

T.R.
GEBZE TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

**QUANTITATIVE KINETIC MODELLING OF SIGNALING
PATHWAYS REGULATING PLURIPOTENCY**

SİMGE ŞENGÜL BABAL
A THESIS SUBMITTED FOR THE DEGREE OF
MASTER OF SCIENCE
DEPARTMENT OF MOLECULAR BIOLOGY AND GENETICS

GEBZE
2021

T.R.
GEBZE TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

**QUANTITATIVE KINETIC MODELLING OF
SIGNALING PATHWAYS REGULATING
PLURIPOTENCY**

SİMGE ŞENGÜL BABAL
**A THESIS SUBMITTED FOR THE DEGREE OF
MASTER OF SCIENCE**
DEPARTMENT OF MOLECULAR BIOLOGY AND GENETICS

THESIS SUPERVISOR
ASSIST. PROF. DR. PINAR PİR

GEBZE
2021

T.C.
GEBZE TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

PLURİPOTENSİYİ REGÜLE EDEN
SİNYAL YOLAKLARININ KANTİTATİF
KİNETİK MODELLENMESİ

SİMGE ŞENGÜL BABAL
YÜKSEK LİSANS TEZİ
MOLEKÜLER BİYOLOJİ VE GENETİK ANABİLİM DALI

DANIŞMANI
DR. ÖĞR. ÜYESİ PINAR PİR

GEBZE
2021



YÜKSEK LİSANS JÜRİ ONAY FORMU

GTÜ Fen Bilimleri Enstitüsü Yönetim Kurulu'nun 08/07/2021 tarih ve 2021/32 sayılı kararıyla oluşturulan jüri tarafından 05/08/2021 tarihinde tez savunma sınavı yapılan Simge Şengül BABAL tez çalışması Moleküler Biyoloji ve Genetik Anabilim Dalı Moleküler Biyoloji ve Genetik Programı YÜKSEK LİSANS tezi olarak kabul edilmiştir.

JÜRİ

ÜYE

(TEZ DANIŞMANI): Dr. Öğr. Üyesi Pınar PİR

ÜYE

: Prof. Dr. Işıl KURNAZ

ÜYE

: Dr. Öğr. Üyesi Hilal Taymaz NİKEREL

ONAY

Gebze Teknik Üniversitesi Fen Bilimleri Enstitüsü Yönetim Kurulu'nun

...../...../..... tarih ve/..... sayılı kararı.

SUMMARY

Stem cells are a special family of cells that have the ability to self-renew and differentiate into all types of tissues and cells. Stem cell studies contribute to the creation of models of various diseases, the explanation of many phenomena in areas such as developmental biology and medical biotechnology, and the development of new technologies.

Somatic cells can be transformed into induced pluripotent stem cells (iPSCs) into the stem cell state by ‘reprogramming’ processes. Thanks to reprogramming techniques, stem cells can be obtained by overcoming difficult processes such as isolation of rare stem cells from primary cells populations. Signaling pathways control the cellular processes of stem cells and by regulation of target gene expression, fate of the cell is drawn. Therefore, investigating the molecular interaction networks of signaling pathways in detail is of vital importance in stem cell studies.

The purpose of mathematical modeling is to explain important phenomena in biology, systems biology approaches, and complex processes under specific conditions by using mathematical equations. Generally, mathematical models are used to simulate states and dynamics of these biological processes and to predict different outcomes that have not been experimentally tested before.

In this thesis, quantitative kinetic modeling of cellular signal transduction pathways and their interaction networks in the reprogramming of somatic cells has been reconstructed and investigated. As a result of the analyzes based on molecular interactions of the 9 signaling pathways, the dynamic structure of the reprogramming of mouse embryonic fibroblast cells into induced pluripotent stem cells was investigated with deterministic and stochastic simulations. This model offers a comprehensive *in silico* working environment for stem cell studies.

Key Words: Stem cell, pluripotency, iPSCs, Signaling pathways, quantitative kinetic modeling.

ÖZET

Kök hücreler her türlü doku ve hücre tipine farklılaşma ve kendini yenileme özelliklerine sahip özel bir hücre grubudur. Kök hücre çalışmaları çeşitli hastalıkların modellerinin oluşturulmasında, gelişim biyolojisi ve medikal biyoteknoloji gibi alanlarda bir çok fenomenin açıklanmasında ve yeni teknolojiler geliştirilmesine katkı sağlamaktadır.

Somatik hücreler, 'yeniden programlama' süreçleriyle indüklenmiş pluripotent kök hücrelere (iPSC'ler) kök hücre durumuna dönüştürülebilir. Yeniden programlama teknikleri sayesinde, birincil hücre popülasyonlarından nadir bulunan kök hücrelerin izolasyonu gibi zorlu süreçlerin üstesinden gelinerek kök hücreler elde edilebilmektedir.. Sinyal yolları, kök hücrelerin hücresel işlemlerini kontrol eder ve hedef gen ekspresyonunun düzenlenmesiyle hücrenin kaderi çizilir. Bu nedenle sinyal yollarının moleküler etkileşim ağlarının detaylı olarak araştırılması kök hücre çalışmalarında hayati önem taşımaktadır.

Matematiksel modellemenin amacı, biyolojideki önemli fenomenleri, sistem biyolojisi yaklaşımları ile spesifik durumlar altındaki kompleks süreçleri açıklamaktır. Genellikle matematiksel modeller, bu biyolojik süreçleri denklemler ile simüle etmek ve önceden deneysel olarak test edilmemiş farklı çıktıları tahmin etmek için kullanılır.

Bu tezde, somatik hücrelerin yeniden programlanmasında hücresel sinyal iletim yollarının ve bunların etkileşim ağlarının kantitatif kinetik modellemesi yeniden yapılandırıldı ve araştırıldı. 9 sinyal yolunun moleküler etkileşimlerine dayalı analizler sonucunda fare embriyonik fibroblast hücrelerinin uyarılmış pluripotent kök hücrelere yeniden programlanmasının dinamik yapısı deterministik ve stokastik simülasyonlarla araştırıldı. Bu model, kök hücre çalışmaları için kapsamlı bir in silico çalışma ortamı sunmaktadır.

Anahtar Kelimeler: Kök hücre, pluripotensi, iPSCs, sinyal yolları, kantitatif kinetik modelleme.

ACKNOWLEDGEMENTS

I am excited and proud to complete my Master's Thesis, which I have been working on for a long time with serious effort and devotion. I will use this section as an opportunity to thank people who have helped and encouraged me.

First of all, I would like to express my sincere gratitude to my esteemed supervisor Asst. Prof. Pınar Pir who allowed me to explore in my own research interests, and guided me with her support and knowledge. I also would like to thank TUBITAK for funding this project.

In addition, I would like to thank valuable professors Işıl Kurnaz and Hilal Taymaz Nikerel, who participated in my thesis defense jury and provided constructive criticism at critical points to help to finalize the shape of my thesis.

I would like to thank all my professors who have shed light on me during my education. Their critiques in lectures and assessments have helped me improve my research skills immensely.

I would also like to thank our research team and my beloved colleagues. I have always felt their support, as well as joy, thanks to their amazing energy.

Finally, I would like to express my endless gratitude to my dear family, who have been by my side all my life, for their support and encouragement in all my decisions, without them nothing would have been possible...

I would like to express my deepest love to my dear husband Yiğit Koray BABAL, who has started walking with me at the beginning of this path and whose support and presence I have always felt.

TABLE OF CONTENT

	<u>Page</u>
SUMMARY	v
ÖZET	vi
ACKNOWLEDGEMENTS	vii
LIST of ABBREVIATIONS	x
LIST of FIGURES	xi
LIST of TABLES	xiii
1. INTRODUCTION	1
2. BACKGROUND ASPECTS	2
2.1. Stem Cells	2
2.2. Pluripotency and Reprogramming Mechanisms	5
2.3. Signaling Pathways Regulating Pluripotency	7
2.4. Quantitative Kinetic Modelling	11
3. MATERIAL AND METHODS	13
3.1. Materials	13
3.1.1. Literature resources and databases	13
3.1.2. Datasets	13
3.2. Methods	14
3.2.1. Quantitative kinetic modeling of signaling pathways	14
4. RESULTS	21
4.1. Estimating protein levels from mRNA level	21
4.2. Parameter Estimation	23
4.3. Time Course Simulations	25
4.4. <i>In Silico</i> Experiment of Doxycycline Induction	31
4.5. The Effect of Small-molecule Inhibitors on Pluripotency	32
4.6. Single and Combined Effects of OSKM Factors on Pluripotency	33
5. DISCUSSION AND CONCLUSION	35
5.1. The Model Topology	35
5.2. Parameter Estimation and Mathematical Stability	36
5.3. Signaling Crosstalk in Pluripotency	37

5.4. Conclusions	39
REFERENCES	41
BIOGRAPHY	46
APPENDIX	47

LIST of ABBREVIATIONS

<u>Abbreviations</u>	<u>Descriptions</u>
BMP4	: Bone Morphogenetic Protein 4
COPASI	: Complex Pathway Simulator
Dox	: Doxycycline
ESC	: Embryonic Stem Cell
GSK3 β	: Glycogen Synthase Kinase 3 Beta
hESC	: Human Embryonic Stem Cell
ICM	: Inner Cell Mass
iPSC	: Induced Pluripotent Stem Cell
JAK	: Janus Kinase
KEGG	: Kyoto Encyclopedia of Genes and Genomes
Klf4	: Krüppel-like factor 4
LIF	: Leukemia Inhibitory Factor
MEF	: Mouse Embryonic Fibroblast
mESC	: Mouse Embryonic Stem Cell
ODE	: Ordinary Differential Equation
RA	: Retinoic Acid
Shh	: Sonic Hedgehog
STAT3	: Signal Transducer and Activator Of Transcription 3
TGF β	: Transforming Growth Factor- β
WNT	: Wingless-related integration site

LIST of FIGURES

<u>Figure No:</u>	<u>Page</u>
2.1: Representation of stem cells.	2
2.2: Classification of stem cells.	3
2.3: Clinical application processes of induced pluripotent stem cells.	5
2.4: Four views of pluripotency.	6
2.5: An illustration of cell signaling.	7
2.6: Signaling pathways in the pluripotency and cell differentiation	11
3.1: Topological schematics of the kinetic model and reaction types.	15
3.2: Experimental procedures of reprogramming from MEF to iPSC by Velychko et al. 2019.	18
4.1: Linear regression model of protein concentration and mRNA level.	21
4.2: Correlation matrix of reference conversion factor proteins with model species using expression values from MEF to iPSC.	22
4.3: Comparisons of estimating parameter sets. Rows indicate the model parameters and columns indicate the different estimation sets 1-20.	24
4.4: Time course simulation of OSKM factor from MEF to iPSC.	26
4.5: Relative fold change of pluripotency- (red) and differentiation- (blue) related proteins.	27
4.6: Time-scale comparison of pluripotency and differentiation-related proteins.	27
4.7: Relative fold changes proteins of individual signaling pathways during MEF reprogramming.	28
4.8: Comparison of relative fold changes between RNAseq study and deterministic time course simulation.	29
4.9: Relative pluripotency and differentiation level during deterministic time course simulation.	30
4.10: Stochastic simulation of OSKM factors during reprogramming.	30
4.11: <i>in silico</i> experiment of transfection efficiency and doxycycline induction.	31

4.12: Different doxycycline induction levels by parameter scan of z parameters effect on pluripotency level	32
4.13: Small molecule inhibitors' effect on pluripotency level.	33
4.14: Single and combinatory effect of OSKM factor induction on pluripotency.	34

LIST of TABLES

<u>Table No:</u>	<u>Page</u>
3.1: Generalized rate laws of reaction types.	16
3.2: Lower and upper limits in parameters estimation.	19
4.1: Reference conversion factor of model species	22

1. INTRODUCTION

Induced pluripotent stem cells (iPSCs) are stem cells that can differentiate into any type of mature cells and are produced by the reprogramming of somatic cells. Thus, they have an important role in tissue engineering, artificial organ design, and disease modeling as well as shedding light on the treatment of most diseases in the field of regenerative medicine [Chagastelles & Nardi, 2011].

Differentiation and reprogramming mechanisms of iPSCs start with the external signals triggering the hierarchical regulation of the cell signaling network. External signaling molecules stimulate their target signal pathways from the cell membrane to the nucleus with a series of biochemical cascades. As a next step, proteins called transcription factors, which are the terminal elements of these cascades, increase or decrease the expression of the genes involved in differentiation processes by binding to their target DNA regions in the nucleus [Widłak, 2013]. Eventually, the dynamics of these signal pathways determine the phenotype and fate of the cell.

The purpose of mathematical modeling in biology is to investigate important biological phenomena, and complex biological processes under specific conditions with a system biology approach. Mathematical models are used to simulate the behavior of these biological processes and predict different outcomes which were not tested experimentally before [Le Novère, 2015].

This project aims to investigate the reprogramming processes of fibroblast cells to iPSCs and the effectiveness of the signal pathways involved in these processes by constructing mathematical kinetic models. Therefore, the methodology of this project is based on the construction of an ODE-based quantitative kinetic model of several signaling pathways by combining transcriptome data and literature information in COPASI software [Hoops et al., 2006]. In the construction of the kinetic model, reactions, initial conditions, and parameters related to signaling pathways were retrieved from literature and transcriptome studies. By using this kinetic model, different conditions were tested by time-course simulations to understand the dynamics of distinct processes in the reprogramming of fibroblast cells into iPSCs.

2. BACKGROUND ASPECTS

2.1. Stem Cells

Stem cells are the population of non-specialized and undifferentiated cells that are capable of differentiating into at least one type of mature cell of an organism and have the capacity for self-renewing (Figure 2.1) [Chagastelles & Nardi, 2011]. They exist in the embryonic, fetal, and adult phases of an organism and differentiate and proliferate to the main components of tissues and organs by changing their phenotype. Thanks to their unique properties, they are used as a potential medical treatment in regenerative medicine to regenerate damaged or lost cells and replace them with new cells. Over the years, stem cells have shed light on studies conducted to repair non-functioning parts of organs, test the effects of potential drugs, create and test disease models *in vitro* and investigate developmental processes [Kolios & Moodley, 2013].

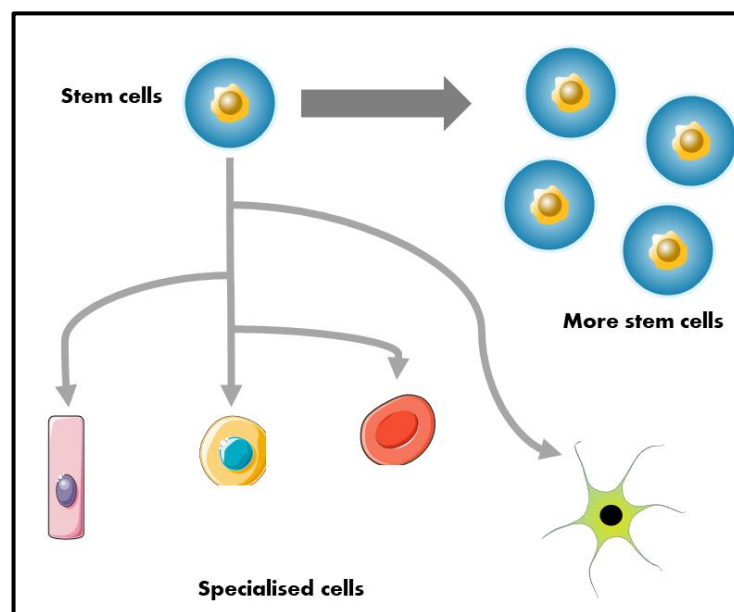


Figure 2.1:Representation of stem cells.

In the mid-1800s, it was realized that some cells could differentiate into different cells. In the early 1900s, stem cells were discovered, and they also were revealed that they could differentiate into blood cells. At the beginning of the 20th-century Russian histology Alexander A. Maximow, who is known for his experimental work on the

theory of hematopoiesis, discovered that all blood cells differentiate from precursor cells [Konstantinov, 2000]. In the 1960s, the specifications of stem cells were firstly defined by Ernest McCulloch and James Till [Sornberger, 2014]. Embryonic stem cells were first isolated by James Thomson for new transplantation applications and treatments. Pluripotent stem cells (also known as induced pluripotent stem cells) were reprogrammed from fibroblast cells with changing expression profiles of only four genes by Shinya Yamanaka and his team in Kyoto, 2006 [Takahashi & Yamanaka, 2006].

Stem cells are classified according to their ability to differentiate into different types of mature cells and the source from which they originate. The differentiation ability of stem cells into specialized cells is also known as “potency” and based on this ability, they are divided into 5 main groups. In addition to this classification, stem cells are also categorized into 3 main groups based on their sources (Figure 2.2).

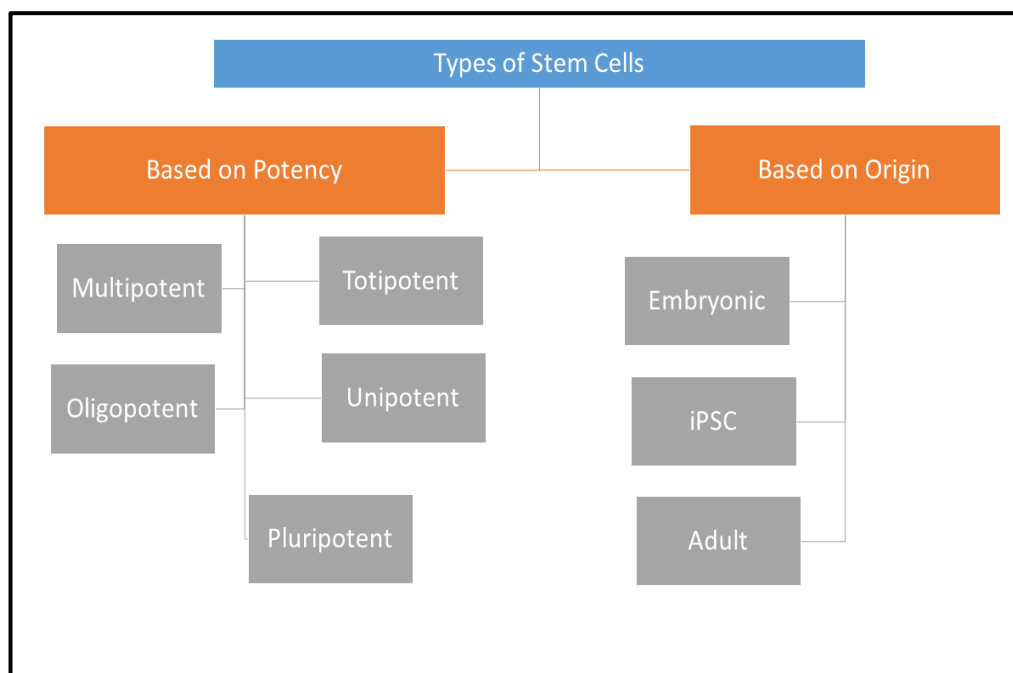


Figure 2.2: Classification of stem cells.

Totipotent stem cells have the ability to turn into any kind of body cell with an unlimited number of differentiation capabilities. Zygote and the first few cells that resulting from the division of the zygote are examples of totipotent cells [Kalra & Tomar, 2014]. Pluripotent stem cells can differentiate into cell types of the three different germ layers which are ectoderm, mesoderm, and endoderm that are found at

the beginning of embryonic development. Embryonic stem cells are pluripotent cells isolated from a blastocyst (an early stage embryo)(Plews & Wu, 013). Multipotent stem cells can differentiate into multiple lineages and an interrelated family of mature cell types. For example, hematopoietic stem cells can turn into most of the blood cells [Sobhani et al., 2017], but not other cell types in the organism. Oligopotent stem cells turn into only a few cell types, for example, myeloid and lymphoid stem cells. Unipotent or progenitor stem cells can form a single type of cell and differentiate in a single direction, for example, stem cells that are obtained from the brain can only turn into nerve cells [Kolios & Moodley, 2013].

Stem cells are divided into 2 categories according to the source from which they were obtained: Embryonic and adult stem cells [El Barky et al., 2017]. Embryonic stem cells (ESC) are totipotent stem cells found in the early embryonic stage and are obtained from the part of the blastocyst called inner cell mass (ICM). ESCs are pluripotent cells and can differentiate into cells of all 3 germ layers therefore they are valuable to the field of regenerative medicine and provide a renewable resource for investigating diseases and testing treatments [Rathjen & Rathjen, 2013]. Adult stem cells are undifferentiated and unspecialized cells found among differentiated or specialized cells in a tissue or organ. Adult stem cells, also known as somatic stem cells, can generally transform into a limited number of tissue types (multipotent) and are to maintain and repair the tissue in which they are localized. Adult stem cells emerge over time after embryonic development [Kalra & Tomar, 2014]. The third type of stem cell has also emerged in recent years. This type of stem cells called induced pluripotent stem cells (iPSCs), have similar properties to embryonic stem cells and are obtained from adult stem cells [Ye et al., 2013]. iPSCs are derived by manipulating the four genes (Oct4, Sox2, Klf4, Myc) called Yamanaka factors or cocktail, overexpression of these four transcription factors reprogram the somatic cells back into the pluripotent stage [Takahashi & Yamanaka, 2006]. iPSCs are the most preferred stem cell type for research purposes, due to their ability to transform into all cell types and decreasing the need for use of embryos. iPSCs have shed light on the field of regenerative medicine as they are easy to produce, manipulate, and culture [Sugimoto & Eto, 2021] (Figure 2.3). Although they have some disadvantages such as uncontrolled proliferation and integration, they are still one of the most used methods in stem cell therapies [Medvedev et al., 2010].

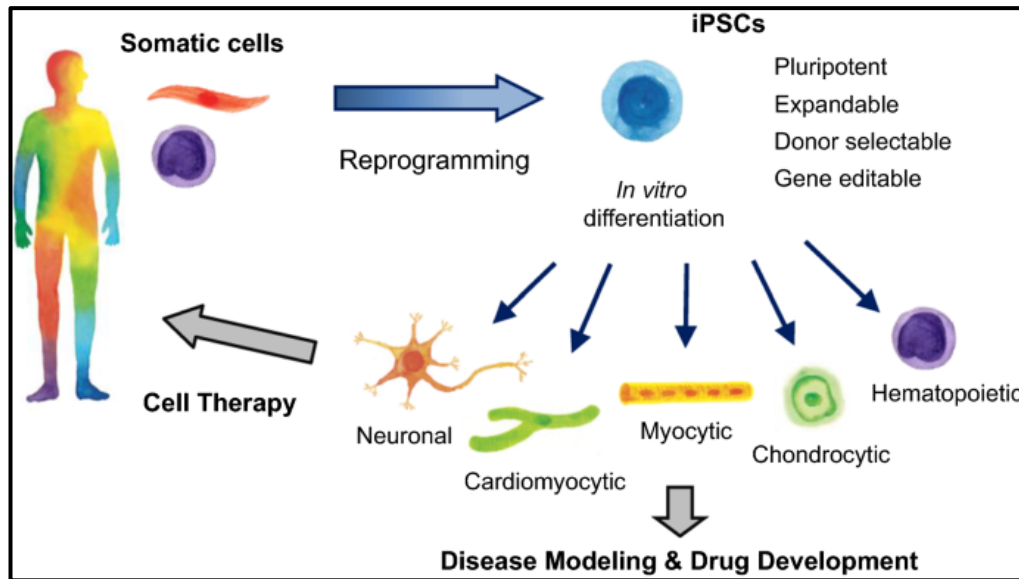


Figure 2.3: Clinical application processes of induced pluripotent stem cells.

2.2. Pluripotency and Reprogramming Mechanisms

The revolutionary discovery in 2006 which includes the process of returning somatic cells to a pluripotent state and transforming them back into different germ layers, has paved the road for many studies. This pioneering work enables the production of both embryonic stem cells and induced pluripotent stem cells, was awarded the Nobel Prize in 2007 and 2012. Understanding pluripotency and reprogramming processes have a pivotal role in the fields of developmental biology, regenerative medicine, and drug discovery.

Pluripotent stem cells have an unlimited ability to regenerate and differentiate into other cell types [Smith, 2001]. The concept of pluripotency is specific to pluripotent cells and is used for cells that have not yet differentiated and committed to cells of 3 germ layers [Yilmaz & Benvenisty, 2019]. Pluripotency is provided by a system created with signaling molecules and supported by gene regulation [Medvedev et al., 2010]. Studies which explain pluripotency have a focus on four characteristic properties [Yilmaz & Benvenisty, 2019] (Figure 2.4). First, pluripotency has features that cover the development process. It means that classical pluripotent stem cells (PSCs) are embryonic stem cells, derived from the inner cell mass of blastocyte and can differentiate into 3 germ layers [De Los Angeles et al., 2015]. The second feature refers to transcriptomic and epigenetic profiles and defines marker genes that are unique for pluripotency and differentiation [Weinberger et al., 2016]. The third point

of view defines that the reprogramming process is induced in somatic cells by certain factors and as a result, induced pluripotent stem cells (iPSCs) are produced. The state of being induced and regulated by these factors is called “induced pluripotency”. Takahashi and Yamanaka explained that with these transcription factors - Oct4, Sox2, Klf4, c-Myc (OSKM)-, somatic cells are reprogramed into iPSCs [Takahashi & Yamanaka, 2016]. The last and fourth feature is essentialome, deals with the expression and analysis of specific genes necessary for the survival of pluripotent stem cells [Yilmaz & Benvenisty, 2019].

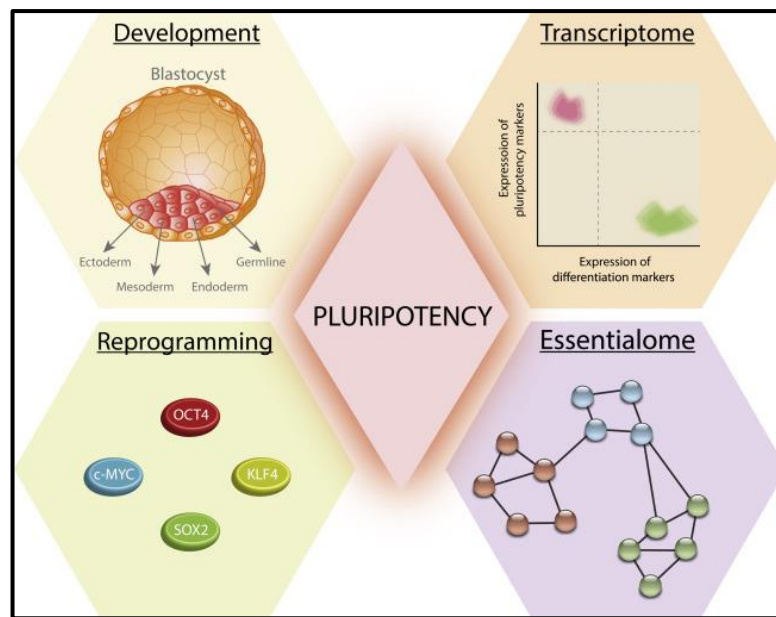


Figure 2.4: Four views of pluripotency.

In reprogramming, somatic cells are reprogrammed and brought back to the pluripotent state, which is ready to proliferate and differentiate into various cell and tissue lineages. Patient-specific iPSCs can be produced by the genetic manipulation of OSKM with retrovirus-mediated transfection [Takahashi et al., 2007]. Then, iPSCs are used in stem cell therapy, to contribute to drug screening, tissue engineering, disease modeling, etc. [Teshigawara et al., 2017]. It is also possible to convert any somatic cell type to higher efficiency and higher quality iPSCs compared to embryonic stem cells [Hochedlinger & Plath, 2009].

2.3. Signaling Pathways Regulating Pluripotency

Signaling pathways have important roles in determining the fate of cells. These pathways, which govern the cellular communication processes, create the coordination network for transferring hormones or other signal molecules from the outside of the cell to the cell nucleus and orchestrate the cellular response. Hence, signal molecules carry information as a cascade to the DNA and regulate cell growth, cell division, proliferation, apoptosis, or regulation of gene expression [Widłak, 2013] (Figure 2.5).

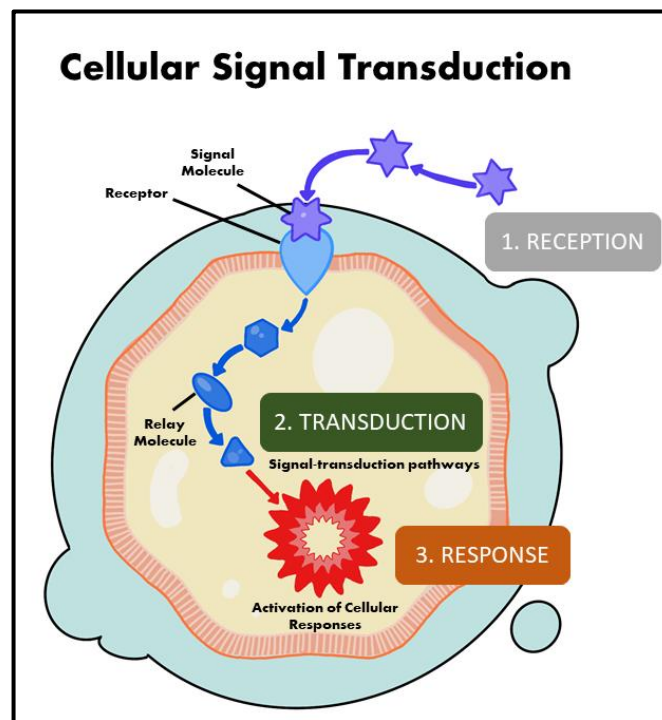


Figure 2.5: An illustration of cell signaling.

Signaling pathways play a critical part in processes such as differentiation of stem cells, maintaining pluripotency, and reprogramming of somatic cells by targeting a subset of genes. Changes in gene expression required for the differentiation of embryonic stem cells are the target of signaling pathways, on the other hand, pluripotency state is also related to the regulation of the gene expression that is the target of these signaling pathways [Okita & Yamanaka, 2006]. In addition, it is also possible to artificially design the necessary regulation of these signaling pathways with various feedback and feedforward loops to reprogram the cell [Padala et al., 2017].

The first pluripotency-associated signal transduction pathway identified in naïve mouse embryonic stem cell (mESC) is LIF (leukemia inhibitory factor) pathway. LIF signaling pathway has an important role in supporting naïve pluripotency together with other pathways in mESC [Hassani et al., 2013]. LIF ligand, which belongs to the interleukin family, binds to the LIF receptor and activates the heterodimer receptor complex. This complex then phosphorylates Janus kinase (Jak) which is a member of the cytoplasmic tyrosine kinase family and activates 3 intracellular signaling pathways: Jak/STAT3 pathway, PI3K/Akt pathway, and MAPK/Erk pathway [Graf et al., 2011]. Activated Jak promotes Jak/STAT3 pathway and phosphorylates STAT3 (Signal Transducer And Activator Of Transcription 3) first and phosphorylated STAT3 is transported into the nucleus to mediate the target gene expressions. Previous genome-wide studies have shown that STAT3 activates transcription of c-Myc and Klf4 (Krüppel-like factor 4) which play important roles in the reprogramming of somatic cells [Niwa et al., 2009]. PI3K/Akt intracellular pathway is also activated by LIF in mESC and hESC. Nanog and c-Myc expressions are increased by PI3K/Akt signaling pathway through glycogen synthase kinase 3 beta (GSK3 β) thereby maintain pluripotency [Bechard & Dalton, 2009]. LIF can also activate MAPK/Erk signaling pathway, which is one of the most important pathways for cellular communication and regulation, to induce differentiation of mESCs [Burdon et al., 1999]. To cultivate mESCs in serum, in addition to LIF, BMP4 (bone morphogenetic protein 4) also is a significant component of the cell culture conditions and maintains pluripotency [Ying et al., 2003].

BMP proteins are members of the transforming growth factor- β (TGF β) superfamily and are discovered as an element involved in the differentiation of bone cells. Their ligands, particularly BMP4, activate the BMP signaling pathway with Smads (a fusion of *Caenorhabditis legans* *Sma* genes and the *Drosophila* *Mad*, Mothers against decapentaplegic), a well-conserved transcription factor family, and supports bone differentiation in embryo formation [Massagué et al., 2005]. BMP ligands bind to type I and II serine/threonine kinase membrane receptor and establish an active heterotetramer complex which then activates Smads and inhibitors of differentiation (*Id*) genes to suppress neural differentiation. In addition, BMP4 pathway inhibits MAPK signaling pathway through the dual-specificity phosphatase 9 (DUSP9) to sustain pluripotency in mESCs [Hassani et al., 2013]. In summary, BMP

and LIF pathways interoperate a cascade towards the nucleus to support the self-renewal of mESC.

TGF β superfamily plays a vital role in tissue homeostasis with 33 different ligands and regulates important cellular processes such as proliferation, differentiation, and apoptosis. Many important diseases such as cancer or vascular disorders emerge when mutations and dysregulations occur in TGF β pathway [Nicklas & Saiz, 2013]. Fate decisions, development of tissues and organs which are important for embryogenesis are also regulated TGF β signaling pathway. Recent studies have revealed that TGF β pathway has a function in naïve and primed pluripotency [Park, 2011], inhibition of the TGF β signaling pathway increases the efficiency of mouse-induced pluripotent stem cell formation [Galvin et al., 2010]. TGF β , Activin, Nodal, and BMP bind to the membrane receptors type I and II, then phosphorylation cascades progress via Smads proteins into the nucleus to transmit the signal. Activated Smads then regulate the expression of pluripotency-related gene Nanog to sustain human ESC self-renewal [Xu et al., 2008].

Wnt (Wingless-related integration site) signaling pathway is initiated and directed by members of the large family of Wnt proteins, also known as GSK3 β or β -catenin signaling pathway. The Wnt ligand molecule binds to the receptor called Frizzled and initiates the canonical phosphorylation cascade. Each phosphorylated molecule becomes active and triggers the signal transduction of its next element. GSK3 β molecule then suppresses activities of its main targets β -catenin, cMyc, and Nanog. Wnt signaling pathway is involved in the maintenance of self-renewal and supporting naïve pluripotency of the stem cell [Sokol, 2011].

Sonic Hedgehog (Shh) signaling pathway has a part in cellular processes such as embryonic cell differentiation, cancer metastasis, proliferation, and stem cell regeneration. Shh ligand activates GLI family transcriptional effectors and induces the expression of target genes PTCH, GLI, BCL2, and CCND. Upregulation of these genes play a role in the stability of stem cell and neuronal cell fates [Hooper & Scott, 2005]. Recent studies indicated that PTCH1 and GLI3 are key regulators for the core pluripotency transcriptional network of OSKM [Wu et al., 2010].

The Notch signaling pathway is involved in determining cell fate, especially direct differentiation into the neuronal lineage, and neurogenesis in embryonic development is regulated by this pathway [Hansson et al., 2004]. Different types of evolutionarily conserved Notch ligands such as Delta-like1/4 and Jagged bind to the

Notch receptor, and activate the intracellular region of the Notch protein, and lead to activation of important transcription factors such as the Hes1 and Pax7 [Lowell et al., 2006].

The retinoic acid (RA) signaling pathway is activated by the entry of retinol from the blood into the cell and then activates its targets and is involved in the specialization and differentiation of cells. It triggers anterior-posterior development by targeting Hox genes, which are a very large and diverse family, and regulate the formation of an embryo along the head-tail axis [Nolte et al., 2019]. When other signaling pathways work together, they coordinate the balance of differentiation and roots in embryogenesis. Retinoic acid also activates the Ascl1 gene, which has a role in neuronal differentiation and triggers cell differentiation. Due to these two functions, RA is used in cancer treatments [Ghyselinck & Duester, 2019].

Small-molecule inhibitors against MEK1/2 and GSK3B, referred to as 2i in this study, have been used for supporting pluripotency of mouse embryonic stem cell culture even there are not any LIF and BMP4 (the main component in serum that sustains pluripotency). These inhibitors -PD0325901 and CHIR99021- inhibit the FGF and GSK3 β pathways respectively (Figure 2.6).

All these signaling pathways provide a balance between differentiation and reprogramming by creating crosstalk both individually and together. The reason why LIF signal pathway is very important in stem cells is because it initiates 3 different intracellular signaling pathways and establishes a crosstalk network between them. This balance is maintained by both feedback mechanisms and inhibitor molecules. Especially activity of transcription factors and epigenetic profiles are keys that play an important role in maintaining this balance and stability.

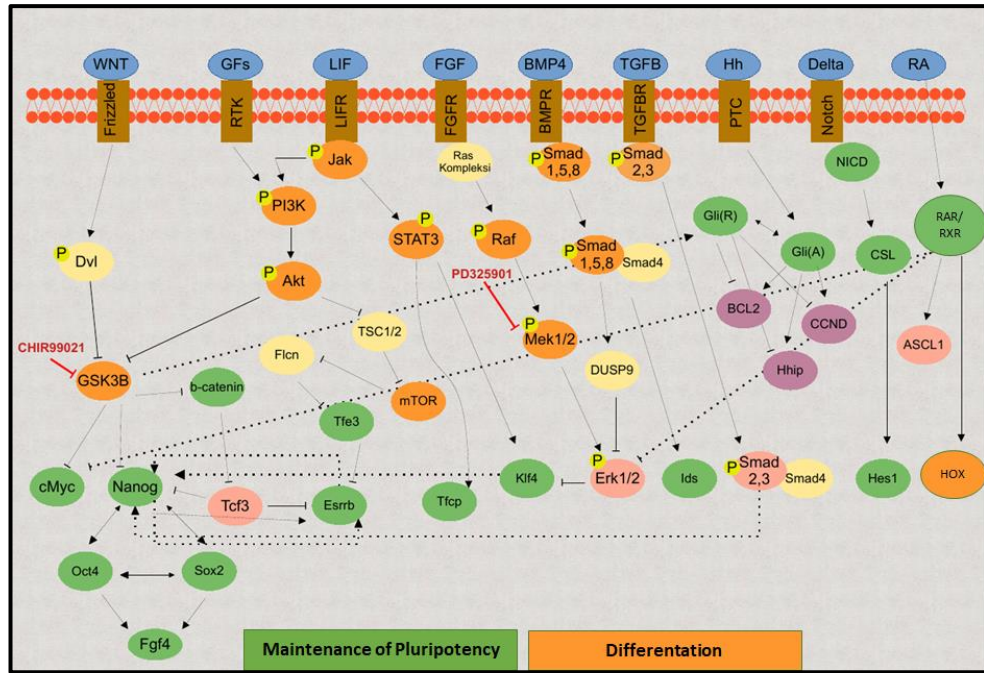


Figure 2.6: Signaling pathways in the pluripotency and cell differentiation

2.4. Quantitative Kinetic Modelling

Mathematical modeling is the representation of any system by using mathematical expressions. In the field of biology, these systems could be a group of chemical reactions, signaling pathways, tissue organizations, and even the whole organism. In systems biology, there are several types of mathematical modeling such as logical, constraints-based, agent-based, and quantitative kinetic modeling to investigate diverse biological processes or systems. The most common type of mathematical modeling in biology is quantitative kinetic modeling which is based on ordinary differential equation (ODE) representation of chemical reactions in the cell with chemical kinetics. In this kind of model, biological molecules are referred to as species (ie. metabolites, genes, proteins), the environmental space where chemical reactions take place is referred to as components (ie. cell, tissue). Reactions are constructed by the integration of these species and components with chemical kinetics referred to as the rate law. All these model structures are represented as ODEs to mimic the behavior of the system as a time course simulation.

There is a life cycle of construction quantitative kinetic modeling. As an initial step, the system boundary or structure should be determined (core biological system of the model such as cell, some signaling pathways, or tissues). After determining the

system boundary, mathematical relations, constraints, and assumptions are specified. In this step, reactions are constructed with chemical kinetics such as Mass Actions, Hill Functions, Michaelis-Menten Kinetics. Then, initial conditions (concentration, molecule level, etc.), parameters, and constants are retrieved from literature or experimental procedures. Integration of all these building blocks leads to a draft model, deterministic or stochastic time course simulations are performed to show the system behaviors as a next step. If there is a concern about the mathematical expression of the system or any kind of problems resulting from the time course simulations, all these steps should be optimized to construct mathematically corrected systems. After these steps, the draft model could be further optimized by performing parameter estimation algorithms for estimating condition-specific parameter sets with the integration of experimental data. As a final step, the optimized model is tested or validated by external experimental datasets [Le Novère, 2015].

Quantitative kinetic modeling of signaling pathways is one of the most common applications for studying the different biological processes such as cancer progression, cell cycle, stem cell differentiation, and somatic cell reprogramming *in silico*. For instance, MAPK and WNT signaling pathways have a critical role in cancer progression, and investigation of this pathway in cancers is crucial [Kim et al., 2007]. During embryonic development and tissue homeostasis, the TGF β signaling pathway plays an important role in cell specificity and tissue heterogeneity [Strasen et al., 2018]. In the stem cell differentiation and reprogramming processes, understanding of signaling pathways and their gene regulatory networks are considerable components to develop new therapeutic approaches [Pir & Le Novère, 2016], hence quantitative kinetic modeling of signaling pathways is often used for a mechanistic understanding of these cellular processes.

Although there are kinetic models of signaling pathways and their crosstalks available in the literature, it has not been performed on a large-scale quantitative kinetic modeling to investigate the pluripotency behavior of stem cells. In this thesis project, a large-scale quantitative kinetic model of 9 signaling pathways with their crosstalks was constructed to represent somatic cell reprogramming processes mathematically. Our main question is whether it is possible to co-examine the kinetic behavior of all major signaling pathways during the establishment of stem cell pluripotency together using both information and data-driven-based approaches.

3.MATERIAL AND METHODS

3.1.Materials

For constructing a quantitative kinetic model of signaling pathways regulating the reprogramming processes, literature information, transcriptomic dataset, and reference kinetic models were used, these resources were introduced in the following sections.

3.1.1. Literature resources and databases

For determining the biological boundaries of the kinetic model, a review article of the signaling pathways active during mouse pluripotency and differentiation processes was used as a starting point. From this review article [Hassani et al., 2013], information on molecular interactions in 6 signaling pathways were retrieved (WNT, Gfs, LIF, FGF, BMP4, and TGF β signaling pathways). Additional molecular interactions of 3 other signaling pathways, Hedgehog, Notch, and Retinoic Acid signaling pathways, were obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [M. Kanehisa & Goto, 2000; Minoru Kanehisa et al., 2021].

The chemical kinetics equations included in the model were retrieved from the reference systems biology book of Uri Alon such as Mass Action, Hill Equations, Michaelis-Menten Kinetics [Alon, 2006].

Previously reported helper quantitative kinetic models were retrieved from the BioModels database [Malik-Sheriff et al., 2020]. In particular, phosphorylation and dephosphorylation kinetics were obtained from the kinetic models of crosstalk WNT and ERK pathways with accession number BIOMD0000000149 [Kim et al., 2007].

3.1.2. Datasets

Integrating the transcriptomic profile of somatic cell reprogramming to iPSCs into the model, for determining the initial conditions and parameter estimation, a gene

expression matrix (TPM values) from the RNAseq study were retrieved from the Gene Expression Omnibus database with GSE137001 accession number [Barrett et al., 2013; Velychko et al., 2019]. To convert mRNA level to protein copy number, the reference conversion factors were provided from the transcriptomic and proteomic integrative study [Edfors et al., 2016].

3.2.Methods

To examine the behavior of the somatic cell reprogramming processes with a systems biology perspective, a quantitative kinetic model of 9 major signaling pathways was constructed with information and data-driven-based approaches. In this approach, different sources of information and datasets were integrated into the life cycle of the kinetic model construction. As a main model construction tool, COPASI (Complex Pathway Simulator) software was used [Hoops et al., 2006]. For data rearrangements and visualization, the R programming language was used in most of the cases.

3.2.1.Quantitative kinetic modeling of signaling pathways

The main characteristic of the model is integrative crosstalks of signaling pathways. In particular, each ligand molecule of specific signaling pathways interacts with its receptor molecule on the cell membrane. Initiation of ligand-receptor complex molecule activates the intracellular cell signaling cascade with diverse molecular kinetics such as activation, inhibition, or phosphorylation. At the end of the signaling pathway, the output molecule such as transcription factors, phospho-proteins, or other signal molecules decide the cell fate. The topological characteristics of the model were visualized in Figure 3.1.

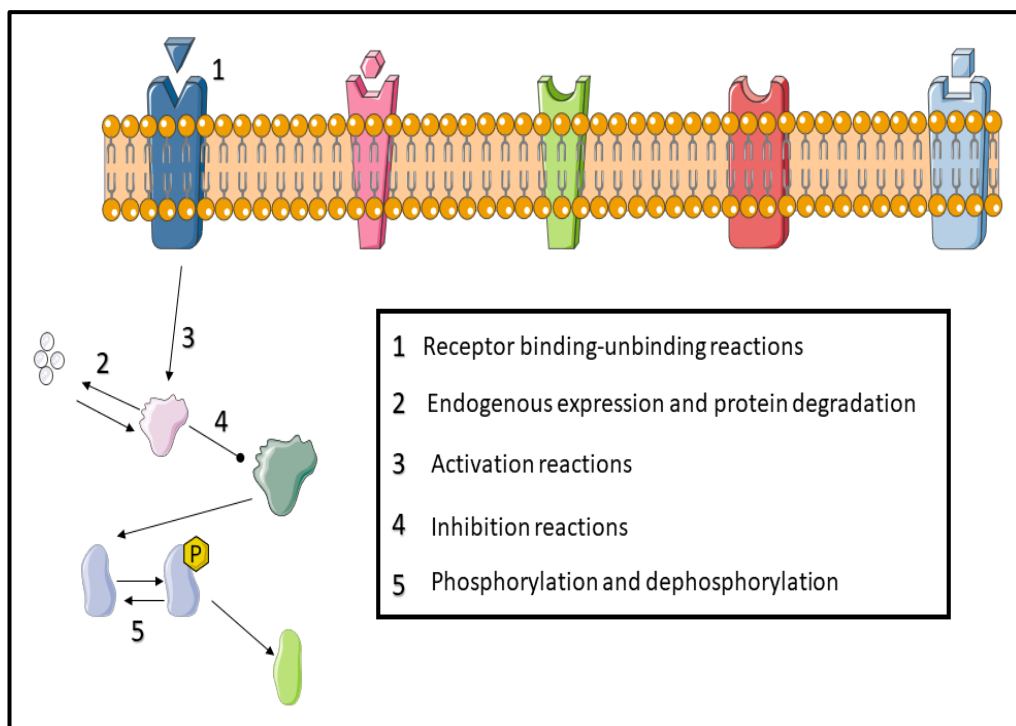


Figure 3.1: Topological schematics of the kinetic model and reaction types.

Based on the boundaries of the model and reaction types, the model is constructed using chemical kinetic equations, also known as rate laws. Each type of reaction was represented with its rate law kinetic using Mass actions, Hill equations, and Michaelis-Menten kinetics. Generalized rate laws were listed in Table 3.1, all reactions in the model were constructed using the rate laws as defined in the table. Ordinary differential equations of all species are listed in the Appendix.

The model is composed of 84 species representing the proteins and genes, and 144 reactions (in Appendix) in which these species are involved. The model time unit is minute (min), volume unit is the liter (l) and quantity units are nanomolar (nM) and copy numbers.

Table 3.1: Generalized rate laws of reaction types.

Reaction type & Example Reaction	Rate law	Origin
Receptor binding: WNT + Frizzled $\rightarrow$ Frizzled-Receptor active complex	$[Ligand] \times [Receptor] \times k$	Mass Action
Receptor unbinding: Frizzled-Receptor active complex $\rightarrow$ WNT + Frizzled	$[Active\ receptor\ complex] \times k$	Mass Action
Endogenous protein expression: $\emptyset \rightarrow Dvl$	k_{endo}	Constant Flux
Protein degradation: $Dvl \rightarrow \emptyset$ $pDvl \rightarrow \emptyset$	$k_{deg} \times [Protein]$	Mass Action
Activation by modifiers: $\rightarrow TfcP; pSTAT3(activator)$	$\frac{[Activator\ molecule]}{k_a + [Activator\ molecule]}$	Hill Function
Inhibition by modifiers: $\rightarrow GSK3\beta; pDvl(inhibitor)$	$\frac{k_i}{k_i + [Inhibitor\ molecule]}$	Hill Function
Phosphorylation: $Dvl \rightarrow pDvl$; Frizzled-Receptor active complex	$\frac{k_{cat} \times [Modifier] \times [Protein]}{k_m + [Protein]}$	Michaelis-Menten
Dephosphorylation: $pDvl \rightarrow Dvl$	$\frac{v_{max} \times [Phosphorylated\ protein]}{k_m + [Phosphorylated\ protein]}$	Michaelis-Menten

3.2.1.1. Defining initial conditions

After constructing the backbone of the kinetic model, the initial conditions were set after comparing different methodologies. Two types of approaches were tested to determine protein concentration/level from RNA sequencing (RNAseq) data. The first one is linear regression modeling which is aimed at predicting the protein concentration (nm/l) from TPM values and reference protein concentration of another kinetic model from Biomodels database. The second one is using RNA for the protein

conversion factor that indicates the multiplication of TPM values from RNAseq data with a constant to convert TPM values to protein copy numbers.

Within the scope of the linear regression modeling method, protein concentrations from the study [Kim et al., 2007] including some of our model species, also TPM values of these species from mouse fibroblast sample in RNAseq data (referred to as the day 0 in the study) [Velychko et al., 2019] were integrated to build a linear regression model. The rest of the initial concentrations of model species were determined by solving the equation resulting from this regression model. The linear regression model was constructed using functions in MS Excel.

In the conversion factor approach, 55 gene-specific pre-defined RNA to protein conversion factors resulting from the comparison of proteomic and transcriptomic datasets from diverse cell types and tissue were used as a reference set [Edfors et al., 2016]. However, most proteins of our model are not included in this reference list. To predict the conversion factor of our proteins, the correlation matrix was calculated from TPM values of the RNAseq dataset between these 55 reference genes (which have conversion factors from Edfors et al., 2016) and 117 genes in our model. After that, reference conversion factors were paired with our proteins by the biggest correlation value between our genes and reference set. All initial concentrations were calculated using individual reference conversion factors (available in Table 4.1).

In KEGG database, proteins were annotated in corresponding signaling pathways with multiple gene names and, therefore the estimated protein concentrations/levels resulting from both approaches were incorporated into the model by calculating means of genes of the RNAseq dataset for corresponding proteins into the model separately. For instance, there are 10 types of Frizzled genes in the RNAseq dataset, but Frizzled species were used in our model as single species. Therefore, to predict the initial protein level of Frizzled, the mean of expression of all Fizzled genes was calculated. Complete list of genes from RNAseq dataset were listed in Appendix.

3.2.1.2. Parameter estimation and time course simulations

After determining the model reactions and initial conditions, the 146 model parameters were estimated by converted protein levels from MEF (mouse embryonic fibroblast) to the iPSC reprogramming process using a time course RNAseq dataset. Before performing parameter estimation, several assumptions were made to overcome technical limitations and obtain much more predictive results. To mimic MEF-iPSC reprogramming of experimental procedures of transcriptomic data [Velychko et al., 2019], the level of LIF molecule was kept constant as LIF was continuously available to the cells in the growth medium. In experimental procedures of RNAseq data, OSKM factors were induced with a common expression vector in cell culture and activated by the doxycycline, (dox)-inducible operator molecule (Figure 3.2). To represent the over-expression of OSKM factor with doxycycline inducible vector in the experimental procedures in the model, a constant parameter “z” were integrated into rate law of OSKM factor production as a multiplication factor (28,29,70 and 87th reactions in Appendix) to mimic the induction effect over the endogenous OSKM expression. Then z parameter was set at a value larger than 1 from day 0 to day 10, when the z is set to 1 (ineffective element in multiplication), induction is absent. Using this adjustment, z parameters were constructed to mimic dox induction as same as experimental procedures.

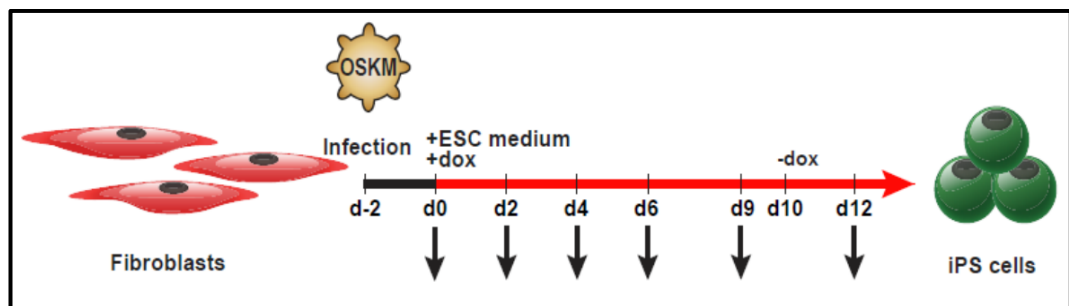


Figure 3.2: Experimental procedures of reprogramming from MEF to iPSC by Velychko et al. 2019.

With all these technical assumptions, parameter estimation was performed by using “Genetic Algorithm SR” pre-defined in COPASI Software with default parameter setup (number of generations = 200, population size = 20, Pf = 0.475, and

the starting value of parameters were randomized) and custom lower and upper bound for estimating parameter sets. Lower and upper bounds were listed in Table 3.2.

Table 3.2: Lower and upper limits in parameters estimation.

Parameter	Lower Limit	Upper Limit
k_{endo} (protein copy number/min*L)	1	1e+6
k_{deg} (1/min)	0.001	1
k_a (protein copy number /L)	1e-6	1e+6
k_i (protein copy number/L)	1e-6	1e+6
K_m (phosphorylation) (protein copy number/L)	1e-6	1e+6
K_{cat} (phosphorylation) (1/min)	1e-6	1e+6
K_m (dephosphorylation) (protein copy number/L)	1e-6	1e+6
V_{max} (dephosphorylation) (protein copy number/min)	1e-6	1e+6
Z (dox induction factor, unitless)	1	1e+6

Total of 146 parameters for all reactions and events were estimated and model parameters were updated with the estimation results. k_{deg} and k_{endo} parameter of OSKM factor and K_m and K_{cat} parameters of all phosphorylation and dephosphorylation reactions were determined globally. All other parameters were determined specifically for each reaction. Then, model behavior was tested by time-course simulations with Deterministic simulation (LSODA) pre-defined in COPASI Software for 12 days long cultivation (from day 0 to day 10 referred to as doxycycline induction, day 10 to day 12 referred as iPSC formation). Additionally, the model behavior was investigated with Stochastic simulation (Adaptive SSA/T-Leap). Time course simulation results were imported into the R programming language for visual representation, normalization, and comparisons.

To check the sensitivity of parameter estimation results, parameter estimation algorithm was performed 20 times with the same settings. All parameter sets were compared to each other for each parameter.

3.2.1.3.Determining pluripotency and differentiation levels

To determine the pluripotency and differentiation levels from day 0 to day 12 in time-course simulation results, firstly the mean of pluripotency and differentiation-related genes/proteins shown in Figure 2.6 were calculated for each time points. Then mean values of pluripotency and differentiation were pooled and normalized to the range [0-1] and multiplied with 100. This normalization provides that the highest value of time course results indicates 100% and the lowest value indicates 0%. The normalization equation was shown in below: x' indicates normalized value of pluripotency and differentiation for individual time point.

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \times 100 \quad (2.1)$$

3.2.1.4.*In silico* Experiment for doxycycline induction

Generally, expression vectors cannot transfect all cells in the experimental procedures. After performing vector transfection in an experiment, only a the proportion of the cells are transfected with the expression vector. This situation is expressed as a percentage called “transfection efficiency”. To mimic this phenomenon in the kinetic model, *in silico* doxycycline induction experiments were constructed, to test if the model can predict that only transfected cells with OSKM vectors can reprogramme to iPSC. In our model z parameter indicates the induction level and affects the expression change of OSKM factors. Therefore, a hypothetical cell culture environment was planned: 200 MEF cells were seeded into cell culture, and only 50% were transfected with doxycycline inducible OSKM vector. After parameter estimations, all parameters were determined for MEF to iPSC transition. To mimic behavior of multiple cell in the same time simulations were run 200 time (each simulation run indicates one cell) with 2 different distribution of z parameters: 100 simulations were run with z parameters equal to 1 (28,29,70 and 87th reactions in Appendix) and 100 simulations were run with z parameters equals to estimated value (15.465). Then average protein levels of pluripotency-related proteins were investigated. For this *in silico* experiment, the Parameter Scan task was used in COPASI.

4.RESULTS

4.1.Estimating protein levels from mRNA level

Determining the protein level using TPM values from the RNAseq study, a linear regression model was constructed with reference protein concentration (nmol/L) on the y-axis and corresponding TPM values on the x-axis. According to the linear regression model, reference protein concentrations and their TPM values were not fitted very well (Figure 4.1).

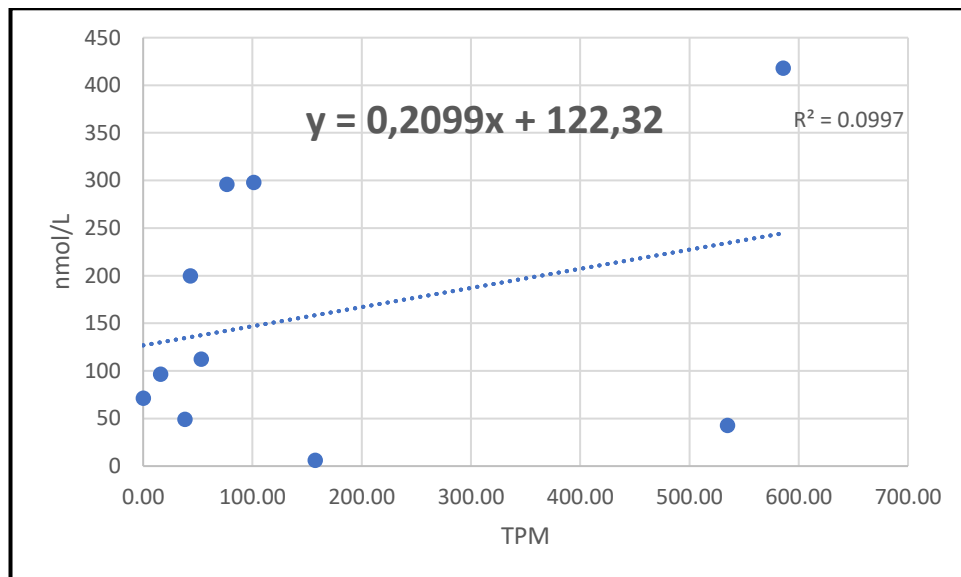


Figure 4.1:Linear regression model of protein concentration and mRNA level.

As a second approach, protein conversion factors were determined with the correlation of expression changes during MEF reprogramming to iPSC of the model species (Figure 4.2). Maximum correlation between reference protein and model proteins was selected to determine the corresponding conversion factor for model species (Table 4.1).

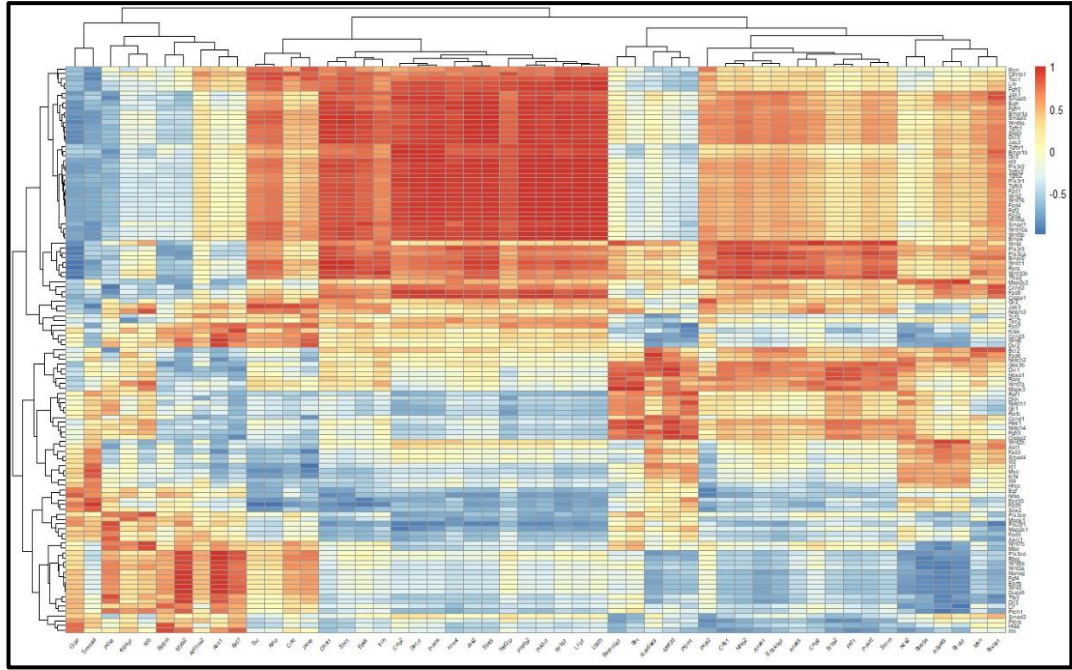


Figure 4.2:Correlation matrix of reference conversion factor proteins with model species using expression values from MEF to iPSC.

Table 4.1:Reference conversion factor of model species

Model Genes	Reference Genes	Correlation	RNA to PROTEIN (RTP) Conversion Factors for Reference Genes
Fzd2	Phldb2	0.9989668	1040
Fzd1	Phldb2	0.99817248	1040
Fzd7	Pak1	0.84037343	16140
Fzd5	Pdk1	0.71157414	30810
Fzd10	Timm44	0.84897757	4660
Fzd4	Lcp1	0.9985504	19020
Fzd6	Rps6ka3	0.96093749	23000
Fzd8	Decr1	0.95962389	53290
Fzd3	Agpat3	0.74500569	1580
Fzd9	Timm44	0.97583762	4660
Wnt5a	Phldb2	0.99909663	1040
Wnt7a	Sfn	0.90500232	71960
Wnt4	Capg	0.98096322	22440
Wnt9a	Stat3	0.98996325	35460
Wnt2b	Agpat3	0.95947456	1580
Wnt7b	Mb	0.93502865	20070
Wnt2	Phldb2	0.99594167	1040
Wnt8a	Ass1	0.95110598	31570
Wnt16	Lcp1	0.99556138	19020
Wnt5b	Rrbp1	0.99630934	14300
Wnt11	Cant1	0.98491087	1240
Wnt10b	Itch	0.97061558	11170
Wnt6	Ass1	0.96711731	31570
Wnt3	Mybl2	0.97694018	260
Wnt10a	Rrbp1	0.99345137	14300
Wnt3a	Ass1	0.92557331	31570
Dvl2	Ass1	0.92183447	31570
Dvl1	Mef2d	0.96686012	11500
Dvl3	Jak2	0.9968019	2850

Table 4.1:Continued.

Model Genes	Reference Genes	Correlation	RNA to PROTEIN (RTP) Conversion Factors for Reference Genes
Gsk3b	Prkcd	0.85659185	29840
Myc	Timm44	0.89972371	4660
Nanog	Mybl2	0.97686158	260
Ctnnb1	Xiap	0.9117278	7040
Tcf3	Pak1	0.76031412	16140
Egf	Clpp	0.72648224	5480
Egfr	Stat3	0.99124508	35460
Pik3ca	Nfkb2	0.97010927	8040
Pik3cb	Mb	0.8843258	20070
Pik3cd	Ass1	0.99474848	31570
Pik3r1	Pik3r1	1	12520
Pik3r2	Phldb2	0.99266745	1040
Pik3r3	Hnmt	0.90749082	7950
Tsc1	Rrbp1	0.93958694	14300
Tsc2	Hnmt	0.68947643	7950
Mtor	Mb	0.90638691	20070
Flcn	Pak1	0.865243	16140
Tfe3	Mybl2	0.90221922	260
Esrrb	Mybl2	0.97170285	260
Lif	Bpgm	0.81645231	11840
Lifr	Ugdh	0.98043103	14130
Jak1	Jak2	0.96659206	2850
Jak2	Jak2	1	2850
Jak3	Src	0.92639655	19270
Stat3	Stat3	1	35460
Tfcp2	Nfkb2	0.97516447	8040
Klf4	Timm44	0.85079311	4660
Fgf2	Lcp1	0.99822302	19020
Fgfr1	Hnmt	0.97979375	7950

Using corresponding conversion factors, protein levels were calculated from TPM value as protein copy numbers. All converted and adjusted protein levels were listed in Appendix. These results suggest that the conversion factor estimation for protein level is much more suitable than the linear regression model. Therefore, all other analyses were performed using the converted protein levels resulting from the conversion factor approach.

4.2.Parameter Estimation

After constructing the kinetic model with converted protein levels, model parameters were estimated using parameter estimation algorithms. To evaluate the feasible range of each parameter, estimation algorithms were performed 20 times and all parameter sets were mapped into a heatmap (Figure 4.3, sets 1-20). Additionally, mean and standard deviation values of parameters were listed in Appendix.

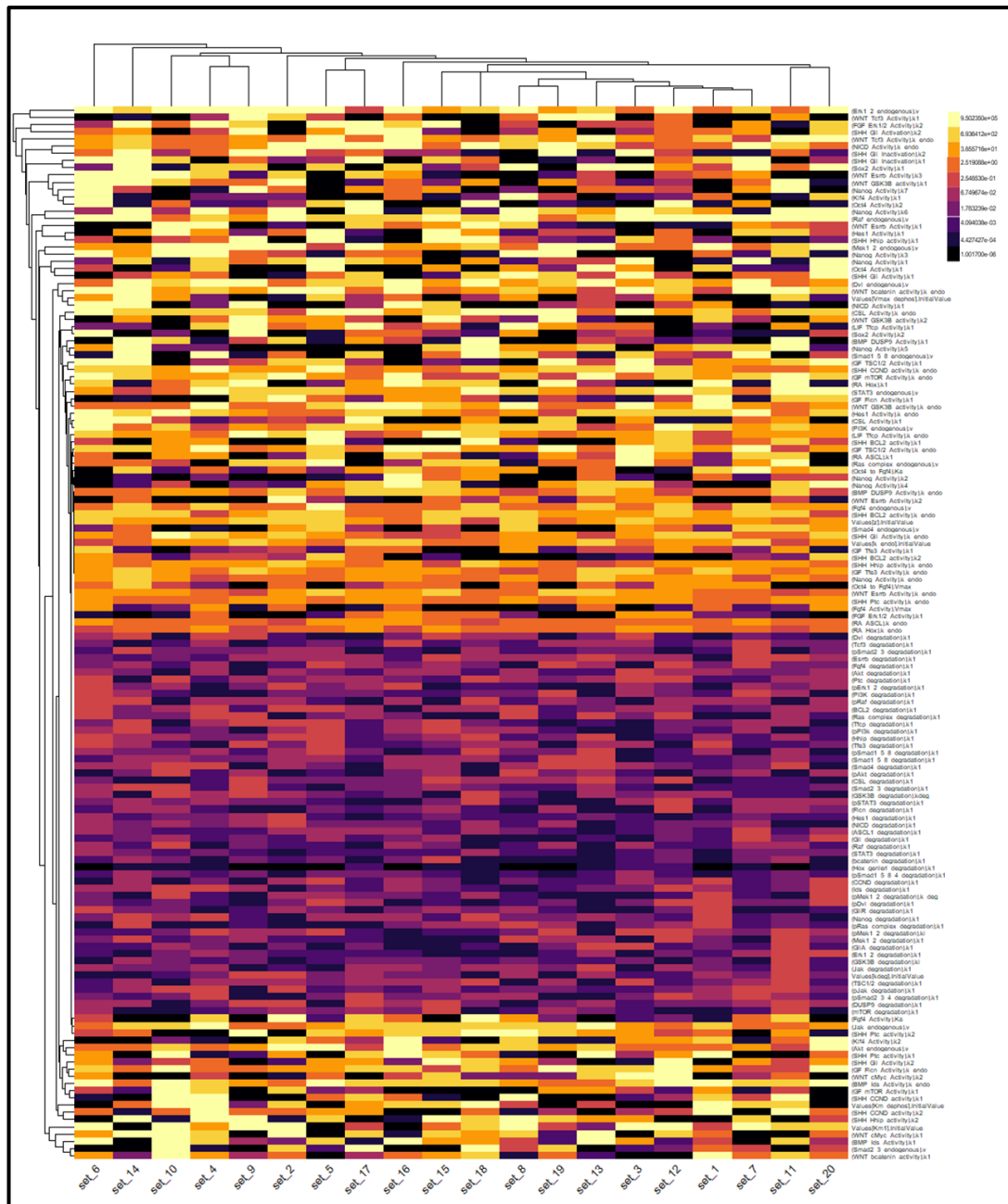


Figure 4.3: Comparisons of estimating parameter sets. Rows indicate the model parameters and columns indicate the different estimation sets 1-20. Each estimation sets refer to individual estimation run with same parameters in genetic algorithms.

Parameter estimation results suggest that protein degradation parameters were estimated much more expectedly homogenously between different estimation runs due to narrow range of upper and lower bounds of the degradation parameters. Mathematical stabilization of the model was conserved with degradation parameters in each run of parameter estimations. However, other parameters such as endogenous expression, activity, and inhibition parameters show heterogeneity between different

estimations due to complex behavior of the kinetic model. On the other hand, diverse range of activation and inhibition parameters could be the cause of the dynamic changes of protein levels for each run. For instance, Nanog is targeted by multiple proteins in the model. When the variation of parameters of Nanog Activity reaction (reaction 93 in Appendix) was investigated, all activation parameters have higher standard deviation. According to estimation results, when one or two activator parameters of Nanog was predicted to be higher, the other parameters were predicted to be lower level (*i.e.*, when k_1 of Klf4 activator is higher, k_2 of Esrrb activator parameter were estimated at a lower value). To stabilize the Nanog production-degradation rates for better fit to experimental data, activity or inhibitory parameters (K_a and K_i) have higher variation across parameter estimation runs. These variations lead to the heterogeneity in estimated values of most of the parameters, more data from other experimental conditions should be used in parameter estimation to be able to achieve more robust parameter ranges, nevertheless, the model behavior was conserved for each estimation set in time course simulations, indicating that same phenotype can be produced by the pluripotency network in a wide range of activity levels. Therefore to simulate the model behavior, one parameter set was selected randomly, and time course simulations were performed.

4.3. Time Course Simulations

For investigating the MEF to iPSC transition dynamics in the kinetic model, deterministic and stochastic time course simulations were performed individually. In the deterministic simulation, firstly, protein levels of OSKM factors were checked (Figure 4.4). According to time-course simulation, protein copy numbers of OSKM factors were increased from day 0 to day 10 (14400 mins). When the doxycycline induction was ended after day 10, protein levels of OSKM factors were decreased.

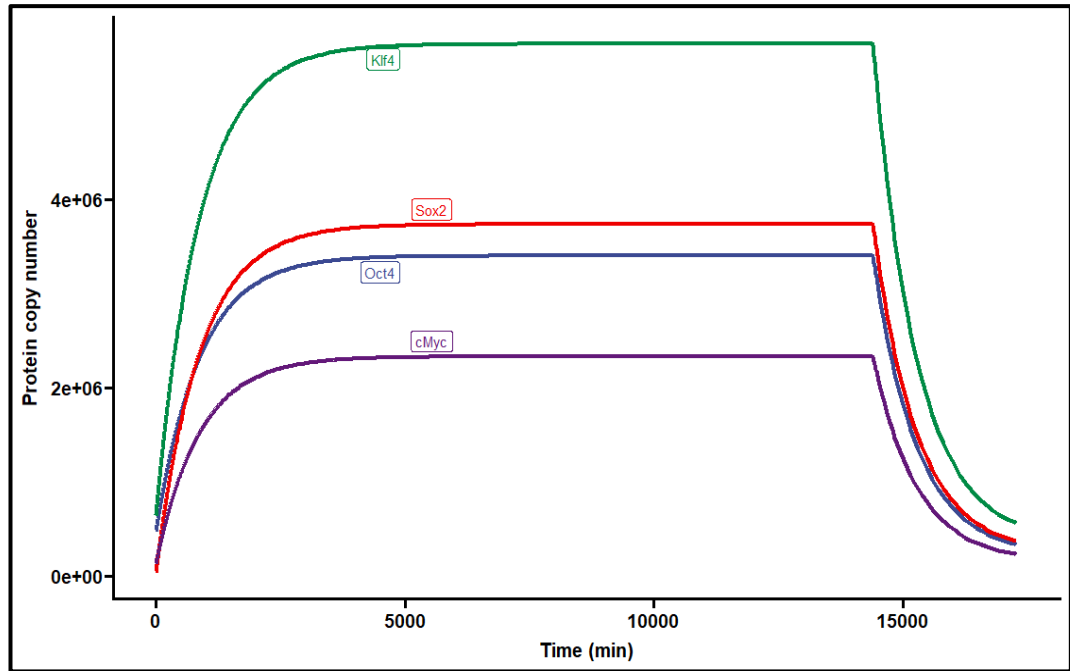


Figure 4.4: Time course simulation of OSKM factor from MEF to iPSC.

Relative fold change of pluripotency and differentiation-related proteins from MEF to iPSC are investigated. In order to do this, from the simulation result, the 12. Day (corresponding to iPSC) data was taken and divided by 0. Day (corresponding to MEF) data and then log2 fold change was calculated. The most pluripotency-associated proteins were up-regulated except for β -catenin and Tfcp. On the other hand, only phospho-Smad 2-3 and Hox proteins were up-regulated in differentiation-related proteins. These results suggest that pluripotency was activated through MEF reprogramming processes (Figure 4.5-4.6).

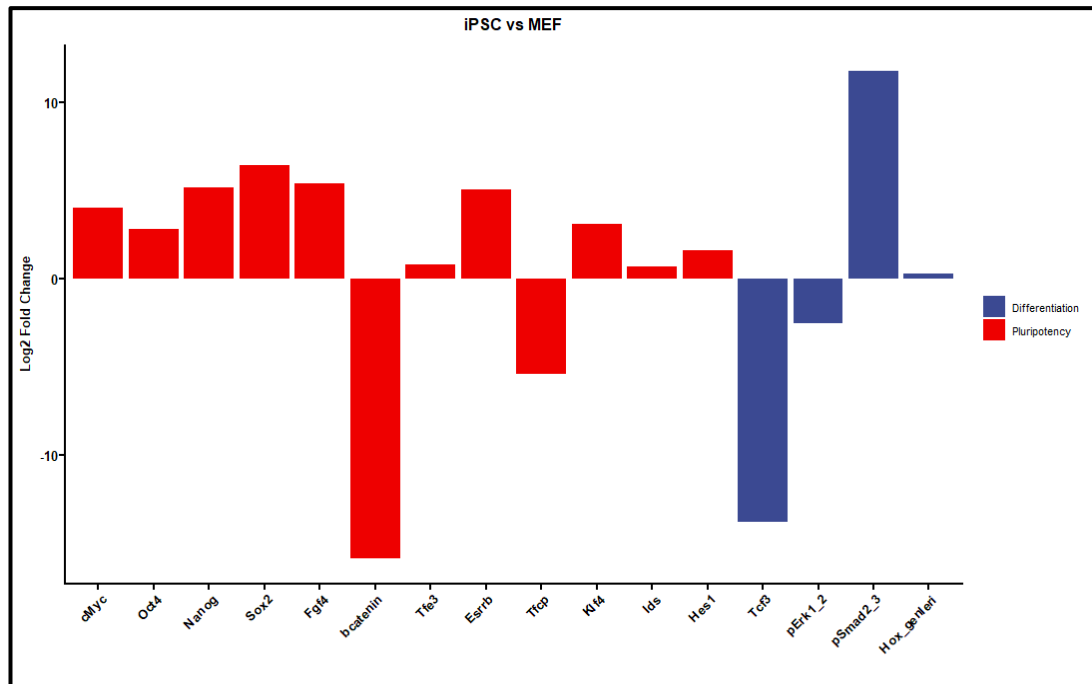


Figure 4.5:Relative fold change of pluripotency- (red) and differentiation- (blue) related proteins.

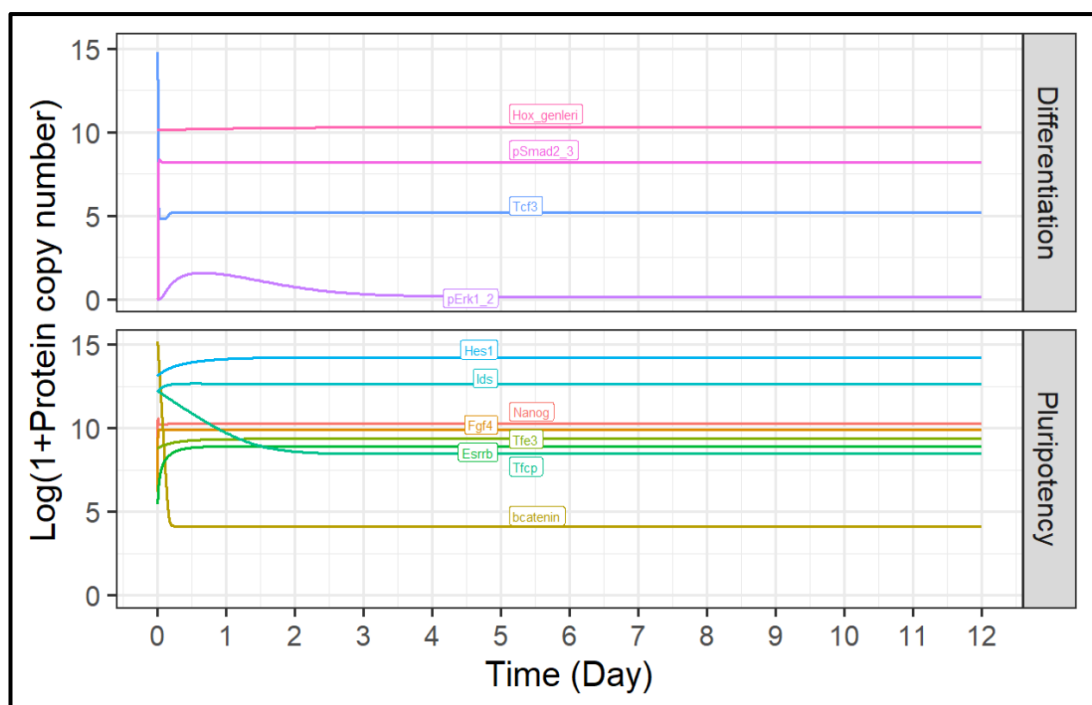


Figure 4.6:Time-scale comparison of pluripotency and differentiation-related proteins. Proteins copy numbers were log-scaled.

Focusing on the individual signaling pathways, relative fold changes of the proteins were observed consistent with mouse pluripotency behavior represented in Figure 2.6 (Figure 4.7).

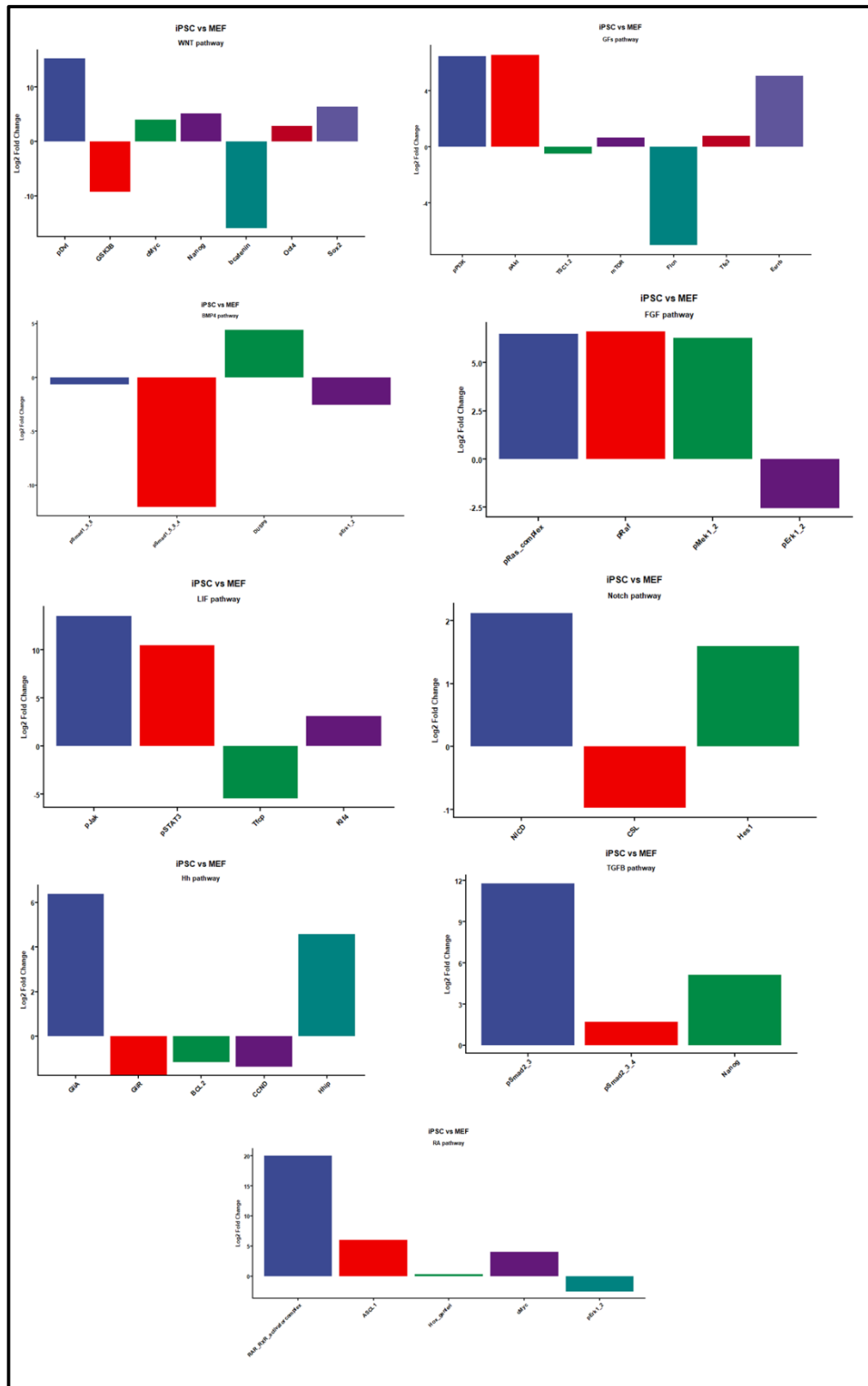


Figure 4.7: Relative fold changes proteins of individual signaling pathways during MEF reprogramming. Relative fold change was calculated by log2 of the Day 12 / Day 0

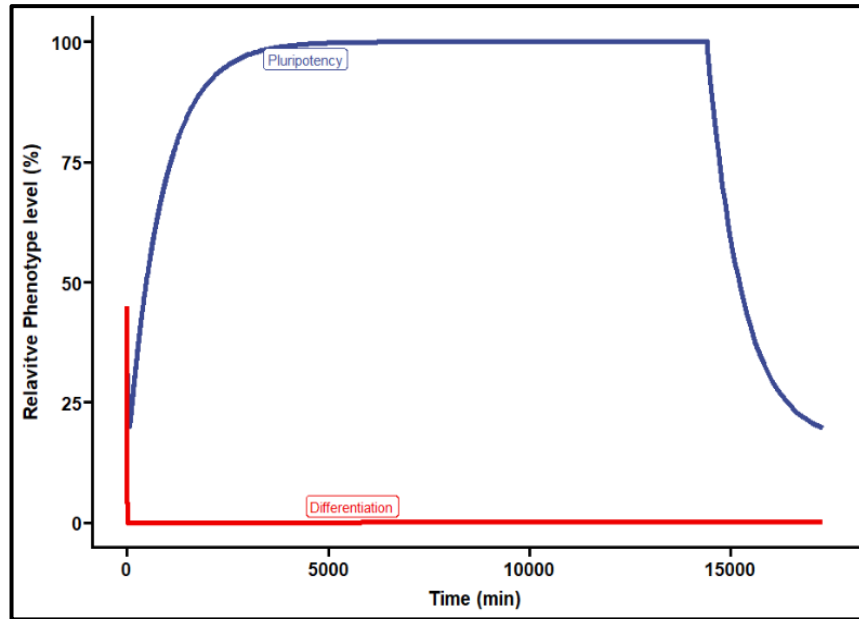


Figure 4.9: Relative pluripotency and differentiation level during deterministic time course simulation. The relative level of pluripotency and differentiation were calculated with normalization of the average protein copy number for the corresponding stage (pluripotency or differentiation).

In stochastic simulations, protein levels of OSKM factors were investigated taking the stochasticity of gene expression into consideration. This simulation result suggests that the iPSC transition could be observed with OSKM activation, however, any effect of doxycycline withdrawal was not observed after day 10 (14400 min) (Figure 4.10).

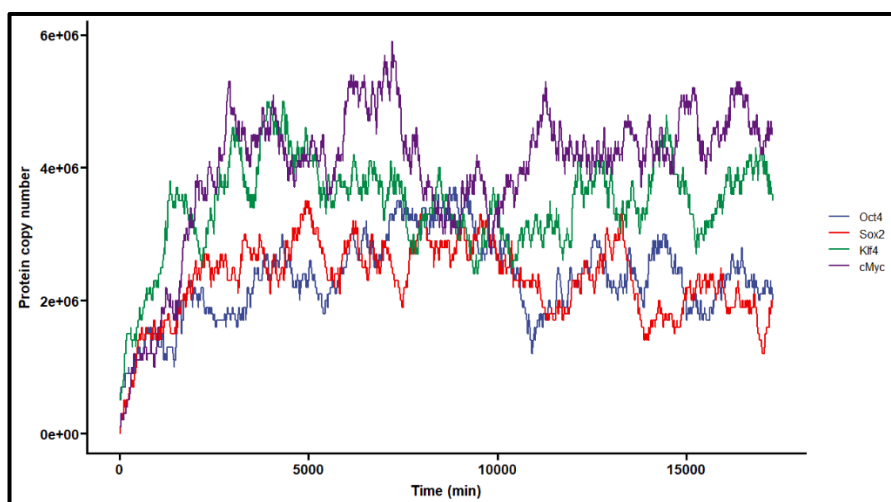


Figure 4.10: Stochastic simulation of OSKM factors during reprogramming.

4.4. *In Silico* Experiment of Doxycycline Induction

To mimic the transfection efficiency and doxycycline induction in the kinetic model, stochastic time course simulation was performed with 200 hypothetical cells by 50% transfection efficiency (50% of cells have $z = 15.465$ estimated parameter value, others 50% of cells have $z = 1$). As a result, 50% of cells which correspond to transfected cells were reprogrammed to iPSC indicating the high level of OSKM factors and other pluripotency-related proteins. Oppositely, the average pluripotency level of not transfected cells was under the level of initial protein level indicating the cells were staged into the MEF stage (Figure 4.11).

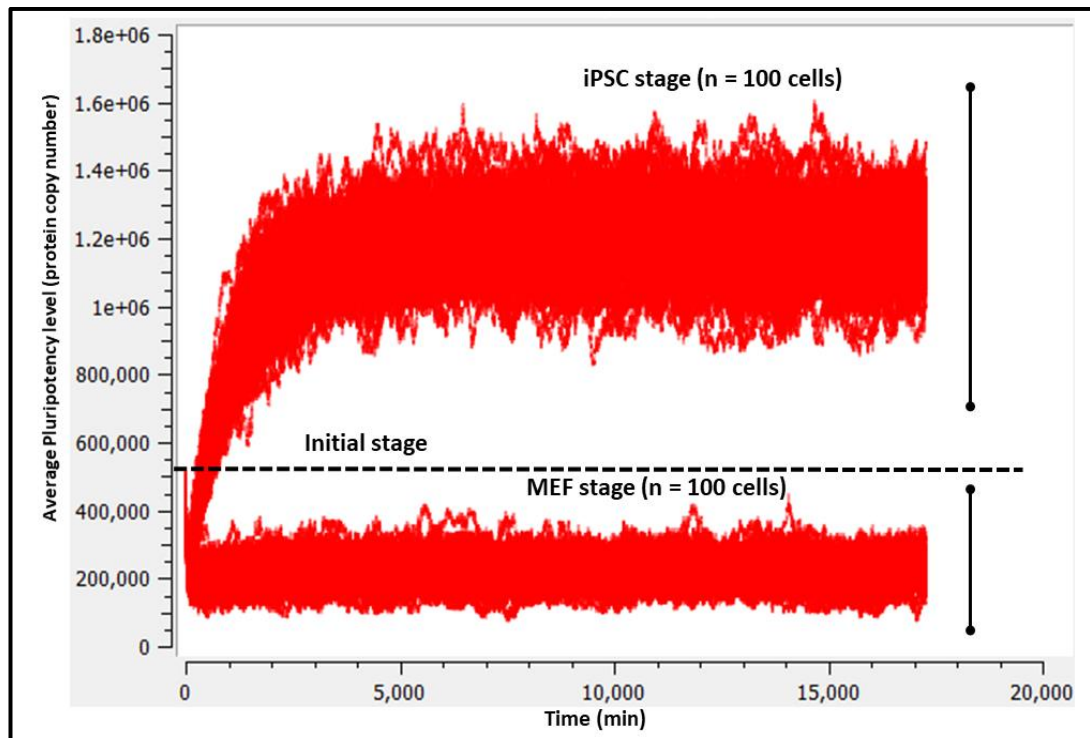


Figure 4.11: *in silico* experiment of transfection efficiency and doxycycline induction. The initial stage indicates the average protein level of pluripotency-related species in the model. Above the initial stage protein level was determined as iPSC stage, below was determined as MEF stage.

Additionally, different doxycycline induction levels were checked by changing the z parameter from 1 to 9. When the z parameter rises, cells can enter the pluripotency stage hypothetically due to the increased effect of doxycycline induction (Figure 4.12).

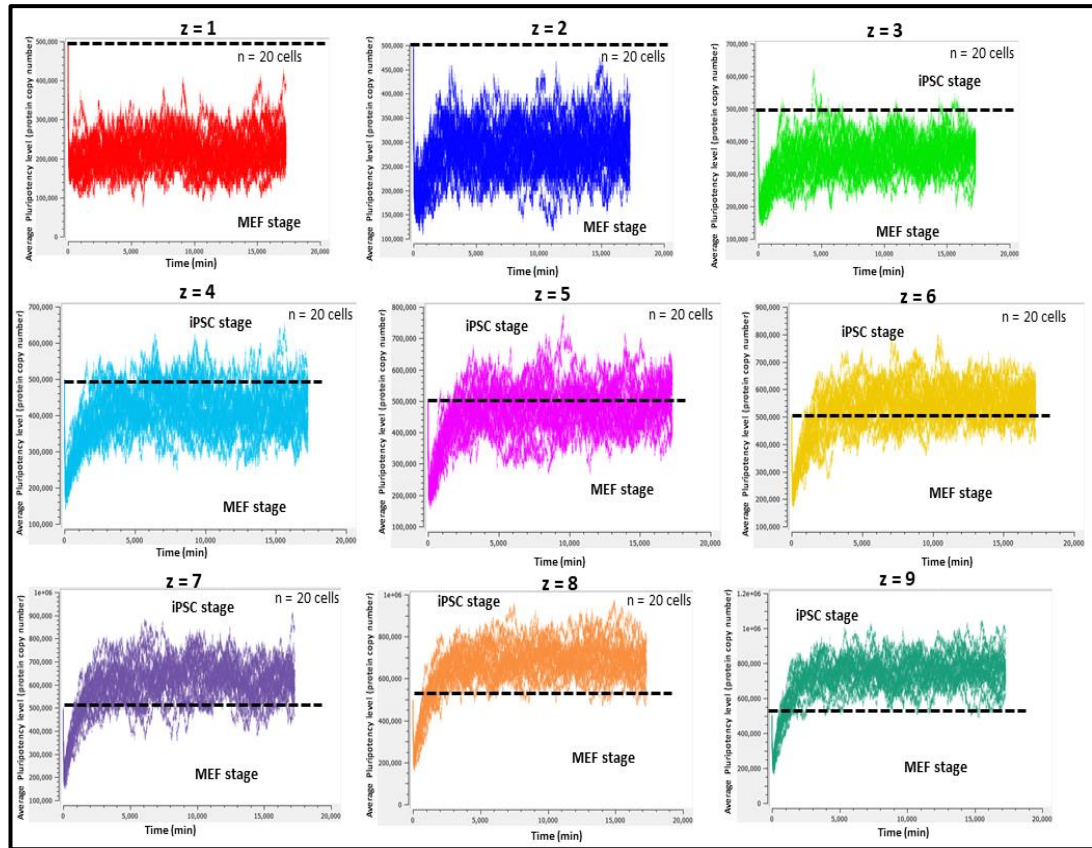


Figure 4.12: Different doxycycline induction levels by parameter scan of z parameters effect on pluripotency level

4.5. The Effect of Small-molecule Inhibitors on Pluripotency

Another *in silico* experiment is about the effect on pluripotency of small molecule inhibitors CHIR99021 and PD0325901 (together known as 2i) which plays role in maintaining pluripotency by inhibition of GSK3B on the WNT pathway and phospho Mek 1/2 on the FGF pathways, respectively. To investigate the effect of these inhibitors on pluripotency, CHIR99021 and PD0325901 were introduced to the system on day 10 right after the doxycycline withdrawal. According to stochastic simulation results, there is no effect of these 2 inhibitors on the pluripotency level (Figure 4.13). Because after the 10th day, GSK3B and phospho Mek 1/2 protein levels were observed to be nearly zero. Therefore, although the inhibition reactions of small molecules worked mathematically, their effect on the protein levels of pluripotency were not observed in simulation results.

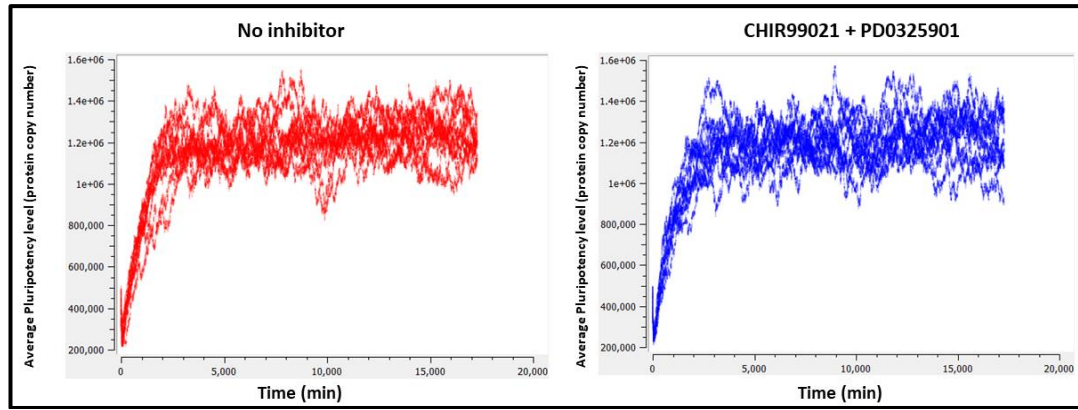


Figure 4.13: Small molecule inhibitors' effect on pluripotency level.

4.6. Single and Combined Effects of OSKM Factors on Pluripotency

In order to check the induction effect on pluripotency with gene deletions of OSKM factors individually and in a combination of them, all possible combinations were analyzed in COPASI. To be able to mimic the gene deletion in the *in silico* model and check the effect of pluripotency, each individual and combined version of OSKM factors were fixed to zero one by one. For example, to check the Oct4 gene effect in pluripotency level and transition to the iPSCs stage, other Yamanaka's factor genes which are Sox2, Klf4, and c-Myc were fixed to zero which means that these genes were deleted in the model (Figure 4.14). According to experimental results in Velychko dataset, single factors of OSKM were not enough to iPSC transition. They also showed that dual induction of Sox2 and Klf4 were not enough too. However, in dual induction of Sox2 and cMyc, SKM and OSKM inductions, cells have passed into iPSC stage. Therefore, transition threshold was determined (around 600.000-800.000 average protein copy number) to be higher than SM induction effect and lower the SKM and OSKM induction level as experimental reference point.

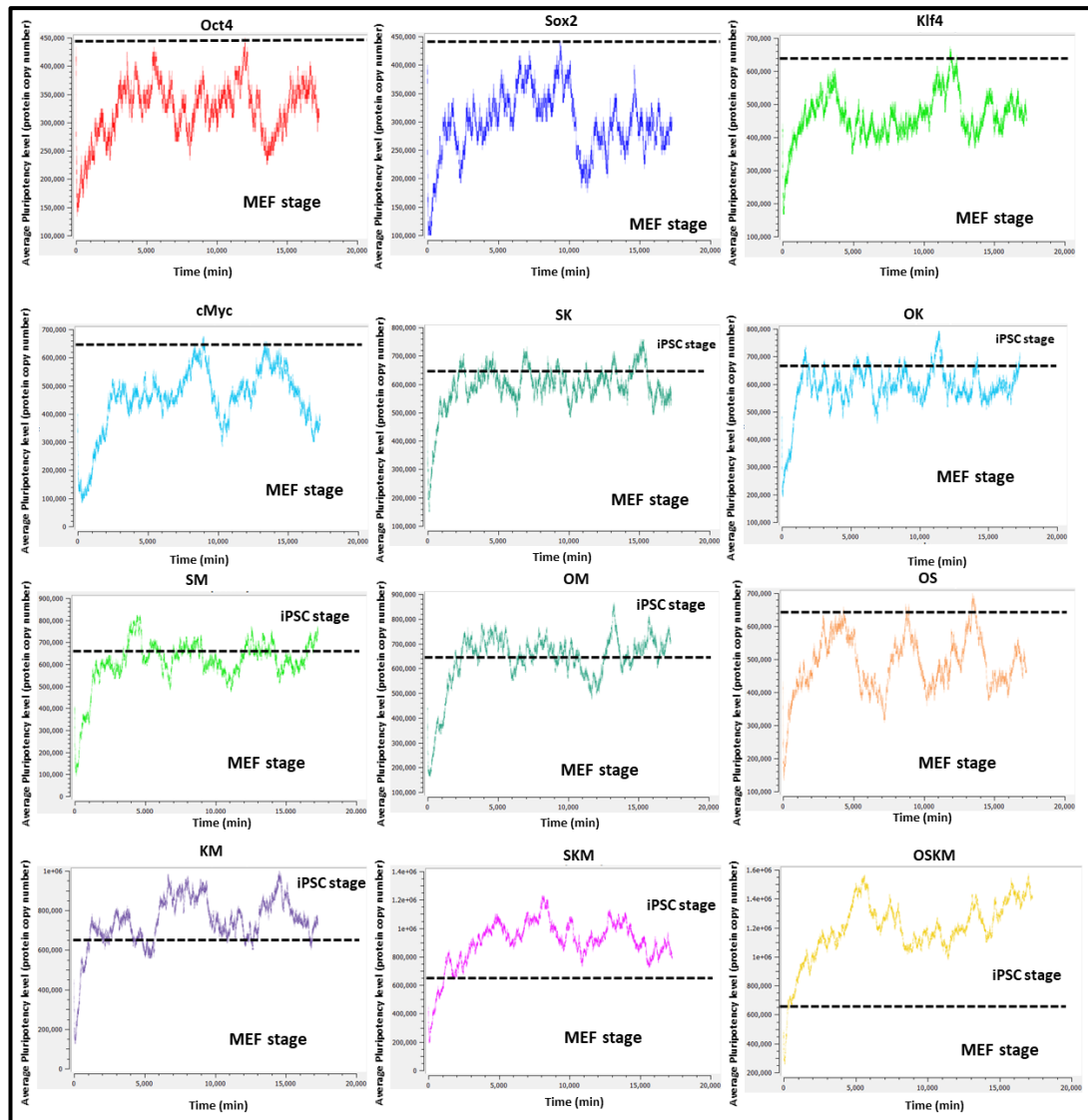


Figure 4.14: Single and combinatory effect of OSKM factor induction on pluripotency.

5.DISCUSSION AND CONCLUSION

Quantitative kinetic modeling is an emerging research area in systems biology. Mathematical modeling of biological systems provides researchers the cost-efficient predictions for understanding biological processes and allows the creation of new hypotheses. Signaling pathways are one of the major regulatory elements of the biological network in different aspects such as cancer progression, cell differentiation, and stem cell pluripotency. Understanding of complicated processes of somatic cell reprogramming and stem cell pluripotency is crucial to develop new therapeutic strategies. When the quantitative kinetic modeling literature was investigated, there is no integrative, multi-signaling pathway modeling. Although some of the models were only focused on 1 or 2 signaling pathways and their cross-talks, there is a still gap for an inclusive kinetic model. In this study, we tried to construct inclusive, integrative quantitative kinetic modeling of all signaling pathways associated with somatic cell reprogramming using different systems biology approaches.

5.1. The Model Topology

Within the framework of the constructed kinetic model, we focused on major cascades of the signaling pathways and affected proteins during pluripotency and differentiation. Although kinetic models are limited with increased complexity as the model size is increased, molecular interactions of the 9 signaling pathways were modeled with remarkable predictions. In accordance with the life cycle of the modeling, different approaches were investigated to construct an optimal quantitative kinetic model.

For determining the initial values of the species, 2 different approaches were tested. Linear regression modeling did not provide good predictions to determine protein levels from mRNA level due to limitations of the reference protein concentrations and lack of protein concentration measurements in literature (Figure 4.1). To overcome this limitation, more protein concentrations should be measured as nmol/L level and/or found from literature. However, the protein conversion factor methodology has been identified as a good alternative for determining protein levels (Figure 4.2 and Table 4.1). Due to correlation analysis between reference conversion

factor genes (Table 4.1) and the counterpart of genes in the RNAseq study, conversion factors of genes in the model which are not on the reference list were determined. Since there was no factor for the MEF or iPSC cell types in the reference dataset, the median values of the factors calculated for 20 different cells and tissues were used in the analysis. When the calculated initial protein copy number of the model proteins were compared with literature, the distributions of protein copy numbers were observed to be consistent with a study in which protein copy numbers were calculated in fibroblast cells for a large set of proteins [Schwanhäusser et al., 2011]. As a result of this observation, it is concluded that the conversion factor algorithm was more useful and realistic than linear regression models to calculate protein level using RNAseq TPM values.

5.2. Parameter Estimation and Mathematical Stability

After 20 parameter estimation sets trials, the heterogeneity of some parameter sets raises different questions (Figure 4.3 and Appendix). The lower and upper boundaries of some parameters were set in a narrow range to mathematically stabilize the kinetic model. For instance, when the k_{deg} parameters are too low or high, the expression-degradation balance of the proteins in the model is seriously disturbed. The standard deviations of the estimated parameters were found to be dependent on reaction type. Because of the narrow upper and lower bounds of the protein endogenous expression and degradation parameters, these parameters had the lowest level of standard deviation. K_{deg} (degradation parameter) could be counted as a limiting parameter for this model. On the other hand, activation and inhibition parameters were observed at the highest level of standard deviation as seen in Appendix, due to the complexity of signaling pathway reactions and the wide range of lower and upper bound of these reaction parameters (Figure 4.3). For instance, *Essrb* in WNT signaling pathway has 2 activators and inhibitors. In this kind of multi interaction species of the model, each estimation run finds a wide range of parameters sets for individual activator or inhibitor parameters. Therefore, parameters of activator and inhibitor reactions show high plasticity.

For a much more precise parameter estimation of the model, the different lower and upper bound combinations should be tested. On the other hand, each estimation

runs were performed with same default settings., Initial parameter values were randomized within the range of their upper and lower limits prior to parameter estimation. Therefore, each run provides the diverse parameter estimation results, parameter estimation was performed 20 times to investigate plasticity of the parameter values. The effect of other settings (such as generation size, generation number) in genetic algorithm were not tested. As a future aspect, this optimization will also be performed.

5.3. Signaling Crosstalk in Pluripotency

When the molecular functioning of the signaling pathways is examined with deterministic simulations, it has been observed that the signaling pathways work in a mathematically consistent manner in iPSC reprogramming individually and together (Figure 4.4-7). In the comparison of relative fold change between the RNAseq study and model time-course simulations, most of the proteins were observed to have consistent protein level changes during iPSC reprogramming. These results suggest that our model accurately mimics the experimental procedures under the tested conditions (Figure 4.8). When the deterministic simulations results are compared with the literature, the dynamic behavior of the model proteins was observed to be consistent with molecular studies. For instance, Stat3, phospho PI3K, and phospho Akt were studied in triggering pluripotency of mESCs by activation of the LIF/Stat3 signaling pathway [Matsuda et al., 1999, p. 3; Niwa et al., 2009]. In our simulations, these proteins were indeed up-regulated during iPSCs formation. Another main protein of mouse pluripotency is Nanog which plays a role in the core transcriptional regulation of pluripotency [Chen et al., 2008] which was also observed to be up-regulated in our model. On the BMP4 signaling pathway inhibition of Erk by DUSP9 was identified with the molecular study to sustain pluripotency [Li et al., 2012, p. 4]. Esrrb was also identified as a novel regulatory element of pluripotency in mESC [Storm et al., 2009]. On the other hand, chemically suppression of mTOR was used to exit mESC from its pluripotent stage. Therefore it is suggested that the high-level expression of mTOR maintains the pluripotency stage in mESCs [Welham et al., 2007]. TSC1/2, known as a negative regulator of mTOR, plays a role in the exit from the pluripotency stage of mESC which was also observed as down-regulated in our

model [Tsuneyoshi & Dunn, 2013]. Interestingly, although bcatenin is responsible for the differentiation of mouse embryonic stem cells, a few different studies show that bcatenin deficiency is also involved in the self-renewal of embryonic stem cells with LIF cultivation [Lyashenko et al., 2011; Wray et al., 2011]. In our model bcatenin was observed to show the latter behavior with LIF induction.

All these molecular studies show that most of the model dynamics in deterministic simulations were consistent with literature information. However, changes in protein levels in iPSC versus MEF stages were not consistent between model simulations and experimental dataset for some of the proteins. Because too many parameters were tried to be estimated and some proteins have very complex crosstalk patterns with other proteins in signaling pathways, several protein levels did not agree well with experimental data in parameter estimation. To overcome this inconsistency, parameter estimation will be performed with different experimental datasets for cross validation and mathematical expression of multi-interaction reactions will be optimized to stabilize the model behavior.

When the dox induction was withdrawn after day 10, pluripotency-related proteins including OSKM factors were observed to remain as up-regulated compared to the MEF stage. These results suggest that pluripotency was successfully maintained *in silico* after dox withdrawal. Therefore, our model could be used for further *in silico* experiments to investigate dynamic changes in mouse pluripotency.

Both deterministic and stochastic simulations offer a platform to research dynamic interactions. In the *in silico* experiments we performed using the model, we have shown that the model can mimic consistently the transfection efficiency and doxycycline induction applied in the experimental procedure. In Figure 4.11, we observed the transition between MEF and iPSC stages on hypothetically seeded 200 cells with 50% transfection efficiency. When the cells do not have OSKM expression plasmid (represented with $z = 1$) cells were stuck in the MEF state although they were provided with LIF. On the other hand, transfected cells exit the MEF state and reprogram to the iPSC state. Additionally, doxycycline induction levels were tested by stochastic simulations with different z parameter levels. This result shows that when the effect of transfection or doxycycline increases, cells are more likely to reach the iPSC state (Figure 4.12).

According to results of the small molecule inhibition (Figure 4.13), the expected effect on pluripotency were not observed because of the absence of the differentiation

stage of iPSCs to desired cell. These inhibitors sustain the stem cells in pluripotency stage. After induction of these inhibitors after 10 days to day 12 may not be proper time space for testing effect of the inhibitor reactions. When the integration of the iPSC differentiation dataset into the model, GSK3B and phosphorylated MEK 1/2 protein levels will be estimated much more precisely, then the inhibition effect of the small molecules will be tested more accurately.

On the other hand, individual effect of OSKM factors on the pluripotency were observed by testing the combination of OSKM factors (Figure 4.14). As a result, we observed that single gene of OSKM factors were not enough for iPSC transitions. On the dual OSKM factor inductions, cells were not observed as fully transformed into iPSC, same as shown by the experimental dataset. Oppositely, SKM and OSKM induction has passed through into iPSC stage successfully. Based on the reference threshold point determined using the experimental data, these results suggest that our model can mimic the results of the experimental procedures in the study by Velychko et. al (2019).

5.4. Conclusions

The results presented in the previous sections suggest that our model could be used in other *in silico* experiments and provide predictions to understand the complex processes of somatic cell reprogramming as a result of signaling network activity and crosstalk. In this thesis project, we focused on the construction of dynamic behavior of the signaling pathway during somatic cell reprogramming with a holistic approach by including 9 signaling pathways together in the same model. Molecular crosstalk of these pathways were not previously collected in one model in the literature. Therefore, our model fits a major gap in the kinetic model literature. Additionally, experimental costs could be decreased using our model predictions as preliminary data for experimental studies.

In quantitative kinetic modelling literature, most of the initial conditions are determined with experimental measurements or obtained from other models. Our RNA to protein conversion factor (RTP) approach was provided to obtain initial conditions with high throughput and data-driven manner.. Via implementation of RTP factor into

model, other experimental datasets can be integrated into model accurately for further analysis.

When the kinetic modelling of signaling pathways were investigated in the literature, most of the published models attempted integration of one or two signaling pathways together. For instance WNT and ERK pathways were studied to investigate progression of cancer [Kim et al., 2007]. Another study were investigated the mechanism of 4 signaling pathways, however the kinetic models were constructed independently from each other [Sivakumar et al., 2011]. Hence, our model provides a novel mathematical platform to understand complex behavior of pluripotency.

Further, our kinetic model was constructed in a structure both for deterministic and stochastic simulations. Most of models in the literature did not have this flexibility as stochastic simulations in COPASI is possible only if the model consists of solely irreversible reactions.

In the comparison of pluripotency and differentiation phenotypes, average level of pluripotency and differentiation related proteins were investigated. Most of the highly expressed proteins in our model are also used iPSC markers in the literature such as Nanog [Xu et al., 2008], Oct4, Sox2 and Klf4 [Storm et al., 2009].

Integration of small inhibitory molecules did not produce the expected results in simulation of our kinetic model. To overcome this limitation, different reaction mechanism were tested for inhibitory reactions. More extensive investigation will be tested with optimizing the mathematical expression and crosstalk reaction in the model and will be tested via integration of relevant experimental data.

As a future aspect, more molecular interactions will be included in the model such as transcription factor and target gene interactions, epigenetic alterations, much more complex protein association, and dimerization. To achieve more accurate model behavior, a more extensive parameter estimation investigation will be performed. Also, the kinetic behavior of individual signaling pathways that affect pluripotency and differentiation by developing more *in silico* experimental procedures will be tested to enhance our model consistency. Additionally, iPSC cell differentiation datasets will also be integrated with the kinetic model to mimic differentiation procedures to enhance the dynamic behavior of the iPSC mathematically. By integration of both somatic cells reprogramming of iPSC and differentiation will provide us a much more extensive *in silico* platform for further studies to understand the dynamic behavior of stem cells.

REFERENCES

- Alon U., (2006), "An Introduction to Systems Biology: Design Principles of Biological Circuits", Chapman & Hall/CRC.
- Barrett T., Wilhite S. E., Ledoux P., Evangelista C., Kim I. F., Tomashevsky M., Marshall K. A., Phillippy K. H., Sherman P. M., Holko M., Yefanov A., Lee H., Zhang N., Robertson C. L. Serova N., Davis S., Soboleva A., (2013), "NCBI GEO: archive for functional genomics data sets—update", *Nucleic Acids Research*, *41*(D1), D991–D995.
- Bechard M., Dalton S., (2009), "Subcellular localization of glycogen synthase kinase 3 β controls embryonic stem cell self-renewal", *Molecular and Cellular Biology*, *29*(8), 2092–2104.
- Burdon T., Chambers I., Stracey C., Niwa H., & Smith A., (1999), "Signaling Mechanisms Regulating Self-Renewal and Differentiation of Pluripotent Embryonic Stem Cells", *Cells Tissues Organs*, *165*(3–4), 131–143.
- Chagastelles P. C., Nardi N. B., (2011), "Biology of stem cells: an overview", *Kidney International Supplements*, *1*(3), 63–67.
- Chen X., Vega V. B., (2008), "Transcriptional regulatory networks in embryonic stem cells", *Cold Spring Harbor Symposia on Quantitative Biology*, *73*, 203–209.
- De Los Angeles A., Ferrari F., Xi R., Fujiwara Y., Benvenisty N., Deng H., Hochedlinger K., Jaenisch R., Lee S., Leitch H. G., Lensch M. W., Lujan E., Pei D., Rossant J., Wernig M., Park P. J., Daley G. Q., (2015), "Hallmarks of pluripotency", *Nature*, *525*(7570), 469–478.
- Edfors F., Danielsson F., Hallström B. M., Käll L., Lundberg E., Pontén F., Forsström B., Uhlén M., (2016), "Gene-specific correlation of RNA and protein levels in human cells and tissues", *Molecular Systems Biology*, *12*(10), 883.
- El Barky A., Ali E., Mohamed T., (2017), "Stem Cells, Classifications and their Clinical Applications", *1*, 001–007.
- Galvin K. E., Travis E. D., Yee D., Magnuson T., Vivian J. L., (2010), "Nodal Signaling Regulates the Bone Morphogenic Protein Pluripotency Pathway in Mouse Embryonic Stem Cells", *The Journal of Biological Chemistry*, *285*(26), 19747–19756.
- Ghyselinck N. B., Duyster G., (2019), "Retinoic acid signaling pathways", *Development*, *146*(dev167502).
- Graf U., Casanova E. A., Cinelli P., (2011), "The Role of the Leukemia Inhibitory Factor (LIF) — Pathway in Derivation and Maintenance of Murine Pluripotent Stem Cells", *Genes*, *2*(1), 280–297.
- Hansson E. M., Lendahl U., Chapman G., (2004), "Notch signaling in development and disease", *Seminars in Cancer Biology*, *14*(5), 320–328.

- Hassani S.-N., Totonchi M., Gourabi H., Schöler H. R., Baharvand H., (2013), "Signaling Roadmap Modulating Naive and Primed Pluripotency", *Stem Cells and Development*, 23(3), 193–208.
- Hochedlinger K., Plath K., (2009), "Epigenetic reprogramming and induced pluripotency", *Development*, 136(4), 509–523.
- Hooper J. E., Scott M. P., (2005), "Communicating with Hedgehogs", *Nature Reviews Molecular Cell Biology*, 6(4), 306–317.
- Hoops S., Sahle S., Gauges R., Lee C., Pahle J., Simus N., Singhal M., Xu, L., Mendes P., Kummer U., (2006), "COPASI—a COMplex PATHway SIMulator", *Bioinformatics*, 22(24), 3067–3074.
- Kalra K., Tomar P. C., (2014), "Stem Cell: Basics, Classification and Applications", *American Journal of Phytomedicine and Clinical Therapeutics*, 2(7).
- Kanehisa M., Goto S., (2000), "KEGG: kyoto encyclopedia of genes and genomes", *Nucleic Acids Research*, 28(1), 27–30.
- Kanehisa Minoru Furumichi M., Sato Y., Ishiguro-Watanabe M., Tanabe M., (2021), "KEGG: integrating viruses and cellular organisms", *Nucleic Acids Research*, 49(D1), D545–D551.
- Kim D., Rath O., Kolch W., Cho K.-H., (2007), "A hidden oncogenic positive feedback loop caused by crosstalk between Wnt and ERK pathways", *Oncogene*, 26(31), 4571–4579.
- Kolios G., Moodley Y., (2013), "Introduction to Stem Cells and Regenerative Medicine", *Respiration*, 85(1), 3–10.
- Konstantinov I. E., (2000), "In Search of Alexander A. Maximow: The Man Behind the Unitarian Theory of Hematopoiesis", *Perspectives in Biology and Medicine*, 43(2), 269–276.
- Le Novère N., (2015), "Quantitative and logic modelling of molecular and gene networks", *Nature Reviews. Genetics*, 16(3), 146–158.
- Li Z., Fei T., Zhang J., Zhu G., Wang L., Lu D., Chi X., Teng Y., Hou N., Yang X., Zhang H., Han J. D. J., Chen, Y. G., (2012), "BMP4 Signaling Acts via dual-specificity phosphatase 9 to control ERK activity in mouse embryonic stem cells", *Cell Stem Cell*, 10(2), 171–182.
- Lowell S., Benchoua A., Heavey B., Smith A. G., (2006), "Notch Promotes Neural Lineage Entry by Pluripotent Embryonic Stem Cells", *PLOS Biology*, 4(5), e121.
- Lyashenko N., Winter M., Migliorini D., Biechele T., Moon R. T., Hartmann C., (2011), "Differential requirement for the dual functions of β -catenin in embryonic stem cell self-renewal and germ layer formation", *Nature Cell Biology*, 13(7), 753–761.
- Malik-Sheriff R. S., Glont M., Nguyen T. V. N., Tiwari K., Roberts M. G., Xavier A., Vu M. T., Men J., Maire M., Kananathan S., Fairbanks E. L., Meyer J. P., Arankalle

C., Varusai T. M., Knight-Schrijver V., Li L., Dueñas-Roca C., Dass G., Keating S. M., Hermjakob H., (2020), "BioModels—15 years of sharing computational models in life science", *Nucleic Acids Research*, 48(D1), D407–D415.

Massagué J., Seoane J., Wotton D., (2005), "Smad transcription factors", *Genes and Development*, 19(23), 2783–2810.

Matsuda T., Nakamura T., Nakao K., Arai T., Katsuki M., Heike T., Yokota T., (1999), "STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells", *The EMBO Journal*, 18(15), 4261–4269.

Medvedev S. P., Shevchenko A. I., Zakian S. M., (2010), "Induced Pluripotent Stem Cells: Problems and Advantages when Applying them in Regenerative Medicine", *Acta Naturae*, 2(2), 18–28.

Nicklas D., Saiz L., (2013), "Characterization of Negative Feedback Network Motifs in the TGF- β Signaling Pathway", *PLOS ONE*, 8(12), e83531.

Niwa H., Ogawa K., Shimosato D., Adachi K., (2009), "A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells", *Nature*, 460(7251), 118–122.

Nolte C., De Kumar B., Krumlauf R., (2019), "Hox genes: Downstream “effectors” of retinoic acid signaling in vertebrate embryogenesis", *Genesis*, 57(7–8), e23306.

Okita K., Yamanaka S., (2006), "Intracellular Signaling Pathways Regulating Pluripotency of Embryonic Stem Cells", *Current Stem Cell Research and Therapy*, 1(1), 103–111.

Padala R. R., Karnawat R., Viswanathan S. B., Thakkar A. V., Das A. B., (2017), "Cancerous perturbations within the ERK, PI3K/Akt, and Wnt/ β -catenin signaling network constitutively activate inter-pathway positive feedback loops", *Molecular BioSystems*, 13(5), 830–840.

Park K.-S., (2011), "TGF-beta Family Signaling in Embryonic Stem Cells", *International Journal of Stem Cells*, 4(1), 18–23.

Pir P., Le Novère N., (2016), "Mathematical Models of Pluripotent Stem Cells: At the Dawn of Predictive Regenerative Medicine", *Methods in Molecular Biology*, 1386, 331–350.

Plews J. R., Wu J. C., (2013), "Pluripotent Stem Cells", *Brenner’s Encyclopedia of Genetics (Second Edition)*

Rathjen J., Rathjen P., (2013), "Embryonic Stem Cells", *Brenner’s Encyclopedia of Genetics (Second Edition)*

Schwanhäusser B., Busse D., Li N., Dittmar G., Schuchhardt J., Wolf J., Chen W., Selbach M., (2011), "Global quantification of mammalian gene expression control", *Nature*, 473(7347), 337–342.

Sivakumar K. C., Dhanesh S. B., Shobana S., James J., Mundayoor S., (2011), "A systems biology approach to model neural stem cell regulation by notch, shh, wnt, and EGF signaling pathways", *Omics: A Journal of Integrative Biology*, 15(10), 729–737.

Smith A. G., (2001), "Embryo-derived stem cells: of mice and men", *Annual Review of Cell and Developmental Biology*, 17, 435–462.

Sobhani A., Khanlarkhani N., Baazm M., Mohammadzadeh F., Najafi A., Mehdinejadi S., Aval F. S., (2017), "Multipotent Stem Cell and Current Application", *Acta Medica Iranica*, 6–23.

Sokol S. Y., (2011), "Maintaining embryonic stem cell pluripotency with Wnt signaling", *Development*, 138(20), 4341–4350.

Sornberger J., (2014), "JAMES TILL AND ERNEST McCULLOCH: THE DISCOVERY OF STEM CELLS", In G. Thompson, *Pioneers of Medicine Without a Nobel Prize*, IMPERIAL COLLEGE PRESS.

Storm M. P., Kumpfmüller B., Thompson B., Kolde R., Vilo J., Hummel O., Schulz H., Welham M. J., (2009), "Characterization of the phosphoinositide 3-kinase-dependent transcriptome in murine embryonic stem cells: identification of novel regulators of pluripotency", *Stem Cells*, 27(4), 764–775.

Strasen J., Sarma U., Jentsch M., Bohn S., Sheng C., Horbelt D., Knaus P., Legewie S., Loewer A., (2018), "Cell-specific responses to the cytokine TGFβ are determined by variability in protein levels", *Molecular Systems Biology*, 14(1), e7733.

Sugimoto N., Eto K., (2021), "Generation and manipulation of human iPSC-derived platelets", *Cellular and Molecular Life Sciences*, 78(7), 3385–3401.

Takahashi K., Tanabe K., Ohnuki M., Narita M., Ichisaka T., Tomoda K., Yamanaka S., (2007), "Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors", *Cell*, 131(5), 861–872.

Takahashi K., Yamanaka S., (2006), "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors", *Cell*, 126(4), 663–676.

Takahashi K., Yamanaka S., (2016), "A decade of transcription factor-mediated reprogramming to pluripotency", *Nature Reviews Molecular Cell Biology*, 17(3), 183–193.

Teshigawara R., Cho J., Kameda M., Tada T., (2017), "Mechanism of human somatic reprogramming to iPS cell", *Laboratory Investigation*, 97(10), 1152–1157.

Tsuneyoshi N., Dunn N. R., (2013), "Guards at the gate to embryonic stem cell differentiation", *Cell*, 153(2), 281–283.

Velychko S., Adachi K., Kim K.-P., Hou Y., MacCarthy C. M., Wu G., Schöler H. R., (2019), "Excluding Oct4 from Yamanaka Cocktail Unleashes the Developmental Potential of iPSCs", *Cell Stem Cell*, 25(6), 737-753.e4.

Weinberger L., Ayyash M., Novershtern N., Hanna J. H., (2016), "Dynamic stem cell states: naive to primed pluripotency in rodents and humans", *Nature Reviews Molecular Cell Biology*, 17(3), 155–169.

Welham M. J., Storm M. P., Kingham E., Bone H. K., (2007), "Phosphoinositide 3-kinases and regulation of embryonic stem cell fate", *Biochemical Society Transactions*, 35(Pt 2), 225–228.

Widłak W., (2013), "Cell Signaling Pathways", *Molecular Biology: Not Only for Bioinformaticians*, Springer.

Wray J., Kalkan T., Gomez-Lopez S., Eckardt D., Cook A., Kemler R., Smith A., (2011), "Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation", *Nature Cell Biology*, 13(7), 838–845.

Wu S. M., Choo A. B. H., Yap M. G. S., Chan K. K., (2010), "Role of Sonic hedgehog signaling and the expression of its components in human embryonic stem cells", *Stem Cell Research*, 4(1), 38–49.

Xu R. H., Sampsel-Barron T. L., Gu F., Root S., Peck R. M., Pan G., Yu J., Antosiewicz-Bourget J., Tian S., Stewart R., Thomson J. A., (2008), "NANOG is a direct target of TGFbeta/activin-mediated SMAD signaling in human ESCs", *Cell Stem Cell*, 3(2), 196–206.

Ye L., Swingen C., Zhang J., (2013), "Induced Pluripotent Stem Cells and Their Potential for Basic and Clinical Sciences", *Current Cardiology Reviews*, 9(1), 63–72.

Yilmaz A., Benvenisty N., (2019), "Defining Human Pluripotency", *Cell Stem Cell*, 25(1), 9–22.

Ying Q. L., Nichols J., Chambers I., Smith A., (2003), "BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3", *Cell*, 115(3), 281–292.

BIOGRAPHY

Simge Şengül Babal completed her undergraduate studies in the Department of Molecular Biology and Genetics at Üsküdar University between 2013-2018. In 2018, she started to study for her master's degree in Molecular Biology and Genetics Programme at Gebze Technical University. In 2019, she did an internship in Yesil Science Company as a developer in İstanbul, TR. During her master's education, she participated in a stem cell research project funded by TUBITAK, project number 116S388.

APPENDIX

Appendix A: Supplementary Files

Supplementary files mentioned in the thesis are provided in the electronic format and listed below:

- Supplementary File 1: Model Reactions and Kinetics
- Supplementary File 2: Ordinary Differential Equations of the Kinetic Models.
- Supplementary File 3: Initial Conditions and Data for Parameter Estimation
- Supplementary File 4: Parameter Estimations Summary Table
- Supplementary File 5: Raw Protein Copy Number Dataset for Parameter Estimation