculation of velocity vectors along the transcript trajectory**

From our dynamic measurements summarized in Figure 3.7, we can compute the ‘average transcript velocity’ along the transcript trajectory. To do so, we need to compute both the Nanog-component velocity and the scaled Oct4-Sox2 component velocity. Together, these two vectors give the full velocity vector along each point on the trajectory. The velocity in the 3-dimensional transcript space has three components:

\[ V_{\text{Sox2}}(t) = \frac{x(t + \Delta t) - x(t)}{\Delta t} \]

\[ V_{\text{Oct4}}(t) = \frac{y(t + \Delta t) - y(t)}{\Delta t} \]  \[ V_{\text{Nanog}}(t) = \frac{z(t + \Delta t) - z(t)}{\Delta t} \]

where \( t \) is the time at the current position on the trajectory, \( \Delta t \) is the time duration between the two (current and the next) measurements along the given perturbation (i.e., for a given [doxycyline] and cell line). \((x(t), y(t), z(t))\) is the position in the three-dimensional transcript space, namely the population average transcript count at time \( t \) in the perturbation (recall that \( x, y, \) and \( z \) are the population average of Sox2, Oct4, and Nanog transcript counts respectively). The ‘scaled Oct4-Sox2’-component of the velocity
(moving along the regression fit line in Oct4-Sox2 plane; Fig. 3.10a) is the temporal rate of change of the displacement (Equation [3]) along the regression line that we computed above. It is

\[ \mathbf{v}_{\text{scaled oct4 sox2}} = \frac{d \mathbf{v}_{\text{oct4}} + \mathbf{v}_{\text{sox2}}}{1 + a^2} \]  

[5]

Hence the velocity vector, at each data point shown in Fig. 3.10c, is given by

\[ \mathbf{v}_{\text{trajectory}} = (\mathbf{v}_{\text{nanog}}, \mathbf{v}_{\text{scaled oct4-sox2}}) \]  

[6]

At each data point on the transcript trajectory in figure 3.10c, this velocity vector is calculated and plotted.

3.6 References


Chapter 4 Gene regulation of the core pluripotency circuit in embryonic stem cells

4.1 Introduction

ESCs maintain their pluripotent state by expressing a battery of transcription factors (Huang et al., 2005; Wang et al., 2006; Chen et al., 2008; Jaenisch and Young, 2008; Kim et al., 2008; Liu et al., 2008; Orkin et al., 2008; Chambers and Tomlinson, 2009). The core regulatory circuit formed by Nanog, Oct4 and Sox2 (NOS) has been identified to be essential to the ESCs identity (Boyer et al., 2005; Loh et al., 2006; Chambers et al., 2007; Loh et al., 2011). In this circuit, all three factors regulate themselves, as well as each other. The frequency with which these three factors co-occupy within the same gene region is very high. It is known that Oct4 and Sox2 form heterodimers which regulate a wide range of ESCs-specific genes, including Oct4 and Sox2 themselves as well as Nanog (Chew et al., 2005). Among Oct4 bound genes, half of them are also bound by Sox2. More than 90% of the promoter region bound by Oct4 and Sox2 are also bound by Nanog (Chen et al., 2008; Jaenisch and Young, 2008; Thomson et al., 2011; Navarro et al., 2012). Besides forming the regulatory circuit, the three core factors contribute to the hallmark characteristics of ESCs by activation of target genes that are involved in maintaining pluripotency and self-renewal of the ESCs, and repression of signaling pathways that promote differentiation. Although the importance of the three factors for the regulation of pluripotency and differentiation is well established, their mutual interaction network and the resulting regulatory dynamics are still not well understood. Here, we examined the architecture of the NOS network using inducible gain- and loss-of-function approaches. By quantitatively measuring mRNA transcripts and proteins in individual ESCs, we identified a few unique features of the pluripotency network in ESCs.

In Chapter 3, we directly measured the enormous heterogeneity in the levels of the NOS circuit’s transcripts among single cells and managed to reveal a restricted transcript space formed by the core circuit in ESCs. However, as discussed in Section 3.4 there were a few pitfalls in the experimental design. In this chapter, we improved our experimental system by 1) growing the ESCs on feeders in
serum containing medium or in 2i medium without feeders for shorter period of time (up to 24hrs); 2) including control samples at exactly the same time points as those perturbed samples; 3) measuring protein levels of Oct4, Sox2 and Nanog via both immunofluorescence (IF) and western blot approaches; 4) performing smFISH and IF simultaneously in the same single cell.

4.2 Results

4.2.1 mRNA expression levels of Sox2, Oct4 and Nanog were doxycycline dose-dependent

For the pilot experiment, the ESCs were treated with various concentrations of doxycycline (0, 0.2, 0.5 or 2µg/ml) for 12 or 24 hours. Fig. 4.1 suggested that the gene responses of Sox2, Oct4 and Nanog in ESCs upon perturbation were doxycycline dose-dependent. More specifically, the extent of the ectopic induction of a factor was directly proportional to the amount of doxycycline added to the cells. The expression of the other two factors besides the one that was ectopically induced demonstrated similar patterns across various doxycycline conditions and in a doxycycline dose-dependent manner. In future experiments, we could simplify the system by including fewer doxycycline concentration conditions since all the three factors responded to doxycycline-induced perturbations in a dose-dependent way.

4.2.2 The regulatory patterns of Sox2 and Oct4 were symmetric

One interesting observation was that Sox2 and Oct4 behaved in a symmetric way in response to perturbations. In the Sox2-inducible cells, the expression level of Sox2 went up first and then gradually returned to almost baseline. Oct4 was down-regulated monotonically when Sox2 was ectopically induced (Fig. 4.1a, Fig. 4.2b, c and d). Similar patterns were observed in the Oct4-inducible cells: Oct4 expression increased initially and followed by a decrease by 24hr, expression of Sox2 decreased monotonically (Fig. 4.2a, Fig. 4.2b, c and d).
4.1b, Fig. 4.2 a, c and d). The “mirroring behavior” of Sox2 and Oct4 might be explained by the fact that they physically form heterodimers. This was further supported by our single-cell data that Sox2 and Oct4 transcript levels had a strong positive correlation in the unperturbed NOS network (Fig. 3.3g). However, on the other hand our data demonstrated that the population expression levels of Sox2 and Oct4 were anti-correlated. When Sox2 were over-expressed, Oct4 decreased, and vice versa (Fig. 4.1 and 4.2). This suggested that Sox2-Oct4 positive interaction might become uncoupled when one of them was perturbed. We will revisit this issue in greater details later this chapter.
Figure 4.1: Doxycycline dose-dependent mRNA expression profiles of Sox2, Oct4 and Nanog in ESCs grown in serum containing medium. Each data point represents the population average (total mRNA counts normalized by cell volume) measured by smFISH at a particular time point and doxycycline concentration. Error bars are 95% confidence intervals of the mean of 1000 bootstrap samples. a, Expression profiles of Sox2-inducible cells. b, Expression profiles of Oct4-inducible cells. c, Expression profiles of Nanog-inducible cells.
Figure 4.2: Two-dimensional transcriptional trajectory of the NOS circuit. Each data point represents the population mean (total mRNA transcript counts measured by smFISH) at a particular condition. Error bars are 95% confidence intervals of the mean of 1000 bootstrap samples. To read the graphs: Follow dashed line (1st point = 0hr, 2nd point = 12hr, 3rd point = 24hr) starting at the (= day 0, [dox]=0) star-shaped sign. **a-c**, Population average data from the three inducible cell lines (Nanog, Sox2 and Oct4) combined. Sox2 versus Nanog (**a**), Oct4 versus Nanog (**b**), and Oct4 versus Sox2 (**c**). **d**, Population average data from Sox2- and Oct4-inducible cell lines.
Figure 4.3: smFISH measurement of population average of Sox2, Oc4 and Nanog for the wild type KH2 cells. Each data point represents the population average (total mRNA transcript counts) at a particular time point. Black dotted lines correspond to data obtained from KH2 cells in the absence of doxycycline. Colored lines correspond to data obtained from KH2 cells treated with 2 µg/ml of doxycycline.
Figure 4.4: Single-cell histograms of Sox2, Oct4 and Nanog transcripts in wild type KH2 cells at t = 12hr. a-c., Transcript counts of KH2 cells in the absence of doxycycline, the single-cell averages of the three transcripts are (Mean±S.E.M; 607 cells) 115±2.1 for Sox2, 224±2.6 for Oct4 and 59.7±2.4 for Nanog. d-f., Transcript counts of KH2 cells treated with 2 µg/ml of doxycycline, the single-cell averages of the three transcripts are (Mean±S.E.M; 580 cells) 114±2.3 for Sox2, 224±2.4 for Oct4 and 55.7±2.2 for Nanog.
Results from our pilot experiment (Fig. 4.1 and 4.2) demonstrated that the gene responses of the NOS circuit upon perturbations were doxycycline dose-dependent. To simplify the setups and to save resources, in the subsequent experiments the ESCs were only treated with 2 µg/ml of doxycycline. Before conducting the downstream studies, one important control experiment required was to test whether the addition of doxycycline would have any abnormal effect on gene expressions besides turning on the doxycycline-inducible promoter $P_{TET}$. The wild type KH2 cells, used to construct the three stable cell lines in our study, were cultured in medium with the presence of 2 µg/ml of doxycycline for 12 or 24hrs. Fig. 4.3 and 4.4 showed that the presence of 2 µg/ml of doxycycline did not significantly affect the gene expression profiles of Sox2, Oct4 and Nanog. Moreover, our pilot experiment suggested that rewiring of the NOS network upon perturbations could happen earlier than 12 hours. Thus, we decided to include earlier time points (2, 4 and 8hrs) to further fine-tune the perturbation experiments in order to capture greater details of the gene regulatory dynamics of the core pluripotency circuit in ESCs.

It has been suggested that serum contains a large number of factors that might influence the gene regulatory activities in the ESCs (Ying, et al., 2008; Marks et al., 2012). For example, some studies have showed that Nanog transcripts are more highly expressed and less heterogeneous in 2i medium compared with in serum containing medium (Hamazaki et al., 2006; Niwa et al., 2009; Theunissen et al., 2011). It is important to show that whatever gene responses upon perturbations observed in the serum containing medium are not an artifact of serum. Therefore, we conducted comparative studies of the perturbation experiments in both serum containing and 2i media. Our results demonstrated that the gene expression profiles of the NOS circuit in both culture media were consistent (Fig. 4.5, Fig 4.7, Fig. 4.15 and Fig. 4.17).
4.2.3 NOS circuit seemed to have an upper bound to keep the gene expression below a certain level

In both serum containing and 2i media, the NOS network exhibited an upper bound for the gene expression (Fig. 4.5 and Fig. 4.7). The three inducible cell lines were treated with 2 µg/ml of doxycycline that was supposed to be potent enough for near-maximal induction of the target genes. Surprisingly, Oct4 mRNA was over-expressed by only 10-20% compared with its expression level in the control samples (Fig. 4.5b and Fig. 4.7b). The induction effect seemed to be very subtle given the amount of doxycycline was added to the system. It had been speculated that the inducible promoter used in our study was not working properly or was silenced after a few hours. However, one of the control experiments we did earlier on showed that the \( P_{TET} \) was turned on as early as 2hr and stayed active at least up to 24hr (Fig. 3.5). These results led us to start search for the underlying mechanisms that helped explain the very subtle induction of genes in our system. We speculated that when the expression level of one of the factors in the NOS circuit rose above a certain level, it repressed its endogenous promoter thus exerting a negative feedback regulation loop to keep its expression at a steady level (further discussions in Chapter 5).

Our results were consistent with other groups’ findings that Nanog is more highly expressed in 2i medium compared with in serum containing medium. The ESCs grown in 2i medium (third row of Fig. 4.7) expressed higher level of Nanog transcripts than those cultured in serum containing medium (third row of Fig. 4.5). The wild type KH2 cells on average expressed 56.6 Nanog transcripts in serum medium, and 122.3 Nanog transcripts in 2i medium (Fig. 4.9). Therefore, we expected the Nanog-inducible cells grown in 2i medium to express twice amount of total Nanog transcripts compared with those cultured in serum medium when the cells were ectopically perturbed with 2 µg/ml of doxycycline. However, in contrast to a moderate amount of induction of Nanog (~110% increase) for cells grown in serum medium, those cultured in 2i medium only showed a very subtle induction (~20% increase) (Fig. 4.5c and Fig. 4.7c). In both cell culture conditions, Nanog expression level became saturated at approximate 8-9E-04 transcripts after normalization by cell volume.
In both serum containing and 2i media, the NOS network exhibited an upper bound for its gene expression: Sox2 (~1.0E-03 transcripts after normalization), Oct4 (~1.0E-03 transcripts after normalization), and Nanog (~8-9E-04 transcripts after normalization). These results suggested that the core circuit’s transcript space might have a “fence” that restricted how high the transcription levels could overshoot beyond those in the unperturbed cells. We will revisit this “upper bound” issue and the potential underlying mechanisms in Chapter 5.

4.2.4 Nanog seemed to minimally regulate Sox2 and Oct4

Quite a number of studies have explored the importance of the NOS circuit for the pluripotency regulation in ESCs. However, the role of Nanog in this core regulatory circuit is still controversially discussed. A recent study has suggested that the Nanog-Sox2 interaction is critical for Sox2 function (Gagliardi et al., 2013). Conversely, it has been reported that the maintenance of pluripotency does not necessarily require continuously high Nanog levels. New experimental results provide evidence that Nanog is not a critical regulator of Oct4/Sox2 expression (Chambers et al., 2007).

In our study, the expression level of Nanog transcripts went down first and then returned to almost baseline in both Sox2 and Oct4 induction experiments (Fig. 4.1a-b, Fig. 4.2 a-b, Fig. 4.5a-b, and Fig. 4.7 a-b). It suggested that induction of Sox2 or Oct4 had little effect on Nanog transcript level on a time scale of more than 24 hours. These results were consistent with what we had observed in Chapter 3 (Fig. 3.6). Moreover, our results (Fig. 4.1c, Fig. 4.5c and Fig. 4.7c) demonstrated that over-expression of Nanog seemed to have minimal influence on Sox2 and Oct4 expression. For the ESCs grown in serum containing medium, when Nanog was ectopically up-regulated by more than two folds the expression levels of Sox2 and Oct4 did not change significantly compared with those in the unperturbed cells (Fig. 4.5c). The above results all together suggested that Nanog might not play a critical role in regulating Oct/Sox2 expression.
Figure 4.5: smFISH measurement of population average of Sox2, Oct4 and Nanog for the three inducible cell lines grown in serum containing medium. Each data point represents the population average (total mRNA counts normalized by cell volume) at a particular time point. Black dotted lines correspond to data obtained from the unperturbed cells ([doxycycline] = 0 µg/ml). Colored lines correspond to data obtained from the perturbed cells ([doxycycline] = 2 µg/ml). a, Expression profiles of Sox2-inducible cells. b, Expression profiles of Oct4-inducible cells. c, Expression profiles of Nanog-inducible cells.
Figure 4.6: Single-cell histograms of Sox2, Oct4 and Nanog transcripts (after normalization by cell volume) in three inducible cell lines grown in serum containing medium. Transcript counts in Sox2-inducible cells (a), Oct4-inducible cells (b), and Nanog-inducible cells (c). 1st row in each subfigure: Single-cell histograms of Sox2 transcripts. 2nd row in each subfigure: Single-cell histograms of Oct4 transcripts. 3rd row in each subfigure: Single-cell histograms of Nanog transcripts. “[0]” represents the condition in which no doxycycline is added to the culture medium. “[2]” represents the condition in which 2 µg/ml of doxycycline is added to the medium. “[2] 2hr” means the cells are treated with 2 µg/ml of doxycycline for 2hr.
Figure 4.6 (Continued): Single-cell histograms of Sox2, Oct4 and Nanog transcripts (after normalization by cell volume) in three inducible cell lines grown in serum containing medium. Transcript counts in Sox2-inducible cells (a), Oct4-inducible cells (b), and Nanog-inducible cells (c). 1st row in each subfigure: Single-cell histograms of Sox2 transcripts. 2nd row in each subfigure: Single-cell histograms of Oct4 transcripts. 3rd row in each subfigure: Single-cell histograms of Nanog transcripts. “[0]” represents the condition in which no doxycycline is added to the culture medium. “[2]” represents the condition in which 2 µg/ml of doxycycline is added to the medium. “[2] 2hr” means the cells are treated with 2 µg/ml of doxycycline for 2hr.
Figure 4.7: smFISH measurement of population average of Sox2, Oc4 and Nanog for the three inducible cell lines grown in 2i medium. Each data point represents the population average (total mRNA counts normalized by cell volume) at a particular time point. Black dotted lines correspond to data obtained from the unperturbed cells ([doxycycline] = 0 µg/ml). Colored lines correspond to data obtained from the perturbed cells ([doxycycline] = 2 µg/ml). a, Expression profiles of Sox2-inducible cells. b, Expression profiles of Oct4-inducible cells. c, Expression profiles of Nanog-inducible cells.
Figure 4.8: Single-cell histograms of Sox2, Oct4 and Nanog transcripts (after normalization by cell volume) in three inducible cell lines grown in 2i medium. Transcript counts in Sox2-inducible cells (a), Oct4-inducible cells (b), and Nanog-inducible cells (c). 1st row in each subfigure: Single-cell histograms of Sox2 transcripts. 2nd row in each subfigure: Single-cell histograms of Oct4 transcripts. 3rd row in each subfigure: Single-cell histograms of Nanog transcripts. “[0]” represents the condition in which no doxycycline is added to the culture medium. “[2]” represents the condition in which 2 µg/ml of doxycycline is added to the medium. “[2] 2hr” means the cells are treated with 2 µg/ml of doxycycline for 2hr.
Figure 4.8 (Continued): Single-cell histograms of Sox2, Oct4 and Nanog transcripts (after normalization by cell volume) in three inducible cell lines grown in 2i medium. Transcript counts in Sox2-inducible cells (a), Oct4-inducible cells (b), and Nanog-inducible cells (c). 1st row in each subfigure: Single-cell histograms of Sox2 transcripts. 2nd row in each subfigure: Single-cell histograms of Oct4 transcripts. 3rd row in each subfigure: Single-cell histograms of Nanog transcripts. “[0]” represents the condition in which no doxycycline is added to the culture medium. “[2]” represents the condition in which 2 μg/ml of doxycycline is added to the medium. “[2] 2hr” means the cells are treated with 2 μg/ml of doxycycline for 2hr.
Figure 4.9: Single-cell histograms of Nanog transcripts in unperturbed population of cells. Transcript counts of Nanog in ESCs grown in serum containing medium (a), and in 2i medium (b). The single-cell average of Nanog in serum medium is 56.6±2.2 (Mean±S.E.M; 775 cells), 122.3±3.1 (742 cells) in 2i medium.

So far, we presented the mRNA expression profiles of Sox2, Oct4 and Nanog upon perturbations. To obtain a more comprehensive picture of the core pluripotency regulatory circuit, we measured mRNA and protein expression simultaneously in single ESCs (Fig. 4.10, see information in the method section for details of smFISH/IF protocol). Fig. 4.11-4.13 demonstrated that the IF approach (antibody staining, image acquiring and image analysis) used in our study was quantitative for protein measurement. First, the high linear correlation between the IF intensity and the amount of mRNA transcripts expressed (Fig. 4.11a-c, Fig. 4.12a-c, Fig. 4.13a-c and Fig. 4.14a-c) suggested that the antibodies were sensitive to detect a small increase in the protein expression level that corresponded to a small increase in the mRNA transcript level. Second, there was a nearly perfect linear relationship between the IF intensity and the amount of antibody added to the cells (Fig. 4.11d, Fig. 4.12d and Fig. 4.13d). Third, measurement of protein expression in single cells enabled us to quantify the variation in heterogeneity in the expression levels of the core circuit due to perturbation. Besides IF, western blot
(WB) was also performed to measure the protein expression level (Fig. 4.18). Results from both approaches were mostly consistent besides that the expression level of Oct4 protein stayed almost constant in Fig. 4.15c (IF) but showed a down-regulation in Fig. 4.18c (WB).

4.2.5 Protein and mRNA expression profiles of the NOS circuit correlated well with each other

IF measurement of Sox2, Oct4 and Nanog proteins for the three inducible cell lines revealed very similar expression patterns as smFISH measurement of those three pluripotent factors (Fig. 4.5, Fig. 4.7, Fig. 4.15 and Fig. 4.17). On the protein level, we also observed that Oct4 decreased monotonically when Sox2 was ectopically induced, and vice versa. The expression level of Nanog proteins decreased first and then returned to almost baseline in both Sox2 and Oct4 induction experiments. Also, over-expression of Nanog proteins seemed to have minimal influence on Sox2 and Oct4 protein expression. Table 4.3 summarized the correlations between mRNA and protein population average expression profiles for the NOS network. The overall pattern was that mRNA and protein expressions on the population level seemed to be well correlated. The consistency between mRNA and protein expression profiles was further supported by the observation that there existed a high linear correlation between the IF intensity and the amount of mRNA transcripts expressed (Fig. 4.11a-c, Fig. 4.12a-c, Fig. 4.13a-c and Fig. 4.14a-c).

These high correlations could be potentially explained by the fact that all the three factors have relatively similar half-lives for their mRNA transcripts and the corresponding proteins. It has been suggested that Sox2 transcripts have a half-life of ~1.09hrs whereas its proteins have a few hours half-life. Oct4 mRNAs’ half-life has been suggested to be ~7.4hrs which is on a similar time scale (a few hours) as the half-life of its proteins. The half-lives of Nanog mRNAs and proteins have been found to be ~7hrs and ~5.2hrs respectively.
4.2.6 Upon perturbation Sox2 and Oct4 expression profiles became anti-correlated at both mRNA and protein levels

As discussed in 3.2.1, Sox2 and Oct4 transcript levels have a strong positive correlation in the unperturbed NOS network (Fig. 3.3g). This is supported by the fact that those two factors form heterodimers which regulate a wide-range of ESCs-specific genes (Chew et al., 2005). One study has proposed that the heterodimer complex maintains the constantly high Sox2 and Oct4 expression through a positive auto-regulation (Glauche et al., 2010). However, our results demonstrated that upon perturbation of either Sox2 or Oct4 the expression profiles of the two factors became anti-correlated at both mRNA and protein levels.

When Sox2 was over-expressed, Oct4 decreased monotonically at both mRNA and protein levels, and vice versa. This was observed in both serum containing medium and 2i medium (Fig. 4.5 a-b, Fig. 4.7 a-b, Fig. 4.15a-b and Fig. 4.17a-b). Moreover, in the Oct4 knockdown cells, Sox2 transcripts and proteins both increased monotonically when Oct4 expression was down-regulated (Fig. 4.19). Table 4.1 showed that the pairwise gene correlation between Sox2 and Oct4 on the transcript population average level became negative when one of the two factors was ectopically induced. Similarly, Table 4.2 demonstrated that protein expression profiles of Sox2 and Oct4 became negatively correlated when one of them was perturbed.

The anti-correlation between Sox2 and Oct4 expression profiles was similar to what had been observed in the RA-induced differentiation assay described in Chapter 2. When these two factors are differentially regulated leading to high Oct4 and low Sox2 levels, or low Oct4 and high Sox2 levels, either the mesendodermal (ME) fate or the neural ectodermal (NE) fate becomes available to the cell (Ema et al., 2008; Jiang et al., 2008; Silva and Smith, 2008; Peng et al., 2009; Schuettengruber and Cavalli, 2009; Han et al., 2010; Pasini et al., 2010; Thomson et al., 2011). One interesting experiment could be done in the future is to examine the expression pattern of various differentiation markers in the
Sox2- or Oct4 inducible cells and then compare the results with the control samples. For example, we might expect to observe co-expression of Sox2 and NE lineage specific markers such as Sox1, as well as co-expression of Oct4 and ME lineage specific markers such as Brachyury.

**Figure 4.10: smFISH/IF images of individual unperturbed ESCs.** smFISH and IF were performed simultaneously in the same single cell (see procedures described in the method section). Images in the left panel are example images of smFISH in individual ESCs. Each bright diffraction-limited spot is a single Sox2 mRNA (a), Oct4 mRNA (d), and Nanog mRNA (g) molecule. Images in the middle panel are example IF images in individual ESCs, Sox2 IF (b), Oct4 IF (e) and Nanog IF (h). Images in the right panel are merged smFISH and IF images that are shown in the left and middle panels.
Figure 4.11: Sox2 antibody optimization and quantification. 

a, Scatter plot of Sox2 IF intensity and mRNA counts normalized by cell volume. Each data point represents a single cell within a population. Each color corresponds to an Sox2 antibody concentration.

b, Apply moving average algorithm to (a). Each data point represents the average in that subset of data.

c, Collapse the plots in (b) into a single curve by multipliers that correspond to the dilution factors of the Sox2 antibody.

d, Scatter plot of average Sox2 IF intensity and Sox2 antibody concentration. The black solid line indicates the best fitted line. a.u., arbitrary units. \( r^2 \), the coefficient of determination.
Figure 4.12: Oct4 antibody optimization and quantification. a, Scatter plot of Oct4 IF intensity and mRNA counts normalized by cell volume. Each data point represents a single cell within a population. Each color corresponds to an Oct4 antibody concentration. b, Apply moving average algorithm to (a). Each data point represents the average in that subset of data. c, Collapse the plots in (b) into a single curve by multipliers that correspond to the dilution factors of the Oct4 antibody. d, Scatter plot of average Oct4 IF intensity and Oct4 antibody concentration. The black solid line indicates the best fitted line. a.u., arbitrary units. $r^2$, the coefficient of determination.
Figure 4.13: Nanog antibody optimization and quantification. a, Scatter plot of Nanog IF intensity and mRNA counts normalized by cell volume. Each data point represents a single cell within a population. Each color corresponds to an Nanog antibody concentration. b, Apply moving average algorithm to (a). Each data point represents the average in that subset of data. c, Collapse the plots in (b) into a single curve by multipliers that correspond to the dilution factors of the Nanog antibody. d, Scatter plot of average Nanog IF intensity and Nanog antibody concentration. The black solid line indicates the best fitted line. a.u., arbitrary units. $r^2$, the coefficient of determination.
Figure 4.14: Scatter plots of IF intensity and mRNA counts normalized by cell volume for Sox2, Oct4 and Nanog in unperturbed ESCs. Each data point represents a single ESC within a population. a, Sox2 IF versus mRNA, a dilution factor of 1:40 was used for the Sox2 antibody. b, Oct4 IF versus mRNA, a dilution factor of 1:400 was used for the Oct4 antibody. c, Nanog IF versus mRNA, a dilution factor of 1:10 was used for the Nanog antibody. \( r \), the coefficient of determination.
Figure 4.15: IF measurement of population average of Sox2, Oc4 and Nanog for the three inducible cell lines grown in serum containing medium. Each data point represents the population average (total IF intensity normalized by cell volume) at a particular time point. Black dotted lines correspond to data obtained from the unperturbed cells ([doxycycline] = 0 µg/ml). Colored lines correspond to data obtained from the perturbed cells ([doxycycline] = 2 µg/ml). a, Expression profiles of Sox2-inducible cells. b, Expression profiles of Oct4-inducible cells. c, Expression profiles of Nanog-inducible cells.
Figure 4.16: Single-cell histograms of Sox2, Oct4 and Nanog IF intensity (after normalization by cell volume) in three inducible cell lines. IF intensity in Sox2-inducible cells (a), Oct4-inducible cells (b), and Nanog-inducible cells (c). 1st row in each subfigure: Single-cell histograms of Sox2 IF intensity. 2nd row in each subfigure: Single-cell histograms of Oct4 IF intensity. 3rd row in each subfigure: Single-cell histograms of Nanog IF intensity. “[0]” represents the condition in which no doxycycline is added to the culture medium. “[2]” represents the condition in which 2 µg/ml of doxycycline is added to the medium. “[2] 2hr” means the cells are treated with 2 µg/ml of doxycycline for 2hr.
Figure 4.16 (Continued): Single-cell histograms of Sox2, Oct4 and Nanog IF intensity (after normalization by cell volume) in three inducible cell lines. IF intensity in Sox2-inducible cells (a), Oct4-inducible cells (b), and Nanog-inducible cells (c). 1st row in each subfigure: Single-cell histograms of Sox2 IF intensity. 2nd row in each subfigure: Single-cell histograms of Oct4 IF intensity. 3rd row in each subfigure: Single-cell histograms of Nanog IF intensity. “[0]” represents the condition in which no doxycycline is added to the culture medium. “[2]” represents the condition in which 2 µg/ml of doxycycline is added to the medium. “[2] 2hr” means the cells are treated with 2 µg/ml of doxycycline for 2hr.
Figure 4.17: IF measurement of population average of Sox2, Oc4 and Nanog for the three inducible cell lines grown in 2i medium. Each data point represents the population average (total IF intensity normalized by cell volume) at a particular time point. Black dotted lines correspond to data obtained from the unperturbed cells ([doxycycline] = 0 µg/ml). Colored lines correspond to data obtained from the perturbed cells ([doxycycline] = 2 µg/ml). a. Expression profiles of Sox2-inducible cells. b. Expression profiles of Oct4-inducible cells. c. Expression profiles of Nanog-inducible cells.
Figure 4.18: Western blot analysis of Oct4, Sox2 and Nanog expression in the three inducible cell lines grown in serum containing medium for various durations with the indicated concentrations of doxycycline. a, Results from Oct4-inducible cell line. b, Results from Sox2-inducible cell line. c, Results from Nanog-inducible cell line.
Table 4.1: Correlation coefficient of transcript profiles for all pairs of genes in the NOS circuit for the three inducible cell lines. Correlation of the population average of mRNA expressions for a pair of genes.

<table>
<thead>
<tr>
<th>Population average of mRNA</th>
<th>Ectopic perturbation of Sox2</th>
<th>Ectopic perturbation of Oct4</th>
<th>Ectopic perturbation of Nanog</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sox2 vs Oct4</td>
<td>-0.69</td>
<td>-0.23</td>
<td>0.65</td>
</tr>
<tr>
<td>Sox2 vs Nanog</td>
<td>-0.64</td>
<td>0.49</td>
<td>0.73</td>
</tr>
<tr>
<td>Oct4 vs Nanog</td>
<td>0.83</td>
<td>-0.67</td>
<td>0.47</td>
</tr>
</tbody>
</table>

Table 4.2: Correlation coefficient of protein profiles for all pairs of genes in the NOS circuit for the three inducible cell lines. Correlation of the population average of protein expressions for a pair of genes.

<table>
<thead>
<tr>
<th>Population average of protein</th>
<th>Ectopic perturbation of Sox2</th>
<th>Ectopic perturbation of Oct4</th>
<th>Ectopic perturbation of Nanog</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sox2 vs Oct4</td>
<td>-0.68</td>
<td>-0.86</td>
<td>0.74</td>
</tr>
<tr>
<td>Sox2 vs Nanog</td>
<td>-0.14</td>
<td>0.84</td>
<td>0.72</td>
</tr>
<tr>
<td>Oct4 vs Nanog</td>
<td>0.37</td>
<td>-0.90</td>
<td>0.42</td>
</tr>
</tbody>
</table>
Table 4.3: Correlation coefficient between smFISH and IF population average data for the three inducible cell lines grown in serum containing medium.

<table>
<thead>
<tr>
<th>Population average</th>
<th>Ectopic perturbation of Sox2</th>
<th>Ectopic perturbation of Oct4</th>
<th>Ectopic perturbation of Nanog</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sox2 mRNA vs protein</td>
<td>0.88</td>
<td>0.73</td>
<td>0.88</td>
</tr>
<tr>
<td>Oct4 mRNA vs protein</td>
<td>0.70</td>
<td>0.28</td>
<td>0.52</td>
</tr>
<tr>
<td>Nanog mRNA vs protein</td>
<td>0.55</td>
<td>0.83</td>
<td>0.94</td>
</tr>
</tbody>
</table>
Figure 4.19: smFISH and Immunofluorescence (IF) population average data for Oct4 knockdown cells grown in serum containing medium. 

a, smFISH data. Each data point represents the population average (total mRNA counts normalized by cell volume) at a particular time point. Black dotted lines correspond to data obtained from the unperturbed cells ([doxycycline] = 0 µg/ml). Red lines correspond to data obtained from the perturbed cells ([doxycycline] = 2 µg/ml).

b, IF data. Each data point represents the population average (total integrated immunofluorescence intensity normalized by cell volume) at a particular time point. Black dotted lines correspond to data obtained from the unperturbed cells ([doxycycline] = 0 µg/ml). Blue lines correspond to data obtained from the perturbed cells ([doxycycline] = 2 µg/ml).
Figure 4.20: Plots of coefficients of variation (CV) of Sox2, Oct4 and Nanog for the three inducible cell lines grown in 2i medium. a, CV profiles of Sox2-inducible cells. b, CV profiles of Oct4-inducible cells. c, CV profiles of Nanog-inducible cells. The blue dashed lines correspond to data obtained from the perturbed cells ([doxycycline] = 2 µg/ml). The red dashed lines correspond to data obtained from the unperturbed cells ([doxycycline] = 0 µg/ml). The CV is defined as the ratio of the standard deviation σ to the mean μ, and is calculated based on the formula CV = σ/μ. Error bars are 95% confidence intervals of the mean of 1000 bootstrap samples.
Figure 4.20 (Continued): Plots of coefficients of variation (CV) of Sox2, Oc4 and Nanog for the three inducible cell lines grown in 2i medium. **a**, CV profiles of Sox2-inducible cells. **b**, CV profiles of Oc4-inducible cells. **c**, CV profiles of Nanog-inducible cells. The blue dashed lines correspond to data obtained from the perturbed cells ([doxycycline] = 2 µg/ml). The red dashed lines correspond to data obtained from the unperturbed cells ([doxycycline] = 0 µg/ml). The CV is defined as the ratio of the standard deviation \( \sigma \) to the mean \( \mu \), and is calculated based on the formula \( CV = \sigma / \mu \). Error bars are 95% confidence intervals of the mean of 1000 bootstrap samples.
Figure 4.20 (Continued): Plots of coefficients of variation (CV) of Sox2, Oc4 and Nanog for the three inducible cell lines grown in 2i medium. a, CV profiles of Sox2-inducible cells. b, CV profiles of Oct4-inducible cells. c, CV profiles of Nanog-inducible cells. The blue dashed lines correspond to data obtained from the perturbed cells ([doxycycline] = 2 µg/ml). The red dashed lines correspond to data obtained from the unperturbed cells ([doxycycline] = 0 µg/ml). The CV is defined as the ratio of the standard deviation σ to the mean μ, and is calculated based on the formula CV = σ/μ. Error bars are 95% confidence intervals of the mean of 1000 bootstrap samples.
4.2.7 Overexpressing Sox2 or Oct4 increased transcription heterogeneity of the other two factors in the NOS circuit

To measure transcription heterogeneity of Sox2, Oct4 and Nanog in the ESCs grown in 2i medium, coefficients of variation (CVs) were computed based on the ratio of the standard deviation $\sigma$ to the mean $\mu$. Our results suggested that perturbing the core pluripotency circuit over a temporal range tuned the heterogeneity in transcript levels. Most interestingly, when Sox2 was overexpressed, the heterogeneity of Sox2 transcripts did not change much compared with that in the unperturbed cells (third plot of Fig. 4.20a). On the other hand, Oct4 and Nanog transcripts became more heterogeneous upon Sox2 overexpression (first and second plots of Fig. 4.20a). It suggested that when the cells sensed an overshoot of Sox2 transcripts they had the tendency to maintain the heterogeneity of Sox2 despite its increased expression level. The NOS circuit then responded by slightly expanding the transcript space occupied by Oct4 and Nanog. Similarly, the transcription heterogeneity of Oct4 stayed almost constant upon ectopic induction of Oct4 whereas the population of both Sox2 and Nanog became more heterogeneous (Fig. 4.20b).

4.2.8 Overexpressing Nanog transiently decreased the heterogeneity of Nanog transcripts while having minimal influence on Sox2 and Oct4

Navarro et al. (2012) reported an increased Nanog transcription heterogeneity upon Nanog overexpression. However, our results in 2i medium suggested otherwise: upon perturbation the heterogeneity of Nanog transcripts transiently decreased in the first 8hrs and then returned to baseline by 24hrs (first plot of Fig. 4.20c). The single-cell histograms (third row of Fig. 4.8c) helped us better visualize the shift in distribution of Nanog transcripts over time. The discrepancy between our results and those from Navarro et al. (2012) might be explained by the fact that different cell culture medium were used in the two studies. In contrast to the serum containing medium used by their group, for this part of
our study data from the cells grown in 2i medium were analyzed. It has been shown by other groups that Nanog transcripts are less heterogeneous in 2i medium compared with in serum containing medium. Therefore, the increased heterogeneity of Nanog transcripts might be caused by the presence of serum. In fact, our observations from serum containing medium (last row of Fig. 4.6c) were consistent with their results that Nanog transcripts became more heterogeneous over time upon ectopic induction of Nanog.

Upon Nanog overexpression, the heterogeneity of Sox2 and Oct4 remained almost constant throughout the entire perturbation assay (first and second rows of Fig. 4.8c; second and third plots of Fig. 4.20c). Together with the smFISH population average results, it suggested that perturbation of Nanog had minimal influence on the expression level and transcription heterogeneity of Sox2 and Oct4.

4.3 Discussion

In this chapter, we identified a few interesting features of the NOS network by quantitatively measuring mRNA transcripts and proteins in individual ESCs. One of the most striking observations was that the NOS network exhibited an upper bound for the gene expression of Nanog, Oct4 and Sox2 for the cells grown in either serum containing or 2i medium. These results suggested that the core pluripotency circuit’s transcript space might have a threshold that restricted how high the transcription levels could overshoot beyond those in the unperturbed cells. It thus naturally led us to explore the underlying mechanisms that helped explain the “upper bound” phenomenon (discussed in Chapter 5). Our initial hypothesis was that when the expression level of one of the factors in the NOS circuit rose above certain level, it repressed its own endogenous promoter to maintain its expression at a steady level.

Our results also demonstrated that Sox2 and Oct4 behaved in a symmetric way in response to perturbations. This “mirroring behavior” might be well explained by the fact that Sox2 and Oct4 physically form heterodimers that regulate a wide-range of ESCs-specific genes. It was further supported
by our single-cell data that the transcript levels of these two factors had a strong positive correlation in the unperturbed NOS network. However, on the other hand, results from the induction and knockdown experiments showed that Sox2 and Oct4 expression profiles became anti-correlated at both transcript and protein levels. This suggested that Sox2-Oct4 positive interaction might become uncoupled when one of them was perturbed.

Last but not least, our results seemed to be consistent with what some of the existing studies had reported that Nanog might not be a critical regulator of Oct4/Sox2 expression. The smFISH population average data together with the CV plots suggested that perturbation of Nanog has minimal influence on the expression level and transcription heterogeneity of Oct4 and Sox2. However, a recent study reported that Sox2 was identified as a robust interacting partner of Nanog, and the Nanog-Sox2 interaction was critical for Sox2 function (Gagliardi et al., 2013). A potential follow-up experiment would be to mutate the binding sites of Nanog-Sox2 and to examine the subsequent effect of Nanog perturbation on Sox2 expression.

4.4 Methods

4.4.1 ESC culture medium

The recipe of serum containing medium was the same as in 2.5.1. The 2i medium was made as previously described (Silva et al., 2008b). In brief, the ESCs were maintained without feeders in serum-free N2B27 prepared as described (Ying and Smith, 2003) and supplemented with LIF and 2i inhibitors, CHIR99021 (3μM), and PD0325901 (1 μM).

4.4.2 ESC adherent monoculture for doxycycline-induced perturbation assays

The cells were first cultured with irradiated MEFs on gelatinized culture plates for 2-3days till reached ~70-80% confluency. The cells were then trypsinized, and the mixture of both the ESCs and
MEFs were resuspended in 10ml of culture medium. The cell number was counted using hemocytometer. The cells were then seeded onto gelatinized culture plates in the presence of MEFs and cultured in serum containing medium for 12hrs. By the end of 12hr, the medium was removed and fresh medium with 2µg/ml of doxycycline (Invitrogen) was added to the plates. The cells were then collected by the end of each time point for imaging.

For the ESCs grown in 2i medium, they first went through a serial dilution of FBS (HyClone) for a few passages before getting ready for the perturbation experiments. In brief, the cells were first cultured in 2i medium supplemented with 10% FBS in the absence of feeders on gelatinized plates for 2-3 days. Then, the cells were passaged and grown in 2i medium supplemented with 5% FBS on gelatinized culture plates for another 2-3 days. The cells were subsequently passaged and cultured in 2i medium with 3%, 1%, 0.1% and eventually 0% FBS. The cell number was counted using hemocytometer. The cells were then seeded onto gelatinized culture plates in the absence of MEFs and cultured in 2i medium for 12hrs. By the end of 12hr, the medium was removed and fresh medium with 2µg/ml of doxycycline was added to the plates. The cells were then collected by the end of each time point for imaging.

4.4.3 Sample preparations for smFISH/IF

The samples were first prepared for smFISH (as described in 2.5.4). The cells were then washed with 2xSSC once for 5 minutes followed by two washes with 1xPBS each for 5 minutes at room temperature. The cells were incubated with 300µl of 5% BSA (in 1xPBS) solution containing 7.5µl of RNAsin (Ambion) and 15µl of SUPERasin (Ambion) for 1 hour at room temperature or overnight at 4º C. The cells were spin down and the pellet was resuspended in 100µl of 5% BSA (in 1xPBS) solution containing appropriate amount of the primary antibodies (1:40 for Sox2 antibody from Cell Signaling, 1:400 for Oct4 antibody from Abcam, and 1:10 for Nanog antibody from BD Pharmingen), 2.5µl of RNAsin, 5µl of SUPERasin and 0.5µl of 20% of Tween 20. The cells were incubated for 1 hour at room temperature and then washed with 0.1% Tween 20 (in 1xPBS) three times each for 10 minutes at room
temperature. The cells were spin down and the pellet was incubated with 100µl of 5% BSA (in 1xPBS) solution containing appropriate amount of the secondary antibodies (1:300 for all the secondary antibodies from Life Technologies), 2.5µl of RNAsin, 5µl of SUPERasin and 0.5µl of 20% of Tween 20 for 1 hour at room temperature. The cells were washed with 0.1% Tween 20 (in 1xPBS) three times each for 10 minutes at room temperature.

4.4.4 Immunoblot

For immunoblot analysis, samples were resolved on SDS-PAGE gels. Proteins were transferred to a PVDF membrane, blocked with 5% milk in PBST (0.1% Tween-20 in phosphate-buffered saline, pH 7.4), blotted overnight with primary antibody in PBST, and binding was detected by HRP-conjugated secondary antibody. Western blotting was performed with the following antibodies: Rabbit anti-Sox2 antibody (Santa Cruz, sc-20088X, 1:200); Rabbit anti-Nanog (Bethyl, A300-397A, 1:500); Mouse anti-Oct4 (Santa Cruz, sc-5279, 1:200); Mouse anti-β-actin (Abcam, ab8226, 1:1000).

4.4.5 Image acquisition and analysis (same as in 2.5.5 and 2.5.6)

To quantify fluorescence signal in each cell, we used ImageJ to obtain a Z-projection of 20 optical slices of the cell. For each image, which covered up to 100 correctly segmented cells, the mean fluorescence per pixel of each cell was computed. The minimum of mean fluorescence was taken to be the background. Then for each cell in the image, the total fluorescence of the cell was computed as the sum of the fluorescence at each pixel subtracting the background. If this value was negative, zero was used instead.
4.5 References


Chapter 5 Autorepression helps maintain the expression level of Oct4, Sox2 and Nanog in the ESCs

5.1 Introduction

Oct4, Sox2 and Nanog form a regulatory feedback circuit that maintains pluripotency in the ESCs. In this core circuit, all three transcription factors regulate themselves, as well as each other. Studies have suggested that the three factors bind to their own promoters, thus forming an interconnected autoregulation loop to maintain the ESC identity (Chen et al., 2008; Jaenisch and Young, 2008; Thomson et al., 2011; Navarro et al., 2012). However, nature of this autoregulation of the NOS network is still debatable. The previous view of the core pluripotency network proposed that Oct4, Sox2 and Nang form a self-reinforcing circuit (Jaenisch and Young, 2008) and was inferred from genome-wide analysis (Loh et al., 2006; Ivanova et al., 2006; Chen et al., 2008; Marson et al., 2008; Kim et al., 2008). A positive Nanog autoregulation was proposed to be mediated by Nanog dimer molecules; Oct4-Sox2 heterodimer complex was thought to maintain the constantly high Oct4 and Sox2 expression through a positive and self-reinforcing regulatory loop in pluripotent cells (Chew et al., 2005; Loh et al., 2006; Mullin et al., 2008; Wang et al., 2008; Glauche et al., 2010). Although based on the commonly believed assumption that binding of a transcription factor to a regulatory region of an active gene suggests that the transcription factor acts as an activator, experimental evidence supporting the inferred gene network architecture had not been generated. Moreover, some recent studies argued against the generally accepted idea of self-reinforcing feedback loop of the NOS network mentioned above. Navarro et al. (2012) suggested that Nanog activity is autorepressive, and OCT4/SOX2-independent. When the expression of Oct4 rose above a steady level, it was found to repress its own promoter thus exerting a negative feedback regulation loop to limit its own expression.

In Chapter 4, we showed that in both serum containing and 2i media the NOS network exhibited an upper bound for its gene expression. We speculated that the upper bound concept and autorepression might go hand in hand. Upon activating a transgene, expression level of the corresponding endogenous
gene went down via the autorepression mechanism thus keeping the total amount of that particular factor at a steady level. To comprehensively examine how endogenous genes respond to the presence of their regulators, in this chapter we applied smFISH to measure the integer counts of both the exo- and endogenous mRNAs for Oct4, Sox2 and Nanog in single ESCs whose core pluripotency circuit was perturbed by the addition of 2 µg/ml of doxycycline that turned on the inducible promoter $P_{TET}$. Doxycycline removal experiment was also performed to investigate whether or not the autorepression could be reversed when the transgene expression was turned off. Our results might shed light on the architecture of the NOS network as well as the underlying mechanisms of the autoregulation of Oct4, Sox2 and Nanog.

5.2 Results

5.2.1 smFISH probes were designed to accurately measure the exo- and endo-genous expression of Oct4, Sox2 and Nanog

Initially, we did not know the exact sequence of the construct used to generate the three inducible cell lines in our study. The only available information we had was that each of the three stable cell lines had an extra copy of one of the three pluripotency genes controlled by the inducible promoter $P_{TETO1}$, and stably integrated at the ColA1 locus (Fig. 3.4 and 3.5). A hypothetical structure of the construct (Fig. 5.1) was then proposed with the hope of finding sequences that were differentially expressed in exo- and endo-genous genes of a particular factor. The “5’ approach” aimed to design sets of primers to target sequences that spanned between the Tet-on promoter (TetOP) and part of the coding region of a gene of interest. On the other hand, the “3’ approach” aimed to target sequences that spanned between part of the coding region of a gene of interest and the SV40 polyadenylation signal (SV40 pA) or the poly(A) tail. Ideally, the “5’ approach” would generate a single polymerase chain reaction (PCR) product that corresponds to the exogenous transcripts of a particular factor whereas the “3’ approach” would give rise to two PCR products that could be either the exo- or endo-genous transcripts of a gene of
interest. Both approaches were tested, and the “5’ approach” successfully revealed the sequence differences between the exo- and endo-genous transcripts for Oct4, Sox2 and Nanog (Table 5.1). The 87 nucleotides (nts) of the mini-CMV promoter (part of the TetOP) were only expressed in the exogenous Oct4, Sox2 and Nanog. The 5’UTR regions were transcribed exclusively in the endogenous transcripts of Oct4 (20 nts), Sox2 (401nts) and Nanog (132nts). smFISH probes (Table 5.2) were then designed based on the sequence differences. Fig. 5.2 illustrated the approach to measure exo- and endo-genous transcripts of a particular factor. In brief, 3 smFISH probes were designed to target the mini-CMV promoter that was common among exogenous Oct4, Sox2 and Nanog. For the endogenous transcripts, probes were designed to target the 5’UTR region of the three transcription factors. Given the low number of smFISH probes used to measure the exo- and endo-genous gene expressions, the images acquired had the problems of low signal-to-noise ratio, high false-positive as well as high false-negative rate. Thus, it is critical to develop quantitative approaches for image acquisition and downstream analysis. An independent study from our group demonstrated an improved smFISH method allowing quantitative and accurate measurement of transcripts that differed by only a few nucleotides (Hansen and van Oudenaarden, 2013). We adopted the similar method and demonstrated that our measurements of the exo- and endo-genous transcripts were quantitative (Fig. 5.4–5.7; see Methods section for details of analysis).

![Diagram](image.png)

**Figure 5.1:** Schematic representation of the construct used to generate the three inducible cell lines in this study and two potential approaches to measure exo- and endo-genous gene expressions of Sox2, Oct4 and Nanog. The inducible promoter $P_{TETO1}$ is stably integrated at the ColA1 locus. “5’ approach” aims to target sequences that spanned between the Tet-on promoter (TetOP) and part of the coding region of a gene of interest. “3’ approach” aims to target sequences that spanned between part of the coding region of a gene of interest and the SV40 polyadenylation signal (SV40 pA) or the poly(A) tail.
<table>
<thead>
<tr>
<th>mini-CMV promoter</th>
<th>CGTCAGATCGCTGGAGAGCGCCATCCACGCTGTGACCTCCATA GAAGACACCCGGGACCGATCCAGCCTCCGCGGCCC GAATTC</th>
</tr>
</thead>
<tbody>
<tr>
<td>endo-Oct4</td>
<td>GAGGTTGAACCGTCCCTAGG</td>
</tr>
<tr>
<td>endo-Sox2</td>
<td>CTATTAACCTTGTTCAAAAAAGTATCAGGAGGTTGCAAGGCAGAGAGAGAGGAGAGGAGAGGAGAGAAGAAAGAGGA GAAGTTTGGAGCCCGAGGCTTAAGCCTTTCCAAACAAACTAATCACA ACAAATCAGCGCCGCGCCGAGGGAGAGCAGCGCCTGTGTTTTTTCTATCCC AATTTCACTTCGCCCCTCTGAGCTCCGCTTCCCTCCTCCCCCCAACTATTCT CGCCAGATCTCCGCGAGGGGCAGGTCAGCAGCAGCCCGGCCCCGCCCCG GCAGCCCCCTGCAATCGCCGCCGAGGCGCGGCCAGCCCGCCGCCAGCAGTCCCAG GCCGGGCGGAGGGTTGGCAGCCGCGGCGCAGGCAGGCAGGCAGGCCGCAAGCCGCAAGG</td>
</tr>
<tr>
<td>endo-Nanog</td>
<td>TCTATCGCCTTTGAGGCGGCTTCAGATAGGCTGATTGTTGATG TGTTTCTCTTCTTTGTTGGAAGGCTGCGCCTCAGTCCCTTTTCTGACC TTCTTGATATAATTGCTGAGCACATTTAACTCTTTCTTCT</td>
</tr>
</tbody>
</table>

Table 5.1 Sequences used to design smFISH probes to measure exo- and endo-genous gene expressions of Sox2, Oct4 and Nanog.
Table 5.2: smFISH probe sequences used in this study to measure exo- and endo-genous gene expressions of Sox2, Oct4 and Nanog.

<table>
<thead>
<tr>
<th>mini-CMV promoter (3 probes):</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>tctccagggcgatctgagt</td>
<td>ggtcaaaacagctggagt</td>
<td>cggcccgtgctttctatg</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>endo-Sox2 (11 probes):</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Gacaactctctgatactttt</td>
<td>caaacatcttcctctgc</td>
<td>gcaaaagtcttttccttt</td>
<td>cagcctagcttttaagagg</td>
</tr>
<tr>
<td>Ttctctctctctctctctt</td>
<td>gctccaaactctctctcctt</td>
<td>tttgaaagcttttaagctc</td>
<td>gcaagctttgttggattg</td>
</tr>
<tr>
<td>aaaaaacagggctctctc</td>
<td>agacgccccagaagcctttg</td>
<td>atctgccggagaagttgg</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>endo-Oct4 (1 probe):</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>cctagggagggctcctc</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>endo-Nanog (6 probes):</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>caacggtcagggcatatga</td>
<td>ccaaatcagctctctgaa</td>
<td>cacagaaagcaagacacc</td>
<td>aggaatggcggcagcttc</td>
</tr>
<tr>
<td>gaaaaattatcaagaagttca</td>
<td>gaaaagagttaaatgtct</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Figure 5.2: Schematic of smFISH probe design to detect exo- and endo-genous gene expressions of Oct4, Sox2 and Nanog. 3 smFISH probes are designed to target the mini-CMV promoter that is only expressed in the exogenous transcripts. To measure the endogenous gene expressions, smFISH probes are designed to specifically target the 5’UTR region of the three factors. Identification library (48 probes) contains probes complementary to both exo- and endo-genous transcripts.
Figure 5.3: Using unperturbed cells as a negative control to test binding specificity of the “CMV promoter” smFISH probe. a, Example images of smFISH in individual Sox2-, Oc4- or Nanog-inducible ESCs when no doxycycline was added. We used the “CMV promoter” probe in those unperturbed cells which did not ectopically express any of the three pluripotency genes. In the unperturbed cells, no spots were detected (a), as quantified by single-cell counts (Mean±S.E.M) for transcripts of Sox2 (0.46±0.05, 235 cells) (b), Oct4 (0.37±0.04, 261 cells) (c), and Nanog (0.70±0.08) (d).
Figure 5.4: Distribution of mRNA transcripts using identification and endogenous probes in the three inducible cell lines grown in serum containing medium with no doxycycline added. 

- **Data obtained from the Sox2-inducible cell line.**
  - **a**, Histogram of Sox2 transcripts detected using identification probe (48 probes) in the unperturbed cells ([doxycycline] = 0 µg/ml). 
  - **b**, Histogram of Sox2 transcripts detected using endo-Sox2 probe (11 probe) in the unperturbed cells ([doxycycline] = 0 µg/ml). 
  - **c**, Histogram of the endo-Sox2 probe detection rate compared with the identification probe. The average detection rate is 0.513.

- **Data obtained from the Oct4-inducible cell line.**
  - **d**, Histogram of Oct4 transcripts detected using identification probe (48 probes) in the unperturbed cells ([doxycycline] = 0 µg/ml). 
  - **e**, Histogram of Oct4 transcripts detected using endo-Oct4 probe (1 probe) in the unperturbed cells ([doxycycline] = 0 µg/ml). 
  - **f**, Histogram of the endo-Oct4 probe detection rate compared with the identification probe. The average detection rate is 0.495.

- **Data obtained from the Nanog-inducible cell line.**
  - **g**, Histogram of Nanog transcripts detected using identification probe (48 probes) in the unperturbed cells ([doxycycline] = 0 µg/ml). 
  - **h**, Histogram of Nanog transcripts detected using endo-Nanog probe (6 probe) in the unperturbed cells ([doxycycline] = 0 µg/ml). 
  - **i**, Histogram of the endo-Nanog probe detection rate compared with the identification probe. The average detection rate is 0.402.
Figure 5.4 (Continued): Distribution of mRNA transcripts using identification and endogenous probes in the three inducible cell lines grown in serum containing medium with no doxycycline added. 

- **a-c**, Data obtained from the Sox2-inducible cell line. 
  - **a**, Histogram of Sox2 transcripts detected using identification probe (48 probes) in the unperturbed cells ([doxycycline] = 0 µg/ml). 
  - **b**, Histogram of Sox2 transcripts detected using endo-Sox2 probe (11 probe) in the unperturbed cells ([doxycycline] = 0 µg/ml). 
  - **c**, Histogram of the endo-Sox2 probe detection rate compared with the identification probe. The average detection rate is 0.513.

- **d-f**, Data obtained from the Oct4-inducible cell line. 
  - **d**, Histogram of Oct4 transcripts detected using identification probe (48 probes) in the unperturbed cells ([doxycycline] = 0 µg/ml). 
  - **e**, Histogram of Oct4 transcripts detected using endo-Oct4 probe (1 probe) in the unperturbed cells ([doxycycline] = 0 µg/ml). 
  - **f**, Histogram of the endo-Oct4 probe detection rate compared with the identification probe. The average detection rate is 0.495.

- **g-i**, Data obtained from the Nanog-inducible cell line. 
  - **g**, Histogram of Nanog transcripts detected using identification probe (48 probes) in the unperturbed cells ([doxycycline] = 0 µg/ml). 
  - **h**, Histogram of Nanog transcripts detected using endo-Nanog probe (6 probe) in the unperturbed cells ([doxycycline] = 0 µg/ml). 
  - **i**, Histogram of the endo-Nanog probe detection rate compared with the identification probe. The average detection rate is 0.402.
Figure 5.5: Portion of detected dots. The average proportion of detected dots for endo-Oct4 (1 probe), endo-Sox2 (11 probes), and endo-Nang (6 probes), as compared with using the “identification” probe (48 probes) are 0.495, 0.513 and 0.402 respectively.
Figure 5.6: Scatter plots of the quantified relative intensities of Alexa 594 and Cy5 signals for Sox2 (a), Oct4 (b) and Nanog (c) transcripts respectively in cells that express both the exogenous (green) and endogenous (magenta) form of a particular factor. The black dashed lines indicate the optimal segmentation for exogenous versus endogenous transcripts using Support Vector Machine algorithm. a.u., arbitrary units.
Figure 5.7: The average correct assignment rates for Oct4 (4 probes), Sox2 (14 probes) and Nanog (9 probes) quantified with (orange) and without (blue) information from the “identification” channel.
5.2.2 Autorepression of Oct4, Sox2 and Nanog in both serum containing and 2i media

In Chapter 4, our results showed that the NOS network exhibited an upper bound for its gene expression in both serum containing and 2i media: Sox2 (~ 1.0E-03 transcripts normalized by cell volume), Oct4 (~ 1.0E-03 transcripts normalized by cell volume), and Nanog (~ 8.9E-04 transcripts normalized by cell volume). We speculated that when the expression level of one of the factors in the NOS circuit rose above a certain level, it repressed its endogenous promoter thus exerting a negative feedback regulation loop to maintain its overall expression at a steady level. To test our hypothesis of the autorepression mechanism of the NOS network, we measured the amount of total (exo + endo), exo- and endo-genous transcripts of a particular pluripotent factor when the ESCs were perturbed by 2 µg/ml of doxycycline for various durations. The population average plot of smFISH measurements demonstrated a clear autorepression for all three factors (Fig. 5.8). In serum containing medium, by 24hrs endogenous Oct4 transcripts decreased by 63% compared with the expression level in the unperturbed cells, 75% and 50% down-regulation for endogenous Sox2 and Nanog transcripts respectively (Fig. 5.8). Autorepression of the NOS network was also operational in 2i+LIF medium, suggesting independence of the MEK/GSK3 signaling pathways.

The expression profiles of the total, exo- and endo-genous transcripts of a particular pluripotent factor in the NOS network were very similar for the ESCs grown in serum containing or 2i medium. In both culture systems, the transgene expression (dashed lines with square markers in Fig. 5.8) was turned on as early as 2hrs, reached a plateau around 8-12hrs and then stayed at a constant level till 24hrs. On the other hand, the endogenous transcripts (dotted lines with diamond markers in Fig. 5.8) of all three factors decreased monotonically. Thus, it was no longer surprising that with the addition of 2 µg/ml of doxycycline Oct4 mRNA was over-expressed by only 10-20% compared with its expression level in the control samples. The negative feedback regulation loop might help keep the expression of each of the NOS network components at a steady level, thus maintaining the ESCs identity. Indeed, our results have demonstrated that the upper bound concept and autorepression go hand in hand.
Figure 5.8: smFISH population average data for the three inducible cell lines. Each data point represents the population average (total mRNA counts normalized by cell volume) at a particular time point. Solid lines with circle markers correspond to total (exogenous + endogenous) mRNA expression profiles. Dashed lines with square markers correspond to exogenous gene expression profiles. Dotted lines with diamond markers correspond to endogenous gene expression profiles. a-c, smFISH data for the three cell lines grown in serum containing medium. d-f, smFISH data for the three cell lines grown in 2i medium. 1st column: Sox2 expression profiles. 2nd column: Oct4 expression profiles. 3rd column: Nanog expression profiles.
Figure 5.9: Single-cell histograms of Sox2, Oct4 and Nanog transcripts (total, exogenous, endogenous) in three inducible cell lines grown in serum containing medium. Transcript counts of Sox2 in Sox2-inducible cells (a), Oct4 in Oct4-inducible cells (b), and Nanog in Nanog-inducible cells (c). 1st row in each subfigure: Single-cell histograms of total transcripts. 2nd row in each subfigure: Single-cell histograms of exogenous transcripts. 3rd row in each subfigure: Single-cell histograms of endogenous transcripts. “[2]” represents the condition in which 2 µg/ml of doxycycline is added to the medium. “[2] 2hr” means the cells are treated with 2 µg/ml of doxycycline for 2hr.
Figure 5.9 (Continued): Single-cell histograms of Sox2, Oct4 and Nanog transcripts (total, exogenous, endogenous) in three inducible cell lines grown in serum containing medium. Transcript counts of Sox2 in Sox2-inducible cells (a), Oct4 in Oct4-inducible cells (b), and Nanog in Nanog-inducible cells (c). 1st row in each subfigure: Single-cell histograms of total transcripts. 2nd row in each subfigure: Single-cell histograms of exogenous transcripts. 3rd row in each subfigure: Single-cell histograms of endogenous transcripts. “[2]” represents the condition in which 2 µg/ml of doxycycline is added to the medium. “[2] 2hr” means the cells are treated with 2 µg/ml of doxycycline for 2hr.
Figure 5.10: Single-cell scatterplots of Sox2, Oct4 and Nanog mRNA transcripts (exogenous versus endogenous) in three inducible cell lines grown in serum containing medium. a, Scatterplots of exogenous versus endogenous Sox2 in Sox2-inducible cells. b, Scatterplots of exogenous versus endogenous Oct4 in Oct4-inducible cells. c, Scatterplots of exogenous versus endogenous Nanog in Nanog-inducible cells. “[2] 2hr” means the cells are treated with 2 µg/ml of doxycycline for 2 hr.
Figure 5.11: Single-cell histograms of Sox2, Oct4 and Nanog transcripts (total, exogenous, endogenous) in three inducible cell lines grown in 2i medium. Transcript counts of Sox2 in Sox2-inducible cells (a), Oct4 in Oct4-inducible cells (b), and Nanog in Nanog-inducible cells (c). 1st row in each subfigure: Single-cell histograms of total transcripts. 2nd row in each subfigure: Single-cell histograms of exogenous transcripts. 3rd row in each subfigure: Single-cell histograms of endogenous transcripts. “[2]” represents the condition in which 2 µg/ml of doxycycline is added to the medium. “[2] 2hr” means the cells are treated with 2 µg/ml of doxycycline for 2hr.
Figure 5.11 (Continued): Single-cell histograms of Sox2, Oct4 and Nanog transcripts (total, exogenous, endogenous) in three inducible cell lines grown in 2i medium. Transcript counts of Sox2 in Sox2-inducible cells (a), Oct4 in Oct4-inducible cells (b), and Nanog in Nanog-inducible cells (c). 1st row in each subfigure: Single-cell histograms of total transcripts. 2nd row in each subfigure: Single-cell histograms of exogenous transcripts. 3rd row in each subfigure: Single-cell histograms of endogenous transcripts. “[2]” represents the condition in which 2 µg/ml of doxycycline is added to the medium. “[2] 2hr” means the cells are treated with 2 µg/ml of doxycycline for 2hr.
Figure 5.12: Single-cell scatterplots of Sox2, Oct4 and Nanog mRNA transcripts (exogenous versus endogenous) in three inducible cell lines grown in 2i medium. 

a, Scatterplots of exogenous versus endogenous Sox2 in Sox2-inducible cells. 

b, Scatterplots of exogenous versus endogenous Oct4 in Oct4-inducible cells. 

c, Scatterplots of exogenous versus endogenous Nanog in Nanog-inducible cells. “[2] 2hr” means the cells are treated with 2 µg/ml of doxycycline for 2hr.
5.2.3 Autorepression of Sox2, Oct4 and Nanog could be reversed upon doxycycline removal

In the previous section (5.2.2), we showed that when the expression level of one of the factors in the NOS network overshoot beyond a certain level, it exerted a negative feedback regulation to limit its own expression. Similar to the rationales for the RA removal experiments described in Chapter 2, the next question we would like to explore was whether or not the autorepression could be reversed upon doxycycline removal. The answers could potentially help us understand the following two important questions: 1) Is the autorepression of the endogenous transcripts terminal? 2) If the autorepression could be reversed, how long does it take for expression level of the endogenous transcripts return to the baseline level?

In this part of study, the ESCs were first treated with 2 µg/ml of doxycycline for 24hrs. By the end of 24hrs, doxycycline was removed and the cells were cultured in serum containing medium for another 24hrs. The population average plots of smFISH measurements in Fig. 5.14 showed that autorepression of Sox2, Oct4 and Nanog could be reversed upon doxycycline removal. The transgene expression (dashed lines with square markers in Fig. 5.14) was turned off as early as 4hrs post doxycycline removal and stayed at a near-zero level afterwards. The endogenous transcripts (dotted lines with diamond markers in Fig. 5.14) of all three factors were gradually up-regulated and eventually returned to baseline expression level 24hrs post doxycycline removal. This was further supported by Table 5.3 which provided a summary of the percentage of endogenous transcripts of the three factors compared with those in the unperturbed cells across various time points in the doxycycline removal assay. The total (exo + endo) amount of Sox2, Oct4 or Nanog transcripts in the perturbed cells eventually leveled off around baseline 24hrs after doxycycline was removed from the culture medium (Fig. 5.13). All together these results suggested that the ESCs seemed to prefer keeping the expression of each of the NOS network components at a steady level, thus maintaining the stem cell identity.
5.2.4 Reversal of autorepression of Sox2, Oct4 and Nanog occurred at different rates

As discussed in the previous section, the endogenous transcripts of all three factors were up-regulated and eventually returned to the expression levels in the unperturbed cells upon doxycycline removal. However, Table 5.3 suggested that endogenous Sox2, Oct4 and Nanog responded to doxycycline removal at different rates. Within the first 4hrs after doxycycline was removed, endogenous Oct4 was up-regulated most rapidly (106% - 46% = 60%), followed by Nanog (71% - 26% = 45%) and then Sox2 (61% - 38% = 23%). The dotted lines with diamond markers in Fig. 5.14 helped us better visualize the different rates of reversing autorepression: the upward slope for the first 4hrs post doxycycline removal was the steepest for endogenous Oct4 and the gentlest for endogenous Sox2. Coincidentally, in Chapter 2 we reported that Sox2 was the most gradually down-regulated during RA-induced differentiation whereas Oct4 responded to RA treatment at the fastest rate among the three factors. We used to speculate that the different response rates might be explained by the possibility that the mRNA transcripts of the three factors have different half-lives. According to the existing literature, the half-lives of Sox2, Oct4 and Nanog mRNA transcripts are 1.09hrs, 7.4hrs and 7hrs respectively. If based on the assumption that a transcription factor with a shorter half-life would respond to external regulatory signal at a faster speed compared with those factors with longer half-lives, we would expect to see Sox2 mRNA transcripts responded to doxycycline removal or RA treatment at the fastest rate followed by Nanog and then Oct4. Therefore, the different response rates might not be explained well by the fact that the transcripts of the three factors have different half-lives. An interesting follow-up experiment would be to investigate other possibilities such as different binding rates of a transcription factor to its own promoter.
5.2.5 Anti-correlation between Oct4 and Sox2 disappeared upon doxycycline removal

As discussed in 4.2.6, upon perturbation of either Sox2 or Oct4 the expression profiles of the two factors became anti-correlated at both mRNA and protein levels. When Sox2 was over-expressed, Oct4 decreased monotonically, and vice versa. Upon doxycycline removal, the anti-correlation relationship between the two factors disappeared. Fig. 5.13a showed that between 4 to 24hrs post doxycycline removal in the Sox2-inducible cells, the expression profiles of Sox2 and Oct4 followed similar pattern that both factors gradually returned to expression levels in the unperturbed cells. Similar trend was observed in the Oct4-inducible cells (Fig. 5.13b). These results suggested that in the absence of perturbation signals the ESCs had the tendency to return to the ground state in which Sox2 and Oct4 transcript levels had a strong positive correlation as showed in Fig. 3.3g.
Figure 5.13: smFISH population average data for the doxycycline removal assay in serum containing medium. Each data point represents the population average (total mRNA counts normalized by cell volume) at a particular time point. Black dotted lines correspond to data obtained from the unperturbed cells ([doxycycline] = 0 µg/ml). Colored lines correspond to data obtained from the doxycycline-treated cells ([doxycycline] = 2 µg/ml). a, Expression profiles of Sox2-inducible cells. b, Expression profiles of Oct4-inducible cells. c, Expression profiles of Nanog-inducible cells.
Figure 5.14: smFISH population average of total, exo- and endo-genous transcripts of Sox2, Oct4 and Nanog for the doxycycline removal assay in serum containing medium. Each data point represents the population average (total mRNA counts normalized by cell volume) at a particular time point. Solid lines with circle markers correspond to total (exogenous + endogenous) mRNA expression profiles. Dashed lines with square markers correspond to endogenous gene expression profiles. Dotted lines with diamond markers correspond to endogenous gene expression profiles. a, Sox2 expression profiles in the Sox2-inducible ESCs. b, Oct4 expression profiles in the Oct4-inducible ESCs. c, Nanog expression profiles in the Nanog-inducible ESCs.

Table 5.3: Percentage of endogenous transcripts of Sox2, Oct4 and Nanog compared with those in the unperturbed cells across various time points in the doxycycline removal assay. “dox treatment” denotes the ESCs were treated with 2 µg/ml of doxycycline. “post dox removal” denotes the cells were exposed to 2 µg/ml of doxycycline for 24hrs and then the doxycycline was removed from the medium. The percentage is calculated based on the formula ((number of endogenous transcripts in the treatment sample – number of endogenous transcripts in the control sample) / number of endogenous transcripts in the control sample)*100

<table>
<thead>
<tr>
<th>hr</th>
<th>endo-Sox2</th>
<th>endo-Oct4</th>
<th>endo-Nanog</th>
</tr>
</thead>
<tbody>
<tr>
<td>dox treatment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>12</td>
<td>55%</td>
<td>70%</td>
<td>44%</td>
</tr>
<tr>
<td>24</td>
<td>38%</td>
<td>46%</td>
<td>26%</td>
</tr>
<tr>
<td>post dox removal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>61%</td>
<td>106%</td>
<td>71%</td>
</tr>
<tr>
<td>12</td>
<td>74%</td>
<td>104%</td>
<td>80%</td>
</tr>
<tr>
<td>24</td>
<td>106%</td>
<td>108%</td>
<td>98%</td>
</tr>
</tbody>
</table>
Figure 5.15: Schematic of the core pluripotency circuit formed by Nanog, Oct4 and Sox2. a, Architecture of the network inferred from genome-wide analyses (Jaenisch and Young, 2008). b, Model architecture with autorepression of the NOS network.

Figure 5.16: Schematic representation of the potential binding sites of Sox2, Oct4 and Nanog to their own promoters. a, Sox2 binding enriched at the +3.7kb (downstream) and ~ -3.5kb (upstream) of the Sox2 locus. b, Oct4 binding enriched at the -1.8kb (upstream), -2.3kb (upstream) and -3.4kb (upstream) of the Oct4 locus. c, Nanog binding located within -5kb (upstream) of the Nanog transcription start site through unknown binding sites.
5.3 Discussion

Our study argues against the generally accepted idea that Oct4, Sox2 and Nanog form a self-reinforcing circuit (Jaenisch and Young, 2008; Fig. 5.15a): The NOS circuit exerts a negative feedback regulation loop to keep the expression of each of its components at a steady level (Fig. 5.15b). In Chapter 4, we showed that in both serum containing and 2i media the NOS network exhibited an upper bound for its gene expression. The experimental evidence of autorepression in this chapter suggested that the upper bound concept and autorepression might go hand in hand. When the expression level of one of the factors in the NOS circuit rose above a certain threshold, it repressed its endogenous promoter thus exerting a negative feedback regulation loop to maintain its overall expression at a steady level. One interesting idea worth pondering is that how the cells sense the threshold given that a subpopulation of the unperturbed cells are already expressing Sox2, Oct4 and Nanog above certain threshold. That might be explained by the possibility that the ESCs make the decision of gene regulation of a particular factor based on the global expression level of that factor in the whole system rather the local expression level in individual cells. The doxycycline removal experiment demonstrated that autorepression of Oct4, Sox2 and Nanog could be reversed, suggesting that the ESCs preferred keeping the expression of each of the NOS network components at a steady level thus maintaining the stem cell identity. Moreover, Oct4 and Sox2 became positively correlated upon doxycycline removal, again suggesting that the ESCs had the tendency to stay at the ground state.

In this chapter, we provided solid experimental evidence on the autorepression of the NOS circuit. Based on the existing literature, Fig. 5.16 illustrated the potential binding site(s) of Sox2, Oct4 and Nanog to their own promoter. It has been suggested that Sox2 binding was enriched at the +3.7kb (downstream) and ~ -3.5kb (upstream) of the Sox2 locus; Oct4 bound to its own promoter at the -1.8kb (upstream), -2.3kb (upstream) and -3.4kb (upstream) of the Oct4 locus. So far, there was no conclusive evidence on the exact binding site(s) of Nanog to its own promoter (Chen et al., 2008; Marson et al., 2008; Thomson et al., 2011). It has been suggested that Nanog repressed its endogenous promoter through
unknown binding sites within the -5kb (upstream) of the Nanog transcription start site (Navarro et al., 2012). To get a more complete picture of the underlying mechanisms about how each transcription factor represses its own endogenous promoter, one interesting follow-up experiment would be to mutate the binding site(s) and to study the subsequent effects.

5.4 Methods

5.4.1 ESC culture medium (same as in 4.5.1)

5.4.2 ESC adherent monoculture for doxycycline-induced perturbation assays (same as in 4.5.2)

5.4.3 Sample preparations for smFISH (same as in 2.5.4)

5.4.4 Image analysis for quantifying exogenous and endogenous transcripts

We applied the method developed by Hansen (Hansen and van Oudenaarden, 2013) to accurately quantify the exo- and endo-genous transcripts. Here, we extended the smFISH method (Raj et al., 2008) to detect and to quantify expression of mRNA variants that differ by very short sequence(s).

Based on the sequencing results, we obtained the sequence differences between the exo- and endo-genous transcripts for Sox2, Oct4 and Nanog (Table 5.1). For the exogenous transcripts, 3 smFISH probes were designed to target the mini-CMV promoter that was the common feature for exogenous Oct4, Sox2 and Nanog. One important control experiment was conducted to test the binding specificity of the “CMV promoter” smFISH probes. Fig. 5.3 demonstrated that no spots were detected when the “CMV promoter” probes were added to those unperturbed ESCs that did not ectopically express any of the three pluripotency factors.

For the endogenous transcripts, smFISH probes were designed to target the 5’UTR region of the three transcription factors (Table 5.2). First, we compared the detection rates of those endogenous probes
with the traditional 48 smFISH probes (“identification probes”) used to detect the total (exo + endo) amount of mRNA transcripts. The endogenous probes were added to the three inducible cell lines grown in serum containing medium with no doxycycline added. Ideally, we would expect the amount of smFISH dots detected by the endogenous probes to be the same as by those identification probes since only the endogenous transcripts were expressed in the absence of doxycycline. Figure 5.4 and 5.5 showed that compared with the identification probes the average detection rates of endogenous probes for Oct4, Sox2 and Nanog were 49.5%, 51.3% and 40.2% respectively. Given the low number of smFISH probes used to detect the endogenous gene expressions, it was not surprising that the false-negative detection rate was relatively high. However, this should not be a problem as long as we could do a proper control for the false-negative detection rate and adjust accordingly (discussed below). Also, if we only care about the relative fold changes of the endogenous gene expression, the problem of false-negative detection rate becomes even less prominent.

As mentioned earlier, the short sequence differences between exo- and endo-genous transcripts limited the number of probes that could be designed. For each transcription factor, the CMV-promoter probes were coupled to Cy5 whereas the endogenous probes were coupled to Alexa594. We utilized a third probe library coupled to TMR that contained identification probes complementary to both exo- and endo-genous transcripts (Fig. 5.2). The identification probe library was used to accurately identify the three-dimensional positions of mRNA transcripts. At each position, exo- or endo-genous specific information was obtained by quantifying the relative intensities between the Cy5 and Alexa594 channels (see Online Methods in Hansen and van Oudenaarden, 2013). A high percentage of independently detected Cy5 and Alexa594 spots colocalized with the identification TMR spots, suggesting that the majority of detected spots by the CMV-promoter and endogenous probes were real transcripts. To compute the correct assignment rate, we measured the relative intensity distributions in the ESCs expressing both exo- and endo-genous transcripts of a factor. More specifically, the spots from the exogenous transcripts formed a cloud along the Cy5 axis, while dots from the endogenous transcripts
formed a cloud along the Alexa594 axis (Fig. 5.6). The correct assignment rate was then determined by the local overlap in density between the exo- and endo-genous transcripts (Fig. 5.6 and 5.7; see Online Methods in Hansen and van Oudenaarden, 2013). Without using the information of identification probes, the correct assignment rates for Oct4, Sox2 and Nanog by the CMV-promoter and endogenous probes alone were 87.1%, 92.7% and 95.8% respectively (blue bars in Fig. 5.7). Colocalization with the identification TMR spots helped correct for the false-positive detection. With the additional information provided by the identification probes, the average correct assignment rates were further improved: 93.1%, 96.1% and 98.4% for Oct4, Sox2 and Nanog respectively (orange bars in Fig. 5.7).

5.5 References


Chapter 6 Potential network connectivity of the core pluripotency circuit upon perturbation

6.1 Introduction

ES cells self-renewal efficiency depends on the level of expression of hundreds of interconnected genes of the pluripotency regulatory network (Fig. 6.1). Among these, Oct4, Sox2 and Nanog play central roles (Fig. 6.2). Although the importance of the three factors for the regulation of pluripotency in ESCs is well established, their mutual interaction and the resulting regulatory dynamics are still incompletely understood and controversially discussed (Fig. 6.3). As mentioned in Chapter 5, some earlier studies have proposed that Oct4, Sox2 and Nanog form a self-reinforcing circuit (Fig. 6.3a; Loh et al., 2006; Ivanova et al., 2006; Chen et al., 2008; Jaenisch and Young, 2008; Marson et al., 2008; Kim et al., 2008). A positive autoregulation of the three factors has been reported by a few research groups (Chew et al., 2005; Loh et al., 2006; Mullin et al., 2008; Wang et al., 2008; Glauche et al., 2010). On the other hand, some recent studies have argued against the generally accepted idea of self-reinforcing feedback loop of the NOS circuit. For example, Navarro et al. (2012) has suggested a negative autoregulation of Nanog activity (Fig. 6.3b).

In Chapter 3, we have showed a simplified model of the NOS circuit with nine regulatory links among them (Fig. 3.3a). As mentioned earlier, each regulatory link represents the combination of direct and indirect interactions between the two factors it connects. When the network is perturbed, the strengths of these interaction links will change over time and the whole circuit will be rewired in complicated ways that are not yet well understood. With feedback, perturbation to a single factor can propagate to affect other components in the network, making it difficult to quantitatively unravel even a simple 3-factor network. In this chapter, we tackled this general challenge in network inference by employing Modular Response Analysis (MRA) that retrieves the topology and strength of network connections from experimentally measured network responses to successive perturbations of all modules (Kholodenko et al., 2002; see Method for details of analysis). All the experimental data used for analysis in this chapter are
from the perturbation experiments done in 2i medium (Chapter 4 and 5). Our network inference here is based on solid and quantitative experimental measurements, thus will shed light on the architecture and dynamics of the NOS network.

6.2 Results

6.2.1 NOS network upon perturbation over time

Before diving into details of the connectivity of the NOS network, we first computed a simple matrix (Fig. 6.4a) which was a direct read out from our experimental measurements of both mRNA and protein expressions of Nanog, Oct4 and Sox2 in the cells grown in 2i medium upon perturbation. Network components listed on top of the matrix represent the putative regulators, and those listed on the left represent the targets of those regulators. Matrix C was computed twice based on the population average data from smFISH and IF measurement respectively, and the outputs from the two measurements were consistent. Given the high positive correlation between mRNA and protein expressions for the three factors (discussed in Chapter 4), a single matrix was used to represent the gene regulatory output of the NOS network from each perturbation over time.

Our results suggested that overexpression of Nanog had a negative effect on all the three factors of the NOS circuit including itself (the three “-1”s in the first column of Matrix C in Fig. 6.4a). An increased expression of total (exo + endo) Nanog led to a down-regulation of endogenous Nanog across the entire 24hr perturbation assay while the expression levels of all the other components in the network remained nearly constant. On the other hand, Oct4 and Sox2 perturbations had almost no overall effect on Nanog expression throughout the time course experiments (the two “0”s in Matrix C in Fig. 6.4a). Moreover, Sox2 and Oct4 became negatively correlated at both mRNA and protein levels when one of them was perturbed (the two “-1”s under the Oct4” and “Sox2” columns of Matrix C in Fig. 6.4a). When either Oct4 or Sox2 was overexpressed, all the other factors in the network were down-regulated
simultaneously at \( t = 2 \)hrs. Therefore, we could not conclude the overall effect of Oct4 or Sox2 induction on their endogenous transcripts throughout the entire perturbation assay (the two “?”s in Matrix C in Fig. 6.4a).

It is important to note that Matrix C is a not a connectivity matrix of the NOS network, instead it is just a read out from the experimental data and not computed. There are many possible network diagrams that could generate Matrix C (Fig. 6.4b). It simply says that whatever the NOS network looks like, it has to yield the results summarized in Matrix C in the end.

### 6.2.2 MRA revealed network connectivity of the NOS circuit upon perturbation

In the previous section, we generated a matrix that summarized an overall picture of the NOS circuit upon each perturbation. Here, we used MRA (Kholodenko et al., 2002) to compute the connections among the three nodes of the network that could eventually yield phenotypes summarized in Matrix C. The average transcription profiles from smFISH measurements were used for network inference. All transcript counts were normalized to the wild-type mean, and the MRA algorithm was iteratively applied to bootstrap samples of the normalized data. The resulting connectivity matrices were used to determine the interaction strengths within the NOS network (Fig. 6.5).

There were a few really interesting observations that we could infer from the connectivity matrices. In Chapter 4, we reported that when Oct4 was ectopically induced in the cells grown in 2i medium, the transcript level of Nanog went down in the first 12hrs and then returned to almost the wild-type means by 24hrs (third plot in Fig. 4.7b). Our connectivity matrices may provide an explanation for that phenomenon. Oct4 on average exerted a strong negative regulatory effect on Nanog during the first 8hrs of the perturbation and then a strong positive influence by 24hrs (Fig. 6.5).

Here, we revisited one of our earlier findings that the strong positive interaction between Oct4 and Sox2 became anti-correlated when either of the two factors was perturbed (Chapter 4). Now with the
connectivity matrices, we could retrieve much more detailed information about the Oct4-Sox2 interaction. Our results suggested that the interaction between the two factors became asymmetric upon perturbation. Oct4 kept exerting a strong regulatory effect (thick arrows pointing from Oct4 to Sox2 in Fig. 6.5) on Sox2 throughout the entire 24hrs. On the other hand, Sox2 had a negligible or very weak influence on Oct4 (absence of arrows or thin arrows pointing from Sox2 to Oct4 in Fig. 6.5). These results suggested that in the Oct4-Sox2 pair Oct4 might be the more dominant player.

6.3 Discussion

There have been a lot of research and studies been conducted to determine the architecture of the NOS network. However, even till now the nature of the mutual interactions among Oct4, Sox2 and Nanog are still controversially discussed. Some studies have proposed positive regulations among the three factors as well as positive autoregulation for each of them (Chen et al., 2008; Jaenisch and Young, 2008; Marson et al., 2008; Kim et al., 2008). In contrast, a recent study has suggested a negative autoregulation of Nanog gene expression (Navarro et al., 2012). Moreover, some studies have reported that the regulatory loop among the three genes become uncoupled prior to differentiation. For example, when Oct4 and Sox2 are differentially regulated leading to high Oct4 and low Sox2 levels, or low Oct4 and high Sox2 levels, the cells are prone to differentiate towards ME or NE lineage respectively (Ema et al., 2008; Jiang et al., 2008; Silva and Smith, 2008; Peng et al., 2009; Schuettengruber and Cavalli, 2009; Han et al., 2010; Pasini et al., 2010; Thomson et al., 2011). In this chapter, we built simple yet robust models based on accurate and quantitative experimental measurements to illustrate the interconnection structure of the NOS network. Our results suggested that most of the regulatory interactions were inhibitory among the three factors (Fig. 6.4). Overexpression of Nanog had a negative influence on Oct4 and Sox2 as well as endogenous Nanog itself. Oct4 and Sox2 negatively regulated each other when either of them was perturbed. This seemed to contradict the generally accepted fact that those two factors
physically form heterodimers which regulate a wide-range of ESC-specific genes. However, this anti-correlation between Oct4 and Sox2 expression profiles suggest that the cells might be primed for differentiation into certain lineage upon perturbation. The network connectivity generated by the MRA algorithm helped us retrieve much more detailed information about the architecture of the NOS circuit. For example, in addition to the existing literature on the Oct4-Sox2 interaction, our connectivity matrices suggested that Oct4 might be the more dominant player.

One thing worth mentioning is that so far we used only the population means for model building in this chapter, and did not use any single cell information in getting those network topologies. Therefore, another possible model to try would be a Langevin stochastic simulation of the transcript levels in single cells using the inferred network topology generated by the MRA algorithm. If the simulation correctly reproduces the single cell scatters that measured by our smFISH approach, more confidence would be built in the inferred network diagrams.
Figure 6.1: Stem cell regulatory networks (adapted from MacArthur et al., 2009). Schematic showing the stem cell transcriptional regulatory circuit. This network was reconstructed from the data presented in various high-throughput chromatin immunoprecipitation (ChIP) experiments.

Figure 6.2: Core pluripotency circuit formed by Nanog, Oct4 and Sox2. a, OCT4, SOX2 and NANOG are central to the maintenance of ESC identity (adopted from MacArthur et al., 2009). b, Core transcriptional regulatory network in human ESCs (adopted from Boyer et al., 2005).
Figure 6.3: Examples of proposed architecture of the NOS circuit. a, Architecture of the network inferred from genome-wide analyses (Jaenisch and Young, 2008). b, Model architecture of the core regulatory network of ESCs with Nanog autorepression (adopted from Navarro et al., 2012). c, Conceptual scheme of the assumed interaction network (adopted from Glauche et al., 2010). Core network for the interaction between Oct4-Sox2 heterodimer and Nanog: positive auto-regulation of the Oct4-Sox2 complex, transcription activation of Nanog by Oct4-Sox2 and auto-regulation of Nanog.
Figure 6.4: Matrix to generate NOS network diagram. a, Matrix C is the read out from the smFISH and IF population average data for the three inducible cell lines grown in 2i medium. It summarizes the overall change of population average over time in each perturbation. $C_{ij} = 1$ means that when gene “i” is overexpressed, then gene “j” goes up. $C_{ij} = -1$ means that when gene “i” is overexpressed, then gene “j” goes down. $C_{ij} = 0$ means that when gene “i” is overexpressed, then gene “j” does not change. b, An example of network diagram that can produce matrix C.
Figure 6.5: Inferring the regulatory network within the NOS network using the MRA algorithm. Network components listed on top of the matrices represent putative regulators, and those listed on the left represent putative regulatory targets. Thickness of the arrows symbolizes the interaction strength. All the raw data are from the ESCs grown in 2i medium.
6.4 Calculation and Analysis

Summary of algorithm used for inferring the transcriptional network

We applied the algorithm developed by Kholodenko et al. (Kholodenko et al., *PNAS* 2002) to our RNA FISH measurements to infer the connections in a network formed by Nanog, Oct4, and Sox2. The resulting wiring diagram describes how perturbing the transcription of one gene (e.g., Nanog) affects the transcription of another gene (e.g., Sox2). Since this algorithm accounts for the sum of both direct and indirect effect of perturbing one node on another node of the wiring diagram. An example of a direct effect is, for example, overexpression of Nanog causing more Nanog proteins to directly bind to the promoter of Sox2, which then affects the transcription of Sox2. An example of an indirect effect is, for example, overexpression of Nanog causing changes in many other genes, not measured in our study, which in turn causes changes in the transcription of Sox2. Thus each arrow connecting two nodes in the wiring diagram represents all possible conglomerate interactions.

We first summarize the algorithm by Kholodenko et al. Our goal is to compute a 3x3 “connectivity matrix” $r$. An entry of this matrix, $r_{ij}$ is defined as

$$r_{ij} = \frac{x_i/x_i}{\Delta x_j/x_j} \quad (i,j = 1,2,3) \quad [1]$$

where $x_i$ represents the absolute counts of a given mRNA specie-i. $r_{ij}$ represents the fractional change in transcript level of gene “i” due to a fractional change in transcript level of gene “j”. The sign (plus or minus) of $r_{ij}$ dictates whether an activation or repression occurs to transcript-i when transcript-j level is increased.
But we cannot measure the elements of \( r \) directly. Instead, we perturb gene-\( j \), then measure what happens gene-\( i \) and gene-\( k \). Doing this for Nanog, Oct4, and Sox2 separately, we obtain a matrix \( R \), which is defined as

\[
R_{ij} = \frac{x_i(t) - x_i(0)}{(x_i(t) + x_i(0))/2} \tag{2}
\]

This represents the fractional change in the transcript level of gene-\( i \) from initial time to a later time \( t \) when transcription level of \( j \) is perturbed (the algorithm does not distinguish whether \( j \) is increased or decreased as a perturbation). Here, \( i,j = 1 \) (Nanog), 2 (Oct4), and 3 (Sox2). The algorithm by Kholodenko et al. shows that

\[
r = -(D(R^{-1}))^{-1} \cdot R^{-1} \tag{3}
\]

where \( r \) is the connectivity matrix and \( D(R^t) \) is the 3x3 diagonal matrix whose diagonal entries are given by the diagonal entries of \( R^t \). Although the original algorithm suggests that the network reach a steady-state at time \( t \) after the perturbation, this is not required and their method is actually more general.

**Our implementation of the network inference algorithm**

We measured the absolute mRNA counts in single cells, took their means, and used these population means to compute the connectivity matrix using equation [3]. Specifically, we subjected the three cell lines (Exogenous Nanog-, Sox2-, and Oct4-inducible cell lines) to zero and 2 \( \mu \)g/ml of doxycycline for various lengths of time (\( t = 2 \)hr, 4hr, 8hr, 12hr, and 24hr). Thus for each time point, we
have 3 data sets, one data set for each perturbed gene. Our \( t=0 \) (used in Equation 2) corresponds to the “unperturbed” state.

For a given time \( t \), we computed \( \mathbf{R} \) using equation [2]. To do so, we subsampled 50 cells from each data set (i.e., from the data set for perturbed and unperturbed cells). Then we computed the mean of the subsampled data sets for perturbed and unperturbed cells, which are \( x_i(t) \) and \( x_i(0) \) respectively. Using equations [2] and [3], we obtain \( r \) for the given time interval. We repeated this 10,000 times for the same time interval. This yielded 10,000 versions of the connectivity matrix \( r \). To obtain each matrix element \( r_{ij} \), we computed the mean of the distribution of the 10,000 values for that matrix element as well as the standard error (standard deviation / sqrt(10,000)).
6.5 References


Chapter 7 Final Conclusions

In summary, this thesis focused on adapting a high resolution single-cell measurement technique to investigate the dynamics of gene regulatory network formed by Oct4, Sox2 and Nanog in ESCs. By incorporating novel modifications into the existing smFISH technique, we managed to directly measure the enormous heterogeneity in the levels of the core pluripotency circuit’s mRNA transcripts as well as proteins in individual cells upon various types of perturbations such as differentiation, gain- and loss-of-function of one of the three factors. On top of the accurate and quantitative single-cell measurements, we also built robust network connectivity matrices to identify a few novel features of the core pluripotency circuit. We believe that the results of this work represent an important step in extending gene expression analysis to a variety of biological systems. Below we discuss some of ideas for possible directions that may result from our work.

7.1 Future Directions

Chapter 2 explores the precise effects of retinoic acid signaling on the differentiation and cell lineage decisions of ESCs. Our results disagree with one exiting study that has suggested that short-term treatment with RA in the early stage of differentiation is able to prevent spontaneous differentiation of ESCs and mains self-renewal (Wang et al., 2008). It would be interesting to treat the cells with 0.5 µM of RA for less than 24 hours (e.g. 2, 4, 8 and 12 hours) to quantify the extent of gene down-regulation and hopefully to find a “threshold” of the RA-induced differentiation effect on the stem cells. To add one more dimension to the differentiation assay, we could also expose the cells to a range of concentrations of RA (e.g. 0.05 – 1 µM).

Additionally, our results show that two mutually exclusive subpopulations coexist upon prolonged RA treatment. Even after treating the cells with RA for a long period of time (e.g. 7 days), a
small population of the cells still express some or all of the three pluripotency factors. One potential interesting experiment is to isolate that “persistent” subpopulation and to conduct a genome-wide assay to compare its gene expression profiles with the subpopulation that has undetectable Oct4/Sox2/Nanog expression as we all as with non-RA treated control samples.

More importantly, our observation of Sox2 being expressed in a complementary pattern to Oct4 upon RA treatment is consistent with what other groups have suggested that differential modulation of Oct4 and Sox2 precedes cell fate selection (Thomson et al., 2011; Ema et al., 2008; Han et al., 2010; Jiang et al., 2008; Pasini et al., 2010; Peng et al., 2009; Schuettengruber and Cavalli, 2009; Silva and Smith, 2008). Thus, it would be interesting to examine the expression pattern of various differentiation markers in the ESCs treated with RA and then compare the results with those non-RA-treated cells.

Finally, as mentioned in Section 2.4, there are a few potential pitfalls in the experimental design. To further improve the experimental setups, in the future studies we should consider effects of cell size, cell density, serum and feeders.

**Chapter 3** focuses on revealing dynamics on transcript pathway formed by Oct4, Sox2 and Nanog. Again, as mentioned in the “Potential pitfalls and caveats” section, we should take into account of the effect of off-feeders and also include proper control samples. To gain a deeper understanding of the architecture of the NOS circuit in ESCs, it would be more informative if we could examine the expression profiles of both mRNA and protein in single cells. Last but not least, it is critical to tease apart the transgene and endogenous gene expressions to fully comprehend the underlying mechanisms of NOS circuit regulation upon perturbations.

**Chapter 4** is the continuation of Chapter 3 by addressing the potential pitfalls of experimental design discussed in Chapter 3. Our results demonstrate that upon perturbation of either Sox2 or Oct4 the expression profiles of the two factors become anti-correlated at both mRNA and protein levels. The anti-correlation between Sox2 and Oct4 expression profiles is similar to what have been observed in the RA-
induced differentiation assay described in Chapter 2. A follow up study would be to examine the expression pattern of various differentiation markers in the Sox2- or Oct4 –inducible cells and then compare the results with the control samples.

In addition, our results seem to be consistent with what some of the existing studies have reported that Nanog might not be a critical regulator of Oct4/Sox2 expression. However, a recent study reports that Sox2 has been identified as a robust interacting partner of Nanog, and the Nanog-Sox2 interaction is critical for Sox2 function (Gagliardi et al., 2013). A potential follow-up experiment would be to mutate the binding sites of Nanog-Sox2 and to examine the subsequent effect of Nanog perturbation on Sox2 expression.

**Chapter 5** provides solid experimental evidence to convincingly validate the hypothesis mentioned in Chapter 4 that the NOS circuit exerts a negative autoregulation loop to restrict how high the transcription levels can overshoot beyond those in the unperturbed cells. Our results demonstrate that reversal of autorepression of Sox2, Oct4 and Nanog occur at different rates. We used to speculate that the different response rates might be explained by the possibility that the mRNA transcripts of the three factors have different half-lives. However, the different response rates seem not to be explained well by the fact that the transcripts of the three factors have different half-lives. An interesting follow-up experiment would be to investigate other possibilities such as different binding rates of a transcription factor to its own promoter.

The existing literature proposes the potential binding site(s) of Sox2, Oct4 and Nanog to their own promoter (Chen et al., 2008; Marson et al., 2008; Thomson et al., 2011). To get a more complete picture of the underlying mechanisms about how each transcription factor represses its own endogenous promoter, one interesting follow-up experiment would be to mutate the binding site(s) and to study the subsequent effects.
Chapter 6 focuses on building models based on experimental measurements to illustrate the interconnection structure of the NOS network. As mentioned in the discussion section, another possible model to try would be a Langevin stochastic simulation of the transcript levels in single cells using the inferred network topology generated by the MRA algorithm to check whether the simulation correctly reproduces the single cell scatters that measured by our smFISH approach.

7.2 References


